8.3 Mean, Median and Mode

Figure 8.29 What does it mean to say someone has average height? (credit: modification of work “I’m the tallest” by Jenn Durfey/Flickr, CC BY 2.0)

Learning Objectives

After completing this section, you should be able to:

1. Calculate the mode of a dataset.
2. Calculate the median of a dataset.
3. Calculate the mean of a dataset.
4. Contrast measures of central tendency to identify the most representative average.
5. Solve application problems involving mean, median, and mode.

What exactly do we mean when we describe something as "average"? Is the height of an average person the height that more people share than any other? What if we line up every person in the world, in order from shortest to tallest, and find the person right in the middle: Is that person’s height the average? Or maybe it’s something more complicated.

Imagine a game where you and a friend are trying to guess the typical person’s height. Once the guesses are made, you bring in every person and measure their height. You and your friend figure out how far off each of your guesses were from the actual value, then square that number. The result is the number of points you earn for that person. After we check every height and award points accordingly, the person with the lower score wins (because a lower score means that person’s guess was, overall, closer to the actual values). Could we define the average height to be the number that you should guess to give you the smallest possible score?

Each of these three methods of determining the “average” is commonly used. They are all methods of measuring *centrality* (or *central tendency*). Centrality is just a word that describes the middle of a set of data. All give potentially different results, and all are useful for different reasons. In this section, we’ll explore each of these methods of finding the “average.”

The Mode

In our discussion of average heights, the first possible definition we offered was the height that more people share than any other. This is the **mode**, or the value that appears most often. If there are two modes, the data are **bimodal**.

Let’s look at some examples.

EXAMPLE 8.15

Finding the Mode Using a Stem-and-Leaf Plot

In Example 8.9, we looked at a stem-and-leaf plot of the sale prices (in dollars) of a particular collectible trading card:

0	5 8 9
1	0 0 0 3 4 4 5 5 5 5 6 9 9
2	0 0 0 0 5 5 9 9
3	0 0 0 5 5
4	0 0 5
5	
6	0

What is the mode price?

Solution

The mode is the price that appears most often. Both 15 and 20 appear 4 times, more than any other values. So, they are the modes (and we can conclude that this set of data is bimodal).

YOUR TURN 8.15

1. In Example 8.10, we constructed a stem-and-leaf plot for the number of times in one minute that different crickets chirped:

8	2 2 4 5 9 9 9
9	1 1 2 3 7 9
10	0 1 2 3 3 4 4 4 5 6 6 7 9 9
11	5 6
12	0

What is the mode of the number of chirps in a minute?

When we have a complete list of the data or a stem-and-leaf plot, it's pretty straightforward to find the mode; we just need to find the number that appears most often. If we're given a frequency distribution instead, the technique is different (but just as straightforward): we're looking for the number with the highest frequency.

EXAMPLE 8.16

Finding the Mode Using a Frequency Distribution

In Example 8.3, we created a frequency distribution of the number of siblings of conflict resolution class attendees.

Number of Siblings	Frequency
0	5
1	13
2	6
3	3
4	2
5	1

What is the mode of the number of siblings?

Solution

The mode is the value that appears the most often, which means it has the greatest frequency. Thirteen of the respondents have one sibling, more than any other number. So, the mode is 1.

YOUR TURN 8.16

1. In Your Turn 8.3, you found a frequency distribution for the number of people who shared a residence with people in a sample.

Number of People in the Residence	Frequency
1	12
2	13
3	8
4	6
5	1

What's the mode of the number of people in these residences?

What happens if there is no number in the data that appears more than once? In that case, by our definition, every data value is a mode. But according to some other definitions, the data would have no mode. In practice, though, it doesn't really matter; if no data value appears more than once, then the mode is not helpful at all as a measure of centrality.

The Median

Let's revisit our example of trying to identify the height of the "average" person. If we lined everyone up in order by height and found the person right in the middle, that person's height is called the **median**, or the value that is greater than no more than half and less than no more than half of the values.

