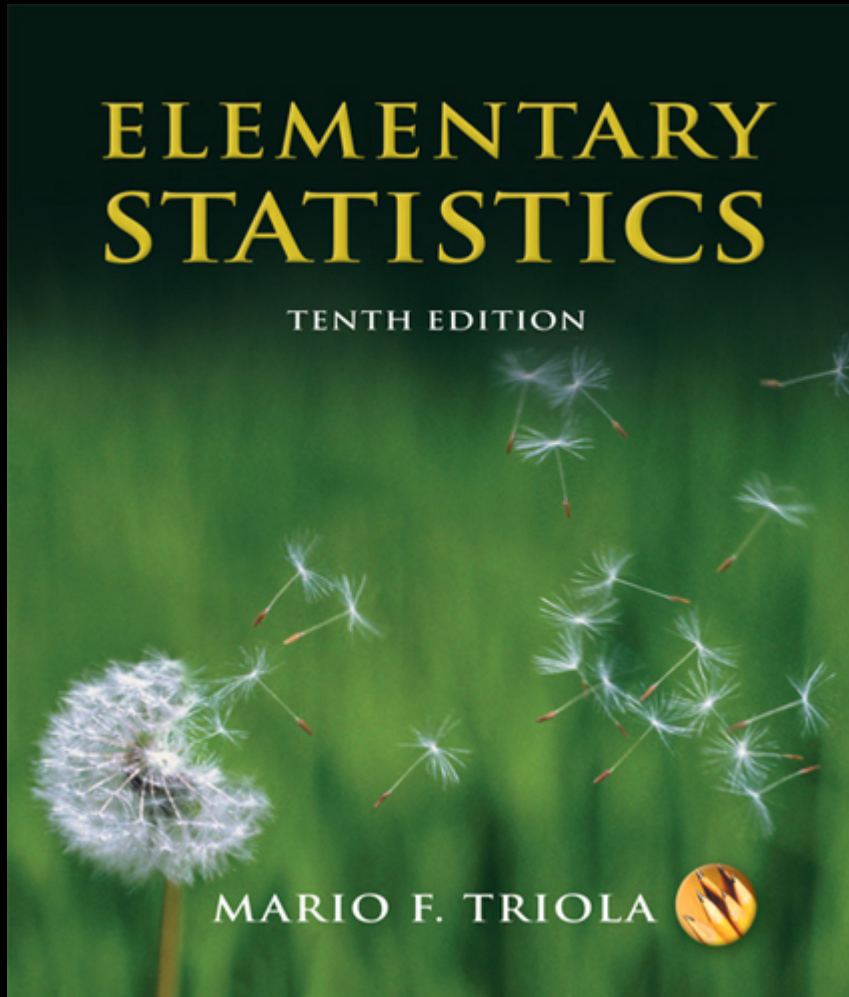# Lecture Slides

Elementary Statistics
Tenth Edition

and the Triola Statistics Series

by Mario F. Triola

Chapter 6
Normal Probability Distributions
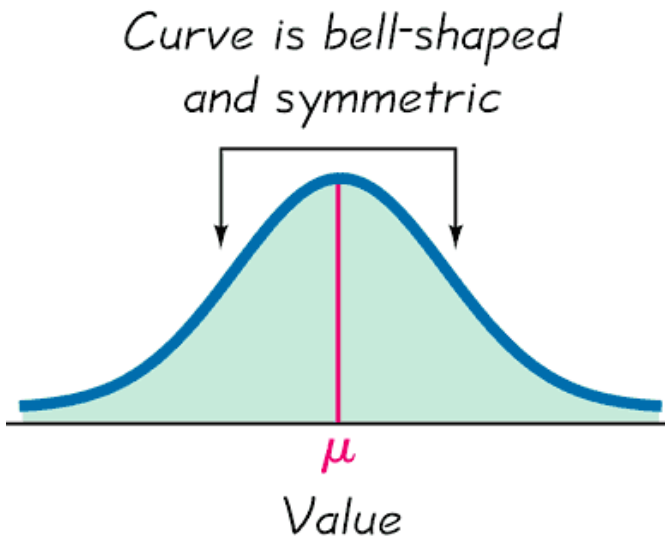
Section 6-1 Overview

Created by Erin Hodgess, Houston, Texas
Revised to accompany 10th Edition, Tom Wegleitner, Centreville, VA

Overview

Chapter focus is on:

- ❖ **Continuous random variables**

- ❖ **Normal distributions**

Curve is bell-shaped and symmetric

μ

Value

Figure 6-1

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

Formula 6-1

Section 6-2
The Standard Normal Distribution

Created by Erin Hodgess, Houston, Texas
Revised to accompany 10th Edition, Tom Wegleitner, Centreville, VA

Key Concept

This section presents the standard normal distribution which has three properties:

1. It is bell-shaped.

2. It has a mean equal to 0.

3. It has a standard deviation equal to 1.

It is extremely important to develop the skill to find areas (or probabilities or relative frequencies) corresponding to various regions under the graph of the standard normal distribution.

Definition

❖ **A continuous random variable has a uniform distribution if its values spread evenly over the range of probabilities. The graph of a uniform distribution results in a rectangular shape.**

Definition

❖ **A density curve** is the graph of a continuous probability distribution. It must satisfy the following properties:

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. (That is, the curve cannot fall below the *x*-axis.)

Area and Probability

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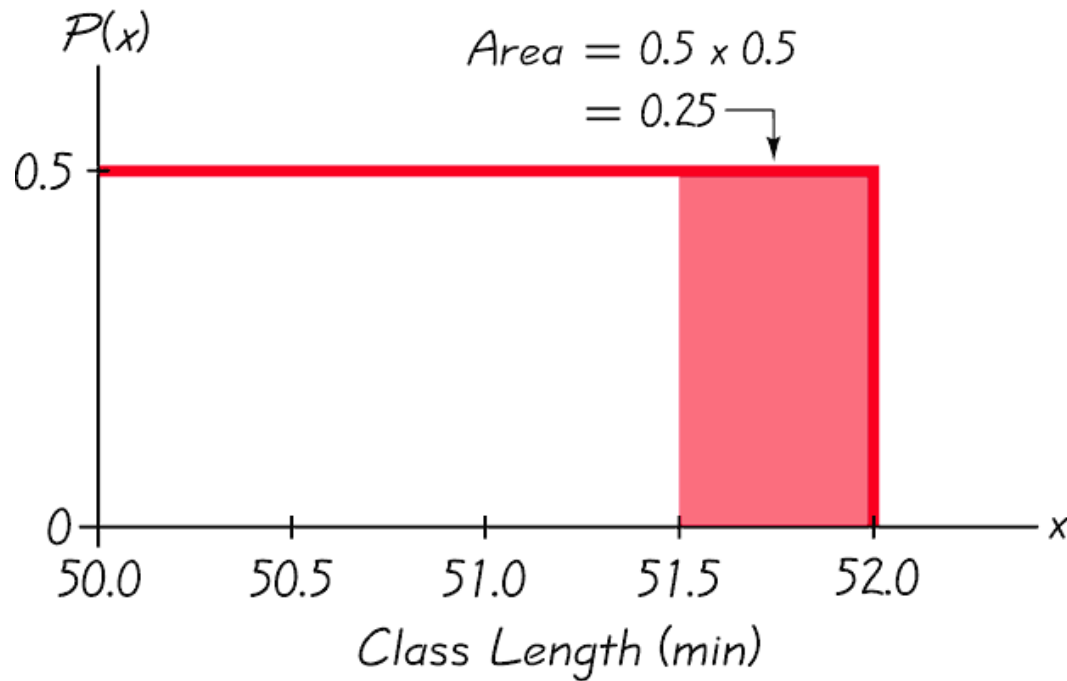
Using Area to Find Probability
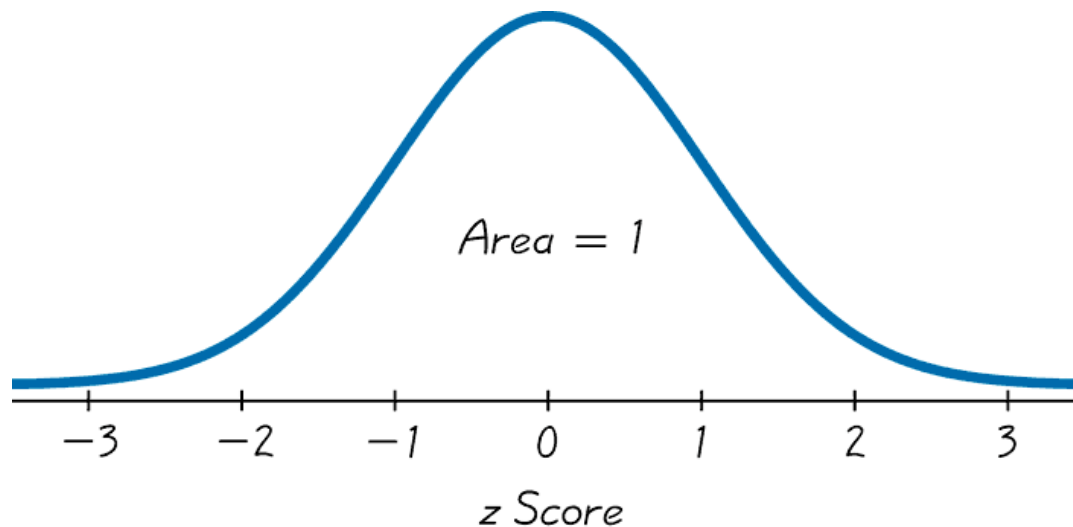

Figure 6-3

Definition

❖ **The standard normal distribution is a probability distribution with mean equal to 0 and standard deviation equal to 1, and the total area under its density curve is equal to 1.**

Area = 1

−3 −2 −1 0 1 2 3

z Score

Finding Probabilities - Table A-2

❖ **Inside back cover of textbook**

❖ **Formulas and Tables card**

❖ **Appendix**

Finding Probabilities – Other Methods

❖ **STATDISK**

❖ **Minitab**

❖ **Excel**

❖ **TI-83/84**

Table A-2 - Example

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

Using Table A-2

z Score

Distance along horizontal scale of the standard normal distribution; refer to the <u>leftmost column and top row</u> of Table A-2.

Area

Region under the curve; refer to the values in the <u>body</u> of Table A-2.

