CHAPTER 8

Samples, Sampling Distributions, and Confidence Intervals

Summary

In this chapter you were introduced to sampling (random and research samples), sampling distributions, a new use for z scores, a statistic called t, and confidence intervals.

Random samples are a subset of the population selected using a method that allows every possible sample of size N an equal chance of being selected. Randomly select a sample using a computer program or random number table. Random samples are the best way to ensure a sample mirrors the population. However, in research situations, random samples are usually not feasible. Instead, ***research samples*** are used and other methods, such as ***replication***, help establish the validity of the results. Avoid ***biased samples***, in which certain segments of the population are systematically over- or under-selected.

Sampling distributions are theoretical distributions of a statistic based on an infinite number of random samples with the same size N. A sampling distribution is always about a particular statistic (e.g., the mean, variance, or correlation coefficient). The sampling distribution about the *mean*, for example, is a theoretical distribution of all possible sample *means* of same sized samples from a particular population. Sampling distributions are useful because they give information about the probability of a particular statistical outcome. For example, consulting a sampling distribution can tell you how likely it is to randomly select a sample of 16 people with a mean IQ of 110 from a population with a mean IQ of 100 and standard deviation of 15. Probability of statistical outcomes is important information for inferential statistics, and you will use sampling distributions throughout the rest of the course. The standard deviation of a sampling distribution is called the ***standard error***; the mean of a sampling distribution is called the ***expected value***.

Sampling distributions for different statistics come in many shapes (as you will see in later chapters). However, the ***Central Limit Theorem*** states that regardless of the form of the population from which the samples are drawn, the sampling distribution about the *mean* a) approaches a normal curve as sample size increases, b) has a mean (expected value) equal to μ, and c) has a standard deviation (standard error) equal to $\frac{\sigma}{\sqrt{N}}$.

Examining the formula for the standard error of the mean demonstrates the sampling distribution of the mean becomes narrower as N is increased. This is important because the sample means will cluster more closely to the expected value and therefore, the true population mean.

If the Central Limit Theorem applies, you can use the normal distribution to find the probability of selecting a sample with a specific mean from a particular population. Convert the mean to a z score and use the z table to determine the probability associated with that z score. If the probability is very low, this is reason to believe the sample may have been selected from a different population.

You are justified using the normal curve to determine probabilities when you know σ and you have an adequate sample. If you cannot meet those requirements (a very common problem), estimate σ using $\hat{s}$ and use the ***t*** distribution instead. The shape of the t distribution gets closer to normal as degrees of freedom increase (for t, degrees of freedom are based on sample size).

A ***confidence interval*** is an inferential statistical technique that identifies, with a specified degree of confidence (such as 95%), an interval of scores (defined by a lower and upper limit) that contain the value of an unknown parameter. For this chapter, the parameter of interest is the mean. Calculate a confidence interval about the mean using a sample mean, a t value from the t table, and an estimate of the standard error. Confidence intervals can be added to a graph to illustrate how much confidence one has that population means are different from each other when only sample data is available.

Multiple-Choice Questions

1. According to your text, if you draw a random sample, you are assured that _____.

    a.   the sample will always mirror the population
    b.   you will be somewhat uncertain about the population
    c.   the conclusions you draw will be correct
    d.   none of the above

2. A biased sample is one that _____.

    a.   is too small
    b.   will always lead to a wrong conclusion
    c.   has certain groups from the population overrepresented or underrepresented
    d.   is always nonrepresentative

3. Uncertainty regarding conclusions about a population can be eliminated by _____.

    a.   drawing a sample
    b.   drawing a large sample
    c.   drawing a large, random sample
    d.   none of the above

4. As N increases, the mean of the sample is _____.

    a.   likely to increase in size
    b.   likely to decrease in size
    c.   less likely to be representative of the population mean
    d.   more likely to be representative of the population mean

5. The difference between a random sample and a biased sample is _____.

    a.   biased samples are larger
    b.   random samples more accurately reflect the population
    c.   biased samples more accurately reflect the population
    d.   random samples are larger

6. A standard error is a measure of _____.

    a.   central tendency
    b.   variability
    c.   correlation
    d.   none of the above

7. The mean of a sampling distribution is called the _____, and the standard deviation of a sampling distribution is called the _____.

    a.   mean; standard deviation
    b.   mean; standard error
    c.   expected value; standard deviation
    d.   expected value; standard error

8. As N becomes larger, $\sigma_{\bar{X}}$ _____.

    a.   becomes smaller

b. becomes larger
c. gets closer in value to the mean
d. gets farther in value from the mean

9. A professor has a large number of sample means taken from the same population. If the professor counts the number of sample means, takes the square root, and divides that into the standard deviation of the population, he will have _____.

a. the expected value
b. the standard error of the mean
c. a sampling distribution
d. none of the above

10. A sampling distribution of the mean is most likely to be normal when _____.

a. N is large
b. the population is normal
c. both a. and b.
d. neither a. nor b.

11. Sampling distributions are useful because _____.

a. they give information about the probability of a particular statistical outcome
b. they eliminate uncertainty about populations when using samples
c. both a. and b.
d. neither a. nor b.

12. Suppose you had a rectangular distribution (like that of the playing cards, pictured in your textbook). Suppose you drew many, many random samples of 35 scores and found the mean for each sample. If you arranged these means into a frequency distribution, you would expect the distribution to be _____.

a. rectangular
b. bimodal
c. both a. and b.
d. neither a. nor b.

13. The Central Limit Theorem states that a sampling distribution of the mean approaches the normal curve if the _____.

a. sample size is large
b. standard deviation is large
c. standard error is below 30
d. any of the above is sufficient

14. In this chapter, z scores were used to determine the probability of _____; in the last chapter, z scores were used to determine the probability of _____.

a. sample means, raw scores
b. raw scores, sample means
c. neither a. nor b.; z scores do not indicate information about probability
d. there is no difference; this chapter just reviewed the same procedures of the last chapter

15. Under which of the following conditions is the t distribution a normal curve?

a. when $df = 1$

b. when the population from which the sample is drawn is normal
c. both a. and b.
d. neither a. nor b.

16. The *t* distribution was invented to handle _____.

a. confidence intervals with more than 99% confidence
b. very large samples
c. not knowing the value of σ
d. common arithmetic errors

17. Confidence intervals are _____.

a. experimental designs
b. descriptive statistics
c. inferential statistics
d. sampling distributions

18. For a 95% confidence interval about the mean, there is a 95% probability that the confidence interval will contain _____.

a. $\bar{X}$
b. μ
c. both a. and b.
d. neither a. nor b.

