05-Sirkin-4731.qxd 6/9/2005 6:40 PM Page 127

CHAPTER 5

Measuring Dispersion

▼ PROLOGUE ▼

Comparing two groups by a measure of central tendency may run the risk for each group of failing to reveal valuable information. In particular, information about the distribution of the scores within each group may be useful to us but not revealed by the mean, median, or mode. In some groups, the scores may all fall near the middle score, whereas in other groups, the scores may be more widely spread above and below the central scores. Accordingly, it is possible that the more bigoted group of the two we compared, using a measure of central tendency, might contain some highly bigoted individuals but possibly also several less bigoted people than could be found in the less bigoted group. So in addition to central tendency, we should examine the dispersion of the scores in each group as well.

05-Sirkin-4731.qxd 6/9/2005 6:40 PM Page 128

INTRODUCTION

In addition to finding measures of central tendency for a set of scores, we also calculate measures of dispersion to aid us in describing the data. **Measures of dispersion,** also called *measures of variability,* address the degree of clustering of the scores about the mean. Are most scores relatively close to the mean, or are they scattered over a wider interval and thus farther from the mean? The extent of clustering or spread of the scores about the mean determines the amount of **dispersion.** In the instance where all scores are exactly at the mean, there is no dispersion at all; dispersion increases from zero as the spread of scores widens about the mean. In this chapter, we will cover four measures of dispersion: the range, the mean deviation, the variance, and the standard deviation.

Measures of dispersion Measures of variability that address the degree of clustering of the scores about the mean.

Dispersion The extent of clustering or spread of the scores about the mean.

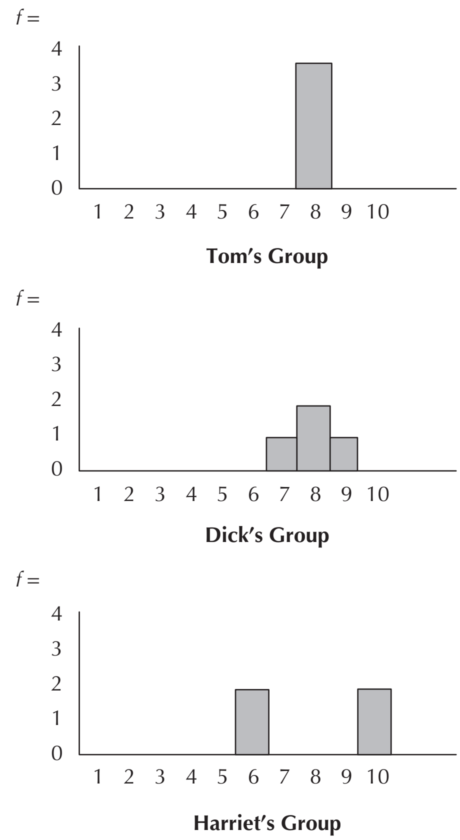
VISUALIZING DISPERSION

To begin our discussion, let us suppose that in a penology class, three teaching assistants—Tom, Dick, and Harriet—had their respective discussion groups role-play court-employed social case workers who read the files of convicted criminals and recommended to the judge the penalty to be imposed for each criminal. The teaching assistants then compared each student's recommended sentence to the one actually imposed by the real judge. The teaching assistants then rated each student on a 0 to 10 scale, with 10 being a totally accurate reproduction of the sentences that were actually handed down. There were four students in each discussion group. The results were as follows:

Tom's Group	*Dick's Group*	*Harriet's Group*
$x =$	$x =$	$x =$
8	9	10
8	8	10
8	8	6
8	7	6
$\sum x = 32$	$\sum x = 32$	$\sum x = 32$
$\bar{x}_{\text{Tom}} = \frac{32}{4} = 8$	$\bar{x}_{\text{Dick}} = \frac{32}{4} = 8$	$\bar{x}_{\text{Harriet}} = \frac{32}{4} = 8$

05-Sirkin-4731.qxd 6/9/2005 6:40 PM Page 129

The three groups share the same mean, but the dispersion of the scores varies from none in Tom's group to some in Dick's group to even more in Harriet's group. This is illustrated in the histograms to the left. Because the distribution of individual scores clearly differed from each other in terms of their dispersion, we need to measure that dispersion in addition to measuring central tendency.



