

# NORMAL DISTRIBUTION: INTRODUCTION\*

Susan Dean  
Barbara Illowsky, Ph.D.

This work is produced by The Connexions Project and licensed under the  
Creative Commons Attribution License †

## 1 Student Learning Objectives

By the end of this chapter, the student should be able to:

- Recognize the normal probability distribution and apply it appropriately.
- Recognize the standard normal probability distribution and apply it appropriately.
- Compare normal probabilities by converting to the standard normal distribution.

## 2 Introduction

The normal, a continuous distribution, is the most important of all the distributions. It is widely used and even more widely abused. Its graph is bell-shaped. You see the bell curve in almost all disciplines. Some of these include psychology, business, economics, the sciences, nursing, and, of course, mathematics. Some of your instructors may use the normal distribution to help determine your grade. Most IQ scores are normally distributed. Often real estate prices fit a normal distribution. The normal distribution is extremely important but it cannot be applied to everything in the real world.

In this chapter, you will study the normal distribution, the standard normal, and applications associated with them.

## 3 Optional Collaborative Classroom Activity

Your instructor will record the heights of both men and women in your class, separately. Draw histograms of your data. Then draw a smooth curve through each histogram. Is each curve somewhat bell-shaped? Do you think that if you had recorded 200 data values for men and 200 for women that the curves would look bell-shaped? Calculate the mean for each data set. Write the means on the x-axis of the appropriate graph below the peak. Shade the approximate area that represents the probability that one randomly chosen male is taller than 72 inches. Shade the approximate area that represents the probability that one randomly chosen female is shorter than 60 inches. If the total area under each curve is one, does either probability appear to be more than 0.5?

The normal distribution has two parameters (two numerical descriptive measures), the mean ( $\mu$ ) and the standard deviation ( $\sigma$ ). If  $X$  is a quantity to be measured that has a normal distribution with mean ( $\mu$ ) and the standard deviation ( $\sigma$ ), we designate this by writing

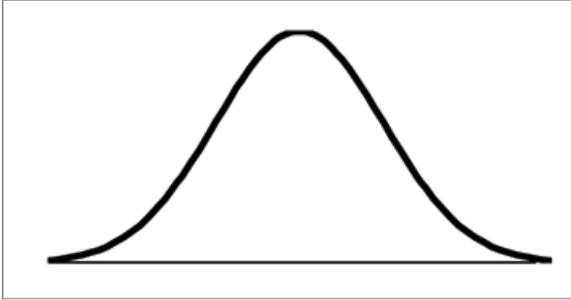
---

\*Version 1.9: Feb 8, 2009 10:20 am US/Central

†<http://creativecommons.org/licenses/by/2.0/>



**NORMAL:**  $X \sim N(\mu, \sigma)$



The probability density function is a rather complicated function. **Do not memorize it.** It is not necessary.

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

The cumulative distribution function is  $P(X < x)$  It is calculated either by a calculator or a computer or it is looked up in a table

The curve is symmetrical about a vertical line drawn through the mean,  $\mu$ . In theory, the mean is the same as the median since the graph is symmetric about  $\mu$ . As the notation indicates, the normal distribution depends only on the mean and the standard deviation. Since the area under the curve must equal one, a change in the standard deviation,  $\sigma$ , causes a change in the shape of the curve; the curve becomes fatter or skinnier depending on  $\sigma$ . A change in  $\mu$  causes the graph to shift to the left or right. This means there are an infinite number of normal probability distributions. One of special interest is called the **standard normal distribution**.

## Glossary

### Definition 1: Normal Distribution

A continuous random variable (RV) with pdf  $f(x) = \frac{1}{\sigma \sqrt{2\pi}} e^{-(x-\mu)^2/2\sigma^2}$ , where  $\mu$  is the mean of the distribution and  $\sigma$  is the standard deviation. Notation:  $X \sim N(\mu, \sigma)$ . If  $\mu = 0$  and  $\sigma = 1$ , the RV is called **the standard normal distribution**.

# NORMAL DISTRIBUTION: STANDARD NORMAL DISTRIBUTION\*

Susan Dean

Barbara Illowsky, Ph.D.

This work is produced by The Connexions Project and licensed under the  
Creative Commons Attribution License †

The **standard normal distribution** is a normal distribution of **standardized values called z-scores**. **A z-score is measured in units of the standard deviation**. For example, if the mean of a normal distribution is 5 and the standard deviation is 2, the value 11 is 3 standard deviations above (or to the right of) the mean. The calculation is:

$$x = \mu + (z)\sigma = 5 + (3)(2) = 11 \quad (1)$$

The z-score is 3.

