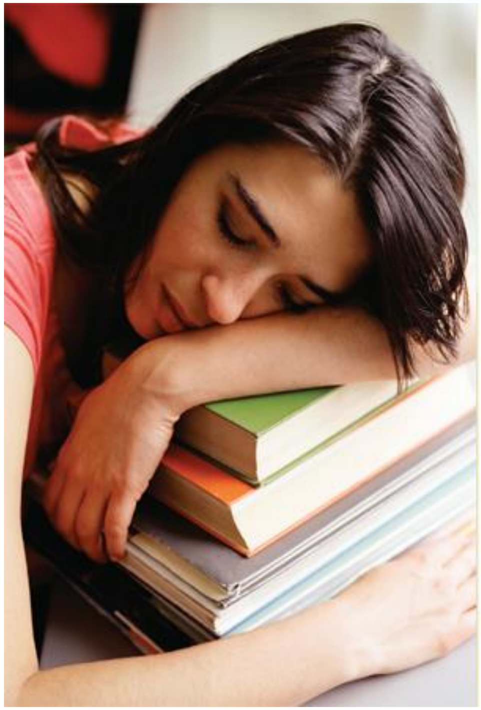
2 Descriptive Analysis and Presentation of Single-Variable Data



2.4 Measures of Dispersion

Measures of Dispersion

The **measures of dispersion** include the *range, variance*, and *standard deviation*. These numerical values describe the amount of spread, or variability, that is found among the data: Closely grouped data have relatively small values, and more widely spread-out data have larger values.

The closest possible grouping occurs when the data have no dispersion (all data are the same value); in this situation, the measure of dispersion will be zero.

Measures of Dispersion

There is no limit to how widely spread out the data can be; therefore, measures of dispersion can be very large. The simplest measure of dispersion is the range.

Range The difference in value between the highest-valued data, H, and the lowest-valued data, L:

$$\text{range} = \text{high value} - \text{low value}$$

$$\text{range} = H - L \qquad \textbf{(2.4)}$$

Measures of Dispersion

Deviation from the mean A deviation from the mean, $x - \bar{x}$, is the difference between the value of *x* and the mean, $\bar{x}$.

Each individual value of *x* deviates from the mean by an amount equal to $(x - \bar{x})$. This deviation $(x - \bar{x})$ is zero when *x* is equal to the mean, $\bar{x}$. The deviation $(x - \bar{x})$ is positive when *x* is larger than $\bar{x}$ and negative when *x* is smaller than $\bar{x}$.

Measures of Dispersion

The sum of the deviations, $\Sigma(x - \overline{x})$, is always zero because the deviations of *x* values smaller than the mean (which are negative) cancel out those *x* values larger than the mean (which are positive).

We can remove this neutralizing effect if we do something to make all the deviations positive. This can be accomplished by squaring each of the deviations; squared deviations will all be nonnegative (positive or zero) values. The squared deviations are used to find the *variance*.

Measures of Dispersion

Sample variance The sample variance, s^2, is the mean of the squared deviations, calculated using n – 1 as the divisor:

$$\text{sample variance:} \quad s\ squared = \frac{sum\ of\ (deviations\ squared)}{number - 1}$$

$$s^2 = \frac{\sum(x - \bar{x})^2}{n - 1} \qquad \textbf{(2.5)}$$

where n is the sample size—that is, the number of data in the sample.

Measures of Dispersion

Notes

1. The sum of all the *x* values is used to find $\bar{x}$.
2. The sum of the deviations, $\Sigma(x - \bar{x})$, is always zero, provided the exact value of $\bar{x}$ is used.
3. If a rounded value of $\bar{x}$ is used, then $\Sigma(x - \bar{x})$ will not always be exactly zero. It will, however, be reasonably close to zero.
4. The sum of the squared deviations is found by squaring each deviation and then adding the squared values.

Measures of Dispersion

To graphically demonstrate what variances of data sets are telling us, consider a second set of data: {1, 3, 5, 6, 10}. Note that the data values are more dispersed than the data values in Table 2.12.

