

Section 3-2

Measures of Variation

WAITING TIMES AT DIFFERENT BANKS

Jefferson Valley Bank (single waiting line)	6.5	6.6	6.7	6.8	7.1	7.3	7.4	7.7	7.7	7.7
Bank of Providence (multiple waiting lines)	4.2	5.4	5.8	6.2	6.7	7.7	7.7	8.5	9.3	10.0

All the measures of center are equal for both banks.

Mean = 7.15 min

Median = 7.20 min

Mode = 7.7 min

Midrange = 7.10 min

RANGE

The **range** of a set of data values is the difference between the highest value and the lowest value.

$$\text{range} = (\text{maximum data value}) - (\text{minimum data value})$$

EXAMPLE:

Jefferson Valley Bank range = $7.7 - 6.5 = 1.2$ min

Bank of Providence range = $10.0 - 4.2 = 5.8$ min

STANDARD DEVIATION FOR A SAMPLE

The **standard deviation** of a set of sample values, denoted by s , is a measure of how much the data values deviate away from the mean

STANDARD DEVIATION FORMULAS

Sample Standard
Deviation

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

Shortcut Formula
for Sample
Standard Deviation

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

EXAMPLE

Use both the regular formula and shortcut formula to find the standard deviation of the following.

3 7 4 2

PROPERTIES OF THE STANDARD DEVIATION

- The standard deviation is a measure of variation of all values from the **mean**.
- The value of the standard deviation s is never negative. It is zero only when all of the data values are the same number.
- Larger values of s indicate greater amounts of variation.
- The standard deviation s can increase dramatically with the inclusion of one or more outliers.
- The units of the standard deviation s (such as minutes, feet, pounds) are the same as the units of the original data values.
- The sample standard deviation s is a **biased estimator** of the population standard deviation σ , which means that the values of the sample standard deviation do not center around the value of σ .

STANDARD DEVIATION OF A POPULATION

The standard deviation for a **population** is denoted by σ and is given by the formula

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

FINDING THE STANDARD DEVIATION ON THE TI-83/84

1. Press **STAT**; select **1:Edit...**
2. Enter your data values in **L1**. (You may enter the values in any of the lists.)
3. Press **2ND, MODE** (for **QUIT**).
4. Press **STAT**; arrow over to **CALC**. Select **1:1-Var Stats**.
5. Enter L1 by pressing **2ND, 1**.
6. Press **ENTER**.
7. The **sample** standard deviation is given by **Sx**. The **population** standard deviation is given by **σx** .

FINDING THE STANDARD DEVIATION ON THE TI-84 WITH NEW OS

1. Press **STAT**; select **1:Edit...**
2. Enter your data values in **L1**. (You may enter the values in any of the lists.)
3. Press **2ND, MODE** (for **QUIT**).
4. Press **STAT**; arrow over to **CALC**. Select **1:1-Var Stats**.
5. For "List", enter L1 by pressing **2ND, 1**.
6. Leave "FreqList" blank.
7. Highlight "Calculate" and press **ENTER**.
8. The **sample** standard deviation is given by **Sx**. The **population** standard deviation is given by **σx** .

SYMBOLS FOR STANDARD DEVIATION

Sample

s ← textbook

Sx ← TI-83/84 Calculators

Population

σ ← textbook

σx ← TI-83/84 Calculators

EXAMPLE

Use your calculator to find the standard deviation for waiting times at the Jefferson Valley Bank and the Bank of Providence.

Jefferson Valley Bank (single waiting line)	6.5	6.6	6.7	6.8	7.1	7.3	7.4	7.7	7.7	7.7
Bank of Providence (multiple waiting lines)	4.2	5.4	5.8	6.2	6.7	7.7	7.7	8.5	9.3	10.0

VARIANCE

The **variance** of set of values is a measure of variation equal to the square of the standard deviation.

sample variance = s^2

population variance = σ^2

PROPERTIES OF THE VARIANCE

- The units of the variance are the *squares* of the units of the original data values. (If the original data values are in seconds, the variance will have units of sec^2 .)
- The value of the variance can increase dramatically with the inclusion of one or more outliers. (The variance is not *resistant*.)
- The value of the variance is never negative. It is zero only when all of the data values are the same number.
- The sample variance is an **unbiased estimator** of the population variance σ^2 .

ROUND-OFF RULE FOR MEASURES OF VARIATION

Carry one more decimal place than is present in the original data set.

**STANDARD DEVIATION FROM A
FREQUENCY DISTRIBUTION**

$$s = \sqrt{\frac{n[\sum(f \cdot x^2)] - [\sum(f \cdot x)]^2}{n(n - 1)}}$$

Use the class midpoints as the x values.

NOTE: We will *not* use this formula but will use the TI-83/84 calculator.

**FINDING THE STANDARD DEVIATION
FROM A FREQUENCY TABLE
ON THE TI-83/84**

1. Enter the class midpoints in **L1**.
2. Enter the frequencies in **L2**.
3. Press **STAT**, arrow over to **CALC**, and select **1:1-Var Stats**.
4. Press **L1,L2** followed by **ENTER**.
5. The sample standard deviation will be **Sx**.
The population standard deviation will be **σx** .

