Standard Deviation and z-Scores

Lesson objectives

- I can use technology to calculate the variance and standard deviation of a data set
- I can calculate and understand the significance of a z-score
- I can relate the positive or negative scores to their locations in a histogram
- I can develop significant conclusions about a data set

1.1

Lesson objectives | Teachers' notes | Lesson notes

MHR Page 286 #s 1 - 6, 10 & 11

Day	Thickness	Deviation (x - x bar)	Squared Deviation
1	152	-2.7143	7.3673
2	158	3.2857	10.7959
3	151	-3.7143	13.7959
4	153	-1.7143	2.9388
5	159	4.2857	18.3673
6	158	3.2857	10.7959
7	152	-2.7143	7.3673
Sum		0.0000	71.4286
Mean =	154.7142857	Variance =	10.2041
		Standard Deviation =	3.1944

Variance = mean of the squared deviations

Variance = Sum of Squared deviation ÷ # of data points

Standard deviation = $\sqrt{\text{variance}}$

Definitions

Variance

- The average squared difference of the scores from the mean

Standard Deviation

- The square root of the variance
- The average distance of the scores from the mean

z-Score

- The number of standard deviations an observation is from the mean

The variance and standard deviation of a data set allow you to determine how close the values in a distribution are to the middle of the distribution. You can calculate the variance and standard deviation of a data set using the following formulas.

Population Variance

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

Sample Variance

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

Population Standard Deviation

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

Sample Standard Deviation

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

where the population deviation is represented by $(x-\mu)$ and the sample deviation is represented by $(x-\overline{x})$.

Samples rarely contain extreme values when compared to entire populations. As a result, the variance and standard deviation are less than would be expected. To use the sample variance and standard deviation to model a population, divide by $n-1$ instead of n. This slightly increases their values.

Calculating the variance alone is not a perfect measure of spread. First, because the deviations of each value are squared in the formula, more "weight" is given to extreme values. Therefore, data sets with extreme values or outliers will skew the validity of the result. Second, the variance is calculated in units squared, which is not the same units as the scores in the data set. This means that you cannot show the variance on a frequency distribution and cannot make a direct correlation between its value and the values of your data set. This problem is easily corrected by calculating the standard deviation.

Example 1

Visualizing the Spread of Marks

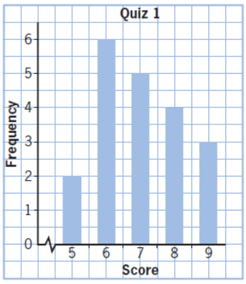
The graphs represent the scores on two quizzes. The mean score for each quiz is 7.0.

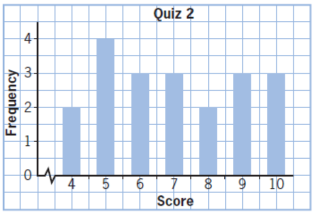
a) Which quiz would have a greater standard deviation? Why?

b) The variance of Quiz 1 is 1.5. What is the standard deviation?

c) What would the Quiz 1 graph look like if the standard deviation were 1.6?

d) What would the graph look like if the standard deviation were 0?





a) Despite the *frequencies* being closer in value for Quiz 2, there is less spread in the *scores* on Quiz 1. This will lead to a greater standard deviation on Quiz 2 because the data is more spread out.

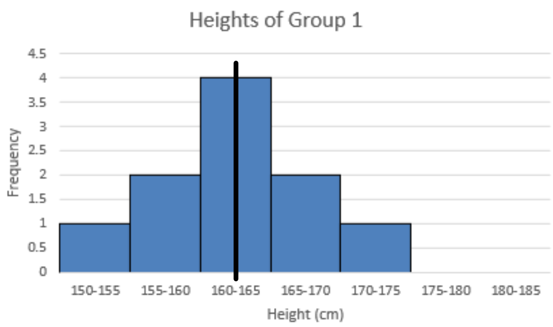
b) Standard deviation = √variance. Therefore $\sqrt{1.5} \approx 1.22$