In this chapter, we will discuss measures of dispersion in an order that will ultimately bring us to the two measures used to the virtual exclusion of the others, the *variance* and its positive square root, the *standard deviation.* The first two measures we will discuss, the *range* and the *mean deviation,* may be thought of as building blocks for understanding the variance and standard deviation. Since such measures are rarely used with data having a level of measurement less sophisticated than interval level, they are usually calculated along with the calculation of the mean. With the mean as our measure of central tendency, we then calculate a measure of dispersion, most often the standard deviation.

THE RANGE

The **range** is the simplest measure of dispersion. It compares the highest score and the lowest score achieved for a given set of scores. The range can be expressed in two ways: (a) with a statement such as, "The scores ranged from (the lowest score) to (the highest score)," or (b) with a single number representing the difference between the highest and lowest score.

Range The simplest measure of dispersion that compares the highest score and the lowest score achieved for a given set of scores.

In the case of Harriet's group, whose scores were 6, 6, 10, and 10, we would say, "The scores ranged from 6 to 10." Or we could express the range as the difference between 6 and 10 (10 – 6) or 4. "The scores in Harriet's group had a mean of 8 and range of 4." Now we can compare the ranges of the three groups.

05-Sirkin-4731.qxd 6/9/2005 6:40 PM Page 130

- Harriet's Group: Scores ranged from 6 to 10. Range = 10 − 6 = 4.
- Dick's Group: Scores ranged from 7 to 9. Range = 9 − 7 = 2.
- Tom's Group: Scores ranged from 8 to 8. Range = 8 − 8 = 0.

These ranges correspond to the spread on the histograms for the three groups, with Harriet's group's scores being most dispersed about the mean, Dick's being less dispersed, and Tom's having no dispersion at all.

Although we commonly make use of the range in our day-to-day discourse, it really is not a very meaningful measure of dispersion. Because only the highest and lowest scores are taken into consideration in finding the range, the other scores have no impact. Just as in the case of the mean where an extreme value of x can distort the mean and lessen its usefulness, the use of only the extreme values can render the range less useful. Our next measure, the mean deviation, rectifies this situation.

THE MEAN DEVIATION

The **mean deviation** (*M.D.*) (also called the *average deviation* or the *mean absolute deviation*) is sensitive to every score in the set. It is based on a strategy of first finding out how far each score deviated from the mean of the scores (the distance from each score to the mean), summing these distances to find the total amount of deviation from the mean in the entire set of scores, and dividing by the number of scores in the set. The result is a mean, or "average," distance that a score deviates from the mean.

Mean deviation An average distance that a score deviates from the mean.

To get the mean deviation, we first find the distance between each score and the mean by subtracting the mean from each score. Let us use Harriet's group as an example.

Harriet's Group

$x =$	$\bar{x} =$	$x - \bar{x} =$
10	8	2
10	8	2
6	8	–2
6	8	–2
$\sum x = 32$		

$$\bar{x} = \frac{32}{4} = 8$$

At this juncture, we encounter a problem: We cannot add up the $x - \bar{x}$ column to get the total amount of deviation in the system. Recalling that the mean is the value of x that satisfies the expression $\sum(x - \bar{x}) = 0$, we can see that if $\bar{x} = 8$, adding algebraically, the $x - \bar{x}$'s for each student in Harriet's group produce a sum of zero:

$$\sum(x - \bar{x}) = 2+2-2-2 = 4-4 = 0$$

This is because the positive deviations (where x is greater than the mean) exactly balance the negative deviations (where x is less than the mean).

Recall that we currently are seeking the distance from each score to the mean, without regard to direction; that is, we do not care whether x is greater or less than $\bar{x}$. Like a car's odometer, we want to count the distances traveled, disregarding the direction or directions in which we drove. We do this by taking the **absolute value** of each $x - \bar{x}$, the distance disregarding its sign (in effect treating all $x - \bar{x}$ s as if they were positive numbers). We symbolize the absolute value of a deviation as $|x - \bar{x}|$. When we add up all these absolute values, $\sum|x - \bar{x}|$, we get the total amount of deviation of the scores from the mean. When we divide that sum by the total number of scores, we get the "average" amount (the mean amount) that a score deviated from the mean of all of the scores: the mean deviation.

Absolute value The distance or difference disregarding its sign. Here, the distance between each value of *x* and the mean, regardless of whether *x* is greater than the mean (a positive distance) or less than the mean (a negative distance).

