

Section 7-1: The Normal Distribution

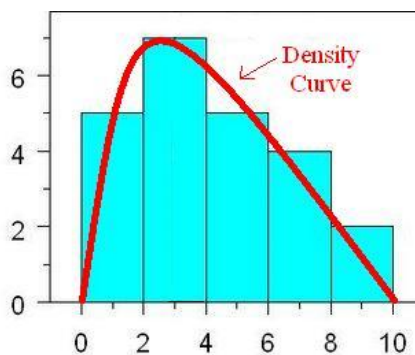
Learning Objectives

Understand how a density curve can be used to approximate the data in a histogram

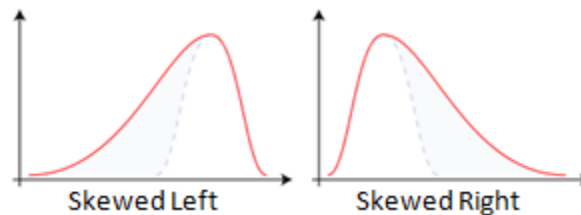
Understand how to visually identify the mean and standard deviation of a normal distribution

Be able to tie the concepts of percentages in the 68-95-99.7 empirical rule to normal distributions

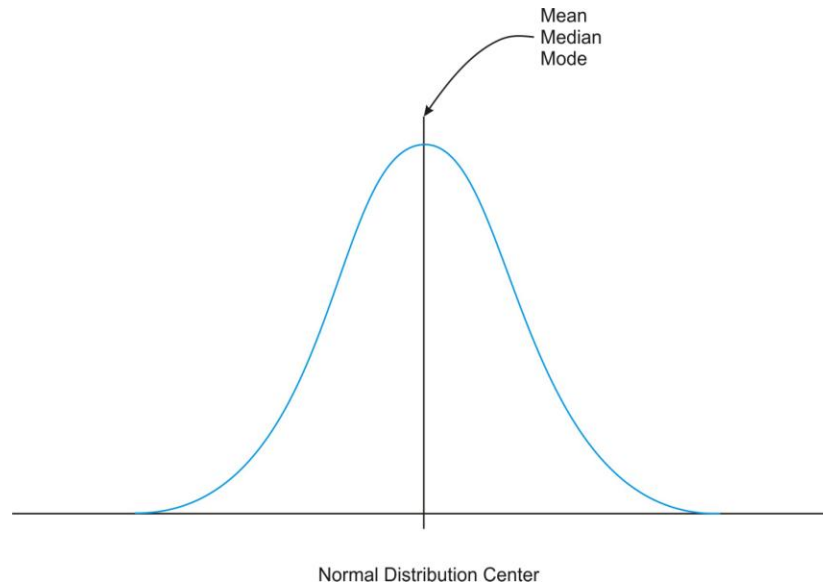
In previous chapters we have seen how data can be represented by histograms. A **density curve** is a curve that gives an approximate description of a distribution. The curve is smooth, so any small irregularities in the data are ignored. A density curve for a particular histogram is shown below. Perhaps the most important thought to remember about a density curve is that it represents 100% of the data. In other words, the area under any density curve is equal to 1. This is important because it allows us to ask probability questions about a population. For example, we might ask how likely is it that a teenager has a shoe size of 8 or larger.



In our chapter, we will focus on a special density curve called the normal curve. Have you ever wondered if you are 'normal'? You probably are normal in most ways, but there may be some things about you that might not be considered normal by the mathematical definition. If you are on the high school baseball team, do you throw the baseball at a 'normal' speed? Is your hair a 'normal' length? Do you drive at a 'normal' speed on the freeway? Our goal this chapter is to gain an understanding of what 'normal' really is and how to properly calculate within the Normal Distribution. We have seen skewed distributions before. The density curves in the following figure show one density curve that is skewed left and one that is skewed right.