19. You have a 95% confidence interval about the mean with lower and upper limits of 14 to 17. The correct interpretation of this information is that _____.

a. 95% of the time $\bar{X}$ will be between 14 and 17
b. 95% of the μ's will be between 14 and 17
c. there is a 95% chance that the interval from 14 to 17 will contain the population mean (μ)
d. all of the above are correct

20. Given a 95% confidence interval about the mean of with lower and upper limits of 10 to 20, your interpretation should be _____.

a. there is a 95% chance that the interval from 10 to 20 will contain μ
b. $\bar{X}$ has a 95% chance of being between 10 and 20
c. 95% of the μ's will be between 10 and 20
d. all of the above are appropriate interpretations of this confidence interval

Short-Answer Questions

1. Describe the steps for obtaining a random sample using a random number table.

2. Distinguish between a sampling distribution and a frequency distribution.

3. State the Central Limit Theorem.

4. Suppose you know 20-29 year old men have an average weight of 166 pounds, a standard deviation of 22 pounds, and their weights are positively skewed not normally distributed. Can you use the techniques described in this chapter of the textbook to determine the probability of randomly selecting

49 20-29 year old men with a mean weight of 163 pounds or less from the population of 20-29 year old men? Write your conclusion and your reasoning.

5. What is the purpose of calculating a confidence interval about a mean?

Problems

1. Use a random number table to draw a random sample with $N = 6$ from the following scores.
21, 31, 17, 13, 02, 09, 57, 26, 72, 140, 27, 27, 27, 21, 13, 57, 43, 22, 18, 19, 18, 18, 18, 23, 25, 66, 75, 89, 99, 112

2. Calculate the standard error of the mean for a sample of 6 for the population in Problem #1. The ΣX = 1205 and $\Sigma X^2 = 82091$.

3. In short-answer question #4 you were asked if the techniques you learned in this chapter could be used to determine the probability of a sample ($N = 49$) being drawn from a population even though that population was positively skewed. Hopefully, you realized, or were later convinced, that they could.

Here is a reminder of the information from the problem. *Suppose you know 20-29 year old males have an average weight of 166 pounds, a standard deviation of 22 pounds, and their weights are positively skewed not normally distributed.* Now actually use the techniques described in this chapter to determine the probability of randomly selecting 49 20-29 year old males with a mean weight of 163 pounds or less from the population of 20-29 year old males.

4. In earlier chapters of this study guide, you had the opportunity to work with data from a sample of 38 children about the age (in months) at which they first stood without support. The mean you calculated was 13.00 months.
If the population mean for children standing on their own is 13.20 months with a standard deviation of 2, what is the probability of randomly selecting a sample of 38 children who stood on their own by the average age of 13.00 months or earlier?

5. Here's another problem about the age that children first stand on their own without support, remember that the population mean is 13.20 months with a standard deviation of 2. What is the probability of randomly selecting a sample of 25 children who stood on their own by the average age of 12.2 months or earlier? Write an interpretation about this probability.

6. A teacher was interested in the mathematical ability of graduating high school seniors in her state. She gave a 32-item test to a random sample of 75 seniors from her state with the following results: The $\Sigma X = 1275$ and $\Sigma X^2 = 23525$. Establish a 95% confidence interval about the mean and write a sentence that explains the interval you found to a high school principal.

7. The instructor of a popular course on health uses a number of videos in her course. After showing a new video, the instructor had a sample of 36 students fill out a questionnaire to indicate how valuable the video had been in helping them understand the topic. The instructor got the following statistics: ΣX = 1224 and $\Sigma X^2 = 41818$. High scores indicated the students judged the video as valuable. Over the previous five years, student evaluations of her other videos were negatively skewed with a mean of 33 and a standard deviation of 2.40. Calculate a 98% confidence interval about the mean and tell what the students thought of the new video compared to past videos.

8. In Chapter 4 of this study guide, you had the opportunity to determine the variability in employee satisfaction for a sample of 12 employees from a particular company. The following data were given: $\Sigma X = 30$, $\Sigma X^2 = 350$. Possible scores were between -10 and +10 with higher scores indicating greater satisfaction. Now, calculate a 98% confidence interval to estimate employee satisfaction for all employees in the company based on this sample of 12 and write an interpretation of your findings.

9. In past chapters of this study guide, you had the opportunity to learn about the Clock Test and do some calculations related to it. Here is the description of the test again if you need it. *The Clock Test is a technique for studying human vigilance. In the Clock Test, a hand moves regularly at one step per second, but sometimes, at random intervals, it jumps two steps. The participant's task is to notice the two-step jumps and press a button.*
The statistics that follow are the percentages of two-step jumps that were missed by 100 participants during the first 15 minutes and the last 15 minutes of a two-hour watch.

Statistics	Time of Test	
	First 15 min of watch	Last 15 min of watch
$\bar{X}$	12	25
$\hat{s}$	3.54	9.62

    a.   Calculate a 95% confidence interval about the mean for lapses during the first 15 minutes of a two-hour watch.
    b.   Calculate a 95% confidence interval about the mean for lapses during the last 15 minutes of a two-hour watch.
    c.   Draw a graph that illustrates the difference between the population means for the beginning and end of a watch.
    d.   Write a short interpretation of your findings.

10. In Chapter 6 of this study guide, you had the opportunity to work a correlation problem about the relationship of wise reasoning about social conflicts (self and other) and the pursuit of virtue. Here is a description of the study if you need it. *Huynh, Oakes, Shay, and McGregor (2017) studied the relationship between the pursuit of virtue and one's ability to reason wisely about social conflicts. The researchers assessed pursuit of virtue by asking participants their agreement with statements like, "I would like to do what I believe in." They assessed participants' endorsement of wise reasoning strategies by asking them how useful they found strategies like "making an effort to take the other person's perspective" when facing a social conflict. Participants answered the wise reasoning questions about either their own conflicts or other people's conflicts. Below are summary data from Study 1 of Huynh et al. (2017; retrieved from* ***https://osf.io/4chuy/*** *on July 6, 2018) that will allow close replications of the outcomes of this paper.*

Statistics	Wise Reasoning	
	Self	Other
$\bar{X}$	3.69	4.00
$\hat{s}$	0.70	0.62
N	130	126

Using these data complete the following.
    a.   Calculate a confidence interval about the mean for wise reasoning during one's own social conflicts.
    b.   Calculate a confidence interval about the mean for wise reasoning during other people's social

conflicts.
c. Draw a graph that illustrates the difference between wise reasoning during conflicts of one's own and during other people's conflicts.
d. Write a short interpretation of your findings.