Thus,

$$M.D. = \frac{\sum|x - \bar{x}|}{n}$$

For Harriet's Group:

	$x =$	$\bar{x} =$	$x - \bar{x} =$	$\lvert x - \bar{x}\rvert =$
	10	8	2	2
	10	8	2	2
	6	8	–2	2
$n = 4$	6	8	–2	2
	$\sum x = 32$			$\sum\lvert x - \bar{x}\rvert = 8$

$$\bar{x} = \frac{32}{4} = 8 \qquad M.D. = \frac{\sum|x - \bar{x}|}{n} = \frac{8}{4} = 2.0$$

05-Sirkin-4731.qxd 6/9/2005 6:40 PM Page 132

For Dick's Group:

	$x =$	$\bar{x} =$	$x - \bar{x} =$	$\lvert x - \bar{x} \rvert =$
	9	8	2	1
	8	8	0	0
	8	8	0	0
$n = 4$	7	8	–1	1
	$\sum x = 32$			$\sum \lvert x - \bar{x} \rvert = 2$

$$\bar{x} = \frac{32}{4} = 8 \qquad M.D. = \frac{\sum \lvert x - \bar{x} \rvert}{n} = \frac{2}{4} = 0.5$$

For Tom's Group:

	$x =$	$\bar{x} =$	$x - \bar{x} =$	$\lvert x - \bar{x} \rvert =$
	8	8	0	0
	8	8	0	0
	8	8	0	0
$n = 4$	8	8	0	0
	$\sum x = 32$			$\sum \lvert x - \bar{x} \rvert = 0$

$$\bar{x} = \frac{32}{4} = 8 \qquad M.D. = \frac{\sum \lvert x - \bar{x} \rvert}{n} = \frac{0}{4} = 0$$

These results are in keeping with our expectations: Harriet's group has the largest mean deviation, Dick's has a smaller one, and Tom's has the smallest (a value of zero).

THE VARIANCE AND STANDARD DEVIATION

The formula for the **variance** resembles that of the mean deviation except that $\sum \lvert x - \bar{x} \rvert$ is replaced by the expression $\sum (x - \bar{x})^2$. Instead of taking the absolute value of each deviation, we square it to get rid of negative numbers. (Remember that a negative number times itself is a positive number, just as a positive number times itself is a positive number.) Since the squares of the deviations greater than one unit will be much larger than their respective absolute values, $\sum (x - \bar{x})^2$ will usually be larger than $\sum \lvert x - \bar{x} \rvert$, and the final variance will usually be larger than the mean deviation. To adjust for this and produce a result more comparable to the mean deviation (more like an "average" amount of deviation), we often take the positive square root of the variance, thus producing the **standard deviation,** indicated for now by the letter *s*.

Thus,

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

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

Variance An "average" or mean value of the squared deviations of the scores from the mean.

Standard deviation The positive square root of the variance, which provides a measure of dispersion closer in size to the mean deviation.

Let us calculate s^2 and s for our three groups—Tom's, Dick's, and Harriet's—whose mean deviations were 0, 0.5, and 2.0, respectively.

Tom's Group

$x =$	$\bar{x} =$	$x - \bar{x} =$	$(x-\bar{x})^2 =$
8	8	0	0
8	8	0	0
8	8	0	0
8	8	0	0
			$\sum(x-\bar{x})^2 = 0$

Thus,

$$s^2 = \frac{\sum(x-\bar{x})^2}{n} = \frac{0}{4} = 0$$

$$s = \sqrt{\frac{\sum(x-\bar{x})^2}{n}} = \sqrt{\frac{0}{4}} = \sqrt{0} = 0$$

The variance and standard deviation both equal zero, as does the mean deviation, for this group in which there is *no* dispersion at all.

05-Sirkin-4731.qxd 6/9/2005 6:40 PM Page 134

Dick's Group

$x =$	$\bar{x} =$	$x - \bar{x} =$	$(x - \bar{x})^2 =$
9	8	1	1
8	8	0	0
8	8	0	0
7	8	–1	1
			$\sum(x - \bar{x})^2 = 2$

Thus,

$$s^2 = \frac{\sum(x - \bar{x})^2}{n} = \frac{2}{4} = \frac{1}{2} = 0.5$$

$$s = \sqrt{\frac{\sum(x - \bar{x})^2}{n}} = \sqrt{\frac{2}{4}} = \sqrt{\frac{1}{2}} = 0.70$$

Remember that it is the standard deviation (0.7), not the variance, which substitutes for the mean deviation (0.5).

