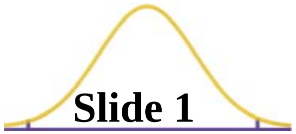
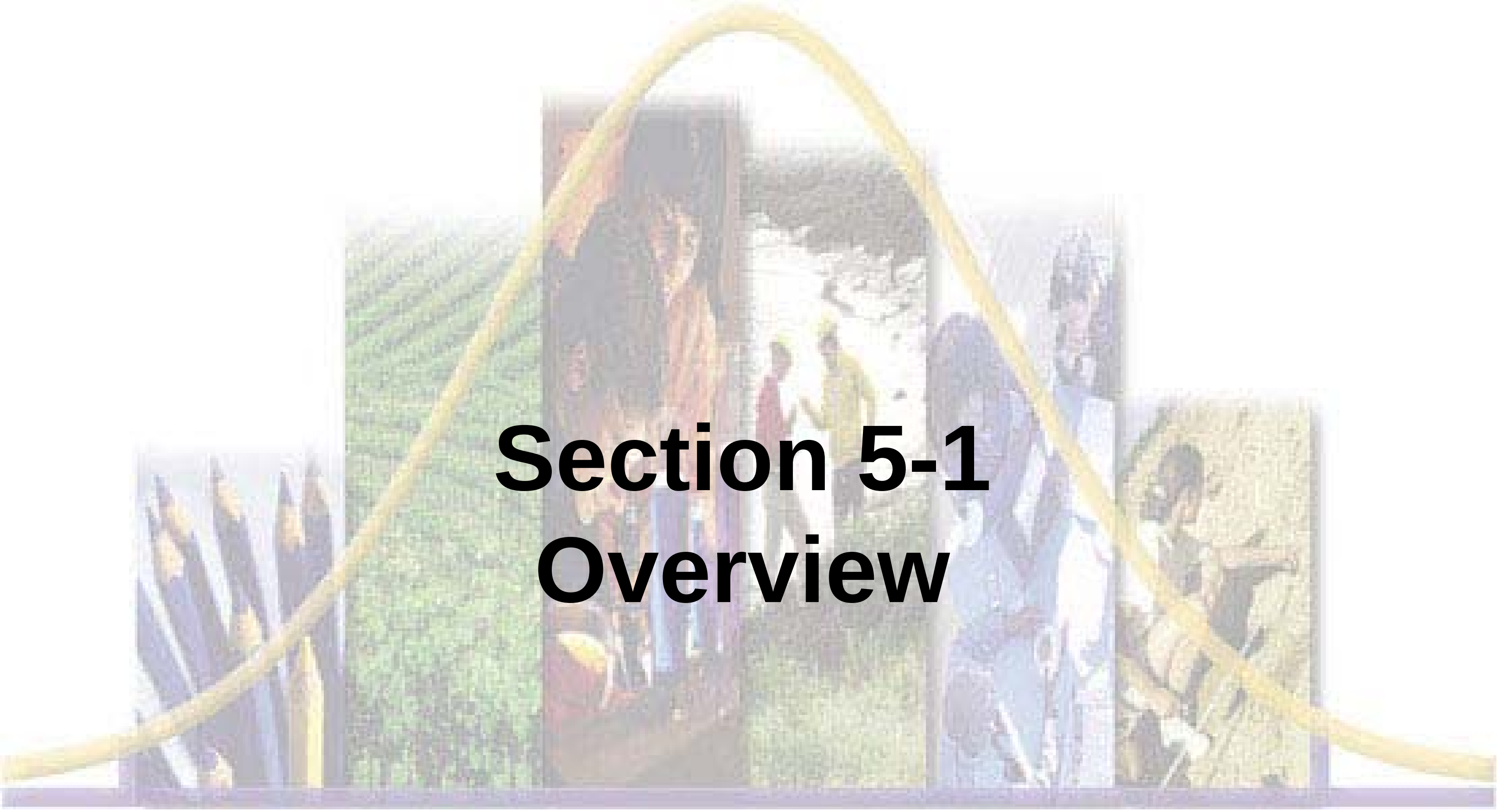
Chapter 5
Normal Probability Distributions





Section 5-1 Overview

Created by Erin Hodgess, Houston, Texas

Overview

- **Continuous random variable**
- **Normal distribution**

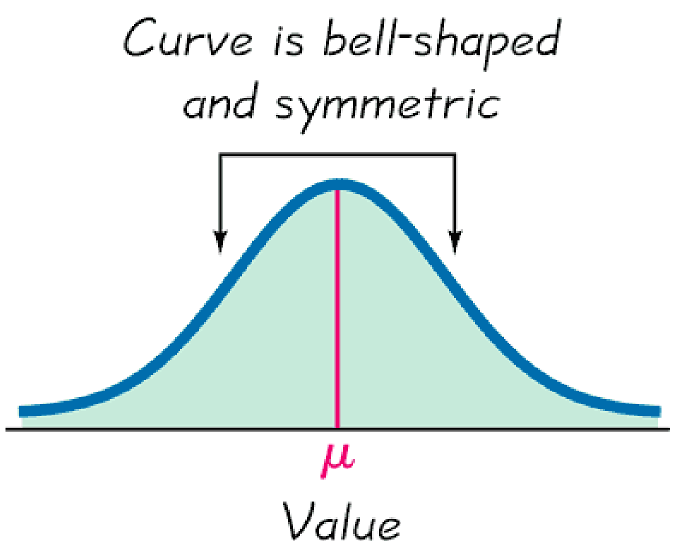
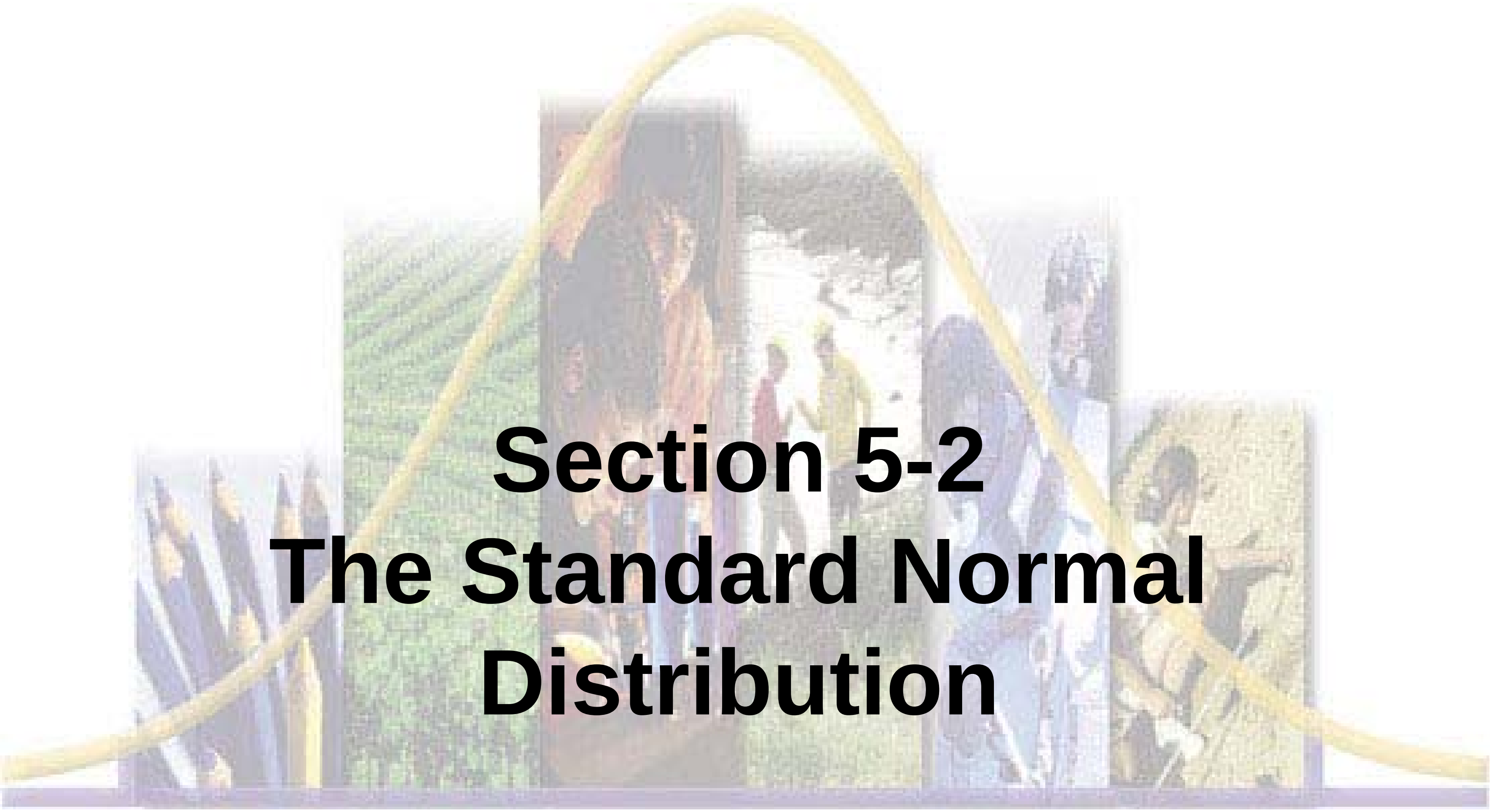


Figure 5-1 (p.226)

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

Formula 5-1



Section 5-2
The Standard Normal Distribution

Definitions

- **Uniform Distribution is a probability distribution in which the continuous random variable values are spread evenly over the range of possibilities; the graph results in a rectangular shape.**

Definitions (p.228)

- **Density Curve** (or probability density function is the graph of a continuous probability distribution.

1. The total area under the curve must equal 1.

2. Every point on the curve must have a vertical height that is 0 or greater.

Because the total area under the density curve is equal to 1, there is a correspondence between area and probability.

Using Area to Find Probability

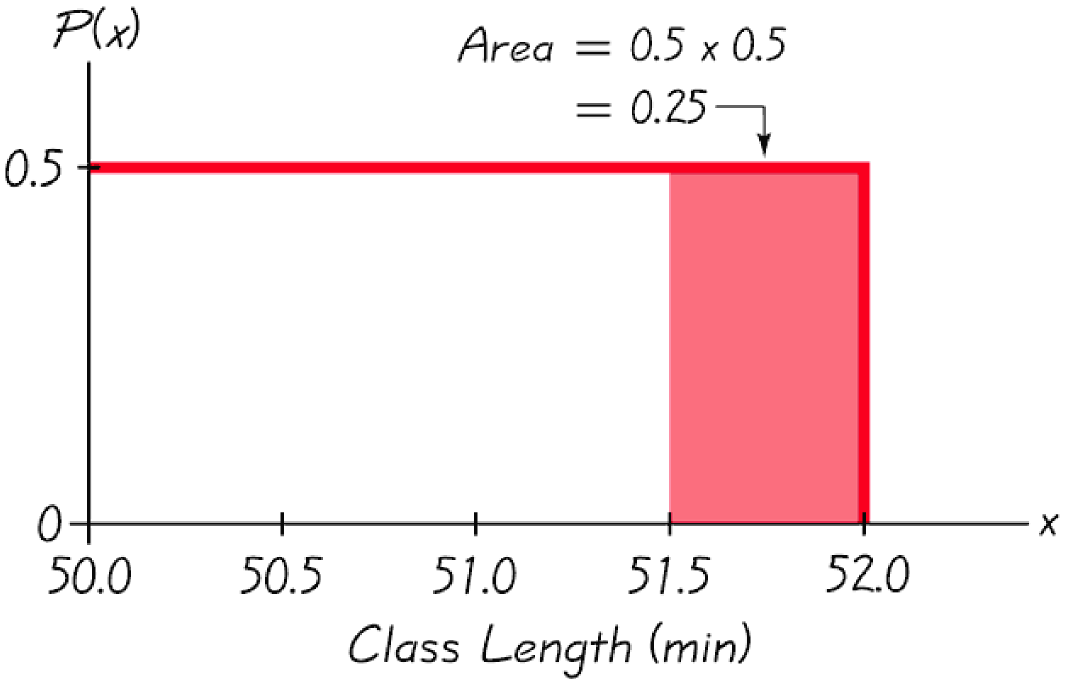


Figure 5-3 (p.228)

Heights of Adult Men and Women

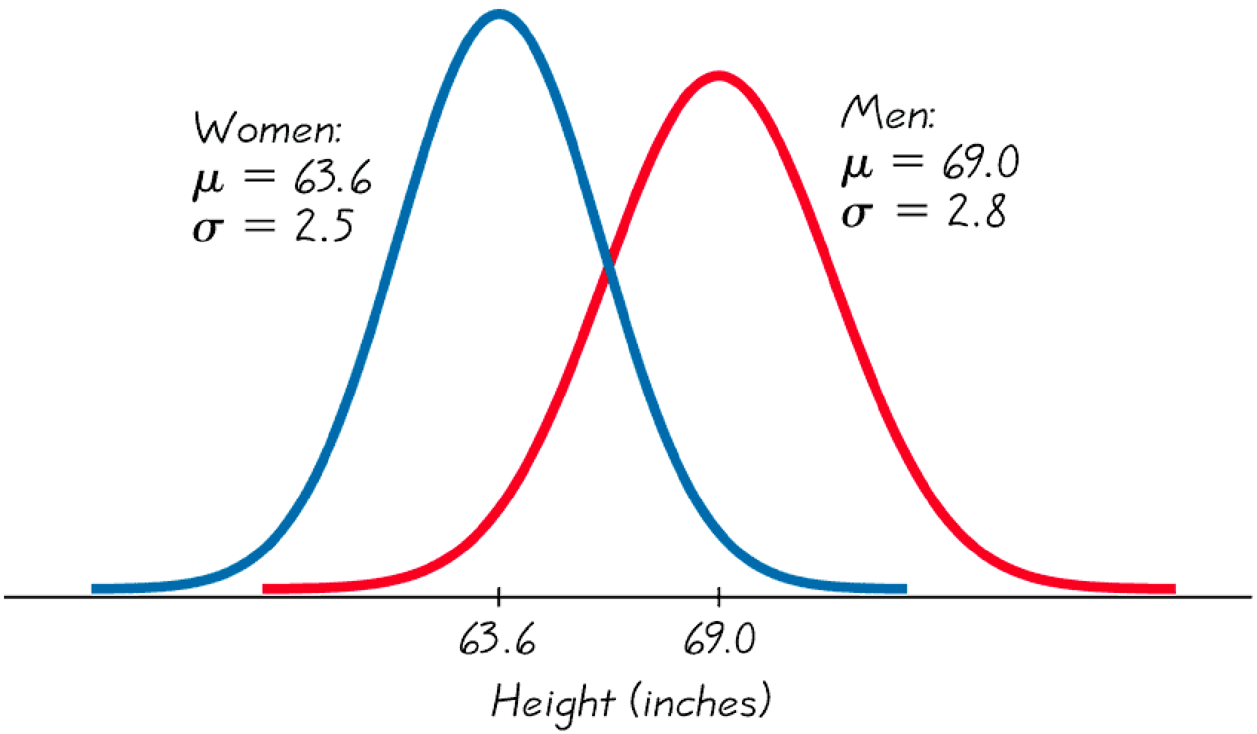


Figure 5-4 (p.229)

Definition

Standard Normal Distribution:

a normal probability distribution that has a mean of 0 and a standard deviation of 1.

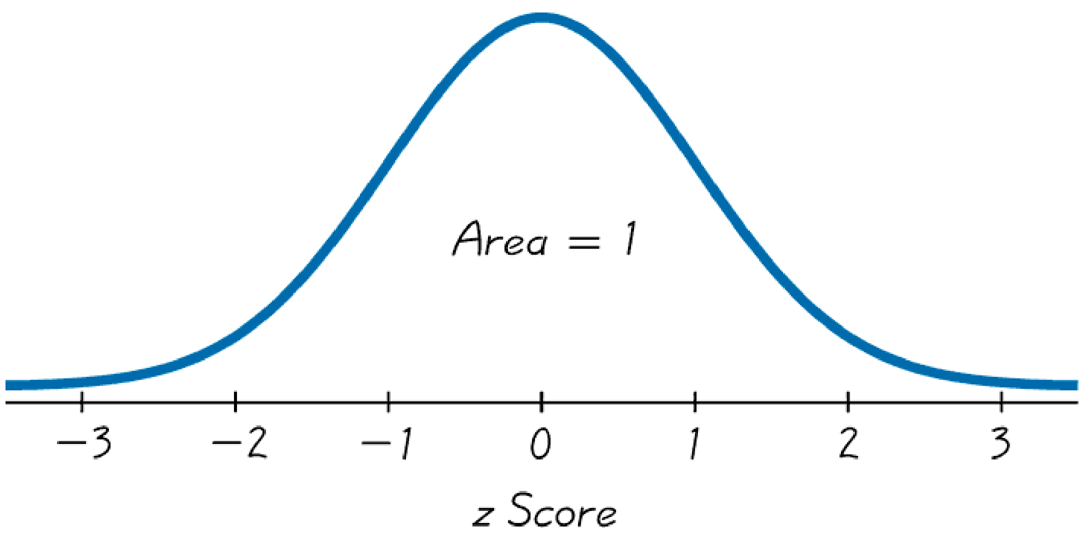


Figure 5-5 (p.231)

Table A-2

- Inside front cover of text book
- Formulas and Tables Card
- Appendix (p.734)

TABLE A-2 Standard Normal (z) Distribution: Cumulative Area from the LEFT

z	.00	.01	.02	.03	.04	.05	.06	.07	.08	.09
−3.50 and lower	.0001									
−3.4	.0003	.0003	.0003	.0003	.0003	.0003	.0003	.0003	.0003	.0002
−3.3	.0005	.0005	.0005	.0004	.0004	.0004	.0004	.0004	.0004	.0003
−3.2	.0007	.0007	.0006	.0006	.0006	.0006	.0006	.0005	.0005	.0005
−3.1	.0010	.0009	.0009	.0009	.0008	.0008	.0008	.0008	.0007	.0007
−3.0	.0013	.0013	.0013	.0012	.0012	.0011	.0011	.0011	.0010	.0010
−2.9	.0019	.0018	.0018	.0017	.0016	.0016	.0015	.0015	.0014	.0014
−2.8	.0026	.0025	.0024	.0023	.0023	.0022	.0021	.0021	.0020	.0019
−2.7	.0035	.0034	.0033	.0032	.0031	.0030	.0029	.0028	.0027	.0026
−2.6	.0047	.0045	.0044	.0043	.0041	.0040	.0039	.0038	.0037	.0036
−2.5	.0062	.0060	.0059	.0057	.0055	.0054	.0052	.0051 *	.0049	.0048
−2.4	.0082	.0080	.0078	.0075	.0073	.0071	.0069	.0068	.0066	.0064
−2.3	.0107	.0104	.0102	.0099	.0096	.0094	.0091	.0089	.0087	.0084
−2.2	.0139	.0136	.0132	.0129	.0125	.0122	.0119	.0116	.0113	.0110
−2.1	.0179	.0174	.0170	.0166	.0162	.0158	.0154	.0150	.0146	.0143
−2.0	.0228	.0222	.0217	.0212	.0207	.0202	.0197	.0192	.0188	.0183
−1.9	.0287	.0281	.0274	.0268	.0262	.0256	.0250	.0244	.0239	.0233
−1.8	.0359	.0351	.0344	.0336	.0329	.0322	.0314	.0307	.0301	.0294
−1.7	.0446	.0436	.0427	.0418	.0409	.0401	.0392	.0384	.0375	.0367
−1.6	.0548	.0537	.0526	.0516	.0505 *	.0495	.0485	.0475	.0465	.0455
−1.5	.0668	.0655	.0643	.0630	.0618	.0606	.0594	.0582	.0571	.0559

To find:

z Score

__the distance along horizontal scale of the standard normal distribution; refer to the **leftmost column and top row** of Table A-2.

