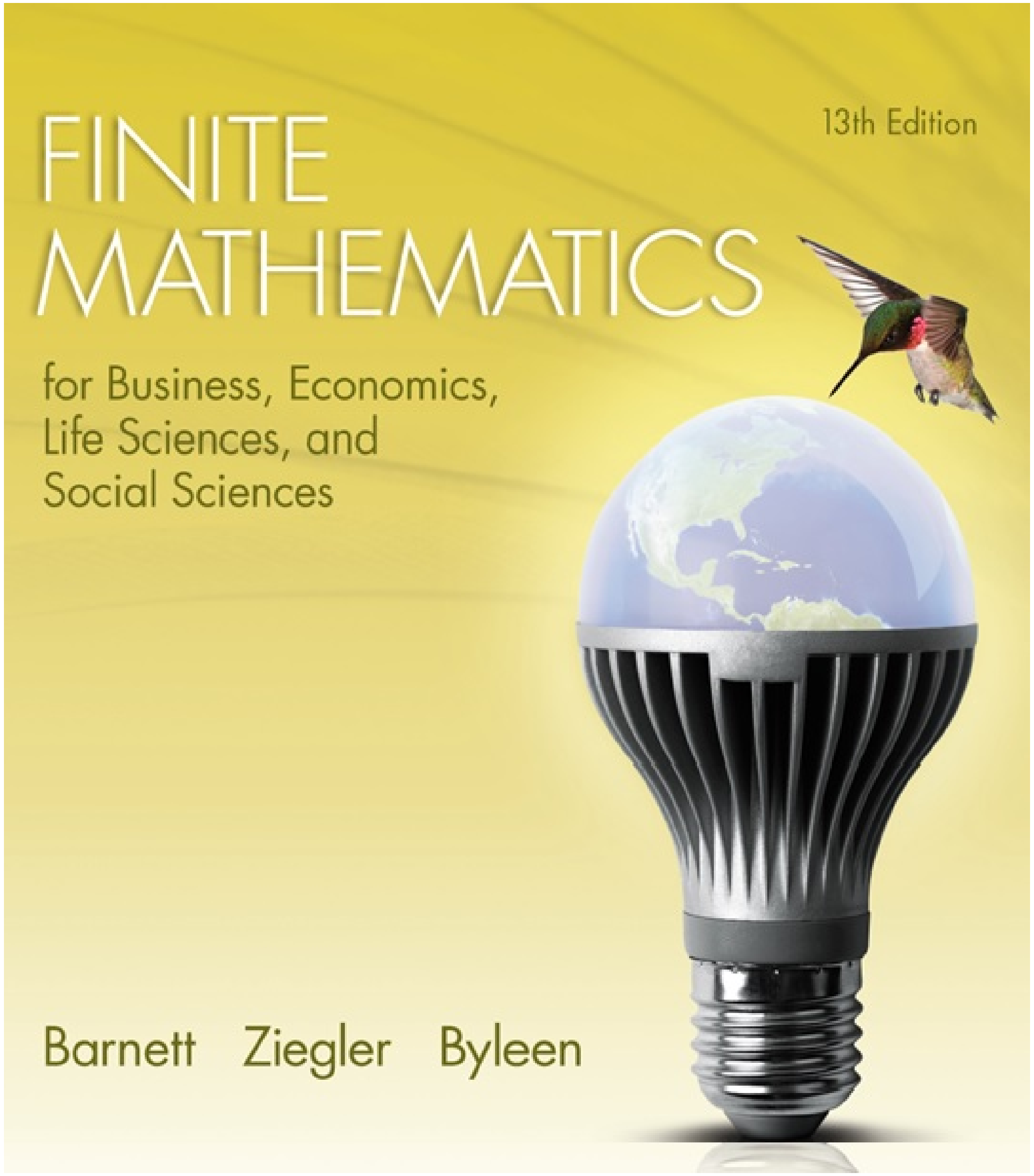


Chapter 11

Data Descriptions and Probability Distributions

Section 3

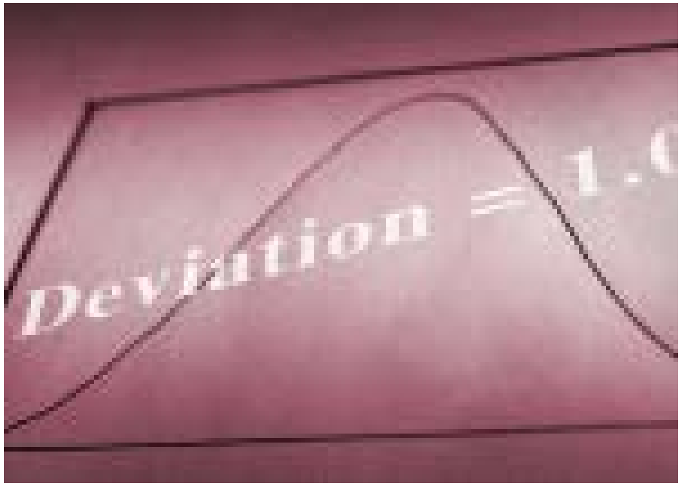
Measures of Dispersion

Learning Objectives for Section 11.3 Measures of Dispersion

- The student will be able to compute the range of a set of data.
- The student will be able to compute the standard deviation for both grouped and ungrouped data.
- The student will be able to interpret the significance of standard deviation.

11.3 Measures of Dispersion

In this section, you will study measures of variability of data. In addition to being able to find measures of central tendency for data, it is also necessary to determine how "spread out" the data is. Two measures of variability of data are the **range** and the **standard deviation**.

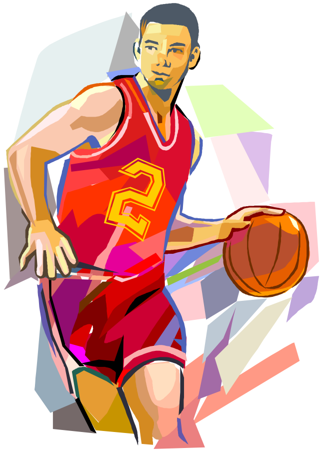


Measures of Variation

- **Example 1.** Heights (in inches) of 5 starting players from two basketball teams:

  A: 72 , 73, 76, 76, 78

  B: 67, 72, 76, 76, 84



- Verify that the two teams have the same mean heights, the same median and the same mode.

Range

- To describe the difference in the two data sets, we use a descriptive measure that indicates the amount of spread or variablility or dispersion in a data set.
