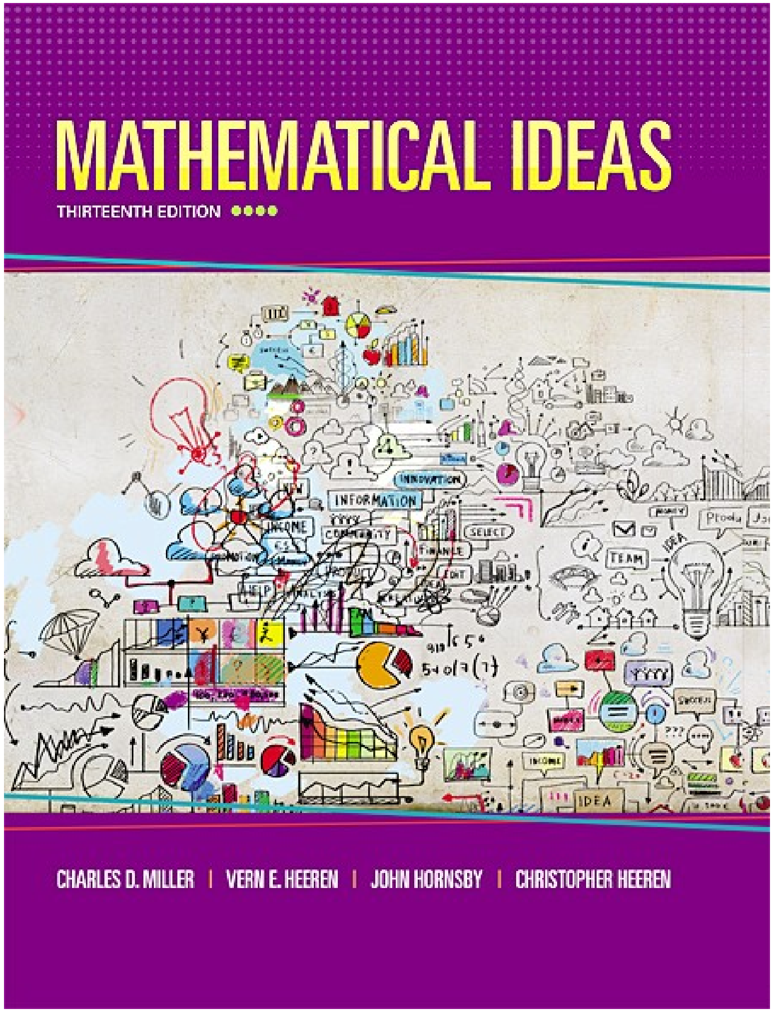
Chapter 12

Statistics



Chapter 12: Statistics

Section 12-3

Measures of Dispersion

Measures of Dispersion

- Find the range of a data set.
- Calculate the standard deviation of a data set.
- Interpret measures of dispersion.
- Calculate the coefficient of variation.

Measures of Dispersion

Sometimes we want to look at a measure of **dispersion**, or *spread*, of data. Two of the most common measures of dispersion are the *range* and the *standard deviation*.

Range

For any set of data, the **range** of the set is given by

Range = (greatest value in set) – (least value in set).

Example: Finding Range Values

The two sets below have the same mean and median (7). Find the range of each set.

Set A	1	2	7	12	13
Set B	5	6	7	8	9

Solution

Range of Set A: 13 – 1 = 12.

Range of Set B: 9 – 5 = 4.

Standard Deviation

One of the most useful measures of dispersion, the *standard deviation*, is based on *deviations from the mean* of the data.

Example: Finding Deviations from the Mean

Find the deviations from the mean for all data values of the sample 1, 2, 8, 11, 13.

Solution

The mean is 7. Subtract to find deviation.

Data Value	1	2	8	11	13
Deviation	–6	–5	1	4	6

13 – 7 = 6

The sum of the deviations for a set is always 0.

