

# The Central Limit Theorem: Homework

## EXERCISE 1

$X \sim N(60, 9)$ . Suppose that you form random samples of 25 from this distribution. Let  $\bar{X}$  be the random variable of averages. Let  $\Sigma X$  be the random variable of sums. For **c - f**, sketch the graph, shade the region, label and scale the horizontal axis for  $\bar{X}$ , and find the probability.

- Sketch the distributions of  $X$  and  $\bar{X}$  on the same graph.
- $\bar{X} \sim$
- $P(\bar{X} < 60) =$
- Find the 30<sup>th</sup> percentile.
- $P(56 < \bar{X} < 62) =$
- $P(18 < \bar{X} < 58) =$
- $\Sigma X \sim$
- Find the minimum value for the upper quartile.
- $P(1400 < \Sigma X < 1550) =$

## EXERCISE 2

Determine which of the following are true and which are false. Then, in complete sentences, justify your answers.

- When the sample size is large, the mean of  $\bar{X}$  is approximately equal to the mean of  $X$ .
- When the sample size is large,  $\bar{X}$  is approximately normally distributed.
- When the sample size is large, the standard deviation of  $\bar{X}$  is approximately the same as the standard deviation of  $X$ .

## EXERCISE 3

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 about 10. Suppose that 16 individuals are randomly chosen.

Let  $\bar{X}$  = average percent of fat calories.

- $\bar{X} \sim$  \_\_\_\_\_ ( \_\_\_\_\_ , \_\_\_\_\_ )
- For the group of 16, find the probability that the average percent of fat calories consumed is more than 5. Graph the situation and shade in the area to be determined.
- Find the first quartile for the average percent of fat calories.

#### EXERCISE 4

Previously, De Anza statistics students estimated that the amount of change daytime statistics students carry is exponentially distributed with a mean of \$0.88. Suppose that we randomly pick 25 daytime statistics students.

- In words,  $X =$
- $X \sim$
- In words,  $\bar{X} =$
- $\bar{X} \sim$  \_\_\_\_\_ ( \_\_\_\_\_ , \_\_\_\_\_ )
- Find the probability that an individual had between \$0.80 and \$1.00. Graph the situation and shade in the area to be determined.
- Find the probability that the average of the 25 students was between \$0.80 and \$1.00. Graph the situation and shade in the area to be determined.
- Explain the why there is a difference in (e) and (f).

#### EXERCISE 5

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. We randomly sample 49 fly balls.

- If  $\bar{X}$  = average distance in feet for 49 fly balls, then  $\bar{X} \sim$  \_\_\_\_\_ ( \_\_\_\_\_ , \_\_\_\_\_ )
- What is the probability that the 49 balls traveled an average of less than 240 feet? Sketch the graph. Scale the horizontal axis for  $\bar{X}$ . Shade the region corresponding to the probability. Find the probability.
- Find the 80th percentile of the distribution of the average of 49 fly balls.

#### EXERCISE 6

Suppose that the weight of open boxes of cereal in a home with children is uniformly distributed from 2 to 6 pounds. We randomly survey 64 homes with children.

- In words,  $X =$
- $X \sim$
- $\mu_x =$
- $\sigma_x =$
- In words,  $\Sigma X =$
- $\Sigma X \sim$
- Find the probability that the total weight of open boxes is less than 250 pounds.
- Find the 35th percentile for the total weight of open boxes of cereal.

### EXERCISE 7

Suppose that the duration of a particular type of criminal trial is known to have a mean of 21 days and a standard deviation of 7 days. We randomly sample 9 trials.

- In words,  $\Sigma X =$
- $\Sigma X \sim$
- Find the probability that the total length of the 9 trials is at least 225 days.
- 90 percent of the total of 9 of these types of trials will last at least how long?

### EXERCISE 8

According to the Internal Revenue Service, the average length of time for an individual to complete (record keep, learn, prepare, copy, assemble and send) IRS Form 1040 is 10.53 hours (without any attached schedules). The distribution is unknown. Let us assume that the standard deviation is 2 hours. Suppose we randomly sample 36 taxpayers.

