Section 6-2
The Standard Normal Distribution
Continuous Random Variables

- Continuous random variable
  A random variable $X$ takes infinitely many values, and those values can be associated with measurements on a continuous scale without gaps or interruptions
Density Curve

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.)
Using Area to Find Probability

Given the uniform distribution illustrated, find the probability that a randomly selected voltage level is greater than 124.5 volts.

Shaded area represents voltage levels greater than 124.5 volts.

Correspondence between area and probability: 0.25.
Area and Probability

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

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

Normal distribution represents:

- Continuous random variable
- Bell-shaped density curve

Figure 6-1
The **standard normal distribution** is a normal probability distribution with $\mu = 0$ and $\sigma = 1$. The total area under its density curve is equal to 1.
Example - Thermometers

Thermometers are supposed to give readings of 0ºC at the freezing point of water, but they are not necessarily accurate. Assume that the mean reading is 0ºC and the standard deviation of the readings is 1.00ºC. Also assume that the readings are normally distributed. If one thermometer is randomly selected, find the probability that, at the freezing point of water, the reading is less than 1.27º.
Example - Thermometers

\[ P(z < 1.27) = 0.8980 \]

The *probability* of randomly selecting a thermometer with a reading less than 1.27° is 0.8980.
Using Table A-2

1. It is designed only for the *standard* normal distribution, which has a mean of 0 and a standard deviation of 1.

2. It is on two pages, with one page for *negative* \(z\)-scores and the other page for *positive* \(z\)-scores.

3. Each value in the body of the table is a *cumulative area from the left* up to a vertical boundary above a specific \(z\)-score.
Using Table A-2

4. When working with a graph, avoid confusion between z-scores and areas.

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

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

5. The part of the z-score denoting hundredths is found across the top.
Look at Table A-2

<table>
<thead>
<tr>
<th>$z$</th>
<th>.00</th>
<th>.01</th>
<th>.02</th>
<th>.03</th>
<th>.04</th>
<th>.05</th>
<th>.06</th>
<th>.07</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.0</td>
<td>.5000</td>
<td>.5040</td>
<td>.5080</td>
<td>.5120</td>
<td>.5160</td>
<td>.5199</td>
<td>.5239</td>
<td>.5279</td>
</tr>
<tr>
<td>0.1</td>
<td>.5398</td>
<td>.5438</td>
<td>.5478</td>
<td>.5517</td>
<td>.5557</td>
<td>.5596</td>
<td>.5636</td>
<td>.5675</td>
</tr>
<tr>
<td>0.2</td>
<td>.5793</td>
<td>.5832</td>
<td>.5871</td>
<td>.5910</td>
<td>.5948</td>
<td>.5987</td>
<td>.6026</td>
<td>.6064</td>
</tr>
<tr>
<td>1.0</td>
<td>.8413</td>
<td>.8438</td>
<td>.8461</td>
<td>.8485</td>
<td>.8508</td>
<td>.8531</td>
<td>.8554</td>
<td>.8577</td>
</tr>
<tr>
<td>1.1</td>
<td>.8643</td>
<td>.8665</td>
<td>.8686</td>
<td>.8708</td>
<td>.8729</td>
<td>.8749</td>
<td>.8770</td>
<td>.8790</td>
</tr>
<tr>
<td>1.2</td>
<td>.8849</td>
<td>.8869</td>
<td>.8888</td>
<td>.8907</td>
<td>.8925</td>
<td>.8944</td>
<td>.8962</td>
<td>.8980</td>
</tr>
<tr>
<td>1.3</td>
<td>.9032</td>
<td>.9049</td>
<td>.9066</td>
<td>.9082</td>
<td>.9099</td>
<td>.9115</td>
<td>.9131</td>
<td>.9147</td>
</tr>
<tr>
<td>1.4</td>
<td>.9192</td>
<td>.9207</td>
<td>.9222</td>
<td>.9236</td>
<td>.9251</td>
<td>.9265</td>
<td>.9279</td>
<td>.9292</td>
</tr>
</tbody>
</table>
Finding probability using Excel

\[ = \text{NORMSDIST}(z) \]

It gives the cumulative area from the left up to a vertical line above the value \( z \).
Example - Thermometers (2)

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.