Harriet's Group

$x =$	$\bar{x} =$	$x - \bar{x} =$	$(x - \bar{x})^2 =$
10	8	2	4
10	8	2	4
6	8	–2	4
6	8	–1	4
			$\sum(x - \bar{x})^2 = 16$

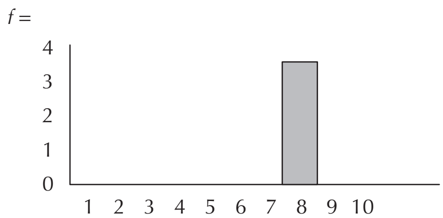
Thus,

$$s^2 = \frac{\sum(x - \bar{x})^2}{n} = \frac{16}{4} = 4.0$$

$$s = \sqrt{\frac{\sum(x - \bar{x})^2}{n}} = \sqrt{\frac{16}{4}} = \sqrt{4} = 2.0$$

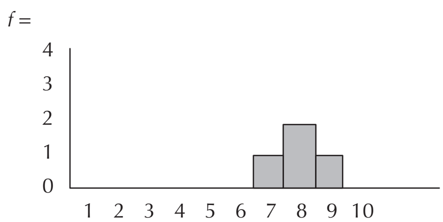
Let us compare our measures. See the histograms at the top of the next page.

05-Sirkin-4731.qxd 6/9/2005 6:40 PM Page 135



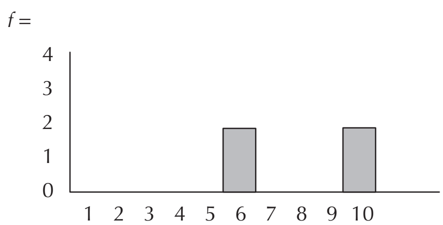
Tom's Group

Range = 0
Mean Deviation = 0
Variance = 0
Standard Deviation = 0



Dick's Group

Range = 2.0
Mean Deviation = 0.5
Variance = 0.5
Standard Deviation = 0.7



Harriet's Group

Range = 4.0
Mean Deviation = 2.0
Variance = 4.0
Standard Deviation = 2.0

Below are the dispersion measures for artistic freedom for the non–liberal arts majors, Group A, presented in Chapter 4.

Group A

	$x =$	$\bar{x} =$	$x - \bar{x} =$	$\lvert x - \bar{x} \rvert =$	$(x - \bar{x})^2 =$
	8	7	1	1	1
	8	7	1	1	1
	8	7	1	1	1
	7	7	0	0	0
	7	7	0	0	0
	7	7	0	0	0
	6	7	–1	1	1
	6	7	–1	1	1
$n = 9$	6	7	–1	1	1
	$\sum x = 63$			$\sum \lvert x - \bar{x} \rvert = 6$	$\sum (x - \bar{x})^2 = 6$

The scores range from 6 to 8. Range = 8 – 6 = 2.

05-Sirkin-4731.qxd 6/9/2005 6:40 PM Page 136

$$\bar{x} = \frac{\sum x}{n} = \frac{63}{9} = 7.0$$

$$M.D. = \frac{\sum |x - \bar{x}|^2}{n} = \frac{6}{9} = \frac{2}{3} \cong 0.67$$

$$\text{Variance} = s^2 = \frac{\sum (x - \bar{x})^2}{n} = \frac{6}{9} = \frac{2}{3} \cong 0.67$$

$$\text{Standard Deviation} = s = \sqrt{0.67} = 0.82$$

Summary	*Group A*
Range	2.00
Mean Deviation	0.67
Variance	0.67
Standard Deviation	0.82

As mentioned, the variance and standard deviation are the most widely used measures of dispersion in statistics, even though on the face of it, the mean deviation would appear to be the most logical measure (and easiest to calculate) of the three. The reason is that the standard deviation has meaning in terms of a common frequency distribution known as the *normal curve,* which we will encounter later in this text.

THE COMPUTATIONAL FORMULAS FOR VARIANCE AND STANDARD DEVIATION

The variance formula $s^2 = \sum (x - \bar{x})^2/n$ is often referred to as the **definitional formula** since it not only calculates the variance but also defines or explains what the variance is: the mean amount of the squared deviations of the scores from the mean. (It is often quite difficult for those long away from algebraic formulas to "see" that definition, but it is there.)

Definitional formula A formula that not only calculates the variance but also defines or explains what the variance is: the mean amount of the squared deviations of the scores from the mean.