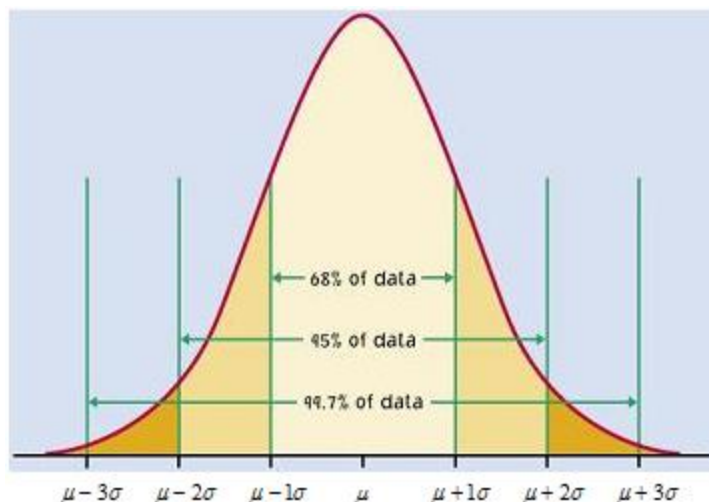
A **normal curve** is neither skewed left nor right and is often referred to as 'the bell curve' because of its shape. It is symmetrical. In addition, as you get closer and closer to the middle of the curve, there is a higher frequency of results. The **mean** (along with the median and mode) always lands at the center of a normal distribution. When dealing with the mean in previous chapters, we have used the symbol \bar{x} because the data came from a sample. Normal distributions deal with an entire population instead of just a sample and we will use the symbol μ (Greek letter mu) to mark the mean of a normal distribution for an entire population. The mean is one of two key values needed to make a proper sketch and analysis of a normal distribution. The curve shown below represents a normal distribution and is a good representation of what a normal curve looks like.



Note that the amount of data to the right of the mean is the same as the amount of data to the left of the mean. Thinking about the definition of the median, this suggests that the mean and median are located at the same point. The other key component used to construct and analyze a normal distribution is the **standard deviation**. The standard deviation is a measure of spread and can be loosely thought of as a 'typical' distance from the mean. You may have calculated the standard deviation before for data sets either by hand or by using your calculator and looked for the S_X in the statistical calculations summary screen. The symbol S_X is used for the standard deviation whenever data is collected through the use of a sample from a population. When dealing with the normal distribution, we will use the symbol σ (Greek letter sigma) to represent the standard deviation. The symbol σ indicates that the standard deviation of the entire population is known. Visually, the standard deviation can be seen as the distance from the mean to an **inflection point**. An inflection point is located on a curve at the point where the curve changes from **concave up** (bent up) to **concave down** (bent down) or vice versa. On the normal curve in Figure 7.1, the mean is 23 and the standard deviation is 3.

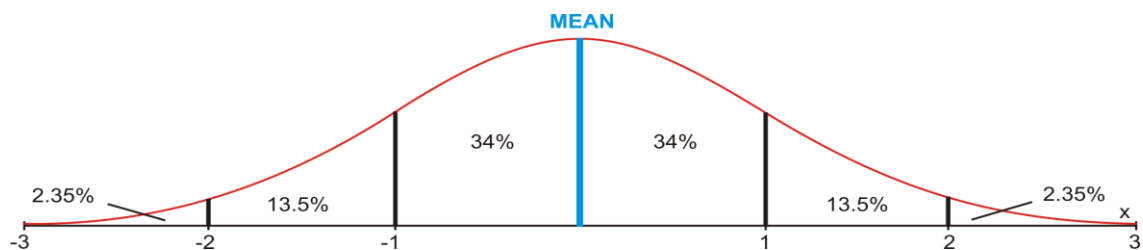
The Empirical Rule (68-95-99.7 Rule)

It is now time to make use of some of the special characteristics of the normal curve. As mentioned earlier, 100% of all results fall somewhere under the normal curve. It turns out that approximately 68% of all results are within one standard deviation of the mean, 95% of all results are within 2 standard deviations of the mean, and 99.7% of all results land within three standard deviations of the mean. These percentages are illustrated in the graphic below.



The numbers on the bottom represent the number of standard deviations from the mean. For example, the marks the point one standard deviation below the mean. Some simple addition and subtraction allows us to be very specific in the percents of the data that land in the sections of the normal curve as shown below.

Can you see the 68-95-99.7 rule here?



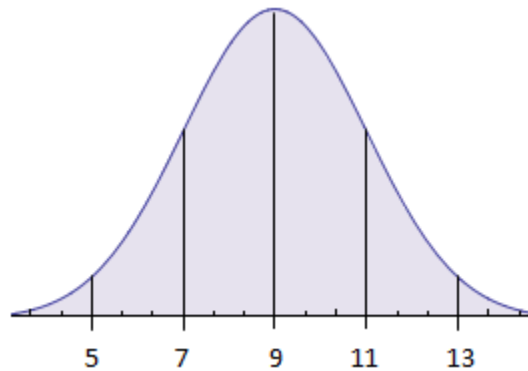
Example 1

Suppose the mathematics portion of the SAT exam is normally distributed with a mean of 500 and a standard deviation of 100.

