

CHAPTER 6

MEASURES OF DISPERSION AND STANDARD SCORES

OBJECTIVES

After completing this chapter, you should

- be able to define and compute the measures of dispersion covered in the chapter—the range, average deviation, variance, and standard deviation.
- be able to define and compute z scores or standard scores.

CHAPTER REVIEW

Several statistics are introduced in this chapter to measure the spread or dispersion of the scores around the mean in a distribution. The measures discussed are the range, the average deviation, the variance, and the standard deviation.

The *range* (R) is the difference between the highest and lowest scores in a distribution. It is useful in a descriptive sense because it is determined so easily. The major problem with the range is that it is based on the two extreme scores in the distribution. If the extremes are not representative of the distribution, the range will be unrepresentative as well.

The *average deviation* (AD) is the average of the deviations from the mean of the distribution of each score in the distribution. It is computed by dividing the sum of the absolute value of the deviations by the number of deviations. The absolute value of the deviations is used to prevent the sum of the deviations from always being zero, because of the property of the mean as the balancing point in the distribution that we discussed in Chapter 5. AD is introduced as a prelude to the variance and the standard deviation.

The *variance* is the average of the squared deviations. Population variance is symbolized by σ^2 , and sample variance is symbolized by s^2 . The numerator of the equation for the variance is sometimes called the sum of squares or *SS*. The sample variance (s^2) is an unbiased estimate of σ^2 .

The *standard deviation* is the square root of the variance or the square root of the average squared deviation. Symbols for the population standard deviation and the sample standard deviation are σ and s , respectively. The standard deviation can be visualized as another unit of measurement on the baseline of a frequency polygon. A useful approximation of s can be found by dividing the range by 4.

A *standard score* or *z score* is the deviation of a raw score from the mean in standard deviation units. The sign of the *z score* tells the direction of the score relative to the mean: Negative *z scores* represent raw scores below the mean, and positive *z scores* indicate scores above the mean.

SYMBOLS

<i>Symbol</i>	<i>Stands For</i>
<i>AD</i>	average deviation
<i>R</i>	range
σ^2	population variance
s^2	sample variance
σ	population standard deviation
<i>s</i>	sample standard deviation
s_{approx}	an approximation of <i>s</i> ; $s_{\text{approx}} = \frac{R}{4}$
<i>SS</i>	sum of squares or the numerator of variance
<i>z</i>	standard score or <i>z score</i>

FORMULAS

Formula 6-1. Formula for calculating the range

$$R (\text{range}) = HS - LS$$

Formula 6-3. Formula for computing *AD* from a frequency distribution

$$AD = \frac{\sum f |X - \bar{X}|}{N}$$

To find *AD*, the first step is to subtract the mean from each score. Next, multiply the absolute values of the differences by their frequencies and sum the results. Finally, divide the numerator by *N* to find the average deviation.

Formula 6-8. Computational formula for s^2 , the sample variance

$$s^2 = \frac{\sum X^2 - \frac{(\sum X)^2}{N}}{N - 1}$$

Formula 6-9. Computational formula for s^2 , the sample variance, for a frequency distribution

$$s^2 = \frac{\sum fX^2 - \frac{(\sum fX)^2}{N}}{N-1}$$

The numerator is the computational form of the sum of squares. To get it, sum the squared scores and subtract from this value the sum of the scores squared divided by sample size. Then divide the whole thing by $N - 1$.

Formula 6-14. Computational formula for sample standard deviation

$$s = \sqrt{\frac{\sum X^2 - \frac{(\sum X)^2}{N}}{N-1}}$$

Formula 6-15. Computational formula for sample standard deviation for a frequency distribution

$$s = \frac{\sum fX^2 - \frac{(\sum fX)^2}{N}}{N-1}$$

s is simply the square root of the formula for sample variance.

Formula 6-17. Formula for finding a z score from a raw score using sample statistics

$$z = \frac{X - \bar{X}}{s}$$

Formula 6-19. Formula for finding a raw score from a z score using sample statistics

$$X = zs + \bar{X}$$

Formula 6-19 is obtained by solving Formula 6-17 for X .

TERMS TO DEFINE AND/OR IDENTIFY

dispersion

range

average deviation

absolute value of a score

variance

standard deviation

sum of squares

standard score

z score

FILL-IN-THE-BLANK ITEMS

Introduction

To describe a frequency distribution, we need to know a measure of the (1) _____ of the data in addition to measures of central tendency. Four measures discussed in this chapter are the range, the average deviation, the (2) _____, and the (3) _____.

The Range

The (4) _____ is the difference between the highest score in the distribution and the lowest score. The range is easy to determine but not very useful for further statistical procedures.

The Average Deviation

The average deviation, symbolized by (5) _____, is useful as a prelude to the most commonly used measures of dispersion, the (6) _____ and the standard deviation. In order to keep from obtaining zero each time the deviations around the sample mean are summed, we take the (7) _____ of the deviations before summing.

