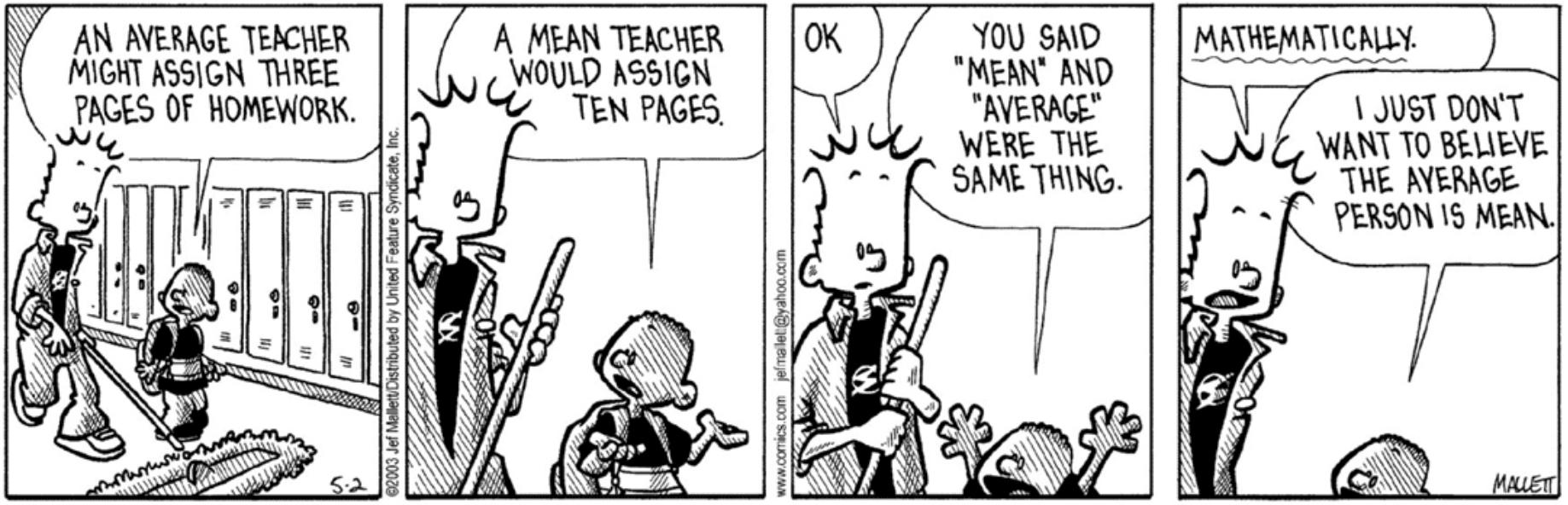
Mean vs. Average



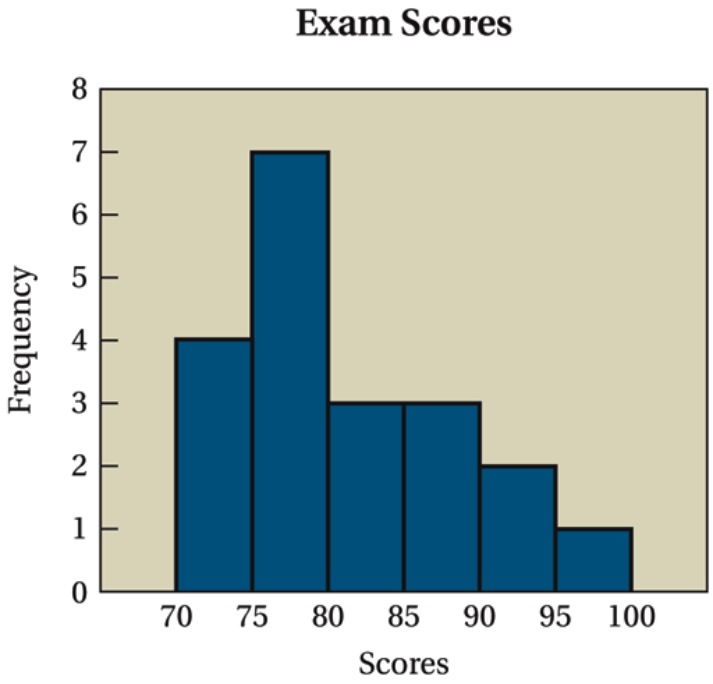
FRAZZ reprinted by permission of United Feature Syndicate, Inc.