Let's look at a really simple example. Consider the following list of numbers: 11, 12, 13, 13, 14. Is the first number on the list, 11, the median? There are no values less than 11 (that's 0%), and there are four values greater than 11 (that's 80%). Since more than 50% of the data are greater than 11, the definition is violated; it's not the median. Here's a chart with the rest of the data, with red shading to show where the definition is violated:

Data Value	Number of Values Below	Percentage of Values Below	Number of Values Above	Percentage of Values Above
11	0	0%	4	**80%**
12	1	20%	3	**60%**
13	2	40%	1	20%
14	4	**80%**	0	0%

Table 8.10

Only 13 has no violations, so it's the median according to the definition. In practice, we find the median just like we described in the average height example: by lining up all the data values in order from smallest to largest and picking the value in the middle. For our easy example (with data values 11, 12, 13, 13, 14), that first 13 is right in the middle; there are two values to the left and two values to the right. If there's not one value right in the middle, we pick the two closest, then choose the number exactly between them. For example, let's say we have the data 41, 44, 46, 53. Since there are an even number of data values in our list, we can't pick the one right in the middle. The two closest to the middle are 44 and 46, so we'll choose the number halfway between those to be the median: 45. As this example shows, the median (unlike the mode) doesn't have to be a number in our original set of data.

In the examples we've looked at so far, it's been pretty easy to identify which number is right in the middle. If we had a very large dataset, though, it might be harder. Fortunately, we have some formulas to help us with that.

FORMULA

Suppose we have a set of data with n values, ordered from smallest to largest. If n is odd, then the median is the data value at position $\frac{n+1}{2}$. If n is even, then we find the values at positions $\frac{n}{2}$ and $\frac{n}{2}+1$. If those values are named a and b, then the median is defined to be $\frac{a+b}{2}$.

Let's put those formulas to work in an example.

EXAMPLE 8.17

Finding the Median Using a Stem-and-Leaf Plot

In Example 8.9, we looked at a stem-and-leaf plot that contained 33 sale prices (in dollars) of a particular collectible trading card:

0	5 8 9
1	0 0 0 3 4 4 5 5 5 5 6 9 9
2	0 0 0 0 5 5 9 9
3	0 0 0 5 5
4	0 0 5
5	
6	0

What is the median price?

Solution

Step 1: Since 33 is odd, the median is the data value at position $\frac{n+1}{2}$, where n is the number of values in the dataset. There are 33 total values, so our formula becomes $\frac{33+1}{2} = 17$. That means we want to look for the 17th number in the dataset.

Step 2: We'll want to count from the lowest value to the 17th number. We can use our stem and leaf plot to do this.

	1 2 3
0	5 8 9
	4 5 6 7 8 9 10 11 12 13 14 15 16
1	0 0 0 3 4 4 5 5 5 5 6 9 9
	17
2	0 0 0 0 5 5 9 9
3	0 0 0 5 5
4	0 0 5
5	
6	0

The seventeenth number is 20, so the median is 20.

YOUR TURN 8.17

1. Consider the data given in this stem-and-leaf plot (there are 17 data values):

12	1 2 2 5
13	0 3 4 4 6 8
14	2 5 9 9
15	0 3
16	
17	0

What is the median of this data?

Now, let's tackle an example with an even number of values.

EXAMPLE 8.18

Finding the Median

In Example 8.10, we looked at the number of times different crickets (of differing species, genders, etc.) chirped in a one-minute span. That data is again provided below:

89	97	82	102	84	99
115	105	89	109	107	89
101	109	116	103	100	91
93	103	120	91	85	104
104	82	106	92	104	106

Find the median.

Solution

Step 1: In order to find the median, we first need to sort the data so that they're in order, smallest to largest:

82	82	84	85	89	89
89	91	91	92	93	97
99	100	101	102	103	103
104	104	104	105	106	106
107	109	109	115	116	120

Step 2: Next, we figure out how many data values we have. Counting them up, we see there are 30, which is even.

Step 3: Since we have an even number of data values, we need to find the values in positions $\frac{30}{2} = 15$ and $\frac{30}{2} + 1 = 16$. These are 101 and 102.

