5.3 Standard Deviation

YOU WILL NEED

- calculator OR computer with spreadsheet software

GOAL

Determine the standard deviation for sets of data, and use it to solve problems and make decisions.

EXPLORE...

- A teacher has two chemistry classes. She gives the same tests to both classes. Examine the mean mark for each of the first five tests given to both classes. Compare the results for the two classes.

Test	Class A (%)	Class B (%)
1	94	84
2	56	77
3	89	76
4	67	81
5	84	74

INVESTIGATE the Math

The coach of a varsity girls' basketball team keeps statistics on all the players. Near the end of one game, the score is tied and the best starting guard has fouled out. The coach needs to make a substitution. The coach examines the field goal stats for five guards on the bench in the last 10 games.

Player	Field Goal Percent in Last 10 Basketball Games									
Anna	36	41	43	39	45	27	40	37	31	28
Patrice	36	39	36	38	35	37	35	36	38	34
Morgan	34	41	38	37	48	19	33	43	21	44
Paige	34	35	33	35	33	34	33	35	34	33
Star	41	33	39	36	38	36	29	34	38	39

? How can the coach use the data to determine which player should be substituted into the game?

A. Which player seems to be the most consistent shooter? Explain.

B. Analyze the data for Paige using a table like the one shown on the next page. Determine the mean of the data, $\bar{x}$, for Paige, and record this value in the first column.

C. Complete the second column for the **deviation** of each field goal percent: $(x - \bar{x})$

D. Complete the third column for the squares of the deviations.

Communication *Tip*

The symbol $\bar{x}$ (read as "*x* bar") represents the mean of the data.

deviation

The difference between a data value and the mean for the same set of data.

Paige's Field Goal (%)	Deviation $(x - \bar{x})$	Square of Deviation $(x - \bar{x})^2$
34	0.1	0.01
35	1.1	1.21
33		

E. Determine the **standard deviation** of Paige's data by following these steps:

Step 1: Determine the mean of the squares of the deviations.
Step 2: Determine the square root of the mean from Step 1. This number is the standard deviation.

standard deviation

A measure of the dispersion or scatter of data values in relation to the mean; a low standard deviation indicates that most data values are close to the mean, and a high standard deviation indicates that most data values are scattered farther from the mean.

F. Analyze all the data using a spreadsheet like the one below. Enter the field goal percent for Anna in row 2, and for Patrice, Morgan, Paige, and Star in rows 3, 4, 5, and 6, as shown.

	A	B	C	D	E	F	G	H	I	J	K	L	M
1	Game Player	1	2	3	4	5	6	7	8	9	10	Mean	Standard Deviation
2	Anna	36	41	43	39	45	27	40	37	31	28		
3	Patrice	36	39	36	38	35	37	35	36	38	34		
4	Morgan	34	41	38	37	48	19	33	43	21	44		
5	Paige	34	35	33	35	33	34	33	35	34	33		
6	Star	41	33	39	36	38	36	29	34	38	39		

G. Using the features of the spreadsheet software, determine the mean and standard deviation for each set of data.

H. Compare your result from part E with your results for Paige from part G. What do you notice?

I. Examine the means of the players. Would you use the most consistent player identified in part A as a substitute? Explain.

J. Which player has the greatest percent range? Which player has the least? How do the standard deviations of these players compare?

K. Compare the means and standard deviations of the data sets for all the players. Which player's data has the lowest standard deviation? What does this imply about her shooting consistency?

L. Based on past performance, which player has the potential to shoot most poorly? Which player has the potential to shoot most successfully?

M. If you were the coach, which player would you substitute into the game? Explain why.

Reflecting

N. The mean, $\bar{x}$, can be expressed using symbols:

$$\bar{x} = \frac{\Sigma x}{n}$$

Based on your understanding of the mean, what does the symbol Σ represent?

O. The standard deviation, $\boldsymbol{\sigma}$, can also be expressed using symbols:

$$\sigma = \sqrt{\frac{\Sigma(x - \bar{x})^2}{n}}$$

Interpret this expression verbally.

Communication | *Tip*

The symbol $\boldsymbol{\sigma}$ (read as "sigma") represents the standard deviation of the data.

P. Standard deviation is a measure of dispersion, of how the data in a set is distributed. How would a set of data with a low standard deviation differ from a set of data with a high standard deviation?

APPLY the Math

EXAMPLE 1 Using standard deviation to compare sets of data

Brendan works part-time in the canteen at his local community centre. One of his tasks is to unload delivery trucks. He wondered about the accuracy of the mass measurements given on two cartons that contained sunflower seeds. He decided to measure the masses of the 20 bags in the two cartons. One carton contained 227 g bags, and the other carton contained 454 g bags.