Example - Thermometers

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.
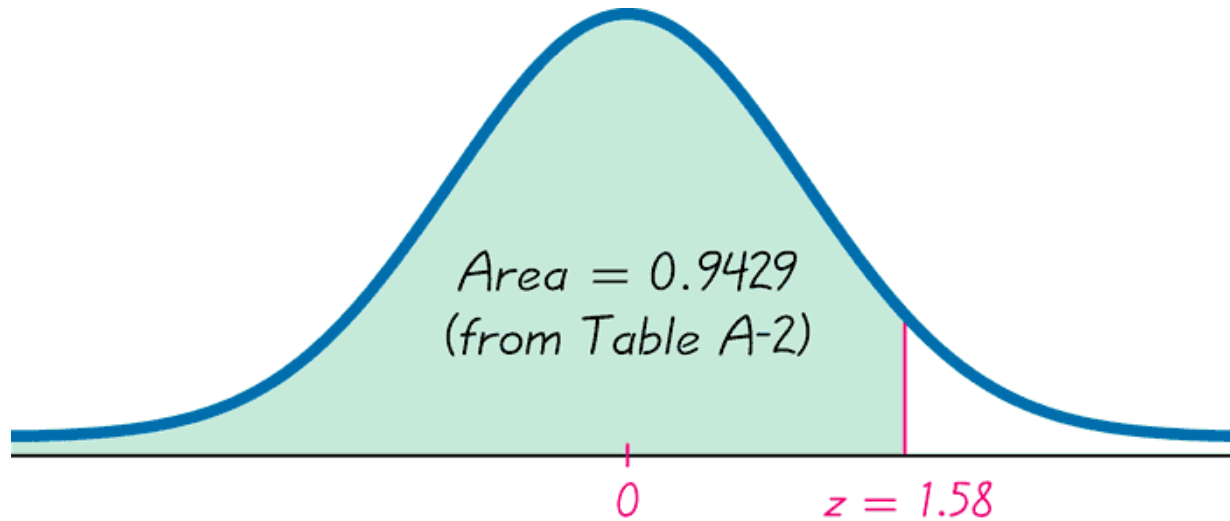
Example - Cont

P(z < 1.58) =

Area = 0.9429
(from Table A-2)

0

z = 1.58

Figure 6-6

Look at Table A-2

TABLE A-2	(*continued*) Cumulative Area from the LEFT								
z	.00	.01	.02	.03	.04	.05	.06	.07	.08	.09
0.0	.5000	.5040	.5080	.5120	.5160	.5199	.5239	.5279	.5319	.5359
0.1	.5398	.5438	.5478	.5517	.5557	.5596	.5636	.5675	.5714	.5753
0.2	.5793	.5832	.5871	.5910	.5948	.5987	.6026	.6064	.6103	.6141
0.3	.6179	.6217	.6255	.6293	.6331	.6368	.6406	.6443	.6480	.6517
0.4	.6554	.6591	.6628	.6664	.6700	.6736	.6772	.6808	.6844	.6879
0.5	.6915	.6950	.6985	.7019	.7054	.7088	.7123	.7157	.7190	.7224
0.6	.7257	.7291	.7324	.7357	.7389	.7422	.7454	.7486	.7517	.7549
0.7	.7580	.7611	.7642	.7673	.7704	.7734	.7764	.7794	.7823	.7852
0.8	.7881	.7910	.7939	.7967	.7995	.8023	.8051	.8078	.8106	.8133
0.9	.8159	.8186	.8212	.8238	.8264	.8289	.8315	.8340	.8365	.8389
1.0	.8413	.8438	.8461	.8485	.8508	.8531	.8554	.8577	.8599	.8621
1.1	.8643	.8665	.8686	.8708	.8729	.8749	.8770	.8790	.8810	.8830
1.2	.8849	.8869	.8888	.8907	.8925	.8944	.8962	.8980	.8997	.9015
1.3	.9032	.9049	.9066	.9082	.9099	.9115	.9131	.9147	.9162	.9177
1.4	.9192	.9207	.9222	.9236	.9251	.9265	.9279	.9292	.9306	.9319
1.5	.9332	.9345	.9357	.9370	.9382	.9394	.9406	.9418	.9429	.9441
1.6	.9452	.9463	.9474	.9484	.9495	.9505	.9515	.9525	.9535	.9545
1.7	.9554	.9564	.9573	.9582	.9591	.9599	.9608	.9616	.9625	.9633
1.8	.9641	.9649	.9656	.9664	.9671	.9678	.9686	.9693	.9699	.9706
1.9	.9713	.9719	.9726	.9732	.9738	.9744	.9750	.9756	.9761	.9767
2.0	.9772	.9778	.9783	.9788	.9793	.9798	.9803	.9808	.9812	.9817
2.1	.9821	.9826	.9830	.9834	.9838	.9842	.9846	.9850	.9854	.9857
2.2	.9861	.9864	.9868	.9871	.9875	.9878	.9881	.9884	.9887	.9890

Example - cont

P (*z* < 1.58) = 0.9429

Area = 0.9429
(from Table A-2)

0   z = 1.58
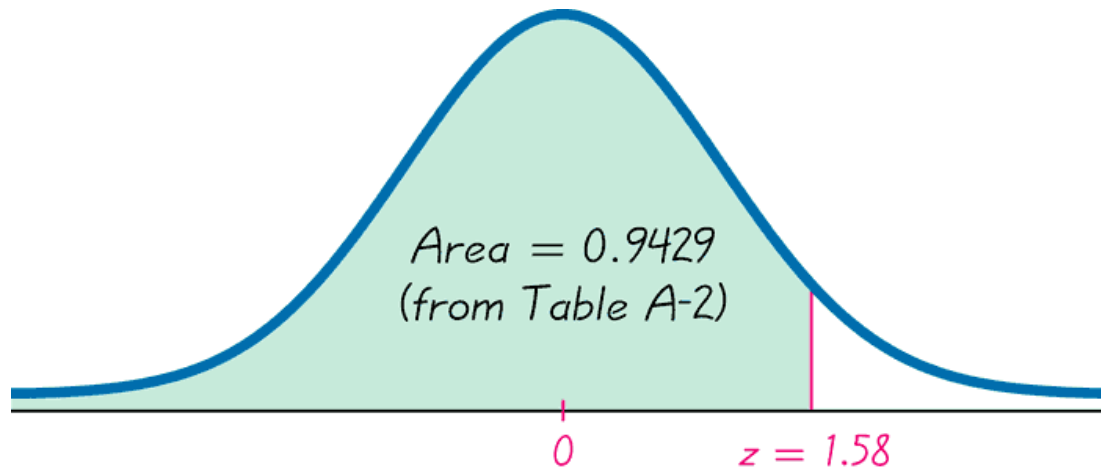
Figure 6-6

Example - cont

P (z < 1.58) = 0.9429

Area = 0.9429
(from Table A-2)

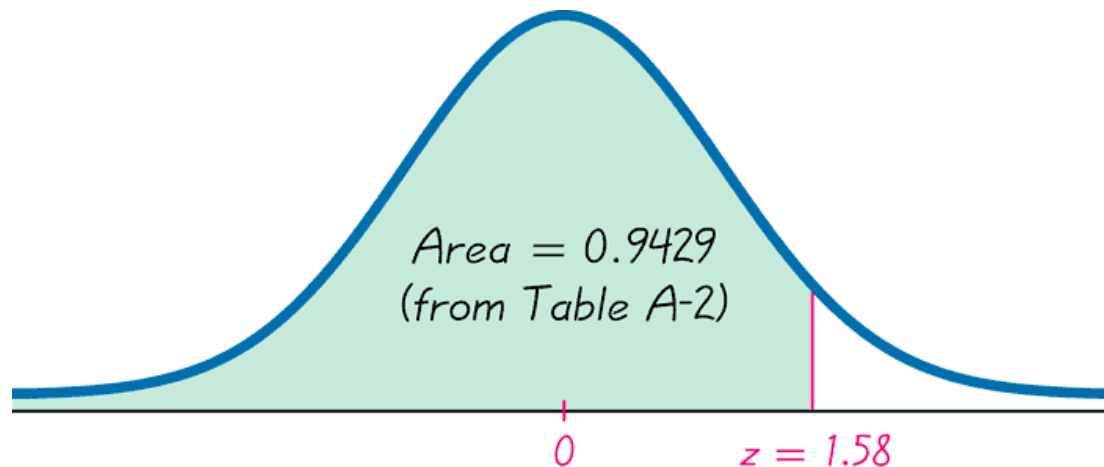0   z = 1.58
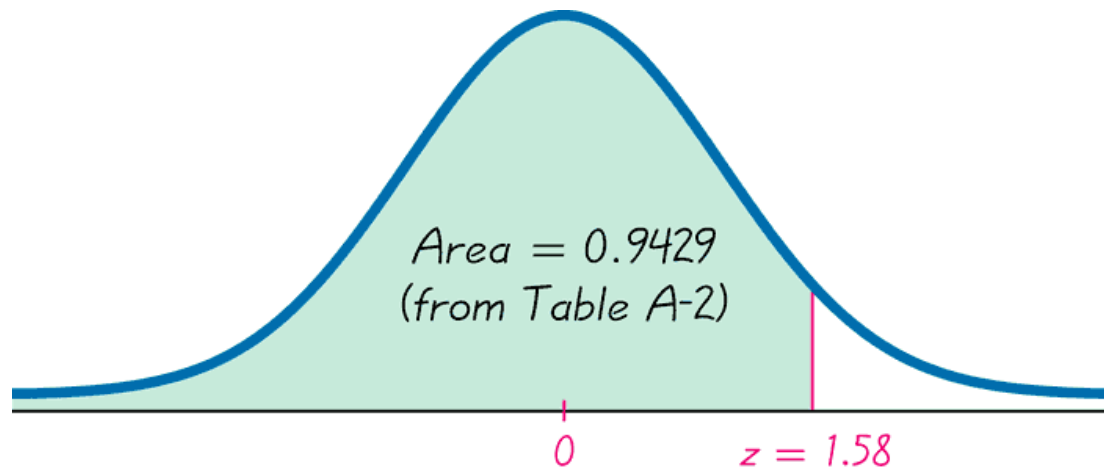
The probability that the chosen thermometer will measure freezing water less than 1.58 degrees is 0.9429.
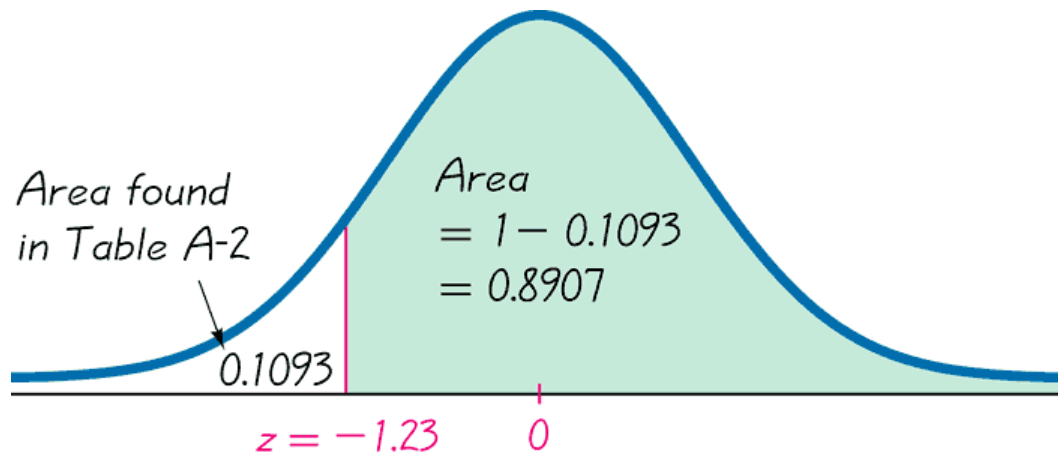
Example - cont

P (z < 1.58) = 0.9429

Area = 0.9429
(from Table A-2)

0   z = 1.58

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

Example - cont

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.
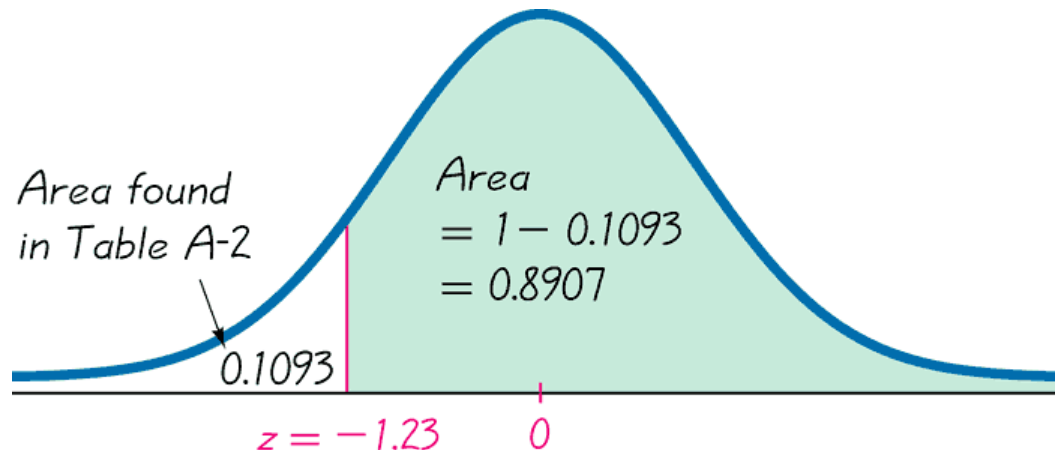
P (z > –1.23) = 0.8907