Probability of randomly selecting a thermometer with a reading above −1.23º is 0.8907.
Example - Thermometers (3)

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

The probability that the chosen thermometer has a reading between –2.00 and 1.50 degrees is 0.9104.
Finding “percentile” Given Probabilities

Finding the 95th Percentile

\( z = ? \)

Area = 0.95

5% or 0.05

(z score will be positive)
Finding a “percentile” given a probability using Table A-2

1. Draw a bell-shaped curve 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 “percentile” Given Probabilities

Finding the 95th Percentile

Area = 0.95

5% or 0.05

(z score will be positive)

1.645
Find “percentile” using Excel

= NORMSINV(p)

It finds the value $z$ with the area of $p$ to its left.
Definition of “critical value”

The expression

\[ Z_\alpha \]

denotes the \((1- \alpha)\)th percentile, that is, The z-score with an area of \(a\) to the left.
Finding “critical value”
Given Probabilities

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

Finding the Bottom 2.5% and Upper 2.5%
Finding “critical value”
Given Probabilities - continue

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

Finding the Bottom 2.5% and Upper 2.5%
Section 6-3
Applications of Normal Distributions
Conversion Formula

Given the mean (mu) and the standard deviation (sigma) we can convert the value x into the Z score.

\[ Z = \frac{X - \mu}{\sigma} \]

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
Example – Weights of Water Taxi Passengers

Assume 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

\[ z = \frac{174 - 172}{29} = 0.07 \]

\[ \mu = 172 \]
\[ x = 174 \]
\[ z = 0 \quad z = 0.07 \]
Finding probability using Excel

= NORMDIST(x, mean, SD, 1)

It gives the cumulative area from the left of the value x, given mean and standard deviation (SD).
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

μ = 172

x = ?

z = 0

z = 2.575
Procedure for Finding Values

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 $z$ score found in step 2, then solve for $x$.

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

(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

\[ x = 172 + (2.575)(29) \]
\[ x = 246.675 \text{ (247 rounded)} \]
Example – Lightest and Heaviest - cont

The weight of 247 pounds separates the lightest 99.5% from the heaviest 0.5%
Finding the value using Excel

\[ = \text{NORMINV}(p, \text{mean, SD}) \]

It finds the value \( x \) with the area of \( p \) to the left, given mean and standard deviation (SD).
Section 6-4
Sampling Distributions and Estimators
The sampling distribution of a statistic (such as the sample mean or sample proportion) is the distribution of all values of the statistic when all possible samples of the same size $n$ are taken from the same population. (The sampling distribution of a statistic is typically represented as a probability distribution in the format of a table, probability histogram, or formula.)
Sample Mean

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.)
Example - Sampling Distribution

Consider repeating this process: Roll a die 5 times, find the mean $\bar{x}$, variance $s^2$, and the proportion of odd numbers of the results. What do we know about the behavior of all sample means that are generated as this process continues indefinitely?
Sampling Distribution of Mean

Specific results from 10,000 trials

All outcomes are equally likely so the population mean is 3.5; the mean of the 10,000 trials is 3.49. If continued indefinitely, the sample mean will be 3.5. Also, notice the distribution is "normal."
Properties of Sample Means

- Sample means target the value of the population mean. (That is, the mean of the sample means is the population mean. The sample mean is an unbiased estimator of the population mean.)

- The distribution of the sample means tends to be a normal distribution.
Unbiased Estimators

Sample means, variances and proportions are **unbiased estimators**.

That is, the mean of the sample statistic is exactly the population parameter; thus, they target the population parameter.

These statistics are better in estimating the population parameter.
Biased Estimators

Sample medians, ranges and standard deviations are biased estimators.

That is they do NOT target the population parameter.