The mean for the standard normal distribution is 0 and the standard deviation is 1. The transformation  $z = \frac{x-\mu}{\sigma}$  produces the distribution  $Z \sim N(0, 1)$ . The value  $x$  comes from a normal distribution with mean  $\mu$  and standard deviation  $\sigma$ .

## Glossary

### Definition 1: Standard Normal Distribution

A continuous random variable (RV)  $X \sim N(0, 1)$ .. When X follows the standard normal distribution, it is often noted as  $Z \sim N(0, 1)$ .

### Definition 2: z-score

The linear transformation of the form  $z = \frac{x-\mu}{\sigma}$ . If this transformation is applied to any normal distribution  $X \sim N(\mu, \sigma)$ , the result is the standard normal distribution  $Z \sim N(0, 1)$ . If this transformation is applied to any specific value  $x$  of the RV with mean  $\mu$  and standard deviation  $\sigma$ , the result is called the z-score of  $x$ . Z-scores allow us to compare data that are normally distributed but scaled differently.

\*Version 1.7: Feb 7, 2009 8:34 am US/Central

†<http://creativecommons.org/licenses/by/2.0/>



# NORMAL DISTRIBUTION: Z-SCORES\*

Susan Dean

Barbara Illowsky, Ph.D.

This work is produced by The Connexions Project and licensed under the Creative Commons Attribution License †

If  $X$  is a normally distributed random variable and  $X \sim N(\mu, \sigma)$ , then the z-score is:

$$z = \frac{x - \mu}{\sigma} \quad (1)$$

**The z-score tells you how many standard deviations that the value  $x$  is above (to the right of) or below (to the left of) the mean,  $\mu$ .** Values of  $x$  that are larger than the mean have positive z-scores and values of  $x$  that are smaller than the mean have negative z-scores. If  $x$  equals the mean, then  $x$  has a z-score of 0.

## Example 1

Suppose  $X \sim N(5, 6)$ . This says that  $X$  is a normally distributed random variable with mean  $\mu = 5$  and standard deviation  $\sigma = 6$ . Suppose  $x = 17$ . Then:

$$z = \frac{x - \mu}{\sigma} = \frac{17 - 5}{6} = 2 \quad (2)$$

This means that  $x = 17$  is **2 standard deviations** ( $2\sigma$ ) above or to the right of the mean  $\mu = 5$ . The standard deviation is  $\sigma = 6$ .

Notice that:

$$5 + 2 \cdot 6 = 17 \quad (\text{The pattern is } \mu + z\sigma = x.) \quad (3)$$

Now suppose  $x = 1$ . Then:

$$z = \frac{x - \mu}{\sigma} = \frac{1 - 5}{6} = -0.67 \quad (\text{rounded to two decimal places}) \quad (4)$$

**This means that  $x = 1$  is 0.67 standard deviations ( $-0.67\sigma$ ) below or to the left of the mean  $\mu = 5$ . Notice that:**

$5 + (-0.67)(6)$  is approximately equal to 1 (This has the pattern  $\mu + (-0.67)\sigma = 1$ )

Summarizing, when  $z$  is positive,  $x$  is above or to the right of  $\mu$  and when  $z$  is negative,  $x$  is to the left of or below  $\mu$ .

## Example 2

Some doctors believe that a person can lose 5 pounds, on the average, in a month by reducing his/her fat intake and by exercising consistently. Suppose weight loss has a normal distribution. Let  $X$  = the amount of weight lost (in pounds) by a person in a month. Use a standard deviation of 2 pounds.  $X \sim N(5, 2)$ . Fill in the blanks.

\*Version 1.7: Jan 23, 2009 10:12 am US/Central

†<http://creativecommons.org/licenses/by/2.0/>

**Problem 1***(Solution on p. 3.)*

Suppose a person **lost** 10 pounds in a month. The z-score when  $x = 10$  pounds is  $z = 2.5$  (verify). This z-score tells you that  $x = 10$  is \_\_\_\_\_ standard deviations to the \_\_\_\_\_ (right or left) of the mean \_\_\_\_\_. (What is the mean?).