Area

__the region under the curve; refer to the values in the **body** of Table A-2.

Example (p.232): **If thermometers have an average (mean) reading of 0 degrees and a standard deviation of 1 degree for freezing water, and if one thermometer is randomly selected, find the probability that, at the freezing point of water, the reading is less than 1.58 degrees.**

P(z < 1.58) =

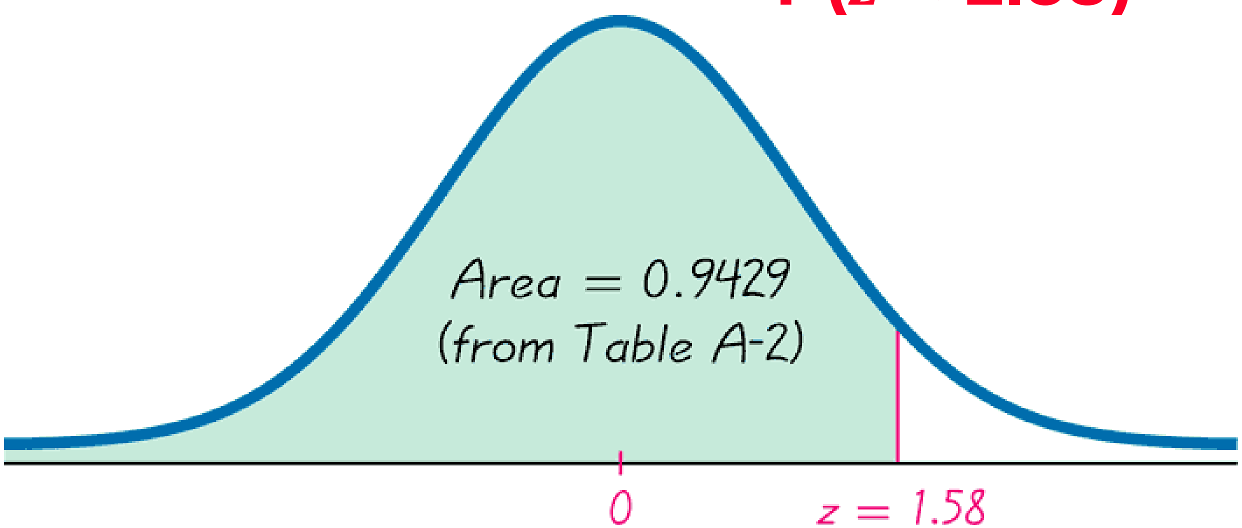


Figure 5-6 (p.232)

Example: **If thermometers have an average (mean) reading of 0 degrees and a standard deviation of 1 degree for freezing water and if one thermometer is randomly selected, find the probability that, at the freezing point of water, the reading is less than 1.58 degrees.**

$$P(z < 1.58) = 0.9429$$

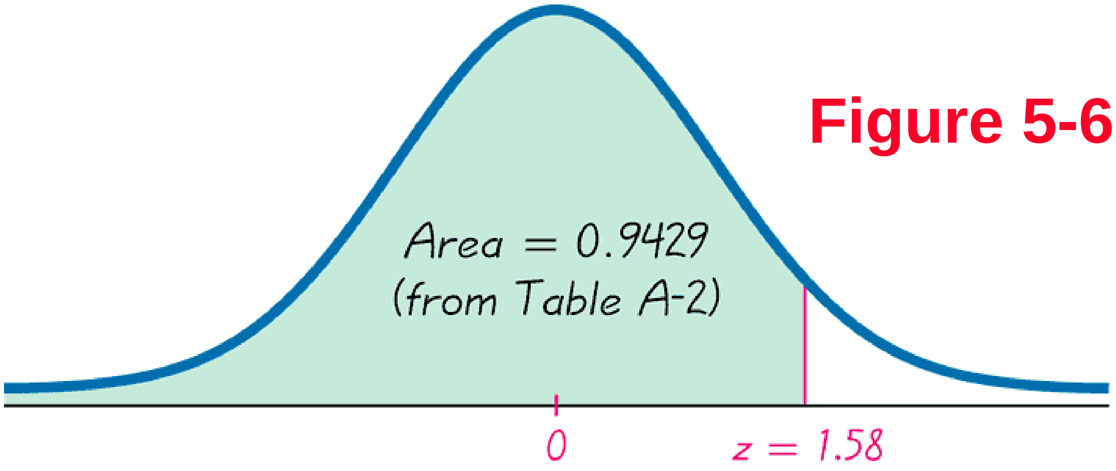
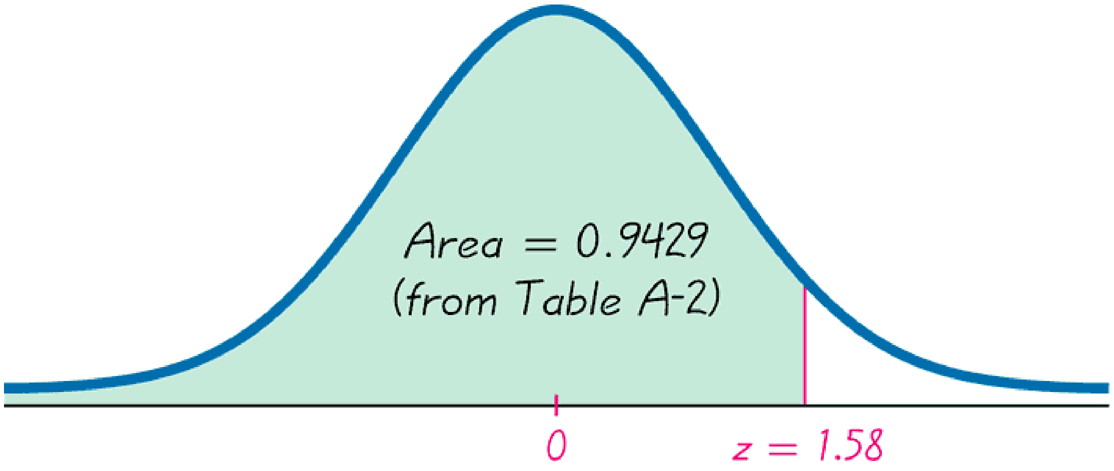


Figure 5-6

The probability that the chosen thermometer will measure freezing water less than 1.58 degrees is 0.9429.

Example: **If thermometers have an average (mean) reading of 0 degrees and a standard deviation of 1 degree for freezing water and if one thermometer is randomly selected, find the probability that, at the freezing point of water, the reading is less than 1.58 degrees.**

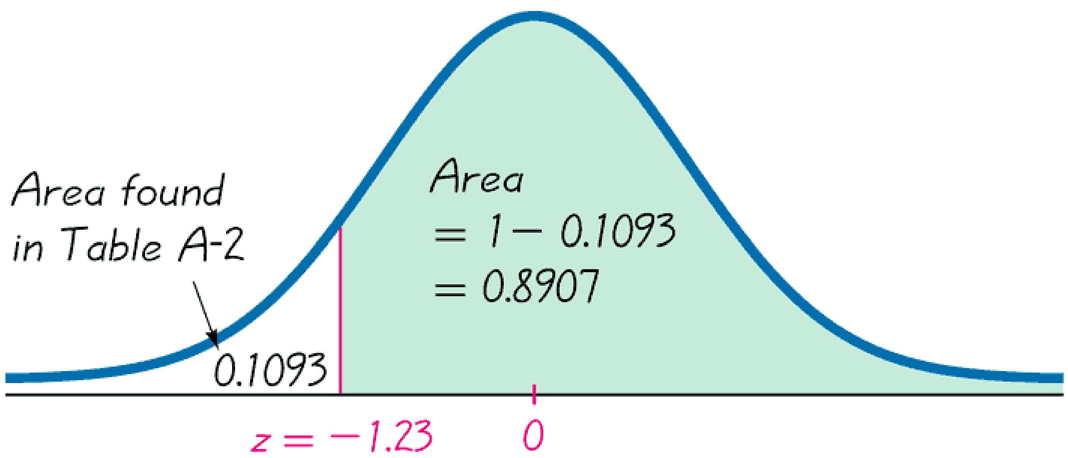
$$P\ (z < 1.58) = 0.9429$$



94.29% of the thermometers have readings less than 1.58 degrees.

Example: **If thermometers have an average (mean) reading of 0 degrees and a standard deviation of 1 degree for freezing water, and if one thermometer is randomly selected, find the probability that it reads (at the freezing point of water) above –1.23 degrees.**

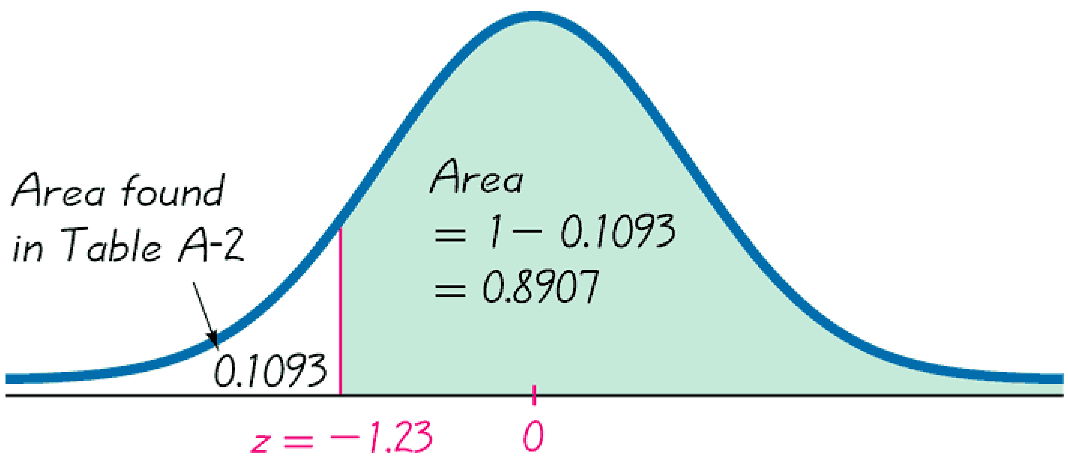
$P(z > -1.23) = 0.8907$



The probability that the chosen thermometer with a reading above –1.23 degrees is 0.8907.

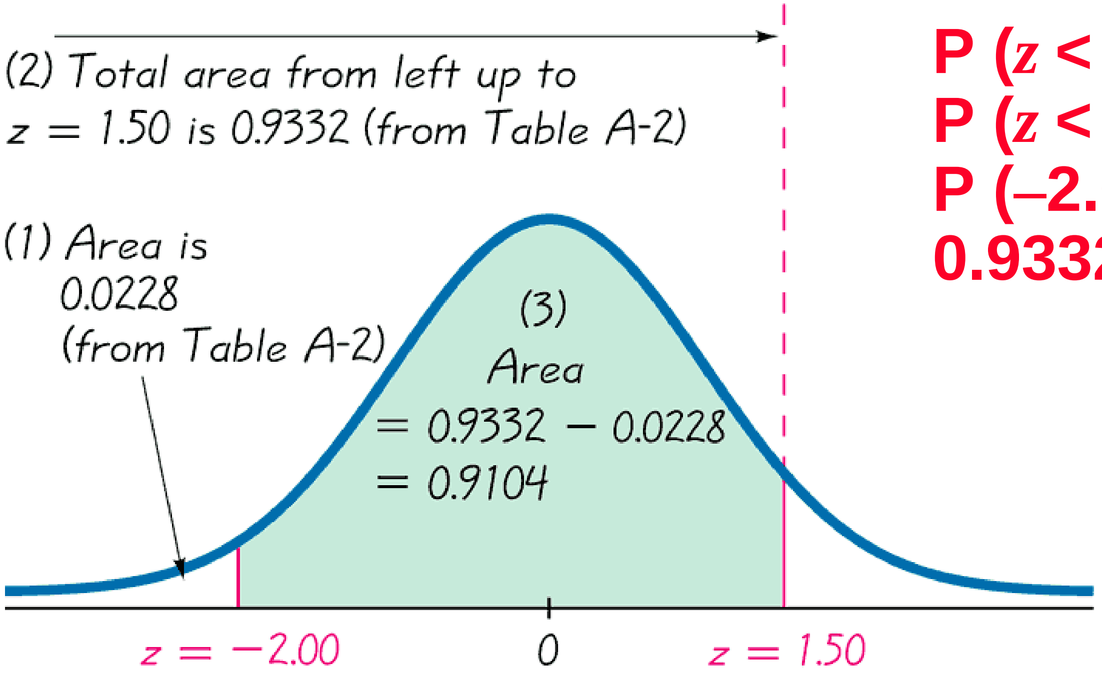
Example: **If thermometers have an average (mean) reading of 0 degrees and a standard deviation of 1 degree for freezing water, and if one thermometer is randomly selected, find the probability that it reads (at the freezing point of water) above –1.23 degrees.**

$$P(z > -1.23) = 0.8907$$



89.07% of the thermometers have readings above –1.23 degrees.

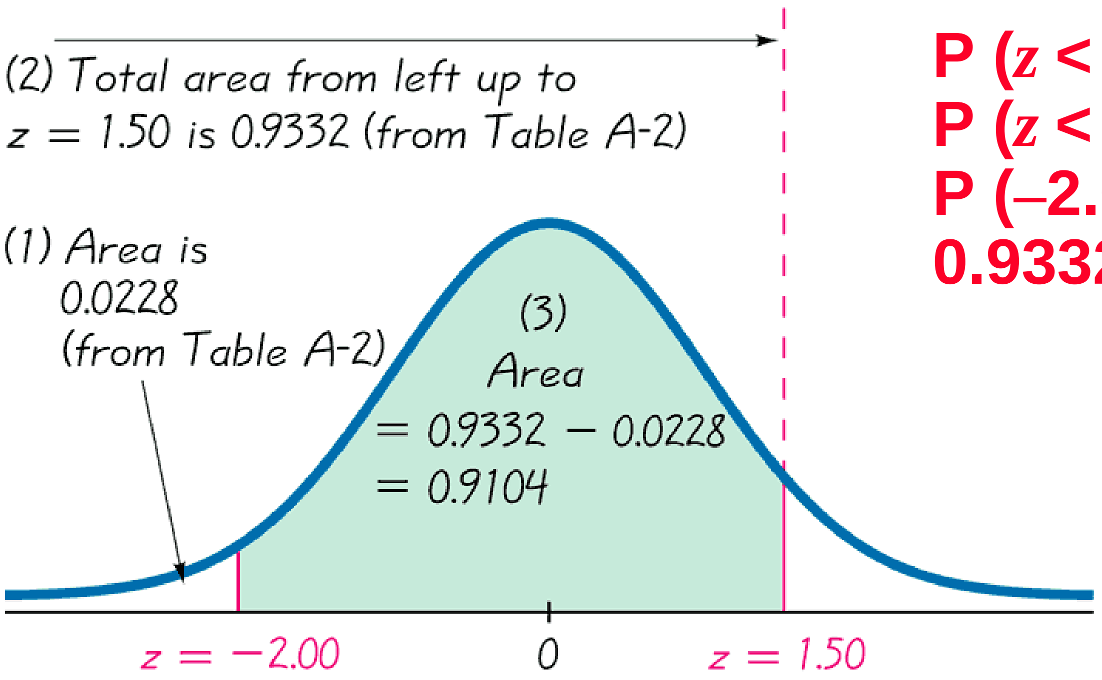
Example (p.233): A thermometer is randomly selected. Find the probability that it reads (at the freezing point of water) between –2.00 and 1.50 degrees.



$P(z < -2.00) = 0.0228$
$P(z < 1.50) = 0.9332$
$P(-2.00 < z < 1.50) =$
$0.9332 - 0.0228 = 0.9104$

The probability that the chosen thermometer has a reading between – 2.00 and 1.50 degrees is 0.9104.

Example: **A thermometer is randomly selected. Find the probability that it reads (at the freezing point of water) between –2.00 and 1.50 degrees.**



P (z < –2.00) = 0.0228
P (z < 1.50) = 0.9332
P (–2.00 < z < 1.50) =
0.9332 – 0.0228 = 0.9104

If many thermometers are selected and tested at the freezing point of water, then 91.04% of them will read between –2.00 and 1.50 degrees.

