Random variables can be either \_\_\_\_\_\_\_\_\_\_\_ or \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_. Continuous variables, such as height of adults, body temperature of rats, or household size, can assume all values between any two given values of the variables. Although no variable fits a normal distribution perfectly, the normal distribution can be used to describe many variables since the \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_ from a normal distribution are small.

# 6-1: Normal Distributions

## Objective 1. Identify the Properties of a Normal Distribution.

### Definition: Normal Distribution

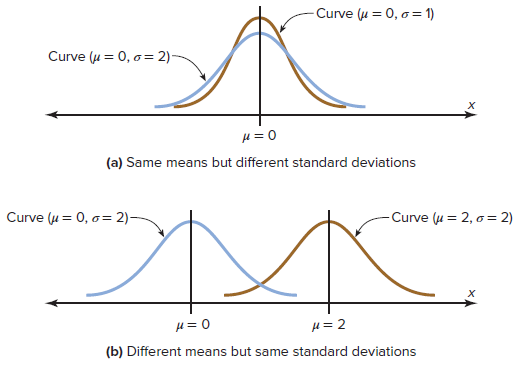
If a random variable has a probability distribution whose graph is continuous, bell-shaped, and symmetric, it is called the \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_ \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_. The graph is called a normal distribution curve.

### Mathematical Equation for a Normal Distribution

where ;

.

The shape and position of the normal distribution curve depend on the parameters \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_ and \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_.



When the population means are the same, but the population standard deviations are different, the curve with the larger population standard deviation is more \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_ \_\_\_\_\_\_\_\_\_\_\_.

When the population standard deviations are the same, but the means are different, the curves have the \_\_\_\_\_\_\_\_\_ \_\_\_\_\_\_\_\_\_\_\_\_\_, but are centered at different points along the x axis.

### Summary of the Properties of the Theoretical Normal Distribution

A normal distribution curve is \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_.

The \_\_\_\_\_\_, \_\_\_\_\_\_\_\_\_\_\_\_, and \_\_\_\_\_\_\_\_\_\_ are located at the center of the distribution.

The normal distribution curve has exactly \_\_\_\_\_\_ mode.

The normal distribution curve is \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_ about the mean.

The normal distribution curve is \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_, having no gaps or holes, for all values of the independent variable.

The curve \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_, but does meet, the x-axis, getting increasingly close while never touching the x-axis.

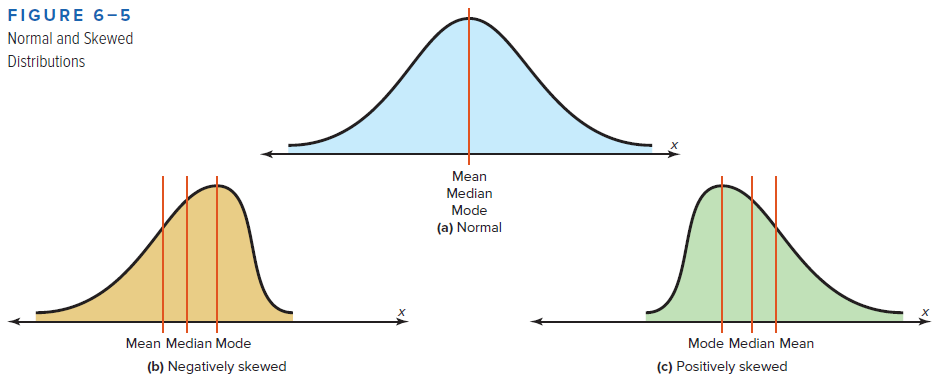
The total area under a normal distribution curve is equal to \_\_\_\_\_\_.

The area under the part of the normal distribution that lies within \_\_\_\_\_\_\_ standard deviation of the mean is approximately \_\_\_\_\_\_%; within \_\_\_\_\_ standard deviations of the mean is approximately \_\_\_\_\_\_%; and within \_\_\_\_\_\_\_ standard deviations, approximately \_\_\_\_\_\_%.

The area under the part of the normal distribution that lies within one standard deviation of the mean is approximately 68%; within two standard deviations of the mean is approximately 95%; and within three standard deviations, approximately 99.7%.  This is also known as the empirical rule.