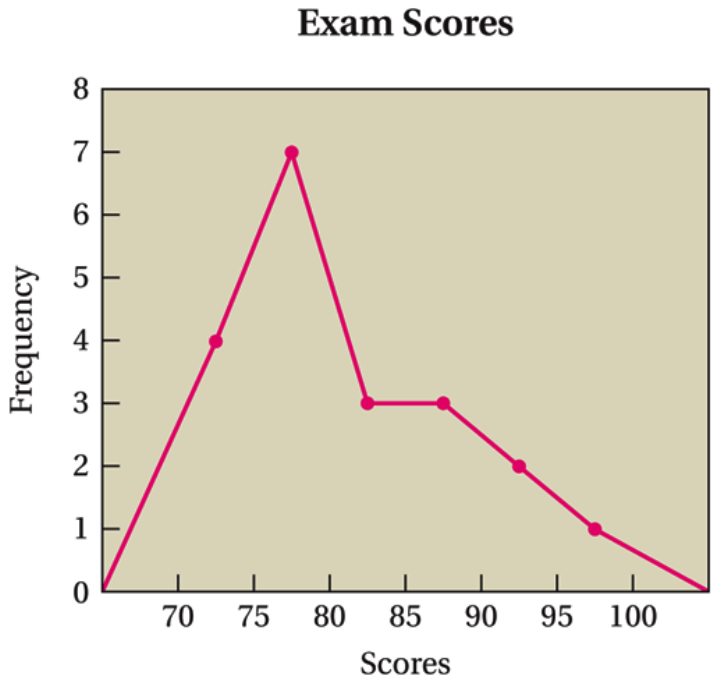
Unit 6A

Characterizing Data

Definition

The **distribution** of a variable (or data set) describes the values taken on by the variable and the frequency (or relative frequency) of these values.





Measures of Center in a Distribution

- The **mean** is what we most commonly call the average value. It is defined as follows:

$$\text{mean} = \frac{\text{sum of all values}}{\text{total number of values}}$$

- The **median** is the middle value in the sorted data set (or halfway between the two middle values if the number of values is even).
- The **mode** is the most common value (or group of values) in a distribution.

Examples: *Calculating the Median*

20 25 25 27 28 31 33 34 36 37 44 50 59 85 86

***n*=15** so use middle value → **Therefore Median=34**

Median

5 7 10 14 18 | 19 25 29 31 33

***n*=10** so average two middle values

Therefore Median $= \frac{18+19}{2} = 18.5$

Finding the Mode

Example: Find the mode of each data set below.

a.	**5**	**5**	**5**	**3**	**1**	**5**	**1**	**4**	**3**	**5**	
b.	**1**	**2**	**2**	**2**	**3**	**4**	**5**	**6**	**6**	**6**	**7**	**9**
c.	**1**	**2**	**3**	**6**	**7**	**8**	**9**	**10**				

✓ Mode is 5

✓ Bimodal (2, 6)

✓ No Mode

Effects of Outliers

An **outlier** is a data value that is much higher or much lower than almost all other values.

