Unit 6: Standard Deviation

PREREQUISITES

Students should be familiar with graphic displays such as stemplots, histograms, and boxplots, which are covered in Units 2, 3, and 5, respectively. From these units, students should be familiar with the five-number summary and measures of spread related to the five-number summary, namely the interquartile range (IQR) and range. They should be able to identify whether or not the shape of a distribution is symmetric and to identify potential outliers from a graphic display. In addition, students need to be able to compute the mean of a set of data, which is covered in Unit 4, Measures of Center.

ADDITIONAL TOPIC COVERAGE

Additional coverage on standard deviation can be found in *The Basic Practice of Statistics*, Chapter 2, Describing Distributions with Numbers.

ACTIVITY DESCRIPTION

The purpose of this activity is to help students visualize the spread of the data by focusing on the concept of deviations from the mean. Students can work on this activity either individually or in groups.

In questions 1 and 2, students make dotplots of the data and then draw horizontal bars that represent deviations from the mean. Based on the lengths of the horizontal bars for the data sets in question 2, it should be obvious which data set has the larger standard deviation.

In questions 3 and 4, students are given histograms of five data sets, all of which have the same mean. In question 3, they compare two data sets at a time and determine which has the larger standard deviation. In one case, the two data sets have histograms that are mirror images of each other about a vertical line at the mean, and hence, these data sets have the same standard deviation. In question 4, students are given the actual standard deviations of the five data sets and are asked to match them to the histograms.

THE VIDEO SOLUTIONS

1. The mean precipitation rates for the two cities were very close – Portland had a mean of 3.32 inches/month and Montreal had a mean of 3.4 inches/month. What differed between the two cities was the variability in the precipitation patterns. Montreal's precipitation rate was fairly constant from month to month; Portland's precipitation was heavy during winter months and light during summer months.

2. The sum of the deviations of individual data values from their mean, $\sum\left(x - \bar{x}\right)$, is always exactly 0.

3. The monthly sales data from the Manhattan Beach location is more variable than the sales data from the South Coast Plaza location.

4. No, standard deviation is always positive or 0. When you square deviations from the mean, they become positive or zero. Their sum is still positive or zero and the quotient after dividing the sum by n – 1 stays positive or zero. This final quantity is the variance. To get the standard deviation, take the square root of the variance, which gives a number greater than or equal to zero.

UNIT ACTIVITY SOLUTIONS

1. a. 1, 1, -3, 3, -2

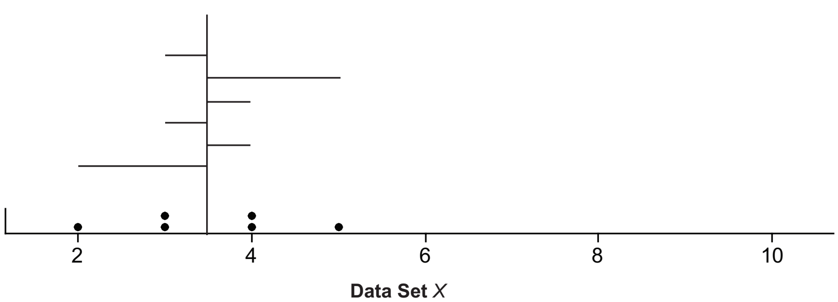
b. The sum is zero.

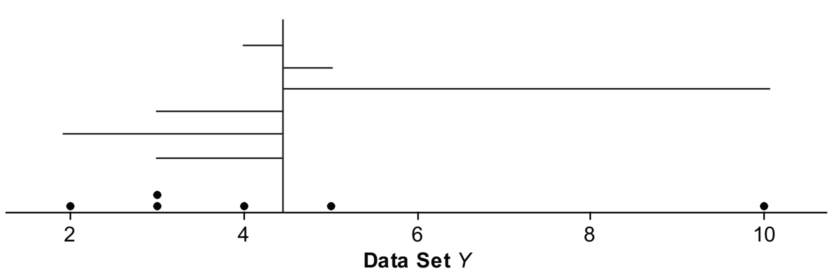
c. 1 + 1 + 9 + 9 + 4 = 24

d. $s^2 = 24/4 = 6$; $s = \sqrt{6} \approx 2.45$

2. a. Data Set *X*: $\bar{x} = 3.5$; Data Set *Y*: $\bar{y} = 4.5$

b.





c. The line segments representing the deviations from the mean tend to be longer for Data Set Y than for Data Set X. Since standard deviation is based on the deviations from the mean, Data Set Y will have the larger standard deviation.

d. For Data Set X: $s^2 = \frac{5.5}{5} = 1.1$; $s = \sqrt{1.1} \approx 1.05$

For Data Set Y: $s^2 = \frac{41.5}{5} = 8.3$; $s = \sqrt{8.3} \approx 2.88$

As was predicted in (c), the standard deviation for Data Set Y is larger than for Data Set X.