ANSWERS

Multiple-Choice Questions

1. b. you will be somewhat uncertain about the population
   Explanation: Drawing random samples are the best way to ensure a good estimate of the population, but as long as you are using a sample and not taking data from the entire population, some degree of uncertainty will remain. This means your conclusions are likely to be correct, but could be incorrect.
2. c. has certain groups from the population overrepresented or underrepresented
3. d. none of the above
   Explanation: The only way to eliminate uncertainty about a population is to collect data from the entire population. Large, random samples increase the chances you will have a representative sample that leads to correct conclusions, but does not eliminate that uncertainty.
4. d. more likely to be representative of the population mean
5. b. random samples more accurately reflect the population
6. b. variability
   Explanation: Standard error is the standard deviation of a sampling distribution. Remember from Chapter 4, standard deviation is a measure of variability.
7. d. expected value; standard error
8. a. becomes smaller
   Explanation: Calculate standard error as $\sigma_{\bar{X}} = \sigma / \sqrt{N}$. Notice the denominator contains N. If $\sigma = 10$. Using 100 in the denominator leads to a smaller outcome than using 10 in the denominator.
9. b. the standard error of the mean
   Explanation: If the professor calculated the mean of all of those means, he would have the expected value. If he had arranged all of those means into a frequency distribution, he would have made a sampling distribution. However, if he took the square root of the number of means he had and divided it into the standard deviation of the population, that is the standard error of the mean.
10. c. both a. and b.
    Explanation: The Central Limit Theorem says that when N is large, the sampling distribution of the mean will be normal regardless of the shape of the original distribution. The sampling distribution of the mean is also likely to be normal if the distribution from which it comes is normal even when N is not large.
11. a. they give information about the probability of a particular statistical outcome
    Explanation: Sampling distributions measure the degree of uncertainty. They do not eliminate it. The only way to eliminate uncertainty about populations is to measure the entire population directly.
12. d. neither a. nor b.
    Explanation: The Central Limit Theorem states that a sampling distribution of the mean approaches the normal curve if the sample size is large. Because a sample of 30 is usually large enough to result in a relatively normal sampling distribution of the mean, you should expect a sample of 35 to produce a normally shaped sampling distribution rather than rectangular or bimodal.
13. a. sample size is large

14. a. sample means, raw scores
15. d. neither a. nor b.

    Explanation: The *t* distribution is a normal curve when $df = \infty$.
16. c. not knowing the value of σ
17. c. inferential statistics
18. b. μ

    Explanation: Confidence intervals indicate with a specified degree of confidence (here 95%), an interval of scores (defined by a lower and upper limit) containing the value of an unknown parameter (here the mean, μ). The sample mean ($\bar{X}$) will always be in the center of the interval (100% of the time) because to build a confidence interval about the mean, you add and subtract from the sample mean.
19. c. there is a 95% chance that the interval from 14 to 17 will contain the population mean (μ)
20. a. there is a 95% chance that the interval from 10 to 20 will contain μ

Short-Answer Questions

1. 1. Give each score in the population a unique identification number. Often, it is a good idea to use a
        two-digit number for this (i.e., 01, 02, 03, etc.) but it depends on how big your population is. If you have more than 100 scores in your population, you would need three-digit ID numbers so each score in the population has a unique ID number.
   2. Go to a random number table. You have one available to you in Table B of Appendix C.
   3. Haphazardly place your finger down on the random number table.
   4. If the number you are touching matches the ID number you have assigned to a score (not the score itself) mark that ID number and include its corresponding score (not the ID number) as part of your sample. If the number you are touching is too big to match an ID number, skip it.
   5. In any direction, move from that place in the table to the next number (be sure the direction you move
        is predetermined). If the number you reach matches the ID number you have assigned to a score, mark that score as being included in your sample. If the number you reach is too big or repeats an ID number you have already used, skip it.
   6. Repeat Step 5 until you have selected the number of scores you want in your random sample.

2. A frequency distribution is an *empirical* distribution of *scores* from a sample or population. It tells how often each *score value* appears in the distribution. A sampling distribution is a *theoretical* distribution of *sample statistics* taken from a population. It tells how often a particular *sample statistical outcome* is likely to occur.
3. For any population of scores, regardless of form, the sampling distribution of the mean approaches a normal distribution as *N* gets larger, the expected value is equal to μ, and the standard error is equal to $\frac{\sigma}{\sqrt{N}}$.

    In statistics there are a few things that probably deserve memorizing. The Central Limit Theorem is one of them. If you couldn't state the Central Limit Theorem to answer this question, take a few minutes now to commit the theorem to memory.
4. Yes, because the sample size ($N = 49$) is greater than 30, it will not matter that the population is positively skewed rather than normal. The techniques of this chapter will work to determine

the probability of a sample of 49 men with a mean of 163 being selected from a population with a mean of 166. Because the standard deviation of the population is known, a z score can be calculated and the associated probability obtained from the z table.

5. Confidence intervals provide a range of values μ is expected to fall within with a specified degree of certainty. This is important because in inferential statistics we can estimate μ using $\bar{X}$. However, because our estimate is based on a sample, we know that we are left with an unknown amount of uncertainty about how accurate our estimate is. Confidence intervals with their specified certainty give more information for better conclusions.

Problems

1. There are many possible correct random samples from this population. The key is that you follow the proper procedure.

   1. Give each score in the population a unique identification number. Often, it is a good idea to use a two-digit number for this (i.e., 01, 02, 03, etc.) but it depends on how big your population is. If you have more than 100 scores in your population, you would need three-digit ID numbers. Here two-digit ID numbers will do.

   2. Go to a random number table. You have one available to you in Table B of Appendix C.

   3. Haphazardly place your finger down on the random number table.

   4. If the number you are touching matches the ID number you have assigned to a score (not the score itself) mark that ID number and include its corresponding score (not the ID number) as part of your sample. If the number you are touching is too big, skip it.

