Lesson #3

MAT 1372 Statistics with Probability

Dr. Bonanome

Measures of Central Tendency

Measure of central tendency

- A value that represents a typical, or central, entry of a data set.

- Most common measures of central tendency:
  - Mean
  - Median
  - Mode

Measure of Central Tendency: Mean

Mean (average)

- The sum of all the data entries divided by the number of entries.

- **Sigma notation**: Σx = add all of the data entries (x) in the data set.

- **Population mean**: $\mu = \dfrac{\Sigma x}{N}$

- **Sample mean**: $\bar{x} = \dfrac{\Sigma x}{n}$
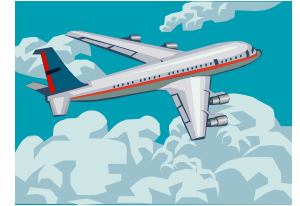
Example: Finding a Sample Mean

The prices (in dollars) for a sample of round-trip flights from Chicago, Illinois to Cancun, Mexico are listed. What is the mean price of the flights?

872 432 397 427 388 782 397
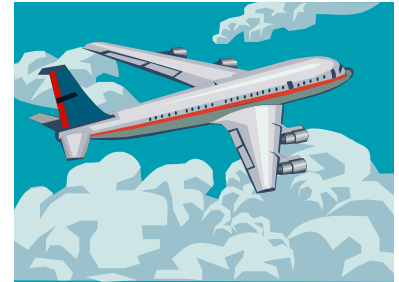
Solution: Finding a Sample Mean

872 432 397 427 388 782 397

- The sum of the flight prices is

$$\Sigma x = 872 + 432 + 397 + 427 + 388 + 782 + 397 = 3695$$

- To find the mean price, divide the sum of the prices by the number of prices in the sample

$$\bar{x} = \frac{\Sigma x}{n} = \frac{3695}{7} \approx 527.9$$

The mean price of the flights is about $527.90.

Measure of Central Tendency: Median

Median

- The value that lies in the middle of the data when the data set is **ordered**.

- Measures the center of an ordered data set by dividing it into two equal parts.

- If the data set has an

  - **odd number of entries**: median is the middle data entry.

  - **even number of entries**: median is the mean of the two middle data entries.
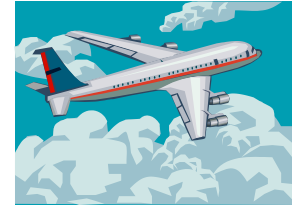
Example: Finding the Median

The prices (in dollars) for a sample of roundtrip flights from Chicago, Illinois to Cancun, Mexico are listed. Find the median of the flight prices.

872   432   397   427   388   782   397

Solution: Finding the Median

872   432   397   427   388   782   397

- First order the data.

  388   397   397   427   432   782   872

  ⬆

- There are seven entries (an odd number), the median is the middle, or fourth, data entry.

The median price of the flights is $427.

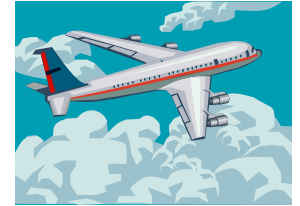# Example: Finding the Median

The flight priced at $432 is no longer available. What is the median price of the remaining flights?

872   397   427   388   782   397

Solution: Finding the Median

872   397   427   388   782   397

- First order the data.

    388   397   397   427   782   872

- There are six entries (an even number), the median is the mean of the two middle entries.

$$\text{Median} = \frac{397 + 427}{2} = 412$$

The median price of the flights is $412.

Measure of Central Tendency: Mode

Mode

- The data entry that occurs with the greatest frequency.

- A data set can have one mode, more than one mode, or no mode.

- If no entry is repeated the data set has no mode.

- If two entries occur with the same greatest frequency, each entry is a mode (**bimodal**).
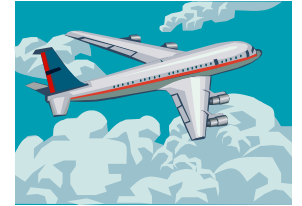
Example: Finding the Mode

The prices (in dollars) for a sample of roundtrip flights from Chicago, Illinois to Cancun, Mexico are listed. Find the mode of the flight prices.

872   432   397   427   388   782   397

Solution: Finding the Mode

872   432   397   427   388   782   397

- Ordering the data helps to find the mode.

  388   397   397   427   432   782   872

- The entry of 397 occurs twice, whereas the other data entries occur only once.

The mode of the flight prices is $397.