Note: the bias with the standard deviation is relatively small in large samples so $s$ is often used to estimate.
Sample Variance

The variance $s^2$ of a set of values is a measure of variation equal to the square of the standard deviation $s$.

$$s^2 = \frac{1}{n-1} \left[ \sum_{i=1}^{n} x_i^2 - \left( \frac{\sum_{i=1}^{n} x_i}{n} \right)^2 \right]$$
Sample Variance

The **sampling distribution of the variance** is the distribution of sample variances, with all samples having the same sample size \( n \) taken from the same population. (The sampling distribution of the variance is typically represented as a probability distribution in the format of a table, probability histogram, or formula.)
All outcomes are equally likely so the population variance is 2.9; the mean of the 10,000 trials is 2.88. If continued indefinitely, the sample variance will be 2.9. Also, notice the distribution is “skewed to the right.”
Properties of Sample Variances

- Sample variances target the value of the population variance. (That is, the sample variances is an unbiased estimator. The expected value of the sample variance is equal to the population variance.)

- The distribution of the sample variances tends to be a distribution skewed to the right.
Sample Proportion

The **sampling distribution of the proportion** is the distribution of sample proportions, with all samples having the same sample size $n$ taken from the same population.
Example - Sampling Distributions

Specific results from 10,000 trials

All outcomes are equally likely so the population proportion of odd numbers is 0.50; the proportion of the 10,000 trials is 0.50. If continued indefinitely, the mean of sample proportions will be 0.50. Also, notice the distribution is “approximately normal.”
Properties of sample proportions

- Sample proportions target the value of the population proportion. (That is, the sample proportion is an unbiased estimator. The expected value of the sample proportion is equal to the population proportion.)

- The distribution of the sample proportion tends to be a normal distribution.
Why Sample with Replacement?

Sampling *without replacement* would have the very practical advantage of avoiding wasteful duplication whenever the same item is selected more than once. However, we are interested in sampling *with replacement* for these two reasons:

1. When selecting a relatively small sample form a large population, it makes no significant difference whether we sample with replacement or without replacement.

2. Sampling with replacement results in independent events that are unaffected by previous outcomes, and independent events are easier to analyze and result in simpler calculations and formulas.
Section 6-5
The Central Limit Theorem
Central Limit Theorem

Given:

1. The random variable $x$ has a distribution (which may or may not be normal) with mean $\mu$ and standard deviation $\sigma$.

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 $\mu$.

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

As we proceed from $n = 1$ to $n = 50$, we see that the distribution of sample means is approaching the shape of a normal distribution.
Example - Uniform Distribution

As we proceed from $n = 1$ to $n = 50$, we see that the distribution of sample means is approaching the shape of a normal distribution.
Example - U-Shaped Distribution

As we proceed from \( n = 1 \) to \( n = 50 \), we see that the distribution of sample means is approaching the shape of a normal distribution.
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 closer to a normal distribution as the sample size $n$ becomes larger.

2. If the original population is *normally distributed*, then for any sample size $n$, the sample means will be normally distributed (not just the values of $n$ larger than 30).
As the sample size increases, the sampling distribution of sample means approaches a normal distribution.
Example – Water Taxi Safety

Use the Chapter Problem. Assume the population of weights of men is normally distributed with a mean of 172 lb and a standard deviation of 29 lb.

a) Find the probability that if an individual man is randomly selected, his weight is greater than 175 lb.

b) Find the probability that 20 randomly selected men will have a mean weight that is greater than 175 lb (so that their total weight exceeds the safe capacity of 3500 pounds).
a) Find the probability that if an individual man is randomly selected, his weight is greater than 175 lb.

\[
z = \frac{175 - 172}{29} = \]

\[\mu = 172, \quad \sigma = 29\]
Example – Water Taxi Safety

b) Find the probability that 20 randomly selected men will have a mean weight that is greater than 175 lb (so that their total weight exceeds the safe capacity of 3500 pounds).

\[ z = \frac{175 - 172}{\frac{29}{\sqrt{20}}} = \]

\( \bar{x} = 175 \)

\( \mu_{\bar{x}} = 172 \)

