CHAPTER 5

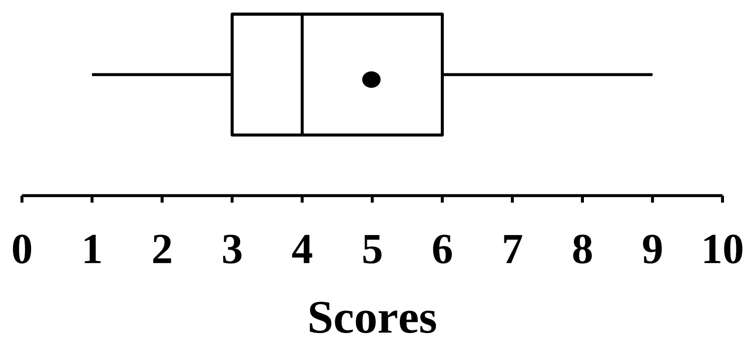
Other Descriptive Statistics

Summary

This chapter covers *z* scores, boxplots, the effect size index *d*, outliers, and Descriptive Statistics Reports. The first three are descriptive statistics that combine two or more simpler descriptive statistics.

A ***z score*** (one kind of ***standard score***) tells how far away a score is from the mean in units of standard deviation. It is the difference between a raw score and the mean of the distribution divided by the standard deviation. As a descriptive statistic, *z* scores describe the relative standing of an individual within a distribution of scores (e.g., How tall am I compared to other people my age?), or compare the relative position of scores in different distributions (e.g., How fast was I in this year's race compared to how fast I was in last year's race?). Typically, descriptive statistic *z* scores range from -3 to +3 whereas inferential statistic *z* scores can occur outside this range.

An ***outlier*** is an extreme score (very small or very large) when compared to other scores in the distribution. It is important to identify them because they influence some statistical outcomes. There are many different ways to identify outliers. The textbook describes finding scores that are 1.5 times the IQR below the 25th percentile or 1.5 times the IQR above the 75th percentile.



There are two parts to a ***boxplot*** (see above). The first, an axis (either horizontal or vertical), shows score values in a distribution. The second, a box with whiskers, has elements corresponding to various descriptive statistics aligned with the corresponding axis values. Within the box, the mean is indicated by a dot and the median by a vertical line. The box width from left to right (or box height from bottom to top, if using a vertical axis) spans the interquartile range. The ends of the whiskers align with the lowest and highest scores, although sometimes outliers are marked with a dot or asterisk and the whisker extends out to the next highest or lowest score. Skew of a distribution can be determined by the relation of the mean to the median (e.g., the mean is larger than the median in a positively skewed distribution) and also by comparing the relative length of the whiskers (longer whiskers are usually on the side of the distribution with the fewest scores). Making a box and whisker on the same axis for more than one distribution makes for convenient comparisons.

Effect size indices are descriptive statistics that describe the size of differences and relationships. They help researchers understand the amount of influence the independent variable has on the dependent variable. The effect size index introduced in this chapter, ***d***, measures the size of the difference between two distributions in units of standard deviation. To find *d*, divide the difference between the means of two distributions by a standard deviation. By convention, interpret *d* values of .20, .50, and .80 as "small", "medium", and "large", respectively; and use adjectives to modify the descriptors for intermediate and extreme values (e.g., "very small" or "moderately large").

A ***Descriptive Statistics Report*** gives the reader an overview of study results. Elements usually include a boxplot and narrative covering central tendency, variability, shape, overlap of distributions, and an effect size index.

Multiple-Choice Questions

1. A *z* score conveys the _______.

   a. central tendency of a distribution
   b. variability of a distribution
   c. relative position of an individual
   d. all of the above

2. A *z* score close to zero indicates a _______.

   a. raw score near the bottom of the distribution
   b. raw score close to the mean
   c. distribution with little variability
   d. distribution that is quite variable

3. A *z* score of 2.5 indicates a raw score that is _______.

   a. close to the mean
   b. much smaller than the mean
   c. much larger than the mean
   d. 2.5 times larger than the mean

4. The advantage of using *z* scores over using raw scores is that *z* scores allow you to _______.

   a. compare a score in one distribution to a score in another
   b. understand the relationship of a score to the mean
   c. both a. and b.
   d. neither a. nor b.