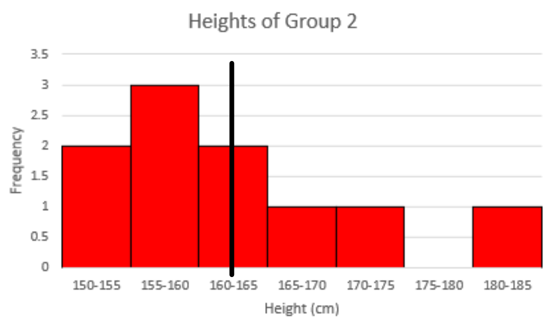
c) With a bigger standard deviation we would expect there to be a greater spread in the scores. It would look more like Quiz 2.

d) With a standard deviation of zero, that means there would be no spread in the marks: everyone would have scored the same.

Your Turn

Sketch examples of two histograms that show the distribution of two sets of girls' heights with the same mean but with different standard deviations. Indicate which histogram will have a greater standard deviation.





Both groups have a mean of 162.5. The data is more spread out for Group 2, so it will have a higher standard deviation.

Example 2

Calculating Variance and Standard Deviation

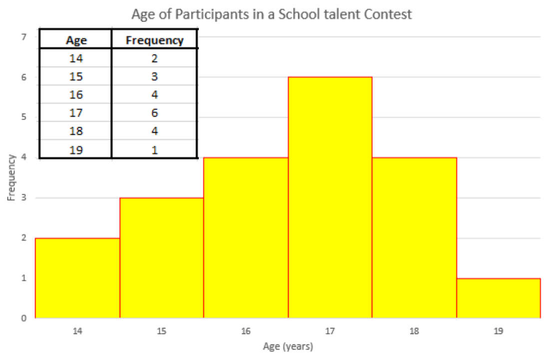
The ages of participants in a school's talent contest are listed below.

16 17 18 16 15 16 17 15 18 14
17 19 18 16 17 17 17 14 15 18

Use technology to answer the questions.

a) Plot a histogram of the data.
b) Calculate the mean and standard deviation.
c) What would happen to the standard deviation if the first person's age were 18?
d) What would happen to the standard deviation if the second person's age were 16 instead of 17?
e) What would happen to the standard deviation if each person were one year older?
f) Which ages are more than one standard deviation from the mean?

a)



b)

	A	B	C	D	E	F	G	H	I	J
246										
247										
248										
249	16	17	18	16	15	16	17	15	18	14
250	17	19	18	16	17	17	17	14	15	18
251										
252		Mean =	16.5		=AVERAGE(A249:J250)					
253										
254		SD =	1.3601471		=STDEV.P(A249:J250)					

Cells that have the data

c)

18	17	18	16	15	16	17	15	18	14
17	19	18	16	17	17	17	14	15	18
	Mean =	16.6		=AVERAGE(A249:J250)					
	SD =	1.39284		=STDEV.P(A249:J250)					

SD would increase as it is further from the mean than the original value of 16.

d)

16	16	18	16	15	16	17	15	18	14
17	19	18	16	17	17	17	14	15	18
	Mean =	16.45		=AVERAGE(A249:J250)					
	SD =	1.35923		=STDEV.P(A249:J250)					

SD DECREASES because 16 is closer to the mean that 17 is.

e)

17	18	19	17	16	17	18	16	19	15
18	20	19	17	18	18	18	15	16	19
	Mean =	17.5		=AVERAGE(A249:J250)					
	SD =	1.36015		=STDEV.P(A249:J250)					

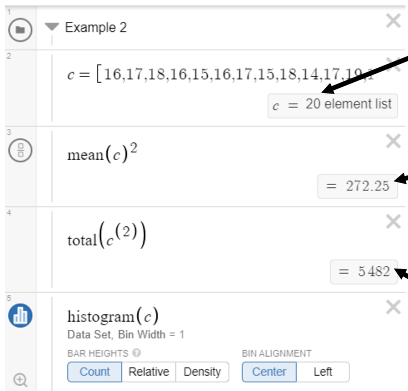
SD would be THE SAME. Mean would increase by one, but the deviation won't change.