Example: Finding the Mode

At a political debate a sample of audience members was asked to name the political party to which they belong. Their responses are shown in the table. What is the mode of the responses?

Political Party	Frequency, f
Democrat	34
Republican	56
Other	21
Did not respond	9

Solution: Finding the Mode

Political Party	Frequency, f
Democrat	34
Republican	56
Other	21
Did not respond	9

The mode is Republican (the response occurring with the greatest frequency). In this sample there were more Republicans than people of any other single affiliation.

Comparing the Mean, Median, and Mode

- All three measures describe a typical entry of a data set.

- Advantage of using the mean:
  - The mean is a reliable measure because it takes into account every entry of a data set.

- Disadvantage of using the mean:
  - Greatly affected by **outliers** (a data entry that is far removed from the other entries in the data set).

Example: Comparing the Mean, Median, and Mode

Find the mean, median, and mode of the sample ages of a class shown. Which measure of central tendency best describes a typical entry of this data set? Are there any outliers?

Ages in a class						
20	20	20	20	20	20	21
21	21	21	22	22	22	23
23	23	23	24	24	65	

Solution: Comparing the Mean, Median, and Mode

Ages in a class						
20	20	20	20	20	20	21
21	21	21	22	22	22	23
23	23	23	24	24	65	

Mean:
$$\bar{x} = \frac{\Sigma x}{n} = \frac{20 + 20 + \ldots + 24 + 65}{20} \approx 23.8 \text{ years}$$

Median:
$$\frac{21 + 22}{2} = 21.5 \text{ years}$$

Mode: 20 years (the entry occurring with the greatest frequency)

Solution: Comparing the Mean, Median, and Mode

Mean ≈ 23.8 years   Median $= 21.5$ years   Mode $=$ 20 years

- The mean takes every entry into account, but is influenced by the **outlier** of 65.
- The median also takes every entry into account, and it is not affected by the outlier.
- In this case the mode exists, but it doesn't appear to represent a typical entry.

Solution: Comparing the Mean, Median, and Mode

Sometimes a graphical comparison can help you decide which measure of central tendency best represents a data set.

Ages of Students in a Class
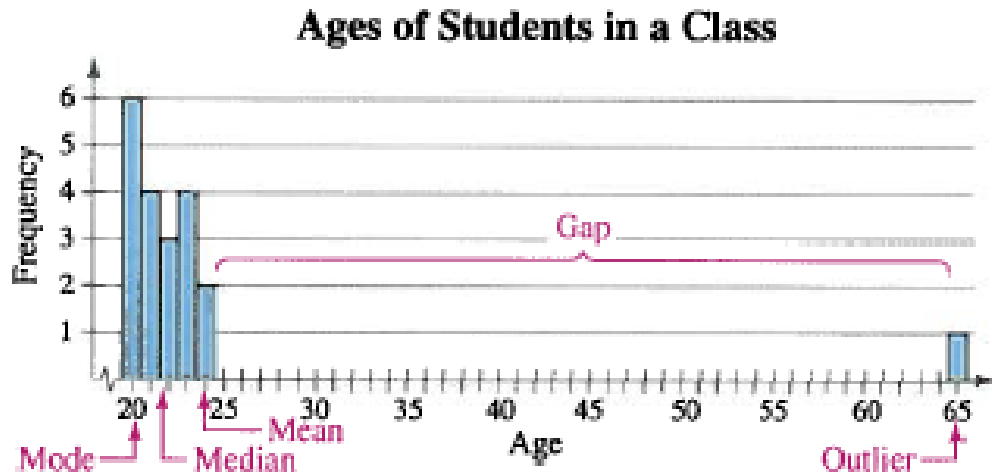
In this case, it appears that the **median** best describes the data set.

Weighted Mean

Weighted Mean

- The mean of a data set whose entries have varying weights.

- $$\bar{x} = \frac{\Sigma(x \cdot w)}{\Sigma w}$$ where *w* is the weight of each entry *x*

Example: Finding a Weighted Mean

You are taking a class in which your grade is determined from five sources: 50% from your test mean, 15% from your midterm, 20% from your final exam, 10% from your computer lab work, and 5% from your homework. Your scores are 86 (test mean), 96 (midterm), 82 (final exam), 98 (computer lab), and 100 (homework). What is the weighted mean of your scores? If the minimum average for an A is 90, did you get an A?