5. For four years in a row at the annual Late 80's Video Game Contest, a different entrant won the contest. The field of contestants was different each year and the game played in the final round was different. The annual winners, amount of time the winner played before losing during the final round, and contest statistics for each year are shown. Screech played the longest in all four years, but because they played different games, who should be considered the best player across all four years?

   a. Year 1: Zach, 32 hours of Donkey Kong; $\bar{X}$ = 28 hours; S = 2 hours
   b. Year 2: Slater, 34 hours of Pac Man; $\bar{X}$ = 29 hours; S = 2 hours
   c. Year 3: Screech, 36 hours of Space Invaders; $\bar{X}$ = 30 hours; S = 6 hours
   d. Year 4: Lisa, 28 hours of Frogger; $\bar{X}$ = 24 hours; S = 3 hours

6. George is 7'8" tall. It is likely that among a group of typical college students, George would be considered a(n) _______ for height.

   a. median
   b. large effect size
   c. outlier
   d. none of the above

7. Boxplots convey information about the _______ of a distribution.

   a. central tendency
   b. variability
   c. shape
   d. all of the above

8. A boxplot of a negatively skewed distribution usually has a whisker over the lower scores that is _______than the whisker over the higher scores.

a. longer
b. shorter
c. cannot be determined from the information given
d. whisker length in a boxplot is not impacted by form

9. In a boxplot of a positively skewed distribution, the mean is _______.

a. greater than the median
b. less than the median
c. outside of the interquartile box
d. in the middle of the interquartile box

10. To determine the range of a distribution using a boxplot, look at the difference between the _______.

a. dot and the line in the box
b. ends of the box
c. ends of the whiskers
d. any of the above will work to tell the range of a distribution

11. Outliers can be presented in a boxplot as points _______.

a. within the box
b. beyond the whiskers
c. both a. and b.
d. neither a. nor b.

12. The measure of central tendency **not** presented in a boxplot is the _______.

a. mean
b. median
c. mode
d. all of the above

13. The percent of scores covered by the box portion of a boxplot is _______.

a. 100 percent
b. 95 percent
c. 50 percent
d. 5 percent

14. Rats in the control group engaged in an average of a whopping 110 aggressive acts during the 10 minute testing period, but it was exceeded by rats being given a new drug with a mean of 120. The effect size index in this experiment, *d*, was 0.20. The difference between the two groups should be described as _______.

a. very large
b. large
c. medium
d. small

15. From an effect size estimate of 0.80, you can determine that the means of the two groups are _______.

a. identical
b. different by 80 percent
c. related to each other
d. very different

16. An effect size index is a measure of _______.

a. the relative position of an individual
b. the effectiveness of a Descriptive Statistics Report
c. the accuracy of the calculations
d. none of the above

17. *d* is a measure of the size of the _______ of two distributions.

a. means
b. difference in the means
c. standard deviations
d. ranges

18. A Descriptive Statistics Report includes a _______ to graphically illustrate the statistical results.

a. boxplot with a box for each distribution
b. line graph with one distribution represented on each axis
c. frequency polygon with one distribution on each axis
d. none of the above, there are no graphs in a Descriptive Statistics Report

19. A Descriptive Statistics Report has information about _______.

a. central tendency
b. effect size
c. shape of the distributions
d. all of the above

20. One use of Descriptive Statistics Reports is to _______.

1. avoid arithmetic errors in reporting descriptive statistics
2. present tables that show descriptive statistics
3. explain the results of an experiment
4. present results without using graphs

Short-Answer Questions

1. The statistics professor hands back the first test. The test was worth 100 points. Instead of a score out of 100, she has written a *z* score. Frank got a 0. What does that mean?

2. Write from memory the *d* values that are considered small, medium, and large.

3. What descriptive statistics are presented in a boxplot?

4. When examining a boxplot, how could you determine the score that marks the upper 25% of the distribution from the rest of the distribution?