Notation (p.234)

$$P(a < z < b)$$

denotes the probability that the z score is between a and b

$$P(z > a)$$

denotes the probability that the z score is greater than a

$$P(z < a)$$

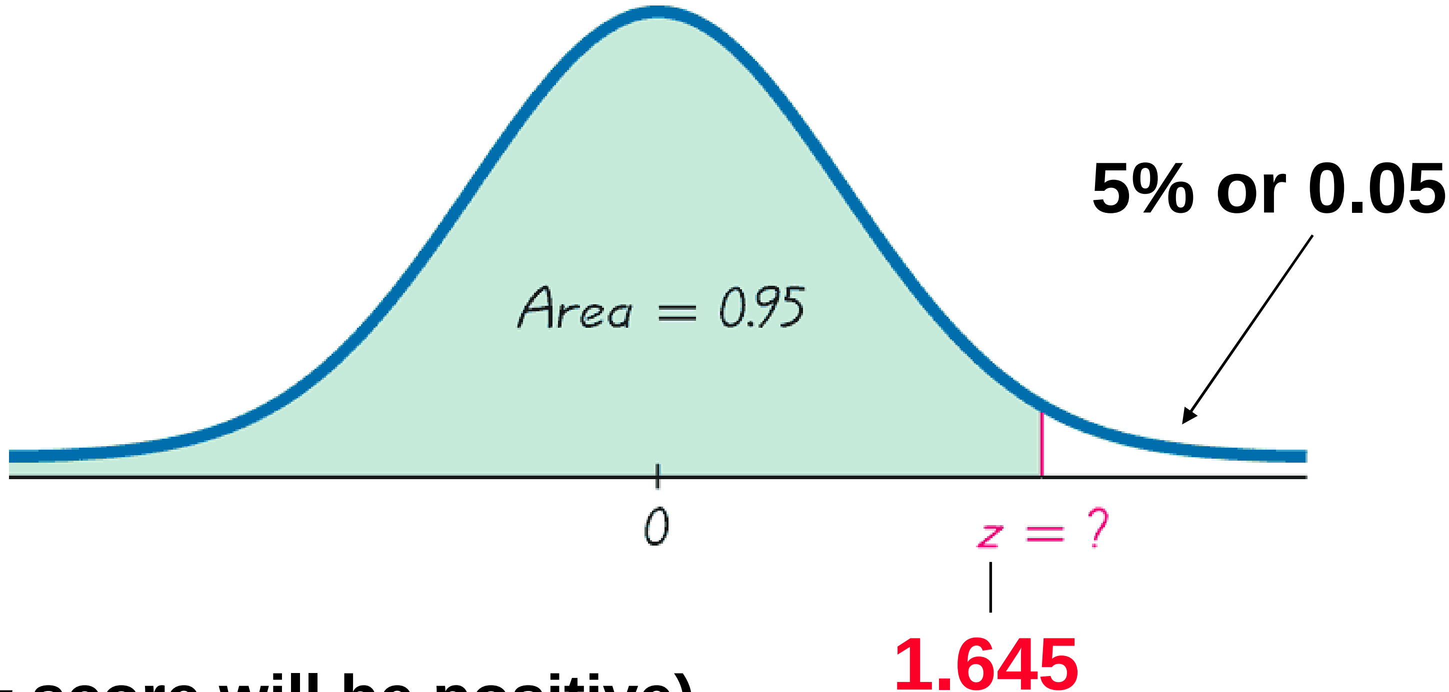
denotes the probability that the z score is less than a

Finding a *z* - score when given a probability Using Table A-2

1. **Draw a bell-shaped curve, draw the centerline, and identify the region under the curve that corresponds to the given probability. If that region is not a cumulative region from the left, work instead with a known region that is a cumulative region from the left.**

2. **Using the cumulative area from the left, locate the closest probability in the *body* of Table A-2 and identify the corresponding *z* score.**

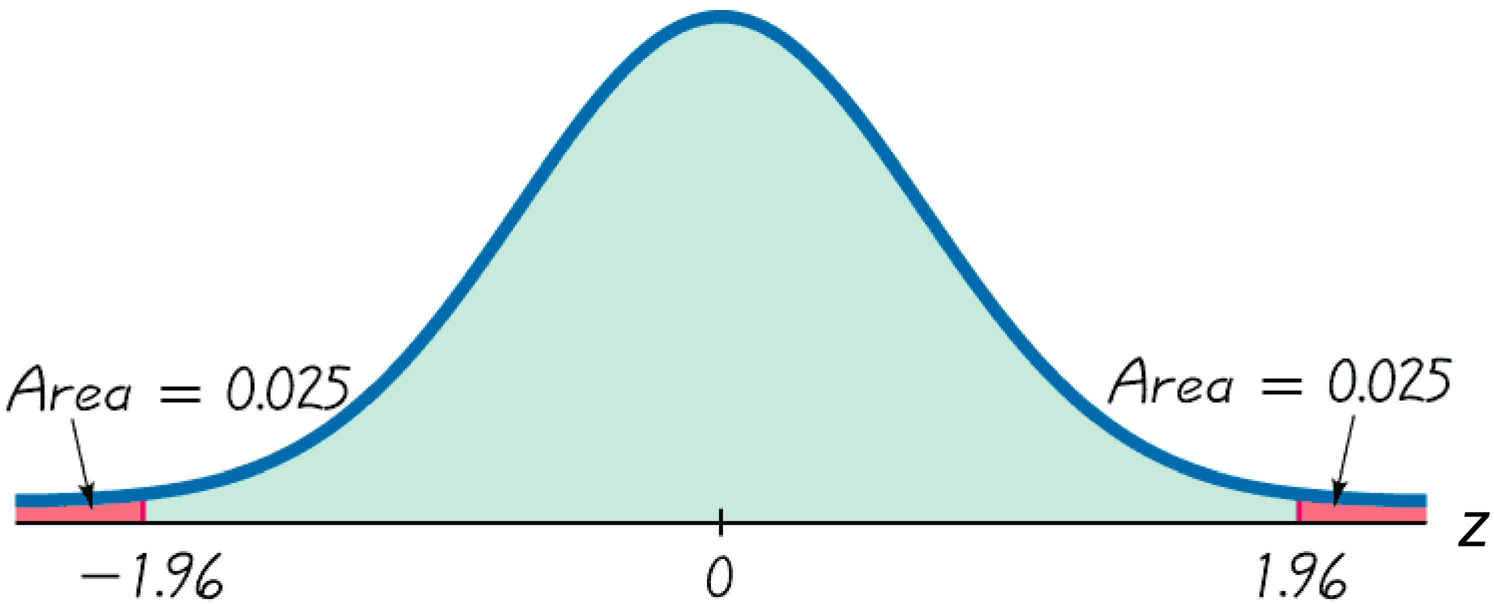
Finding *z* Scores when Given Probabilities



(*z* score will be positive)

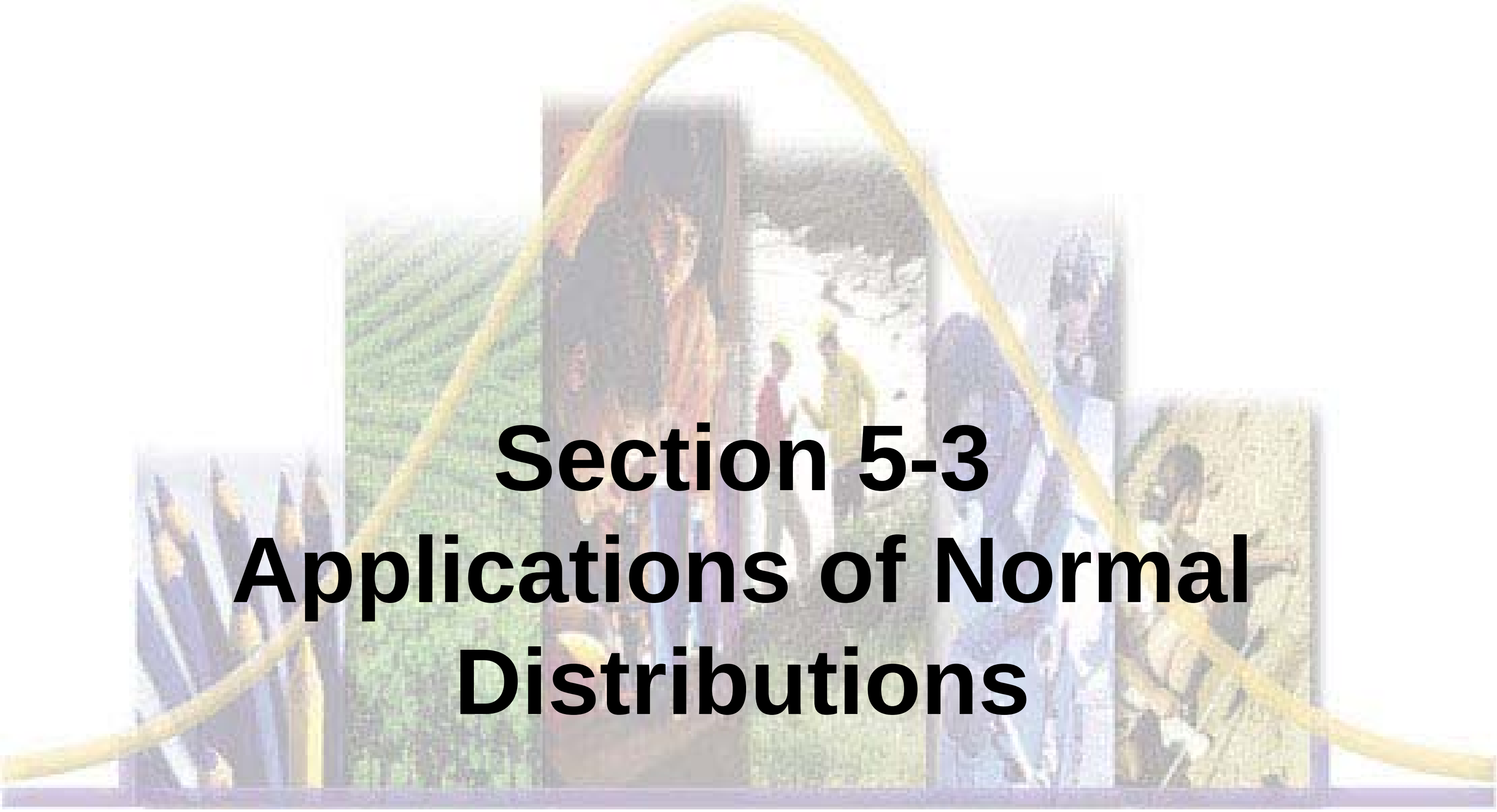
Figure 5-10 (p.236)
Finding the 95th Percentile

Finding *z* Scores when Given Probabilities



(One *z* score will be negative and the other positive)

Figure 5-11 (p.237)
Finding the Bottom 2.5% and Upper 2.5%



Section 5-3 Applications of Normal Distributions

Created by Erin Hodgess, Houston, Texas

Nonstandard Normal Distributions

If $\mu \neq 0$ or $\sigma \neq 1$ (or both), we will convert values to standard scores using Formula 5-2, then procedures for working with all normal distributions are the same as those for the standard normal distribution.

Formula 5-2

(p.240)

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

Converting to Standard Normal Distribution

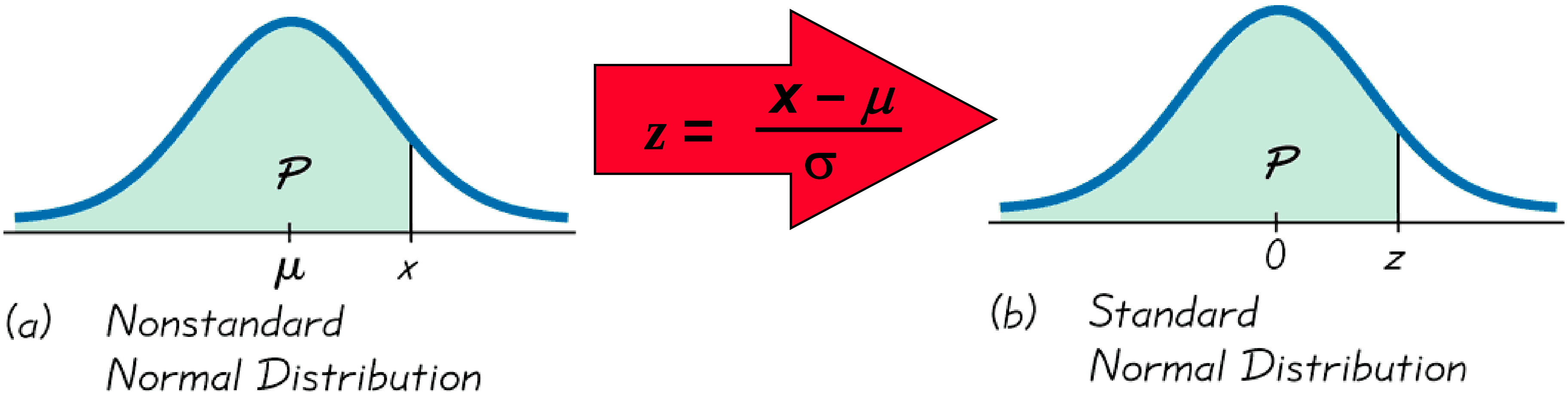


Figure 5-12 (p.240)

Probability of Sitting Heights Less Than 38.8 Inches

- The sitting height (from seat to top of head) of drivers must be considered in the design of a new car model. Men have sitting heights that are normally distributed with a mean of 36.0 in. and a standard deviation of 1.4 in. (based on anthropometric survey data from Gordon, Clauser, et al.). Engineers have provided plans that can accommodate men with sitting heights up to 38.8 in., but taller men cannot fit. If a man is randomly selected, find the probability that he has a sitting height less than 38.8 in. Based on that result, is the current engineering design feasible?

Probability of Sitting Heights Less Than 38.8 Inches

$\mu = 36.0$

$\sigma = 1.4$

$$z = \frac{38.8 - 36.0}{1.4} = 2.00$$

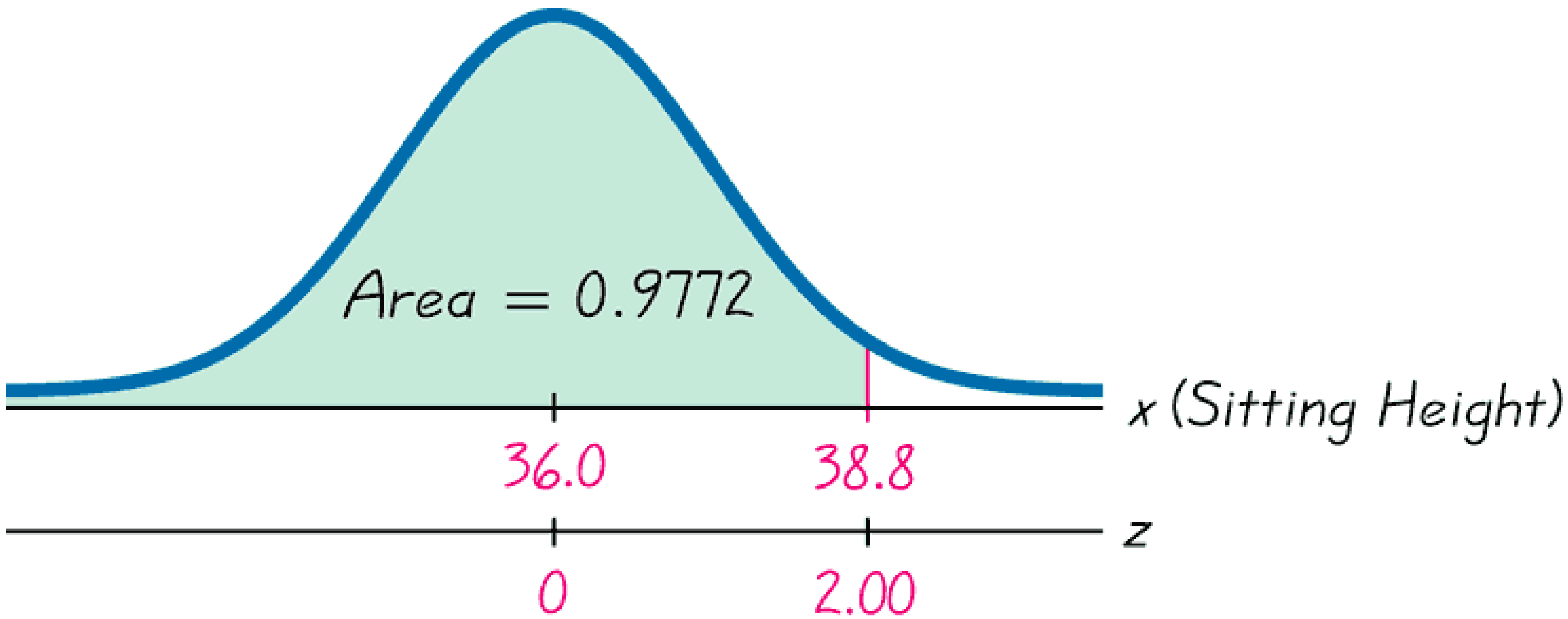


Figure 5-13 (p.241)

Probability of Sitting Heights Less Than 38.8 Inches

μ = 38.8

σ = 1.4

P (*x* < 38.8 in.) = P(*z* < 2)
= 0.9772

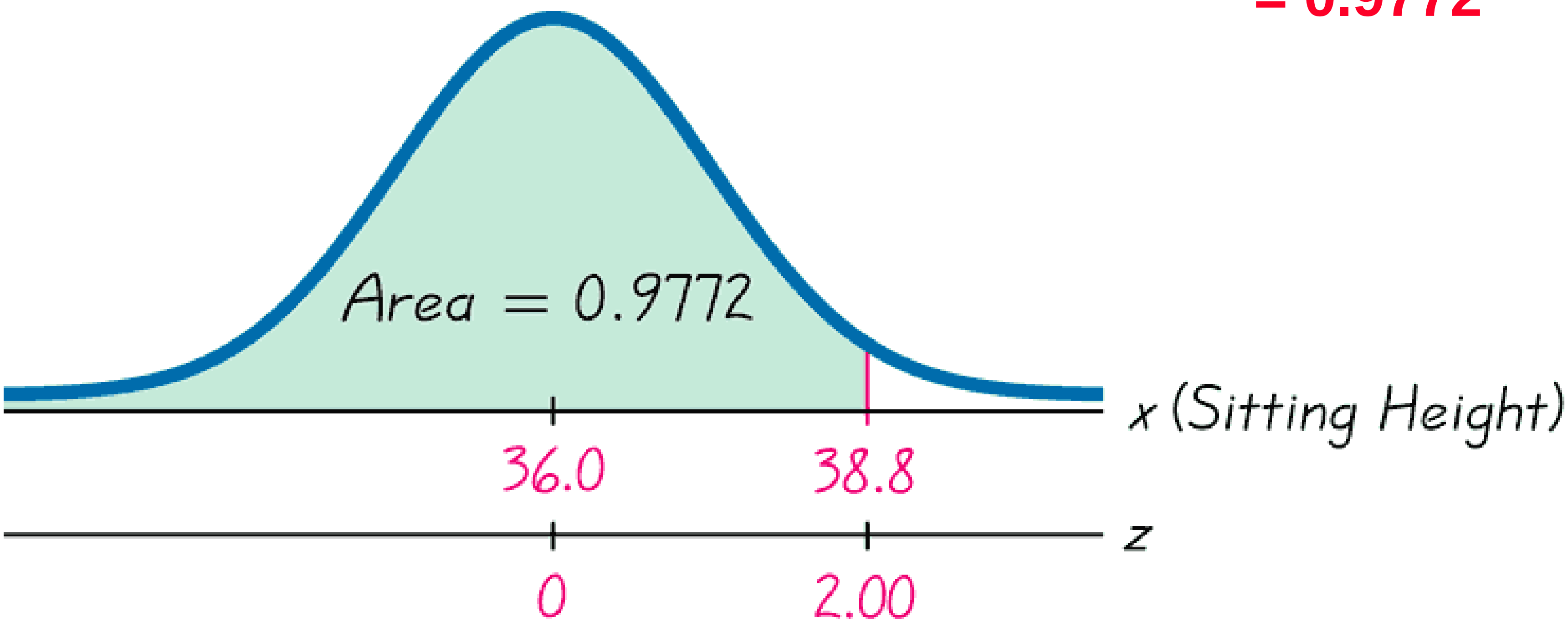


Figure 5-13 (p.241)

Probability of Weight between 140 pounds and 211 pounds

In the Chapter Problem, we noted that the Air Force had been using the ACES-II ejection seats designed for men weighing between 140 lb and 211 lb. Given that women's weights are normally distributed with a mean of 143 lb and a standard deviation of 29 lb (based on data from the National Health survey), what percentage of women have weights that are within those limits?