Consider the following data set of contract offers:

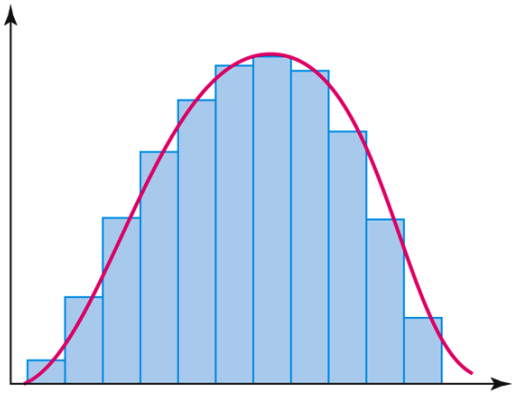
$0 $0 $0 $0 $2,500,000

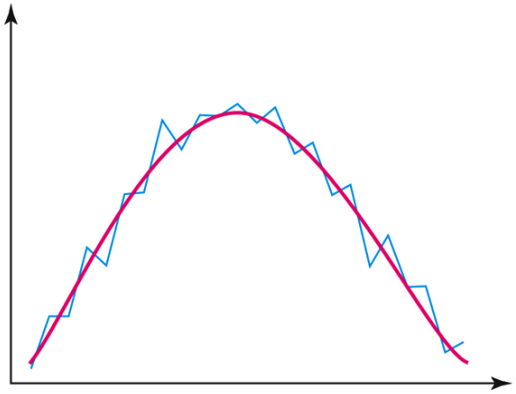
The mean contract offer is

$$\text{mean} = \frac{\$0 + \$0 + \$0 + \$0 + \$2{,}500{,}000}{5} = \$500{,}000$$

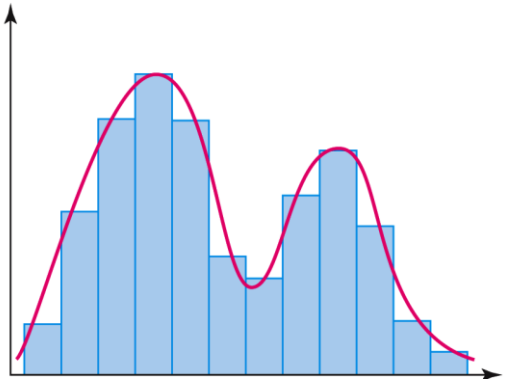
Outliers can pull the mean upward (or downward). The median and mode of the data are unaffected.

Shapes of Distributions





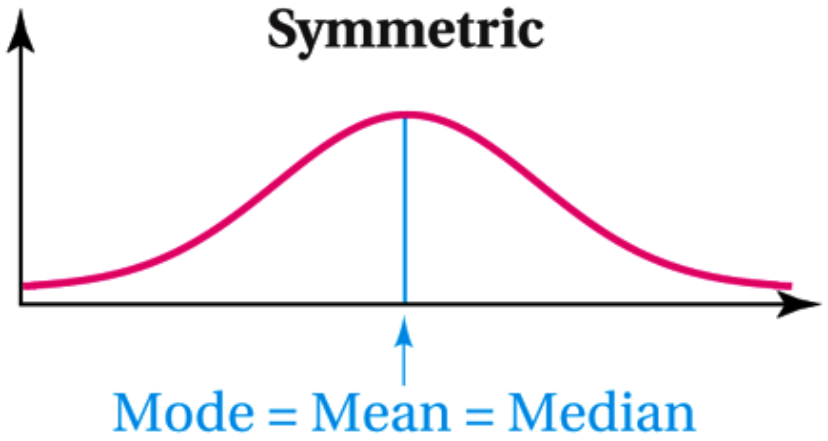
Two single-peaked (unimodal) distributions