- Sketch a normal curve for this situation marking the mean and the values 1, 2, and 3 standard deviations above and below the mean.
- Using the 68-95-99.7 rule, approximately what percent of students scored at least 600 on this test?
- Between approximately which two scores did the middle 95% of students score?
 - Suppose that 4600 students take the exam this month. How many of those students should we expect to obtain a score of at least 700?

Example 2

The normal curve below represents the number of races that a typical racehorse will run in one calendar year.



- Approximately what percent of racehorses will run between 5 and 11 races during a calendar year?
- What are the values of the mean and standard deviation for the distribution shown?

Solution

What is Normal?

Let's now go back and try to think about our original question "What is normal?" In mathematics, the middle 95% is often (but not always) considered our 'normal' group. For example, suppose the ACT exam is normally distributed with a mean of 18 and a standard deviation of 6. Our 'normal' group would be comprised of those students who scored anywhere within two standard deviations of the mean or from 6 to 30 on the exam. A student who scored 31 or higher on the exam would have achieved an exceptional score. We might say that this student was not normal with regards to their ACT score.

Normal distributions are not as common as you might think. What if we measured the lengths of shoes of teenagers? Many students think that this would be normal when in fact, there are a couple of contributing factors that might tip us off that the situation may not be normal. First of all, teenagers encompass a large population. Most of those who are in their upper teen years have finished growing into their adult shoe size length whereas many of the younger teens are still growing. This would tend to give us a slightly larger percentage of smaller shoe lengths than we might expect from a normal distribution. In addition, teenagers include males and females. This may lead to us seeing a situation which might be bi-modal. We might expect to see a peak at the most common male lengths and at the most common female lengths.

Example 3

Which situation below is most likely to produce a normal distribution?

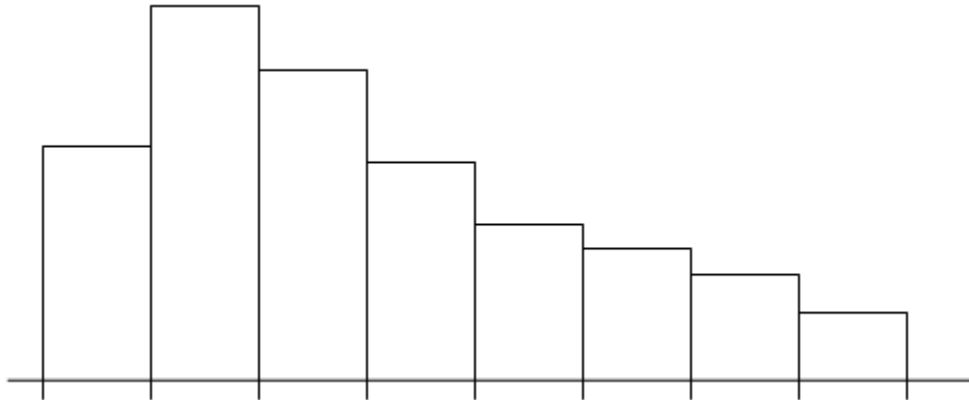
- The heights of all adults.
- The wingspans of three year-old American eagles.
- The number of teeth that Americans adults have.

Solution

Problem Set 7.1

Exercises

- 1) Consider the histogram shown below.



- a) Make a sketch of the histogram and overlay a sketch of a density curve for the histogram.
b) What is the area under your density curve?
c) What is the shape of the density curve?

2) A roadside bait salesman digs up worms to sell to fishermen. It turns out that the worms have a mean length of $\mu = 112$ mm and a standard deviation of $\sigma = 12$ mm.

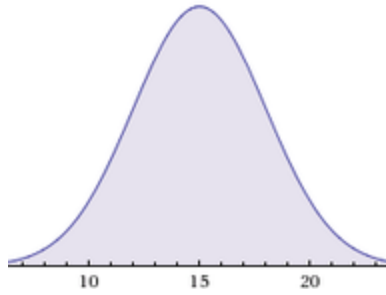
- a) Draw and label a normal curve for this distribution. Include lines for the mean and for 1, 2, and 3 standard deviations above and below the mean.
b) What percentage of the worms will have lengths longer than 112 mm?
c) What percentage of the worms will have lengths between 100 and 124 mm long?
d) What percentage of the worms will have lengths between 100 and 112 mm long?
e) What percentage of the worms are longer than 124 mm?
f) What percentage of the worms are shorter than 88 mm?