Probability of Weight between 140 pounds and 211 pounds

μ = 143
σ = 29

$$z = \frac{211 - 143}{29} = 2.34$$

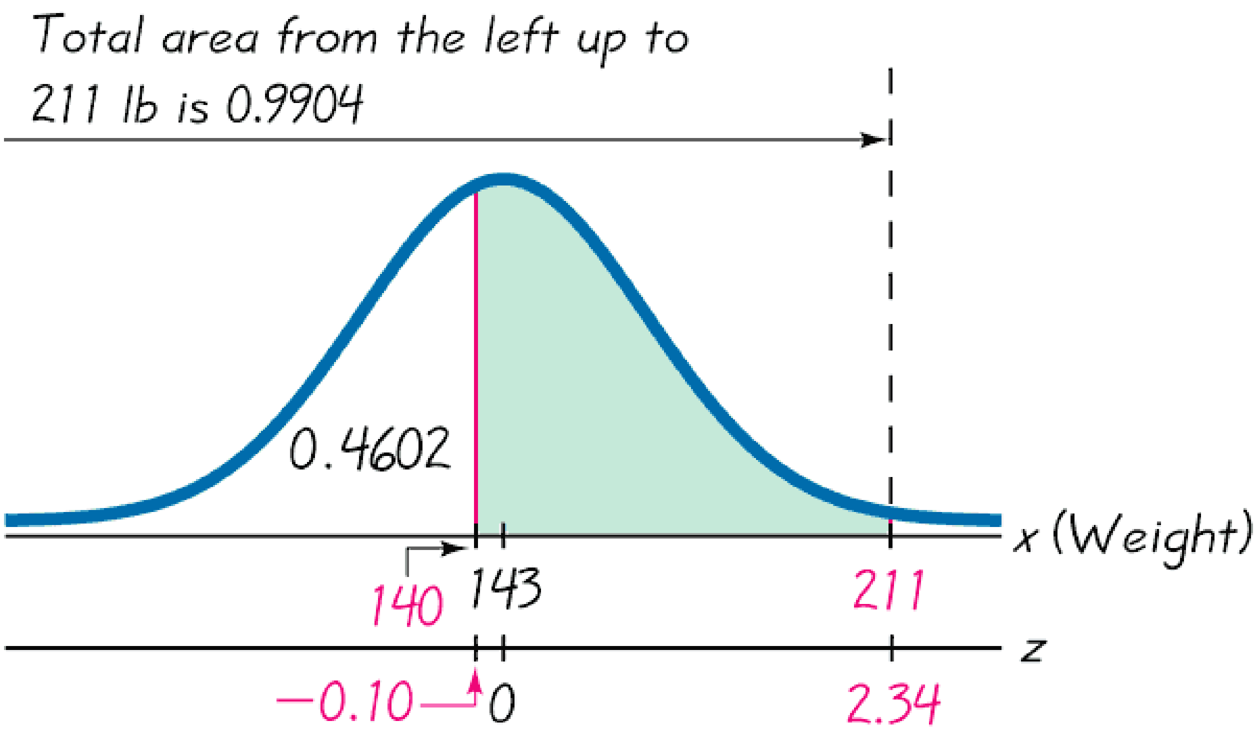


Figure 5-14 (p.242)

Probability of Weight between 140 pounds and 211 pounds

$\mu = 143$
$\sigma = 29$

$$z = \frac{140 - 143}{29} = -0.10$$

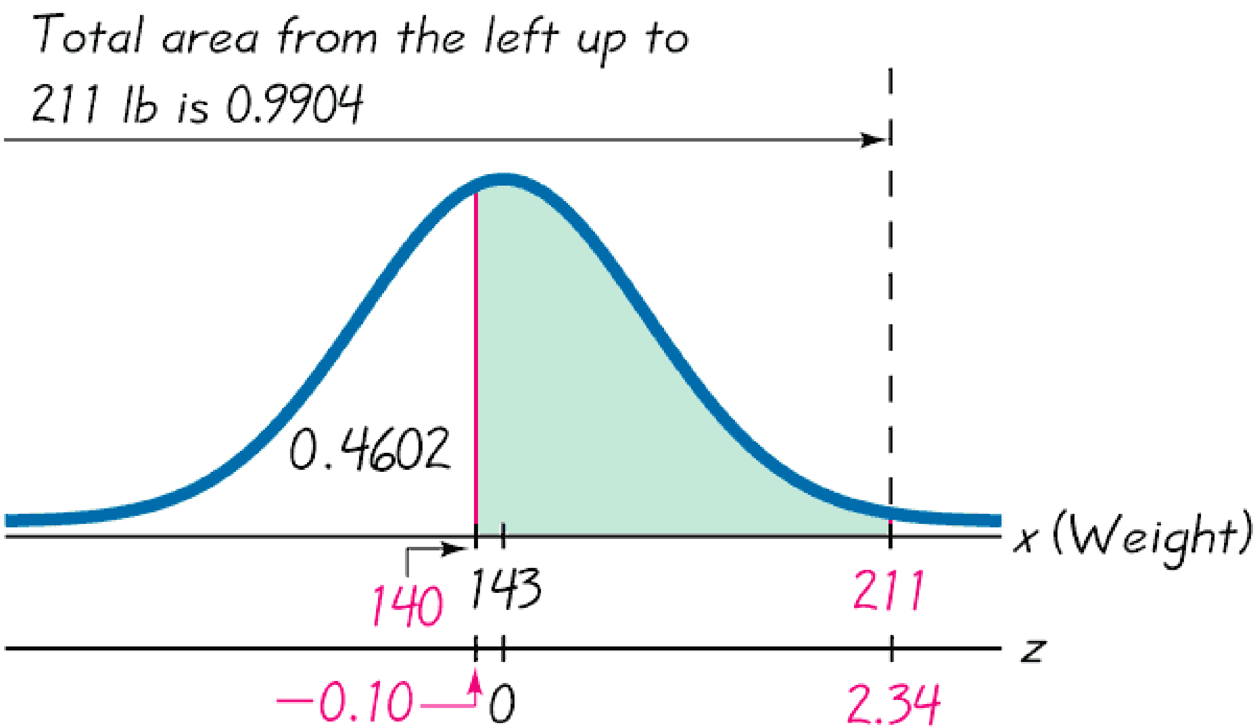


Figure 5-14 (p.242)

Probability of Weight between 140 pounds and 211 pounds

μ = 143
σ = 29

P(–0.10 < z < 2.34) =

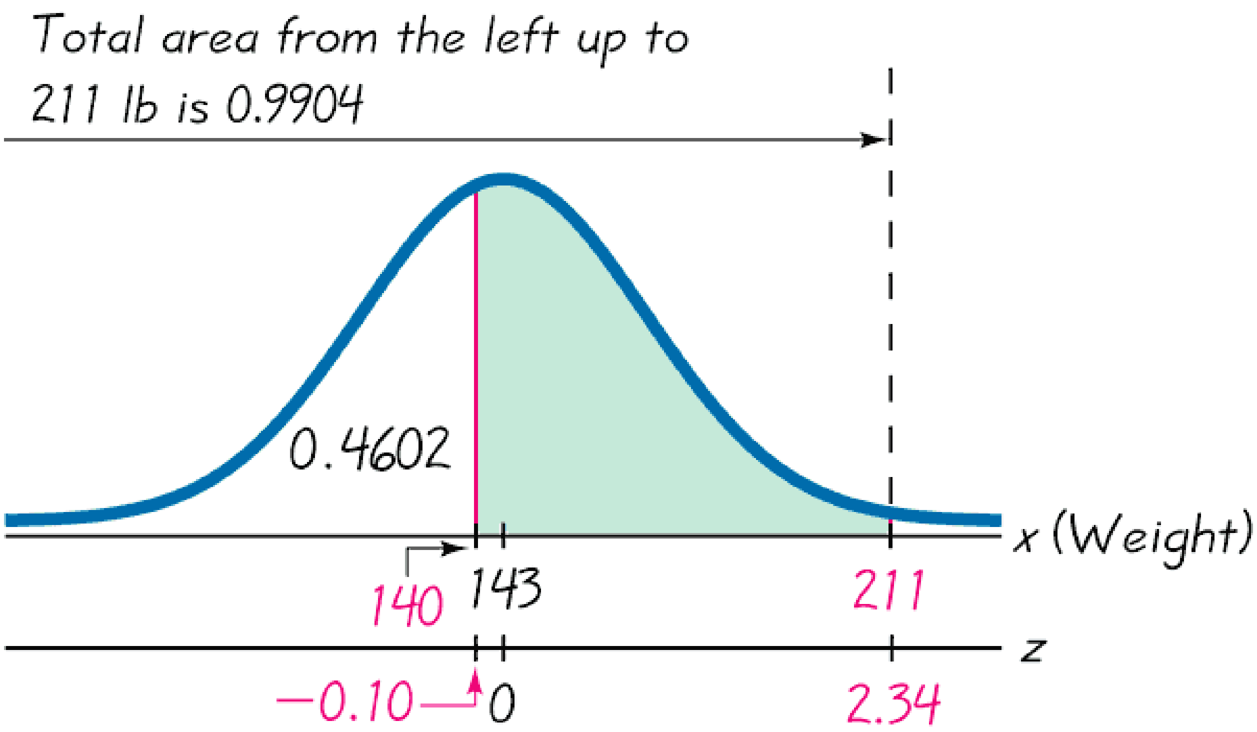


Figure 5-14 (p.242)

Probability of Weight between 140 pounds and 211 pounds

μ = 143
σ = 29

0.9904 – 0.4602 = 0.5302

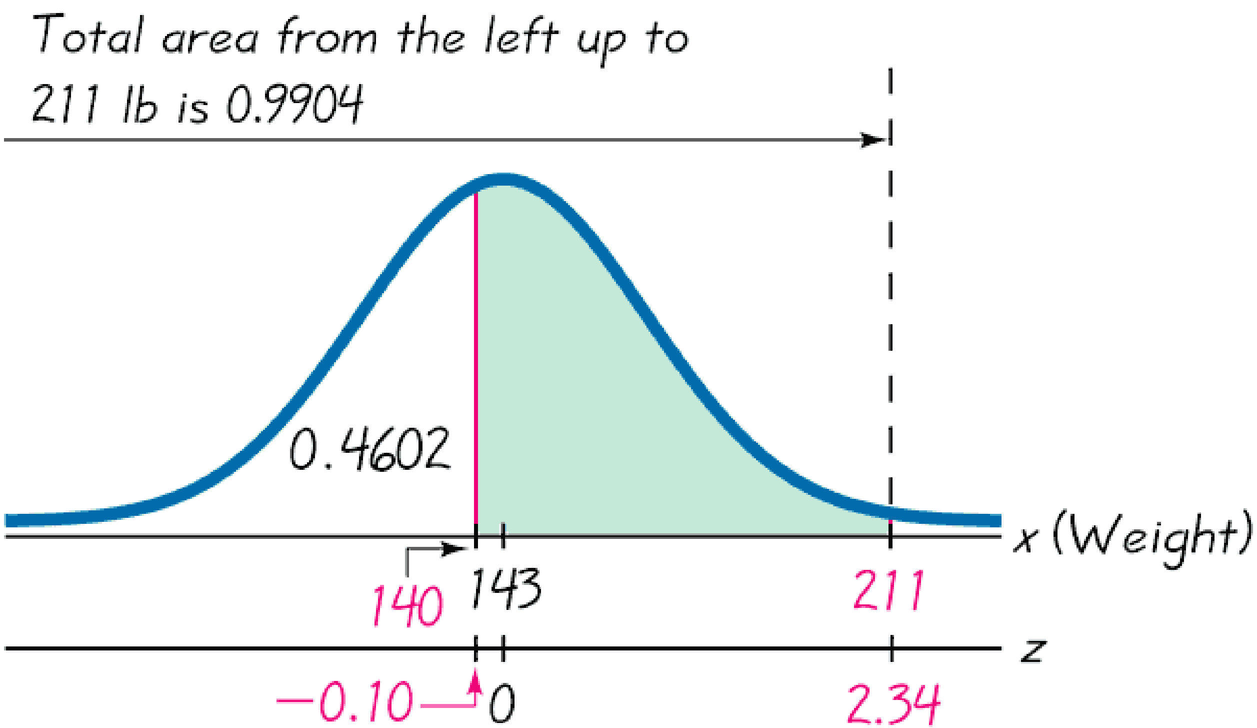


Figure 5-14 (p.242)

Probability of Weight between 140 pounds and 211 pounds

μ = 143
σ = 29

There is a 0.5302 probability of randomly selecting a woman with a weight between 140 and 211 lbs.

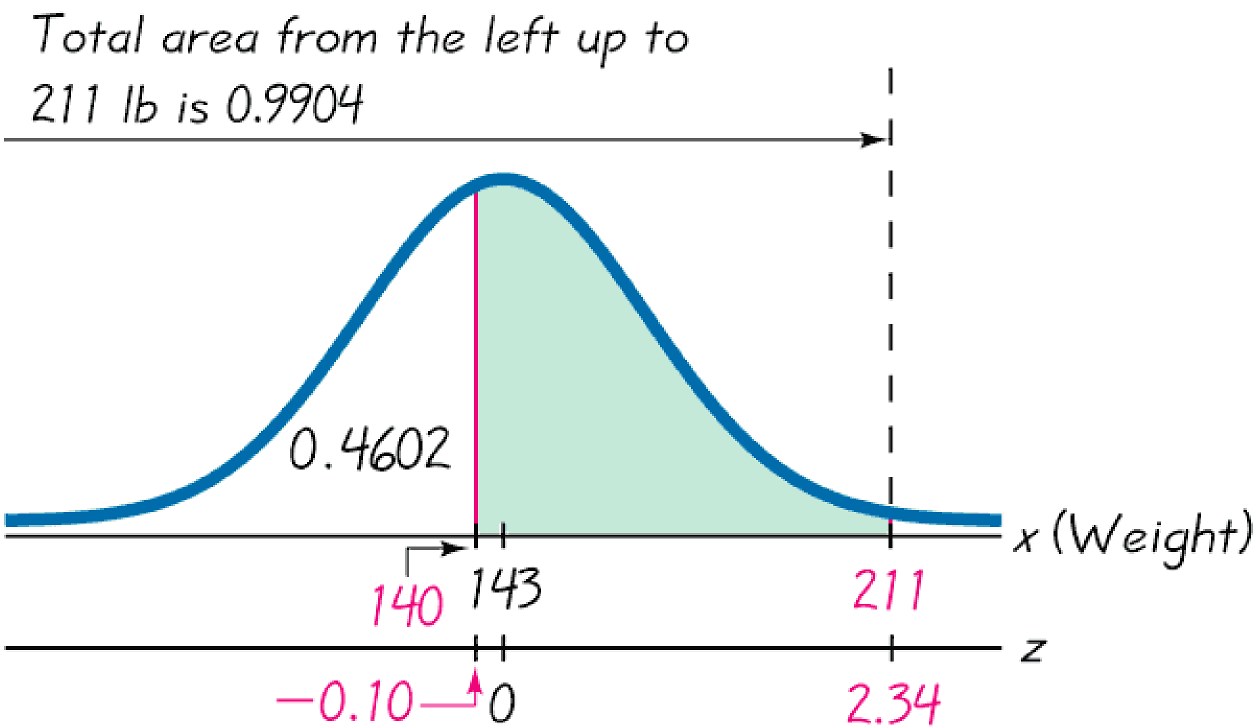


Figure 5-14 (p.242)

Probability of Weight between 140 pounds and 211 pounds

μ = 143
σ = 29

OR - 53.02% of women have weights between 140 lb and 211 lb.

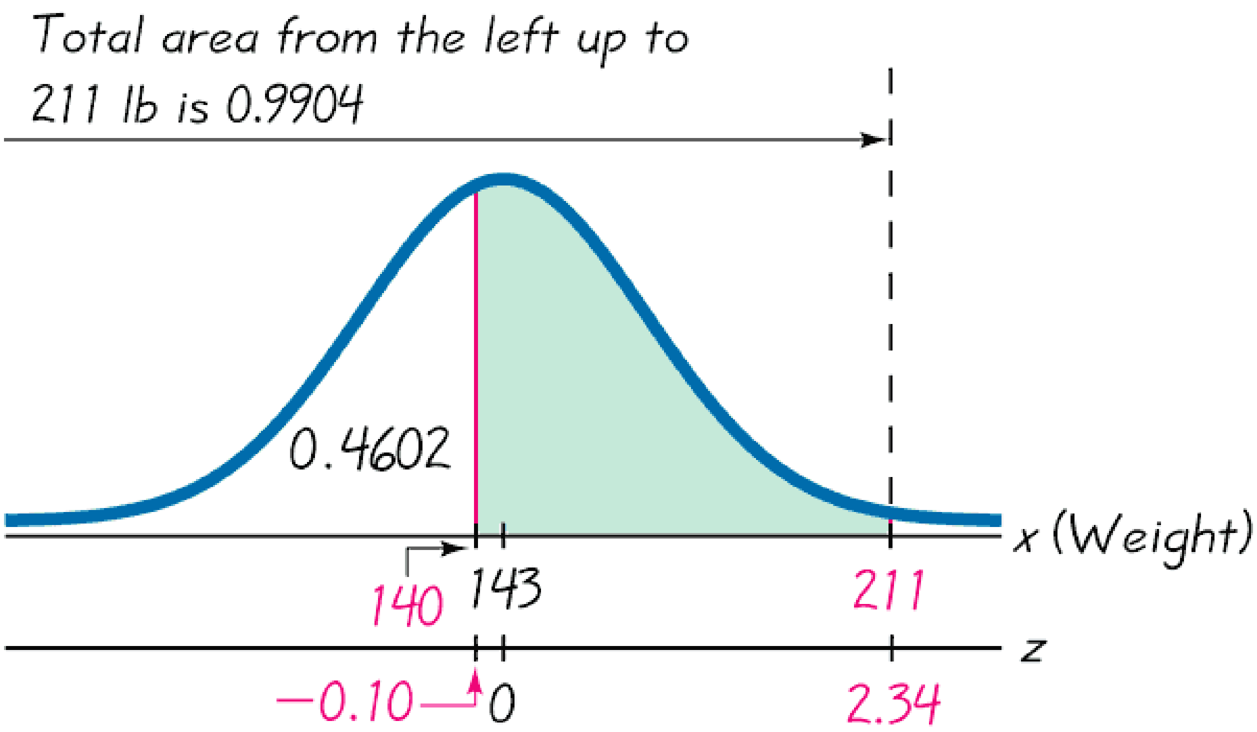


Figure 5-14 (p.242)

Cautions to keep in mind

1. Don’t confuse z scores and areas. z scores are distances along the horizontal scale, but areas are regions under the normal curve. Table A-2 lists z scores in the left column and across the top row, but areas are found in the body of the table.