**Problem 2***(Solution on p. 3.)*

Suppose a person **gained** 3 pounds (a negative weight loss). Then  $z =$  \_\_\_\_\_. This z-score tells you that  $x = -3$  is \_\_\_\_\_ standard deviations to the \_\_\_\_\_ (right or left) of the mean.

Suppose the random variables  $X$  and  $Y$  have the following normal distributions:  $X \sim N(5, 6)$  and  $Y \sim N(2, 1)$ . If  $x = 17$ , then  $z = 2$ . (This was previously shown.) If  $y = 4$ , what is  $z$ ?

$$z = \frac{y - \mu}{\sigma} = \frac{4 - 2}{1} = 2 \quad \text{where } \mu=2 \text{ and } \sigma=1. \quad (5)$$

The z-score for  $y = 4$  is  $z = 2$ . This means that 4 is  $z = 2$  standard deviations to the right of the mean. Therefore,  $x = 17$  and  $y = 4$  are both 2 (of **their**) standard deviations to the right of **their** respective means.

**The z-score allows us to compare data that are scaled differently.** To understand the concept, suppose  $X \sim N(5, 6)$  represents weight gains for one group of people who are trying to gain weight in a 6 week period and  $Y \sim N(2, 1)$  measures the same weight gain for a second group of people. A negative weight gain would be a weight loss. Since  $x = 17$  and  $y = 4$  are each 2 standard deviations to the right of their means, they represent the same weight gain **in relationship to their means**.

## Solutions to Exercises in this Module

### **Solution to Example 2, Problem 1 (p. 2)**

This z-score tells you that  $x = 10$  is **2.5** standard deviations to the **right** of the mean **5**.

### **Solution to Example 2, Problem 2 (p. 2)**

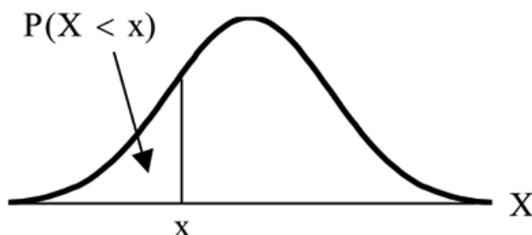
$z = -4$ . This z-score tells you that  $x = -3$  is **4** standard deviations to the **left** of the mean.

# NORMAL DISTRIBUTION: AREAS TO THE LEFT AND RIGHT OF $X^*$

Susan Dean  
Barbara Illowsky, Ph.D.

This work is produced by The Connexions Project and licensed under the Creative Commons Attribution License <sup>†</sup>

The arrow in the graph below points to the area to the left of  $x$ . This area is represented by the probability  $P(X < x)$ . Normal tables, computers, and calculators provide or calculate the probability  $P(X < x)$ .



The area to the right is then  $P(X > x) = 1 - P(X < x)$ .

Remember,  $P(X < x)$  = Area to the left of the vertical line through  $x$ .

$P(X > x) = 1 - P(X < x)$  = Area to the right of the vertical line through  $x$

$P(X < x)$  is the same as  $P(X \leq x)$  and  $P(X > x)$  is the same as  $P(X \geq x)$  for continuous distributions.

\*Version 1.5: Oct 27, 2008 6:41 pm GMT-5

<sup>†</sup><http://creativecommons.org/licenses/by/2.0/>

# NORMAL DISTRIBUTION: CALCULATIONS OF PROBABILITIES\*

Susan Dean  
Barbara Illowsky, Ph.D.

This work is produced by The Connexions Project and licensed under the Creative Commons Attribution License †

Probabilities are calculated by using technology. There are instructions in the chapter for the TI-83+ and TI-84 calculators.

NOTE: In the Table of Contents for **Collaborative Statistics**, entry **15. Tables** has a link to a table of normal probabilities. Use the probability tables if so desired, instead of a calculator.

## Example 1

If the area to the left is 0.0228, then the area to the right is  $1 - 0.0228 = 0.9772$ .

## Example 2

The final exam scores in a statistics class were normally distributed with a mean of 63 and a standard deviation of 5.

## Problem 1

Find the probability that a randomly selected student scored more than 65 on the exam.

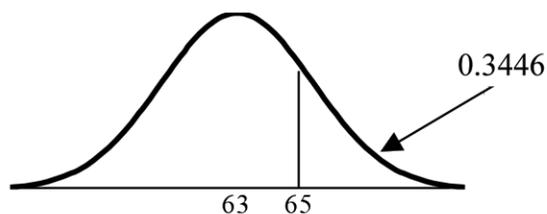
## Solution

Let  $X$  = a score on the final exam.  $X \sim N(63, 5)$ , where  $\mu = 63$  and  $\sigma = 5$

Draw a graph.