5. Why is it important to identify outliers in a distribution?

6. Describe the steps to write a Descriptive Statistics Report.

Problems

1. The average score on an IQ test is 100, and the standard deviation is 15. What *z* score corresponds to a raw score of 115?

2. The average score on an IQ test is 100, and the standard deviation is 15. What *z* score corresponds to a raw score of 76?

3. The average score on an IQ test is 100, and the standard deviation is 15. What *z* score corresponds to a raw score of 97?

4. The average score on an IQ test is 100, and the standard deviation is 15. What *z* score corresponds to a raw score of 101?

5. The average score on an IQ test is 100, and the standard deviation is 15. What *z* score corresponds to a raw score of 87?

6. The average score on an IQ test is 100, and the standard deviation is 15. What raw score corresponds to a *z* score of -0.53?

7. The average score on an IQ test is 100, and the standard deviation is 15. What raw score corresponds to a *z* score of 1.40?

8. The average score on an IQ test is 100, and the standard deviation is 15. What raw score corresponds to a *z* score of 2.73?

9. The average score on an IQ test is 100, and the standard deviation is 15. What raw score corresponds to a *z* score of -1.93?

10. In Chapter 4 of this study guide, you had the opportunity to learn about the Clock Test. Here is the description 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 shown here are the percentages of two-step jumps that were missed by 100 participants during the last 15 minutes of a two-hour watch.

Using these statistics, create a boxplot. Write an explanation of what the boxplot shows.

Statistic	Clock Test Results
Mean	25
Median	27
Minimum	12
Maximum	38
25th percentile	21
75th percentile	29

11. In Chapter 3 and 4 of this study guide, you had the opportunity to find the measures of central tendency and variability for the age at which 38 children are first able to stand alone. The table that follows shows the statistics you may have already calculated.

Using these statistics, create a boxplot and write an explanation for parents who are not familiar with boxplots.

Statistic	Age (months)
Mean	13
Median	13
Minimum	9
Maximum	16
25th percentile	12
75th percentile	14

12. In Chapter 1 of this study guide, you read about a study examining the influence of air pollution on unethical behavior. In a second experiment reported in the same paper, Lu and colleagues (2018) exposed half of their participants to a series of photographs of locations with polluted air; the other half viewed a series of photographs of the same locations, but taken on clear days. They asked participants to imagine that the picture was taken where they lived and what it would feel like and be like walking around and breathing the air. Next, participants were given the opportunity to roll a die and self-report what number they rolled. They were to be paid whatever number they reported (i.e. $1 for a 1, $2 for a 2, etc.). If no participants cheated by reporting a number higher than they actually rolled, the expected average report for the die roll would be 3.5. (This is because all the die outcomes have an equal probability. So, $\left(\frac{1+2+3+4+5+6}{6}\right) = 3.5$). The group exposed to the polluted scenes self-reported an average roll of 4.46 and the group exposed to the clear scenes self-reported an average roll of 3.60. Let σ = 1.72. Find the effect size index for the difference between the two groups and write an interpretation.

13. Earlier you made a boxplot for the percentages of two-step jumps that were missed by 100 participants during the *last* 15 minutes of a two-hour watch on the Clock Test. The table shows the statistics you used to create that boxplot.

Determine if the lowest scoring participant in the distribution is an outlier.
Determine if the highest scoring participant in the distribution is an outlier.

Statistic	Clock Test Results
Mean	25
Median	27
Minimum	12
Maximum	38
25th percentile	21
75th percentile	29

14. Earlier you made a boxplot for the percentages of two-step jumps that were missed by 100 participants during the *last* 15 minutes of a two-hour watch on the Clock Test. The table shows the statistics you used to create that boxplot. Now it includes the percentages of two-step jumps that were missed by those same 100 participants during the *first* 15 minutes of that two-hour watch.

Compose a boxplot, calculate the effect size index, and write a Descriptive Statistics Report about the effect of two hours of vigilance on Clock Test performance. Let σ = 4.