A double-peaked (bimodal) distribution

Symmetry

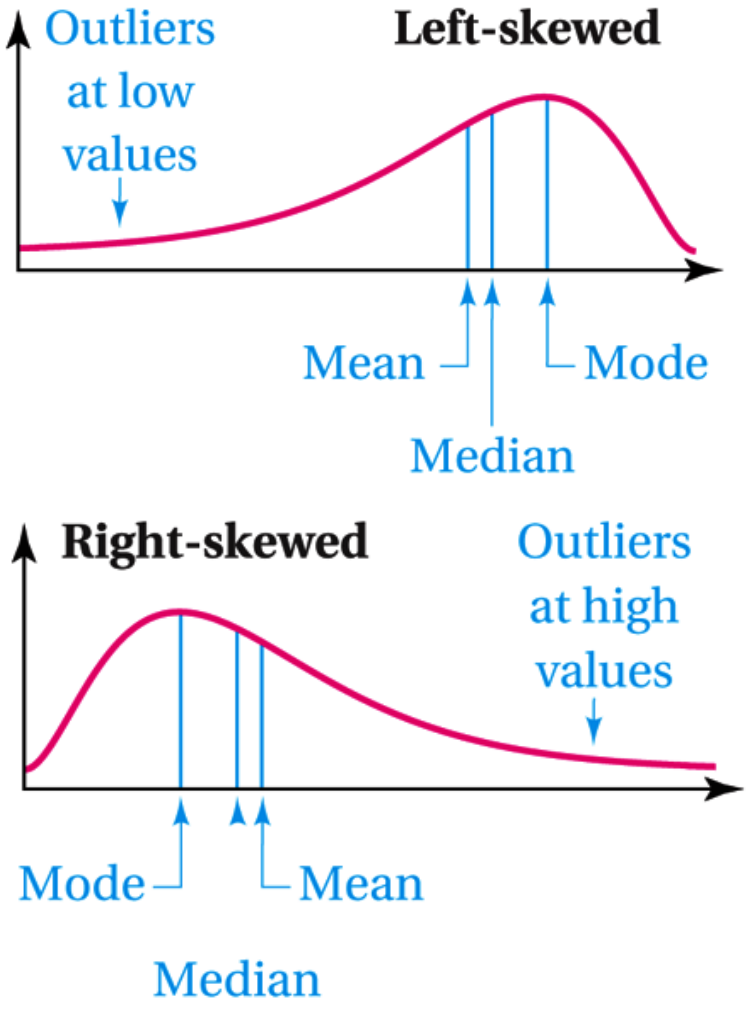
A distribution is **symmetric** if its left half is a mirror image of its right half.



Skewness

A distribution is **left-skewed** if its values are more spread out on the left side.

A distribution is **right-skewed** if its values are more spread out on the right side.

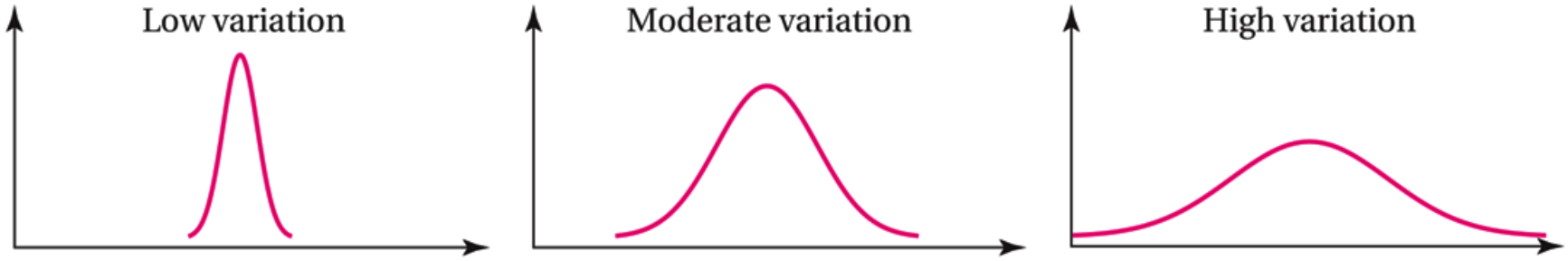


Unit 6B

Measures of Variation

Variation

Variation describes how widely data values are spread out about the center of a distribution.



From left to right, these three distributions have increasing variation.