Solution: Finding a Weighted Mean

Source	Score, x	Weight, w	$x{\cdot}w$
Test Mean	86	0.50	$86(0.50) = 43.0$
Midterm	96	0.15	$96(0.15) = 14.4$
Final Exam	82	0.20	$82(0.20) = 16.4$
Computer Lab	98	0.10	$98(0.10) = 9.8$
Homework	100	0.05	$100(0.05) = 5.0$
		$\Sigma w = 1$	$\Sigma(x{\cdot}w) = 88.6$

$$\overline{x} = \frac{\Sigma(x \cdot w)}{\Sigma w} = \frac{88.6}{1} = 88.6$$

Your weighted mean for the course is 88.6. You did not get an A.

Mean of Grouped Data

Mean of a Frequency Distribution

- Approximated by

$$\bar{x} = \frac{\Sigma(x \cdot f)}{n} \qquad n = \Sigma f$$

where *x* and *f* are the midpoints and frequencies of a class, respectively

Finding the Mean of a Frequency Distribution

In Words	*In Symbols*
1. Find the midpoint of each class.	$x = \dfrac{\text{(lower limit)}+\text{(upper limit)}}{2}$
2. Find the sum of the products of the midpoints and the frequencies.	$\Sigma(x \cdot f)$
3. Find the sum of the frequencies.	$n = \Sigma f$
4. Find the mean of the frequency distribution.	$\overline{x} = \dfrac{\Sigma(x \cdot f)}{n}$

Example: Find the Mean of a Frequency Distribution

Use the frequency distribution to approximate the mean number of minutes that a sample of Internet subscribers spent online during their most recent session.

Class	Midpoint	Frequency, f
7 – 18	12.5	6
19 – 30	24.5	10
31 – 42	36.5	13
43 – 54	48.5	8
55 – 66	60.5	5
67 – 78	72.5	6
79 – 90	84.5	2

Solution: Find the Mean of a Frequency Distribution

Class	Midpoint, x	Frequency, f	$(x \cdot f)$
7 – 18	12.5	6	$12.5 \cdot 6 = 75.0$
19 – 30	24.5	10	$24.5 \cdot 10 = 245.0$
31 – 42	36.5	13	$36.5 \cdot 13 = 474.5$
43 – 54	48.5	8	$48.5 \cdot 8 = 388.0$
55 – 66	60.5	5	$60.5 \cdot 5 = 302.5$
67 – 78	72.5	6	$72.5 \cdot 6 = 435.0$
79 – 90	84.5	2	$84.5 \cdot 2 = 169.0$
		n = 50	$\Sigma(x \cdot f) = 2089.0$

$$\bar{x} = \frac{\Sigma(x \cdot f)}{n} = \frac{2089}{50} \approx 41.8 \text{ minutes}$$