Step 4: We use the formula to compute the median: $\frac{101+102}{2} = 101.5$.

YOUR TURN 8.18

1. This table gives the records of the Major League Baseball teams at the end of the 2019 season:

Team	Wins	Losses	Team	Wins	Losses
HOU	107	55	PHI	81	81
LAD	106	56	TEX	78	84
NYY	103	59	SFG	77	85

Table 8.11 (source: www.mlb.com (http://www.mlb.com))

Team	Wins	Losses	Team	Wins	Losses
MIN	101	61	CIN	75	87
ATL	97	65	CHW	72	89
OAK	97	65	LAA	72	90
TBR	96	66	COL	71	91
CLE	93	69	SDP	70	92
WSN	93	69	PIT	69	93
STL	91	71	SEA	68	94
MIL	89	73	TOR	67	95
NYM	86	76	KCR	59	103
ARI	85	77	MIA	57	105
BOS	84	78	BAL	54	108
CHC	84	78	DET	47	114

Table 8.11 (source: www.mlb.com (http://www.mlb.com))

What is the median number of wins?

EXAMPLE 8.19

Finding the Median Using a Frequency Distribution

In Example 8.3, we created a frequency distribution of the number of siblings of the people who attended a conflict resolution class. Let's review that data again:

Number of Siblings	Frequency
0	5
1	13
2	6
3	3
4	2
5	1

What is the median of the number of siblings?

Solution

There are 30 data values total, so the median is between the 15th and 16th values in the ordered list. There are five 0s and thirteen 1s according to the frequency distribution, so items one through five are all 0s and items six through eighteen are all 1s. Since both items fifteen and sixteen are 1s, the median is 1.

YOUR TURN 8.19

1. In Your Turn 8.3, you found a frequency distribution for the number of people who shared a residence with people in a sample. Let's review that data again:

Number of People in the Residence	Frequency
1	12
2	13
3	8
4	6
5	1

What's the median of the number of people in these residences?

The Mean

Recall our example of ways we could identify the "average" height of an individual. The last method we discussed was also the most complicated. It involved a game where the player guesses a height, then figures out how far off that guess is from every single person's height. Those differences get squared and added together to get a score. Our next measure of centrality gives the lowest possible score: No other guess would beat it in the game. Given a dataset containing n total values, the **mean** of the dataset is the sum of all the data values, divided by n.

This is a computation you have likely done before. In many places, including spreadsheet programs like Microsoft Excel and Google Sheets, this number is called the *average*. For statisticians, though, the word *average* has too many possible meanings, so they prefer the one we'll use: mean.

EXAMPLE 8.20

Finding the Mean

Compute the mean of the numbers 12, 15, 17, 18, 18, and 19.

Solution

The mean is the sum of the values, divided by the number of values on the list. So, we get:

$$\frac{12+15+17+18+18+19}{6} = \frac{99}{6} = 16.5.$$

YOUR TURN 8.20

1. Compute the mean of the numbers 5, 8, 11, 12, 12, 12, 15, 18, and 20. Round your answer to three decimal places.

EXAMPLE 8.21

Finding the Mean Using a Frequency Distribution

Refer again to the frequency distribution of the number of siblings people who attended a conflict resolution class reported:

Number of Siblings	Frequency
0	5
1	13
2	6
3	3
4	2
5	1

What is the mean of the number of siblings?

Solution

Step 1: We compute the mean by adding up all the data values and then dividing by the number of data values on the list.

Step 2: Adding up the frequencies, we get $5 + 13 + 6 + 3 + 2 + 1 = 30$ data values in our list.

Step 3: Now, to find the sum of all the data values, we could simply reconstruct the raw data and add up all the numbers there. But, there's an easier way: Remember that repeatedly adding a number to itself is the definition of multiplication. So, for example, since there are six 2s in our data, the sum of all those 2s must be $6 \times 2 = 12$.