   5. In any direction, move from that place in the table to the next number (be sure the direction you move is predetermined). If the number you reach matches the ID number you have assigned to a score, mark that score as being included in your sample. If the number you reach is too big or repeats an ID number you have already used, skip it. Note: you may have score values that are identical in your sample (e.g., you might have selected two of the three 27's from the population above). This is fine. It is not fine to use a single score more than once in your sample (e.g., you cannot include 140 in your sample more than once because there is only one score that is 140). This is why you skip an ID number that replicates an ID number you have already used and why you use ID numbers in the first place rather than look for actual scores in your random number table.

   6. Repeat Step 5 until you have selected six scores for your random sample.

2. $\sigma_{\bar{X}} = 13.68$

   The standard error of the mean is the standard deviation of the sampling distribution. To find this value, you need to first calculate the standard deviation of the population. A common mistake here is to use $N = 6$. Remember when you are calculating the standard deviation of the population you are using the size of the actual population, $N = 30$, not the size of the sample you are going to select.

$$\sigma = \sqrt{\frac{\sum X^2 - \frac{(\sum X)^2}{N}}{N}} = \sqrt{\frac{82091 - \frac{(1205)^2}{30}}{30}} = \sqrt{\frac{82091 - \frac{1452025}{30}}{30}}$$

$$\sigma = \sqrt{\frac{82091 - 48400.833}{30}} = \sqrt{\frac{33690.167}{30}} = \sqrt{1123.01} = 33.51$$

With σ calculated, you are ready to calculate the standard error of the mean.

$$\sigma_{\bar{X}} = \sigma/\sqrt{N}$$

$$\sigma_{\bar{X}} = 33.51/\sqrt{6}$$

$$\sigma_{\bar{X}} = 33.51/2.449$$

$$\sigma_{\bar{X}} = 13.68$$

A common mistake here is to use $N = 30$. Remember when you are calculating the standard error of the sampling distribution you are using the size of the sample you selected.

There is a lot of variability in this population (standard deviation; $\sigma = 33.51$). With a small sample ($N = 6$), there is also a lot of variability in the possible means that can be obtained from that population (standard error; $\sigma_{\bar{X}} = 13.68$).

3. The probability of randomly selecting 49 20-29 year old males with a mean weight of 163 pounds or less from the population of 20-29 year old males is .1685.

You need to convert your sample mean 163 pounds to a z score using this formula.

$$z = \frac{\bar{X} - \mu}{\sigma_{\bar{X}}}$$

First you will need to find $\sigma_{\bar{X}}$.

Students who get the wrong answer here have often skipped this step and used σ rather than $\sigma_{\bar{X}}$ in the denominator for z.

$$\sigma_{\bar{X}} = \sigma/\sqrt{N}$$

$$\sigma_{\bar{X}} = 22/\sqrt{49}$$

$$\sigma_{\bar{X}} = 22/7$$

$$\sigma_{\bar{X}} = 3.14$$
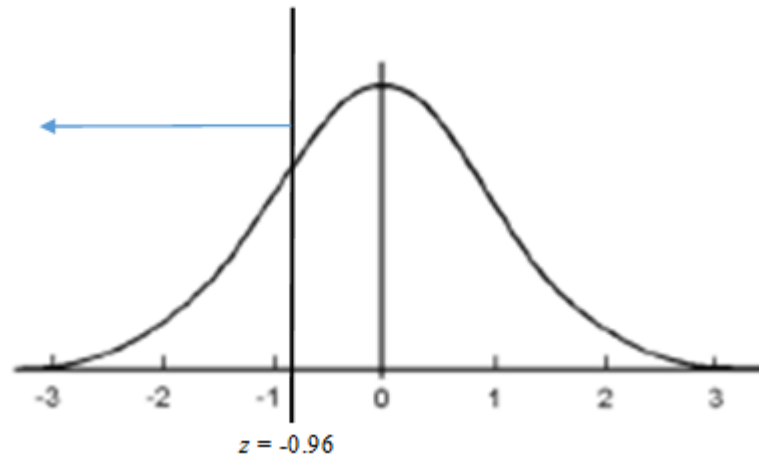
Now you are ready to convert to a z score.

$$z = \frac{\bar{X} - \mu}{\sigma_{\bar{X}}}$$

$$z = \frac{163 - 166}{3.14}$$

$$z = \frac{-3}{3.14}$$

$z = -0.96$

Now take this z score to the z table and determine the probability of randomly selecting a sample of 49 men with a mean of 163 or less from this population.

$z = -0.96$

Find $z = -0.96$ in the z column (There are no negative z's in this table. $+0.96$ will do because you remember that both sides of the normal distribution are the same). The probability in the corresponding c column is .1685.

Some students at this point ask why they would ever need to find such a probability in real life.

There are many reasons, but one common reason in inferential statistics is to help decide the likelihood that a particular sample was really selected from the given population or if instead the sample came from a different population. For instance, say that sample of 49 males were vegetarians, and you wanted to know if vegetarians' average weight is the same as the general population (166 pounds). If the probability of getting a mean of 163 from a population with a mean of 166 is high (i.e., this outcome is very likely), then you might think something like this: "Hey, it is reasonable to believe the population of vegetarian males weigh 166 (the same as other 20-29 year old males) because the probability that I would get a sample mean of 163 from that population is high." However, if the probability of that sample mean of 163 is very low (i.e., this outcome would rarely occur), then you might think something like this: "Either something very rare just happened or the population mean of vegetarian males is something different than 166 because the probability I would get a sample mean of 163 from this population is so low." You will learn a lot more about this in upcoming chapters of the text.

4. The probability of randomly selecting a sample of 38 children who stood on their own by the average age of 13.00 months or earlier is .2643.

You need to convert your sample mean 13.00 months to a z score using this formula:

$$z = \frac{\bar{X} - \mu}{\sigma_{\bar{X}}}$$

First you will need to find $\sigma_{\bar{X}}$.