- Definition: the **range** is the difference between maximum and minimum values of the data set.
- **Example 1 (continued)**
  - Range of team A: 78-72=6
  - Range of team B: 84-67=17
- Advantage of range: it is easy to compute
- Disadvantage: only two values are considered.

Sample Standard Deviation

- Unlike the range, the **sample standard deviation** takes into account all data values. The following procedure is used to find the sample standard deviation.
- **Step 1.** Find the mean of data.

$$\bar{x} = \frac{\sum_i x_i}{n}$$

Example, Team A:

$$\frac{72+73+76+76+78}{5} = 75$$

Sample Standard Deviation (continued)

- **Step 2.** Find the deviation of each score from the mean
- Note that the sum of the deviations is zero:

$$\sum(x-\bar{x}) = 0$$

x	$x-\bar{x}$
72	72-75 = -3
73	73–75 = -2
76	76-75 = 1
76	76-75 = 1
78	78-75= 3

Sample Standard Deviation (continued)

- **Step 3.** Square each deviation from the mean. Find the sum of the squared deviations.

$$\sum_{i=1}^{n}(X_i - \bar{X})^2$$

x_i	$\bar{x} - x_i$	$(\bar{x} - x_i)^2$
72	72-75 = -3	9
73	73–75 = -2	4
76	76-75 = 1	1
76	76-75 = 1	1
78	78-75= 3	9
Σ	0	24

