Standard Deviation

<u>Standard deviation</u> - is the average distance the numbers are from the mean.

How Many Pets

1 3 4 4 4 5 7 8 9

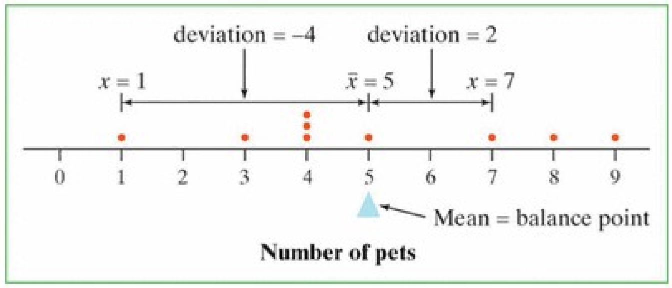


Figure 1.20 Dotplot of the pet data with the mean and two of the deviations marked.

Observations x_i	Deviations $x_i - \bar{x}$	Squared deviations $(x_i - \bar{x})^2$
1	$1 - 5 = -4$	$(-4)^2 = 16$
3	$3 - 5 = -2$	$(-2)^2 = 4$
4	$4 - 5 = -1$	$(-1)^2 = 1$
4	$4 - 5 = -1$	$(-1)^2 = 1$
4	$4 - 5 = -1$	$(-1)^2 = 1$
5	$5 - 5 = 0$	$0^2 = 0$
7	$7 - 5 = 2$	$2^2 = 4$
8	$8 - 5 = 3$	$3^2 = 9$
9	$9 - 5 = 4$	$4^2 = 16$
	sum = 0	sum = 52

Instead of dividing by the number of observations n, we divide by $n - 1$:

$$\text{“average” squared deviation} = \frac{16+4+1+1+1+0+4+9+16}{9-1} = \frac{52}{8} = 6.5$$

This value, 6.5, is called the **variance.**

Because we squared all the deviations, our units are in "squared pets." That's no good. We'll take the square root to get back to the correct units—pets. The resulting value is the **standard deviation:**

$$\text{standard deviation} = \sqrt{\text{variance}} = \sqrt{6.5} = 2.55 \text{ pets}$$

This 2.55 is the "typical" distance of the values in the data set from the mean. In this case, the number of pets typically varies from the mean by about 2.55 pets.

DEFINITION: The standard deviation s_x and variance s_x^2

The **standard deviation** s_x measures the typical distance of the values in a distribution from the mean. It is calculated by finding an average of the squared deviations and then taking the square root. This average squared deviation is called the **variance.** In symbols, the variance s_x^2 is given by

$$s_x^2 = \frac{(x_1 - \bar{x})^2 + (x_2 - \bar{x})^2 + \cdots + (x_n - \bar{x})^2}{n-1} = \frac{1}{n-1}\sum(x_i - \bar{x})^2$$

and the standard deviation is given by

$$s_x = \sqrt{\frac{1}{n-1}\sum(x_i - \bar{x})^2}$$

Here's a brief summary of the process for calculating the standard deviation.

HOW TO FIND THE STANDARD DEVIATION

To find the standard deviation of n observations:

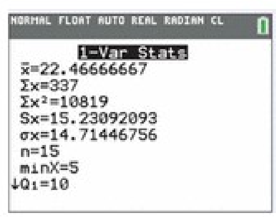
1. Find the distance of each observation from the mean and square each of these distances.
2. Average the distances by dividing their sum by $n - 1$.
3. The standard deviation s_x is the square root of this average squared distance:

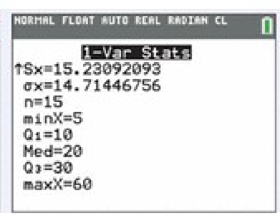
$$s_x = \sqrt{\frac{1}{n-1}\sum(x_i - \bar{x})^2}$$

Many calculators report two standard deviations. One is usually labeled σ_x, the symbol for the standard deviation of a population. This standard deviation is calculated by dividing the sum of squared deviations by n instead of $n - 1$ before taking the square root. If your data set consists of the entire population, then it's appropriate to use σ_x. Most often, the data we're examining come from a sample. In that case, we should use s_x.

TI-83/84

- Press STAT ▶ (CALC) ; choose 1-VarStats.





II. Output from statistical software We used Minitab statistical software to produce descriptive statistics for the New York and North Carolina travel time data. Minitab allows you to choose which numerical summaries are included in the output.

Descriptive Statistics: Travel time to work

Variable	N	Mean	StDev	Minimum	Q_1	Median	Q_3	Maximum
NY Time	20	31.25	21.88	5.00	15.00	22.50	43.75	85.00
NC Time	15	22.47	15.23	5.00	10.00	20.00	30.00	60.00

CHECK YOUR UNDERSTANDING

The heights (in inches) of the five starters on a basketball team are 67, 72, 76, 76, and 84.