Step 4: Let's add a column to our distribution for these products:

Number of Siblings	Frequency	(Number of Siblings) $\times$ (Frequency)
0	5	0
1	13	13
2	6	12
3	3	9
4	2	8
5	1	5

Step 5: So, the sum of all our data values is $0 + 13 + 12 + 9 + 8 + 5 = 47$. The mean is $\frac{47}{30} \approx 1.567$.

YOUR TURN 8.21

1. Refer again to the frequency distribution of the number of people who shared a residence with people in a sample:

Number of People in the Residence	Frequency
1	12
2	13
3	8

What's the mean of the number of people in these residences?

As the number of data values we are considering grows, the computation for the mean gets more and more complicated. That's why people generally trust technology to perform that computation.

VIDEO

Compute Measures of Centrality Using Google Sheets (https://openstax.org/r/Google-sheet)

Note that a recent update to Google Sheets introduced a new function called "MODE.MULT," which will find every mode (not just the first one on the list).

EXAMPLE 8.22

Using Google Sheets to Compute Measures of Centrality

The dataset "InState" (https://openstax.org/r/Chapter8_Data-Sets) contains the in-state tuitions of every college and university in the country that reported that data to the Department of Education. Find the mode, median, and mean of those in-state tuition values.

Solution

Step 1: To find the mode, we select an empty cell type "=MODE(", click on the header of the column to insert a reference to the column into our formula, and then close the parentheses. When we hit the enter key, our formula is replaced with the mode: $13,380.

Step 2: We can find the median and mean using the same process, except using the functions "MEDIAN" and "AVERAGE" in place of "MODE". We find that the median is $11,207 and the mean is $15,476.79.

YOUR TURN 8.22

1. The dataset "MLB2019" (https://openstax.org/r/Chapter8_Data-Sets) gives the number of wins for every Major League Baseball team in the 2019 season. Use Google Sheets to find the mode(s), median, and mean.

Which Is Better: Mode, Median, or Mean?

If the mode, median, and mean all purport to measure the same thing (centrality), why do we need all three? The answer is complicated, as each measure has its own strengths and weaknesses. The mode is simple to compute, but there may be more than one. Further, if no data value appears more than once, the mode is entirely unhelpful. As for the mean and median, the main difference between these two measures is how each is affected by extreme values.

Consider this example: the mean and median of 1, 2, 3, 4, 5 are both 3. But what if the dataset is instead 1, 2, 3, 4, 10? The median is still 3, but the mean is now 4. What this example shows is that the mean is sensitive to extreme values, while the median isn't. This knowledge can help us decide which of the two is more relevant for a given dataset. If it is important that the really high or really low values are reflected in the measure of centrality, then the mean is the better option. If very high or low values are not important, however, then we should stick with the median.

The decision between mean and median only really matters if the data are skewed. If the data are symmetric, then the mean and median are going to be approximately equal, and the distinction between them is irrelevant. If the data are skewed, the mean gets pulled in the direction of the skew (i.e., if the data are right-skewed, then the mean will be bigger

than the median; if the data are left-skewed, the opposite relation is true).

EXAMPLE 8.23

Choosing Which Measure of Centrality to Use

For the following situations, decide which measure(s) of centrality would be best:

1. You found a used car that you like, and you want to know if the price is too high or too low. You find a list of sale prices for that make and model, and you see that the distribution of those prices is skewed to the right. Some of the prices at the high end are close to the original sale price of the car, so you guess that those cars might have really low mileage, or have other enhancements added on that increased the value (but which don't apply to the car you found).
2. You are asked to analyze the responses to a survey. One of the questions asked, "How strongly do you agree with the statement, 'I believe my elected representatives have my best interests in mind when they vote'?" Responses are a number between 1 and 5, with 1 representing "strongly disagree" and 5 representing "strongly agree."
3. You are asked to find the "average" household income for a zip code. Those values are skewed right.
4. Thinking back to the situation at the beginning of this section: you want to find "average" height. The data you've collected seem to be distributed symmetrically.