For computational purposes, however, it is often easier to use one of several alternative formulas, known as **computational formulas,** particularly if a calculator is available. One such computational formula is the following:

05-Sirkin-4731.qxd 6/9/2005 6:40 PM Page 137

Computational formulas A formula that generates the correct variance but does not seek to define what the variance is.

$$\text{Variance} = s^2 = \frac{\sum x^2 - \frac{(\sum x)^2}{n}}{n}$$

$$\text{Standard Deviation} = s = \sqrt{\frac{\sum x^2 - \frac{(\sum x)^2}{n}}{n}}$$

Before we apply these formulas, we should make note of the difference between two parts of the formula: $\sum x$ and $(\sum x)^2$, which are *not* the same. The first, $\sum x^2$, read "summation of x squared," tells us to square each x and then add up all of the x^2s. The second, $(\sum x)^2$, read "summation of x, quantity squared," tells us to first add up all the xs to get $\sum \bar{x}$ and then square $\sum x$ to get $(\sum x)^2$. (This follows the convention of first doing what is *inside* a set of parentheses before doing what is outside of the parentheses.) Thus, we must add the original scores and square the sum, and we must also square each original score and add up the squared values.

Group A

$x =$	$x^2 =$
8	64
8	64
8	64
7	49
7	49
7	49
6	36
6	36
6	36
$\sum x = 63$	$\sum x^2 = 447$

$n = 9$

$$(\sum x)^2 = (63)^2 = 63 \times 63 = 3969$$

$$s^2 = \frac{\sum x^2 - \frac{(\sum x)^2}{n}}{n} = \frac{447 - \frac{(63)^2}{9}}{9}$$

$$= \frac{447 - \frac{3969}{9}}{9} = \frac{447 - 441}{9}$$

$$= \frac{6}{9} = \frac{2}{3} = 0.67$$

and

$$s = \sqrt{0.67} = 0.82$$

The answers are obviously the same as when we use the definitional formula. Often, the two results will differ slightly due to rounding error, particularly if the mean used in the definitional formulas is not a whole number (such as 7, in this case) but possesses several decimals (such as 7.2,

05-Sirkin-4731.qxd 6/9/2005 6:40 PM Page 138

7.23, 7.234, and so on). Notice that the computational formula requires the calculation of several large intermediate figures, such as the $(\sum x)^2 = 3969$. Since such large numbers are not needed when using the definitional formula, we may question the need for a computational formula. If, however, there are many scores (even as few as the 9 scores in Group A), it is faster and easier to use the computational formulas. It is even easier to use the computational formulas with today's advanced scientific, business, and statistical calculators, which usually store $\sum x$ and $\sum x^2$ in their memories for easy retrieval.

BOX 5.1

Another Formula for the Standard Deviation

In Chapter 8, you will encounter another formula for the standard deviation, indicated by the lowercase Greek letter sigma with a circumflex above it and read (believe it or not) as "sigma hat."

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

Note that this formula is the same as the definitional formula we have just been using except that $n - 1$ replaces n in the denominator. When we wish to generalize about some group (called a population) from data taken from fewer people than the entire group (called a sample), we run into a problem. Suppose I wanted to generalize about the ages of all residents of Thousand Oaks, California (the population), from a sample of 20 residents of that town. If I calculate the mean for my sample, I get the best estimate of the mean age of all that community's residents that my data will allow. However, if I estimate the population's standard deviation from my sample, using the formula with n in the denominator, my estimate is inaccurate. In fact, the smaller the size of my sample, the less accurate my estimate of the population's standard deviation will be.

It turns out that the formula with $n - 1$ in the denominator gives us a better estimate of the population's standard deviation than the formula with n. Thus, you will see the $n - 1$ formula widely used in textbooks, calculators, and computer programs. In fact, rarely can we study whole populations directly; so much of the time, we are really using sample data to estimate population data. That is why the formula with $n - 1$ in the denominator appears so often.

(Continued)

05-Sirkin-4731.qxd 6/9/2005 6:40 PM Page 139

(Continued)

Finally, note that many authors will state that the formula with n in the denominator is for a population's standard deviation and the $n - 1$ formula is for a sample's standard deviation. That is not quite correct, but since most of the time what we really are doing is using sample data to estimate population data, we really are not interested in the sample's standard deviation except as an estimate of the population's standard deviation. So, it is easier just to call the $n - 1$ formula the formula for a sample's standard deviation. That practice is not followed in this textbook.