Then, find  $P(X > 65)$ .

$P(X > 65) = 0.3446$  (calculator or computer)



The probability that one student scores more than 65 is 0.3446.

Using the TI-83+ or the TI-84 calculators, the calculation is as follows. Go into 2nd DISTR.

After pressing 2nd DISTR, press 2:normalcdf.

The syntax for the instructions are shown below.

\*Version 1.10: Jun 19, 2011 10:07 pm GMT-5

†<http://creativecommons.org/licenses/by/3.0/>

normalcdf(lower value, upper value, mean, standard deviation) For this problem: normalcdf(65,1E99,63,5) = 0.3446. You get 1E99 (= 10<sup>99</sup>) by pressing 1, the EE key (a 2nd key) and then 99. Or, you can enter 10^99 instead. The number 10<sup>99</sup> is way out in the right tail of the normal curve. We are calculating the area between 65 and 10<sup>99</sup>. In some instances, the lower number of the area might be -1E99 (= -10<sup>99</sup>). The number -10<sup>99</sup> is way out in the left tail of the normal curve.

HISTORICAL NOTE: The TI probability program calculates a z-score and then the probability from the z-score. Before technology, the z-score was looked up in a standard normal probability table (because the math involved is too cumbersome) to find the probability. In this example, a standard normal table with area to the left of the z-score was used. You calculate the z-score and look up the area to the left. The probability is the area to the right.

$$z = \frac{65-63}{5} = 0.4 \quad . \text{ Area to the left is } 0.6554. \quad P(X > 65) = P(Z > 0.4) = 1 - 0.6554 = 0.3446$$

**Problem 2**

Find the probability that a randomly selected student scored less than 85.

**Solution**

Draw a graph.

Then find  $P(X < 85)$ . Shade the graph.  $P(X < 85) = 1$  (calculator or computer)

The probability that one student scores less than 85 is approximately 1 (or 100%).

The TI-instructions and answer are as follows:

$$\text{normalcdf}(0,85,63,5) = 1 \text{ (rounds to 1)}$$

**Problem 3**

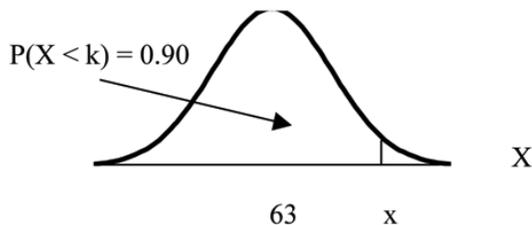
Find the 90th percentile (that is, find the score k that has 90 % of the scores below k and 10% of the scores above k).

**Solution**

Find the 90th percentile. For each problem or part of a problem, draw a new graph. Draw the x-axis. Shade the area that corresponds to the 90th percentile.

**Let  $k =$  the 90th percentile.**  $k$  is located on the x-axis.  $P(X < k)$  is the area to the left of  $k$ . The 90th percentile  $k$  separates the exam scores into those that are the same or lower than  $k$  and those that are the same or higher. Ninety percent of the test scores are the same or lower than  $k$  and 10% are the same or higher.  $k$  is often called a **critical value**.

$$k = 69.4 \text{ (calculator or computer)}$$



The 90th percentile is 69.4. This means that 90% of the test scores fall at or below 69.4 and 10% fall at or above. For the TI-83+ or TI-84 calculators, use invNorm in 2nd DISTR. invNorm(area to the left, mean, standard deviation) For this problem, invNorm(.90,63,5) = 69.4

**Problem 4**

Find the 70th percentile (that is, find the score  $k$  such that 70% of scores are below  $k$  and 30% of the scores are above  $k$ ).

**Solution**

Find the 70th percentile.

Draw a new graph and label it appropriately.  $k = 65.6$

The 70th percentile is 65.6. This means that 70% of the test scores fall at or below 65.5 and 30% fall at or above.

$$\text{invNorm}(.70,63,5) = 65.6$$

**Example 3**

More and more households in the United States have at least one computer. The computer is used for office work at home, research, communication, personal finances, education, entertainment, social networking and a myriad of other things. Suppose that the average number of hours a household personal computer is used for entertainment is 2 hours per day. Assume the times for entertainment are normally distributed and the standard deviation for the times is half an hour.