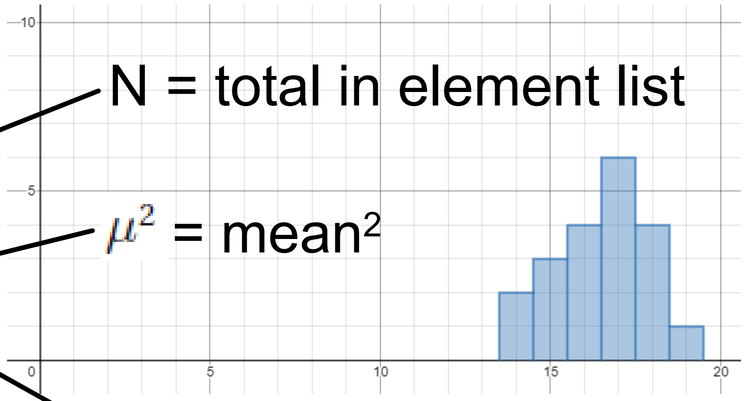
f) The limits are the mean ± 1 standard deviation.

Lower limit = 16.5 - 1.36 = 15.14 & Upper limit = 16.5 + 1.36 = 17.86

So, ages greater than 17 and less then 15 would be more than one standard deviation from the mean.

Using desmos





N = total in element list

μ^2 = mean²

$\sum x^2$ = total(c²)

Population standard deviation:

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

If you call your list by another variable, use that letter instead of *c*.

NOTE: you can't use *e* as a variable.

$$\sigma = \sqrt{\frac{5482 - 20(272.25)}{20}}$$

σ = √(37 ÷ 20)

σ = √1.85 ⟶ σ = 1.36047051

You need to know how to calculate standard deviation using this method.

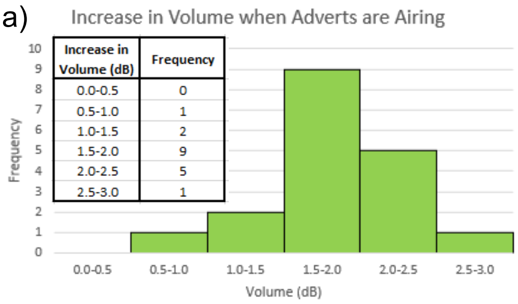
Your Turn

The increases in sound volume from a TV program to the advertisements were measured in decibels during a one-hour TV show. The results were as follows:

1.7 1.9 1.5 2.0 2.1 1.8 2.2 1.9 2.0
1.4 1.7 1.8 1.8 2.1 2.7 1.0 0.6 1.8

a) Plot a histogram of the data.
b) Calculate the population mean and standard deviation.
c) Predict what would happen to the standard deviation if the first measurement were 1.5 dB.
d) Predict what would happen to the standard deviation if the second measurement were 1.7 dB.
e) What would happen to the standard deviation if each measurement were 0.5 dB quieter?
f) Which measurements are within one standard deviation of the mean?

a)



Increase in Volume (dB)	Frequency
0.0-0.5	0
0.5-1.0	1
1.0-1.5	2
1.5-2.0	9
2.0-2.5	5
2.5-3.0	1

b)

1.7	1.9	1.5	2	2.1	1.8	2.2	1.9	2	
1.4	1.7	1.8	1.8	2.1	2.7	1.0	0.6	1.8	
	Mean =	1.77778		=AVERAGE(A249:J250)					
	SD =	0.44666		=STDEV.P(A249:J250)					

Cells that have the data

c) Standard deviation would INCREASE because the new value is further away from the mean.

d) Standard deviation would DECREASE because the new value is closer to the mean.

e) Standard deviation would be THE SAME. The mean would decrease by 0.5dB, but the deviation would be the same.

