CHAPTER 3

Exploring Data: Central Tendency

Summary

This chapter covers ***measures of central tendency***, which are important characteristics of distributions. You have experience with the concept of central tendency, even if you haven't studied the *mean*, *median*, or *mode* before. In fact, you probably understand and use the idea of an average, which is the measure of central tendency represented by the *mean*, regularly when considering your grades or reading online product reviews. Calculating and interpreting measures of central tendency give you a quantitative way to express some familiar ideas.

The ***mean*** is usually the preferred measure of central tendency, but it should not be used if a class interval in a grouped frequency distribution 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. In contrast, you can calculate all measures of central tendency when using interval and ratio data.

Two mathematical characteristics of the mean are a) that the sum of all deviation scores from the mean is 0 and b) the sum of squared deviation scores is minimized. These characteristics are represented by $\sum(X - \bar{X}) = 0$ and $\sum(X - \bar{X})^2$, respectively. Both properties will be referred to later in the text. If you do not feel comfortable with these properties $\sum(X - \bar{X}) = 0$

The ***median*** is the middle score of an ordered distribution, and it is the same whether the scores are ordered highest-to-lowest or lowest-to-highest. It is the score that corresponds to the 50th percentile, meaning ½ the scores are above it and ½ the scores are below it. The median might be an actual number in the distribution, in the case of an odd number of scores, or the hypothetical midpoint between the two numbers in the center, in the case of an even number of scores. The median is in the most appropriate measure of central tendency for skewed distributions, because it is less sensitive to the effects of extreme scores than the mean is.

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.

When trying to find an average using two or more sets of scores with different *N*s, you need to determine the ***weighted mean***. The weighted mean allows the overall measure of central tendency to be determined in a way that assigns more weight to larger samples when calculating the overall mean.

Multiple-Choice Questions

1. Which of the following words could accurately complete this sentence: "That simple frequency distribution has two _____, 13 and 18."

a. means
b. medians
c. modes
d. all of the above

2. Your text noted which of the following as a characteristic of the mean?

a. the sum of the squared deviation scores is a minimum
b. the sum of the squared deviation scores is zero
c. both a. and b.
d. neither a. nor b.

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 (45 mph) or a little over, but there were some that were 10 mph under, even fewer at

20 mph under, and one car that crept by at 30 mph under. 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 _____.

a. population median
b. sample median
c. population mean (μ)
d. sample mean ($\bar{X}$)

4. The mean temperature for January was 30°F. In February the mean was 25°F and for March the mean was 35°F. The weighted mean for these three months is _____.

a. 30°F
b. greater than 30°F
c. less than 30°F
d. cannot be determined from the information given

5. In a positively skewed distribution, _____.

a. high scores are more frequent, and the mean is larger than the median
b. low scores are more frequent, and the mean is larger than the median
c. high scores are more frequent, and the mean is smaller than the median
d. low scores are more frequent, and the mean is smaller than the median

6. The mean is the proper descriptive statistic when _____.

a. you have nominal data
b. you have ordinal data
c. you have severely skewed data
d. none of the above

7. At City College, the five classes of students and their frequencies are: First-year, 500; Sophomores, 400; Juniors, 300; Seniors, 200; Graduate Students, 100. The proper central tendency statistic to describe the student body is the _____ with a value of _____.

a. mean; 300
b. mean; unknown
c. median; sophomore
d. mode; first-year

8. The mean salary for all those with a Bachelor's degree is $30,000; for all those with a Master's degree, $50,000, and for all those with a doctorate, $80,000. The overall mean salary is _____.

a. $50,000
b. less than $50,000
c. more than $50,000
d. cannot be determined from the information given

9. Three samples with *N*s of 5, 5, and 10 produced sample means of 2, 3, and 4 respectively. The overall mean is _____.

a. 3.00
b. 3.25
c. 6.67
d. 7.22

10. You are collecting data about the age of college students. The age options are: 17 and younger, 18, 19, 20, 21, 22, 23 and older. What is the most appropriate measure of central tendency?