The Variance and the Standard Deviation

The (8) _____ is the average of the squared deviations, and the square root of the average is called the (9) _____. The statistic based directly on the formula for the population variance is a (10) _____ estimate of the population variance. To compensate for SD^2 's tendency to (11) _____ the population variance, we have to modify the formula

slightly, which we did by dividing the numerator by (12) _____ rather than by N . Standard deviation is the (13) _____ of the variance.

In addition to defining formulas, (14) _____ or raw-score formulas are introduced. Raw-score formulas require fewer (15) _____ than the defining formulas and are easier to use when computed with a pocket calculator.

s can be visualized as a width measure on the (16) _____ of a frequency polygon. A useful way to estimate the standard deviation is to divide the (17) _____ by (18) _____. The numerator of variance is sometimes called the (19) _____, _____, which is the sum of the squared deviations about the (20) _____. It is symbolized by (21) _____.

Standard Scores (z Scores)

Values on the standard deviation scale are called either (22) _____ scores or (23) _____ scores. A (24) _____ is the deviation of a raw score from the mean in standard deviation units. The (25) _____ of the z score tells us the direction of the score relative to the mean. A (26) _____ z score indicates a raw score below the mean. To convert a z score back to a raw score, multiply it by the standard deviation and add the (27) _____.

Troubleshooting Your Computations

In computing any of the measures discussed in the chapter, it is desirable to have a (28) _____ for a correct answer. For example, if the distribution is large and symmetrical, s should be approximately (29) _____ of the range. You must always get (30) _____ numbers for the standard deviation and the variance. When computing s , don't forget to take the (31) _____.

PROBLEMS

1. Markus says that the AD of the following test scores is 2.5, but Karl says it is 1.45. Who is correct? What are R , s_{approx} , s^2 , and s ?

X	f
25	1
24	2
22	3
21	4
20	4
19	2
18	1

2. In a logic course, 23 students are given 10 word-analogy problems to solve in 5 minutes. The number of problems correctly solved is recorded for each student. The data are as follows:

X	f
10	1
9	2
8	1
7	4
6	6
5	5
4	2
3	1
2	1

- a. Find R , s_{approx} , s^2 , and s .

- b. Plot a frequency polygon of the data and lay off the value of s on it. To simplify your graph, round s and \bar{X} to the nearest whole numbers.

3. Scores on an experimental psychology test were as follows:

X	f
100	1
96	1
92	3
84	2
83	1
80	2
79	1
76	1
63	1
35	1

Find R , s_{approx} , s^2 , and s . What is the z score for a raw score of 96? What score is 2 standard deviation units below the mean?

4. Using the following frequency distribution, demonstrate that $(\sum fX)^2$ is not equivalent to $\sum fX^2$. Compute R , s_{approx} , s^2 , and s .

X	f
5	1
4	2
3	3
2	3
1	2

5. There are two applicants for the job as test pilot for experimental jet aircraft. Both applicants are given a reaction-time test and are found to have the same average times. Which applicant should get the job if consistency of reaction is an important criterion?

<u>Applicant A</u>		<u>Applicant B</u>	
X	f	X	f
1.04	1	1.04	1
1.03	1	0.85	1
0.95	2	0.84	2
0.88	1	0.80	2
0.85	2	0.77	3
0.78	2	0.75	3
0.74	1	0.74	2
0.70	1	0.72	1
0.65	3	0.70	2
0.60	4	0.57	1
0.55	2	0.55	2

6. Participants in an employee training program are tested after 1 week. By company policy, employees who score 1 standard deviation or more below the mean are required to take another week of training. Which employees must receive more training, and which are released from the program?

X	f
92	1
91	1
88	3
87	3
81	4
79	5
77	5
76	3
71	3
69	2
53	1
51	2
47	1
44	1
39	1

7. Using the following data, find s^2 . Determine s and plot it on the baseline of a frequency polygon. For ease in plotting, round the mean and standard deviation to the nearest whole numbers.

X	f
28	1
27	1
26	3
25	5
24	5
23	2
22	1
21	1
20	1

8. Assuming a mean of 2.57 with a standard deviation of 0.83, answer the following questions:

a. What is the z score for a raw score of 3.75?

b. What is the z score for a raw score of 2.10?

c. How far from the mean in standard deviation units is a score of 1.3?

d. What score lies 1.75 standard deviation units above the mean?

e. What score lies 2.17 standard deviation units below the mean?

USING SPSS—EXAMPLES AND EXERCISES

SPSS has several procedures that can provide measures of central tendency and dispersion. For example, Frequencies, Descriptives, or Explore can be used. Using a Compute statement, z scores can be calculated, or they can be obtained “automatically” as an option under the Descriptives procedure.