2. Choose the correct (right/left) side of the graph.
3. A z score must be negative whenever it is located to the *left* half of the normal distribution.
4. Areas (or probabilities) are positive or zero values, but they are never negative.

Procedure for Finding Values Using Table A-2 and Formula 5-2

1. **Sketch a normal distribution curve, enter the given probability or percentage in the appropriate region of the graph, and identify the x value(s) being sought.**

2. **Use Table A-2 to find the z score corresponding to the cumulative left area bounded by x. Refer to the BODY of Table A-2 to find the closest area, then identify the corresponding z score.**

3. **Using Formula 5-2, enter the values for μ, σ, and the z score found in step 2, then solve for x.**

$$x = \mu + (z \cdot \sigma)$$ **(Another form of Formula 5-2)**

(If z is located to the left of the mean, be sure that it is a negative number.)

4. **Refer to the sketch of the curve to verify that the solution makes sense in the context of the graph and the context of the problem.**

Find P_{98} for Hip Breadths of Men

$x = \mu + (z \bullet \sigma)$

$x = 14.4 + (2.05 \bullet 1.0)$

$x = 16.45$

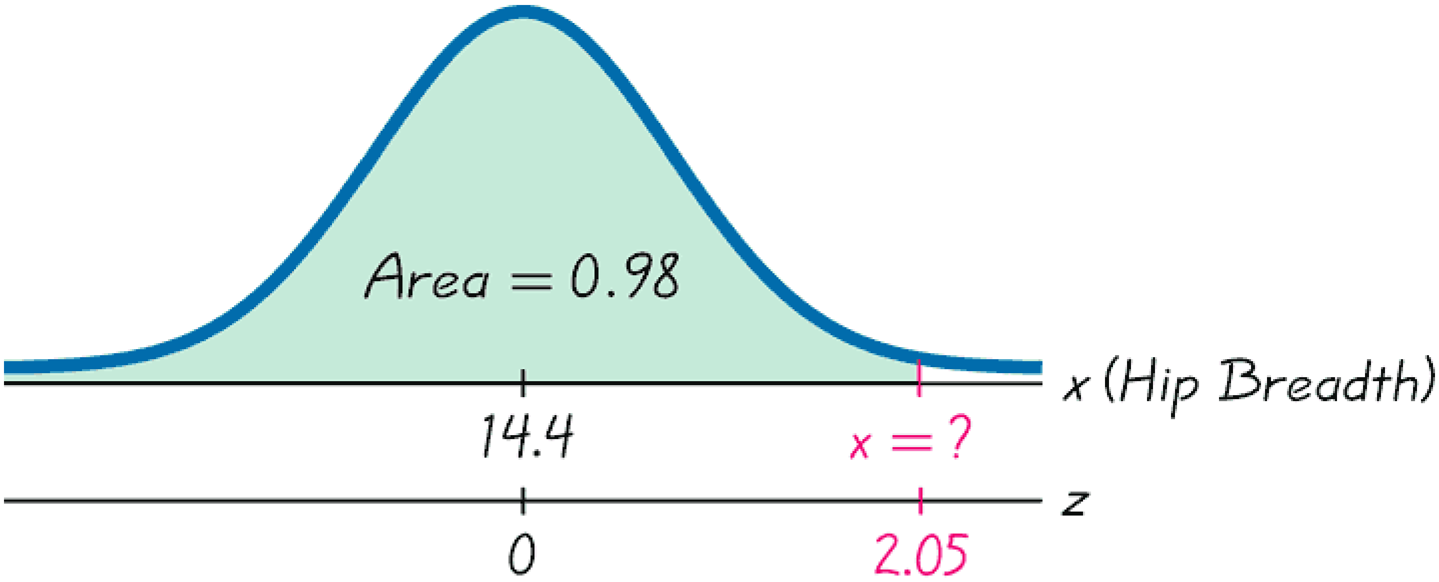


Figure 5-15 (p.244)

Find P_{98} for Hip Breadths of Men

The hip breadth of 16.5 in. separates the lowest 98% from the highest 2%

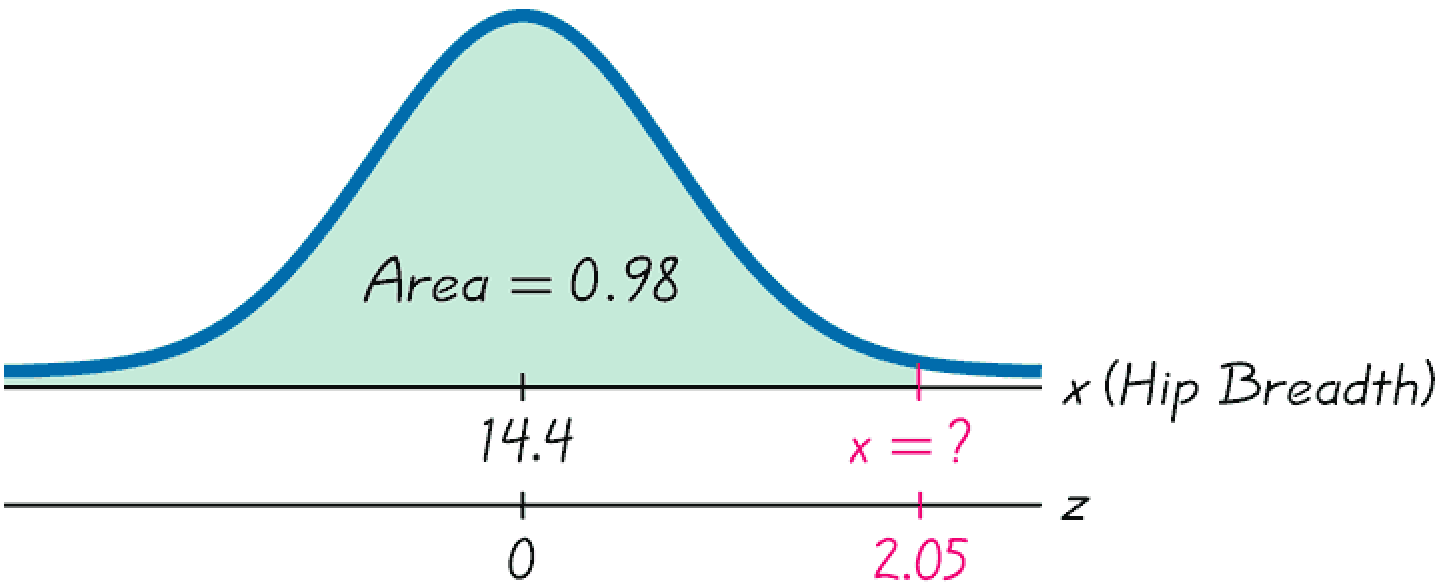


Figure 5-15 (p.244)

Find P_{98} for Hip Breadths of Men

Seats designed for a hip breadth up to 16.5 in. will fit 98% of men.

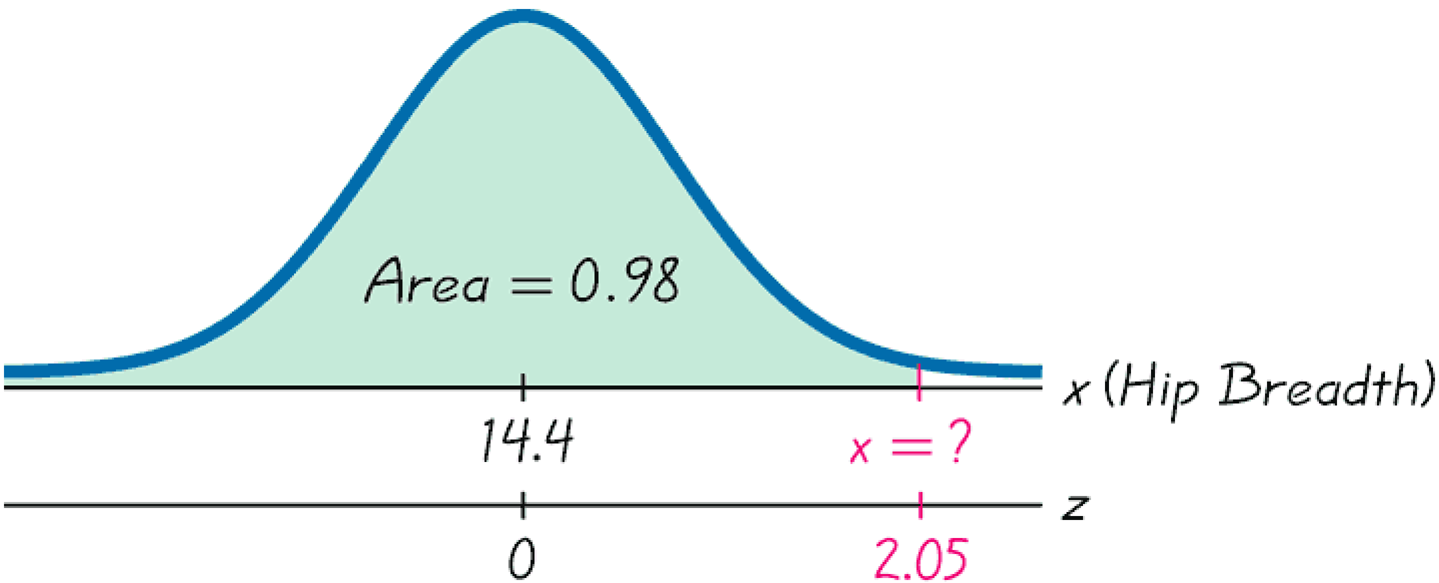


Figure 5-15 (p.244)

Finding P_{05} for Grips of Women

$x = 27.0 + (-1.645 \cdot 1.3) = 24.8615$

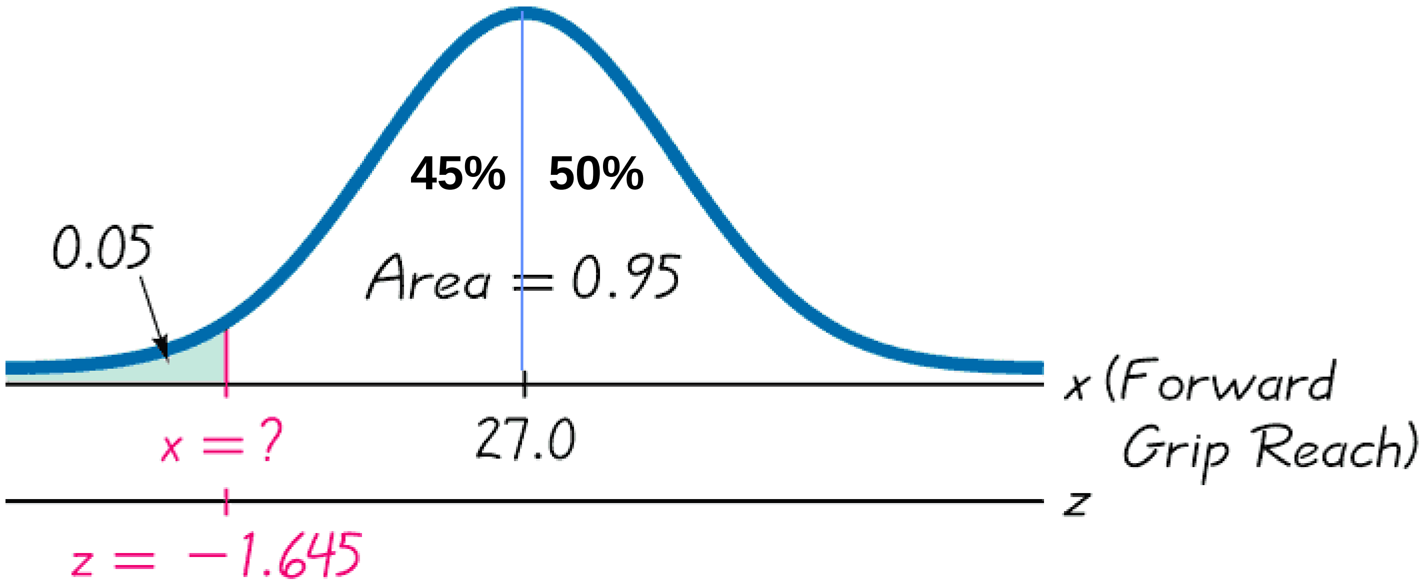


Figure 5-16 (p.245)

Finding P_{05} for Grips of Women

The forward grip of 24.9 in. (rounded) separates the top 95% from the others.

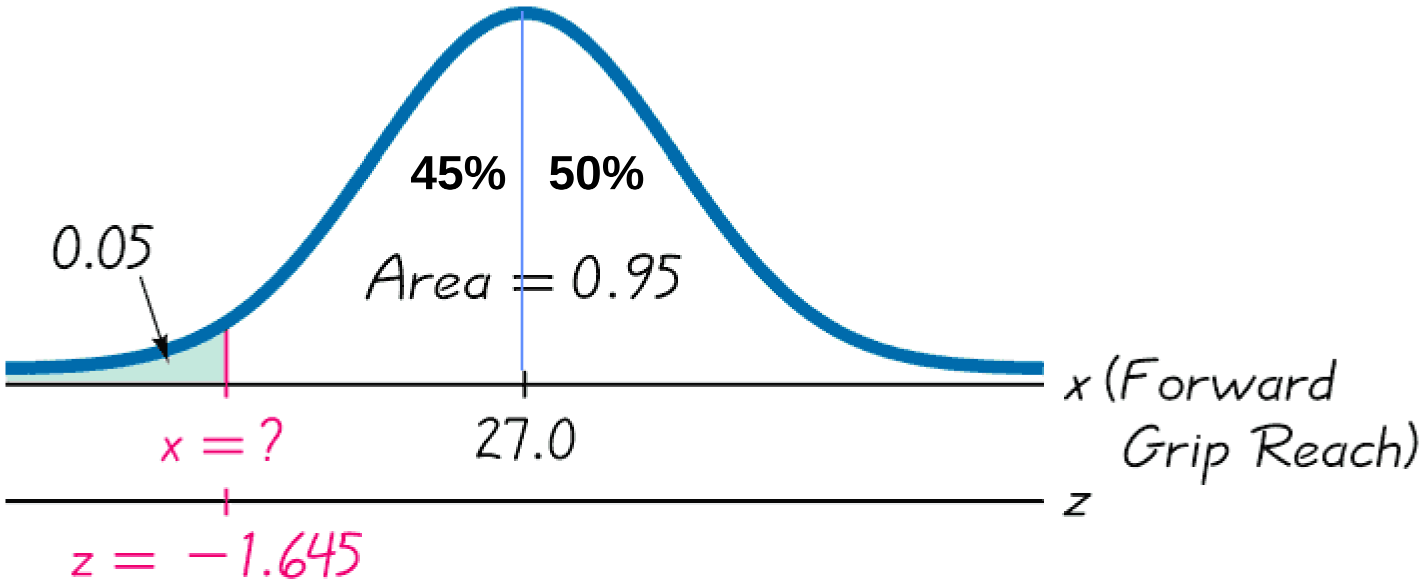
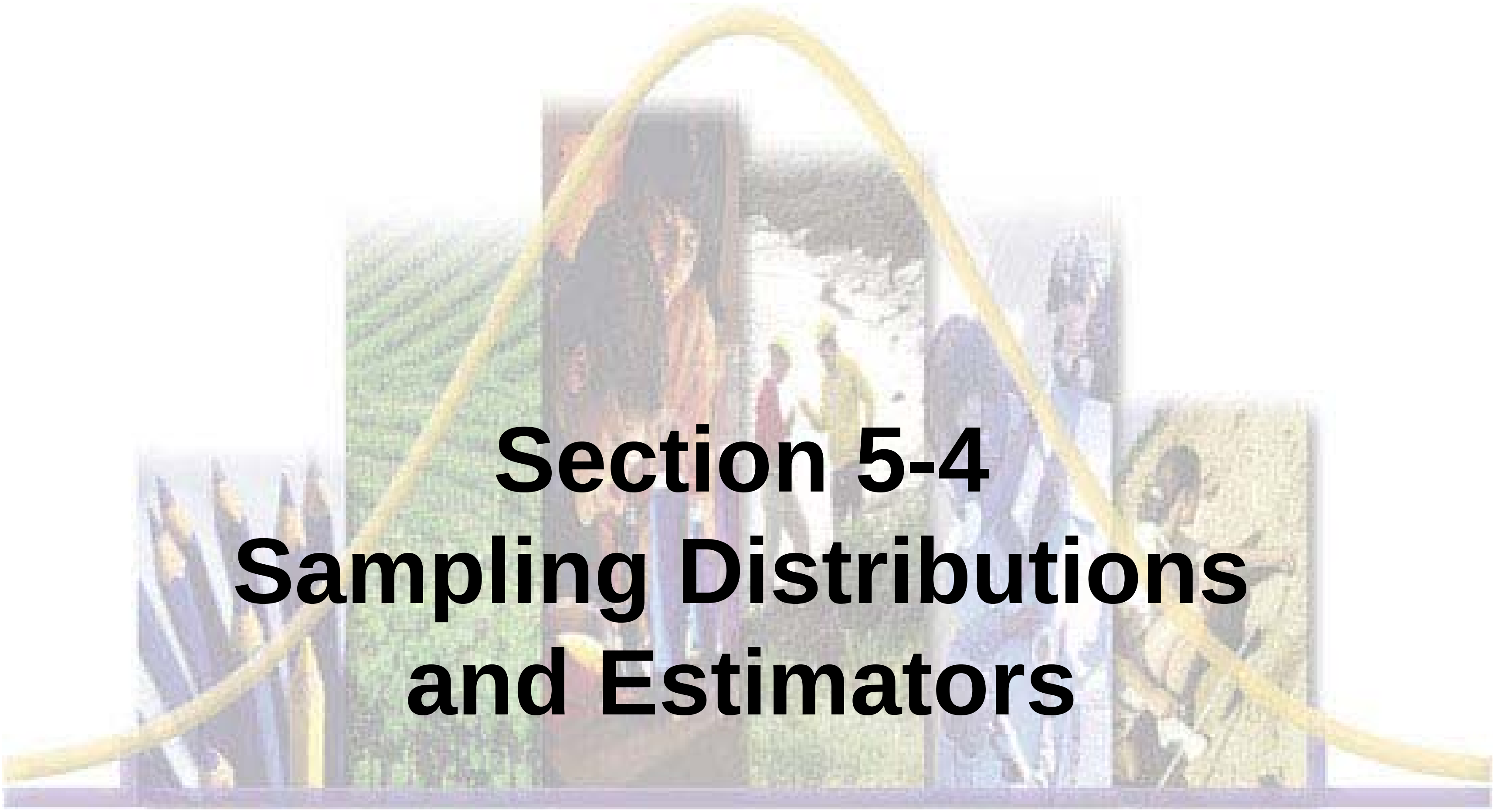


Figure 5-16 (p.245)

REMEMBER!