Solution

1. In this situation, the high values are not comparable to the value of the car you found and we don't want them to affect the results. Also, we're unlikely to find many repeated values, so the mode is probably not useful. Median is best.
2. Here, we want to know what a typical result is. The mean doesn't really make sense; it involves adding the numbers together, so it would treat two "strongly disagree" and two "strongly agree" responses (those add to $1 + 1 + 5 + 5 = 12$) as exactly the same as four "neutral" responses ($3 + 3 + 3 + 3 = 12$). But those are really different situations; the first shows a strong polarization in the responses, while the second represents strong indifference. The mode is probably the best choice (because the data are actually categorical), but the median would be good too.
3. The mode isn't going to be useful in this situation because it's unlikely you will find many households that have exactly the same income. The mean and median will be different because of the skew, so the choice comes down to the extreme values. Remember that the data are skewed right, so high values are prevalent. Because these high values are important for our analysis, we want them to be reflected in the results. Thus, the mean is best. That being said, the median is also useful; it allows us to say something like "50% of the households surveyed make more than" the median.
4. Because we aren't likely to find many people with exactly the same height, the mode won't be useful. Since the data are symmetric, the mean and the median will be about the same. So, it doesn't really matter which of those two we choose.

YOUR TURN 8.23

For the following situations, decide which measure(s) of centrality would be best:

1. The data come from a survey question where the responses range from 1 ("not interested") to 10 ("very interested").

2. The data are fuel efficiency measures (in miles per gallon) for various vehicles. The data are right-skewed.

3. The data are scores on an exam, and they are left-skewed.

Check Your Understanding

13. Use the given stem-and-leaf plot to determine the mode, median, and mean:

8	8 9
9	0 0 7

10	2 5 6 7 8
11	1 2 2 2 4 5 7 9 9
12	0 0 3 5 7
13	0 1 1 4

A survey of college students asked how many courses those students were currently taking. The results are summarized in this frequency distribution:

Number of Classes	Frequency
1	1
2	3
3	16
4	8
5	4

14. Find the mode of the number of classes.
15. Find the median of the number of classes.
16. Find the mean of the number of classes.

Employees at a college help desk track the number of people who request assistance each week, as recorded below:

142	153	158	156	141	143	155	167	143	168
139	158	156	146	137	153	138	156	164	130
136	127	157	148	132	139	133	157	148	136

17. Compute the mode.
18. Compute the median.
19. Compute the mean.

The following table provides admission rates of the different branch campuses in the University of California system, along with the out-of-state tuition and fee cost:

Campus	Admission Rate	Cost ($)
Berkeley	0.1484	43,176
Davis	0.4107	43,394
Irvine	0.2876	42,692
Los Angeles	0.1404	42,218

(source: https://data.ed.gov)

Campus	Admission Rate	Cost ($)
Merced	0.6617	42,530
Riverside	0.5057	42,819
San Diego	0.3006	43,159
Santa Barbara	0.322	43,383
Santa Cruz	0.4737	42,952

(source: https://data.ed.gov)

20. Compute the mode, median, and mean admission rate.
21. Compute the mode, median, and mean cost.

22. Consider the data in "InState," (https://openstax.org/r/Chapter8_Data-Sets), which shows in-state tuition and fees for most institutions of higher learning in the United States. Compute the mode, median, and mean.

23. Workers for a particular company have complained that their wages are too low, while the management says wages are plenty high. If the company's salaries are right-skewed, which measure of centrality will the workers choose to represent salary? Which will management choose?

SECTION 8.3 EXERCISES

Use the steam-and-leaf plot below to answer the questions. Round all answers to three decimal places.

0	1 1 1 2 6 6 7
1	0 0 0 0 2 5 5 9
2	0 0 5 9
3	0 5
4	0

1. Find the mode.
2. Find the median.
3. Find the mean.

Use the stem-and-leaf plot below to answer the following questions. Round all answers to three decimal places.

7	8
8	0 6
9	1 1 4 8
10	0 0 3 5 9
11	0 0 1 1 1 2 6 8
12	0 1 5

4. Find the mode.
5. Find the median.
6. Find the mean.

The following table shows the frequency distribution for a number of countries visited by students. Use this table to answer the questions, rounding all answers to three decimal places.