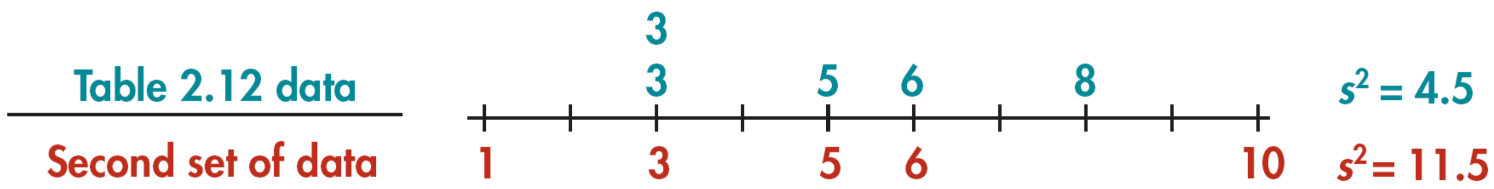
Step 1. Find Σx	Step 2. Find $\bar{x}$	Step 3. Find each $x - \bar{x}$	Step 4. Find $\Sigma(x - \bar{x})^2$	Step 5. Find s^2
6	$\bar{x} = \frac{\Sigma x}{n}$	$6 - 5 = 1$	$(1)^2 = 1$	$s^2 = \frac{\Sigma(x - \bar{x})}{n - 1}$
3		$3 - 5 = -2$	$(-2)^2 = 4$	
8		$8 - 5 = 3$	$(3)^2 = 9$	
5	$\bar{x} = \frac{25}{5}$	$5 - 5 = 0$	$(0)^2 = 0$	$s^2 = \frac{18}{4}$
3		$3 - 5 = -2$	$(-2)^2 = 4$	
$\Sigma x = 25$	$\bar{x} = 5$	$\Sigma(x - \bar{x}) = 0$	$\Sigma(x - \bar{x})^2 = 18$	$s^2 = 4.5$

Calculating Variance Using Formula (2.5)

Table 2.12

Measures of Dispersion

Accordingly, its calculated variance is larger at $s^2 = 11.5$. An illustrative side-by-side graphical comparison of these two samples and their variances is shown in Figure 2.23.



Comparison of Data

Figure 2.23

Measures of Dispersion

Sample standard deviation The standard deviation of a sample, *s*, is the positive square root of the variance:

sample standard deviation: *s* = *square root of sample variance*

$$s = \sqrt{s^2} \quad \textbf{(2.6)}$$

The numerator for the sample variance, $\Sigma(x - \overline{x})^2$, is often called the *sum of squares for x* and symbolized by SS(*x*).

Measures of Dispersion

Thus, formula (2.5) can be expressed as

$$\text{sample variance: } s^2 = \frac{\mathrm{SS}(x)}{n-1} \qquad \textbf{(2.7)}$$

where SS(*x*) = $\Sigma(x - \overline{x})^2$.

Step 1. Find Σx	Step 2. Find $\overline{x}$	Step 3. Find each $x - \overline{x}$	Step 4. Find $\Sigma(x - \overline{x})^2$	Step 5. Find s^2
6	$\overline{x} = \frac{\Sigma x}{n}$	$6 - 4.8 = 1.2$	$(1.2)^2 = 1.44$	$s^2 = \frac{\Sigma(x-\overline{x})^2}{n-1}$
3		$3 - 4.8 = -1.8$	$(-1.8)^2 = 3.24$	
8		$8 - 4.8 = 3.2$	$(3.2)^2 = 10.24$	
5	$\overline{x} = \frac{24}{5}$	$5 - 4.8 = 0.2$	$(0.2)^2 = 0.04$	$s^2 = \frac{22.80}{4}$
2		$2 - 4.8 = -2.8$	$(-2.8)^2 = 7.84$	
$\Sigma x = 24$	$\overline{x} = 4.8$	$\Sigma(x - \overline{x}) = 0$	$\Sigma(x - \overline{x})^2 = 22.80$	$s^2 = 5.7$

Calculating Variance Using Formula (2.5)

Table 2.13

Measures of Dispersion

The arithmetic for this example has become more complicated because the mean contains nonzero digits to the right of the decimal point.

However, the "sum of squares for *x*," the numerator of formula (2.5), can be rewritten so that $\bar{x}$ is not included:

Sum of Squares for *x*

$$SS(x) = \Sigma x^2 - \frac{(\Sigma x)^2}{n} \qquad \textbf{(2.8)}$$

Measures of Dispersion

Sample Variance, “Short-Cut Formula”

$$s \quad squared = \frac{(sum\ of\ x^2) - \left[\frac{(sum\ of\ x)^2}{number}\right]}{number - 1}$$

$$\text{sample variance: } s^2 = \frac{\sum x^2 - \frac{(\sum x)^2}{n}}{n-1} \tag{2.9}$$