Make the *z* score negative if the value is located to the left (below) the mean. Otherwise, the *z* score will be positive.



Section 5-4
Sampling Distributions and Estimators

Created by Erin Hodgess, Houston, Texas

PEARSON
Addison Wesley

Definition

Sampling Distribution of the mean **is the probability distribution of sample means, with all samples having the same sample size *n*.**

Definition

Sampling Variability:

The value of a statistic, such as the sample mean $\bar{x}$, depends on the particular values included in the sample.

Definition

The Sampling Distribution of the Proportion is the probability distribution of sample proportions, with all samples having the same sample size *n*.

Sampling Distributions

A population consists of the values 1, 2, and 5. We randomly select samples of size 2 with replacement. There are 9 possible samples.

a. **For each sample, find the mean, median, range, variance, and standard deviation.**

b. **For each statistic, find the mean from part (a)**

See Table 5-2 (p.251) on the next slide.

Table 5-2 Sampling Distributions of Different Statistics (for Samples of Size 2 Drawn with Replacement from the Population 1, 2, 5)

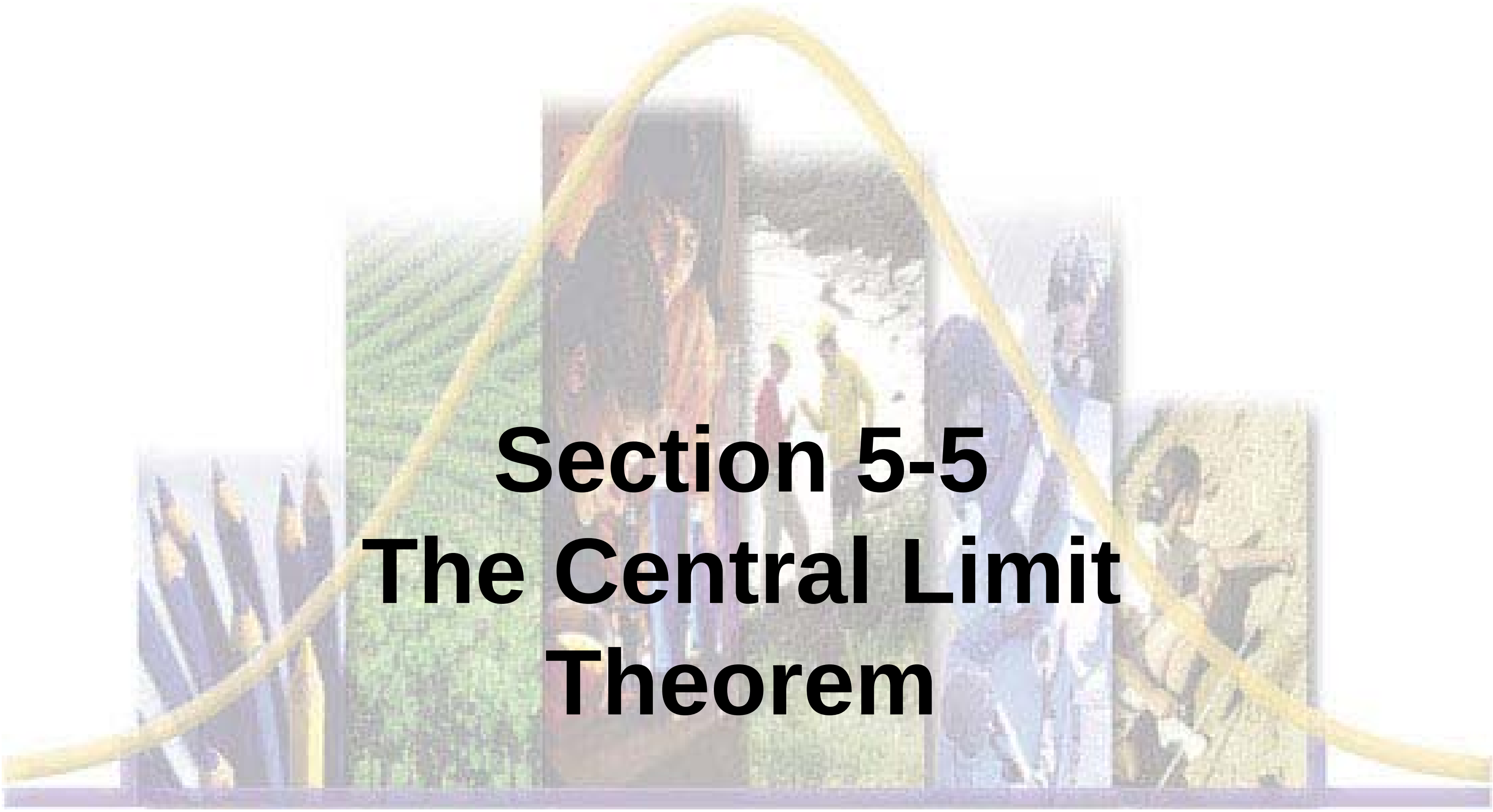
Sample	Mean $\bar{x}$	Median	Range	Variance s^2	Standard Deviation s	Proportion of Odd Numbers	Probability
1, 1	1.0	1.0	0	0.0	0.000	1	1/9
1, 2	1.5	1.5	1	0.5	0.707	0.5	1/9
1, 5	3.0	3.0	4	8.0	2.828	1	1/9
2, 1	1.5	1.5	1	0.5	0.707	0.5	1/9
2, 2	2.0	2.0	0	0.0	0.000	0	1/9
2, 5	3.5	3.5	3	4.5	2.121	0.5	1/9
5, 1	3.0	3.0	4	8.0	2.828	1	1/9
5, 2	3.5	3.5	3	4.5	2.121	0.5	1/9
5, 5	5.0	5.0	0	0.0	0.000	1	1/9
Mean of Statistic Values	2.7	2.7	1.8	2.9	1.3	0.667	
Population Parameter	2.7	2	4	2.9	1.7	0.667	
Does the sample statistic target the population parameter?	Yes	No	No	Yes	No	Yes	

Interpretation of Sampling Distributions

We can see that when using a sample statistic to estimate a population parameter, some statistics are good in the sense that they target the population parameter and are therefore likely to yield good results. Such statistics are called *unbiased estimators*.

Statistics that target population parameters: mean, variance, proportion

Statistics that do not target population parameters: median, range, standard deviation



Section 5-5
The Central Limit Theorem

Created by Erin Hodgess, Houston, Texas

Central Limit Theorem

Given (p.260):

1. **The random variable x has a distribution (which may or may not be normal) with mean μ and standard deviation σ.**

2. **Samples all of the same size n are randomly selected from the population of x values.**

Central Limit Theorem

Conclusions (p.260):

1. The distribution of sample $\bar{x}$ will, as the sample size increases, approach a *normal* distribution.

2. The mean of the sample means will be the population mean μ.

3. The standard deviation of the sample means will approach $\sigma/\sqrt{n}$.

Practical Rules Commonly Used:

1. **For samples of size *n* larger than 30, the distribution of the sample means can be approximated reasonably well by a normal distribution. The approximation gets better as the sample size *n* becomes larger.**

2. **If the original population is itself normally distributed, then the sample means will be normally distributed for <u>any sample size</u> *n* (not just the values of *n* larger than 30).**

Notation (p.261)

the mean of the sample means

$$\mu_{\bar{x}} = \mu$$

the standard deviation of sample mean

$$\sigma_{\bar{x}} = \frac{\sigma}{\sqrt{n}}$$

(often called standard error of the mean)

Distribution of 200 digits from Social Security Numbers

(Last 4 digits from 50 students)

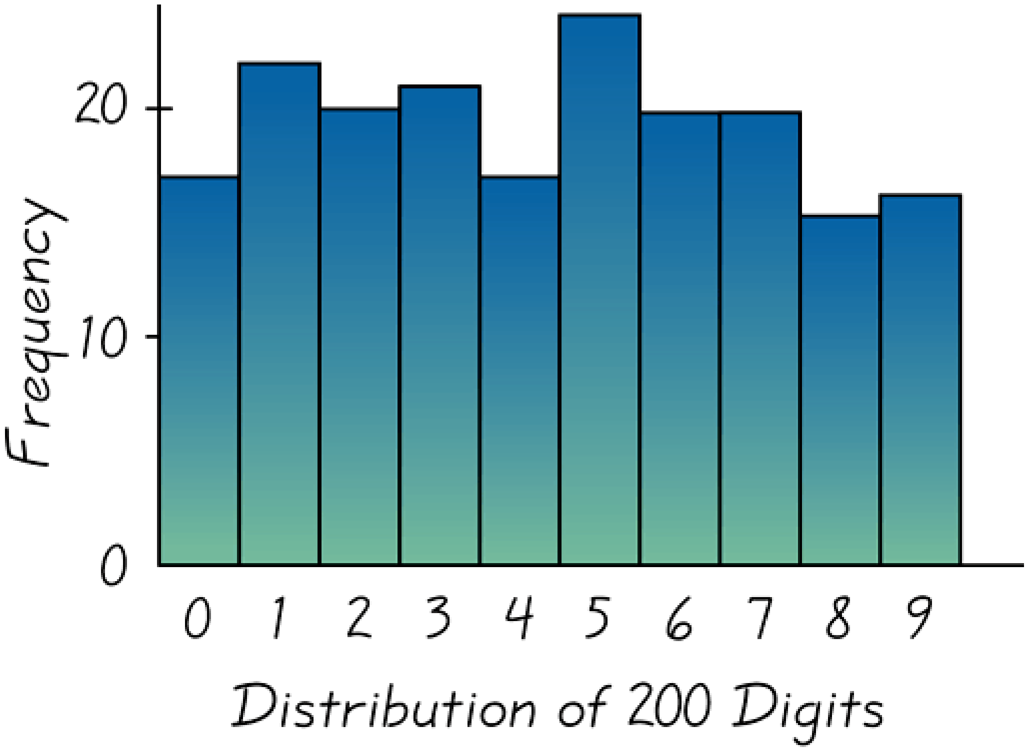


Figure 5-19 (p.262)

Table 5-6

SSN digits				$\bar{x}$
1	8	6	4	4.75
5	3	3	6	4.25
9	8	8	8	8.25
5	1	2	5	3.25
9	3	3	5	5.00
4	2	6	2	3.50
7	7	1	6	5.25
9	1	5	4	4.75
5	3	3	9	5.00
7	8	4	1	5.00
0	5	6	1	3.00
9	8	2	2	5.25
6	1	5	7	4.75
8	1	3	0	3.00
5	9	6	9	7.25
6	2	3	4	3.75
7	4	0	[illegible]	4.50
5	[illegible]	5	[illegible]	5.75
[illegible]	[illegible]	[illegible]	6	[illegible]
[illegible]	2	8	6	[illegible]
2	0	9	7	4.50
5	8	9	0	5.50
6	5	4	9	6.00
4	8	7	6	6.25
7	1	2	0	2.50
2	9	5	0	4.00
8	3	2	2	3.75
2	7	1	6	4.00
6	7	7	1	5.25
2	3	3	9	4.25
2	4	7	5	4.50
5	4	3	7	4.75
0	4	3	8	3.75
2	5	8	6	5.25
7	1	3	4	3.75
8	3	7	0	4.50
5	6	6	7	6.00

Distribution of 50 Sample Means for 50 Students

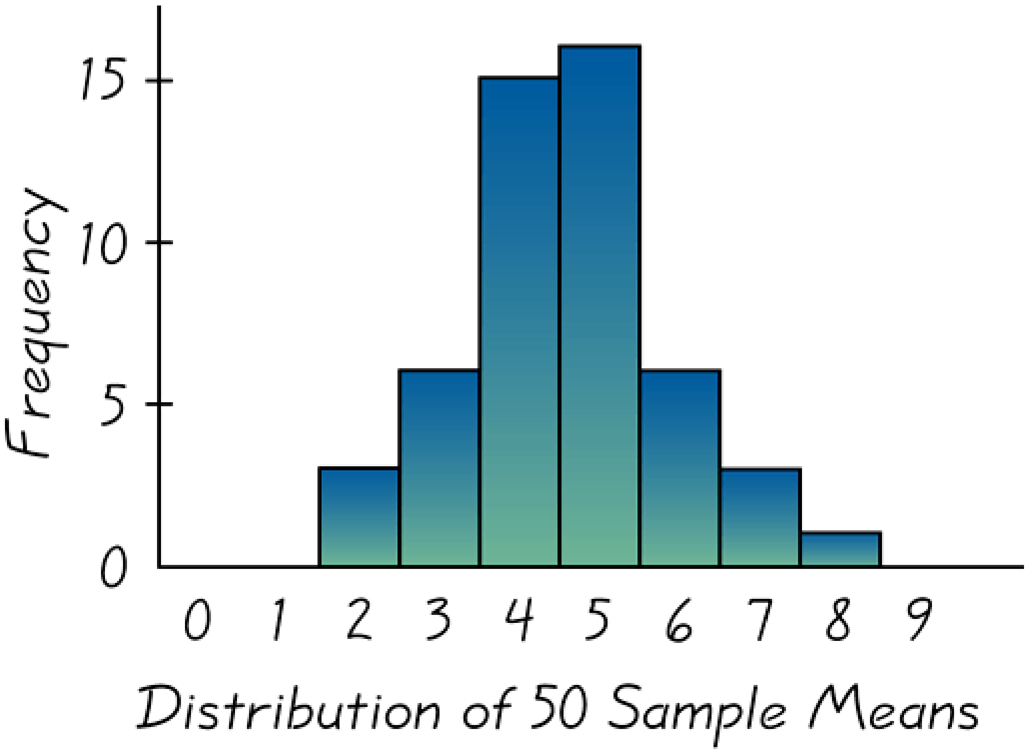


Figure 5-20 (p.262)

As the sample size increases, the sampling distribution of sample means approaches a normal distribution.

Example (p.262): **Given the population of men has normally distributed weights with a mean of 172 lb and a standard deviation of 29 lb,**

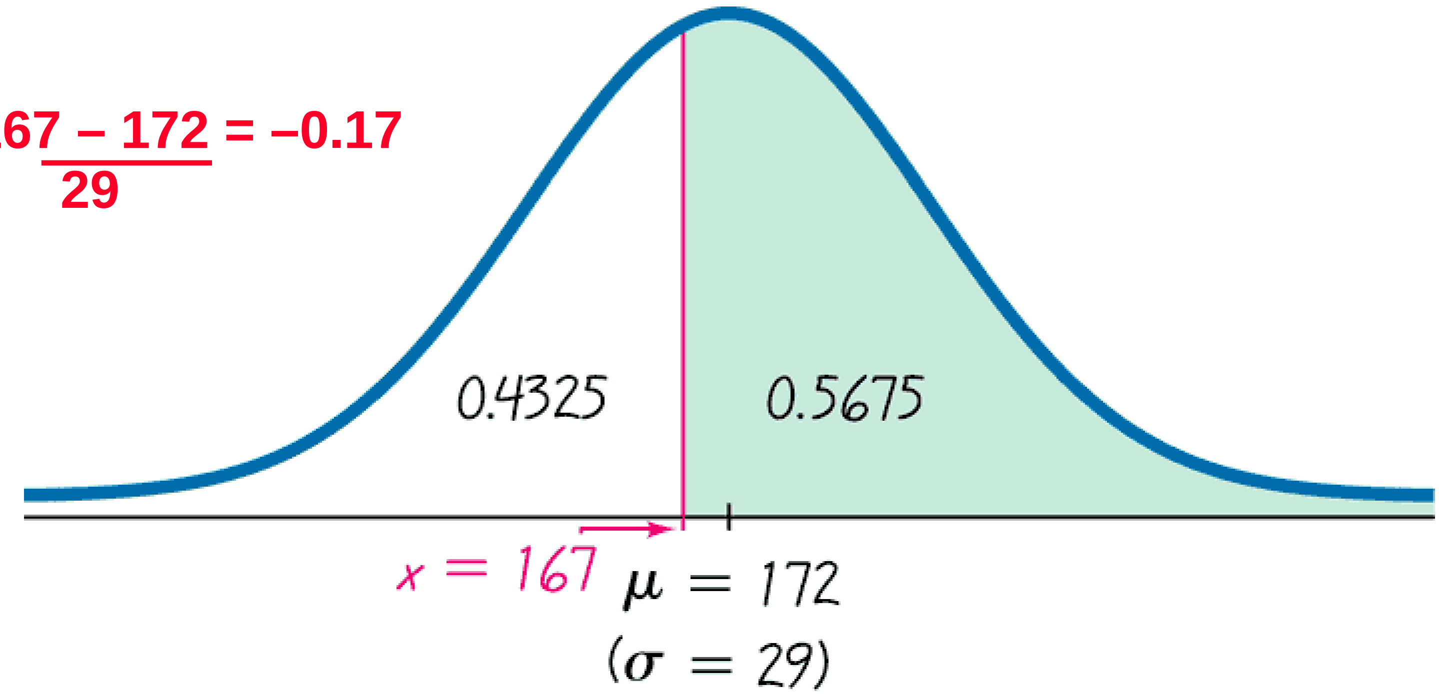
a) if one man is randomly selected, find the probability that his weight is greater than 167 lb.

b) if 12 different men are randomly selected, find the probability that their mean weight is greater than 167 lb.