Masses of 227 g Bags (g)			
228	220	233	227
230	227	221	229
224	235	224	231
226	232	218	218
229	232	236	223

Masses of 454 g Bags (g)			
458	445	457	458
452	457	445	452
463	455	451	460
455	453	456	459
451	455	456	450

How can measures of dispersion be used to determine if the accuracy of measurement is the same for both bag sizes?

Brendan's Solution

227 g bags:
Range = 236 g − 218 g
Range = 18 g

454 g bags:
Range = 463 g − 445 g
Range = 18 g

I examined each set of data and found the greatest mass and least mass. Then I determined the range. The range was the same for both cartons.

227 g bags:
$\bar{x} = 227.15$ g
$\sigma = 5.227...$ g

454 g bags:
$\bar{x} = 454.4$ g
$\sigma = 4.498...$ g

I used my graphing calculator to determine the mean and standard deviation for each set of data.

Both means were above the mass measurements given on the two cartons.

The accuracy of measurement is not the same for both sizes of bag.

The standard deviation for the 454 g bags is less than the standard deviation for the 227 g bags.

The difference in the standard deviations indicates that the masses of the larger bags were closer to their mean mass.

Therefore, the 454 g bags of sunflower seeds have a more consistent mass.

Your Turn

a) Explain why the standard deviations for the masses of the two sizes of bag are different, even though the ranges of the masses are the same.
b) How might standard deviation be used by the company that sells the sunflower seeds for quality control in the packaging process?

EXAMPLE 2 Determining the mean and standard deviation of grouped data

Angèle conducted a survey to determine the number of hours per week that Grade 11 males in her school play video games. She determined that the mean was 12.84 h, with a standard deviation of 2.16 h.

Janessa conducted a similar survey of Grade 11 females in her school. She organized her results in this frequency table. Compare the results of the two surveys.

Gaming Hours per Week for Grade 11 Females	
Hours	**Frequency**
3–5	7
5–7	11
7–9	16
9–11	19
11–13	12
13–15	5

Cole's Solution: Determining $\bar{x}$ and σ manually

Note: The purpose of Cole's Solution is to provide an understanding of what technology does to calculate the mean and standard deviation when working with grouped data. Students are not expected to determine mean and standard deviation manually.

An estimate for the mean of the gaming hours for Grade 11 females is 9.

I predicted that the mean is about 9 h because most of the data is in the 7 to 11 intervals. However, I need to verify my estimate.

A	B	C	D
Hours	**Frequency (*f*)**	**Midpoint of Interval (*x*)**	***f* · *x***
3–5	7	4	28
5–7	11	6	66
7–9	16	8	128
9–11	19	10	190
11–13	12	12	144
13–15	5	14	70
	70		626

I didn't know the actual values in each interval, so I determined the midpoint of each interval. I knew that some values would be greater than the midpoint and some values would be less, but I thought that the midpoint could represent all the values in each interval.

I multiplied the frequency by the midpoint for each interval to determine the number of hours in each interval.

$$\bar{x} = \frac{\Sigma(f)(x)}{n}$$

$$\bar{x} = \frac{626}{70}$$

$$\bar{x} = 8.942\ldots \text{ h}$$

Next, I determined the mean for the data set. I divided the total number of hours by the total number of data values.

C	D	E	F
Midpoint of Interval (*x*)	***f* · *x***	**$(x - \bar{x})^2$**	**$f \cdot (x - \bar{x})^2$**
4	28	24.431...	171.022...
6	66	8.660...	95.264...
8	128	0.888...	14.223...
10	190	1.117...	21.233...
12	144	9.346...	112.153...
14	70	25.574...	127.873...
	626		541.771...
$\bar{x}$ = 8.942... h			

In column E, I squared the deviation from the midpoint for each interval.

In column F, I multiplied each squared value by the frequency, and I added these products to estimate the total square deviations for all the data.

$$\sigma = \sqrt{\frac{\Sigma f(x - \bar{x})^2}{n}}$$

$$\sigma = \sqrt{\frac{541.771...}{70}}$$

$$\sigma = 2.782... \text{ h}$$

I determined the standard deviation by dividing the sum of the squares of the deviations by the number of data values and then taking the square root.

Males:	Females:
$\bar{x} = 12.84$ h	$\bar{x} = 8.942...$ h
$\sigma = 2.16$ h	$\sigma = 2.782...$ h

I compared the results for the two groups.