VARIANCE AND STANDARD DEVIATION FOR DATA IN FREQUENCY DISTRIBUTIONS

If the data are in frequency distributions, the formulas given above will not find the correct variance or standard deviation. In a frequency distribution, we must account not only for each possible value of x but also for the number of times, or frequency, that value occurs. This is the same reason we modified the formula for finding the mean of a frequency distribution in the previous chapter. Recall that in calculating the mean for the liberal arts majors, Group B, we first established an fx column and added it up to get $\sum fx$. We then divided $\sum fx$ by $\sum f$ (our n) to get the mean. For frequency distribution data, the *definitional formula* for the *variance* is also adjusted so that before adding the squared deviations, we multiply each squared deviation by the frequency of that particular value of x.

$$s^2 = \frac{\sum[(x-\bar{x})^2 f]}{n} = \frac{\sum[(x-\bar{x})^2 f]}{\sum f}$$

Therefore,

Group B

$x =$	$f =$	$fx =$	$\bar{x} =$	$x-\bar{x} =$	$(x-\bar{x})^2 =$	$(x-\bar{x})f =$
9	2	18	7.5	1.5	2.25	2.25 × 2 = 4.50
8	3	24	7.5	0.5	0.25	0.25 × 3 = 0.75
7	3	21	7.5	–0.5	0.25	0.25 × 3 = 0.75
6	2	12	7.5	–1.5	2.25	2.25 × 2 = 4.50
	$n = \sum f = 10$	$\sum fx = 75$				$\sum[(x-x)^2 f] = 10.50$

$$\bar{x} = \frac{\sum fx}{n} = \frac{\sum fx}{\sum f} = \frac{75}{10} = 7.5$$

Thus, the variance is

$$s^2 = \frac{\sum[(x-\bar{x})^2 f]}{n} = \frac{\sum[(x-\bar{x})^2 f]}{\sum f} = \frac{10.50}{10} = 1.05$$

and the standard deviation is

$$s = \sqrt{1.05} = 1.0246 = 1.03$$

For data in frequency distributions. there is also an adjusted *computational formula.*

$$s^2 = \frac{\sum x^2 f - \frac{(\sum fx)^2}{n}}{n} = \frac{\sum x^2 f - \frac{(\sum fx)^2}{n}}{\sum f}$$

To apply this to Group B, we must generate columns for x^2 in order to find $\sum x^2$ and $x^2 f$ in order to find $\sum x^2 f$. We have already generated an fx column, but we need to square its summation.

$x =$	$f =$	$fx =$	$x^2 =$	$x^2 f$
9	2	18	81	$81 \times 2 = 162$
8	3	24	64	$64 \times 2 = 192$
7	3	21	49	$49 \times 2 = 147$
6	2	12	36	$36 \times 2 = 72$
	$n = \sum f = 10$	$\sum fx = 75$		$\sum x^2 f = 573$

$$\begin{aligned}(\textstyle\sum fx)^2 &= (75)^2 \\ &= 75 \times 75 \\ &= 5625\end{aligned}$$

Thus, the variance is

$$s^2 = \frac{\sum x^2 f - \frac{(\sum fx)^2}{\sum f}}{n} = \frac{573 - \frac{(75)^2}{10}}{10} = \frac{573 - \frac{5625}{10}}{10}$$

$$= \frac{573 - 562.5}{10} = \frac{10.5}{10} = 1.05$$

and the standard deviation is

$$s = \sqrt{1.05} = 1.03$$

The results are identical to those found using the definitional formulas.

05-Sirkin-4731.qxd 6/9/2005 6:40 PM Page 141

We now know the primary measures for describing a single-interval or ratio-level variable: the mean for central tendency and the standard deviation or variance for dispersion. With the latter two, we generally use the standard deviation for descriptive purposes but retain the variance for use in procedures that will be discussed later in this text.

With the exception of the range, the measures of dispersion presented in this chapter all assume interval level of measurement. (The range may be applied also to ordinal data: "The guests at the $100-a-plate charity fundraiser ranged from middle class to affluent.") While measures of dispersion are widely used with interval-level data, they are only rarely used with lower levels of measurement. Accordingly, such usage will not be covered here.