## Objective 2. Identify Distributions as Symmetric or Skewed

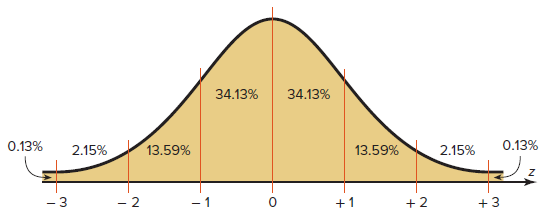
The normal distribution is \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_, that is, the data is evenly distributed about the mean. When the majority of the data values fall either to the left or right of the mean, the distribution is \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_. When the majority of the data is to the **right** of the mean, the distribution is \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_-skewed or left-skewed. The mean is less than or to the left of the \_\_\_\_\_\_\_\_\_\_\_\_\_. When the majority of the data is to the **left** of the mean, the distribution is \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_-skewed or right-skewed. The mean is greater than or to the right of the \_\_\_\_\_\_\_\_\_\_\_\_\_. The “tail” of the curve indicates the direction of the skewness (right is positive and left is negative).



## Objective 3. Find the Area Under the Standard Normal Distribution, Given Various *z* Values.

### Definition: Standard Normal Distribution

The Standard Normal Distribution is a \_\_\_\_\_\_\_\_\_\_\_ \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_ with a mean of \_\_\_\_\_ and a standard deviation of \_\_\_\_.



The formula for the standard normal distribution is .

All normally distributed variables can be transformed into the \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_ \_\_\_\_\_\_\_\_\_\_ \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_ variable by using the formula for the standard score:

.

The standard score, or z-score, is the number of \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_ \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_ that a particular X value is from the mean. Table E in Appendix A gives the area (to four decimal places) under the standard normal curve for z values from -3.49 to 3.49. Technology can also be used to find the area under the standard normal distribution curve.

### Finding Areas Under the Standard Normal Distribution Curve

**Step 1:** Draw the normal distribution curve and shade the area under consideration.

**Step 2:** Choose the appropriate type of problem and either use Table E in Appendix A or use technology to find the shaded area.

| **To find:** | **Sketch of an Example:** |
| --- | --- |
| The area under the standard normal distribution curve to the left a z value, find the z value in Table E (Appendix A) and use the area given. | Find the area to the left of z = 1.32 |
| The area under the standard normal distribution curve to the right of a z value, find the z value in Table E (Appendix A) and subtract the area given from 1. | Find the area to the right of z = −0.67 |
| The area under the standard normal distribution curve between two z values, find both z values in Table E (Appendix A) and subtract the corresponding areas. | Find the area between z = 1.45 and z = 2.32 |

For the following examples, use Table E or technology.

### Example 6 – 1. Find Areas Between Two Given *Z* Values

Find the area under the standard normal distribution curve

1. Between .

*Solution:*

1. To the right of

*Solution:*

1. To the left of .

*Solution:*

1. Between

*Solution:*

### Example 6 – 2. Find Probabilities for Regions Between Two *Z* Values

Find the Probabilities using the standard normal distribution curve

*Solution:*

*Solution*:

*Solution:*

*Solution:*

### Example 6 – 3. Find the *Z* Values for Specific Areas in the Tails

Find two *z* values, one positive and one negative that are equidistant from the mean so that the area in the two tails totals 5%; 2%; and 1%.

*Solutions:*

# 6 – 2 Applications of the Normal Distribution

## Objective 4. Find Probabilities for a Normally Distributed Variable by Transforming it Into a Standard Normal Variable.

The standard normal distribution curve can be used to solve many practical problems. Assume the variables presented here are all normally or approximately \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_ \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_.

First we transform the original variable to a standard normal distribution variable using the formula, , rounding the result to two decimal places.

### To Find the Area Under Any Normal Curve

| **Step 1** | Draw a normal curve and \_\_\_\_\_\_\_\_\_\_ the desired area. |
| --- | --- |
| **Step 2** | Convert the value of *X* to *z* using the formula \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_. |
| **Step 3** | Find the \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_ \_\_\_\_\_\_\_\_\_\_\_\_ using a table, calculator or software. |

### Example 6 – 4. Liters of Blood

An adult has on an average 5.2 liters of blood. Assume the variable is normally distributed and has a standard deviation of 0.3. Find the percentage of people who have more than 5.4 liters of blood in their system.