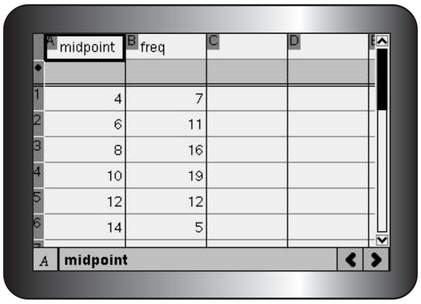
The males played nearly 4 h more per week than the females, on average.

The standard deviation for males is lower than the standard deviation for females. Therefore, the males' playing times are closer to their mean (almost 13 h) and don't vary as much.

The mean playing time for males is higher than the mean playing time for females.

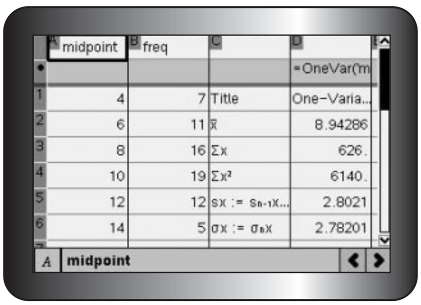
The data for the females is more dispersed than the data for the males.

Danica's Solution: Using technology to determine $\bar{x}$ and σ



	A midpoint	B freq	C	D
1	4	7		
2	6	11		
3	8	16		
4	10	19		
5	12	12		
6	14	5		

First, I determined the midpoints of the intervals for Janessa's data and entered these values in one list on my calculator. Then I entered the frequency in a second list.



	A midpoint	B freq	C	D
•				=OneVar('m
1	4	7	Title	One-Varia...
2	6	11	$\bar{x}$	8.94286
3	8	16	Σx	626.
4	10	19	Σx²	6140.
5	12	12	sx := sn-1x...	2.8021
6	14	5	σx := σnx	2.78201

I determined the mean and the standard deviation.

Gaming hours per week for Grade 11 females:
$\bar{x} = 8.942...$ h $\quad \sigma = 2.782...$ h
Gaming hours per week for Grade 11 males:
$\bar{x} = 12.84$ h $\quad \sigma = 2.16$ h

I compared the results for the two groups.

The standard deviation for the females is higher than the standard deviation for the males.

Therefore, the females' times vary more from their mean of about 9 h.

The standard deviation for the males is lower. Therefore, their data is more consistent, even though their mean is higher.

The females, on average, spent less time playing video games per week. However, some females played a lot more or a lot less than the mean.

The males played more hours per week, and their times were fairly close to their mean.

Your Turn

Could the mean and standard deviation for the female data differ from those determined by Danica and Cole, if the actual data is used? Explain.

In Summary

Key Ideas

- To determine how scattered or clustered the data in a set is, determine the mean of the data and compare each data value to the mean.
- The standard deviation, σ, is a measure of the dispersion of data about the mean.
- The mean and standard deviation can be determined using technology for any set of numerical data, whether or not the data is grouped.

Need to Know

- When data is concentrated close to the mean, the standard deviation, σ, is low. When data is spread far from the mean, the standard deviation is high. As a result, standard deviation is a useful statistic to compare the dispersion of two or more sets of data.
- When determining the standard deviation, σ, for a set of data using technology, this is the process that is followed:
  1. The square of the deviation of each data value (or the midpoint of the interval) from the mean is determined: $(x - \overline{x})^2$
  2. The mean of the squared deviations of all the data values is determined.
  3. The square root of the mean from step 2 is determined. This value is the standard deviation.
- Standard deviation is often used as a measure of consistency. When data is closely clustered around the mean, the process that was used to generate the data can be interpreted as being more consistent than a process that generated data scattered far from the mean.

CHECK Your Understanding

Test	Class A (%)	Class B (%)
1	94	84
2	56	77
3	89	76
4	67	81
5	84	74

1. **a)** Determine, by hand, the standard deviation of test marks for the two chemistry classes shown.
   b) Verify your results from part a) using technology.
   c) Which class had the more consistent marks over the first five tests? Explain.

Use technology to determine the mean and standard deviation, as needed, in questions 2 to 14.

2. Ali bowls in a peewee league. Determine the mean and standard deviation of Ali's bowling scores, rounded to two decimal places.