Example: **Given the population of men has normally distributed weights with a mean of 172 lb and a standard deviation of 29 lb,**

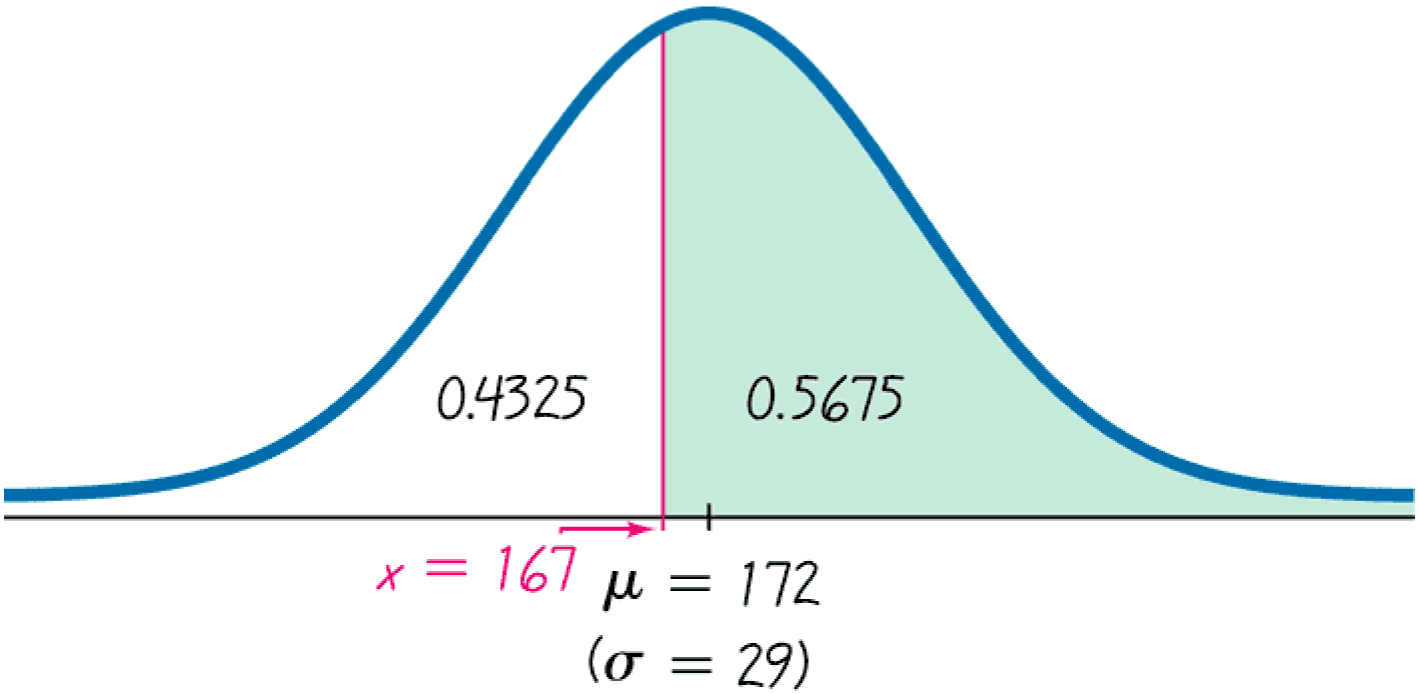
a) if one man is randomly selected, find the probability that his weight is greater than 167 lb.

$$z = \frac{167 - 172}{29} = -0.17$$



Example: **Given the population of men has normally distributed weights with a mean of 172 lb and a standard deviation of 29 lb,**

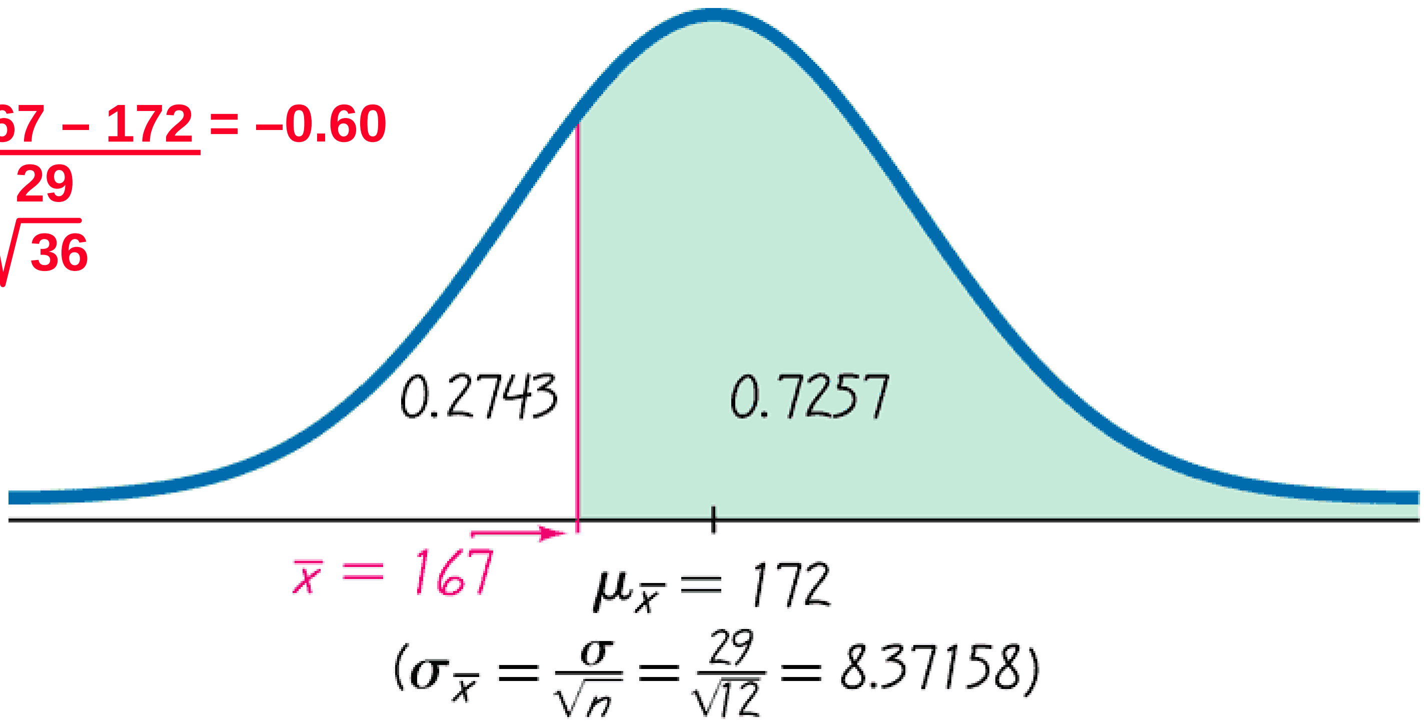
a) if one man is randomly selected, the probability that his weight is greater than 167 lb. is 0.5675.



Example: **Given the population of men has normally distributed weights with a mean of 172 lb and a standard deviation of 29 lb,**

b) if 12 different men are randomly selected, find the probability that their mean weight is greater than 167 lb.

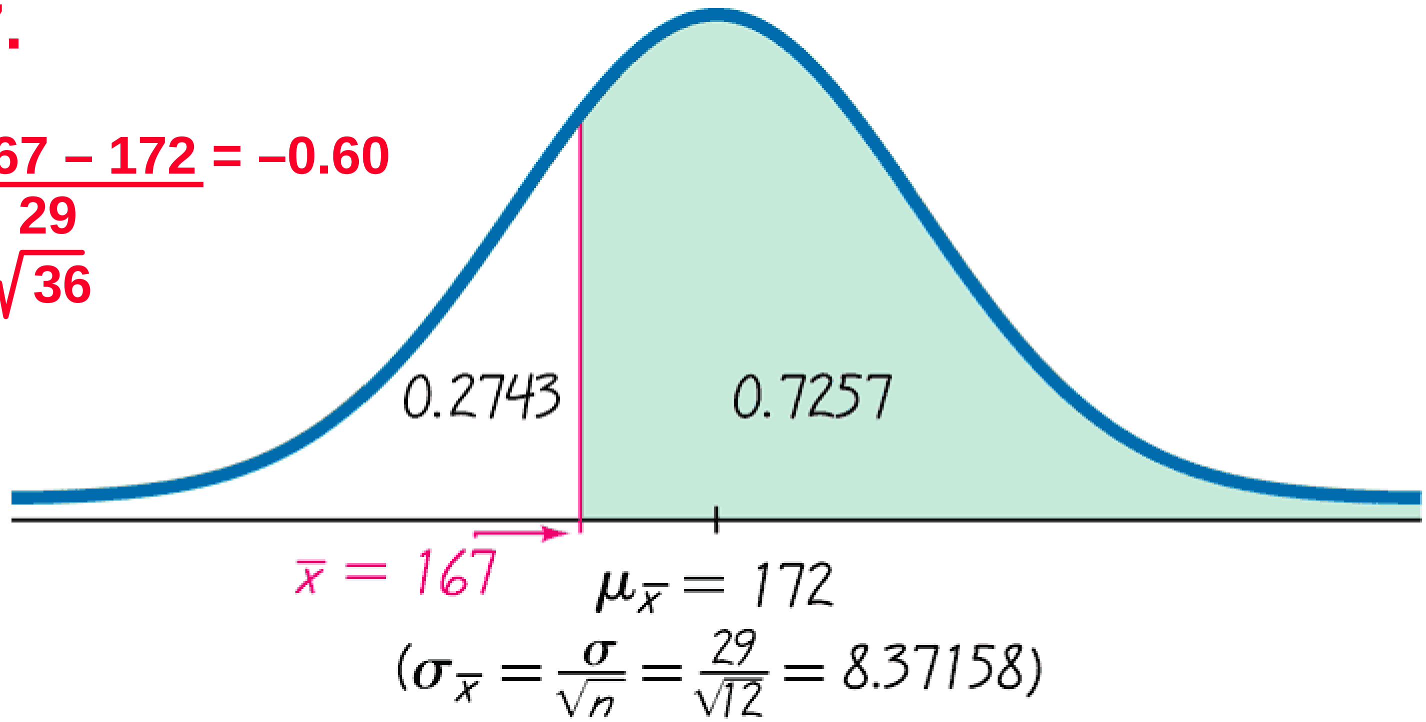
$$z = \frac{167 - 172}{\frac{29}{\sqrt{36}}} = -0.60$$



Example: **Given the population of men has normally distributed weights with a mean of 143 lb and a standard deviation of 29 lb,**

b.) if 12 different men are randomly selected, the probability that their mean weight is greater than 167 lb is 0.7257.

$$z = \frac{167 - 172}{\frac{29}{\sqrt{36}}} = -0.60$$



Example: **Given the population of men has normally distributed weights with a mean of 172 lb and a standard deviation of 29 lb,**

a) if one man is randomly selected, find the probability that his weight is greater than 167 lb.

$$P(x > 167) = 0.5675$$

b) if 12 different men are randomly selected, their mean weight is greater than 167 lb.

$$P(\bar{x} > 167) = 0.7257$$

It is much easier for an individual to deviate from the mean than it is for a group of 12 to deviate from the mean.

Sampling Without Replacement (p.266)

If n > 0.05 N

$$\sigma_{\bar{x}} = \frac{\sigma}{\sqrt{n}}\sqrt{\frac{N-n}{N-1}}$$

finite population correction factor

Section 5-6
Normal as Approximation to Binomial

Created by Erin Hodgess, Houston, Texas

Review

Binomial Probability Distribution

1. The procedure must have fixed number of trials.
2. The trials must be independent.
3. Each trial must have all outcomes classified into two categories.
4. The probabilities must remain constant for each trial.

Solve by binomial probability formula, Table A-1, or technology

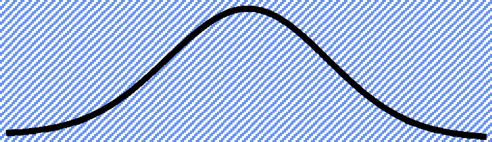
Approximate a Binomial Distribution with a Normal Distribution if:

$np \geq 5$

$nq \geq 5$

then $\mu = np$ and $\sigma = \sqrt{npq}$

and the random variable has

a  distribution.

(normal)

Solving Binomial Probability Problems Using a Normal Approximation

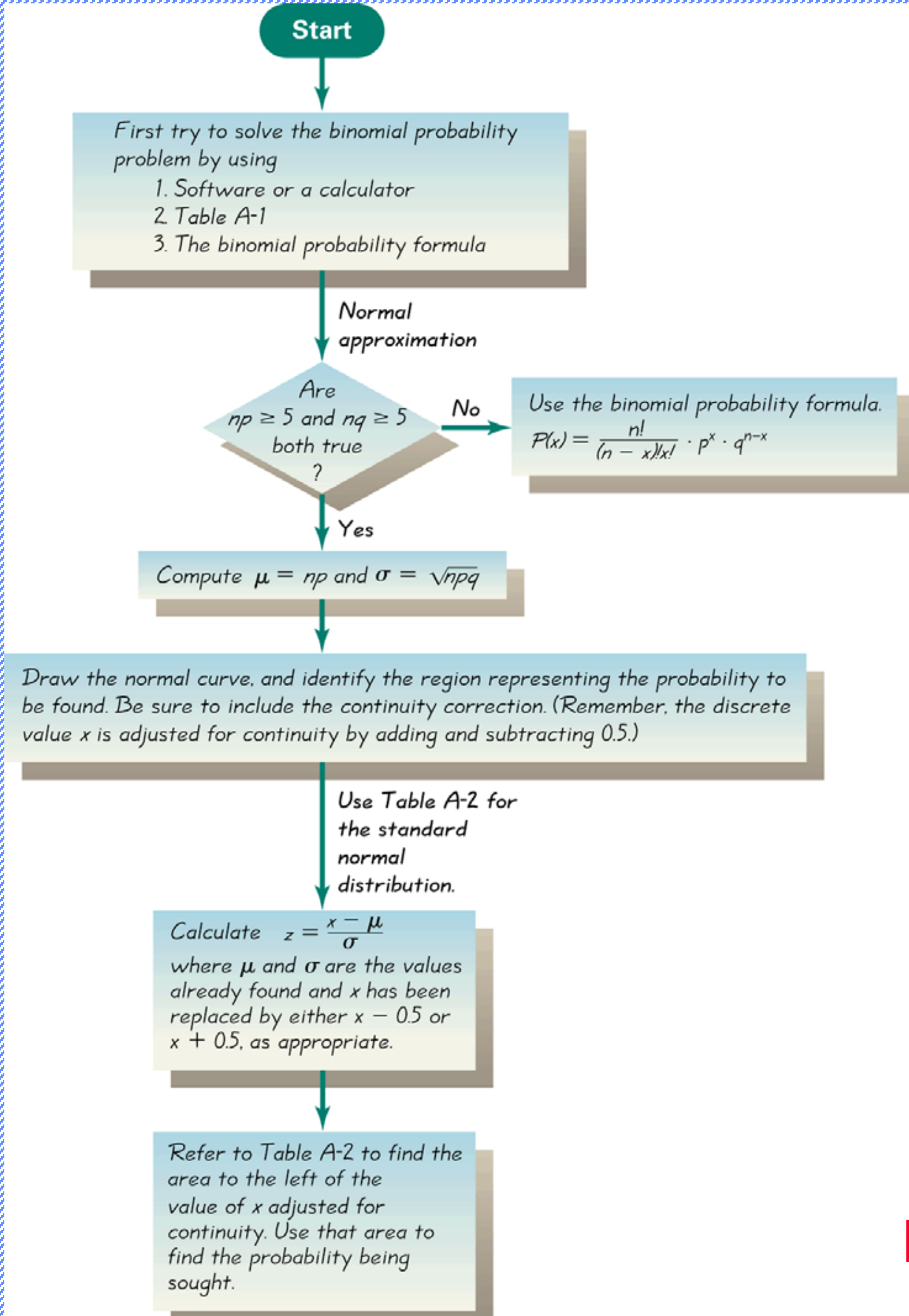
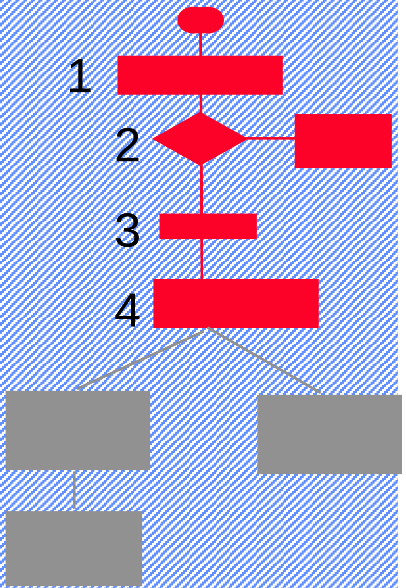