Students who get the wrong answer here have often skipped this step and used σ rather than $\sigma_{\bar{X}}$ in the denominator for z.

$$\sigma_{\bar{X}} = \sigma / \sqrt{N}$$

$$\sigma_{\bar{X}} = 2 / \sqrt{38}$$

$$\sigma_{\bar{X}} = 2 / 6.16$$

$$\sigma_{\bar{X}} = 0.32$$

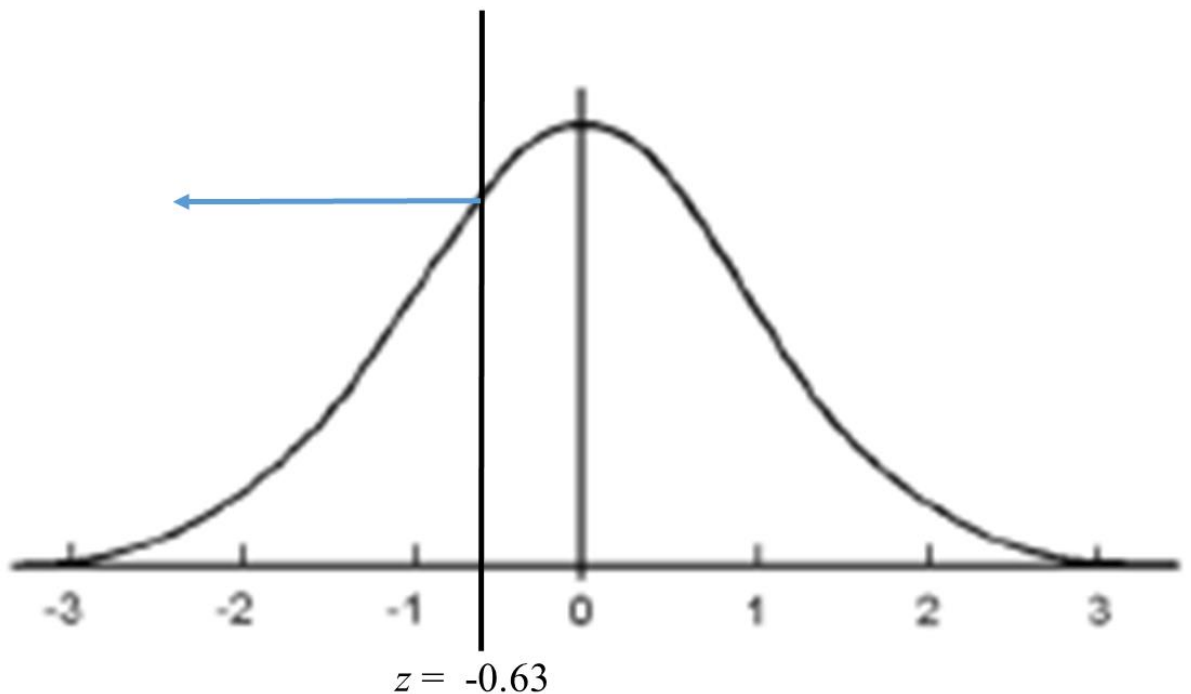Now you are ready to convert to a z score.

$$z = \frac{\bar{X} - \mu}{\sigma_{\bar{X}}}$$

$$z = \frac{13.00 - 13.20}{0.32}$$

$$z = \frac{-0.20}{0.32}$$

$$z = -0.63$$

Now take this z score to the z table and determine the probability of randomly selecting 38 children who stood on their own by the average age of 13.00 months or earlier from this population.

$$z = -0.63$$

Find $z = -0.63$ in the z column (0.63 will do because you remember that both sides of the normal distribution are the same). The probability in the corresponding c column is .2643.

5. The probability of randomly selecting a sample of 25 children who stood on their own by the average age of 12.2 months or earlier is .0062. It would be very unlikely to randomly select 25 children who stood on their own by the average age of 12.2 months or earlier from a population with a mean of 13.2 months and a standard deviation of 2.

You need to convert your sample mean of 12 months to a z score using this formula:

$$z = \frac{\bar{X} - \mu}{\sigma_{\bar{X}}}$$

First you will need to find $\sigma_{\bar{X}}$.
Students who get the wrong answer here have often skipped this step and used σ rather than $\sigma_{\bar{X}}$ in the
denominator for z.

$$\sigma_{\bar{X}} = \sigma / \sqrt{N}$$

$$\sigma_{\bar{X}} = 2 / \sqrt{25}$$

$$\sigma_{\bar{X}} = 2 / 5$$

$$\sigma_{\bar{X}} = 0.40$$

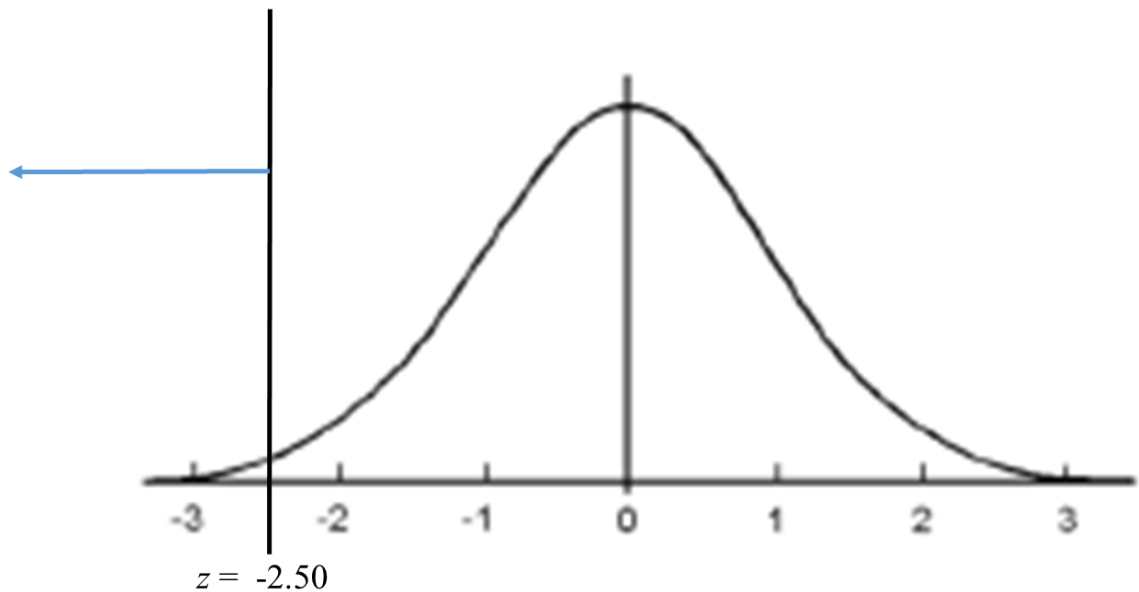Now you are ready to convert to a z score.

$$z = \frac{\bar{X} - \mu}{\sigma_{\bar{X}}}$$

$$z = \frac{12.2 - 13.2}{0.40}$$

$$z = \frac{-1}{0.40}$$

$$z = -2.50$$

Now take this z score to the z table and determine the probability of randomly selecting 25 children who stood on their own by the average age of 12.2 months or earlier from this population.

z = -2.50

Find z = -2.50 in the z column (2.50 will do because you remember that both sides of the normal distribution are the same). The probability in the corresponding c column is .0062.