\( (\sigma_{\bar{x}} = \frac{\sigma}{\sqrt{n}} = \frac{29}{\sqrt{20}} = 6.4845971) \)
Section 6-6
Normal as Approximation to Binomial
Approximation of a Binomial Distribution with a Normal Distribution

\[ np \geq 5 \]
\[ n(1-p) \geq 5 \]

then

\[ \mu = np \quad \sigma = \sqrt{np(1-p)} \]

and the random variable has a (normal) distribution.
Continuity Correction

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 discrete whole number $x$ by the interval from $x - 0.5$ to $x + 0.5$ (that is, adding and subtracting 0.5).
Procedure for Using a Normal Distribution to Approximate a Binomial Distribution

1. Draw a normal distribution centered about \( \mu \) with standard deviation \( \sigma \) then draw a *vertical strip area* centered over \( x \). Mark the left side of the strip with the number equal to \( x - 0.5 \), and mark the right side with the number equal to \( x + 0.5 \). *The entire area of the strip represents the probability of the discrete number.*

2. 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; also shade the interior of the strip *if and only if* \( x \) *itself* is to be included. This total shaded region corresponds to the probability being sought.
Procedure for Using a Normal Distribution to Approximate a Binomial Distribution

3. Using $x - 0.5$ or $x + 0.5$ in place of $x$, find the area of the shaded region: find the $z$ score; use that $z$ score to find the area to the left of the adjusted value of $x$; use that cumulative area to identify the shaded area corresponding to the desired probability.
\(X = \text{at least } 8\)  
(includes 8 and above)

\(X = \text{more than } 8\)  
(doesn’t include 8)

\(X = \text{at most } 8\)  
(includes 8 and below)

\(X = \text{fewer than } 8\)  
(doesn’t include 8)

\(X = \text{exactly } 8\)
Example – Number of Men Among Passengers

Assuming \( p = 0.5 \), find the Probability of “At Least 122 Men” Among 213 Passengers

Figure 6-21
Section 6-7
Assessing Normality
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.
Example

**Normal:** Histogram of IQ scores is close to being bell-shaped, suggests that the IQ scores are from a normal distribution. The normal quantile plot shows points that are reasonably close to a straight-line pattern. It is safe to assume that these IQ scores are from a normally distributed population.
Manual Construction of a Normal Quantile Plot

Step 1. First sort the data by arranging the values in order from lowest to highest.

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

Step 3. Use the standard normal distribution (Table A-2 or software or a calculator) to find the $z$ scores corresponding to the cumulative left areas found in Step 2. (These are the $z$ scores that are expected from a normally distributed sample.)
Manual Construction of a Normal Quantile Plot

Step 4. Match the original sorted data values with their corresponding $z$ scores found in Step 3, then plot the points $(x, y)$, where each $x$ is an original sample value and $y$ is the corresponding $z$ score.

Step 5. Examine the normal quantile plot and determine whether or not the distribution is normal.
Determining Whether It To Assume that Sample Data are From a Normally Distributed Population

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 in one side.

3. **Normal Quantile Plot:** If the histogram is basically symmetric and there is at most one outlier, use technology to generate a normal quantile plot.
Determining Whether To Assume that Sample Data are From a Normally Distributed Population

3. Continued

Use the following criteria to determine whether or not the distribution is normal.

**Normal Distribution:** The population distribution is normal if the pattern of the points is reasonably close to a straight line and the points do not show some systematic pattern that is not a straight-line pattern.
Determining Whether To Assume that Sample Data are From a Normally Distributed Population

3. Continued

Not a Normal Distribution: The population distribution is not normal if either or both of these two conditions applies:

- The points do not lie reasonably close to a straight line.
- The points show some systematic pattern that is not a straight-line pattern.
**Example**

**Skewed:** Histogram of the amounts of rainfall in Boston for every Monday during one year. The shape of the histogram is skewed, not bell-shaped. The corresponding normal quantile plot shows points that are not at all close to a straight-line pattern. These rainfall amounts are not from a population having a normal distribution.
**Example**

**Uniform**: Histogram of data having a uniform distribution. The corresponding normal quantile plot suggests that the points are not normally distributed because the points show a systematic pattern that is not a straight-line pattern. These sample values are not from a population having a normal distribution.