

## Reflect

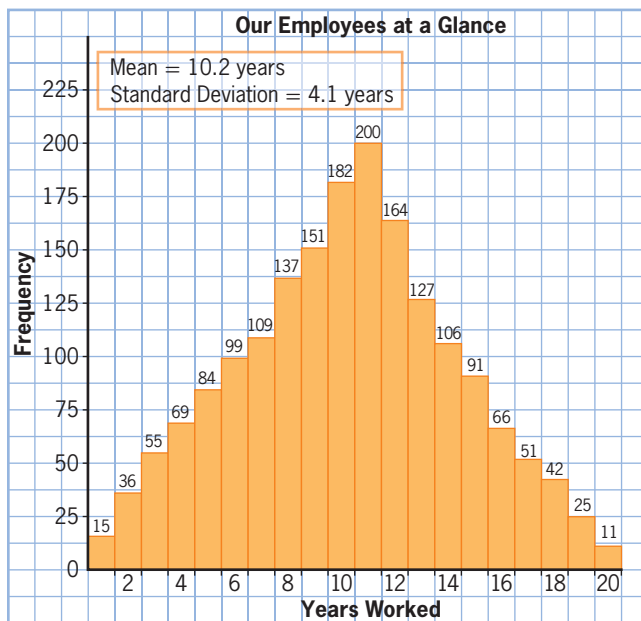
- R1.** The mean of a set of data is 23.5, with standard deviation of 3.1.
- What does a  $z$ -score of  $-2$  mean for a given data point?
  - What does a  $z$ -score of 1.5 mean for a given data point?
- R2.** Before investing in stocks, you read an analysis that includes the standard deviation of its price over a given period of time. Two stocks have the same mean price of \$15.43 over the past 10 days. Stock A has a standard deviation of \$0.56 and stock B has a standard deviation of \$1.22. What does this mean to you as an investor?
- R3.** Explain how  $x$  relates to the mean if the  $z$ -score corresponding to  $x$  is
- positive
  - negative
  - zero
- R4.** Explain how to decide whether the population or sample formulas need to be used for mean and standard deviation.

## Practise

*Choose the best answer for #1 and #2.*

- Adam is building a doorway and wants the height of the door to be three standard deviations above the mean Canadian height. How high must the door be if the mean is 210 cm with a standard deviation of 10 cm?  
**A** 230 cm  
**B** 250 cm  
**C** 200 cm  
**D** 240 cm
- Which is an incorrect statement about standard deviation?  
**A** The variance is the square root of the standard deviation.  
**B** The standard deviation is often called the average distance of the measurements from the mean.  
**C** The standard deviation is expressed in the same units as the data.  
**D** The standard deviation is always a positive quantity.
- The mean of a data set is 25.3 cm, and the standard deviation is 3.6. Determine the  $z$ -score of each of the following and interpret the results.  
**a)** 27.2  
**b)** 24.1  
**c)** 21.9  
**d)** 29.8
- Calculate the standard deviation for each data set and interpret the results.  
**a)** Lengths, in centimetres, of fish caught on a fishing trip.  
15.4 12.3 18.2 9.9  
17.4 12.6 16.3 11.8  
12.3 12.6 16.7  
**b)** Number of home runs in a season by the players on a team.  
3 10 0 12 5 6 10 16 34 11  
6 7 21  
**c)** Final scores by the figure skaters in a competition.  
168.3 178.2 186.1 134.5  
156.7 156.4 167.1 132.0  
154.7 149.8 126.2 134.8  
154.0 175.2 159.2

5. For each of the situations, decide whether you would use the sample or population standard deviation formula. Explain your decisions.
- A researcher recruits females ages 35 to 50 years old for an exercise training study to investigate risk markers for heart disease (e.g., cholesterol).
  - One of the questions on a national survey asks for the respondent's age. Researchers want to describe the variability in all ages received from the survey.
  - A teacher administers a test to her students. The teacher wants to summarize the results the students attained as a mean and standard deviation.
6. As part of a report on its employees, a company published this graph.



- The standard deviation is given as 4.1 years. Identify which numbers of years worked are within one standard deviation of the mean.
- What percent of the employees are within two standard deviations of the mean?
- How does this graph help to explain  $z$ -scores?

## Apply

7. The chart shows the waiting times for customers while having winter tires installed on their cars.

Time (min)	Frequency
30–35	10
35–40	16
40–45	21
45–50	17
50–55	19
55–60	8
60–65	2

When calculating the standard deviation for data in intervals, use the following formulas:

$$\sigma = \sqrt{\frac{\sum(f_i \cdot m_i^2) - N \cdot \mu^2}{N}} \text{ and}$$

$$s = \sqrt{\frac{\sum(f_i \cdot m_i^2) - n \cdot \bar{x}^2}{n - 1}},$$

where  $f$  is the respective frequency of each interval and  $m$  is the interval midpoint.