135	156	118	133
141	127	124	139
109	131	129	123

Bowling Scores	Frequency
101–105	1
106–110	3
111–115	4
116–120	7
121–125	9
126–130	14
131–135	11
136–140	8
141–145	6
146–150	5
151–155	3
156–160	1

3. The bowling scores for the six players on Ali's team are shown at the right.
   a) Determine the mean and standard deviation of the bowling scores for Ali's team, rounded to two decimal places.
   b) Using the mean and standard deviation, compare Ali's data from question 2 to the team's data.

4. Marie, a Métis beadwork artist, ordered packages of beads from two online companies. She is weighing the packages because the sizes seem inconsistent. The standard deviation of the masses of the packages from company A is 11.7 g. The standard deviation of the masses of the packages from company B is 18.2 g.
   a) What does this information tell you about the dispersion of the masses of the packages from each company?
   b) Marie is working on an important project. She needs to make sure that her next order will contain enough beads to complete the project. Should she order from company A or company B?

The Métis are known for floral beadwork. The symmetrical traditional beadwork is illustrated here on the deerskin coat of Louis Riel. Seeds were used to create the beads.

PRACTISING

5. Four groups of students recorded their pulse rates, as given below.

Group 1	63	78	79	75	73	72	62	75	63	77	77	65	70	69	80
Group 2	72	66	73	80	74	75	64	68	67	70	70	69	69	74	74
Group 3	68	75	78	73	75	68	71	78	65	67	63	69	59	68	79
Group 4	78	75	76	76	79	78	78	76	74	81	78	76	79	74	76

Determine the mean and standard deviation for each group, to one decimal place. Which group has the lowest mean pulse rate? Which group has the most consistent pulse rate?

6. Nazra and Diko are laying patio stones. Their supervisor records how many stones they lay each hour.

Hour	1	2	3	4	5	6
Nazra	34	41	40	38	38	45
Diko	51	28	36	44	41	46

a) Which worker lays more stones during the day?
b) Which worker is more consistent?

Milt Stegall

7. Former Winnipeg Blue Bomber Milt Stegall broke several Canadian Football League (CFL) records, including the most touchdowns (TDs) in a season and the most TDs in a career.

Year	'95	'96	'97	'98	'99	'00	'01	'02	'03	'04	'05	'06	'07	'08
TDs	4	6	14	7	7	15	14	23	15	7	17	7	8	3

a) Determine the mean and standard deviation of the TDs that Milt scored in the years he played, to one decimal place.
b) Why do you think that his first and last years had the lowest number of TDs?
c) Determine the mean and standard deviation, to one decimal place, for the years 1996 to 2007.
d) Compare your results from parts a) and c). What do you notice?

8. Milt Stegall also broke the CFL record for most yards receiving.
a) Determine the mean and standard deviation of his statistics, to one decimal place.

Year	'95	'96	'97	'98	'99	'00	'01	'02	'03	'04	'05	'06	'07	'08
Yards	469	613	1616	403	1193	1499	1214	1862	1144	1121	1184	1252	1108	470

b) Allen Pitts, who played for the Calgary Stampeders, held the CFL record for most yards receiving until Milt Stegall surpassed him. Pitts had the following statistics for yards gained per year: mean 1353.7 and standard deviation 357.1. Which player was more consistent in terms of yards gained per year?

9. Two health clubs monitor the number of hours per month that a random sample of their members spend working out.

Fitness Express	
Hours	**Frequency**
8–10	9
10–12	18
12–14	23
14–16	32
16–18	39
18–20	42
20–22	31
22–24	22
24–26	16
26–28	11
28–30	7

Fit for Life	
Hours	**Frequency**
6–9	8
9–12	13
12–15	32
15–18	47
18–21	52
21–24	42
24–27	27
27–30	19

a) Determine the mean and standard deviation of the hours per month for members of each club, to one decimal place.

b) The health clubs believe that workout consistency is more important than workout length. Which club is more successful at encouraging its members to work out consistently?

10. Jaime has 20 min to get to her after-school job. Despite her best efforts, she is frequently late. Her employer says that unless she arrives to work on time consistently, she will lose her job. She has recorded her travel times (in minutes) for the last two weeks: 18, 20, 22, 27, 16, 23, 25, 26, 19, 28. Over the next two weeks, she continues to record her travel times: 22, 20, 19, 16, 20, 23, 25, 18, 19, 17. Do you think Jaime will lose her job? Use statistics to justify your answer.

11. The manager of a customer support line currently has 200 unionized employees. Their contract states that the mean number of calls that an employee should handle per day is 45, with a maximum standard deviation of 6 calls. The manager tracked the number of calls that each employees handles. Does the manager need to hire more employees if the calls continue in this pattern?