Standard Deviation

The **variance** is found by summing the squares of the deviations and dividing that sum by $n – 1$ (since it is a sample instead of a population). The square root of the variance gives a kind of average of the deviations from the mean, which is called a sample **standard deviation**. It is denoted by the letter s. (The standard deviation of a population is denoted σ, the lowercase Greek letter sigma.)

Calculation of Standard Deviation

Let a sample of n numbers $x_1, x_2, \ldots x_n$ have mean $\bar{x}$. Then the **sample standard deviation**, ***s***, of the numbers is given by

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

Calculation of Standard Deviation

The individual steps involved in this calculation are as follows

Step 1 Calculate the mean of the numbers.

Step 2 Find the deviations from the mean.

Step 3 Square each deviation.

Step 4 Sum the squared deviations.

Step 5 Divide the sum in Step 4 by $n - 1$.

Step 6 Take the square root of the quotient in Step 5.

Example: Finding a Sample Standard Deviation

Find the standard deviation of the sample
1, 2, 8, 11, 13.

Solution

The mean is 7.

Data Value	1	2	8	11	13
Deviation	–6	–5	1	4	6
$(\text{Deviation})^2$	36	25	1	16	36

Sum = 36 + 25 + 1 + 16 + 36 = 114

Example: Finding a Sample Standard Deviation

Solution (continued)

Divide by $n - 1$ with $n = 5$:

$$\frac{114}{5-1} = \frac{114}{4} = 28.5.$$

Take the square root:

$$\sqrt{28.5} \approx 5.34.$$

Interpreting Measures of Dispersion

A main use of dispersion is to compare the amounts of spread in two (or more) data sets. A common technique in inferential statistics is to draw comparisons between populations by analyzing samples that come from those populations.

Example: Comparing Measures

Two companies, *A* and *B*, sell small packs of sugar for coffee. The mean and standard deviation for samples from each company are given below. Which company consistently provides more sugar in their packs? Which company fills its packs more consistently?

Company *A*	Company *B*
$\bar{x}_A = 1.013$ tsp	$\bar{x}_B = 1.007$ tsp
$s_A = .0021$	$s_B = .0018$

Example: Comparing Measures

Solution

We infer that Company *A* most likely provides more sugar than Company *B* (greater mean).

We also infer that Company *B* is more consistent than Company *A* (smaller standard deviation).

Chebyshev's Theorem

For any set of numbers, regardless of how they are distributed, the fraction of them that lie within k standard deviations of their mean (where $k > 1$) is *at least*

$$1 - \frac{1}{k^2}.$$

Example: Apply Chebyshev's Theorem

What is the minimum percentage of the items in a data set which lie within 3 standard deviations of the mean?

Solution

With k = 3, we calculate

$$1-\frac{1}{3^2}=1-\frac{1}{9}=\frac{8}{9}\approx .889=88.9\%.$$

Coefficient of Variation

The *coefficient of variation* is not strictly a measure of dispersion; it combines central tendency and dispersion. It expresses the standard deviation as a percentage of the mean.

Coefficient of Variation

For any set of data, the **coefficient of variation** is given by

$$V = \frac{s}{\bar{x}} \cdot 100 \text{ for a sample or}$$

$$V = \frac{\sigma}{\mu} \cdot 100 \text{ for a population.}$$

Example: Comparing Samples

Compare the dispersions in the two samples *A* and *B*.

A: 12, 13, 16, 18, 18, 20

B: 125, 131, 144, 158, 168, 193

Solution

The values on the next slide are computed using a calculator and the formula on the previous slide.

Example: Comparing Samples

Solution (continued)

Sample A	Sample B
$\bar{x}_A = 16.167$	$\bar{x}_B = 153.167$
$s_A = 3.125$	$s_B = 25.294$
$V_A = 19.3$	$V_B = 16.5$

Sample *B* has a larger dispersion than sample *A*, but sample *A* has the larger relative dispersion (coefficient of variation).