Statistic	Time of Test	
	First 15 min of watch	*Last* 15 min of watch
Mean	12	25
Median	13	27
Minimum	8	12
Maximum	17	38
25th percentile	10	21
75th percentile	14	29

15. In Chapter 3 and 4 of this study guide, you had the opportunity to find the measures of central tendency and variability for the opinions of college students and older adults about the importance of social security as a source of income for the elderly. Here's the study description again if you need it.
In a study, college students and older adults were presented with this statement: "Social security is important as a means of income for the elderly." After reading this statement, participants identified the extent to which they agreed or disagreed with the statement using a Likert Scale, which is a common measurement of attitudes. Participants had five response options, ranging from "strongly agree" (1) to "strongly disagree" (5). The table shows the statistics you may have already calculated.

Statistic	Group	
	College Students	Older Adults
Mean	3.75	2.27
Median	4	2
Minimum	1	1
Maximum	5	5
25th percentile	3	1
75th percentile	5	3.25

Compose a boxplot, calculate the effect size index, and write a Descriptive Statistics Report for the opinions of college students and older adults about the importance of social security as a source of income for the elderly. Let $\sigma = 1.75$.

ANSWERS

Multiple-Choice Questions

1. c
   Explanation: Although the *z* score calculation includes a measure of central tendency (the mean) and a measure of variability (the standard deviation), a *z* score, itself, only conveys the relative position of an individual.
2. b
   Explanation: Remember that *z* scores tell you how far from the mean a score is in units of standard deviation, so a *z* score of 0 says that the score is 0 standard deviations away from the mean.
3. c
4. c
5. b
   Explanation: a. $z_{Zach} = \frac{X - \bar{X}}{S} = \frac{32 - 28}{2} = \frac{4}{2} = 2.00$

   b. $z_{Slater} = \frac{X - \bar{X}}{S} = \frac{34 - 29}{2} = \frac{5}{2} = 2.50$

   c. $z_{Screech} = \frac{X - \bar{X}}{S} = \frac{36 - 30}{6} = \frac{6}{6} = 1.00$

   d. $z_{Lisa} = \frac{X - \bar{X}}{S} = \frac{28 - 24}{3} = \frac{4}{3} = 1.33$
   Slater is the farthest above the mean relative to his peers compared to any of the other winners. Therefore, Slater should be considered the overall winner. Perhaps the year Screech won the game was easier than the year that Slater won and that is why Screech played the longest based on raw scores, but not using *z* scores.
6. c
7. d
   Explanation: Boxplots usually give information about the central tendency (mean and median), variability (IQR and range), and shape (symmetrical, positively skewed, negatively skewed).
8. a
9. a
   Explanation: Remember that the mean is more influenced by extreme scores than the median and therefore moves more toward the tail of the distribution than the median. When the longer tail is over the high scores, as it is in a positive distribution, the mean will move toward those higher scores.
10. c
11. b
12. c
13. c
    Explanation: The bottom (or left) of the box represents the 25th percentile and the top (or right) of the box represents the 75th percentile. Therefore the middle 50% of the distribution will be covered by the box portion of a boxplot.
14. d
    Explanation: For *d*,
    Small effect: *d* = 0.20
    Medium effect: *d* = 0.50
    Large effect: *d* = 0.80
15. d
    Explanation: For *d*,
    Small effect: *d* = 0.20
    Medium effect: *d* = 0.50
    Large effect: *d* = 0.80
16. d
    Explanation: An effect size index is a measure of the size of differences between distributions or the size of relationships.
17. b
18. a
19. d
20. c

Short-Answer Questions

1. A *z* score of 0 means Frank got the same score as the class average.
   Explanation: Remember *z* scores tell how far a score is from the mean, so to get a 0 means you scored the mean.

2. Small = 0.20; Medium = 0.50; Large = 0.80

3. Boxplots show the mean, median, range, and IQR of a distribution. Additionally, by comparing the positions of the mean and the median or the length of the whiskers, you can determine shape (i.e. symmetrical, positive skew, or negative skew). Sometimes, boxplots show outliers with symbols beyond the ends of the whiskers.