CONCLUSION

We have now covered the last of the basic tools of descriptive data analysis. With the introduction of dispersion measures, particularly the variance and the standard deviation, we can begin the study of several statistical techniques widely applied in many disciplines. We will see that in addition to their role as useful descriptive tools, the mean and the variance often plug into other formulas. Thus, they do double duty. Armed with the tools introduced so far, we will eventually return to the task of finding and describing relationships between two variables.

Chapter 5: Summary of Major Formulas

Individual Data	
The Mean Deviation $M.D. = \frac{\sum \lvert x - \bar{x} \rvert}{n}$	
The Variance *Definitional* $s^2 = \frac{\sum (x - \bar{x})^2}{n}$	The Variance *Computational* $s^2 = \frac{\sum x^2 - \frac{(\sum x)^2}{n}}{n}$

05-Sirkin-4731.qxd 6/9/2005 6:40 PM Page 142

Frequency Distributions	
Definitional	*Computational*
$s^2 = \frac{\sum[(x-\bar{x})^2 f]}{n} = \frac{\sum[(x-\bar{x})^2 f]}{\sum f}$	$s^2 = \frac{\sum x^2 f - \frac{(\sum fx)^2}{n}}{n} = \frac{\sum^2}{x} f - \frac{(\sum fx)^2}{\sum f} \sum f$

Both Individual and Frequency Distribution

The Standard Deviation

$s = \sqrt{\text{the variance}}$

EXERCISES

Note: For the following exercises, refer to the exercises at the end of Chapter 4 for the definitions of the variables.

Exercise 5.1

In the social worker sample (Exercises 4.7 to 4.9), a group of 9 private agency employees was compared to a group of 16 public employees. Following are the health care cost ratings for the private agency employees. Remember that the higher rating indicates more concern about the issue.

Private Agency Employees
Health
70
55
15
10
5
5
5
0
0

1. Find the mean Health score.
2. Find the median.
3. Find the mean deviation.
4. Find the variance using the definitional formula.
5. Find the variance using the computational formula.
6. Find the standard deviation.

05-Sirkin-4731.qxd 6/9/2005 6:40 PM Page 143

Exercise 5.2

Following are the health care cost ratings for the public employees:

Public Employees Health
95
95
95
90
90
90
90
90
90
80
80
75
75
60
40
35

Form a frequency distribution from the above, and using the appropriate formulas:

1. Find the mean Health score.
2. Find the median.
3. Find the variance using the definitional formula.
4. Find the variance using the computational formula.
5. Find the standard deviation.
6. Compare the mean and standard deviation of the public employees to those of the private agency employees found in Exercise 5.1. Which group's scores cluster more closely about its mean?

Exercise 5.3

Management personnel have been scored on a scale measuring assertiveness of leadership style, where more assertiveness indicates less accommodativeness. Are financial and banking managers more assertive than their colleagues in other service industries? Following are scores for 7 managers in finance- or banking-related firms.

Assertiveness
24
49
92
92
11
68
97

1. Find the mean Assertiveness score.
2. Find the median. (Note that you must first array the data from high to low scores.)
3. Find the mean deviation.
4. Find the variance using the definitional formula.
5. Find the variance using the computational formula.
6. Find the standard deviation.

Exercise 5.4

Following are assertiveness scores for 18 managers from nonfinancial service industries listed in an ungrouped frequency distribution.

x = *Assertiveness*	f =
100	1
97	1
92	1
86	3
54	1
30	1
27	3
24	2
22	1
5	1
3	1
0	2

1. Find the mean Assertiveness score.
2. Find the median.
3. Find the variance using the definitional formula.
4. Find the variance using the computational formula.
5. Find the standard deviation.
6. Compare the means and standard deviations of the nonfinancial institution managers to those found in Exercise 5.3. Which group is more assertive? Which group's scores are more spread out about the mean?

05-Sirkin-4731.qxd 6/9/2005 6:40 PM Page 145

Exercise 5.5

Below are the results, in printout format, for the employee sample of Exercise 4.10 (refer to Exercise 4.10 for a definition of the variables). Please note that this was run using SAS, one of several statistical packages available (we will be discussing the most recent version of SAS later in this book). Like most such packages, data are presented with far more decimal places than social scientists need. While suitable for engineers and some scientists, this level of precision is not suitable for the less exact measures that we use. Thus, when discussing the results, we will round to one or two decimal places.