Area found in Table A-2

Area
= 1 – 0.1093
= 0.8907

0.1093

$z = -1.23 \quad 0$

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

Example - cont

P (z > −1.23) = 0.8907

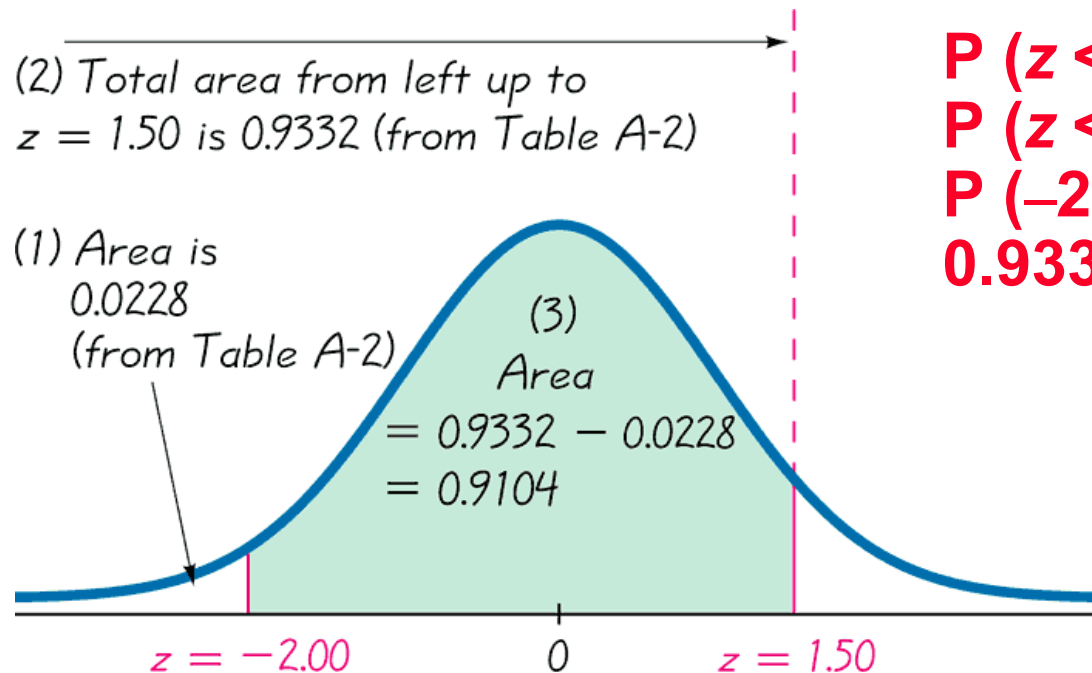Area found in Table A-2

0.1093

Area = 1− 0.1093 = 0.8907

z = −1.23   0

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

Example - cont

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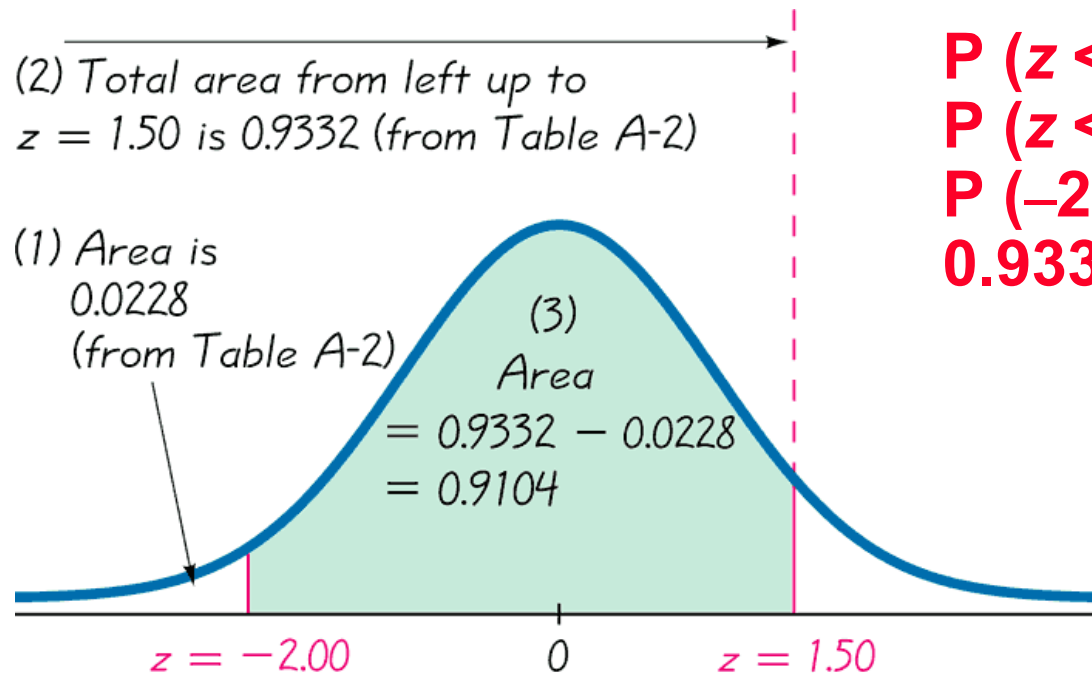

P ($z < $–2.00) = 0.0228
P ($z < $ 1.50) = 0.9332
P (–2.00 $< z < $ 1.50) = 0.9332 – 0.0228 = 0.9104

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

Example - Modified

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

(2) Total area from left up to $z = 1.50$ is 0.9332 (from Table A-2)

(1) Area is 0.0228 (from Table A-2)

(3) Area $= 0.9332 - 0.0228 = 0.9104$

$z = -2.00$ 0 $z = 1.50$

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

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

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
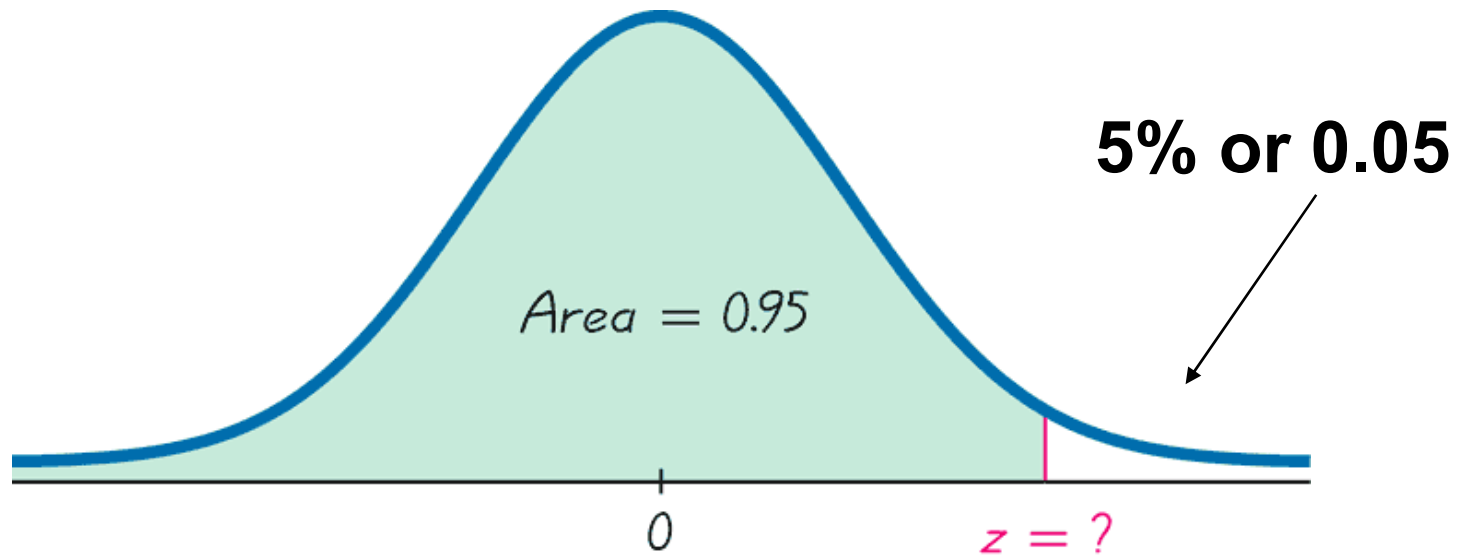
5% or 0.05

Area = 0.95

0

z = ?

(*z* score will be positive)