4. Look at the edge of the box at the top (in a vertical boxplot) or the far right (in a horizontal boxplot) which marks the 75th percentile of the distribution. That score has 75% of the scores below it and 25% of the scores above it.

5. Some statistics (e.g. the mean and the range) are influenced a lot by outliers. Therefore, to interpret your results accurately, you need to know if outliers are present.
   Explanation: You might be interested in knowing that sometimes researchers will remove outliers from calculations in order to have a statistic that better describes the group as a whole. For instance, imagine you are trying to describe the IQ of college students. You take a sample of 30 college students. One of your participants, Jaide, has an IQ of 195. Including Jaide in the sample leads to a mean of 117 and a standard deviation of 16. Removing Jaide leads to a sample mean of 103 with a standard deviation of 9. This leaves the researcher to make a decision about which set of statistics better describes the group.

6. 1) Calculate the relevant statistics (the mean, median, minimum, maximum, 25th percentile, 75th percentile, and effect size index-if you have two distributions) for each distribution of scores.
   2) Make a boxplot.
   3) Examine the boxplot and determine the results of the study.
   4) Write a narrative that explains the study outcome. Include information in the narrative about central tendency (means and medians), variability (ranges and IQR's), shape of the distributions, overlap (or lack thereof) of the distributions, and the size of the difference between the distributions..
   5) Reread and edit the narrative. Make sure you have actually explained the study outcome (e.g. use the terms of the experiment, express where groups differ, and tell the direction of differences) and have not just made a list of the relevant statistics.

Problems

1. $z = \frac{X-\mu}{\sigma} = \frac{115-100}{15} = \frac{15}{15} = 1.00$

2. $z = \frac{X-\mu}{\sigma} = \frac{76-100}{15} = \frac{-24}{15} = -1.60$

3. $z = \frac{X-\mu}{\sigma} = \frac{97-100}{15} = \frac{-3}{15} = -0.20$

4. $z = \frac{X-\mu}{\sigma} = \frac{101-100}{15} = \frac{1}{15} = 0.07$

5. $z = \frac{X-\mu}{\sigma} = \frac{87-100}{15} = \frac{-13}{15} = -0.87$

6. $z = \frac{X-\mu}{\sigma}$

   $-0.53 = \frac{X-100}{15}$

   Multiply both sides by 15.
   -7.95 = *X* - 100

   Add 100 to both sides.
   92.05 = *X*

   You can also use this simplified formula.
   X = μ + (*z*)(σ)
   X = 100 + (-0.53)(15)
   X = 100 + (-7.95)
   X = 92.05

7. $z = \frac{X-\mu}{\sigma}$

$1.40 = \frac{X-100}{15}$

Multiply both sides by 15.
21 = *X* - 100

Add 100 to both sides.
121 = *X*

You can also use this simplified formula.
X = μ + (*z*)(σ)
X = 100 + (1.40)(15)
X = 100 + (21)
X = 121

8. $z = \frac{X-\mu}{\sigma}$

$2.73 = \frac{X-100}{15}$

Multiply both sides by 15.
40.95 = *X* - 100

Add 100 to both sides.
140.95 = *X*

You can also use this simplified formula.
X = μ + (*z*)(σ)
X = 100 + (2.73)(15)
X = 100 + (40.95)
X = 140.95

9. $z = \frac{X-\mu}{\sigma}$

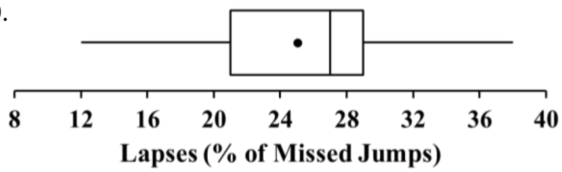
$-1.93 = \frac{X-100}{15}$

Multiply both sides by 15.
-28.95 = *X* - 100

Add 100 to both sides.
71.05 = *X*

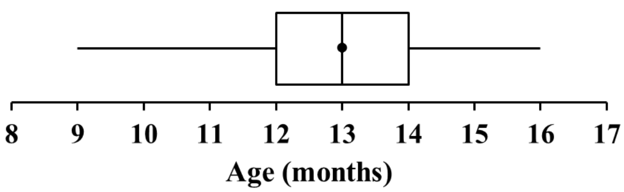
You can also use this simplified formula.
X = μ + (z)(σ)
X = 100 + (-1.93)(15)
X = 100 + (-28.95)
X = 71.05

10.



