8.4 Range and Standard Deviation

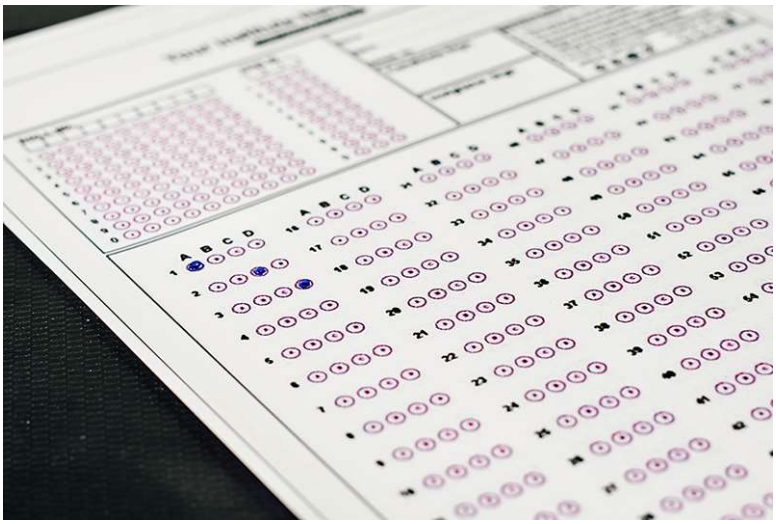

Figure 8.30 Measures of spread help us get a better understanding of test scores. (credit: "Standardized test exams form with answers bubbled" by Marco Verch Professional Photographer/Flickr, CC BY 2.0)

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.

VIDEO

Find the Minimum and Maximum Using Google Sheets (https://openstax.org/r/min-max_Google-Sheet)

In either method, once you've found the maximum and minimum, all you have to do is subtract to find the range.

EXAMPLE 8.25

Finding the Range with Google Sheets

The data in "AvgSAT" (https://openstax.org/r/Chapter8_Data-Sets) contains the average SAT score for students attending every institution of higher learning in the US for which data is available. What is the range of these average SAT scores?

Solution

Step 1: To find the maximum, click on an empty cell in the spreadsheet, type "=MAX(", and then click on the letter that marks the top of the column containing the AvgSAT data. That inserts a reference to the column into our function. Then we close the parentheses and hit the enter key. The formula is replaced with the maximum value in our data: 1566.
Step 2: Using the same process (but with "MIN" instead of "MAX"), we find the minimum value is 785.
Step 3: So, the range is $1566 - 785 = 781$.

YOUR TURN 8.25

1. The file "InState" (https://openstax.org/r/Chapter8_Data-Sets) contains in-state tuition costs (in dollars) for every institution of higher learning in the US for which data is available. What is the range of these costs?

The range is very easy to compute, but it depends only on two of the data values in the entire set. If there happens to be just one unusually high or low data value, then the range might give a distorted measure of dispersion. Our next measure takes every single data value into account, making it more reliable.

The Standard Deviation

The standard deviation is a measure of dispersion that can be interpreted as approximately the average distance of every data value from the mean. (This distance from the mean is the "deviation" in "standard deviation.")

FORMULA

The standard deviation is computed as follows:

$$s = \sqrt{\frac{\sum (x - \overline{x})^2}{n - 1}}$$

Here, x represents each data value, $\overline{x}$ is the mean of the data values, n is the number of data values, and the capital sigma (Σ) indicates that we take a sum.

To compute the standard deviation using the formula, we follow the steps below:

1. Compute the mean of all the data values.
2. Subtract the mean from each data value.
3. Square those differences.
4. Add up the results in step 3.
5. Divide the result in step 4 by $n - 1$
6. Take the square root of the result in step 5.

Let's see that process in action.

EXAMPLE 8.26

Computing the Standard Deviation

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

Solution

Let's follow the six steps mentioned previously to compute the standard deviation.

Step 1: Find the mean: $\overline{x} = \frac{5+20+8+10+35+12}{6} = 15$.