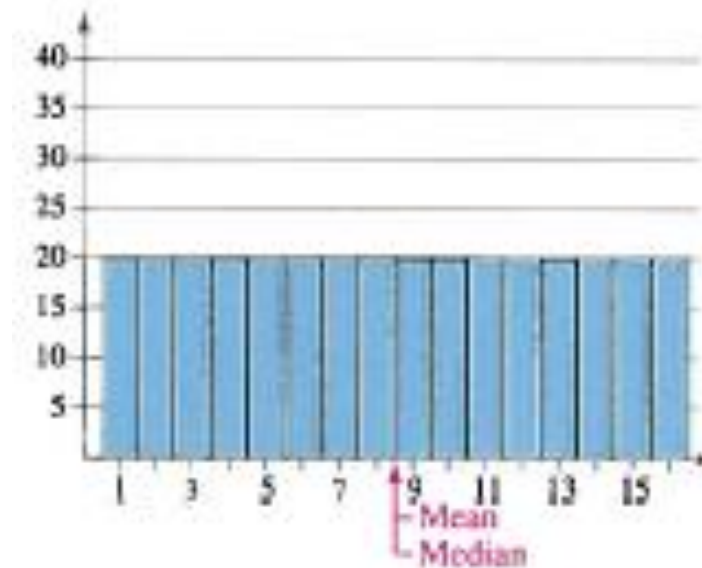# The Shape of Distributions

Symmetric Distribution

- A vertical line can be drawn through the middle of a graph of the distribution and the resulting halves are approximately mirror images.
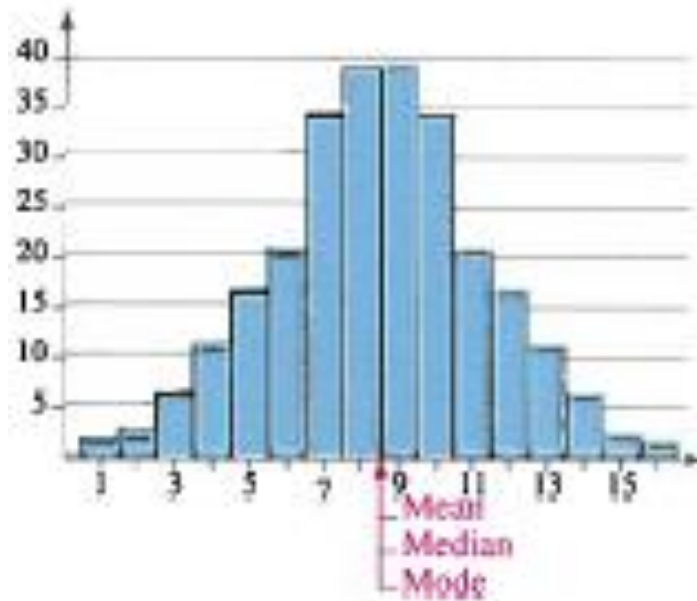
The Shape of Distributions
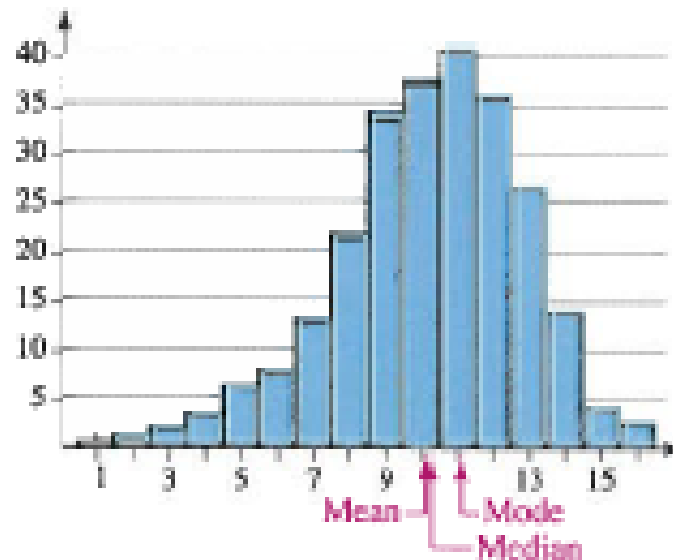
Uniform Distribution (rectangular)

- All entries or classes in the distribution have equal or approximately equal frequencies.

- Symmetric.

The Shape of Distributions
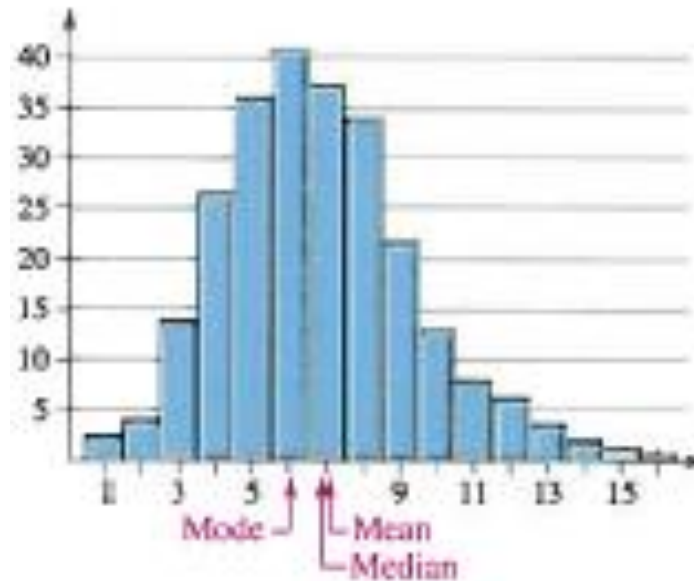
Skewed Left Distribution (negatively skewed)

- The "tail" of the graph elongates more to the left.
- The mean is to the left of the median.

The Shape of Distributions

Skewed Right Distribution (positively skewed)

- The "tail" of the graph elongates more to the right.
- The mean is to the right of the median.