Figure 6-10
Finding the 95th Percentile

Finding *z* Scores
When Given Probabilities - cont

5% or 0.05

Area = 0.95

0
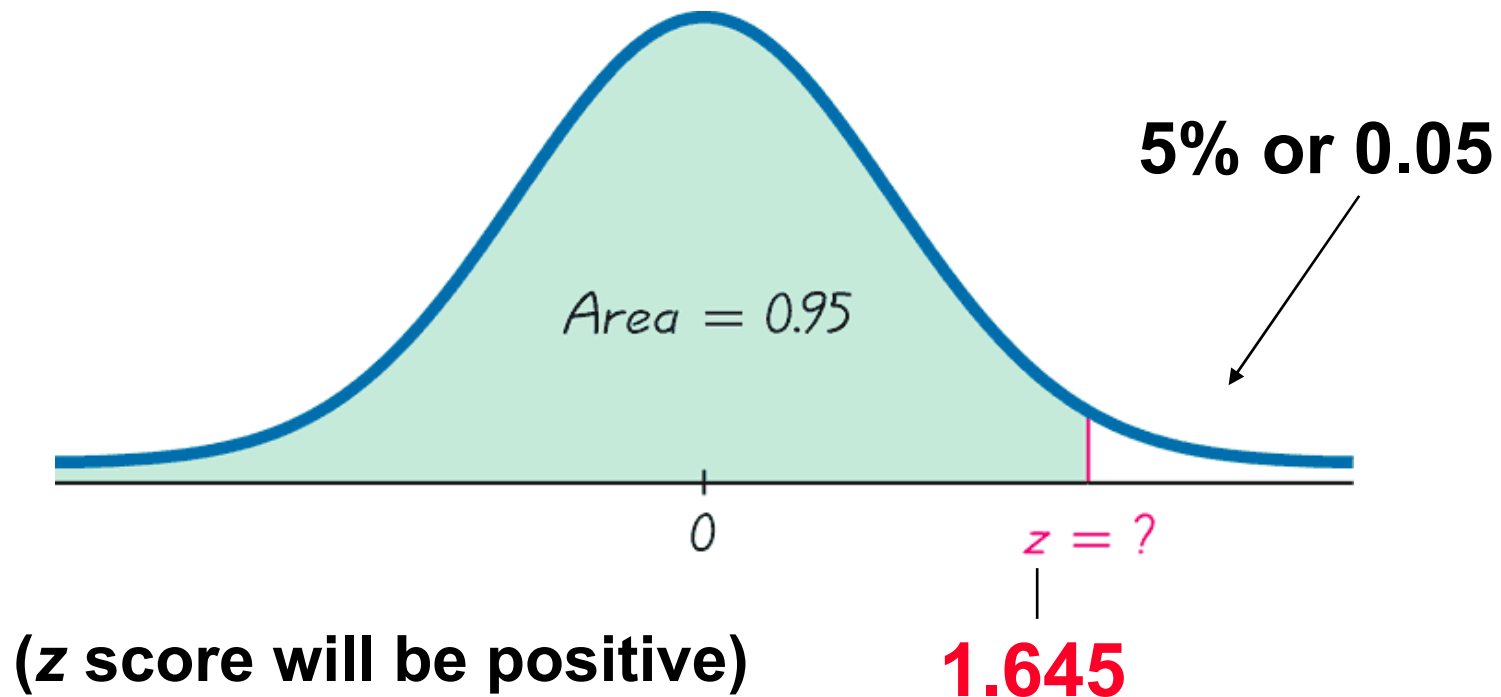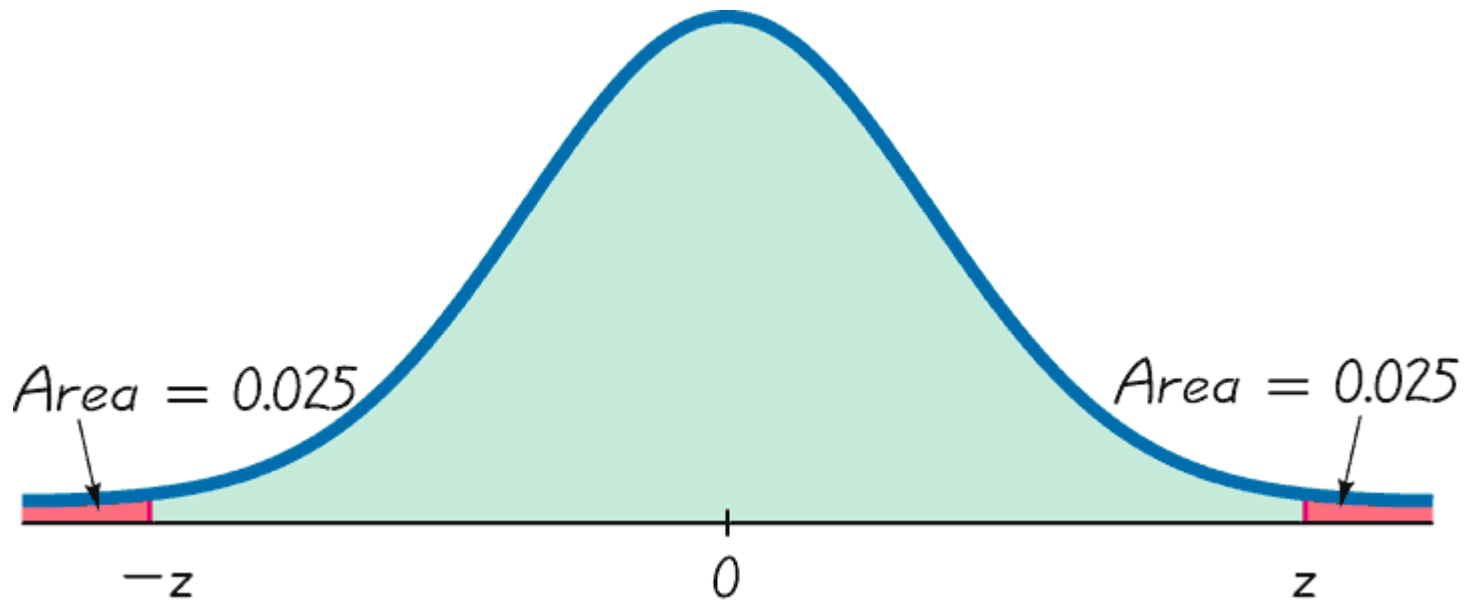
z = ?

1.645

(*z* score will be positive)

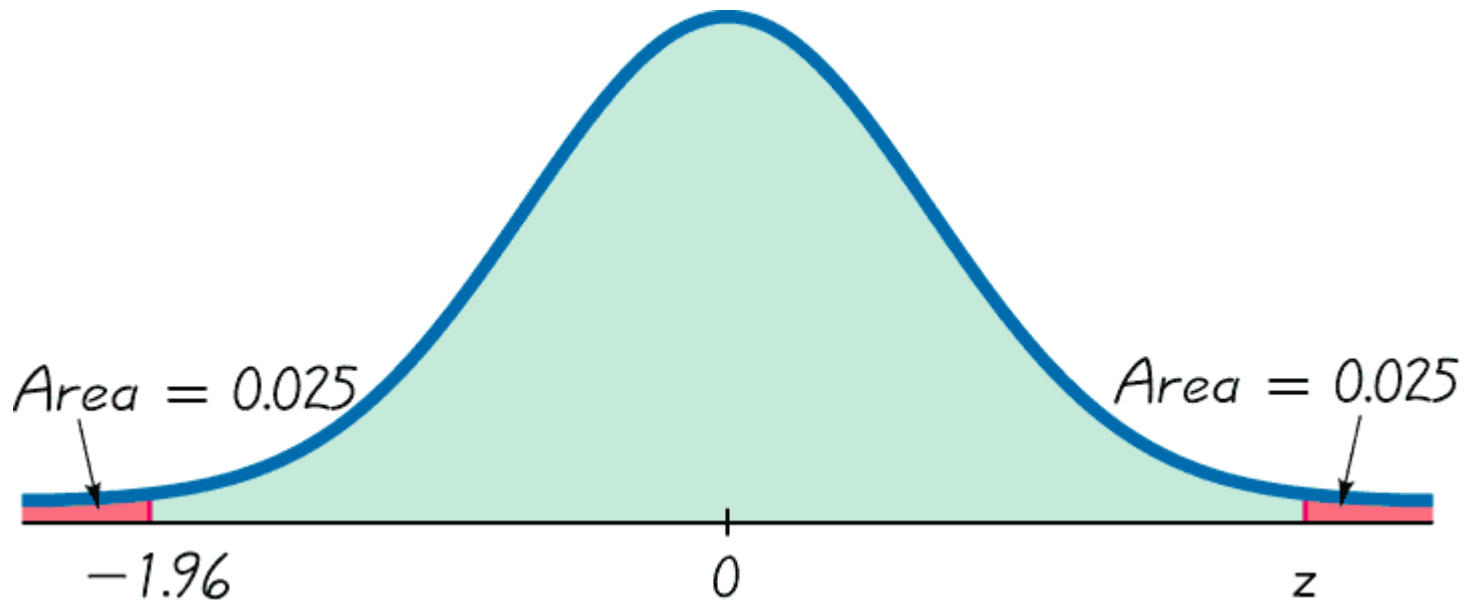**Figure 6-10**
Finding the 95th Percentile

Finding *z* Scores
When Given Probabilities - cont

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

Figure 6-11
Finding the Bottom 2.5% and Upper 2.5%

Slide 30

Finding z Scores
When Given Probabilities - cont

(One z score will be negative and the other positive)
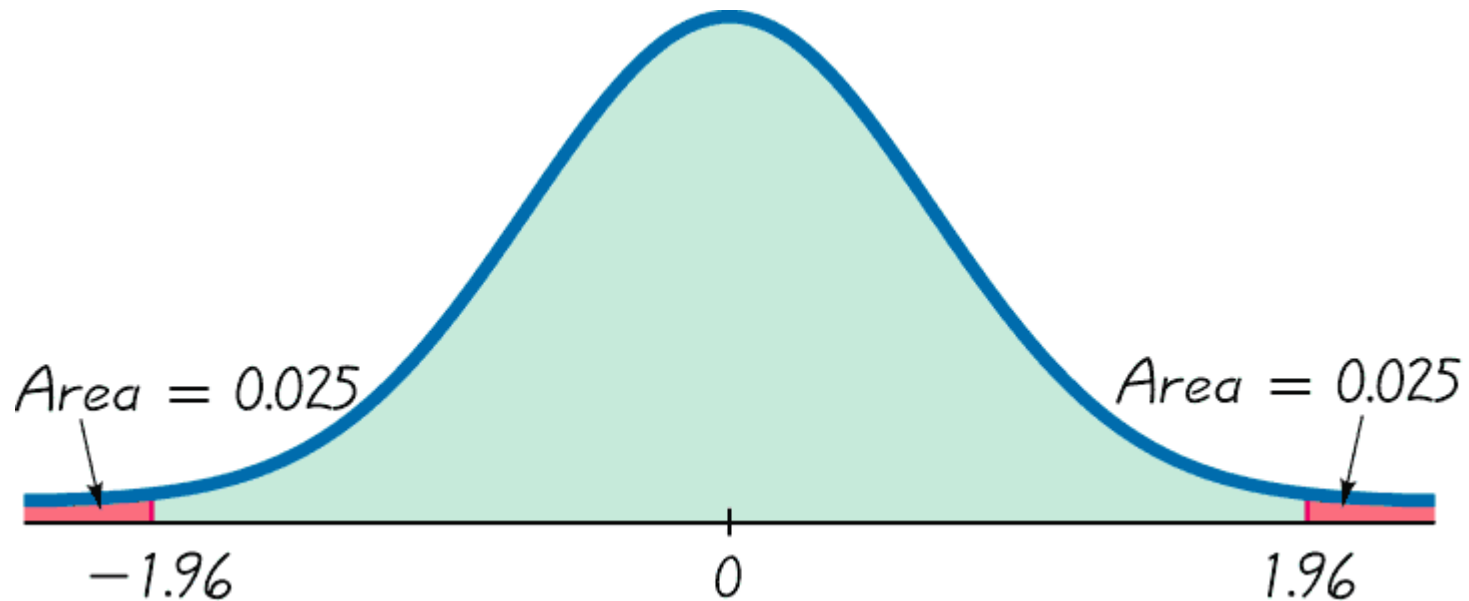
Figure 6-11
Finding the Bottom 2.5% and Upper 2.5%

Finding *z* Scores
When Given Probabilities - cont

Area = 0.025   Area = 0.025

−1.96   0   1.96

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

Figure 6-11
Finding the Bottom 2.5% and Upper 2.5%

Recap

In this section we have discussed:

❖ **Density curves.**

❖ **Relationship between area and probability**

❖ **Standard normal distribution.**

❖ **Using Table A-2.**

Section 6-3
Applications of Normal Distributions

Created by Erin Hodgess, Houston, Texas
Revised to accompany 10th Edition, Tom Wegleitner, Centreville, VA

Key Concept

This section presents methods for working with normal distributions that are not standard. That is, the mean is not 0 or the standard deviation is not 1, or both.

The key concept is that we can use a simple conversion that allows us to standardize any normal distribution so that the same methods of the previous section can be used.