Measures of Dispersion

Step 1. Find Σx	Step 2. Find Σx^2	Step 3. Find SS(x)	Step 4. Find s^2	Step 5. Find s
6	$6^2 = 36$	$SS(x) = \Sigma x^2 - \frac{(\Sigma x)^2}{n}$		$s = \sqrt{s^2}$
3	$3^2 = 9$		$s^2 = \frac{\Sigma x^2 - \frac{(\Sigma x)^2}{n}}{n - 1}$	$s = \sqrt{5.7}$
8	$8^2 = 64$	$SS(x) = 138 - \frac{(24)^2}{5}$		$s = 2.4$
5	$5^2 = 25$		$s^2 = \frac{22.8}{4}$	
2	$2^2 = 4$	$SS(x) = 138 - 115.2$		
$\Sigma x = 24$	$\Sigma x^2 = 138$	$SS(x) = 22.8$	$s^2 = 5.7$	

Calculating Standard Deviation Using the Shortcut Method

Table 2.14

Measures of Dispersion

Standard deviation on your calculator Most calculators have two formulas for finding the standard deviation and mindlessly calculate both, fully expecting the user to decide which one is correct for the given data. How do you decide?

The sample standard deviation is denoted by s and uses the “divide by $n – 1$” formula.

The population standard deviation is denoted by σ and uses the “divide by n” formula.

Measures of Dispersion

When you have sample data, always use the *s* or “divide by *n* – 1” formula. Having the population data is a situation that will probably never occur, other than in a textbook exercise.

If you don’t know whether you have sample data or population data, it is a “safe bet” that they are sample data—use the *s* or “divide by *n* – 1” formula!

Measures of Dispersion

Multiple formulas Statisticians have multiple formulas for convenience—that is, convenience relative to the situation. The following statements will help you decide which formula to use:

1. When you are working on a computer and using statistical software, you will generally store all the data values first.

   The computer handles repeated operations easily and can "revisit" the stored data as often as necessary to complete a procedure.

Measures of Dispersion

The computations for sample variance will be done using formula (2.5), following the process shown in Table 2.12.

Step 1. Find Σx	Step 2. Find $\bar{x}$	Step 3. Find each $x - \bar{x}$	Step 4. Find $\Sigma(x - \bar{x})^2$	Step 5. Find s^2
6	$\bar{x} = \frac{\Sigma x}{n}$	$6 - 5 = 1$	$(1)^2 = 1$	$s^2 = \frac{\Sigma(x - \bar{x})}{n - 1}$
3		$3 - 5 = -2$	$(-2)^2 = 4$	
8		$8 - 5 = 3$	$(3)^2 = 9$	
5	$\bar{x} = \frac{25}{5}$	$5 - 5 = 0$	$(0)^2 = 0$	$s^2 = \frac{18}{4}$
3		$3 - 5 = -2$	$(-2)^2 = 4$	
$\Sigma x = 25$	$\bar{x} = 5$	$\Sigma(x - \bar{x}) = 0$	$\Sigma(x - \bar{x})^2 = 18$	$s^2 = 4.5$

Calculating Variance Using Formula (2.5)

Table 2.12

Measures of Dispersion

2. When you are working on a calculator with built-in statistical functions, the calculator must perform all necessary operations on each data value as the values are entered (most handheld nongraphing calculators do not have the ability to store data).

   Then after all data have been entered, the computations will be completed using the appropriate summations.

Measures of Dispersion

The computations for sample variance will be done using formula (2.9), following the procedure shown in Table 2.14.

Step 1. Find Σx	Step 2. Find Σx^2	Step 3. Find SS(x)	Step 4. Find s^2	Step 5. Find s
6	$6^2 = 36$	$SS(x) = \Sigma x^2 - \frac{(\Sigma x)^2}{n}$		$s = \sqrt{s^2}$
3	$3^2 = 9$		$s^2 = \frac{\Sigma x^2 - \frac{(\Sigma x)^2}{n}}{n - 1}$	$s = \sqrt{5.7}$
8	$8^2 = 64$	$SS(x) = 138 - \frac{(24)^2}{5}$		$s = 2.4$
5	$5^2 = 25$		$s^2 = \frac{22.8}{4}$	
2	$2^2 = 4$	$SS(x) = 138 - 115.2$		
$\Sigma x = 24$	$\Sigma x^2 = 138$	$SS(x) = 22.8$	$s^2 = 5.7$	

Calculating Standard Deviation Using the Shortcut Method

Table 2.14

Measures of Dispersion

3. If you are doing the computations either by hand or with the aid of a calculator, but not using statistical functions, the most convenient formula to use will depend on how many data there are and how convenient the numerical values are to work with.