Range

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

range = highest value (max) – lowest value (min)

Quartiles

- The **lower quartile,** (Q_1 or first quartile) divides the lowest fourth of a data set from the upper three-fourths. It is the median of the data values in the *lower half* of a data set.
- The **middle quartile** (or second quartile) is the overall median (M).
- The **upper quartile** (Q_3 or third quartile) divides the lower three-fourths of a data set from the upper fourth. It is the median of the data values in the *upper half* of a data set.

The Five-Number Summary

- The **five-number summary** for a data set consists of the following five numbers:

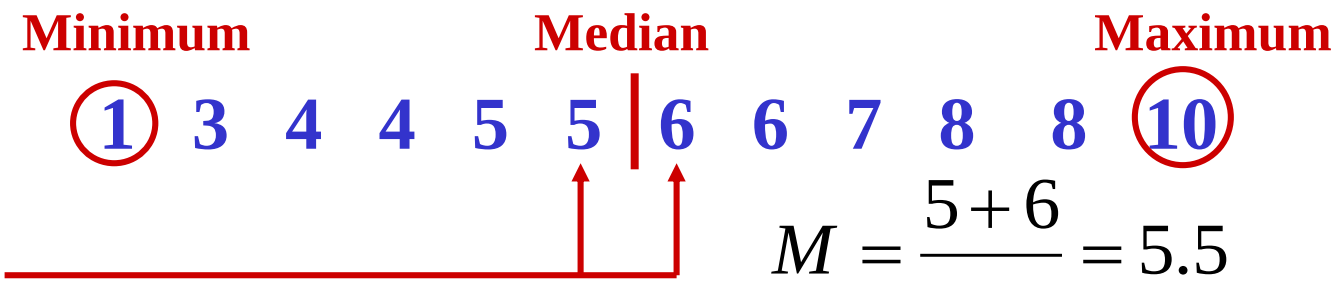
low value	*lower quartile*	*median*	*upper quartile*	*high value*
min	Q_1	M	Q_3	*max*

Calculating the five-number summary

List the *five-number summary* of the given data set:

7 4 10 8 5 6 4 6 1 3 8 5

Solution: **Arrange all observations in order of size:**



$$M = \frac{5+6}{2} = 5.5$$

The median of **1 3 4 4 5 5** is $Q_1 = \frac{4+4}{2} = 4$

The median of **6 6 7 8 8 10** is $Q_3 = \frac{7+8}{2} = 7.5$

Answer: 1 4 5.5 7.5 10

Standard Deviation

The **standard deviation** is the single number most commonly used to describe variation.

$$\text{standard deviation} = \sqrt{\frac{\text{sum of } (\text{deviations from the mean})^2}{\text{total number of data values} - 1}}$$

Calculating the Standard Deviation

The **standard deviation** is calculated by completing the following steps:

1. Compute the mean of the data set. Then find the deviation from the mean for every data value.

   deviation from the mean = data value – mean

2. Find the squares of all the deviations from the mean.
3. Add all the squares of the deviations from the mean.
4. Divide this sum by the total number of data values minus 1.
5. The standard deviation is the square root of this quotient.

Standard Deviation

Let $A = \{2, 8, 9, 12, 19\}$ with a mean of 10. Find the sample standard deviation of the data set A.

x (data value)	x – mean (deviation)	(deviation)^2
2	$2 - 10 = -8$	$(-8)^2 = 64$
8	$8 - 10 = -2$	$(-2)^2 = 4$
9	$9 - 10 = -1$	$(-1)^2 = 1$
12	$12 - 10 = 2$	$(2)^2 = 4$
19	$19 - 10 = 9$	$(9)^2 = 81$
	Total	**154**

Standard Deviation

Let $A = \{2, 8, 9, 12, 19\}$ with a mean of 10. Find the sample standard deviation of the data set A.

$$\text{standard deviation} = \sqrt{\frac{\text{sum of (deviations from the mean)}^2}{\text{total number of data values} - 1}}$$

$$= \sqrt{\frac{154}{5 - 1}} = \sqrt{38.5} \approx 6.2$$