Conversion Formula

Formula 6-2

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

Round *z* scores to 2 decimal places

Converting to a Standard Normal Distribution

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

(a) Nonstandard Normal Distribution
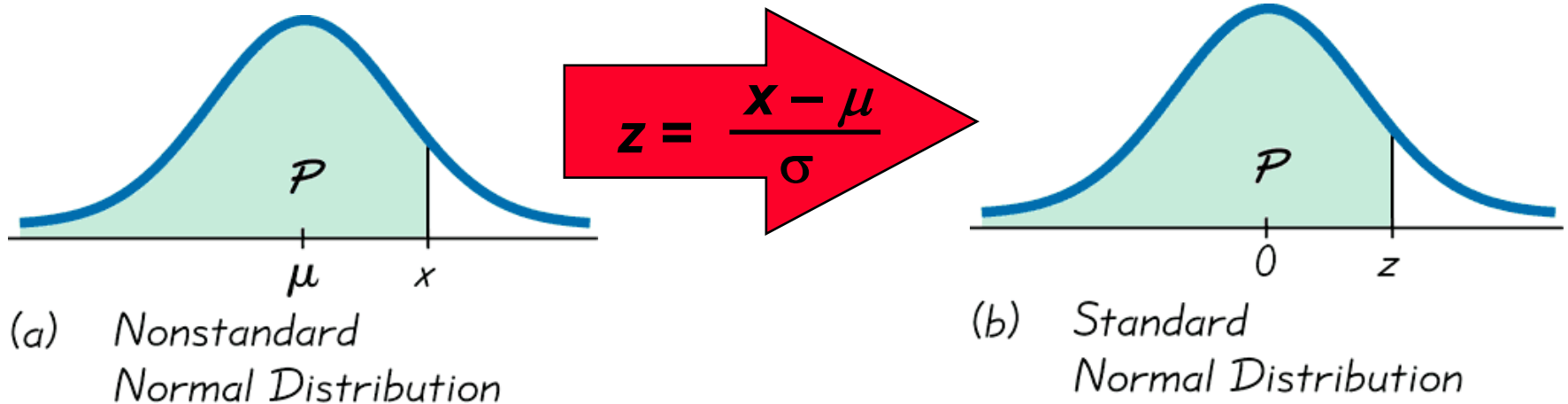
(b) Standard Normal Distribution

Figure 6-12

Example – Weights of Water Taxi Passengers

In the Chapter Problem, we noted that the safe load for a water taxi was found to be 3500 pounds. We also noted that the mean weight of a passenger was assumed to be 140 pounds. Assume the worst case that all passengers are men. Assume also that the weights of the men are normally distributed with a mean of 172 pounds and standard deviation of 29 pounds. If one man is randomly selected, what is the probability he weighs less than 174 pounds?

Example - cont

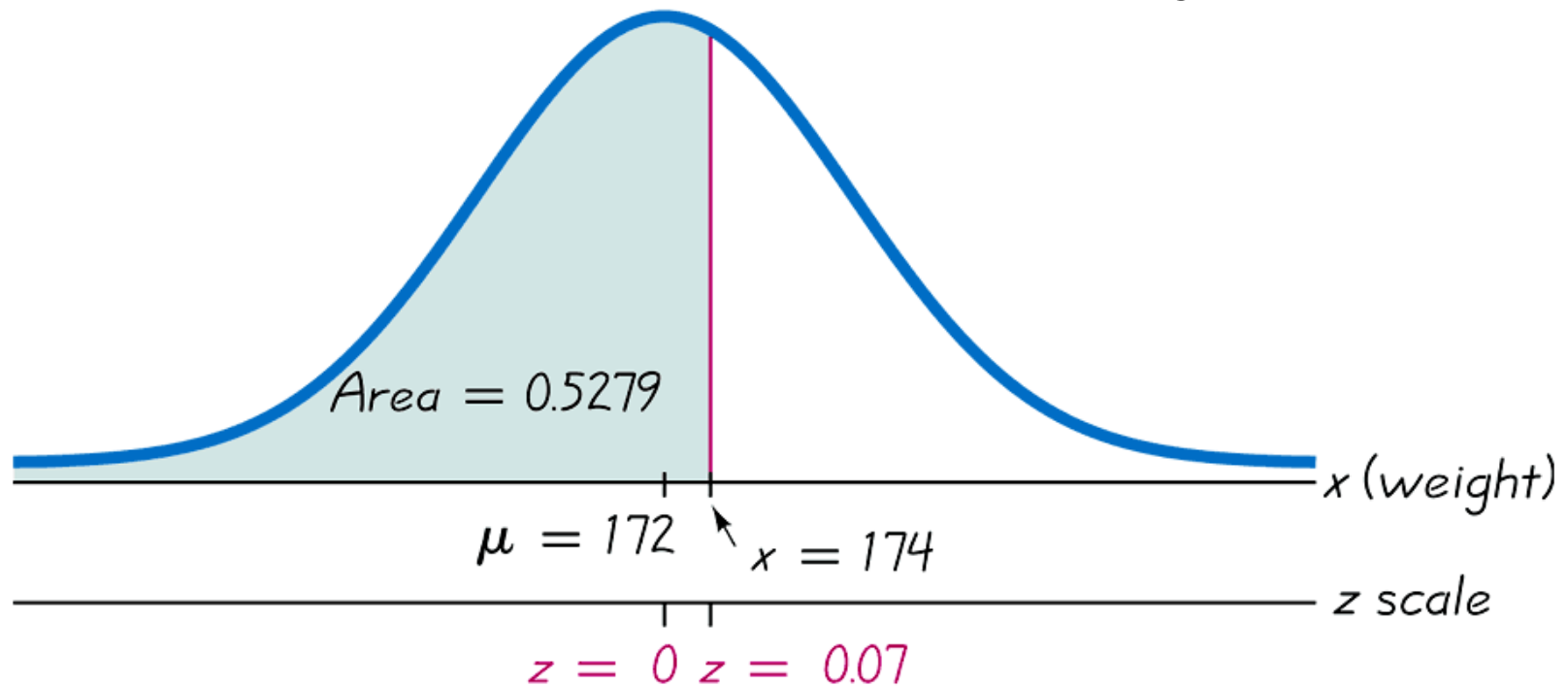$\mu = 172$

$\sigma = 29$

$$z = \frac{174 - 172}{29} = 0.07$$

Area = 0.5279

$\mu = 172$ $x = 174$

x (weight)

z scale

$z = 0$ $z = 0.07$

Figure 6-13

Example - cont

$\mu = 172$

$\sigma = 29$

P (x < 174 lb.) = P(z < 0.07)

Area = 0.5279

$\mu = 172$ $x = 174$

x (weight)

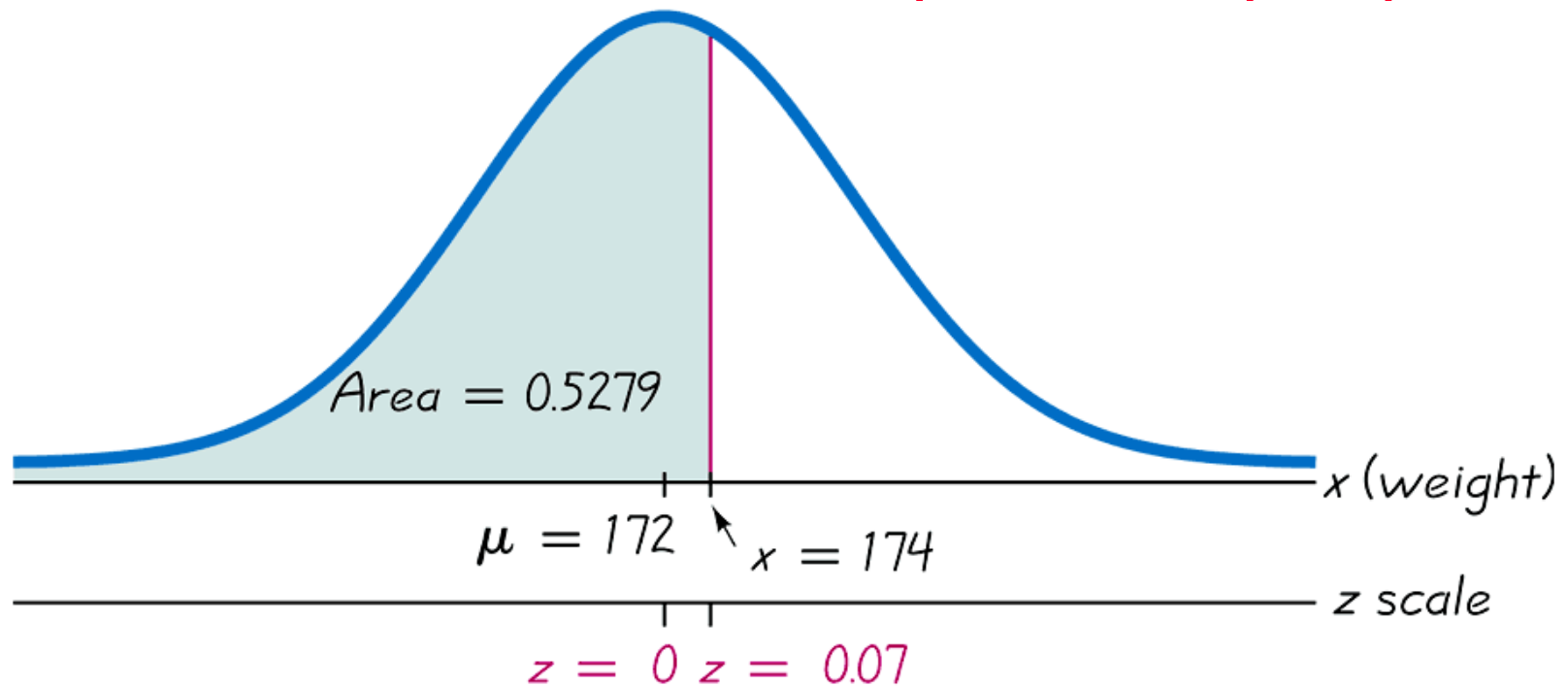z scale

$z = 0$ $z = 0.07$

Figure 6-13

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 in 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 6-2

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 6-2, enter the values for μ, σ, and the *z* score found in step 2, then solve for *x*.

   $$x = \mu + (z \cdot \sigma) \quad \text{(Another form of Formula 6-2)}$$

   (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.

Example – Lightest and Heaviest

Use the data from the previous example to determine what weight separates the lightest 99.5% from the heaviest 0.5%?
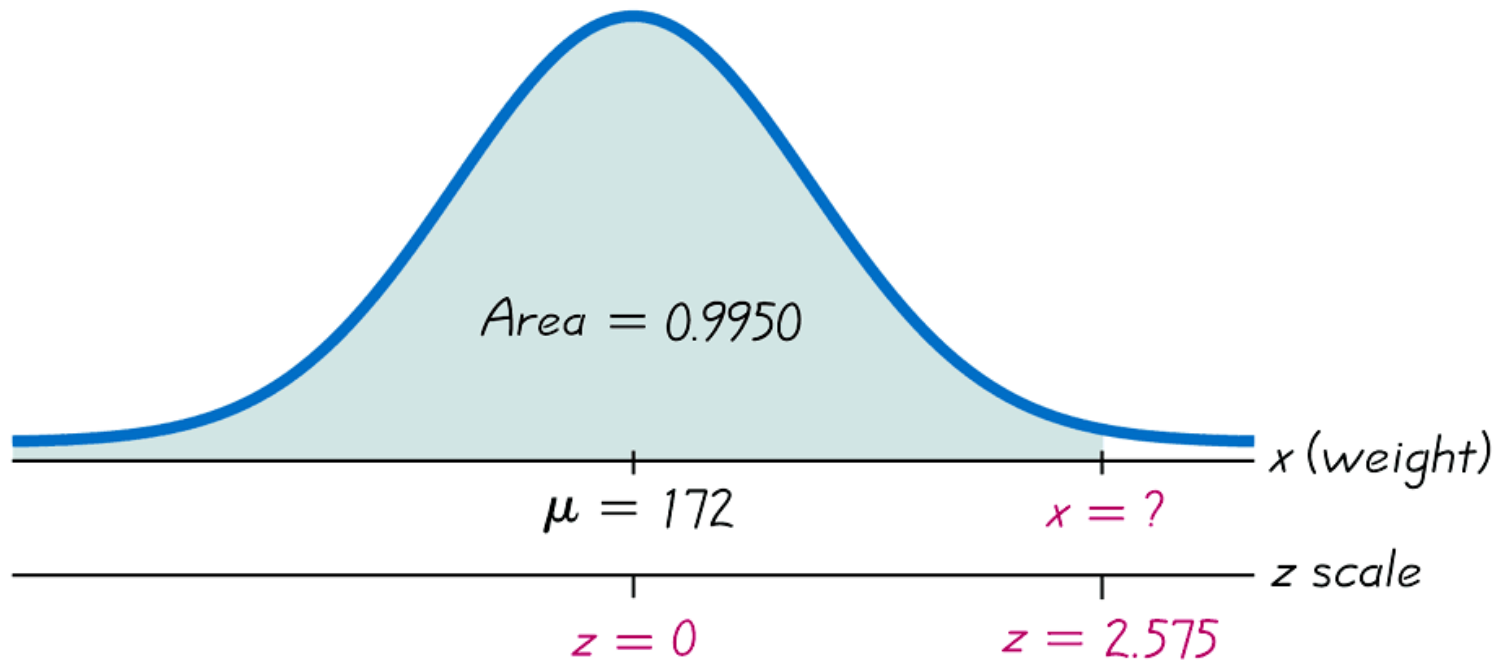
Area = 0.9950

x (weight)