3) Sketch a normal curve which has a mean of 13 pounds and a standard deviation of 3 pounds. Include lines for the mean and for 1, 2, and 3 standard deviations above and below the mean.

4) Not all 12-ounce cans of soda are the same. It turns out that the average 12-ounce can of soda does contain twelve ounces of soda, but the amount of soda is normally distributed with a standard deviation of 0.15 ounces. Fill in the blanks for each statement below.

- a) The middle 68% of all 12-ounce soda cans contain between ____ & ____ ounces of soda.
b) The middle 95% of all 12-ounce soda cans contain between ____ & ____ ounces of soda.
c) The middle 99.7% of all 12-ounce soda cans contain between ____ & ____ ounces of soda.

5) Figure 7.2 on the following page shows an approximate distribution of the number of fish caught by the competitors during a one hour pan-fishing contest. Give the approximate values of the mean and the standard deviation for the distribution.



6) Suppose the weights of adult males of a particular species of whale are distributed normally with a mean of 11,600 pounds and a standard deviation of 640 pounds.

a) Draw a normal curve for this situation. Use vertical lines to mark and label the mean and 1, 2, and 3 standard deviations above and below the mean.

b) What percent of these whales weigh less than 10,320 pounds?

c) Between what two weights do the middle 99.7% of these whales weigh?

d) What percent of these whales weigh between 10,320 pounds and 12,240 pounds?

7) Which situation is most likely to be normally distributed? Explain your reasoning.

i) The hair lengths for all the Statistics and Probability students who have Mr. Johnson as a teacher.

ii) The prices of all new Ipod Touches that are sold in Minnesota this week.

iii) The average running times for all 4th grade boys at Andover Elementary in the 50 yard dash.

8) Suppose a standard incandescent light bulb will run an average of 400 hours before burning out. Of course, some bulbs burn out sooner and some last longer. Suppose that the average lives of these bulbs is normally distributed with a standard deviation of 35 hours.

a) Sketch and label a normal curve to illustrate this situation.

b) What percent of these bulbs will burn out in 400 hours or less?

c) If you are lucky, your bulb will last longer than advertised. What percent of bulbs should last 435 hours or more? What percent of bulbs will last 470 hours or more?

d) If you had 5000 bulbs that you needed for use in a large office building, how many would you expect to last at least 365 hours?

9) Suppose that the time that it takes for a popcorn kernel to pop produces a normal distribution with a mean of 145 seconds and a standard deviation of 13 seconds for a standard microwave oven.

a) It is usually not a good idea to let the microwave oven run until all the kernels are popped because some of the popcorn will start to burn. Suppose the ideal time to shut off the microwave oven is after about 97.5% of the kernels have popped. When will 97.5% of the kernels be popped?

b) Between what two times will we see the middle 68% of kernels popped?

10) After a great deal of surveying, it is determined that the average wait times in the cafeteria line are normally distributed with a mean of 7 minutes and a standard deviation of 2 minutes. Suppose that 400 students are released to the cafeteria for 2nd lunch.

- a) Approximately how many students will have to wait more than 5 minutes for their food?
- b) Approximately how many students will have to wait more than 11 minutes for their food?

11) Sudoku is a popular logic game of number combinations. It originated in the late 1800s in the French press, *Le Siècle*. The mean time it takes the average 11th grader to complete the Sudoku puzzle on the following page was found to be 19.2 minutes, with a standard deviation of 3.1 minutes.

- a) Draw a normal distribution curve to represent this data.
- b) Suppose Andover High School is going to put together a Sudoku team. The coach has decided that she will only consider players who score in the fastest 2.5% of the junior class as she puts together the team. How fast must a student solve a puzzle to be in the top 2.5% of puzzle solvers?
- c) If there are 400 kids in the Andover junior class, how many of them will be able to solve the Sudoku puzzle below in 16.1 minutes or less?

5	3			7			
6			1	9	5		
	9	8					6
8				6			3
4			8		3		1
7				2			6
	6					2	8
			4	1	9		5
				8			7
						7	9

12) In order to qualify for undercover detective training, a police officer must take a stress tolerance test. Scores on this test are normally distributed with a mean of 60 and a standard deviation of 10. Only the top 16% of police officers score high enough on the test to qualify for the detective training. What is the cutoff score that marks the top 16% of all scores?

Review Exercises

13) A pet store must select 2 dogs and 2 cats for display in their front window. In how many ways can this be done if there are 16 dogs and 12 cats available to choose from?