CHAPTER 3

Central Tendency and Variability

Summary

This chapter covers two very important characteristics of distributions: *measures of central tendency* and *variability*. You have experience with the concepts of central tendency and variability even if you haven't studied the *mean* and *standard deviation* before. Calculating and interpreting measures of central tendency and variability such as the mean, median, mode, standard deviation, range, interquartile range, and variance gives you a quantitative way to express some familiar ideas.

Expressing ideas quantitatively can be very powerful. You already have a good bit of experience with this; for years you have used the *mean* when you want to get across the concept of "the typical member of the group." Other measures of central tendency are the *median* and the *mode*.

The mean is usually the preferred measure of central tendency, but it should not be used if a class interval is open ended, if the observations are nominal or ordinal data, or if the distribution is severely skewed. Nominal data limit you to the mode. Ordinal data limit you to the median and the mode.

The *median* and the *mode* are used less frequently than the mean. The median is the middlemost score of a rank ordered distribution. It is the score that corresponds to the 50^{th} percentile. The median is in the appropriate measure of central tendency for skewed distributions.

The mode is the score that occurs most frequently. Some distributions have no modes and some have more than one mode. The mode is more informative when combined with the percentage of times that it occurs in the distribution.

Two mathematical characteristics of the mean are referred to later in the text. If you do not feel comfortable with the expressions $\Sigma(X - \overline{X}) = 0$ and $\Sigma(X - \overline{X})^2$ is a minimum, you should make up a small set of data and perform the operations that

Choosing among these three standard deviation depends on how the data were gathered (was sampling used?) and the purpose of the data gathering (generalization?).

Bars that extend one standard deviation can be added to a bar graph of means. Standard deviation bars indicate the spread of scores in the sample.

If the issue of N or $N-1$ in the formula for a standard deviation has you shaking your head (sideways), I recommend that you work the exercise described in the footnote of the section, " as an estimate of σ" in the textbook.

The *variance* gets two short paragraphs in the text, which is not representative of its importance in the overall field of statistics. You will learn more about variance in Chapters 10, 11, and 12. As a *descriptive index* of variability, however, the variance is not nearly as useful as its square root, the standard deviation.

Multiple-Choice Questions _____

1. Which of the following words could legitimately fit into this sentence: "That simple frequency distribution has two _____, 13 and 18."
   (1) means;
   (2) medians;
   (3) modes;
   (4) all of the above.

2. Your text noted which of the following as a characteristic of the mean?
   (1) The sum of the results of squaring the difference between each score and the mean is a minimum;
   (2) The sum of the results of squaring the difference between each score and the mean is zero;
   (3) Both (1) and (2);
   (4) Neither (1) nor (2).

the formulas describe. Then, be alert for these ideas when they turn up in Chapter 4 and Chapter 5.

Calculating a weighted mean is easy. Recognizing when you can work from the individual means and when you have to calculate the overall sum of X takes practice. Some of the problems that follow give you that practice.

The first quantitative measure of variability in this chapter is the *range*, a simple and easily calculated statistic. The range is simply the numerical distance between the highest and lowest scores.

The *interquartile range* (IQR) gives you the scores in the distribution that are the middle 50 percent of the distribution. The IQR is the 75^{th} percentile minus the 25^{th} percentile. A location value ($.25 \times N$) allows you to find the 75^{th} percentile (by working from the top of the distribution) and the 25^{th} percentile (by working from the bottom of the distribution).

Another important statistic is the *standard deviation.* The standard deviation tells you the average amount that scores differ from the mean. After studying the standard deviation in this chapter and again in Chapter 6, you will find it a powerful and accurate way to express the concept of variability.

The sections on the standard deviation show two ways to arrange the arithmetic—the *deviation-score* method and the *raw-score* method. The principal reason for learning the deviation-score method is that it gives you direct insight into the underlying workings of the standard deviation. However, a raw-score formula and a calculator produce more accurate answers more quickly. Raw-score formulas are given for *unorganized data* and for *frequency distributions.*

In the order of importance in the textbook, the three standard deviations are:

1. ŝ standard deviation of a sample, used to estimate σ ($N - 1$ in denominator)
2. σ standard deviation of a population (N in denominator)
3. S standard deviation of a sample is used when there is no interest in generalizing from the sample to its population. (N in denominator)

Chapter 3

3. For our study of driving habits, we recorded the speed of every fifth vehicle on Drury Lane. Nearly every car traveled right at the speed limit or a little over, but there were some that were 10 mph under, even fewer at 20 mph under, and one care that crept by at just 15 mph. On the basis of the central tendency calculations on our data, we drew conclusions about all drivers on this stretch of road. The proper central tendency value calculated from these data is the
   (1) population median;
   (2) sample median;
   (3) population mean (μ);
   (4) sample mean ($\bar{X}$).