$\mu = 172$

$x = ?$

z scale

$z = 0$

$z = 2.575$
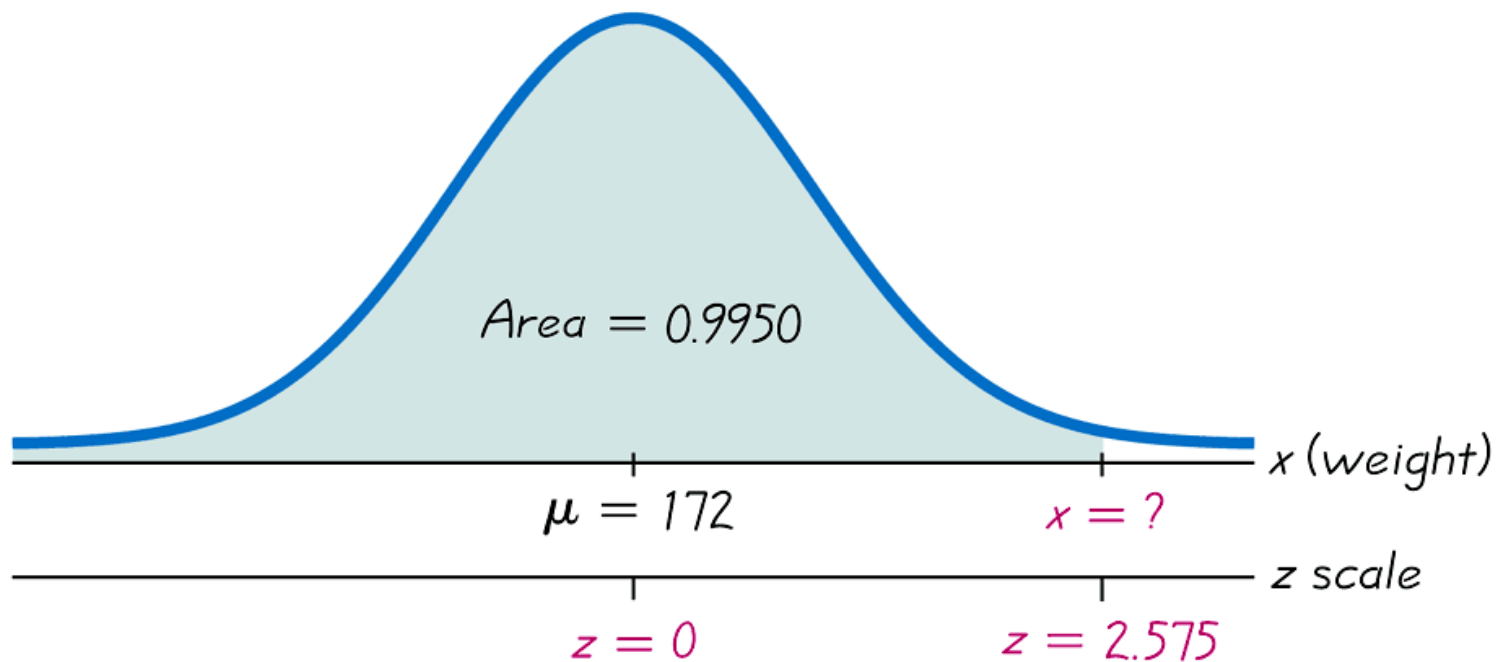
Example –
Lightest and Heaviest - cont

$x = \mu + (z \bullet \sigma)$
$x = 172 + (2.575 \bullet 29)$
$x = 246.675$ (247 rounded)

Area = 0.9950

$\mu = 172$ $x = ?$ x (weight)

                           z scale

$z = 0$ $z = 2.575$

Example – Lightest and Heaviest - cont
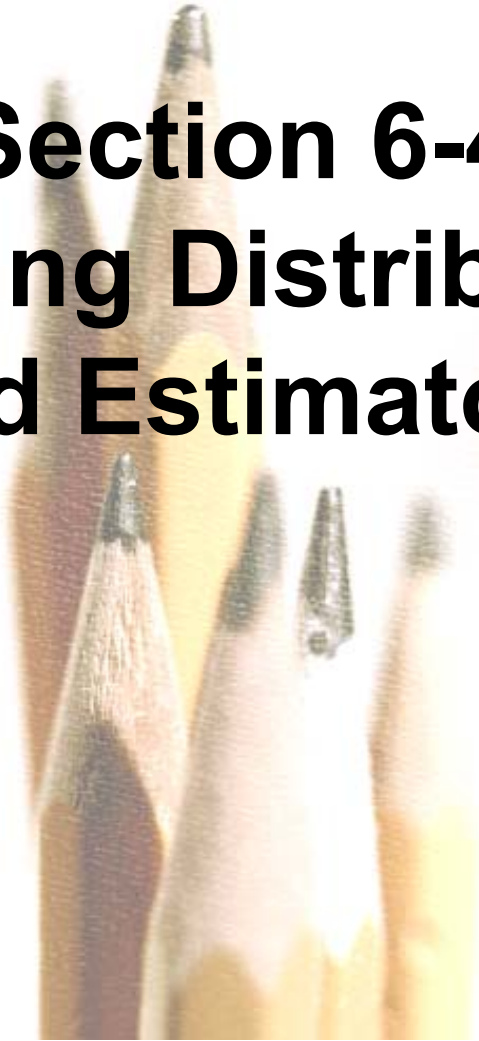
The weight of 247 pounds separates the lightest 99.5% from the heaviest 0.5%

Area = 0.9950

$\mu = 172$ x = ?   x (weight)

z scale

z = 0   z = 2.575
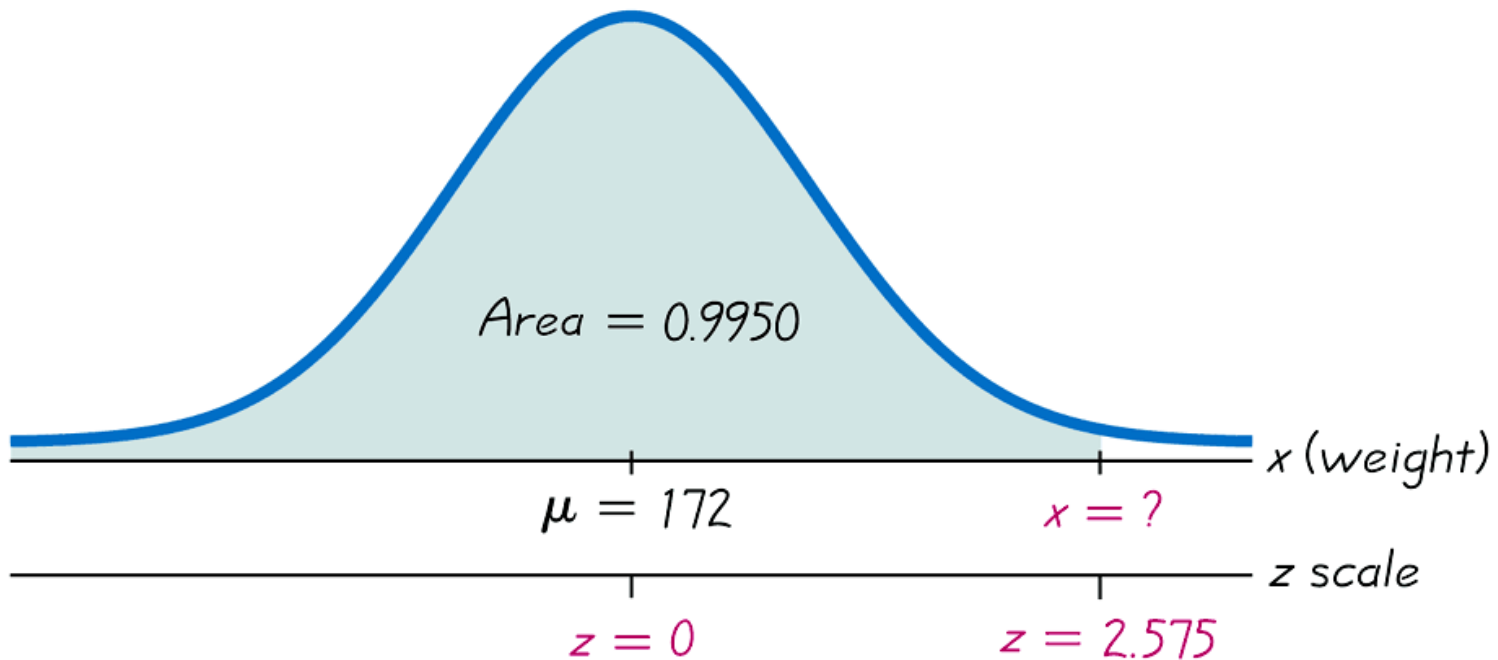
Recap

In this section we have discussed:

❖ **Non-standard normal distribution.**

❖ **Converting to a standard normal distribution.**

❖ **Procedures for finding values using Table A-2 and Formula 6-2.**

Section 6-4
Sampling Distributions and Estimators

Created by Erin Hodgess, Houston, Texas
Revised to accompany 10th Edition, Tom Wegleitner, Centreville, VA

Key Concept

The main objective of this section is to understand the concept of a sampling distribution of a statistic, which is the distribution of all values of that statistic when all possible samples of the same size are taken from the same population.

We will also see that some statistics are better than others for estimating population parameters.

Definition

❖ **The sampling distribution of a statistic (such as the sample proportion or sample mean) is the distribution of all values of the statistic when all possible samples of the same size *n* are taken from the same population.**

Definition

❖ **The sampling distribution of a proportion is the distribution of sample proportions, with all samples having the same sample size *n* taken from the same population.**

Properties

❖ **Sample proportions tend to target the value of the population proportion. (That is, all possible sample proportions have a mean equal to the population proportion.)**

❖ **Under certain conditions, the distribution of the sample proportion can be approximated by a normal distribution.**

Definition

❖ **The sampling distribution of the mean is the distribution of sample means, with all samples having the same sample size *n* taken from the same population. (The sampling distribution of the mean is typically represented as a probability distribution in the format of a table, probability histogram, or formula.)**

Definition

❖ **The value of a statistic, such as the sample mean $\overline{x}$, depends on the particular values included in the sample, and generally varies from sample to sample. This variability of a statistic is called sampling variability.**

Estimators

Some statistics work much better than others as estimators of the population. The example that follows shows this.