**Problem 1**

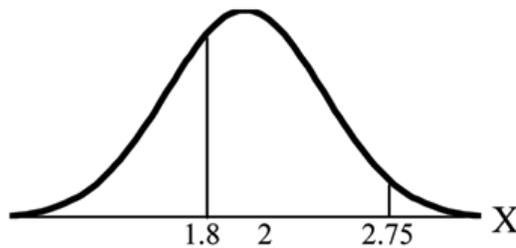
Find the probability that a household personal computer is used between 1.8 and 2.75 hours per day.

**Solution**

Let  $X$  = the amount of time (in hours) a household personal computer is used for entertainment.  $X \sim N(2, 0.5)$  where  $\mu = 2$  and  $\sigma = 0.5$ .

Find  $P(1.8 < X < 2.75)$ .

The probability for which you are looking is the area **between**  $x = 1.8$  and  $x = 2.75$ .  $P(1.8 < X < 2.75) = 0.5886$



$$\text{normalcdf}(1.8,2.75,2,.5) = 0.5886$$

The probability that a household personal computer is used between 1.8 and 2.75 hours per day for entertainment is 0.5886.

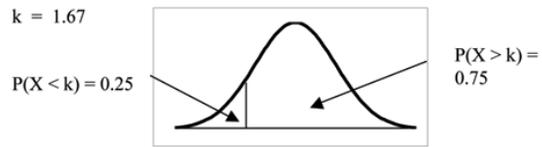
**Problem 2**

Find the maximum number of hours per day that the bottom quartile of households use a personal computer for entertainment.

**Solution**

To find the maximum number of hours per day that the bottom quartile of households uses a personal computer for entertainment, **find the 25th percentile,  $k$** , where  $P(X < k) = 0.25$ .





$$\text{invNorm}(.25, 2, .5) = 1.67$$

The maximum number of hours per day that the bottom quartile of households uses a personal computer for entertainment is 1.67 hours.



# NORMAL DISTRIBUTION: HOMEWORK\*

Susan Dean  
Barbara Illowsky, Ph.D.

This work is produced by The Connexions Project and licensed under the  
Creative Commons Attribution License †

## Exercise 1

(Solution on p. 7.)

According to a study done by De Anza students, the height for Asian adult males is normally distributed with an average of 66 inches and a standard deviation of 2.5 inches. Suppose one Asian adult male is randomly chosen. Let  $X$  = height of the individual.

- $X \sim \text{_____}(\text{_____}, \text{_____})$
- Find the probability that the person is between 65 and 69 inches. Include a sketch of the graph and write a probability statement.
- Would you expect to meet many Asian adult males over 72 inches? Explain why or why not, and justify your answer numerically.
- The middle 40% of heights fall between what two values? Sketch the graph and write the probability statement.

## Exercise 2

IQ is normally distributed with a mean of 100 and a standard deviation of 15. Suppose one individual is randomly chosen. Let  $X$  = IQ of an individual.

- $X \sim \text{_____}(\text{_____}, \text{_____})$
- Find the probability that the person has an IQ greater than 120. Include a sketch of the graph and write a probability statement.
- Mensa is an organization whose members have the top 2% of all IQs. Find the minimum IQ needed to qualify for the Mensa organization. Sketch the graph and write the probability statement.
- The middle 50% of IQs fall between what two values? Sketch the graph and write the probability statement.

## Exercise 3

(Solution on p. 7.)

The percent of fat calories that a person in America consumes each day is normally distributed with a mean of about 36 and a standard deviation of 10. Suppose that one individual is randomly chosen. Let  $X$  = percent of fat calories.

- $X \sim \text{_____}(\text{_____}, \text{_____})$
- Find the probability that the percent of fat calories a person consumes is more than 40. Graph the situation. Shade in the area to be determined.

\*Version 1.20: May 12, 2010 9:14 pm GMT-5

†<http://creativecommons.org/licenses/by/2.0/>

- c. Find the maximum number for the lower quarter of percent of fat calories. Sketch the graph and write the probability statement.

**Exercise 4**

Suppose that the distance of fly balls hit to the outfield (in baseball) is normally distributed with a mean of 250 feet and a standard deviation of 50 feet.