Although it could happen that one would select 25 children who stood on their own by the average age of 12.2 months or earlier from the general population of children. It is very unlikely (probability of .0062). It is more likely this group of children did not come from the regular population. Instead, perhaps they came from a population of children who had particularly good nutrition, their parents coached them to stand rather than letting nature take its course, they were protected from environmental pollution, or some other characteristic that allowed them to reach this milestone early.

6. The teacher can be 95% confident that the interval of scores between 15.84 and 18.16 correct items contains the mean mathematical ability on the test for her state. Thus, the state mean on this exam is likely to be somewhere between 50% and 57% correct for the 32 items.

To calculate a confidence interval, you need the sample mean ($\bar{X}$), an estimate of the standard error ($s_{\bar{X}}$), and a t value from the table in your book.

$$\bar{X} = \frac{\Sigma X}{N} = \frac{1275}{75} = 17.00$$

To calculate $s_{\bar{X}}$, you first need to calculate $\hat{s}$. Don't forget to use N-1 in the denominator because here you are having to estimate the population standard deviation using the sample.

$$\hat{s} = \sqrt{\frac{\Sigma X^2 - \frac{(\Sigma X)^2}{N}}{N-1}} = \sqrt{\frac{23525 - \frac{(1275)^2}{75}}{75-1}} = \sqrt{\frac{23525 - \frac{1625625}{75}}{74}} = \sqrt{\frac{23525 - 21675}{74}} = \sqrt{\frac{1850}{74}} = \sqrt{25} = 5.00$$

Now you are ready to calculate the standard error.

$$s_{\bar{X}} = \hat{s}/\sqrt{N}$$

$$s_{\bar{X}} = 5.00/\sqrt{75}$$

$$s_{\bar{X}} = 5.00/8.66$$

$$s_{\bar{X}} = 0.58$$

Look up the t_α you will use in Table D of Appendix C.
You will need a *df* to do this. Here $df = N - 1$.

$$df = N - 1$$

$$df = 75 - 1$$

$$df = 74$$

Go to Table D in Appendix C of your textbook. Find the column labeled 95% confidence interval. There is no row labeled $df = 74$, so choose the smaller of the two *df* your value falls between, in this case the row labeled $df = 60$. Now find the *t* value at the intersection for that row and column. Here, $t_\alpha = 2.000$.

Now you are ready to calculate the lower limit of the confidence interval.

$$LL = \bar{X} - t_\alpha(s_{\bar{X}})$$

$$LL = 17.00 - 2.000(0.58)$$

$$LL = 17.00 - 1.16$$

$$LL = 15.84$$

Now you are ready to calculate the upper limit of the confidence interval.

$$UL = \bar{X} + t_\alpha(s_{\bar{X}})$$

$$UL = 17.00 + 2.000(0.58)$$

$$UL = 17.00 + 1.16$$

$$UL = 18.16$$

The teacher can be 95% confident that the interval of scores between 15.84 and 18.16 contains the mean mathematical ability on the test for her state. Thus, the state mean on this exam is likely to be somewhere between 50% (15.84/32 items) and 57% (18.16/32 items) correct for the 32 items.

7. The instructor can be 98% confident the interval between 33.02 and 34.98 contains the mean of the population of student judgments for the new video. Because the entire interval is above the previous five-year mean of 33, the instructor can be confident the population of students will judge the new video as more valuable than the average video she has used in her course.

To calculate a confidence interval you need the sample mean ($\bar{X}$), an estimate of the standard error ($s_{\bar{X}}$), and a t value from the table in your book.

$$\bar{X} = \frac{\Sigma X}{N} = \frac{1224}{36} = 34.00$$

To calculate $s_{\bar{X}}$, you first need to calculate $\hat{s}$.

$$\hat{s} = \sqrt{\frac{\Sigma X^2 - \frac{(\Sigma X)^2}{N}}{N-1}} = \sqrt{\frac{41818 - \frac{(1224)^2}{36}}{36-1}} = 2.40$$

Now you are ready to calculate the standard error.

$$s_{\bar{X}} = \hat{s} / \sqrt{N}$$

$$s_{\bar{X}} = 2.40 / \sqrt{36}$$

$$s_{\bar{X}} = 0.40$$

Look up the t_α you will use in Table D of Appendix C.
You will need a df to do this. Here $df = N - 1$.

$$df = N - 1$$

$$df = 36 - 1$$

$$df = 35$$

Go to Table D in Appendix C of your textbook. Find the column labeled *98%* confidence interval (a common mistake is to look up the value corresponding to the 95% confidence interval). There is no row labeled $df = 35$, so choose the smaller of the two df your value falls between, in this case the row labeled $df = 30$. Now find the t value at the intersection for that row and column. Here, $t_\alpha = 2.457$.

Now you are ready to calculate the lower limit of the confidence interval.

$$LL = \bar{X} - t_\alpha(s_{\bar{X}})$$

$$LL = 34.00 - 2.457(0.40)$$

$$LL = 33.02$$

Now you are ready to calculate the upper limit of the confidence interval.

$$UL = \bar{X} + t_\alpha(s_{\bar{X}})$$

$$UL = 34.00 + 2.457(0.40)$$

$$UL = 34.98$$

8. Possible scores were between -10 and +10 with higher scores indicating greater satisfaction. Now, calculate a 98% confidence interval to estimate employee satisfaction for all employees in the company based on this sample of 12 and write an interpretation of your findings.

There is a 98% chance that the interval of scores between -1.41 and 6.41 contains the mean employee satisfaction for all employees in the company. This is a wide range of possible values on a small scale. The employees could be as much as moderately satisfied (6.41on a scale that goes only as high as 10). But, they could also be as low on satisfaction as below the neutral 0 point (-1.41). To get a better idea of employee satisfaction, it might be a good idea to repeat the data collection with a larger sample.