Example - 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)

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.

See Table 6-7 on the next slide.

Table 6-7   Sampling Distributions of 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	8/3	8/3	16/9	26/9	1.3	2/3	
Population Parameter	8/3	2	4	26/9	1.7	2/3	
Does the sample statistic target the population parameter?	Yes	No	No	Yes	No	Yes	

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.

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

The means are found near the bottom of Table 6-7.

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.
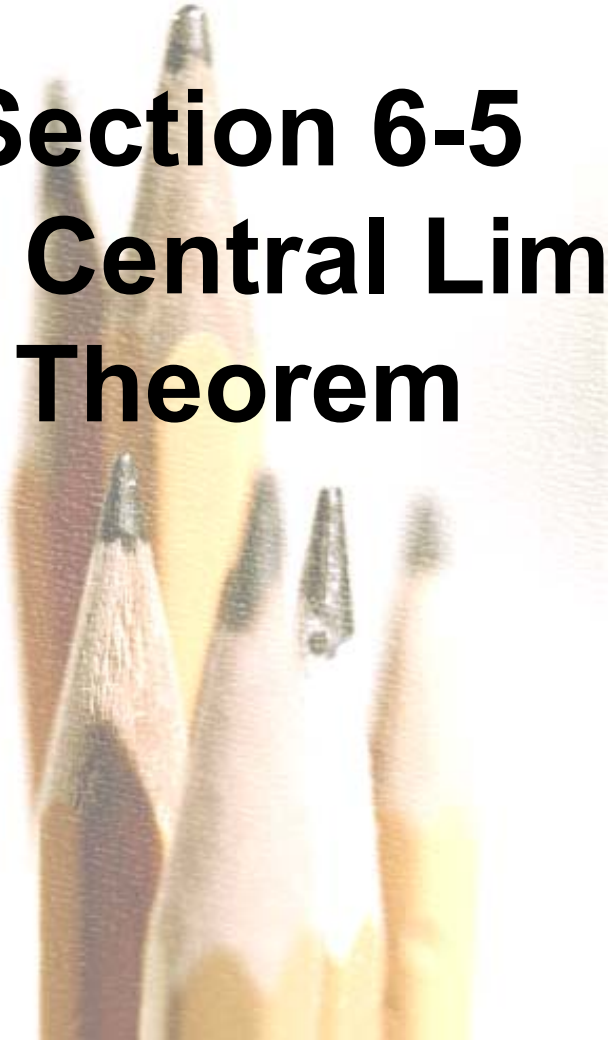
Statistics that target population parameters: mean, variance, proportion

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

Recap

In this section we have discussed:

❖ **Sampling distribution of a statistic.**

❖ **Sampling distribution of a proportion.**

❖ **Sampling distribution of the mean.**

❖ **Sampling variability.**

❖ **Estimators.**

Section 6-5
The Central Limit Theorem

Created by Erin Hodgess, Houston, Texas
Revised to accompany 10th Edition, Tom Wegleitner, Centreville, VA

PEARSON
Addison
Wesley

Key Concept

The procedures of this section form the foundation for estimating population parameters and hypothesis testing – topics discussed at length in the following chapters.

Central Limit Theorem

Given:

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

2. Simple random samples all of size *n* are selected from the population. (The samples are selected so that all possible samples of the same size *n* have the same chance of being selected.)

Central Limit Theorem - cont

Conclusions:

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

2. The mean of the sample means is the population mean μ.

3. The standard deviation of all sample means is $\sigma/\sqrt{n}$.

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 **any** sample size *n* (not just the values of *n* larger than 30).
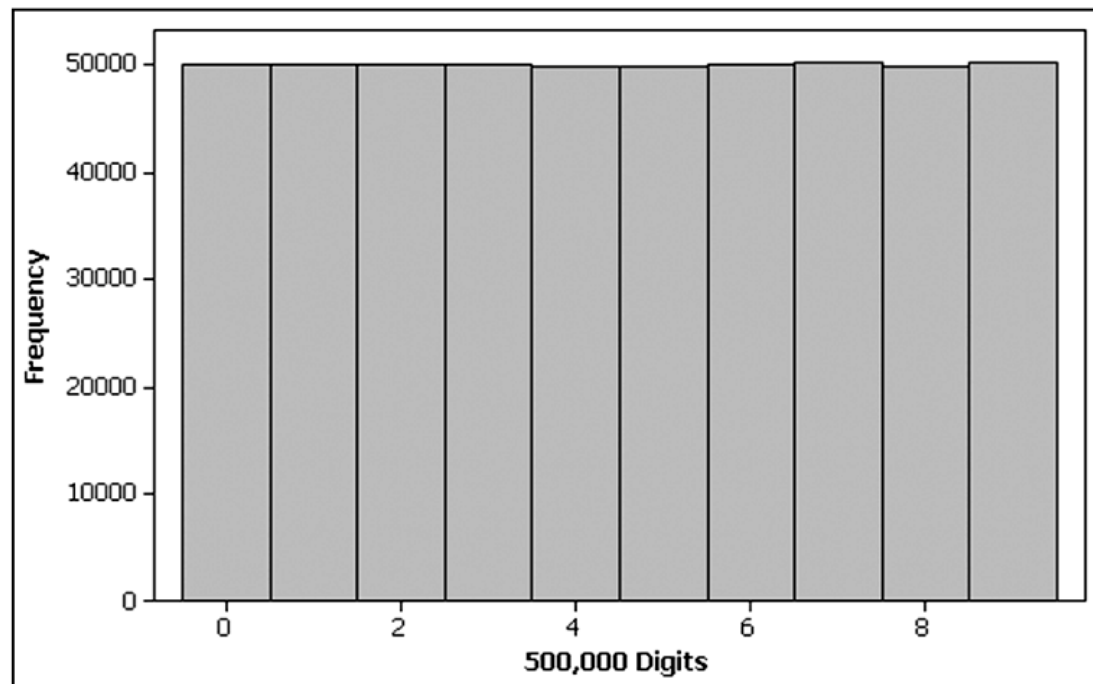
Notation

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 the standard error of the mean)

Simulation With Random Digits

Generate 500,000 random digits, group into 5000 samples of 100 each. Find the mean of each sample.

Even though the original 500,000 digits have a uniform distribution, the distribution of 5000 sample means is approximately a normal distribution!

Important Point

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

Example – Water Taxi Safety

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 175 lb.

b) if 20 different men are randomly selected, find the probability that their mean weight is greater than 175 lb (so that their total weight exceeds the safe capacity of 3500 pounds).

Example – cont

a) if one man is randomly selected, find the probability that his weight is greater than 175 lb.
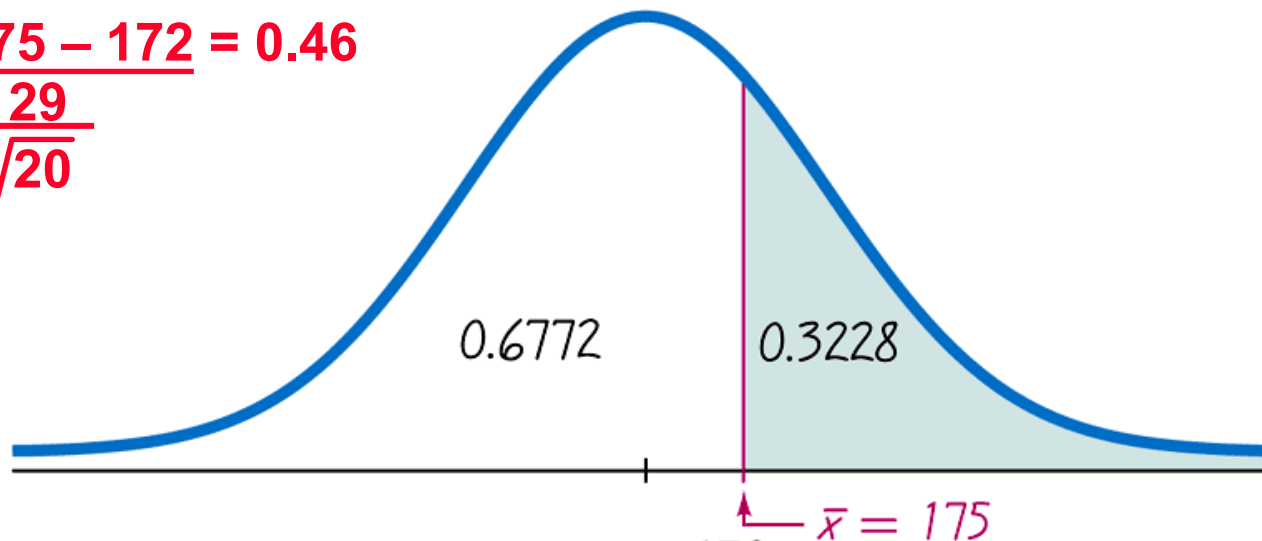
$$z = \frac{175 - 172}{29} = 0.10$$

0.5398   0.4602

$x = 175$

$\mu = 172$
$(\sigma = 29)$
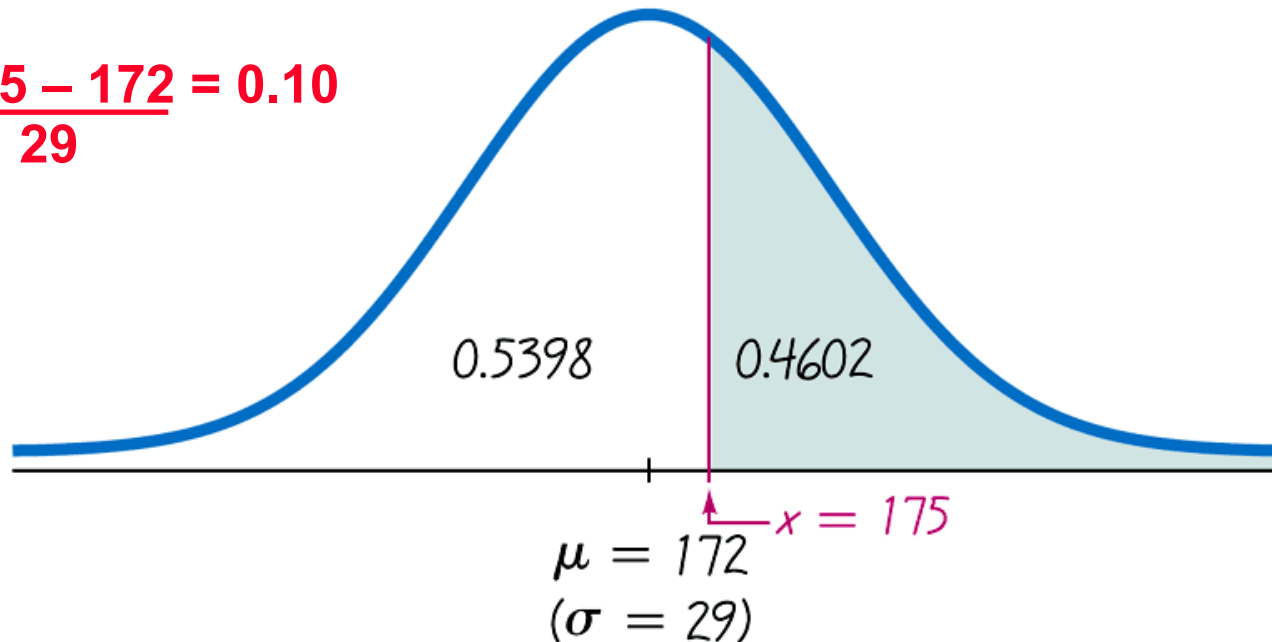
(a)

Example – cont

b) if 20 different men are randomly selected, find the probability that their mean weight is greater than 172 lb.

$$z = \frac{175 - 172}{\frac{29}{\sqrt{20}}} = 0.46$$

0.6772   0.3228

$\bar{x} = 175$

$\mu_{\bar{x}} = 172$

$\left(\sigma_{\bar{x}} = \frac{\sigma}{\sqrt{n}} = \frac{29}{\sqrt{20}} = 6.4845971 \right)$

(b)

Example - cont

a) if one man is randomly selected, find the probability that his weight is greater than 175 lb.

$$P(x > 175) = 0.4602$$

b) if 20 different men are randomly selected, their mean weight is greater than 175 lb.

$$P(\bar{x} > 175) = 0.3228$$

It is much easier for an individual to deviate from the mean than it is for a group of 20 to deviate from the mean.