a. median
b. mean
c. mode
d. any of the above

11. Your text noted two mathematical characteristics of the mean when deviation scores ($X - \bar{X}$) are examined. These two characteristics involved _____.

a. the maximum size of the mean and the minimum size of the mean
b. the size of the deviation scores when the mean is positive and the size when the mean is negative
c. the sum of the deviation scores and the size of the sum of the squared deviations
d. the maximum size of the deviations and the minimum size of the deviations

12. Which of the following is **not** true about the mode?

a. the mode can be more than one number
b. the mode takes into account every score in the distribution
c. the mode is appropriate for any scale of measurement
d. the mode can be a category name or a number

13. Your text noted which of the following as a characteristic of the median?

a. the median is less influenced by skew than the mean
b. the formula to find the value of the median is $\frac{N+1}{2}$
c. both a. and b.
d. neither a. nor b

14. According to the textbook, how can you conceptualize the idea of a measure of central tendency?

a. central tendency could be the average score
b. central tendency could be the most common score
c. central tendency could be the score in the middle of the distribution
d. all of the above

15. The mode is the only appropriate measure of central tendency when _____.

a. you have nominal data
b. you have ordinal data
c. you have severely skewed data
d. none of the above

Short-Answer Questions

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

2. Our company has three divisions. Based on the capital invested in each division, 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 $.18 per dollar invested. Do you agree with this conclusion? Why or why not?

3. At many colleges, there are a small number of introductory courses that have very 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 at a college is 23. Is the median larger or smaller?

Problems

1. Find the mean, median and mode of these numbers: 9, 10, 9, 9, 11, 9, 10, 9, 9, 10

2. Find the mean, median and mode of these numbers: 5, 4, 2, 1, 6, 5, 3, 2, 4, 2, 5, 5, 2

3. Find the mean, median and mode of these numbers: 9, 11, 12, 10, 12, 11, 8, 12, 10, 11, 12

4. Find the mean, median and mode of these numbers: 8, 7, 7, 9, 5, 7, 7, 8, 6, 6, 7, 7

5. In a study, college students and older adults were presented with this statement: "Social security payments are important as a means of income for the elderly." After reading this statement, participants identified the extent to which they agreed or disagreed with the statement using a Likert Scale, which is a common measurement of attitudes. Participants had five response options, ranging from "strongly agree" (1) to "strongly disagree" (5). The two frequency distributions that follow show the responses of college students and older adults to the statement above. Treat these data as population data. For each distribution, find the mean, median and mode.

College Students' Level of Agreement *X*	*f*	Older Adults' Level of Agreement *X*	*f*
5	4	5	2
4	4	4	1
3	2	3	2
2	1	2	4
1	1	1	6

6. Assume the numbers below are the population of representative of human birth weights in 1984 and 2004. For each year, find the mean, median and mode of birth weight.

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

7. The following data are representative of the age (in months) at which 90 percent of babies first stand without support (Shaffer & Kipp, 2010). Treat this data as a population and find the mean, median, and mode. Then, identify whether the data are skewed and identify the direction of the skew (if any).

Age (months) *X*	*f*
16	4
15	4
14	6
13	10
12	6
11	5
10	2
9	1

8. Find the mean, median, and mode of the following set of sample data. If the data are skewed, identify the direction of the skew.