Percentiles and Other Fractiles

Fractiles	Summary	Symbols
Quartiles	Divide a data set into 4 equal parts	Q_1, Q_2, Q_3
Deciles	Divide a data set into 10 equal parts	$D_1, D_2, D_3, \ldots, D_9$
Percentiles	Divide a data set into 100 equal parts	$P_1, P_2, P_3, \ldots, P_{99}$

# Example: Interpreting Percentiles

The ogive represents the cumulative frequency distribution for SAT test scores of college-bound students in a recent year. What test score represents the 62nd percentile? How should you interpret this? *(Source: College Board)*
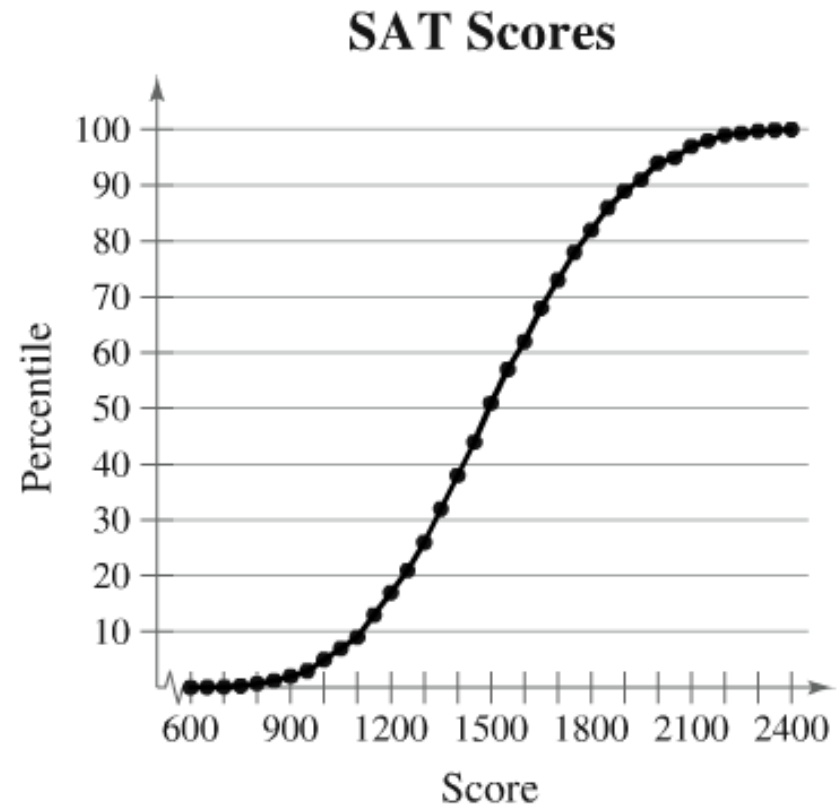
SAT Scores

Solution: Interpreting Percentiles

The 62nd percentile corresponds to a test score of 1600.

This means that 62% of the students had an SAT score of 1600 or less.

SAT Scores
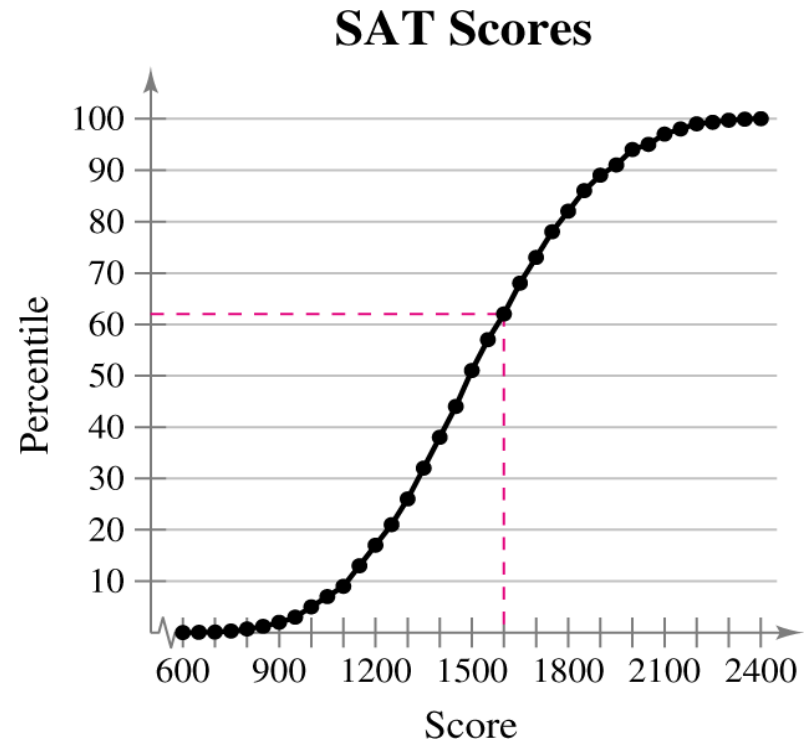
Deviation, Variance, and Standard Deviation

Deviation

- The difference between the data entry, *x*, and the mean of the data set.