Interpretation of Results

Given that the safe capacity of the water taxi is 3500 pounds, there is a fairly good chance (with probability 0.3228) that it will be overloaded with 20 randomly selected men.

Correction for a Finite Population

When sampling without replacement and the sample size *n* is greater than 5% of the finite population of size *N*, adjust the standard deviation of sample means by the following correction factor:

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

finite population correction factor

Recap

In this section we have discussed:

❖ **Central limit theorem.**

❖ **Practical rules.**

❖ **Effects of sample sizes.**

❖ **Correction for a finite population.**

Section 6-6
Normal as Approximation to Binomial

Created by Erin Hodgess, Houston, Texas
Revised to accompany 10th Edition, Tom Wegleitner, Centreville, VA

Key Concept

This section presents a method for using a normal distribution as an approximation to the binomial probability distribution.

If the conditions of $np \geq 5$ and $nq \geq 5$ are both satisfied, then probabilities from a binomial probability distribution can be approximated well by using a normal distribution with mean $\mu = np$ and standard deviation $\sigma = \sqrt{npq}$

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 probability of success remains the same in all trials.

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

Approximation of a Binomial Distribution with a Normal Distribution

$$np \geq 5$$

$$nq \geq 5$$

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

and the random variable has

a distribution.

(normal)

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 $x + 0.5$. (See **continuity corrections** discussion later in this section.) Draw a normal curve and enter the values of μ, σ, and either $x - 0.5$ or $x + 0.5$, as appropriate.

Procedure for Using a Normal Distribution to Approximate a Binomial Distribution - cont

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

5. Using *x* – 0.5 or *x* + 0.5 (as appropriate) in place of *x*, find the area corresponding to the desired probability by first finding the *z* score and finding the area to the left of the adjusted value of *x*.

Example – Number of Men Among Passengers
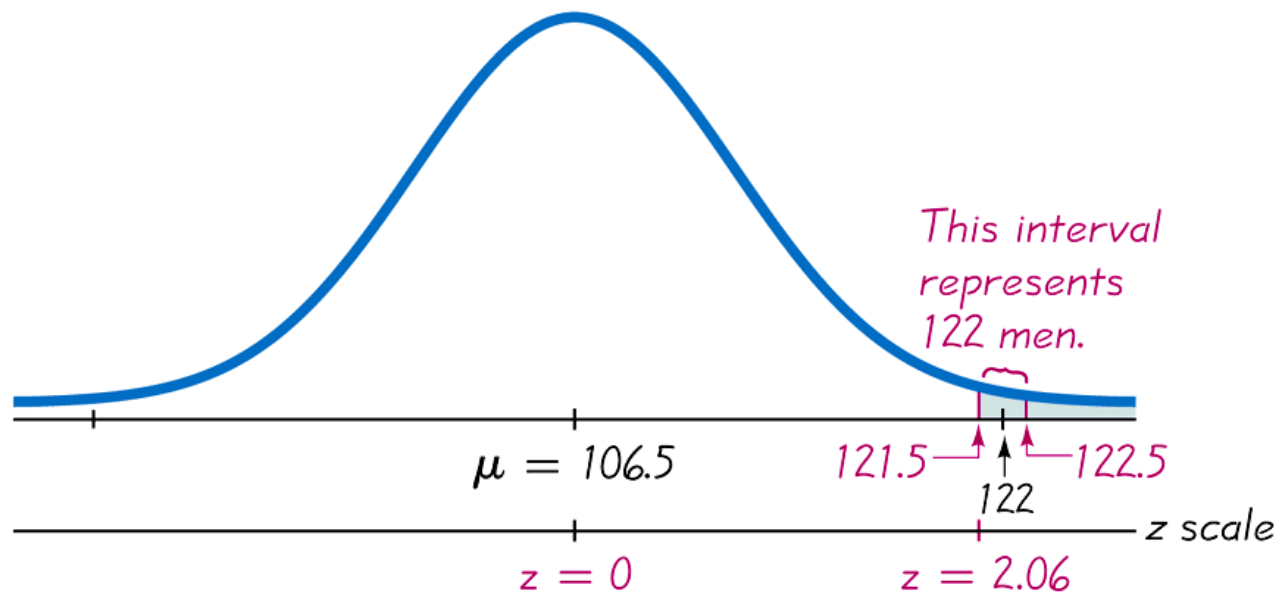
Finding the Probability of "At Least 122 Men" Among 213 Passengers

This interval represents 122 men.

$\mu = 106.5$ 121.5 122.5 122 z scale

$z = 0$ $z = 2.06$

Figure 6-21

Definition

When we use the normal distribution (which is a **continuous** probability distribution) 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

$$x - 0.5 \text{ to } x + 0.5$$

(that is, adding and subtracting 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* = 122, draw a strip from 121.5 to 122.5. **Consider the area of the strip to represent the probability of discrete whole number *x*.**

Procedure for Continuity Corrections - cont

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 the probability being sought.
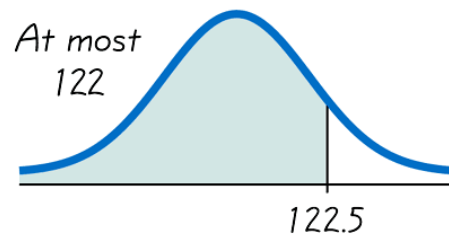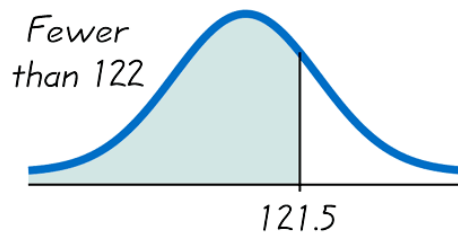
Figure 6-22
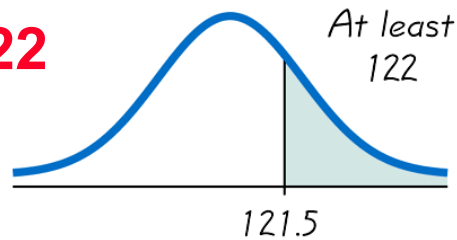
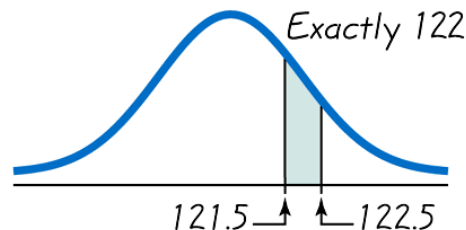*X* = <u>at least</u> 122
(includes 122 and above)
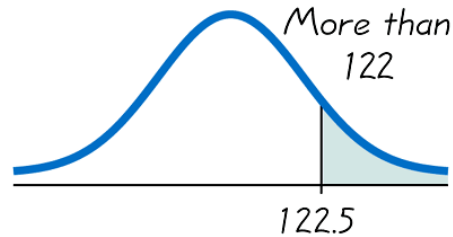
X = <u>more than</u> 122
(doesn't include 122)

X = <u>at most</u> 122
(includes 122 and below)

X = <u>fewer than</u> 122
(doesn't include 122)

X = <u>exactly</u> 122

Recap

In this section we have discussed:

❖ Approximating a binomial distribution with a normal distribution.

❖ Procedures for using a normal distribution to approximate a binomial distribution.

❖ Continuity corrections.

Section 6-7
Assessing Normality

Created by Erin Hodgess, Houston, Texas
Revised to accompany 10th Edition, Tom Wegleitner, Centreville, VA

Key Concept

This section provides criteria for determining whether the requirement of a normal distribution is satisfied.

The criteria involve visual inspection of a histogram to see if it is roughly bell shaped, identifying any outliers, and constructing a new graph called a **normal quantile plot**.

Definition

❖ A **normal quantile plot** (or **normal probability 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 the corresponding *z* score that is a quantile value expected from the standard normal distribution.

Methods 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 - cont

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, software or calculator) to find the z scores corresponding to the cumulative left areas found in Step (b).

Procedure for Determining Whether Data Have a Normal Distribution - cont

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.

Example
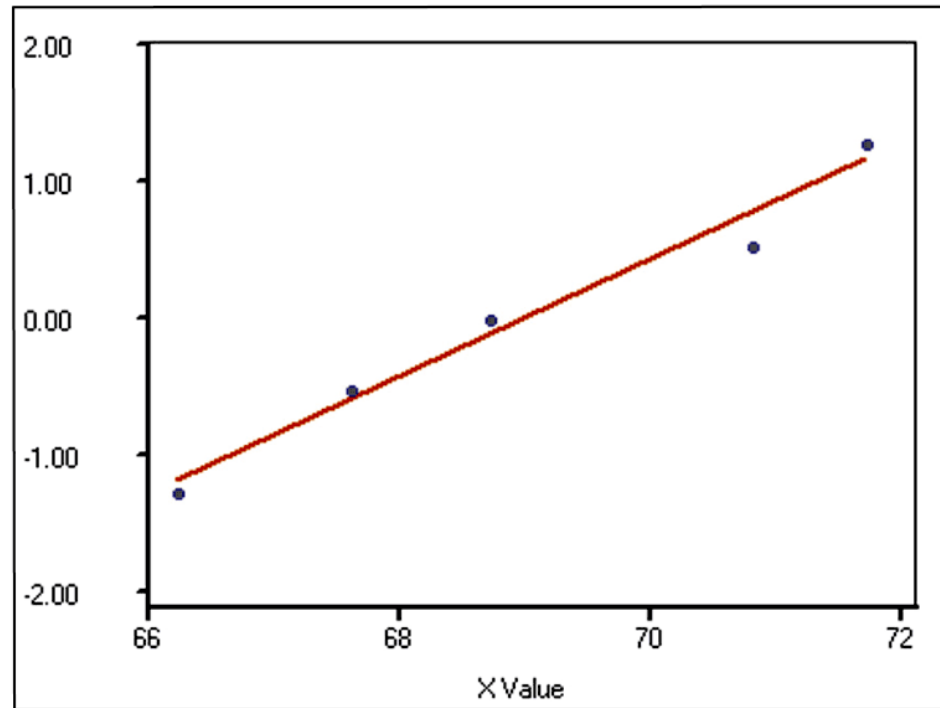
Interpretation: Because the points lie reasonably close to a straight line and there does not appear to be a systematic pattern that is not a straight-line pattern, we conclude that the sample appears to be a normally distributed population.

Recap

In this section we have discussed:

❖ **Normal quantile plot.**

❖ **Procedure to determine if data have a normal distribution.**