Solution.

**Step 1:** Draw a normal curve and **shade** the desired area.

**Step 2:** Use the formula to find the z value corresponding to 5.4.

**Step 3:** Use Table E, or your calculator, or computer software to determine the area to the right of . (Notice this value is rounded.) The area under the standard normal curve to the left of the z-score, then to find the area to the right, subtract the value from the table from 1.

Therefore, \_\_\_\_\_\_\_\_\_% of adults have greater than 5.4 liters of blood in their system.

Example 6 – 5. Monthly Mortgage Payments

The average monthly mortgage payment including principal and interest is $982 in the United States. If the standard deviation is approximately $180 and the mortgage payments are approximately normally distributed, find the probability that a randomly selected monthly payment is a) more than $1000; b) More than $1475; between $800 and $1150?

*Solution:*

**Step 1:** Draw.

a) b) c)

**Step 2:** Calculate z.

a) b) c)

**Step 3:**  Find the appropriate areas.

a) b) c)

## Objective 5. Find Specific Data Values for Given Percentages, Using the Standard Normal Distribution.

Using the formula , solve for *X* and use this new equation to find the specific data value for a given percentage.

Solve for X:

### Formula for Finding the Value of a Normal Variable *X*

### Finding Data Values for Specific Probabilities

| **Step 1** | Draw a normal curve and shade the desired area that represents the probability, proportion or percentile. |
| --- | --- |
| **Step 2** | Find the *z* value corresponding to the desired area from the table, calculator, or computer software. |
| **Step 3** | Calculate the *X* value using the formula . |

### Example 6 – 6. Police Academy Qualifications

To qualify for a police academy, candidates must score in the top 10% on a general abilities test. Assume the test scores are normally distributed and the test has a mean of 200 and a standard deviation of 20. Find the lowest possible score to qualify.

*Solution:*

**Step 1.** Draw a normal distribution curve and shade the right hand 10% of the area.

**Step 2.** The area to the left of the 0.1 shaded on the right is 1 – 0.1 = 0.1. Find the *z* value from Table E (or calculator or computer software) corresponding to 0.9 area to the left.

**Step 3.** Use this value in the formula to find *X.*

*Conclusion:* A score of \_\_\_\_\_\_\_\_\_ should be used for the cutoff. Anyone scoring \_\_\_\_\_\_ or \_\_\_\_\_\_\_\_\_\_\_ qualifies for the academy.

### Example 6 – 7. Qualifying Test Scores

In order to qualify for a medical study, an applicant must have a systolic blood pressure in the 50% of the middle range. If the systolic blood pressure is normally distributed with a mean of 120 and a standard deviation of 4, find the upper and lower limits of blood pressure a person must have in order to qualify for the study.

Solution:

**Step 1.** Draw. Mark the middle 50%. This leaves 25% to the left of the middle and 25% to the right of the middle.

**Step 2.** Use the table or technology to find the *z* scores corresponding with the area of 0.25 to the left and 0.75 to the right.

**Step 3.**  Use the formula to calculate the two values of *X,* using each of the z-scores, 120 as the mean and 4 as the standard deviation.

## Determine Normality

There are a variety of ways to determine if a distribution of values is \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_ or approximately \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_ distributed.

The easiest way to check for normality is to draw a \_\_\_\_\_\_\_\_\_\_\_\_\_ for the data and check its shape.

Another is to find the Pearson coefficient (*PC*) of skewness or the Pearson’s index of skewness:

.

If the index is greater than or equal to or less than or equal to , it can be concluded that the data are significantly \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_.

Check the data for outliers by finding the outlier \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_ as defined by

,

### Example 6 – 8. Technology Innovations

A survey of 18 high-tech firms showed the number of days’ inventory they had on hand. Determine if the data are approximately normally distributed.

| 5 | 29 | 34 | 44 | 45 | 63 | 68 | 74 | 74 |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
| 81 | 88 | 91 | 97 | 98 | 113 | 118 | 151 | 158 |

*Solution:*

Construct a frequency distribution and draw a histogram.

Draw a histogram.

Use the formula for the Pearson Coefficient of Skewness. Determine if there is an indication that the distribution is significantly skewed.

Check for outliers using the formulas for the lower and upper outlier boundaries.