- In words,  $X =$
- In words,  $\bar{X} =$
- $\bar{X} \sim$
- Would you be surprised if the 36 taxpayers finished their Form 1040s in an average of more than 12 hours? Explain why or why not in complete sentences.
- Would you be surprised if one taxpayer finished his Form 1040 in more than 12 hours? In a complete sentence, explain why.

### EXERCISE 9

Suppose that a category of world class runners are known to run a marathon (26 miles) in an average of 145 minutes with a standard deviation of 14 minutes. Consider 49 of the races.

Let  $\bar{X}$  = the average of the 49 races.

- $\bar{X} \sim$
- Find the probability that the runner will average between 142 and 146 minutes in these 49 marathons.
- Find the 80th percentile for the average of these 49 marathons.
- Find the median of the average running times.

### EXERCISE 10

The attention span of a two year-old is exponentially distributed with a mean of about 8 minutes. Suppose we randomly survey 60 two year-olds.

- a. In words,  $X =$
- b.  $X \sim$
- c. In words,  $\bar{X} =$
- d.  $\bar{X} \sim$
- e. Before doing any calculations, which do you think will be higher:
  - the probability that an individual attention span is less than 10 minutes or
  - the probability that the average attention span for the 60 children is less than 10 minutes? Why?
- f. Calculate the probabilities in part (e).
- g. Explain why the distribution for  $\bar{X}$  is not exponential.

#### EXERCISE 11

Suppose that the length of research papers is uniformly distributed from 10 to 25 pages. We survey a class in which 55 research papers were turned in to a professor. We are interested in the average length of the research papers.

- a. In words,  $X =$
- b.  $X \sim$
- c.  $\mu_x =$
- d.  $\sigma_x =$
- e. In words,  $\bar{X} =$
- f.  $\bar{X} \sim$
- g. In words,  $\Sigma X =$
- h.  $\Sigma X \sim$
- i. Without doing any calculations, do you think that it's likely that the professor will need to read a total of more than 1050 pages? Why?
- j. Calculate the probability that the professor will need to read a total of more than 1050 pages.
- k. Why is it so unlikely that the average length of the papers will be less than 12 pages?

#### EXERCISE 12

The length of songs in a collector's CD collection is uniformly distributed from 2 to 3.5 minutes. Suppose we randomly pick 5 CDs from the collection. There is a total of 43 songs on the 5 CDs.

- a. In words,  $X =$
- b.  $X \sim$

- c. In words,  $\bar{X}$  =
- d.  $\bar{X} \sim$
- e. Find the first quartile for the average song length.
- f. The IQR (interquartile range) for the average song length is from \_\_\_\_\_ to \_\_\_\_\_.

### EXERCISE 13

Salaries for teachers in a particular elementary school district are normally distributed with a mean of \$44,000 and a standard deviation of \$6500. We randomly survey 10 teachers from that district.

- a. In words,  $X$  =
- b. In words,  $\bar{X}$  =
- c.  $\bar{X} \sim$
- d. In words,  $\Sigma X$  =
- e.  $\Sigma X \sim$
- f. Find the probability that the teachers earn a total of over \$400,000.
- g. Find the 90th percentile for an individual teacher's salary.
- h. Find the 90th percentile for the average teachers' salary.
- i. If we surveyed 70 teachers instead of 10, graphically, how would that change the distribution for  $\bar{X}$  ?
- j. If each of the 70 teachers received a \$3000 raise, graphically, how would that change the distribution for  $\bar{X}$  ?

### EXERCISE 14

The distribution of income in some Third World countries is considered wedge shaped (many very poor people, very few middle income people, and few to many wealthy people). Suppose we pick a country with a wedge distribution. Let the average salary be \$2000 per year with a standard deviation of \$8000. We randomly survey 1000 residents of that country.