Number of Countries	Frequency
0	13
1	11
2	5
3	1

7. Find the mode of the number of countries visited.
8. Find the median of the number of countries visited.
9. Find the mean of the number of countries visited.

The following table shows the frequency distribution for the number of fumbles by a selection of football players. Use this table to answer the questions, rounding all answers to three decimal places.

Number of Fumbles	Frequency
0	5
1	13
2	3
3	4

10. Find the mode of the number of fumbles.
11. Find the median of the number of fumbles.
12. Find the mean of the number of fumbles.

A public opinion poll about an upcoming election asked respondents, “How many political advertisements do you recall seeing on television in the last 24 hours?” Use this data to answer the questions, rounding all answers to three decimal places.

6	2	5	5	2	2	4	1	3	0	1	2	1	6	2
5	2	4	8	6	3	3	4	2	5	3	4	2	2	3

13. Compute the mode.
14. Compute the median.
15. Compute the mean.

For the following exercises, use the data found in “TNSchools” (https://openstax.org/r/Chapter8_Data-Sets), which has data on many institutions of higher education in the state of Tennessee. Here are what the columns represent:

Column Name	Description
AdmRate	Proportion of applicants that are admitted
UGEnr	Number of undergraduate students
PTUG	Proportion of undergraduates who attend part-time
InState	Tuition and fees for in-state students
OutState	Tuition and fees for out-of-state students
FacSal	Mean monthly faculty salary
Pell	Proportion of students receiving Pell Grants
MedDebt	Median student loan debt at degree completion
StartAge	Mean age at the time of entry
Female	Proportion of students who identify as female

(source: https://data.ed.gov)

16. What is the mean undergraduate enrollment at these schools?
17. What is the median undergraduate enrollment at these schools?
18. Generate a histogram for undergraduate enrollment, and use it to describe the effect of the distribution on the mean and the median as measures of centrality.
19. What is the mean proportion of part-time undergraduates at these schools?
20. What is the median proportion of part-time undergraduates at these schools?
21. Generate a histogram for proportion of part-time undergraduates, and use it to describe the effect of the distribution on the mean and the median as measures of centrality.
22. What is the mean of the monthly faculty salaries at these schools?
23. What is the median of the monthly faculty salaries at these schools?
24. Generate a histogram for the monthly faculty salaries, and use it to describe the effect of the distribution on the mean and the median as measures of centrality.

The dataset "AvgSAT" (https://openstax.org/r/Chapter8_Data-Sets) gives the average SAT score for incoming students (the use of "average" here stands in for "mean" to avoid confusion in the following problems) for every institution in the United States that reported the data (from data.ed.gov). Use that dataset to answer the following questions.

25. What is the mean of the average SAT scores?
26. What is the median of the average SAT scores?
27. Construct a histogram of average SAT scores, and describe how the distribution affects the mean and the median.

For the following exercises, use the table, which gives the final results for the 2021 National Women's Soccer League season. The columns are standings points (PTS; teams earn three points for a win and one point for a tie), wins (W), losses (L), ties (T), goals scored by that team (GF), and goals scored against that team (GA).

Team	Points	W	L	T	GF	GA
Portland Thorns FC	44	13	6	5	33	17
OL Reign	42	13	8	3	37	24

(source: http://www.nwslsoccer.com)

Team	Points	W	L	T	GF	GA
Washington Spirit	39	11	7	6	29	26
Chicago Red Stars	38	11	8	5	28	28
NJ/NY Gotham FC	35	8	5	11	29	21
North Carolina Courage	33	9	9	6	28	23
Houston Dash	32	9	10	5	31	31
Orlando Pride	28	7	10	7	27	32
Racing Louisville FC	22	5	12	7	21	40
Kansas City Current	16	3	14	7	15	36

(source: http://www.nwslsoccer.com)

28. Compute the mode(s), median, and mean of points.
29. Compute the mode(s), median, and mean of wins.
30. Compute the mode(s), median, and mean of losses.
31. Compute the mode(s), median, and mean of ties.
32. Compute the mode(s), median, and mean of goals scored (GF).
33. Compute the mode(s), median, and mean of goals against (GA).