The mean percent of missed jumps during the clock test (lapses) during the last 15 minutes of a two-hour watch was 25, and the median was 27. There was a wide range of lapses among participants (range = 26, from 12 to 38) with the middle 50% of the scores falling in the 8-point IQR between 21 and 29. The distribution of lapses was negatively skewed.
Explanation: There are many ways to explain this boxplot correctly. A correct description includes central tendency (mean and median), variability (range and IQR), and form of the distribution (e.g. symmetrical, positive skew, or negative skew). Remember to use terms specific to the problem (e.g. percent of missed jumps during the clock test during the last 15 minutes of a two-hour watch).

11.



The average age a child first stood alone was 13 months, and 50% of children stood alone by 13 months (median = 13) as well. There is variability in when children stood alone. All of the children in this sample of 38 stood alone between 9 and 16 months (range = 7 months). The middle 50% of the distribution of children stood alone between 12 and 14 months (IQR = 2). The distribution of children's ages when they 1st stood alone is symmetrical.
Explanation: Because this description is intended for people who are not familiar with boxplots, *explanations* of statistics are provided along with the statistics, instead of just the statistics themselves (e.g., saying 50% of children stood alone by 13 months, instead of only saying the median was 13 months). How data are presented should always reflect the audience.

12. d = 0.50
There are many ways to describe the outcome correctly. Here are two examples.
Participants exposed to polluted scenes reported higher die rolls ($\overline{X}$ = 4.46), on average, than participants exposed to clear scenes ($\overline{X}$ = 3.60). The effect was medium (d = 0.50).
OR
There was a medium difference (d = 0.50) between the higher self-reported number for their die roll of those exposed to the polluted scenes ($\overline{X}$ = 4.46) and the lower reports of those exposed to the clear scenes ($\overline{X}$ = 3.60). These data suggest participants exposed to the polluted scenes were more likely to cheat than those exposed to the clear scenes.

Explanation:
$$d = \frac{\mu_1 - \mu_2}{\sigma} = \frac{4.46 - 3.60}{1.72} = 0.50$$
There are many ways to describe the outcome correctly. Your answer should include the size of the difference (e.g. small, medium, large), the direction of the difference (i.e., which group scored higher) and the terms of the experiment (polluted/clear scenes and the self-reported average die roll (or cheating)).

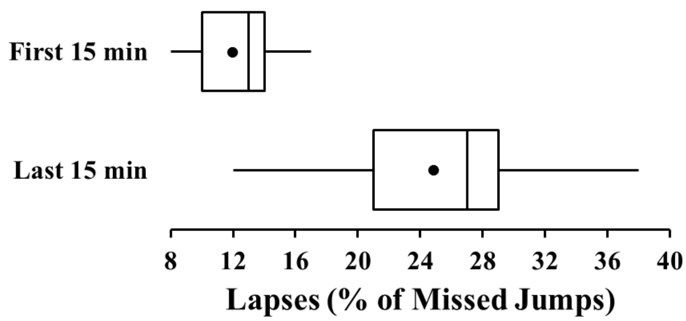
13. Neither the lowest scoring participant (X = 12) nor the highest scoring participant (X = 38) should be considered outliers for this distribution.
Explanation:
First, find the IQR.
IQR = 75th percentile – 25th percentile
IQR = 29 – 21
IQR = 8

Now find 1.5 × IQR
1.5 × 8 = 12

Now find the limit for low outliers.
25th percentile – (1.5 × IQR)
21 – 12 = 9
The minimum score here is 12 so the lowest scoring participant in the distribution is **not** an outlier because 12 does not fall below 9.