- a. If  $X$  = distance in feet for a fly ball, then  $X \sim \text{_____}(\text{_____}, \text{_____})$
- b. If one fly ball is randomly chosen from this distribution, what is the probability that this ball traveled fewer than 220 feet? Sketch the graph. Scale the horizontal axis  $X$ . Shade the region corresponding to the probability. Find the probability.
- c. Find the 80th percentile of the distribution of fly balls. Sketch the graph and write the probability statement.

**Exercise 5**

*(Solution on p. 7.)*

In China, 4-year-olds average 3 hours a day unsupervised. Most of the unsupervised children live in rural areas, considered safe. Suppose that the standard deviation is 1.5 hours and the amount of time spent alone is normally distributed. We randomly survey one Chinese 4-year-old living in a rural area. We are interested in the amount of time the child spends alone per day. (Source: **San Jose Mercury News**)

- a. In words, define the random variable  $X$ .  $X =$
- b.  $X \sim$
- c. Find the probability that the child spends less than 1 hour per day unsupervised. Sketch the graph and write the probability statement.
- d. What percent of the children spend over 10 hours per day unsupervised?
- e. 70% of the children spend at least how long per day unsupervised?

**Exercise 6**

In the 1992 presidential election, Alaska’s 40 election districts averaged 1956.8 votes per district for President Clinton. The standard deviation was 572.3. (There are only 40 election districts in Alaska.) The distribution of the votes per district for President Clinton was bell-shaped. Let  $X$  = number of votes for President Clinton for an election district. (Source: **The World Almanac and Book of Facts**)

- a. State the approximate distribution of  $X$ .  $X \sim$
- b. Is 1956.8 a population mean or a sample mean? How do you know?
- c. Find the probability that a randomly selected district had fewer than 1600 votes for President Clinton. Sketch the graph and write the probability statement.
- d. Find the probability that a randomly selected district had between 1800 and 2000 votes for President Clinton.
- e. Find the third quartile for votes for President Clinton.

**Exercise 7**

*(Solution on p. 7.)*

Suppose that the duration of a particular type of criminal trial is known to be normally distributed with a mean of 21 days and a standard deviation of 7 days.

- a. In words, define the random variable  $X$ .  $X =$
- b.  $X \sim$
- c. If one of the trials is randomly chosen, find the probability that it lasted at least 24 days. Sketch the graph and write the probability statement.
- d. 60% of all of these types of trials are completed within how many days?



**Exercise 8**

Terri Vogel, an amateur motorcycle racer, averages 129.71 seconds per 2.5 mile lap (in a 7 lap race) with a standard deviation of 2.28 seconds . The distribution of her race times is normally distributed. We are interested in one of her randomly selected laps. (Source: log book of Terri Vogel)

- a. In words, define the random variable  $X$ .  $X =$
- b.  $X \sim$
- c. Find the percent of her laps that are completed in less than 130 seconds.
- d. The fastest 3% of her laps are under \_\_\_\_\_ .
- e. The middle 80% of her laps are from \_\_\_\_\_ seconds to \_\_\_\_\_ seconds.

**Exercise 9**

*(Solution on p. 7.)*

Thuy Dau, Ngoc Bui, Sam Su, and Lan Voung conducted a survey as to how long customers at Lucky claimed to wait in the checkout line until their turn. Let  $X =$ time in line. Below are the ordered real data (in minutes):

0.50	4.25	5	6	7.25
1.75	4.25	5.25	6	7.25
2	4.25	5.25	6.25	7.25
2.25	4.25	5.5	6.25	7.75
2.25	4.5	5.5	6.5	8
2.5	4.75	5.5	6.5	8.25
2.75	4.75	5.75	6.5	9.5
3.25	4.75	5.75	6.75	9.5
3.75	5	6	6.75	9.75
3.75	5	6	6.75	10.75

**Table 1**

- a. Calculate the sample mean and the sample standard deviation.
- b. Construct a histogram. Start the  $x -$  axis at  $-0.375$  and make bar widths of 2 minutes.
- c. Draw a smooth curve through the midpoints of the tops of the bars.
- d. In words, describe the shape of your histogram and smooth curve.
- e. Let the sample mean approximate  $\mu$  and the sample standard deviation approximate  $\sigma$ . The distribution of  $X$  can then be approximated by  $X \sim$
- f. Use the distribution in (e) to calculate the probability that a person will wait fewer than 6.1 minutes.
- g. Determine the cumulative relative frequency for waiting less than 6.1 minutes.
- h. Why aren't the answers to (f) and (g) exactly the same?
- i. Why are the answers to (f) and (g) as close as they are?
- j. If only 10 customers were surveyed instead of 50, do you think the answers to (f) and (g) would have been closer together or farther apart? Explain your conclusion.