For the following exercises, use the data in "CUNY" (https://openstax.org/r/Chapter8_Data-Sets) (taken from data.ed.gov, which gives proportions of degrees awarded in various disciplines) to answer the given questions:

34. Compute the median and mean proportions of degrees awarded in Information Science.
35. Compute the median and mean proportions of degrees awarded in Education.
36. Compute the median and mean proportions of degrees awarded in Humanities.
37. Compute the median and mean proportions of degrees awarded in Biology.
38. Compute the median and mean proportions of degrees awarded in Physical Science.
39. Compute the median and mean proportions of degrees awarded in Social Science.

8.4 Range and Standard Deviation

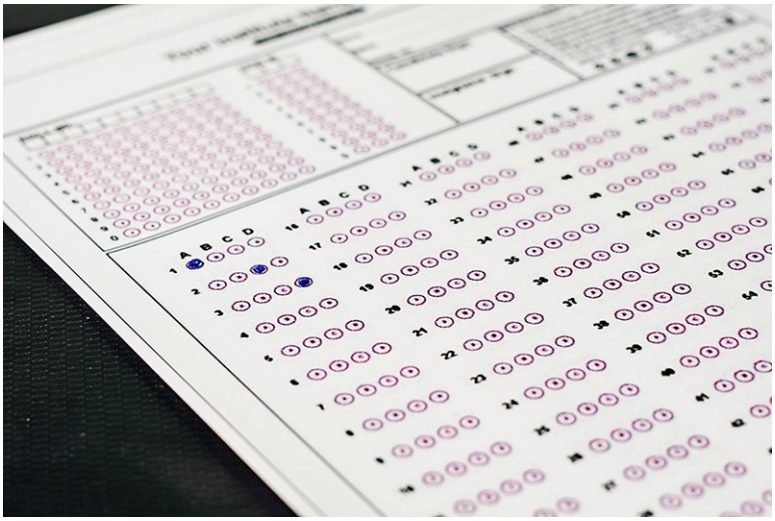

Figure 8.30 Measures of spread help us get a better understanding of test scores. (credit: "Standardized test exams form with answers bubbled" by Marco Verch Professional Photographer/Flickr, CC BY 2.0)

Learning Objectives

After completing this section, you should be able to:

1. Calculate the range of a dataset
2. Calculate the standard deviation of a dataset

Measures of centrality like the mean can give us only part of the picture that a dataset paints. For example, let's say you've just gotten the results of a standardized test back, and your score was 138. The mean score on the test is 120. So, your score is above average! But how good is it *really*? If all the scores were between 100 and 140, then you know your score must be among the best. But if the scores ranged from 0 to 200, then maybe 140 is good, but not great (though still above average). Knowing information about how the data are spread out can help us put a particular data value in better context. In this section, we'll look at two numbers that help us describe the spread in the data: the range and the standard deviation. These numbers are called measures of dispersion.

The Range

Our first measure of dispersion is the **range**, or the difference between the maximum and minimum values in the set. It's the measure we used in the standardized test example above.

Let's look at a couple of examples.

EXAMPLE 8.24

Finding the Range

You survey some of your friends to find out how many hours they work each week. Their responses are: 5, 20, 8, 10, 35, 12. What is the range?

Solution

The maximum value in the set is 35 and the minimum is 5, so the range is $35 - 5 = 30$.

YOUR TURN 8.24

1. On your morning commute, you decide to record how long you have to wait each time you get caught at a red light. Here are the times in seconds: 12, 58, 35, 79, 21. What is the range?

For large datasets, finding the maximum and minimum values can be daunting. There are two ways to do it in a spreadsheet. First, you can ask the spreadsheet program to sort the data from smallest to largest, then find the first and last numbers on the sorted list. The second method uses built-in functions to find the minimum and maximum.