3. a. Data Set B has the larger standard deviation.

Sample answer: Data Set A is more concentrated around the mean of 2.5 and has its highest concentration of data between 2 and 3. Data Set B has data evenly spread from 0 to 5. So, there is a higher concentration of data in class intervals 0 to 1 and 4 to 5 for Data Set B than for Data Set A; these are the class intervals that are farthest from the mean. Therefore, Data Set B has the larger standard deviation.

b. Data Sets C and D have the same standard deviation.

Sample answer: The histograms are mirror images of each other about the vertical line at the mean, 2.5. So, the spread about the mean is the same for both data sets, just in opposite directions.

c. Data Set E has the larger standard deviation.

Sample answer: Data Set E has its highest concentration of data between class intervals 0 to 1 and 4 to 5, the class intervals that are farthest from the mean. A high proportion of the data from Data Set D is concentrated from 1 to 3, close to the mean of 2.5. Therefore, Data Set E is more spread out than Data Set D.

4. Standard deviations: $s_A = 1.124$, $s_B = 1.451$, $s_C = s_D = 1.026$, $s_E = 1.589$

EXERCISE SOLUTIONS

1. (100)2 = 10,000

2. a. Ninth-grade students:

$$\bar{x} = \frac{5+1+2+5+3+8}{6} = 4$$

$$s^2 = \frac{(5-4)^2+(1-4)^2+(2-4)^2+(5-4)^2+(3-4)^2+(8-4)^2}{6-1} = 6.4$$

$$s = \sqrt{6.4} \approx 2.53$$

12th-grade students:

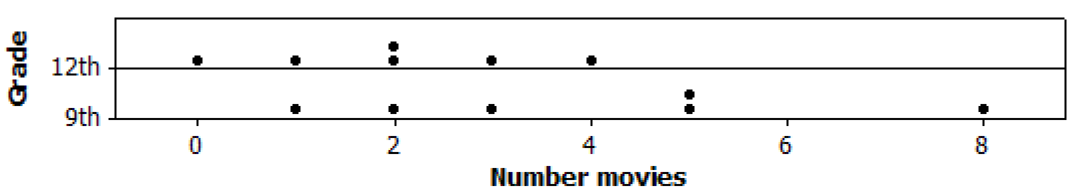
$$\bar{x} = \frac{4+2+0+2+3+1}{6} = 2$$

$$s^2 = \frac{(4-2)^2+(2-2)^2+(0-2)^2+(2-2)^2+(3-2)^2+(1-2)^2}{5} = 2$$

$$s = \sqrt{2} \approx 1.41$$

The data for the ninth-graders is more spread out.

b. The ninth-graders' data appears shifted to the right and more spread out compared to the 12th-graders' data.



3. a. $\bar{x} = \frac{7+3+4+7+5+10}{6} = 6$

$$s = \sqrt{\frac{(7-6)^2+(3-6)^2+(4-6)^2+(7-6)^2+(5-6)^2+(10-6)^2}{6-1}} = \sqrt{\frac{32}{5}} \approx 2.53$$

b. The mean increased by 2, the same amount that was added to the ninth-graders' data. The standard deviation stayed the same.

c. If you add 10 to each of the numbers in the ninth-grade data set, the mean will increase by 10. However, the standard deviation will still be 2.53.

4. a. Critical reading: s = 40.84; mathematics: s = 41.84; writing: s = 39.85

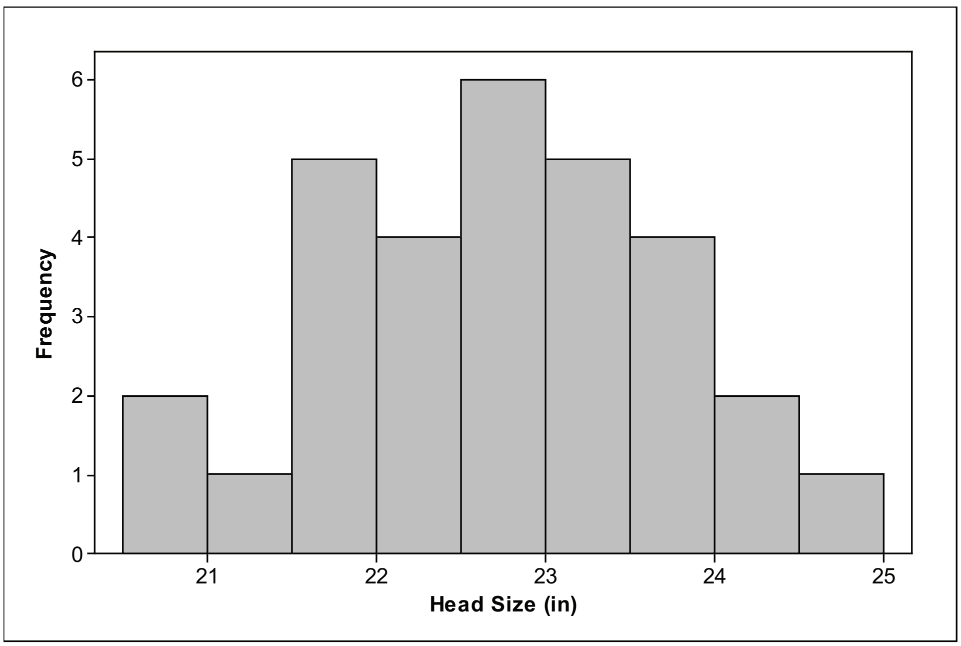
b. The mathematics SAT scores are the most spread out of the three exams.