Sample Variance

- **Step 4.** The **sample variance** is determined by dividing the sum of the squared deviations by (n – 1) (number of scores minus one)

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

Example: For Team A, the sample variance is

$$\frac{24}{5-1} = 6$$

Sample Standard Deviation

- **Step 5.** The **standard deviation** is the square root of the variance.
- The mathematical formula for the **sample standard deviation** is

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

Example: the sample standard deviation for Team A is $\sqrt{6}$

Procedure to Calculate Sample Standard Deviation

1. Find the mean of the data.

2. Set up a table with 3 columns which lists the data in the left hand column and the deviations from the mean in the next column.

3. In the third column, square each deviation and then find the sum of the squares of the deviations.

4. Divide the sum of the squared deviations by (n – 1).

5. Take the positive square root of the result.

Example
Solution

- Find the sample variance and standard deviation of the data
  5, 8, 9, 7, 6 (by hand)
- **Answer:** Variance is approximately 2.496. Standard deviation is approximately 1.581.

Standard Deviation for Grouped Data

1. Find each class midpoint x_i and compute the mean.
2. Find the deviation of each class midpoint from the mean
3. Each deviation is squared and then multiplied by the class frequency.
4. Find the sum of these values and divide the result by (n – 1) (one less than the total number of observations).
5. Take the square root.

$$s = \sqrt{\frac{\sum_{i=1}^{k}(x_i - \bar{x})^2 \cdot f_i}{n-1}}$$

Example of Standard Deviation for Grouped Data

This frequency distribution represents the number of rounds of golf played by a group of golfers. Mean is 2201.5/75=29.3533.

Class	Midpoint	Freq.	$x_i \cdot f_i$	$(\bar{x} - x_i)^2 \cdot f_i$
[0,7)	3.5	0	0	0
[7,14)	10.5	2	21	710.8963556
[14,21)	17.5	10	175	1405.015111
[21,28)	24.5	21	514.5	494.6517333
[28,35)	31.5	23	724.5	105.9880889
[35,42)	38.5	14	539	1171.261156
[42,49)	45.5	5	227.5	1303.574222
Σ		75	2201.5	5191.386667

Example (continued)

$$s = \sqrt{\frac{\sum_{i=1}^{k} (x_i - \overline{x})^2 \cdot f_i}{n-1}} = \sqrt{\frac{5191.386667}{74}}$$