4. The mean temperature for January was 30°. In February the mean was 25° and for March the mean was 35°. The weighted mean for these three months is
   (1) 30°;
   (2) greater than 30°;
   (3) less than 30°.

5. The standard deviation that estimates a population standard deviation from calculations on a sample is
   (1) σ;
   (2) $\hat{s}$;
   (3) S.

6. Two distributions that have the same mean must have the same
   (1) range;
   (2) standard deviation;
   (3) variance;
   (4) none of the above.

7. Which of the following is (are) *not* mathematically possible if only one distribution is being considered?
   (1) range = $\hat{s}$;
   (2) range = $\bar{X}$;
   (3) range = μ;
   (4) all of the above.

8. You will get a negative number as a standard deviation
   (1) when all scores are negative;
   (2) when the mean of the scores is negative;
   (3) either (1) or (2) is sufficient;
   (4) under no circumstances---standard deviations are *always* positive numbers.

9. An experimenter was interested in the variability of SAT scores in the new freshman class at her university. She obtained all the scores from the registrar. She should compute
   (1) σ;
   (2) $\hat{s}$;
   (3) S;
   (4) the 50^{th} percentile.

10. The standard deviation of the IQ's of students at a high school for the gifted and talented is _____ the standard deviation of IQ's of students at a high school that accepts everyone in its district.
   (1) less than;
   (2) greater than;
   (3) equal to.

11. Which of the following distributions is the most variable?
   (1) 1, 2, 3;
   (2) 8, 9, 10;
   (3) 1, 3, 5;
   (4) All of the above are equally variable.

12. Choose the statement that is true, according to your text.
   (1) The pay for people in management is more variable than the pay for those in sales;
   (2) $\hat{s}$ is a biased estimate of σ;
   (3) In reaching puberty, males are more variable than females;
   (4) All of the above are true.

Chapter 3

13. The interquartile range always
    (1) is one-fourth of the range;
    (2) becomes larger if more observations are added;
    (3) contains 50 percent of the frequencies;
    (4) all of the above.

14. For a set of scores, the sum of the deviation scores will be zero
    (1) when half the scores are negative;
    (2) when the mean is zero;
    (3) only when the standard deviation is zero;
    (4) always.

15. From all the employees in a company, a small group was selected to participate in a study of employee satisfaction. The results of the study were to be generalized to all the employees of the company. To find the variability in satisfaction, the investigator should compute
    (1) μ;
    (2) S;
    (3) $\hat{s}$;
    (4) σ.

16. For a sample of 100 or fewer, the standard deviation could be about _____ the range. (Be careful on this one.)
    (1) 2 times;
    (2) 5 times;
    (3) 2-5 times;
    (4) one fourth.

17. On the first test of material on dinosaurs, a class of sixth graders had a mean score of 36 with S = 12. The teacher was disappointed and assigned six pages of homework on dinosaurs and scheduled a second test. The top one-fourth of the class studied the extra material and did even better on the second test. The other three-fourths ignored the material and made the same scores as before. Pick the mean and standard deviation that might be found, given the description.
    (1) $\overline{X} = 36$, S = 12;
    (2) $\overline{X} = 36$, S = 18;
    (3) $\overline{X} = 42$, S = 18;
    (4) $\overline{X} = 42$, S = 12.

18. Which of the following may be a negative number?
    (1) range;
    (2) variance;
    (3) deviation score;
    (4) both (2) and (3).

19. Suppose you had a sample and you wanted to draw conclusions about the variability of a population. You should calculate
    (1) $\hat{s}$;
    (2) S;
    (3) IQR;
    (4) the range.

20. The variance is _____ the standard deviation.
    (1) the sum of;
    (2) the square of;
    (3) the square root of;
    (4) twice as large as.

Short-Answer Questions

1. Your text described three situations in which the median should be used rather than the mean. List them.

Chapter 3

2. Our company has three divisions. Based on our capital invested, Division A made $.10 per dollar, Division B made $.20 per dollar and Division C made $.25 per dollar. Therefore, the profit for the company was $.1833 per dollar invested. Do you agree with this conclusion? Why or why not?