**FINDING THE STANDARD DEVIATION
FROM A FREQUENCY TABLE ON THE TI-84
WITH NEW OS**

1. Enter the class midpoints in **L1**.
2. Enter the frequencies in **L2**.
3. Press **STAT**, arrow over to **CALC**, and select **1:1-Var Stats**.
4. For "List", enter **L1**.
5. For "FreqList", enter **L2**.
6. Highlight "Calculate" and press **ENTER**.
7. The sample standard deviation will be **Sx**. The population standard deviation will be **σx** .

EXAMPLE

Find the sample standard deviation of the following.

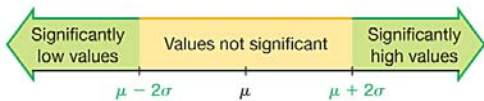
CLASS	FREQUENCY
1-3	4
4-6	9
7-9	6

RANGE RULE OF THUMB FOR IDENTIFYING SIGNIFICANT VALUES

Significantly low values are $(\mu - 2\sigma)$ or lower.

Significantly high values are $(\mu + 2\sigma)$ or higher.

Values that are not significant are between $(\mu - 2\sigma)$ and $(\mu + 2\sigma)$.



RANGE RULE OF THUMB FOR ESTIMATING A VALUE OF THE STANDARD DEVIATION s

To **roughly** estimate the standard deviation, use

$$s \approx \frac{\text{range}}{4}$$

where

$$\text{range} = (\text{maximum data value}) - (\text{minimum data value})$$

EXAMPLES

1. Heights of men have a mean of 69.0 in and a standard deviation of 2.8 in. Use the range rule of thumb to find the limits separating the values that are significantly low and those that are significantly high. In this context, determine if the height 6 ft, 6 in is significantly high.
2. The shortest home-run hit by Mark McGwire was 340 ft and the longest was 550 ft. Use the range rule of thumb to estimate the standard deviation.

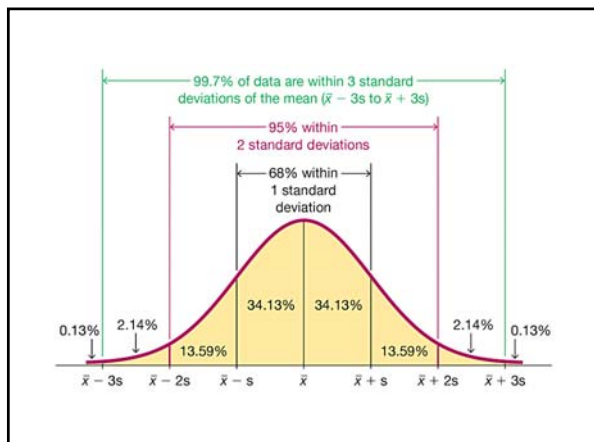
MORE PROPERTIES OF THE STANDARD DEVIATION

- The standard deviation measures the *variation* among the data values.
- Values close together have a small standard deviation, but values with much more variation have a larger standard deviation.
- The standard deviation has the same units of measurement as the original values.
- For many data sets, a value is *significant* if it differs from the mean by more than two standard deviations.
- When comparing variation in two different data sets, compare the standard deviations only if the data sets use the same scale and units and they have means that are approximately the same.

THE EMPIRICAL (OR 68-95-99.7) RULE FOR DATA WITH A BELL-SHAPED DISTRIBUTION

The **empirical** (or **68-95-99.7**) **rule** states that for data sets having a distribution that is approximately **bell-shaped**, the following properties apply.

- About 68% of all values fall within 1 standard deviation of the mean.
- About 95% of all values fall within 2 standard deviations of the mean.
- About 99.7% of all values fall within 3 standard deviations of the mean.



EXAMPLE

A random sample of 50 gas stations in Cook County, Illinois, resulted in a mean price per gallon of \$2.56 and a standard deviation of \$0.07. A histogram indicated that the data follow a bell-shaped distribution.

- Use the Empirical Rule to determine the percentage of gas stations that have prices within three standard deviations of the mean. What are these gas prices?
- Determine the percentage of gas stations with prices between \$2.42 and \$2.70, according to the empirical rule.

CHEBYSHEV'S RULE

The proportion (or fraction) of *any* set of data lying within K standard deviations of the mean is always **at least** $1 - \frac{1}{K^2}$, where K is any positive number greater than 1. For $K = 2$ and $K = 3$, we get the following statements:

- At least $\frac{3}{4}$ (or 75%) of all values lie within 2 standard deviations of the mean.
- At least $\frac{8}{9}$ (or 89%) of all values lie within 3 three standard deviations of the mean.

EXAMPLE

A survey of local companies found that the mean amount of travel allowance for executives was 25¢ per mile. The standard deviation was 2¢. Find the minimum percentage of data values that will fall between 19¢ and 31¢.