Daily Calls	Frequency
26–30	2
31–35	13
36–40	42
41–45	53
46–50	42
51–55	36
56–60	8
61–65	4

Jordin Tootoo, from Rankin Inlet, Nunavut, played for the Brandon Wheat Kings of the WHL from 1999 to 2003. In 2003, he became the first player of Inuit descent to play in a regular-season NHL game.

12. The following table shows Jordin Tootoo's regular season statistics while playing in the Western Hockey League (WHL) and the NHL from 1999 to 2010.

Season	Games Played	Goals	Assists	Points
1999–2000	45	6	10	16
2000–2001	60	20	28	48
2001–2002	64	32	39	71
2002–2003	51	35	39	74
2003–2004	70	4	4	8
2005–2006	34	4	6	10
2006–2007	65	3	6	9
2007–2008	63	11	7	18
2008–2009	72	4	12	16
2009–2010	51	6	10	16

a) Determine the mean and standard deviation for each column of data.

b) In the 2005–2006 season, Jordin played only 34 games. As a result, he had fewer opportunities to score points. Predict the effect on the standard deviation of each column if this season is omitted.

c) Determine the mean and standard deviation for each column, excluding the 2005–2006 season.

d) How do your results for part c) compare with your results for part a)?

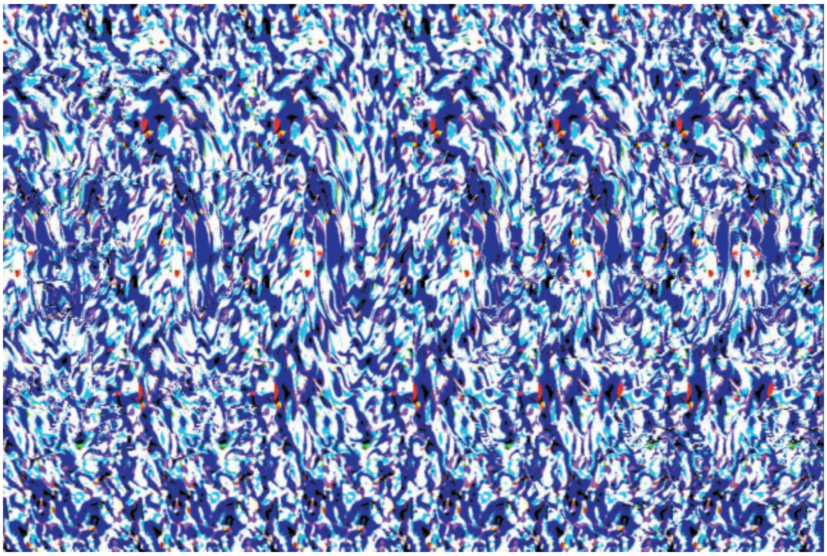
e) Justify the following statement: "Goals + assists = points, whether you are looking at the data for each season or you are looking at the means and standard deviations for many seasons."

Closing

13. Twins Jordana and Jane wrote a total of 10 tests in math class. They have the same mean test score, but different standard deviations. Explain how this is possible.

Extending

14. If you glance at a random dot stereogram (also called a "Magic Eye image"), it looks like a collection of dots or shapes. If you look at it properly, however, it resolves into a 3-D image.



An experiment was done to determine whether people could "fuse" the image faster if they knew what shape they were looking for. The results of the experiment are shown in the table to the right. The people in group A were given no information about the image. The people in group B were given visual information about the image. The number of seconds that the people needed to recognize the 3-D image are listed in the table.

a) Determine the mean and standard deviation for each group, rounded to the nearest hundredth of a second.

b) Were the people who were given visual information able to recognize the image more quickly? Which group was more consistent?

Times for Group A (s)		Times for Group B (s)	
47.2	5.6	19.7	3.6
22.0	4.7	16.2	3.5
20.4	4.7	15.9	3.3
19.7	4.3	15.4	3.3
17.4	4.2	9.7	2.9
14.7	3.9	8.9	2.8
13.4	3.4	8.6	2.7
13.0	3.1	8.6	2.4
12.3	3.1	7.4	2.3
12.2	2.7	6.3	2.0
10.3	2.4	6.1	1.8
9.7	2.3	6.0	1.7
9.7	2.3	6.0	1.7
9.5	2.1	5.9	1.6
9.1	2.1	4.9	1.4
8.9	2.0	4.6	1.2
8.9	1.9	1.0	1.1
8.4	1.7	3.8	
8.1	1.7		
7.9	6.9		
7.8	6.3		
6.1			