Using these three indicators, the distribution is \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_ \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_.

### Example 6 – 9. Number of Runs Made

The data represent the number of runs made each year during Bill Mazeroski’s career. Check for normalcy.

|  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
| 30 | 59 | 69 | 50 | 58 | 71 | 55 | 43 | 66 |
| 52 | 56 | 62 | 36 | 13 | 29 | 17 | 3 |  |

*Solution:*

Construct a frequency distribution and draw a histogram.

Draw a histogram.

Use the formula for the Pearson Coefficient of Skewness. Determine if there is an indication that the distribution is significantly skewed.

Check for outliers using the formulas for the lower and upper outlier boundaries.

Using these three indicators, the distribution is \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_ \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_.

# 6 – 3 The Central Limit Theorem

## Objective 6. Use the Central Limit Theorem to Solve Problems Involving Sample Means for Large Samples.

Select a sample of 30 adult males and find the mean measure of the triglyceride levels for the sample to be 187 mg/dl. Select a second sample, and find the mean to be 192 mg/dl. Continue selecting until you have 100 samples and sample means. Let the mean of these sample means be a random variable. The sample means become a *sampling distribution of sample means.*

### Definition: Sampling Distribution of Sample

A **sampling distribution of sample** means is a distribution using the means computed from all possible random samples of a specific size taken from a population.

The sample means of samples randomly selected with replacement will be somewhat \_\_\_\_\_\_\_\_\_\_\_\_\_\_ from the population \_\_\_\_\_\_\_\_\_\_\_. The differences are caused by sampling error.

### Definition: Sampling Error

\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_ \_\_\_\_\_\_\_\_\_\_\_\_\_\_ is the difference between the sample measure and the corresponding population measure due to the fact that the sample is not a perfect representation of the population.

| **Properties of the Distribution of Sample Means** |
| --- |
| 1. The mean of the sample means will be the same as the population mean. |
| 2. The standard deviation of the sample means will be smaller than the standard deviation of the population , and it will be equal to the population standard deviation divided by the square root of the sample size. |

### Example 6 – 10. The Central Limit Theorem.

This example illustrates the three properties of the Central Limit Theorem.

Suppose an 8-point quiz was given to four students and the results were 2, 4, 6, and 8.

The population mean is \_\_\_\_\_ and the population standard deviation is \_\_\_\_\_\_\_.

Find all possible samples of size 2 taken with replacement and find the mean of each sample. Then using the frequency distribution of the sample means, find the mean of the sample means and the standard deviation of the sample means, also known as the standard error of the mean.

| **Sample** | **Mean** | **Sample** | **Mean** |
| --- | --- | --- | --- |
| 2,2 |  | 6,2 |  |
| 2,4 |  | 6,4 |  |
| 2,6 |  | 6,6 |  |
| 2,8 |  | 6,8 |  |
| 4,2 |  | 8,2 |  |
| 4,4 |  | 8,4 |  |
| 4,6 |  | 8,6 |  |
| 4,8 |  | 8,8 |  |

|  | **frequency** |
| --- | --- |
| 2 |  |
| 3 |  |
| 4 |  |
| 5 |  |
| 6 |  |
| 7 |  |
| 8 |  |

The mean of the sample means, denoted by , is \_\_\_\_\_\_\_\_, which is the mean of the population.

Thus,

The standard deviation of the sample means, denoted by , is \_\_\_\_\_\_\_\_, which is the standard deviation of the population divided by

Thus, .

Draw the histogram of the distribution of sample means. Notice that the shape is approximately \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_.

## The Central Limit Theorem

As the sample size *n* increases without limit, the shape of the distribution of the sample means taken with replacement from a population with mean and standard deviation will approach a normal distribution with a mean of and a standard deviation of .

For samples with sufficient size, the central limit theorem can be used to answer questions about sample means in the same manner that a normal distribution can be used to answer questions about individual values. The difference is that the formula for *z* becomes

1. When the original variable is normally distributed, the distribution of the sample means will be normally distributed for any sample size *n.*
2. When the distribution of the original variable is not normal, a sample size of 30 or more is needed to use a normal distribution to approximate the distribution of the sample means.

### Example 6 – 11. Working Weekends.

The average time spent by construction workers who work on weekends is 7.93 hours (over a period of 2 days). Assume the distribution is approximately normal and has a standard deviation of 0.8 hours.