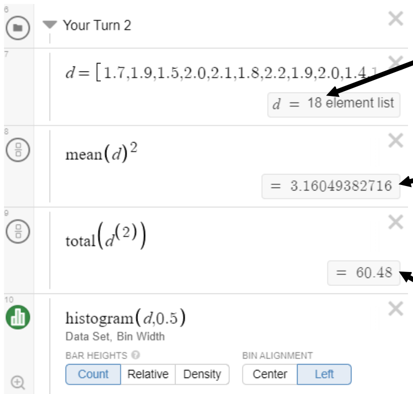
f) The limits are the mean ± 1 standard deviation.

Lower limit = 1.78 - 0.45 = 1.33 & Upper limit = 1.78 + 0.45 = 2.23

So, volumes greater than 2.23 dB and less then 1.33 dB would be more than one standard deviation from the mean.

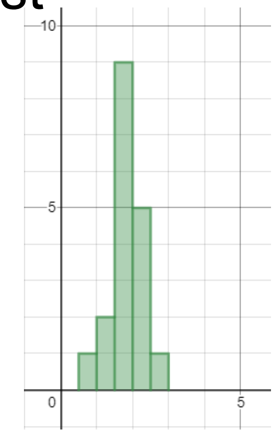
Using desmos

N = total in element list



μ^2 = mean^2

$\sum x^2$ = total(d^2)



Population standard deviation:

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

If you call your list by another variable, use that letter instead of *d*.

NOTE: you can't use *e* as a variable.

$$\sigma = \sqrt{\frac{60.48 - 18(3.16049)}{18}}$$

σ = √(3.59118 ÷ 18)

σ = √0.19951 ⟶ σ = 0.4466654229

You need to know how to calculate standard deviation using this method.

A *z*-score indicates how many standard deviations a data value lies from the mean. In Example 2, part f), the z-score would be 1. You can calculate a z-score using one of the following formulas:

Population *z*-Score	Sample *z*-Score
$z = \frac{x - \mu}{\sigma}$	$z = \frac{x - \overline{x}}{s}$

You can derive computational standard deviation formulas from the given formulas. These formulas simplify the calculations of standard deviation using a scientific calculator.

Population Standard Deviation	Sample Standard Deviation
$\sigma = \sqrt{\frac{\sum x^2 - N \cdot \mu^2}{N}}$	$s = \sqrt{\frac{\sum x^2 - n \cdot \overline{x}^2}{n - 1}}$

Example 3

Analysing *z*-Scores

A food manufacturer makes 2-L jars of pasta sauce. Samples are tested for how close to 2 L the jars are filled. Fifteen samples were taken and their volumes, in litres, were as indicated:

2.11 2.02 2.10 1.99 1.92 2.01 1.89 1.96
2.00 1.96 1.98 2.02 2.08 2.15 2.03

a) Determine the sample mean and standard deviation.
b) Calculate the z-score of the jar that was filled to a volume of 2.02 L. Interpret its meaning.
c) Calculate the z-score of the jar that was filled to a volume of 1.98 L. Compare its distance from the mean to that of 2.02 L.
d) The manufacturer rejects any jars that are filled to less than 1.5 standard deviations below the mean. Which jars would be rejected?

a) s = √[(60.955 - 15(2.01467²)) / 14]

s = √[(60.955 - 60.883226) / 14]

s = √[0.071773736 / 14]

s = √0.0051126695

s = 0.071600945

s = 0.0716

desmos

$a = [2.11, 2.02, 2.10, 1.99, 1.92, 2.01, 1.89, 1.9$

a = 15 element list

mean(a)

= 2.01466666667

mean$(a)^2$

= 4.05888177778

total$(a^{(2)})$

= 60.955

b) $$z = \frac{x - \overline{x}}{s} = \frac{2.02 - 2.014\,67}{0.0716} \approx 0.0744$$

A z-score of about 0.0744 tells us that the volume is more than the mean. However, it is very close to the mean.

c) $$z = \frac{x - \overline{x}}{s} = \frac{1.98 - 2.014\,67}{0.0716} \approx -0.4842$$

A z-score of about -0.4842 tells us that a volume of 1.98 L is less than the mean. It is further from the mean than the 2.02 L jar.

d) $$\overline{x} - 1.5s = 2.014\,67 - 1.5(0.0716) = 1.9073$$

Any jar less than 1.9073 L would be rejected. In this case, the 1.89 L jar would have to go.