- Using the midpoints of the intervals as the measurements, estimate the mean and standard deviation of the wait times.
  - Did you use the population or sample formulas? Why?
  - Calculate the  $z$ -scores of each of the midpoints.
  - Draw a histogram and mark the information from part c) on the graph.
8. **Application** The mean size of Canada's 308 electoral districts or ridings is 102 639.28 people, with a standard deviation of 21 855.384. In 2006, Mississauga-Erindale had a population of 143 361. Parkdale-High Park had a population of 102 142.
- Compare the  $z$ -scores for these ridings.
  - What argument could the citizens of Mississauga-Erindale make about their representation in the House of Commons?

9. Maria handed in her final data management project last week. The class mean was 83% with a standard deviation of 8. If Maria's mark produced a  $z$ -score of 1.09, what was her grade?
10. The actual volume of milk in 1-L cartons of milk was checked by measuring a selection of 120 cartons. The chart shows the results.

Volume (L)	Frequency
0.98	6
0.99	18
1.00	30
1.01	35
1.02	19
1.03	9
1.04	0
1.05	3

- a) Calculate the mean and standard deviation, accurate to three decimal places.
- b) Did you use the population or sample formulas? Why?
- c) The company has decided that a sample that is within two standard deviations of the mean is acceptable. A random sample was taken and the volume was 0.98 L. Would this be an acceptable sample?
- d) On the following day, the mean volume of milk per carton was 1.012 L, with a standard deviation of 0.009 L. Compare the two days' test results.
11. The table shows the lengths of logs shipped to a lumber mill on a particular day.

Length (m)	Frequency
3.5–4.5	3
4.5–5.5	20
5.5–6.5	17
6.5–7.5	38
7.5–8.5	31
8.5–9.5	19
9.5–10.5	15

- a) Calculate the mean and standard deviation of the logs, accurate to three decimal places.
- b) How does this data set compare to the previous day, with a mean of 8.44 m and standard deviation of 1.836 m?
- c) Why would the standard deviation be important to the operators of the lumber mill?

### ✓ Achievement Check

12. Along with their application to a particular university, students were instructed to submit a 700-word essay. The table shows the lengths of 16 of the essays that were submitted.

Student	Number of Words in Essay
Alex	709
Christian	743
Maria	810
Hasanika	900
Daniel	1112
Barb	568
Brian	804
Cathie	951
Wayne	643
Shaniqua	829
Jiang	674
Bill	769
Mohammed	781
Farah	735
Tim	700
Guido	583

- a) Calculate the mean, variance, and standard deviation to the nearest whole number for the data set.
- b) Did you use the sample or population formulas? Why?
- c) Make an appropriate graph of the data. Mark the interval that is within one standard deviation of the mean.

- d) Use  $z$ -scores to determine whether Cathie's or Wayne's essay length is closer to the mean.
- e) Compare this group's essays to the essays in the previous year, with a mean of 712.1 words and standard deviation of 23.2 words.
- 13. Communication** When is it possible for the standard deviation to be larger than the variance?
- 14.** After graduating from university, Yee Ping hopes to get a job in a career with a mean starting salary of \$56 000. Compare the salary ranges for standard deviations of \$15 000 and \$5000, knowing that 95% of the starting salaries are within two standard deviations of the mean.
- 15. Communication** When buying an investment such as a mutual fund, investors look at its volatility. Volatility is measured by calculating the standard deviation of the returns over a given period of time.
- a) What will the standard deviation show an investor if the mean rate of return of a particular mutual fund unit is 14.37% with a volatility of 6.54%?
- b) How would the standard deviation change for a riskier investment?
- 16. Thinking** A set of five whole numbers is arranged in order from least to greatest. The fifth number is decreased by one. Would the interquartile range or standard deviation be more affected? Explain.

## Extend

- 17.** The mean of a sample of  $n$  values is  $\bar{x}$  and the standard deviation is  $s$ . Suppose you add a constant value  $a$  to each observation so that the new data values are

$$x_1 + a, x_2 + a, \dots, x_n + a.$$

Determine the new mean and standard deviation.

- 18.** The mean of a sample of  $n$  values is  $\bar{x}$  and the standard deviation is  $s$ . Suppose the observations are multiplied by a constant value  $c$  so that the new data values are

$$cx_1, cx_2, \dots, cx_n.$$

Determine the new mean and standard deviation.

- 19.** Algebraically derive the computational formula

$$\sigma = \sqrt{\frac{\sum x^2 - N \cdot \mu^2}{N}}$$

from the defined standard deviation formula

$$\sigma = \sqrt{\frac{\sum (x - \mu)^2}{N}}.$$