1. Find the probability that an individual who works at that trade works fewer than 8 hours on the weekend.
2. If a sample of 40 construction workers is randomly selected, find the probability that the mean of the sample will be less than 8 hours.

*Solution*:

1. Find the probability that an individual who works at that trade works fewer than 8 hours on the weekend.

**Step 1:** Draw a normal curve and **shade** the desired area.

**Step 2:** Use the formula to find the z value corresponding to an individual who works fewer than 8 hours.

**Step 3:** Use Table E, or your calculator, or computer software to determine the area to the left of .

Therefore, the probability that an individual who works at that trade works fewer than 8 hours on the weekend is \_\_\_\_\_\_\_\_\_\_\_\_\_\_.

1. If a sample of 40 construction workers is randomly selected, find the probability that the mean of the sample will be less than 8 hours.

**Step 1:** Draw a normal curve and **shade** the desired area for the mean of the sample of 40 construction workers.

**Step 2:** Use the formula to find the z value corresponding to the mean of 40 construction workers has a mean less than 8 hours.

**Step 3:** Use Table E, or your calculator, or computer software to determine the area to the left of .

Therefore, the probability that a sample of 40 construction workers has a mean of fewer than 8 hours on the weekend is \_\_\_\_\_\_\_\_\_\_\_\_\_\_.

## Finite Population Correction Factor

The formula for the standard error of the mean is accurate when samples are drawn with replacement or drawn without replacement from a very large or infinite population. When sampling from a finite population without replacement, a correction factor of , where *N* is the population size and *n* is the sample size.

That is, we use

So the formula for *z* values becomes

If the population is large, but the sample is small, the correction factor is not used. In that case, the correction factor is close to 1.

# 6 – 4 The Normal Approximation to the Binomial Distribution

## Objective 7. Use the Normal Approximation to Compute Probabilities for a Binomial Variable.

Recall that a binomial distribution has the following characteristics:

1. There must be a \_\_\_\_\_\_\_\_\_\_\_\_\_\_ number of trials.
2. The outcome of each trial must be \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_.
3. Each experiment can have only \_\_\_\_\_\_ outcomes, or outcomes that can be reduced to \_\_\_\_\_\_\_ outcomes.
4. The probability of a success must remain the \_\_\_\_\_\_\_\_ for each trial.

Also,

and

When *n* is large, that is 100 or more, the \_\_\_\_\_\_\_\_\_\_\_ distribution can sometimes be used to approximate the binomial distribution. When the probability of success is close to 0.5, the shape of the binomial distribution is approximately normal. Statisticians agree, as a rule of thumb, that the normal distribution should be used only when . However, a continuity correction should also be used when employing the normal distribution to represent a discrete distribution. The continuity correction means that for any specific value of *X,* say 8, the boundaries of *X* in the binomial distribution (in this case, 7.5 to 8.5) must be used.

| **Summary of the Normal Approximation to the Binomial Distribution** | |
| --- | --- |
| **Binomial** | **Normal** |
| When finding: | Use: |
| 1. |  |
| 2. |  |
| 3. |  |
| 4. |  |
| 5. |  |
| For all cases, , , . | |

| Procedure for the Normal Approximation to the Binomial Distribution | |
| --- | --- |
| **Step 1** | Check to see whether the normal approximation can be used. |
| **Step 2** | Find the mean and the standard deviation . |
| **Step 3** | Write the problem in probability notation, using *X.* |
| **Step 4** | Rewrite the problem by using the continuity correction factor, and show the corresponding area under the normal distribution. |
| **Step 5** | Find the corresponding *z* values. |
| **Step 6** | Find the solution. |

### Example 6 –12. Population of College Cities

College students often make up a substantial portion of the population of college cities and towns. State College, Pennsylvania, ranks first with 71.1% of its population made up of college students. What is the probability that in a random sample of 150 people from State College, more than 50 are not college students?

*Solution:*

**Step 1.** Check if the normal approximation can be used.

**Step 2.** Write the probability notation that models the problem.

**Step 3.** Find the mean and standard deviation of the distribution.

**Step 4.** Rewrite the problem using the continuity correction factor.

**Step 5.**  Find the corresponding *z* values.

**Step 6.** Find the probability.