Example 3

Analysing *z*-Scores

A food manufacturer makes 2-L jars of pasta sauce. Samples are tested for how close to 2 L the jars are filled. Fifteen samples were taken and their volumes, in litres, were as indicated:

2.11 2.02 2.10 1.99 1.92 2.01 1.89 1.96
2.00 1.96 1.98 2.02 2.08 2.15 2.03

a) Determine the sample mean and standard deviation.
b) Calculate the z-score of the jar that was filled to a volume of 2.02 L. Interpret its meaning.
c) Calculate the z-score of the jar that was filled to a volume of 1.98 L. Compare its distance from the mean to that of 2.02 L.
d) The manufacturer rejects any jars that are filled to less than 1.5 standard deviations below the mean. Which jars would be rejected?

Excel

a)

2.11	2.02	2.10	1.99	1.92	2.01	1.89	1.96
2.00	1.96	1.98	2.02	2.08	2.15	2.03	
	Mean = 2.01467			=AVERAGE(A249:J250)			
	SD = 0.0716			=STDEV.S(A249:J250)			

As we want the SAMPLE standard deviation, the formula changes from STDEV.P to **STDEV.S**

b) $$z = \frac{x - \overline{x}}{s} = \frac{2.02 - 2.014\,67}{0.0716} \approx 0.0744$$

A z-score of about 0.0744 tells us that the volume is more than the mean. However, it is very close to the mean.

c) $$z = \frac{x - \overline{x}}{s} = \frac{1.98 - 2.014\,67}{0.0716} \approx -0.4842$$

A z-score of about -0.4842 tells us that a volume of 1.98 L is less than the mean. It is further from the mean than the 2.02 L jar.

d) $$\overline{x} - 1.5s = 2.014\,67 - 1.5(0.0716) = 1.9073$$

Any jar less than 1.9073 L would be rejected. In this case, the 1.89 L jar would have to go.

Your Turn

A car manufacturer tested the gap between the doors and the body of a car. Eighteen samples were taken. Their gaps, in millimetres, are shown:

1.7 1.9 1.4 1.4 1.5 1.7 1.1 1.6 1.9
1.4 1.5 1.5 1.6 1.5 1.3 1.8 1.6 1.2

a) Determine the sample mean and standard deviation.
b) Calculate the z-score of a door with gap of 1.6 mm. Interpret its meaning.
c) Calculate the z-score of a door with gap of 1.4 mm. Compare its distance from the mean to that of 1.6 mm.
d) The manufacturer rejects any cars with door gaps that are not within two standard deviations of the mean. Which cars would be rejected?

a) s = √[(43.14 - 18(1.53333²)) / 17]

s = √[(43.14 - 42.319999) / 17]

s = √[0.8200000 / 17]

s = √0.048235295

s = 0.219625352

s = 0.2196

desmos

$b = [1.7,1.9,1.4,1.4,1.5,1.7,1.1,1.6,1.9,1.4,1$
b = 18 element list

mean(b)
= 1.53333333333

mean(b)^2
= 2.35111111111

total($b^{(2)}$)
= 43.14

b) $$z = \frac{x - \bar{x}}{s} = \frac{1.6 - 1.53}{0.2196} \approx 0.3188$$

A gap of 1.6 mm would be about 0.3188 standard deviations above the mean.

c) $$z = \frac{x - \bar{x}}{s} = \frac{1.4 - 1.53}{0.2196} \approx -0.5920$$

A gap of 1.4 mm would be about 0.5920 standard deviations below the mean and be further from it compared to a gap of 1.6 mm.

d) $$\bar{x} - 2s = 1.53 - 2(0.2196) = 1.0908$$

$$\bar{x} + 2s = 1.53 + 2(0.2196) = 1.9692$$

Any cars with door gaps less than 1.0908 mm or more than 1.9692 mm would be rejected. So, they're all fine in this sample.

