

1. Measuring "Spread"

In addition to knowing the center of a distribution of data, it is important to know the "spread" or how far from the center data tends to be.

2. The Standard Deviation (SD or just s)

The Standard Deviation may be thought of as the average size deviations of the individual values of a list from the overall average (the mean) of the list. So it's based on the mean.

a. In general, most numbers in a list are only one standard deviation away from the average. A few numbers will be two standard deviations away. And very few will deviate beyond that. STANDARD DEVIATION is abbreviated as SD or as a lowercase "s".

b. The SD is defined as follows: given a list of n numbers x_1, x_2, \dots, x_n ,

$$s = \sqrt{\frac{(x_1 - \bar{x})^2 + (x_2 - \bar{x})^2 + \dots + (x_n - \bar{x})^2}{n}}$$

or simplified

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

3. An Example

An interview with 5 UCLA students reveals the time in hours spent surfing the web in a given week: 7, 17, 3, 14, 7

| | | | | | | | |
|-----------------------------|------|-------|-------|-------|------|-------------|--------------------------------|
| Value | 7 | 17 | 3 | 14 | 7 | Sum = 48 | Mean = 48/5 = 9.6 |
| Deviation from Mean | -2.6 | +7.4 | -6.6 | +4.4 | -2.6 | Sum = 0 | |
| Squared Deviation from Mean | 6.76 | 54.76 | 43.56 | 19.36 | 6.76 | Sum = 131.2 | SD = $\sqrt{131.2/5} = 5.1225$ |

4. More Properties of the Standard Deviation

- Note that the standard deviation is in the same units as the data. This is why the standard deviation involves the square root, it allows for easier interpretation. It is always a positive value.
- As with the mean, it is not necessary to know HOW MANY items are in a list when computing a standard deviation, only the relative frequency of the values in the list.
- The SD measures how close the numbers in the list are to the average; i.e., not all numbers are equal to the mean; the SD is a measure of the "average" distance between each point and the overall average.
- Another thing about the standard deviation. For many datasets 68% of the entries on a list will fall within one SD of the average. 95% of the entries will fall within two SD of the average. But we'll talk about this some more in Chapter 5.

5. PRINCIPLES:

- It's not enough to know the mean or average of a variable, it is necessary to know something about spread (standard deviation)
- the standard deviation by itself is not particularly interesting, it's more useful when viewed alongside a mean or in comparison with other standard deviations.

Example 1. Which one is the less riskier investment and why?

| Mutual Fund (as of 3/31/01) | Average Annual Return since 1996 | Standard Deviation |
|---|----------------------------------|--------------------|
| Mercantile Growth & Income Fund (AGIIX) | 47.2% | 19.4% |
| Chase Vista Bond (TRBDX) | 39.2% | 3.6% |

Example 2. Here are some statistics on death in Los Angeles in 1999

| Race/Ethnic Category | Mean Age At Death | Standard Deviation |
|----------------------|-------------------|--------------------|
| Non-Hispanic White | 74.7 | 17.8 |
| African American | 65.8 | 20.9 |
| Hispanic (all races) | 61.5 | 24.8 |
| Asian | 70.9 | 18.9 |
| All Others | 63.1 | 22.5 |

Example 3. Coached and Uncoached SATs (see handout)