- a. In words,  $X$  =
- b. In words,  $\bar{X}$  =
- c.  $\bar{X} \sim$
- d. How is it possible for the standard deviation to be greater than the average?
- e. Why is it more likely that the average of the 1000 residents will be from \$2000 to \$2100 than from \$2100 to \$2200?

### EXERCISE 15

The average length of a maternity stay in a U.S. hospital is said to be 2.4 days with a standard deviation of 0.9 days. We randomly survey 80 women who recently bore children in a U.S. hospital.

- a. In words,  $X =$
- b. In words,  $\bar{X} =$
- c.  $\bar{X} \sim$
- d. In words,  $\Sigma X =$
- e.  $\Sigma X \sim$
- f. Is it likely that an individual stayed more than 5 days in the hospital? Why or why not?
- g. Is it likely that the average stay for the 80 women was more than 5 days? Why or why not?
- h. Which is more likely:
  - an individual stayed more than 5 days; or
  - the average stay of 80 women was more than 5 days?
- i. If we were to sum up the women's stays, is it likely that, collectively they spent more than a year in the hospital? Why or why not?

### EXERCISE 16

In 1940 the average size of a U.S. farm was 174 acres. Let's say that the standard deviation was 55 acres. Suppose we randomly survey 38 farmers from 1940. (Source: U.S. Dept. of Agriculture)

- a. In words,  $X =$
- b. In words,  $\bar{X} =$
- c.  $\bar{X} \sim$
- d. The IQR for  $\bar{X}$  is from \_\_\_\_\_ acres to \_\_\_\_\_ acres.

### EXERCISE 17

The stock closing prices of 35 U.S. semiconductor manufacturers are given below. (Source: *Wall Street Journal*)

8.625, 30.25, 27.625, 46.75, 32.875, 18.25, 5, 0.125, 2.9375, 6.875, 28.25, 24.25, 21, 1.5, 30.25, 71, 43.5, 49.25, 2.5625, 31, 16.5, 9.5, 18.5, 18, 9, 10.5, 16.625, 1.25, 18, 12.875, 7, 12.875, 2.875, 60.25, 29.25

- a. In words,  $X =$
- b. Complete the Following:
  - $\bar{x} =$  \_\_\_\_\_
  - $s_x =$  \_\_\_\_\_
  - $n =$  \_\_\_\_\_
- c. Construct a histogram of the distribution. Start at  $x = -0.0005$ . Make bar widths of 10.
- d. In words, describe the distribution of stock prices.
- e. Randomly average 5 stock prices together. (Use a random number generator.) Continue averaging 5 pieces together until you have 10 averages. List those 10 averages.
- f. Use the 10 averages from (e) to calculate:
  - $\bar{x} =$  \_\_\_\_\_
  - $s_{\bar{x}} =$  \_\_\_\_\_
- g. Construct a histogram of the distribution of the averages. Start at  $x = -0.0005$ . Make bar widths of 10.
- h. Does this histogram look like the graph in (c)?
- i. In 1 - 2 complete sentences, explain why the graphs either look the same or look different?
- j. Based upon the theory of the Central Limit Theorem,  $\bar{X} \sim$  \_\_\_\_\_

### EXERCISE 18

Use the Initial Public Offering data (see "Table of Contents") to do this problem.

- a. In words,  $X =$
- b. Complete the following:
  - $\mu_x =$  \_\_\_\_\_
  - $\sigma_x =$  \_\_\_\_\_
  - $n =$  \_\_\_\_\_
- c. Construct a histogram of the distribution. Start at  $x = -0.50$ . Make bar widths of \$5.
- d. In words, describe the distribution of stock prices.
- e. Randomly average 5 stock prices together. (Use a random number generator.) Continue averaging 5 pieces together until you have 15 averages. List those 15 averages.
- f. Use the 15 averages from (e) to calculate:
  - $\bar{x} =$  \_\_\_\_\_
  - $s_{\bar{x}} =$  \_\_\_\_\_

- g. Construct a histogram of the distribution of the averages. Start at  $x = -0.50$ . Make bar widths of \$5.
- h. Does this histogram look like the graph in (c)? Explain any differences.
- i. In 1 - 2 complete sentences, explain why the graphs either look the same or look different?
- j. Based upon the theory of the Central Limit Theorem,  $\bar{X} \sim$  \_\_\_\_\_

Try these multiple choice questions.