To calculate a confidence interval you need the sample mean $(\bar{X})$, an estimate of the standard error $(s_{\bar{X}})$, and a t value from the table in your book.

$$\bar{X} = \frac{\Sigma X}{N} = \frac{30}{12} = 2.5$$

To calculate $s_{\bar{X}}$, you first need to calculate $\hat{s}$.

$$\hat{s} = \sqrt{\frac{\Sigma X^2 - \frac{(\Sigma X)^2}{N}}{N-1}} = \sqrt{\frac{350 - \frac{(30)^2}{12}}{12-1}} = 5.00$$

Now you are ready to calculate the standard error.

$$s_{\bar{X}} = \hat{s}/\sqrt{N}$$

$$s_{\bar{X}} = 5.00/\sqrt{12}$$

$$s_{\bar{X}} = 1.44$$

Look up the t_α you will use in Table D of Appendix C. You will need a df to do this. Here $df = N - 1$.

$df = N - 1$

$df = 12 - 1$

$df = 11$

Go to Table D in Appendix C of your textbook. Find the column labeled *98%* confidence interval. Find the row labeled $df = 11$. Now find the t value at the intersection for that row and column. Here, $t_\alpha = 2.718$.

Now you are ready to calculate the lower limit of the confidence interval.

$$LL = \bar{X} - t_\alpha(s_{\bar{X}})$$
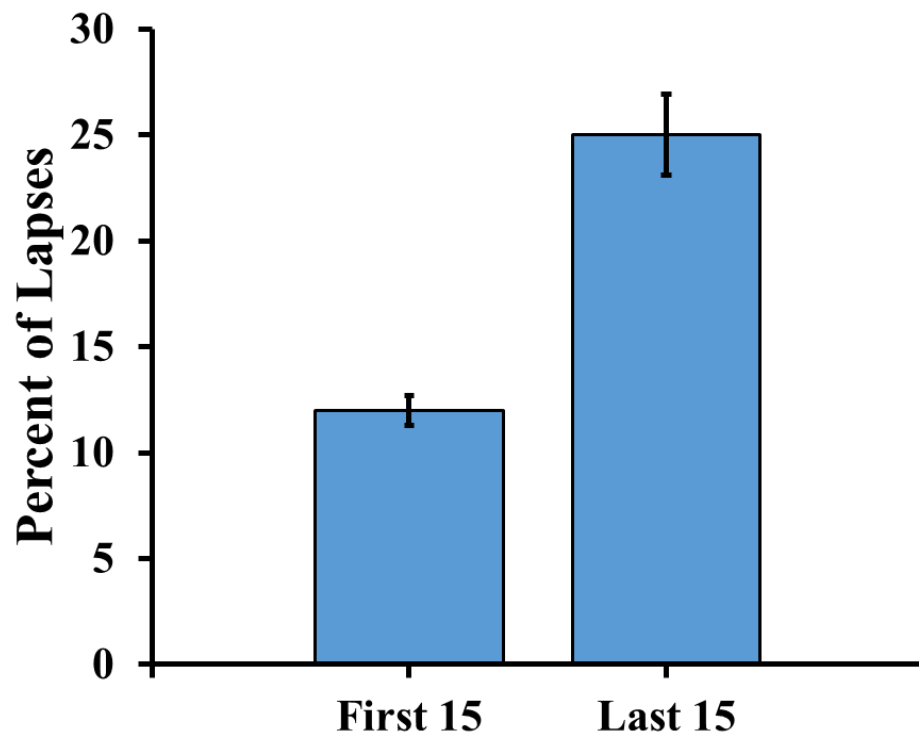
$$LL = 2.5 - 2.718(1.44)$$

$$LL = -1.41$$

Now you are ready to calculate the upper limit of the confidence interval.

$$UL = \bar{X} + t_\alpha(s_{\bar{X}})$$

$$UL = 2.5 + 2.718(1.44)$$

$$UL = 6.41$$

9. a. The 95% confidence interval about the mean for the *first* 15 minutes of a two-hour watch is 11.30 to 12.70 percent lapses.

b. The 95% confidence interval about the mean for the *last* 15 minutes of a two-hour watch is 23.09 to 26.91 percent lapses.

c.

d. There are many acceptable answers for the interpretation. Here are two examples.

These data show a very big difference between the lapses in performance at the beginning and end of a watch. The participants had many fewer lapses during the *first* 15 minutes of a two-hour watch ($\bar{X} = 12$, 95% CI [11.30, 12.70]) than during the *last* 15 minutes of a two-hour watch ($\bar{X} = 25$, 95% CI [23.09, 26.91]).

OR

The researcher can be 95% confident the interval of 11.30 to 12.70 percent lapses contains the true mean percent of lapses that would occur during the *first* 15 minutes of a two-hour watch. This is well below the 95% confidence interval, from 23.09 to 26.91, for the *last* 15 minutes of a

two-hour watch.

We'll start with the *first* 15 minutes of a two-hour watch.

Start with standard error.

$$s_{\bar{X}} = \hat{s}/\sqrt{N}$$

$$s_{\bar{X}} = 3.54/\sqrt{100}$$

$$s_{\bar{X}} = 0.35$$

Next *df*

$$df = N - 1$$

$$df = 100 - 1$$

$$df = 99$$

Go to Table D in Appendix C of your textbook. Find the column labeled 95% confidence interval. Find the row labeled with $df = 99$ (or just below that, here 80). Now find the *t* value at the intersection for that row and column. Here, $t_\alpha = 1.990$.

Now you are ready to calculate the lower limit of the confidence interval for the *first* 15 minutes of a two-hour watch.

$$LL = \bar{X} - t_\alpha(s_{\bar{X}})$$

$$LL = 12 - 1.990(0.35)$$

$$LL = 11.30$$

Now you are ready to calculate the upper limit of the confidence interval for the *first* 15 minutes of a two-hour watch.

$$UL = \bar{X} + t_\alpha(s_{\bar{X}})$$

$$UL = 12 + 1.990(0.35)$$

$$UL = 12.70$$

The 95% confidence interval about the mean for the *first* 15 minutes of a two-hour watch is 11.30 to 12.70 percent lapses.

Now do the 95% confidence interval for the *last* 15 minutes of a two-hour watch.

Start with standard error.

$$s_{\bar{X}} = \hat{s}/\sqrt{N}$$

$$s_{\bar{X}} = 9.62/\sqrt{100}$$

$$s_{\bar{X}} = 0.96$$

The t value is the same for this as for the first 15 minutes because *N* is still 100, $t_\alpha = 1.990$.

Now you are ready to calculate the lower limit of the confidence interval for the *last* 15 minutes of a two-hour watch.