Step 2: Subtract the mean from each data value. To help keep track, let's do this in a table. In the first row, we'll list each of our data values (and we'll label the row x); in the second, we'll subtract $\overline{x} = 15$ from each data value.

x	5	20	8	10	35	12
$x - \overline{x}$	−10	5	−7	−5	20	−3

Step 3: Square the differences. Let's add a row to our table for those values:

x	5	20	8	10	35	12
$x - \overline{x}$	−10	5	−7	−5	20	−3
$(x - \overline{x})^2$	100	25	49	25	400	9

Step 4: Add up those squares: $100 + 25 + 49 + 25 + 400 + 9 = 608$.

Step 5: Divide the sum by $n - 1$. Since we have 6 data values, that gives us $\frac{608}{6-1} = 121.6$.

Step 6: Take the square root of the result: $\sqrt{121.6} \approx 11.027$.

Thus, the standard deviation is $s \approx 11.027$.

YOUR TURN 8.26

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 standard deviation?

The computation for the standard deviation is complicated, even for just a small dataset. We'd never want to compute it without technology for a large dataset! Luckily, technology makes this calculation easy.

VIDEO

Find the Standard Deviation Using Google Sheets (https://openstax.org/r/StanDev_Google-Sheet)

EXAMPLE 8.27

Finding the Standard Deviation with Google Sheets

The data in "AvgSAT" (https://openstax.org/r/Chapter8_Data-Sets) contains the average SAT score for students attending every institution of higher learning in the US for which data is available. What is the standard deviation of these average SAT scores?

Solution

To find the standard deviation, we click in an empty cell in our spreadsheet and then type "=STDEV(". Next, click on the letter at the top of the column containing our data; this will put a reference to that column into our formula. Then close the parentheses with and hit the enter key. The formula is replaced with the result: 125.517.

YOUR TURN 8.27

1. The file "InState" (https://openstax.org/r/Chapter8_Data-Sets) contains in-state tuition costs (in dollars) for every institution of higher learning in the US for which data is available. What is the standard deviation of these costs?

Check Your Understanding

24. Given the data 1, 4, 5, 5, and 10, find the range.

25. Given the data 1, 4, 5, 5, and 10, find the standard deviation using the process outlined in the definition.

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

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

26. Compute the range.

27. Compute the standard deviation.

The following are data on the 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
Merced	0.6617	42,530
Riverside	0.5057	42,819
San Diego	0.3006	43,159

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

Campus	Admission Rate	Cost ($)
Santa Barbara	0.322	43,383
Santa Cruz	0.4737	42,952

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

28. Compute the range of the admission rate.
29. Compute the standard deviation of the admission rate.

Using the data from Table 8.21, find the:
30. Range of the cost.
31. Standard deviation of the cost.

SECTION 8.4 EXERCISES

Given the data 81, 84, 89, 85, 86, 84:
1. Compute the range.
2. Compute the standard deviation using the formula.

Given the data 127, 167, 156, 158, 156, 157:
3. Compute the range.
4. Compute the standard deviation using the formula.

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)

5. Find the range of admissions rates.
6. Find the standard deviation of admissions rates.
7. Find the range of undergraduate enrollments.
8. Find the standard deviation of undergraduate enrollments.
9. Find the range of the proportions of undergraduates attending part-time.

10. Find the standard deviation of the proportions of undergraduates attending part-time.
11. Find the range of in-state tuition costs.
12. Find the standard deviation of in-state tuition costs.
13. Find the range of median debt.
14. Find the standard deviation of median debt.

For the following exercises, use the table below, 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	PTS	W	L	T	GF	GA
Portland Thorns FC	44	13	6	5	33	17
OL Reign	42	13	8	3	37	24
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

15. Compute the standard deviation and range of points.
16. Compute the standard deviation and range of wins.
17. Compute the standard deviation and range of losses.
18. Compute the standard deviation and range of ties.
19. Compute the standard deviation and range of goals scored (GF).
20. Compute the standard deviation and range of goals against (GA).