In this exercise, workers have been broken down by region, Midwest versus all other regions combined. Suppose it had been rumored that the corporation was planning to close several plants and move those jobs to plants in other countries with lower wage scales. Suppose it had also been rumored that only plants in the Midwest would be exempt; in all other regions, some plants would be shut down. Let us compare the attitudes of the employees.

			Reg = Midwest
Variable	*N*	*Mean*	*S.D.*
ATTEND	13	90.6153846	12.2782902
BOARD	13	44.7692308	19.2663517
DIV	13	76.6153846	16.8302231
SECUR	13	67.7692308	28.4580213
PARTIC	13	39.6153846	35.0868885
OPPOR	13	55.4615385	38.5413531
UNION	13	55.3846154	35.5844968
SALARY	13	65.6923077	25.9466909

			Reg ≠ Midwest
Variable	*N*	*Mean*	*S.D.*
ATTEND	37	93.7297297	5.8720082
BOARD	37	34.7837838	18.1615209
DIV	37	78.7837838	16.1832134
SECUR	37	44.7027027	32.6129361
PARTIC	37	67.4324324	30.5646315
OPPOR	37	30.0270270	32.4349661
UNION	37	76.8918919	29.4380553
SALARY	37	49.6486486	27.4764710

1. Compare the means for each variable. What do you conclude?
2. Which region usually has the greater diversity on these dimensions as determined by comparing the standard deviations? In which two scales is that tendency reversed?

Exercise 5.6

Following is a comparison of the managerial group to the employee group.

MGTPOP

Variable	*N*	*Mean*	*S.D.*
ATTEND	89	92.3595506	9.9307228
BOARD	89	57.1123596	15.1840513
DIV	89	74.9438202	16.4305422
SECUR	89	56.1685393	32.4479429
PARTIC	89	48.5955056	34.6737126
OPPOR	89	42.5280899	36.3509065
UNION	89	62.4719101	31.6307136
SALARY	89	53.8764045	24.2575294

EMPLOY

Variable	*N*	*Mean*	*S.D.*
ATTEND	50	92.9200000	8.0997899
BOARD	50	37.3800000	18.7832861
DIV	50	78.2200000	16.2081989
SECUR	50	50.7000000	32.9274093
PARTIC	50	60.2000000	33.7602592
OPPOR	50	36.6400000	35.5486214
UNION	50	71.3000000	32.2118307
SALARY	50	53.8200000	27.7501167

You have already compared the means in Exercise 4.10.

Now compare the standard deviations for each variable. What can you conclude? For which variables are the managers more diverse (have larger standard deviations)? For which variables are the employees more diverse?

Exercise 5.7

The two discontented groups, upper-middle management and white-collar employees, are compared in the following sets of data.

UPPER-MIDDLE MANAGEMENT

Variable	*N*	*Mean*	*S.D.*
ATTEND	50	91.8000000	12.8364914
BOARD	50	47.2200000	10.9195388
DIV	50	78.6400000	15.1600442
SECUR	50	39.8000000	31.0227040
PARTIC	50	72.1000000	22.0178499
OPPOR	50	16.4800000	18.2043278
UNION	50	85.6600000	10.8130873
SALARY	50	36.1000000	11.1158097

05-Sirkin-4731.qxd 6/9/2005 6:40 PM Page 147

WHITE-COLLAR EMPLOYEES

Variable	*N*	*Mean*	*S.D.*
ATTEND	29	93.3103448	5.1137311
BOARD	29	23.8965517	6.9710873
DIV	29	84.2068965	9.4354169
SECUR	29	31.3103448	26.7956598
PARTIC	29	81.8965517	17.8992115
OPPOR	29	13.1724137	16.7333477
UNION	29	92.4482758	10.9628796
SALARY	29	35.0000000	14.7672417

Compare the means and then the standard deviations for each variable. What do you conclude?

Exercise 5.8

For the data in Exercise 4.1, calculate and compare the standard deviations. Use the definitional formula to find the variance for the exporters and the computational formula to find the variance for the nonexporters. Then find and compare the two standard deviations.

Exercise 5.9

For the data in Exercise 4.4, calculate and compare the standard deviations. Use the frequency distribution definitional formula to find the variance for the exporters and the frequency distribution computational formula to find the variance for the nonexporters. Then find and compare the two standard deviations.

06-Sirkin-4731.qxd 6/9/2005 6:53 PM Page 148

▼ KEY CONCEPTS ▼

contingency table	spurious relationships	antecedent variable
control variable	causal models	intervening variable