$LL = \bar{X} - t_\alpha(s_{\bar{X}})$

$LL = 25 - 1.990(0.96)$

$LL = 23.09$

Now you are ready to calculate the upper limit of the confidence interval for the *last* 15 minutes of a two-hour watch.
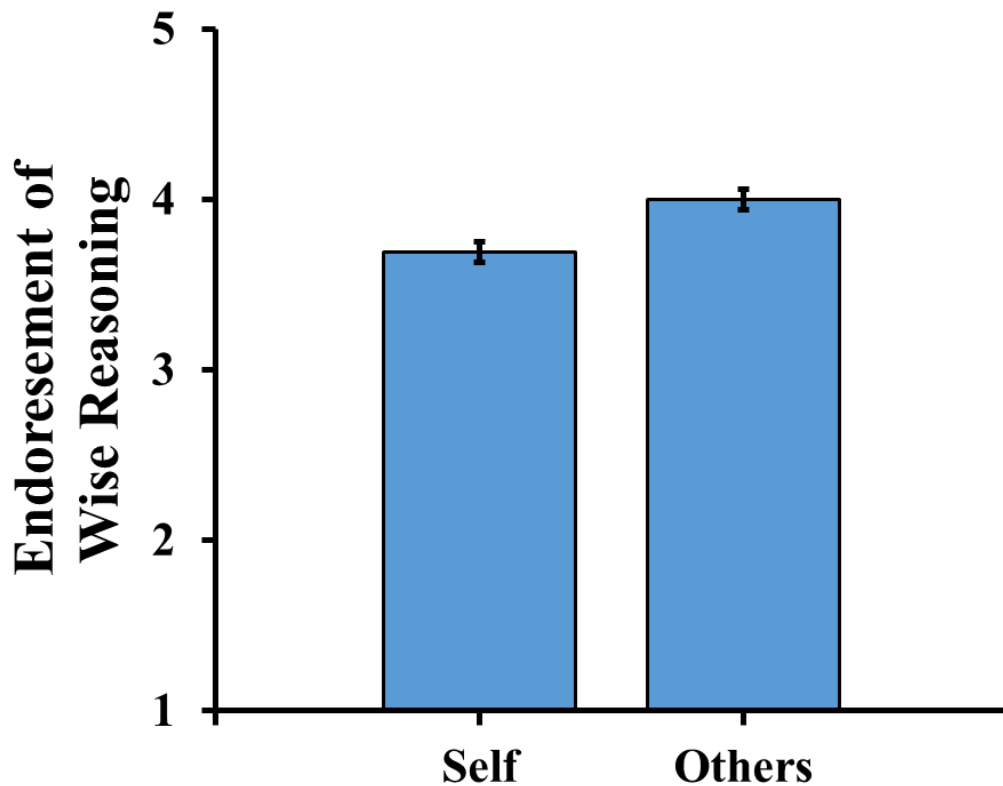
$UL = \bar{X} + t_\alpha(s_{\bar{X}})$

$UL = 25 + 1.990(0.96)$

$UL = 26.91$

The 95% confidence interval about the mean for the *last* 15 minutes of a two-hour watch is 23.09 to 26.91 percent lapses.

10. a. The 95% confidence interval about the mean for wise reasoning for self is 3.57 to 3.81 points.
    b. The 95% confidence interval about the mean for wise reasoning for others is 3.88 to 4.12 points.
    c.

d. There are many acceptable answers for the interpretation. Here are two examples.

It is likely that there is a real difference between the mean level of endorsement between wise reasoning for one's own social conflicts and for social conflicts of others. Participants were less likely to endorse wise reasoning strategies for their own social conflicts ($\bar{X} = 3.69$, 95% CI [3.57, 3.81]) than to endorse wise reasoning strategies for social conflicts of others ($\bar{X} = 4.00$, 95% CI [3.88, 4.12]).

OR

We can be 95% confident that the interval between 3.57 and 3.81 points contains the mean endorsement of wise reasoning strategies for one's own social conflicts and the interval between 3.88 and 4.12 points contains the mean endorsement for wise reasoning strategies for social conflicts of others. Meaning, people are less likely to endorse wise reasoning strategies during social conflict for themselves than for others.

We'll start with wise reasoning for *self*.

Start with standard error.

$$s_{\bar{X}} = \hat{s}/\sqrt{N}$$

$$s_{\bar{X}} = .70/\sqrt{130}$$

$$s_{\bar{X}} = .06$$

Next *df*

$$df = N - 1$$

$$df = 130 - 1$$

$$df = 129$$

Go to Table D in Appendix C of your textbook. Find $t_\alpha = 1.980$.

Now you are ready to calculate the lower limit of the confidence interval for wise reasoning for *self*.

$$LL = \bar{X} - t_\alpha(s_{\bar{X}})$$

$$LL = 3.69 - 1.980(0.06)$$

$$LL = 3.57$$

Now you are ready to calculate the upper limit of the confidence interval for wise reasoning for *self*.

$$UL = \bar{X} + t_\alpha(s_{\bar{X}})$$

$$UL = 3.69 + 1.980(0.06)$$

$$UL = 3.81$$

The 95% confidence interval about the mean for wise reasoning for *self* is 3.57 to 3.81 points.

Now do the 95% confidence interval for wise reasoning for *others*.

Start with standard error.

$$s_{\bar{X}} = \hat{s}/\sqrt{N}$$

$$s_{\bar{X}} = 0.62/\sqrt{126}$$

$$s_{\bar{X}} = 0.06$$

Go to Table D in Appendix C of your textbook. Find $t_\alpha = 1.984$.

Now you are ready to calculate the lower limit of the confidence interval for wise reasoning for *others*.

$$LL = \bar{X} - t_\alpha(s_{\bar{X}})$$

$$LL = 4.00 - 1.984(0.06)$$

$$LL = 3.88$$

Now you are ready to calculate the upper limit of the confidence interval for wise reasoning for *others*.

$$UL = \bar{X} + t_\alpha(s_{\bar{X}})$$

$$UL = 4.00 + 1.984(0.06)$$

$$UL = 4.12$$

The 95% confidence interval about the mean for wise reasoning for *others* is 3.88 to 4.12 points.

References

Huynh, A. C., Oakes, H., Shay, G. R., & McGregor, I. (2017). The wisdom in virtue: Pursuit of virtue predicts wise reasoning about personal conflicts. *Psychological Science, 28*(12), 1848-1856. doi: 10.1177/095679761772621