1. Find the mean. Show your work.

Show Answer

2. Make a table that shows, for each value, its deviation from the mean and its squared deviation from the mean.

Show Answer

3. Show how to calculate the variance and standard deviation from the values in your table.

Show Answer

4. Interpret the standard deviation in this setting.

Show Answer

1. Find the mean. Show your work.

Hide Answer

Correct Answer
The mean is 75.

2. Make a table that shows, for each value, its deviation from the mean and its squared deviation from the mean.

Hide Answer

Correct Answer
The table is given below.

Observation	Deviation	Squared deviation
67	$67 - 75 = -8$	$(-8)^2 = 64$
72	$72 - 75 = -3$	$(-3)^2 = 9$
76	$76 - 75 = 1$	$1^2 = 1$
76	$76 - 75 = 1$	$1^2 = 1$
84	$84 - 75 = 9$	$9^2 = 81$
Total	**0**	**156**

3. Show how to calculate the variance and standard deviation from the values in your table.

Hide Answer

Correct Answer
The variance is $s_x^2 = \frac{156}{5-1} = 39$ inches squared and the standard deviation is $s_x = \sqrt{39} = 6.24$ inches.

4. Interpret the standard deviation in this setting.

Hide Answer

Correct Answer
The players' heights typically vary by about 6.24 inches from the mean height of 75 inches.

HOW TO ORGANIZE A STATISTICS PROBLEM: A FOUR-STEP PROCESS

State: What's the question that you're trying to answer?

Plan: How will you go about answering the question? What statistical techniques does this problem call for?

Do: Make graphs and carry out needed calculations.

Conclude: Give your conclusion in the setting of the real-world problem.

To keep the four steps straight, just remember: Statistics Problems Demand Consistency!

Many examples and exercises in this book will tell you what to do—construct a graph, perform a calculation, interpret a result, and so on. Real statistics problems don't come with such detailed instructions. From now on, you will encounter some examples and exercises that are more realistic. They are marked with the four-step icon. Use the four-step process as a guide to solving these problems, as the following example illustrates.

Example 23 Who Texts More—Males or Females

Putting it all together

STEP 4

For their final project, a group of AP® Statistics students wanted to compare the texting habits of males and females. They asked a random sample of students from their school to record the number of text messages sent and received over a two-day period. Here are their data:

Males:	127	44	28	83	0	6	78	6	5	213	73	20	214	28	11
Females:	112	203	102	54	379	305	179	24	127	65	41	27	298	6	130	0

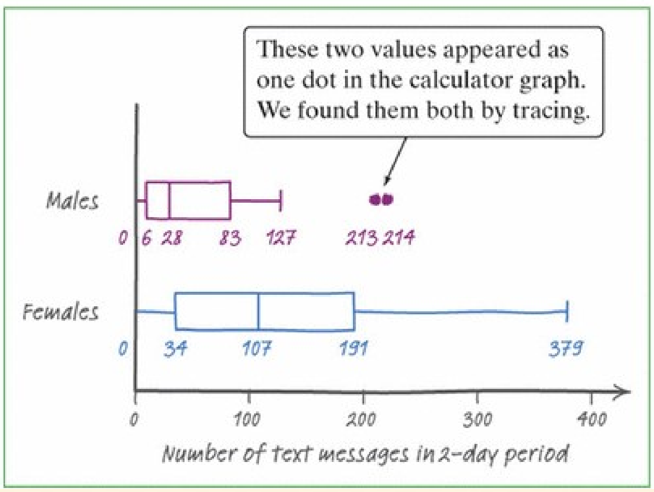
What conclusion should the students draw? Give appropriate evidence to support your answer.

Worked Example Video

STATE: Do males and females at the school differ in their texting habits?

PLAN: We'll begin by making parallel boxplots of the data about males and females. Then we'll calculate one-variable statistics. Finally, we'll compare shape, center, spread, and outliers for the two distributions.

DO: Figure 1.21 is a sketch of the boxplots we got from our calculator. The table below shows numerical summaries for males and females.



	$\bar{x}$	s_x	Min	Q_1	Med	Q_3	Max	*IQR*
Male	62.4	71.4	0	6	28	83	214	77
Female	128.3	116.0	0	34	107	191	379	157

Due to the strong skewness and outliers, we'll use the median and *IQR* instead of the mean and standard deviation when comparing center and spread.

Shape: Both distributions are strongly right-skewed.

Center: Females typically text more than males. The median number of texts for females (107) is about four times as high as for males (28). In fact, the median for the females is above the third quartile for the males. This indicates that over 75% of the males texted less than the "typical" (median) female.

Spread: There is much more variation in texting among the females than the males. The *IQR* for females (157) is about twice the *IQR* for males (77).

Outliers: There are two outliers in the male distribution: students who reported 213 and 214 texts in two days. The female distribution has no outliers.

CONCLUDE: The data from this survey project give very strong evidence that male and female texting habits differ considerably at the school. A typical female sends and receives about 79 more text messages in a two-day period than a typical male. The males as a group are also much more consistent in their texting frequency than the females.