Example: We will use SPSS to obtain the three measures of central tendency (mean, median, and mode), measures of dispersion (range, variance, and standard deviation), and z scores for all test scores using the data in Problem 6. The steps are as follows:

1. Start SPSS, enter the data, and name the variable **test**.
2. We will use Frequencies to obtain the measures of central tendency and dispersion.
3. Go to *Analyze>Descriptive Statistics>Frequencies*.
4. Move **test** to the Variables box; uncheck “display frequency table” so the frequency distribution will not be calculated.
5. Click *Statistics*, select the measures of central tendency and dispersion that you want, and then click *Continue>OK*. This will give the descriptive statistics in the output Viewer window.
6. Next we will compute z scores using the Compute procedure. To do this we need to know the mean (symbolized by $M = 74.3333$ rather than $\bar{X} = 74.3333$, as we symbolized it in the text) and standard deviation ($s = 13.6403$) from the output. The formula for any z score is $z = (X - M)/s$ or $z = (X - 74.3333)/13.6403$. We will now put this formula in the Compute box.
7. From the Data Editor window, click *Transform>Compute*. Then enter “testz” in the Target variable box.
8. In the Numeric Expression box, move **test** in and use the keypad or type and edit so the expression appears as follows: $(\text{test} - 74.3333)/13.6403$.
9. Click *OK*. Check the Data Editor window and find the computed z scores.

Example: Repeat SPSS for the previous example using the Descriptives procedure. Although you will not be able to obtain the median and mode from this procedure, z scores can be computed automatically.

10. Click *Analyze>Descriptive Statistics> Descriptives*, move **test** into the Variables box, and click “Save standardized values as variables.” This causes the z scores to be computed.
11. Click *Options* and select the desired descriptive statistics that are available, then click *Continue>OK*.
12. The output Viewer shows the descriptive statistics. The Data Editor should show another variable **ztest** showing the z scores again, differing only in the number of decimal places displayed.

```

FREQUENCIES
  VARIABLES=test /FORMAT=NOTABLE
  /STATISTICS=STDDEV VARIANCE RANGE MEAN MEDIAN MODE
  /ORDER ANALYSIS .

```

Frequencies

Statistics

TEST		
N	Valid	36
	Missing	0
Mean		74.3333
Median		77.0000
Mode		77.00 ^a
Std. Deviation		13.6403
Variance		186.0571
Range		53.00

a. Multiple modes exist. The smallest value is shown.

```

COMPUTE testz = (test - 74.3333)/13.6403 .
EXECUTE .

```

	test	testz
1	92.00	1.30
2	91.00	1.22
3	88.00	1.00
4	88.00	1.00
5	88.00	1.00
6	87.00	.93
7	87.00	.93
8	87.00	.93
9	81.00	.49
10	81.00	.49
11	81.00	.49
12	81.00	.49
13	79.00	.34
14	79.00	.34
15	79.00	.34
16	79.00	.34
17	79.00	.34
18	77.00	.20
19	77.00	.20
20	77.00	.20

SPSS Data Editor window showing z scores obtained using the Compute statement (first 20 cases).

DESCRIPTIVES

VARIABLES=test /SAVE

/STATISTICS=MEAN STDDEV VARIANCE RANGE MIN MAX .

Descriptives

Descriptive Statistics

	N	Range	Minimum	Maximum	Mean	Std. Deviation	Variance
TEST	36	53.00	39.00	92.00	74.3333	13.6403	186.057
Valid N (listwise)	36						

SPSS Data Editor window showing z scores computed using the Compute statement and using the z -score output option for the Descriptives procedure.

	test	testz	ztest
1	92.00	1.30	1.29518
2	91.00	1.22	1.22187
3	88.00	1.00	1.00193
4	88.00	1.00	1.00193
5	88.00	1.00	1.00193
6	87.00	.93	.92862
7	87.00	.93	.92862
8	87.00	.93	.92862
9	81.00	.49	.48875
10	81.00	.49	.48875
11	81.00	.49	.48875
12	81.00	.49	.48875
13	79.00	.34	.34212
14	79.00	.34	.34212
15	79.00	.34	.34212
16	79.00	.34	.34212
17	79.00	.34	.34212
18	77.00	.20	.19550
19	77.00	.20	.19550
20	77.00	.20	.19550

Exercises Using SPSS

1. Using the data from Problem 2, use the SPSS Descriptives procedure to obtain the mean, standard deviation, variance, and z scores for each score.
2. Using the same data and the necessary statistics from the previous exercise, employ the SPSS Compute procedure to calculate z scores again for each score. Use **correctz** as the target variable name.

CHECKING YOUR PROGRESS: A SELF-TEST

1. Match the following:

- | | |
|------------------|--|
| _____ z | a. symbol for the population standard deviation |
| _____ s | b. average amount that each score in a distribution deviates from the distribution's mean |
| _____ s^2 | c. symbol for a standard score |
| _____ σ^2 | d. symbol for population variance |
| _____ AD | e. numerator of the equation for variance |
| _____ SS | f. symbol for sample variance |
| _____ R | g. symbol for sample standard deviation |
| | h. symbol for the range of a set of scores |

2. What does the size of a z score indicate?

3. What does the sign of a z score indicate?

4. Family size was surveyed in a large sociology class, and the results were as follows:

X	f
13	1
10	1
7	4
6	6
5	12
4	26
3	12
2	8

a. Find R , s^2 , and s .

b. What is the z score for 6? 3?

c. What score is 1.5 standard deviation units above the mean? What score is 2 standard deviation units below the mean?