Your Turn

A car manufacturer tested the gap between the doors and the body of a car. Eighteen samples were taken. Their gaps, in millimetres, are shown:

1.7 1.9 1.4 1.4 1.5 1.7 1.1 1.6 1.9
1.4 1.5 1.5 1.6 1.5 1.3 1.8 1.6 1.2

a) Determine the sample mean and standard deviation.
b) Calculate the z-score of a door with gap of 1.6 mm. Interpret its meaning.
c) Calculate the z-score of a door with gap of 1.4 mm. Compare its distance from the mean to that of 1.6 mm.
d) The manufacturer rejects any cars with door gaps that are not within two standard deviations of the mean. Which cars would be rejected?

Excel

a)

1.7	1.9	1.4	1.4	1.5	1.7	1.1	1.6	1.9
1.4	1.5	1.5	1.6	1.5	1.3	1.8	1.6	1.2
	Mean =	1.53333		=AVERAGE(A249:J250)				
	SD =	0.21963		=STDEV.S(A249:J250)				

Again, as we want the SAMPLE standard deviation, the formula changes from STDEV.P to **STDEV.S**

b) $$z = \frac{x - \bar{x}}{s} = \frac{1.6 - 1.53}{0.2196} \approx 0.3188$$

A gap of 1.6 mm would be about 0.3188 standard deviations above the mean.

c) $$z = \frac{x - \bar{x}}{s} = \frac{1.4 - 1.53}{0.2196} \approx -0.5920$$

A gap of 1.4 mm would be about 0.5920 standard deviations below the mean and be further from it compared to a gap of 1.6 mm.

d) $$\bar{x} - 2s = 1.53 - 2(0.2196) = 1.0908$$

$$\bar{x} + 2s = 1.53 + 2(0.2196) = 1.9692$$

Any cars with door gaps less than 1.0908 mm or more than 1.9692 mm would be rejected. So, they're all fine in this sample.

Key Concepts

- The variance and standard deviation are measures of spread. The standard deviation is the square root of the variance.

Population variance:

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

Population standard deviation:

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

Sample variance:

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

Sample standard deviation:

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

- You can use the following computational formulas to calculate standard deviation more easily.

Population standard deviation:

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

Sample standard deviation:

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

- The standard deviation of a set of data determines the average distance of the measurements from the mean. The larger the value, the greater the spread of the data. The units of the standard deviation are the same as for the mean.
- The z-score tells you the number of standard deviations that an observation in a data set is from the mean.

Population z-score: $z = \frac{x-\mu}{\sigma}$ Sample z-score: $z = \frac{x-\overline{x}}{s}$

R1. The mean of a set of data is 23.5, with standard deviation of 3.1.

a) What does a z-score of –2 mean for a given data point?

b) What does a z-score of 1.5 mean for a given data point?

a)

$$z = \frac{x-\overline{x}}{s}$$

$$-2 = \frac{x-23.5}{3.1}$$

$$-6.2 = x - 23.5$$

$$17.3 = x$$

It means that the data point is 2 standard deviations below the mean. It would have a value of 17.3

b)

$$z = \frac{x-\overline{x}}{s}$$

$$1.5 = \frac{x-23.5}{3.1}$$

$$4.65 = x - 23.5$$

$$28.15 = x$$

It means that the data point is 1.5 standard deviations above the mean. It would have a value of 28.15

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?

Since stock A has a smaller standard deviation, it's price will not fluctuate as much... it's more consistent than stock B. Cautious investors should go with stock A, but more adventurous investors should go with stock B as it has potential for more profit (but also more loss).

R3. Explain how x relates to the mean if the z-score corresponding to x is

a) positive

b) negative

c) zero

a) A positive z-score tells us that x is greater than the mean.

b) A negative z-score tells us that x is less than the mean.

c) A z-score of zero tells us that x is equal to the mean.

R4. Explain how to decide whether the population or sample formulas need to be used for mean and standard deviation.

You should use population formulas when you are using all values from the population. Use the sample formulas when you are using only a sample of the population.