REVIEW QUESTIONS SOLUTIONS

1. a. No. These numerical summaries help describe specific aspects of the distribution, especially center and spread. But they do not describe the exact shape of the distribution.

b. The answer is again No. In the case of the five-number summary, this is easier to see. The observations between, say, the third quartile and the maximum observation are free to move anywhere in that interval without changing the five-number summary.

2. a. From the histogram below, the shape of the histogram appears to be mound-shaped and roughly symmetric. Hence, this is a good distribution to summarize using $\bar{x}$ and s to measure center and spread, respectively.



b. Using technology, $\bar{x} \approx 22.647$ inches and $s \approx 1.026$ inches.

c. $\bar{x} - s \approx 22.647 - 1.026 = 21.621$; $\bar{x} + s \approx 22.647 + 1.026 = 23.673$

Using technology to sort the data from smallest to largest gives the following ordered list of data values. The data values that fall within one standard deviation of the mean have been shaded.

20.8	20.8	21.0	21.5	21.5	21.7	21.8	21.9	22.0	22.2	22.3
22.4	22.5	22.6	22.6	22.7	22.7	22.7	23.0	23.0	23.1	23.3
23.4	23.5	23.5	23.9	23.9	24.0	24.2	24.9			

20/30 × 100% or around 66.7% of the data fall within one standard deviation of the mean.

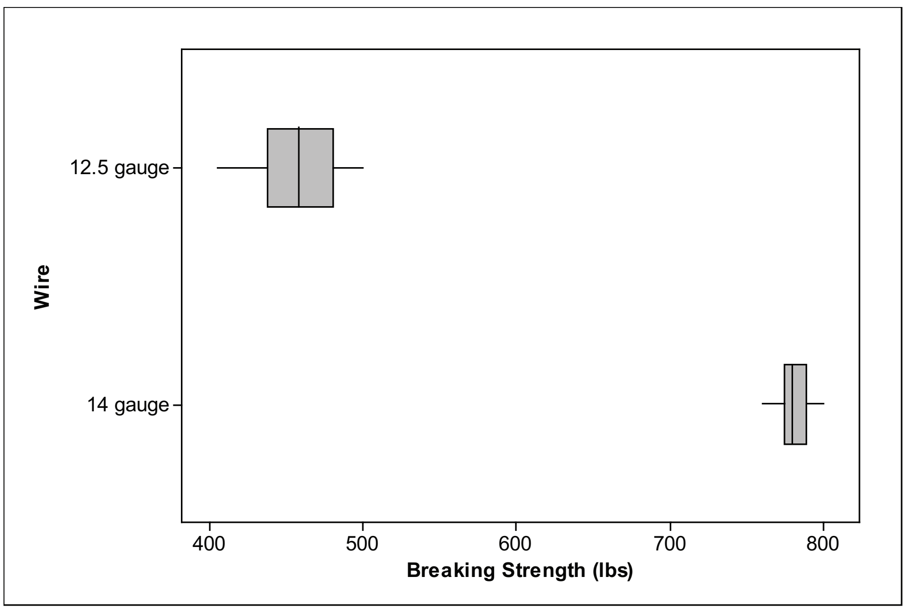
3. a.

58.420	56.388	55.118	55.880	56.642	57.404
57.658	54.610	57.658	63.246	52.832	59.182
61.468	59.690	60.706	59.436	52.832	54.610
58.420	60.960	57.658	57.404	60.706	55.372
58.674	55.626	53.340	56.896	59.690	57.150

b. $s \approx 2.606$

c. The standard deviation in (b) is 2.54 times the standard deviation of the data in 2(b): (1.026)(2.54) ≈ 2.606

4. a. Sample answer: The two boxplots appear roughly symmetric.



b. The mean breaking strength for the 12.5-gauge, low-carbon wire is 458 lbs. The mean breaking strength for the 14-gauge, high-tensile wire is 780.75 lbs. The 14-gauge wire has the larger mean breaking strength.

c. 12.5-gauge wire: s = 26.48 lbs; 14-gauge wire: s = 9.90 lbs. The 12.5-gauge wire's data is more variable than the 14-gauge wire's data.