- Population data set:
  - Deviation of $x = x - \mu$

- Sample data set:
  - Deviation of $x = x - \bar{x}$

Example: Finding the Deviation

A corporation hired 10 graduates. The starting salaries for each graduate are shown. Find the deviation of the starting salaries.

Starting salaries (1000s of dollars)

41   38   39   45   47   41   44   41   37   42

Solution:

- First determine the mean starting salary.

$$\mu = \frac{\Sigma x}{N} = \frac{415}{10} = 41.5$$

Solution: Finding the Deviation

- Determine the deviation for each data entry.

Salary ($1000s), x	Deviation ($1000s) $x - \mu$
41	$41 - 41.5 = -0.5$
38	$38 - 41.5 = -3.5$
39	$39 - 41.5 = -2.5$
45	$45 - 41.5 = 3.5$
47	$47 - 41.5 = 5.5$
41	$41 - 41.5 = -0.5$
44	$44 - 41.5 = 2.5$
41	$41 - 41.5 = -0.5$
37	$37 - 41.5 = -4.5$
42	$42 - 41.5 = 0.5$
$\Sigma x = 415$	$\Sigma(x - \mu) = 0$

Deviation, Variance, and Standard Deviation

Sample Variance

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

Sample Standard Deviation

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

Finding the Sample Variance & Standard Deviation

In Words	*In Symbols*
1. Find the mean of the sample data set.	$\bar{x} = \dfrac{\Sigma x}{n}$
2. Find the deviation of each entry.	$x - \bar{x}$
3. Square each deviation.	$(x - \bar{x})^2$
4. Add to get the **sum of squares**.	$SS_x = \Sigma(x - \bar{x})^2$

Finding the Sample Variance & Standard Deviation

In Words	*In Symbols*
5. Divide by $n - 1$ to get the **sample variance**.	$s^2 = \dfrac{\Sigma(x - \bar{x})^2}{n - 1}$
6. Find the square root of the variance to get the **sample standard deviation**.	$s = \sqrt{\dfrac{\Sigma(x - \bar{x})^2}{n - 1}}$

Example: Finding the Sample Standard Deviation

The starting salaries are for the Chicago branches of a corporation. The corporation has several other branches, and you plan to use the starting salaries of the Chicago branches to estimate the starting salaries for the larger population. Find the *sample* standard deviation of the starting salaries.

Starting salaries (1000s of dollars)

41   38   39   45   47   41   44   41   37   42

Solution: Finding the Sample Standard Deviation

- Determine SS_x
- $n = 10$

Salary, x	Deviation: $x - \mu$	Squares: $(x - \mu)^2$
41	$41 - 41.5 = -0.5$	$(-0.5)^2 = 0.25$
38	$38 - 41.5 = -3.5$	$(-3.5)^2 = 12.25$
39	$39 - 41.5 = -2.5$	$(-2.5)^2 = 6.25$
45	$45 - 41.5 = 3.5$	$(3.5)^2 = 12.25$
47	$47 - 41.5 = 5.5$	$(5.5)^2 = 30.25$
41	$41 - 41.5 = -0.5$	$(-0.5)^2 = 0.25$
44	$44 - 41.5 = 2.5$	$(2.5)^2 = 6.25$
41	$41 - 41.5 = -0.5$	$(-0.5)^2 = 0.25$
37	$37 - 41.5 = -4.5$	$(-4.5)^2 = 20.25$
42	$42 - 41.5 = 0.5$	$(0.5)^2 = 0.25$

$$\Sigma(x - \mu) = 0 \qquad SS_x = 88.5$$

Solution: Finding the Sample Standard Deviation

Sample Variance

- $s^2 = \dfrac{\Sigma(x-\bar{x})^2}{n-1} = \dfrac{88.5}{10-1} \approx 9.8$

Sample Standard Deviation

- $s = \sqrt{s^2} = \sqrt{\dfrac{88.5}{9}} \approx 3.1$

The sample standard deviation is about 3.1, or \$3100.

Interpreting Standard Deviation

- Standard deviation is a measure of the typical amount an entry deviates from the mean.

- The more the entries are spread out, the greater the standard deviation.