$$= 8.37579094$$

Interpreting the Standard Deviation

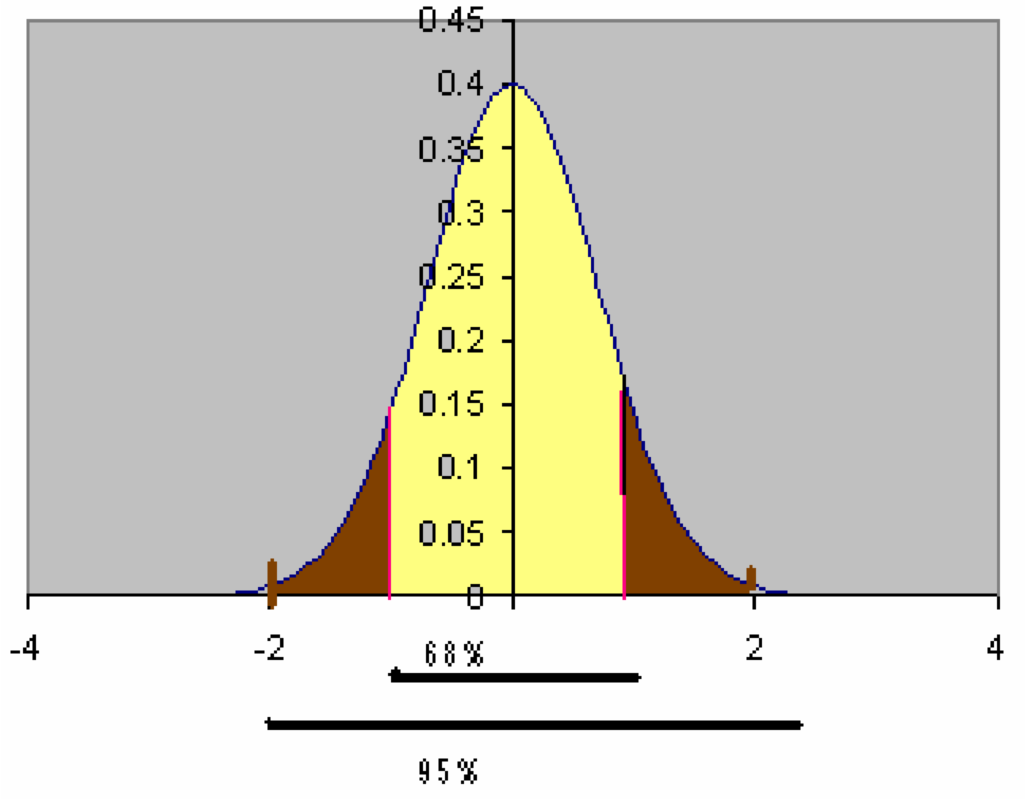
- The more variation in a data set, the greater the standard deviation.
- The larger the standard deviation, the more "spread" in the shape of the histogram representing the data.
- Standard deviation is used for quality control in business and industry. If there is too much variation in the manufacturing of a certain product, the process is out of control and adjustments to the machinery must be made to insure more uniformity in the production process.

Empirical Rule

- If a data set is "mound-shaped" or "bell-shaped", then
  - approximately 68% of the data lies within one standard deviation of the mean
  - about 95% of the data lies within two standard deviations of the mean.
  - about 99.7 % of the data falls within three standard deviations of the mean.
- This means that "almost all" the data will lie within 3 standard deviations of the mean, that is, in the interval determined by $(\bar{x}-3s, \bar{x}+3s)$.

Empirical Rule (continued)

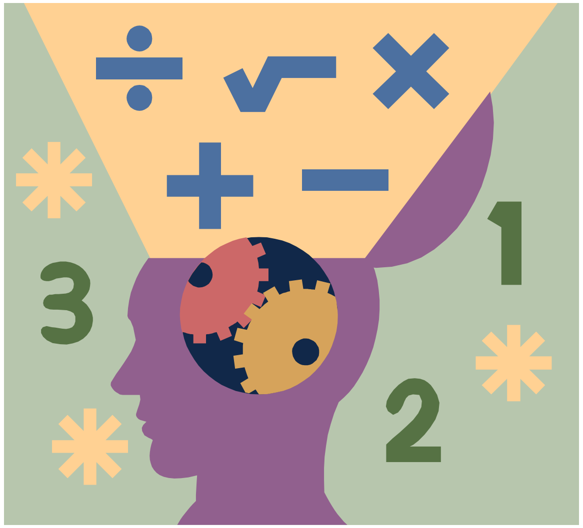
- Yellow region is 68% of the total area. This includes all data within one standard deviation of the mean.
- Yellow region plus brown regions include 95% of the total area. This includes all data that are within two standard deviations from the mean.



Example of Empirical Rule

The shape of the distribution of IQ scores is a mound shape with a mean of 100 and a standard deviation of 15.

A) What proportion of individuals have IQ's ranging from 85 – 115?



B) between 70 and 130?

C) between 55 and 145?

Example of Empirical Rule

The shape of the distribution of IQ scores is a mound shape with a mean of 100 and a standard deviation of 15.

A) What proportion of individuals have IQ's ranging from 85 – 115? (about 68%)

B) between 70 and 130? (about 95%)

C) between 55 and 145? (about 99.7%)