MEASURES OF DISPERSION

Math 1001

Quantitative Skills and Reasoning

Columbus State University

MEASURES OF DISPERSION

- In the preceding section we introduced three types of average values for a data set:
  - Mean
  - Median
  - Mode
- However, some characteristics of a set of data may not be evident from an examination of averages.

MEASURES OF DISPERSION

- Consider a soft-drink dispensing machine that should dispense 8 oz of your selection into a cup.
- The following table shows data for two of these machines:

MACHINE 1	MACHINE 2
8.68	8.21
6.73	7.50
10.39	7.55
5.95	8.32
8.25	8.42
$\bar{x} =$ 8.0	$\bar{x} =$ 8.0

MEASURES OF DISPERSION

- The mean data value for each machine is 8 oz.
- However, look at the variation in data values for Machine 1.
- The quantity of soda dispensed is very inconsistent – in some cases the soda machine overflows the cup, and in other cases too little soda is dispensed.

MACHINE 1	MACHINE 2
8.68	8.21
6.73	7.50
10.39	7.55
5.95	8.32
8.25	8.42
$\bar{x}$ = 8.0	$\bar{x}$ =8.0

- Machine 1 clearly needs to be adjusted.
- Machine 2, on the other hand, is working just fine.
- The quantity dispensed is very consistent, with little variation.
- This example shows that average values do not reflect the *spread* or *dispersion* of data.
- To measure the spread of dispersion of data, we must introduce statistical values known as the *range* and the *standard deviation*.

Columbus State University

THE RANGE

- The **range** of a set of data values is the difference between the greatest data value and the least data value.

- Find the range of the numbers of ounces dispensed by Machine 1.

MACHINE 1
8.68
6.73
10.39
5.95
8.25

The greatest number of ounces dispensed is 10.39 and the smallest is 5.95. The range of the numbers of ounces is 10.39 – 5.95 = 4.44 oz.

THE STANDARD DEVIATION

- The range of a set of data is easy to compute, but it can be deceiving.
- The range is a measure that depends only on the two most extreme values, and as such it is very sensitive.

- A measure of dispersion that is less sensitive to extreme values is the *standard deviation*.
- The standard deviation of a set of numerical data makes use of the individual amount that each data value deviates from the mean.

THE STANDARD DEVIATION

- These deviations, represented by $(x - \bar{x})$, are positive when the data value x is greater than the mean $\bar{x}$, and are negative when x is less than the mean $\bar{x}$.
- The sum of all the deviations $(x - \bar{x})$ is 0 for all sets of data.

x	$x - \bar{x}$
8.21	8.21 – 8.0 = 0.21
7.50	7.50 – 8.0 = -0.5
7.55	7.55 – 8.0 = -0.45
8.32	8.32 – 8.0 = 0.32
8.42	8.42 – 8.0 = 0.42
Sum of deviations	= 0

- This is shown for Machine 2 data here.

THE STANDARD DEVIATION

- Because the sum of all the deviations of the data values from the mean is *always* 0, we cannot use the sum of the deviations as a measure of dispersion for the set of data.
- Instead, the standard deviation uses the sum of the *squares* of the deviations.

STANDARD DEVIATION FOR POPULATIONS AND SAMPLES

- If $x_1, x_2, x_3, \ldots, x_n$ is a *population* of *n* numbers with a mean of μ, then the **standard deviation** of the population is

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

- If $x_1, x_2, x_3, \ldots, x_n$ is a *sample* of *n* numbers with a mean of $\bar{x}$, then the **standard deviation** of the sample is

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

COLUMBUS STATE UNIVERSITY

THE STANDARD DEVIATION

- Most statistical applications involve a sample rather than a population, which is the complete set of data values.
- Sample standard deviations are designated by the lowercase letter *s*.
- In those cases in which we *do* work with a population, we designate the standard deviation of the population by a σ, which is the lowercase Greek letter sigma.

Columbus State University