3. This is a thought question about the distribution of class sizes in a typical college. There are usually a few courses, typically at the first-year level, that have large enrollments. Most of a college's courses, however, are at the junior and senior level, and in these courses, enrollments are much smaller. Suppose the mean class size in such a college is 23. Is the median larger or smaller?

4. The question is, "How can I buy groceries cheaply *and* efficiently?" Suppose there are four equally convenient supermarkets and they all place advertisements in the newspaper on Wednesday. From these advertisements you have noted the prices for Red Delicious apples (per pound), skim milk (per gallon), and Froot Loops cereal (per 12 oz). From your notes you have calculated the standard deviations of the prices of each item. The results are: Apples-----$.10, Milk-----$.02, Cereal-----$.01. Now suppose you are going to buy the three items and you want to do it cheaply and efficiently. You have this week's ads. What do you do?

5. Tell what each of the three types of standard deviations is used for; that is, which are used on samples, and which are used on populations.

Problems _____

1. "Count Avogadro, would you give us your number?" asked a pleasant voice on the phone. "This is your bank and we need your number," the voice continued. Suspecting a scam, Avogadro said, "My number? Here are several: 6.2, 5.8, 5.9, 6.1, and 6.1. Use them to find my number."
   Find the mean, median and mode of the numbers. (To get a psychological bonus from this problem, identify the number that Avogadro might claim as his).

Chapter 3

2. From each distribution of scores, find the mean, median and mode
   A. 9, 10, 9, 9, 11, 9, 10, 9, 9, 10
   B. 5, 4, 2, 1, 6, 5, 3, 2, 4, 2, 5, 5, 2
   C. 9, 11, 12, 10, 12, 11, 8, 12, 10, 11, 12
   D. 8, 7, 7, 9, 5, 7, 7, 8, 6, 6, 7, 7

3. A Likert scale is a standardized method of collecting opinions. A statement is presented such as, "Social security is not important as a means of income for the elderly." Responses are limited to 5 choices anchored by "strongly agree" (scored 1) to "strongly disagree" (scored 5). "No opinion" is scored 3. The two frequency distributions that follow show the responses of College Students and Older Adults to the statement above. For each distribution, find the mean, median and mode.

College Students	*f*	Older Adults	*f*
5	4	5	2
4	4	4	1
3	2	3	2
2	1	2	4
1	1	1	6

4. The numbers below are somewhat representative of human birthweights in 1984 and 2004. For each year, find the mean, median and mode.

1984		2004	
Weight (lbs.)	*f*	Weights (lbs.)	*f*
5	1	5	1
6	3	6	2
7	7	7	4
8	2	8	3
9	1	9	3
10	1	10	1

Chapter 3

5. For the weight data of 20-29 year old Americans (Problem 4 in Chapter 2), find the range and interquartile range.

6. The following data are representative of the age (in months) at which 90 percent of babies first stand without support (Berk, 2004). Compile appropriate statistics, and write a paragraph report to parents about the age they can expect their baby to stand. A report that explains both the typical case and variability is needed.

Age (months)	*f*
16	1
15	2
14	3
13	5
12	7
11	10
10	8
9	2

7. The data you analyzed in Problem 2b in Chapter 2 show the numbers of digits a person can hear and then repeat without error (memory span scores). Treat those data as a population and use the deviation-score method to find σ. Also, calculate the range and interquartile range.

8. The data in Problem 3 above reveal the opinions of college students and older adults on the value of social security. Find the range, IQR, and $\hat{s}$.

9. The Clock Test is a technique for studying human vigilance. In the Clock Test, a hand moves regularly at one step per second, but sometimes, at random intervals, it jumps two steps. The participant's task is to notice the two-step jumps and press a button. The numbers that follow are the percentages of two-step jumps that were missed by 5 participants during the first 15 minutes of a two-hour watch. Estimate the size of the standard deviation. Calculate an estimate of σ and σ^2 using the raw-score method.

11 8 17 10 14

Chapter 3

10. In the Clock Test experiment described in Problem 9, the percentages of missed jumps during the *last* 15 minutes of the two-hour watch are much higher than the first 15 minutes. Estimate the standard deviation of the scores that follow. Calculate estimates of σ and σ^2. Using this answer and your answer from Problem 6, write a sentence about the effect of two hours of vigilance on variability.

25 12 38 21 29