**Exercise 10**

Suppose that Ricardo and Anita attend different colleges. Ricardo's GPA is the same as the average GPA at his school. Anita's GPA is 0.70 standard deviations above her school average. In complete sentences, explain why each of the following statements may be false.



- a. Ricardo’s actual GPA is lower than Anita’s actual GPA.
- b. Ricardo is not passing since his z-score is zero.
- c. Anita is in the 70th percentile of students at her college.

**Exercise 11**

*(Solution on p. 7.)*

Below is a sample of the maximum capacity (maximum number of spectators) of sports stadiums.

The table does not include horse racing or motor racing stadiums. (Source: [http://en.wikipedia.org/wiki/List\\_of\\_sports\\_stadiums](http://en.wikipedia.org/wiki/List_of_sports_stadiums))

40,000	40,000	45,050	45,500	46,249	48,134
49,133	50,071	50,096	50,466	50,832	51,100
51,500	51,900	52,000	52,132	52,200	52,530
52,692	53,864	54,000	55,000	55,000	55,000
55,000	55,000	55,000	55,082	57,000	58,008
59,680	60,000	60,000	60,492	60,580	62,380
62,872	64,035	65,000	65,050	65,647	66,000
66,161	67,428	68,349	68,976	69,372	70,107
70,585	71,594	72,000	72,922	73,379	74,500
75,025	76,212	78,000	80,000	80,000	82,300

**Table 2**

- a. Calculate the sample mean and the sample standard deviation for the maximum capacity of sports stadiums (the data).
- b. Construct a histogram of the data.
- c. Draw a smooth curve through the midpoints of the tops of the bars of the histogram.
- d. In words, describe the shape of your histogram and smooth curve.
- e. Let the sample mean approximate  $\mu$  and the sample standard deviation approximate  $\sigma$ . The distribution of  $X$  can then be approximated by  $X \sim N(\mu, \sigma)$ .
- f. Use the distribution in (e) to calculate the probability that the maximum capacity of sports stadiums is less than 67,000 spectators.
- g. Determine the cumulative relative frequency that the maximum capacity of sports stadiums is less than 67,000 spectators. Hint: Order the data and count the sports stadiums that have a maximum capacity less than 67,000. Divide by the total number of sports stadiums in the sample.
- h. Why aren’t the answers to (f) and (g) exactly the same?

**1 Try These Multiple Choice Questions**

**The questions below refer to the following:** The patient recovery time from a particular surgical procedure is normally distributed with a mean of 5.3 days and a standard deviation of 2.1 days.

**Exercise 12**

*(Solution on p. 7.)*

What is the median recovery time?

- A. 2.7



- B. 5.3
- C. 7.4
- D. 2.1

**Exercise 13***(Solution on p. 7.)*

What is the z-score for a patient who takes 10 days to recover?

- A. 1.5
- B. 0.2
- C. 2.2
- D. 7.3

**Exercise 14***(Solution on p. 7.)*

What is the probability of spending more than 2 days in recovery?

- A. 0.0580
- B. 0.8447
- C. 0.0553
- D. 0.9420

**Exercise 15***(Solution on p. 7.)*

The 90th percentile for recovery times is?

- A. 8.89
- B. 7.07
- C. 7.99
- D. 4.32

**The questions below refer to the following:** The length of time to find a parking space at 9 A.M. follows a normal distribution with a mean of 5 minutes and a standard deviation of 2 minutes.

**Exercise 16***(Solution on p. 7.)*

Based upon the above information and numerically justified, would you be surprised if it took less than 1 minute to find a parking space?

- A. Yes
- B. No
- C. Unable to determine

**Exercise 17***(Solution on p. 7.)*

Find the probability that it takes at least 8 minutes to find a parking space.

- A. 0.0001
- B. 0.9270
- C. 0.1862
- D. 0.0668

**Exercise 18***(Solution on p. 7.)*

Seventy percent of the time, it takes more than how many minutes to find a parking space?

- A. 1.24
- B. 2.41
- C. 3.95

**D. 6.05**

**Exercise 19**

*(Solution on p. 8.)*

If the mean is significantly greater than the standard deviation, which of the following statements is true?