Figure 5-23

Solving Binomial Probability Problems Using a Normal Approximation



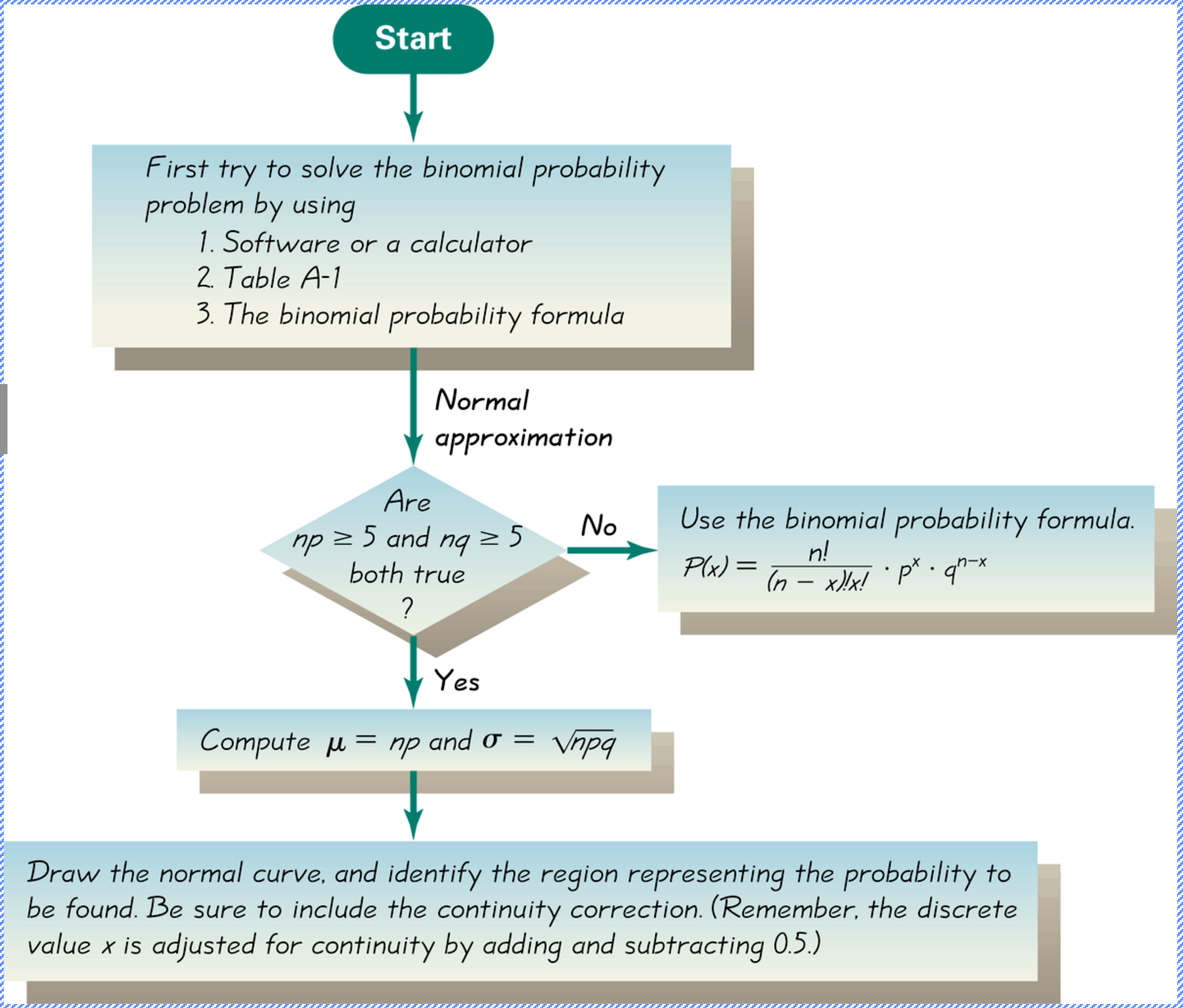


Figure 5-23

Solving Binomial Probability Problems Using a Normal Approximation

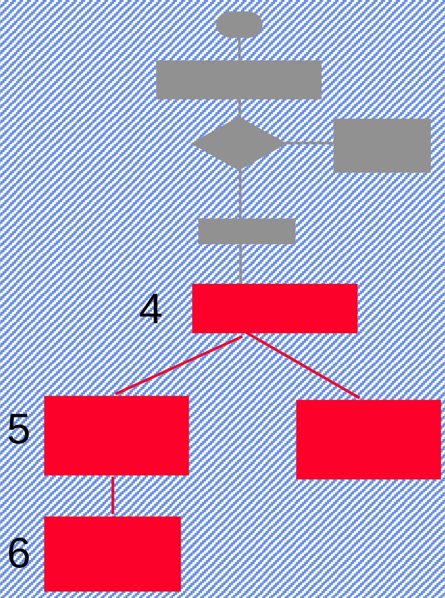


Figure 5-23

Draw the normal curve, and identify the region representing the probability to be found. Be sure to include the continuity correction. (Remember, the discrete value x is adjusted for continuity by adding and subtracting 0.5.)

Use Table A-2 for the standard normal distribution.

Use a TI-83 calculator

Calculate $z = \dfrac{x - \mu}{\sigma}$ where μ and σ are the values already found and x has been replaced by either $x - 0.5$ or $x + 0.5$, as appropriate.

Refer to Table A-2 to find the area to the left of the value of x adjusted for continuity. Use that area to find the probability being sought.

Procedure for Using a Normal Distribution to Approximate a Binomial Distribution

1. Establish that the normal distribution is a suitable approximation to the binomial distribution by verifying $np \geq 5$ and $nq \geq 5$.

2. Find the values of the parameters μ and σ by calculating $\mu = np$ and $\sigma = \sqrt{npq}$.

3. Identify the discrete value of x (the number of successes). Change the discrete value x by replacing it with the interval from $x - 0.5$ to $x + 0.5$. Draw a normal curve and enter the values of μ, σ, and either $x - 0.5$ or $x + 0.5$, as appropriate.

continued

Procedure for Using a Normal Distribution to Approximate a Binomial Distribution

continued

4. **Change *x* by replacing it with *x* – 0.5 or *x* + 0.5, as appropriate.**

5. **Find the area corresponding to the desired probability.**

Finding the Probability of "At Least" 120 Men Among 200 Accepted Applicants

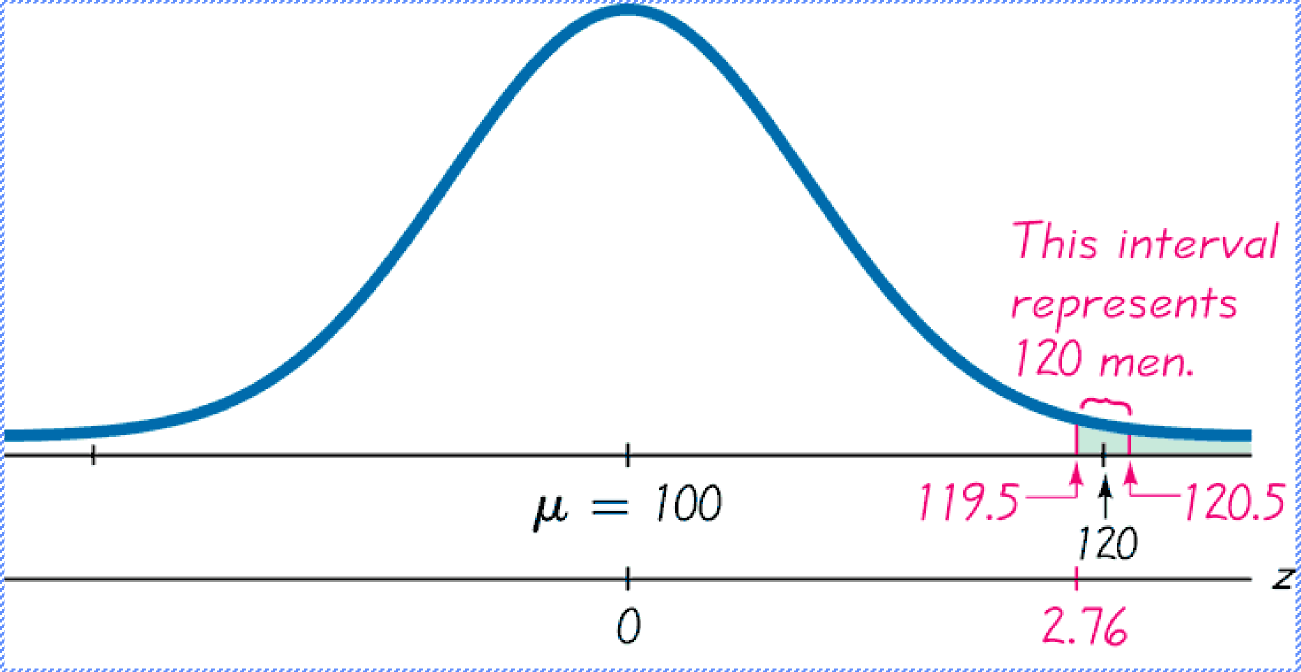


Figure 5-24

Definition

When we use the normal distribution (which is continuous) as an approximation to the binomial distribution (which is discrete), a continuity correction is made to a discrete whole number x in the binomial distribution by representing the single value x by the interval from

$x – 0.5$ to $x + 0.5$.

Procedure for Continuity Corrections

1. **When using the normal distribution as an approximation to the binomial distribution, *always* use the continuity correction.**

2. **In using the continuity correction, first identify the discrete whole number X that is relevant to the binomial probability problem.**

3. **Draw a normal distribution centered about μ, then draw a vertical strip area centered over X . Mark the left side of the strip with the number X – 0.5, and mark the right side with X + 0.5. For X =120, draw a strip from 119.5 to 120.5. Consider the area of the strip to represent the probability of discrete number X.________________ ________________**

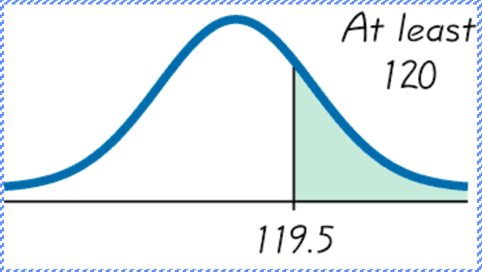
continued

Procedure for Continuity Corrections

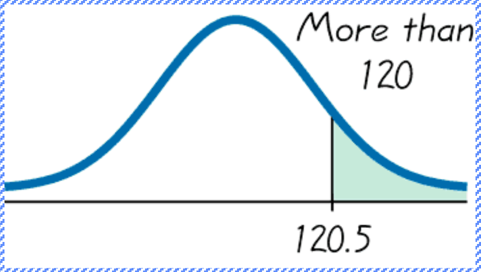
continued

4. **Now determine whether the value of *X* itself should be included in the probability you want. Next, determine whether you want the probability of at least *X*, at most *X*, more than *X*, fewer than *X*, or exactly *X*. Shade the area to the right or left of the strip, as appropriate; also shade the interior of the strip itself if and only if *X itself* is to be included. The total shaded region corresponds to probability being sought.**

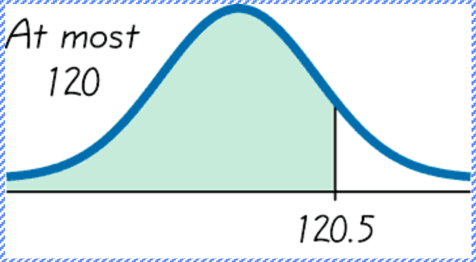
x **= <u>at least</u> 120**
= 120, 121, 122, . . .



x **= <u>more than</u> 120**
= 121, 122, 123, . . .



x **= <u>at most</u> 120**
= 0, 1, . . . 118, 119, 120



x **= <u>fewer than</u> 120**
= 0, 1, . . . 118, 119

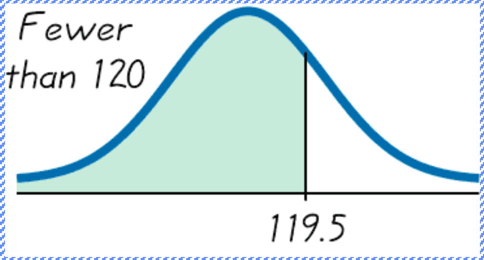
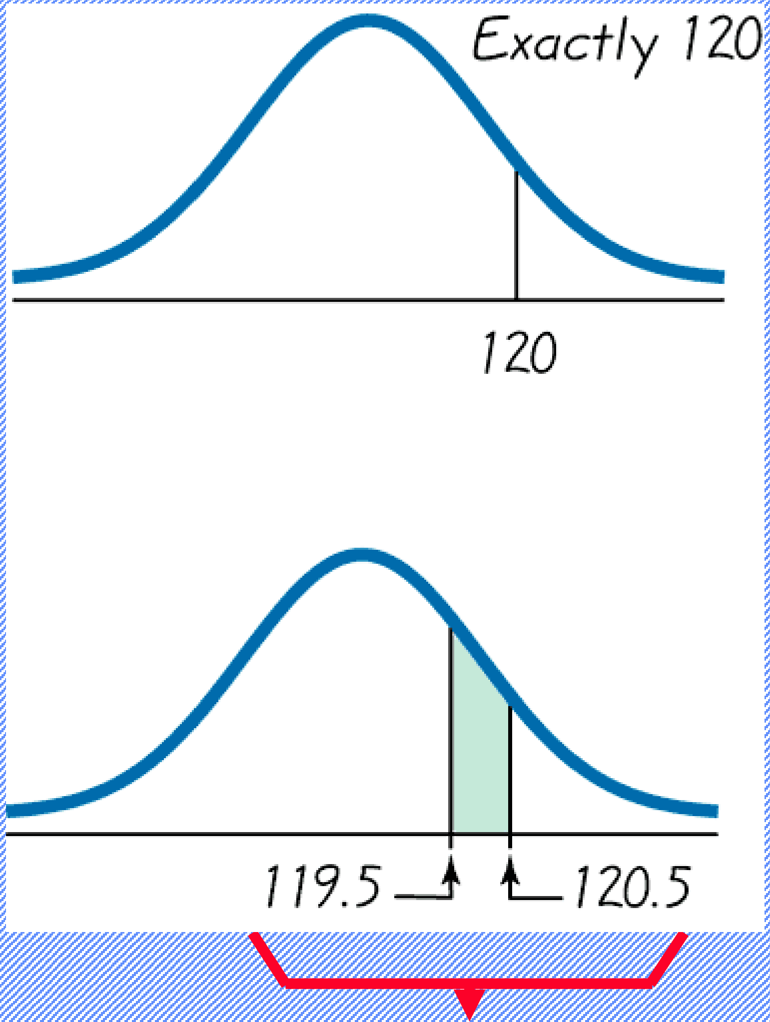
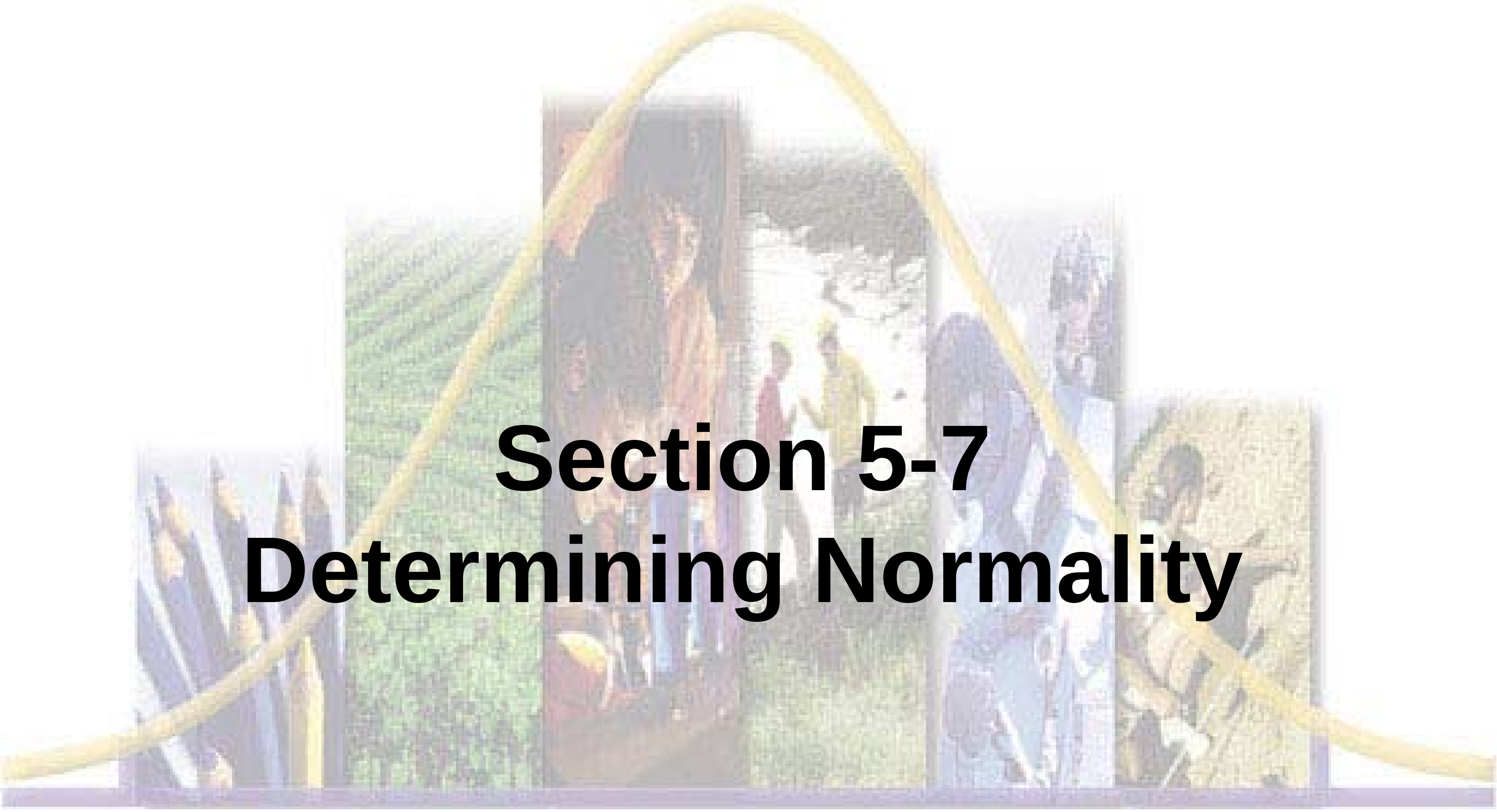


Figure 5-25

x = <u>exactly</u> 120



Interval represents discrete number 120



Section 5-7
Determining Normality

Created by Erin Hodgess, Houston, Texas

Definition

A Normal Quantile Plot is a graph of points (*x*,*y*), where each *x* value is from the original set of sample data, and each *y* value is a *z* score corresponding to a quantile value of the standard normal distribution.

Procedure for Determining Whether Data Have a Normal Distribution

1. ***Histogram:*** **Construct a histogram. Reject normality if the histogram departs dramatically from a bell shape.**

2. ***Outliers:*** **Identify outliers. Reject normality if there is more than one outlier present.**

3. ***Normal Quantile Plot:*** **If the histogram is basically symmetric and there is at most one outlier, construct a *normal quantile plot* as follows:**

Procedure for Determining Whether Data Have a Normal Distribution

3. **Normal Quantile Plot**

a. Sort the data by arranging the values from lowest to highest.

b. With a sample size *n*, each value represents a proportion of 1/*n* of the sample. Using the known sample size *n*, identify the areas of 1/2n, 3/2n, 5/2n, 7/2n, and so on. These are the cumulative areas to the left of the corresponding sample values.

c. Use the standard normal distribution (Table A-2) to find the *z* scores corresponding to the cumulative left areas found in Step (b).

continued

Procedure for Determining Whether Data Have a Normal Distribution

continued

d. Match the original sorted data values with their corresponding z scores found in Step (c), then plot the points (*x*, *y*), where each *x* is an original sample value and *y* is the corresponding z score.

e. Examine the normal quantile plot using these criteria:

If the points do not lie close to a straight line, or if the points exhibit some systematic pattern that is not a straight-line pattern, then the data appear to come from a population that does *not* have a normal distribution. If the pattern of the points is reasonably close to a straight line, then the data appear to come from a population that has a normal distribution.

Procedure for Determining Whether Data Have a Normal Distribution

Examine the normal quantile plot and <u>reject normality</u> if the the points do not lie close to a straight line, or if the points exhibit some systematic pattern that is not a straight-line pattern.