**Questions 19 and 20 refer to the following:** The time to wait for a particular rural bus is distributed uniformly from 0 to 75 minutes. 100 riders are randomly sampled to learn how long they waited.

#### EXERCISE 19

The 90<sup>th</sup> percentile sample average wait time (in minutes) for a sample of 100 riders is:

- A. 315.0
- B. 40.3
- C. 38.5
- D. 65.2

#### EXERCISE 20

Would you be surprised, based upon numerical calculations, if the sample average wait time (in minutes) for 100 riders was less than 30 minutes?

- A. Yes
- B. No
- C. There is not enough information.

#### EXERCISE 21

Which of the following is NOT TRUE about the distribution for averages?

- A. the mean, median and mode are equal
- B. the area under the curve is one
- C. the curve never touches the x-axis
- D. the curve is skewed to the right

**For the next three questions, use the following information:** The average cost of unleaded gasoline in the Bay Area once followed an unknown distribution with a mean of \$2.59 and a standard deviation of



\$0.10. Sixteen gas stations from the Bay Area are randomly chosen. We are interested in the average cost of gasoline for the 16 gas stations.

**EXERCISE 22**

The distribution to use for the average cost of gasoline for the 16 gas stations is

- A.  $\bar{X} \sim N(2.59, 0.10)$
- B.  $\bar{X} \sim N\left(2.59, \frac{0.10}{\sqrt{16}}\right)$
- C.  $\bar{X} \sim N\left(2.59, \frac{0.10}{16}\right)$
- D.  $\bar{X} \sim N\left(2.59, \frac{16}{0.10}\right)$

**EXERCISE 23**

What is the probability that the average price for 16 gas stations is over \$2.69?

- A. almost zero
- B. 0.1587
- C. 0.0943
- D. Unknown

**EXERCISE 24**

Find the probability that the average price for 30 gas stations is less than \$2.55.

- A. 0.6554
- B. 0.3446
- C. 0.0142
- D. 0.9858
- E. 0

**EXERCISE 25**

For the Charter School Problem (Example 3) in **Central Limit Theorem: Using the Central Limit Theorem**, calculate the following using the normal approximation to the binomial.

- A. Find the probability that less than 100 favor a charter school for grades K – 5
- B. Find the probability that 170 or more favor a charter school for grades K – 5
- C. Find the probability that no more than 140 favor a charter school for grades K – 5

- D. Find the probability that there are fewer than 130 that favor a charter school for grades K – 5.
- E. Find the probability that exactly 150 favor a charter school for grades K – 5.

If you either have access to an appropriate calculator or computer software, try calculating these probabilities using the technology. Try also using the suggestion that is at the bottom of **Central Limit Theorem: Using the Central Limit Theorem** for finding a website that calculates binomial probabilities.

#### EXERCISE 26

Four friends, Janice, Barbara, Kathy and Roberta, decided to carpool together to get to school. Each day the driver would be chosen by randomly selecting one of the four names. They carpool to school for 96 days. Use the normal approximation to the binomial to calculate the following probabilities.

- A. Find the probability that Janice is the driver at most 20 days.
- B. Find the probability that Roberta is the driver more than 16 days.
- C. Find the probability that Barbara drives exactly 24 of those 96 days.

If you either have access to an appropriate calculator or computer software, try calculating these probabilities using the technology. Try also using the suggestion that is at the bottom of **Central Limit Theorem: Using the Central Limit Theorem** for finding a website that calculates binomial probabilities.