- I . The data cannot follow the uniform distribution.
- II . The data cannot follow the exponential distribution..
- III . The data cannot follow the normal distribution.

- A. I only
- B. II only
- C. III only
- D. I, II, and III



## Solutions to Exercises in this Module

### Solution to Exercise 1 (p. 1)

- a.  $N(66, 2.5)$
- b. 0.5404
- c. No
- d. Between 64.7 and 67.3 inches

### Solution to Exercise 3 (p. 1)

- a.  $N(36, 10)$
- b. 0.3446
- c. 29.3

### Solution to Exercise 5 (p. 2)

- a. the time (in hours) a 4-year-old in China spends unsupervised per day
- b.  $N(3, 1.5)$
- c. 0.0912
- d. 0
- e. 2.21 hours

### Solution to Exercise 7 (p. 2)

- a. The duration of a criminal trial
- b.  $N(21, 7)$
- c. 0.3341
- d. 22.77

### Solution to Exercise 9 (p. 3)

- a. The sample mean is 5.51 and the sample standard deviation is 2.15
- e.  $N(5.51, 2.15)$
- f. 0.6081
- g. 0.64

### Solution to Exercise 11 (p. 4)

- a. The sample mean is 60,136.4 and the sample standard deviation is 10,468.1.
- e.  $N(60136.4, 10468.1)$
- f. 0.7440
- g. 0.7167

### Solution to Exercise 12 (p. 4)

B

### Solution to Exercise 13 (p. 5)

C

### Solution to Exercise 14 (p. 5)

D

### Solution to Exercise 15 (p. 5)

C

### Solution to Exercise 16 (p. 5)

A

### Solution to Exercise 17 (p. 5)

D

**Solution to Exercise 18 (p. 5)**

C

**Solution to Exercise 19 (p. 6)**

B



# NORMAL DISTRIBUTION: PRACTICE\*

Susan Dean  
Barbara Illowsky, Ph.D.

This work is produced by The Connexions Project and licensed under the  
Creative Commons Attribution License †

## 1 Student Learning Outcomes

- The student will explore the properties of data with a normal distribution.

## 2 Given

The life of Sunshine CD players is normally distributed with a mean of 4.1 years and a standard deviation of 1.3 years. A CD player is guaranteed for 3 years. We are interested in the length of time a CD player lasts.

## 3 Normal Distribution

### Exercise 1

Define the Random Variable  $X$  in words.  $X =$

### Exercise 2

$X \sim$

### Exercise 3

(Solution on p. 4.)

Find the probability that a CD player will break down during the guarantee period.

- a. Sketch the situation. Label and scale the axes. Shade the region corresponding to the probability.

---

\*Version 1.9: Feb 18, 2009 7:53 pm US/Central

†<http://creativecommons.org/licenses/by/2.0/>



Figure 1

b.  $P(0 < X < \text{-----}) = \text{-----}$

**Exercise 4**

*(Solution on p. 4.)*

Find the probability that a CD player will last between 2.8 and 6 years.

- a. Sketch the situation. Label and scale the axes. Shade the region corresponding to the probability.



Figure 2

b.  $P(\text{-----} < X < \text{-----}) = \text{-----}$

**Exercise 5**

*(Solution on p. 4.)*

Find the 70th percentile of the distribution for the time a CD player lasts.

- a. Sketch the situation. Label and scale the axes. Shade the region corresponding to the lower 70%.

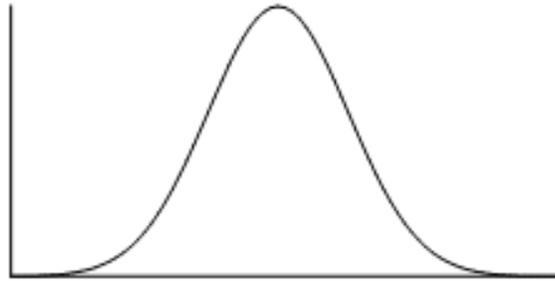


Figure 3

b.  $P(X < k) = \text{-----}$ . Therefore,  $k = \text{-----}$ .

## Solutions to Exercises in this Module

### Solution to Exercise 3 (p. 1)

b. 3, 0.1979

### Solution to Exercise 4 (p. 2)

b. 2.8, 6, 0.7694

### Solution to Exercise 5 (p. 2)

b. 0.70, 4.78years