Now find the limit for high outliers.
75th percentile + (1.5 × IQR)
29 + 12 = 41
The maximum score here is 38 so the highest scoring participant in the distribution is **not** an outlier because 38 does not fall above 41.

14.



There was a huge difference (d = 3.25) in the number of lapses participants made at the beginning and the end of the two-hour watch. The participants made many fewer lapses during the first 15 minutes of the watch, with a mean of 12 lapses and a median of 13, compared to the last 15 minutes of the watch, with a mean of 25 lapses and a median of 27. Over 75% of the participants made more lapses during the last 15 minutes of the watch than any participant made during the first 15 minutes of the watch. The distribution of lapses at the beginning of the watch is slightly negatively skewed and more negatively skewed by the end of the watch. There is much more variability in performance after the two-hour watch (range = 26, IQR = 8) than before the watch (range = 9, IQR = 4). These data show that, while the two-hour watch had a large negative impact on performance, there is a lot of individual variation of that effect.

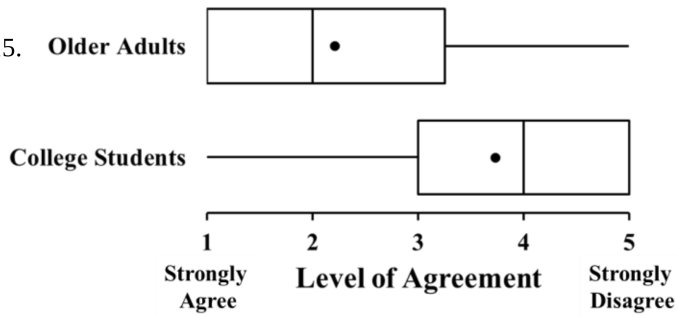
Explanation: There are many good Descriptive Statistics Reports. The one provided in the answer is just an example. You should make sure you follow the steps that follow.
First, calculate the effect size index you need.

$$d = \frac{\mu_1 - \mu_2}{\sigma} = \frac{25 - 12}{4} = 3.25$$

Remember to interpret d values of 0.20, 0.50, and 0.80 as "small", "medium", and "large", respectively, and use adjectives to modify the descriptors for intermediate and extreme values; 3.25 should be considered extremely large!

Next, make a boxplot. Be sure a) the boxes share the same axis to make for easier comparisons, b) you have labeled your axis with labeled tick marks and an axis title, and c) you have labeled the boxes. Then write a narrative that explains the study outcome. Many possible narratives could explain this well. Make sure yours includes information about central tendency (means and medians), variability (ranges and IQR's), shape of the distributions, overlap (or lack thereof) of the distributions, and the size of the difference between the distributions. Last, reread and edit the narrative. Be sure that you have actually explained the study outcome (e.g. use the terms of the experiment, describe differences in the distributions, and tell the direction of differences) and have not just made a list of the relevant statistics.

15.



Older adults were much more likely to agree that social security was an important source of income for the elderly than college students were. This large difference (d = 0.85) was reflected by the higher student mean of 3.75 and median of 4 compared to the lower older adult mean of 2.27 and median of 2 (remember that 5 corresponded to "strongly disagree"). The top scoring 75% of the college students more strongly disagreed with social security's importance to the elderly than 50% of the older adults. There is very similar variability in scores for the two age groups (range for both = 4, IQR for college students = 2 and for older adults = 2.25). However, the college student distribution is negatively skewed, with most of the scores falling at the end of the scale representing disagreement, whereas the older adult distribution was positively skewed, with most of the scores falling at the end of the scale representing agreement.

Explanation: There are many good Descriptive Statistics Reports. The one provided in the answer is just an example. Make sure to follow the steps from your textbook and include all relevant details in your narrative.

$$d = \frac{\mu_1 - \mu_2}{\sigma} = \frac{3.75 - 2.27}{1.75} = 0.85$$

References

Lu, J. G., Lee, J. J., Gino, G., & Galinsky, A. D. (2018). Polluted morality: Air pollution predicts criminal activity and unethical behavior. *Psychological Science, 29*(3), 340–355. doi: 10.1177/0956797617735807