X	*f*
30	2
29	6
28	4
27	3
26	2
25	0
24	1

ANSWERS
Multiple-Choice Questions

1. c. modes
2. a. the sum of the squared deviation scores is a minimum
   Explanation: If deviation scores are squared and then summed, which is expressed as $\sum(X - \bar{X})^2$, the result will be smaller than the sum you get if any number other than $\bar{X}$ is used. That is, using the mean minimizes the sum of squared deviations.
3. b. sample median
   Explanation: In this study, a sample has been collected. And the data described is very negatively skewed (a small number of cars going slowly or very slowly, with most scores being right around 45 mph), which makes the median a more appropriate measure of central tendency.
4. b. greater than 30°F
   Explanation: To answer this question, you have to take into account the number of days in each month. Because February is the shortest month of the year, February's average will make up a smaller portion of the weighted mean than March's average, which would pull the weighted mean above 30°F.
5. b. low scores are more frequent, and the mean is larger than the median
   Explanation: In the case of a positive skew, scores are clustered around the low end of the distribution and the mean is larger than the median because it is more affected by the extreme scores in the tail of the distribution than the median.
6. d. none of the above
   Explanation: You can only calculate the mean if you have an interval or ratio scale of measurement. And, the median is a more appropriate measure of central tendency for skewed data.
7. d. mode, freshmen
8. b. less than $50,000
   Explanation: These three groups do not have equal *N*s, so you need to estimate a weighted mean. More people have Bachelor's degrees than Master's degrees. And, more people have Master's degrees than doctorates. As a result, the weight of the salary for Bachelor's degrees will be the largest and will pull the weighted mean down below $50,000.
9. b. 3.25
   Formula:
   $$\bar{X}_W = \frac{N_1\bar{X}_1 + N_2\bar{X}_2 + \cdots N_K\bar{X}_K}{N_1 + N_2 + \cdots + N_K} = \frac{(5)(2) + (5)(3) + (10)(4)}{5 + 5 + 10}$$
   $$= \frac{10 + 15 + 40}{20} = \frac{65}{20} = 3.25$$
10. a. median
    Explanation: You have open-ended categories, so a mean is not appropriate.
11. c. the sum of the deviation scores and the size of the sum of the squared deviations
12. b. the mode takes into account every score in the distribution
13. a. the median is less influenced by skew than the mean
    Explanation: b. is incorrect because that formula identifies the *position* of the median, not the value of the median.
14. d. all of the above
    Explanation: Choice a. is the mean, b. is the mode, and c. is the median, so all of the options are ways to conceptualize central tendency.
15. a. you have nominal data

Short-Answer Questions

1. The median should be used instead of the mean when a) you have ordinal data, b) the data are skewed, or c) there are open-ended class intervals.
2. The conclusion is correct only if the number of dollars invested in each of the three divisions is the same, which is highly unlikely. If the amount of money invested was different, you would need to create a weighted mean, instead of just averaging the three averages, to take into account the different sample sizes.
3. The median would be smaller. The mean will be pulled in the direction of the skew, which would be towards the larger class sizes. Because the median isn't as affected by skew, it remains closer to the center of the distribution.

Problems

1. Mode = 9, Median = 9, Mean = 9.50
2. Mode = 2 and 5, Median = 4, Mean = 3.54
3. Mode = 12, Median = 11, Mean = 10.73
4. Mode = 7, Median = 7, Mean = 7.00
5. College Students:
   Mode = 4 and 5 (Notice that, although this distribution has two modes, it is not bimodal. For a distribution to be described as bimodal, the two high-frequency scores need to be separated by one or more low-frequency scores.)
   Median = 4, Mean = 3.75.

   Older Adults:
   Mode = 1, Median = 2, Mean = 2.27.

College Students' Level of Agreement X	f	fX	Older Adults' Level of Agreement X	f	fX
5	4	20	5	2	10
4	4	16	4	1	4
3	2	6	3	2	6
2	1	2	2	4	8
1	1	1	1	6	6
	Σf = N = 12	Σf X = 45		Σf = N = 15	Σf X = 34

For College Students

$$\mu = \frac{\sum fX}{N} = \frac{45}{12} = 3.75$$

For Older Adults

$$\mu = \frac{\sum fX}{N} = \frac{34}{15} = 2.27$$

6. 1984: Mode = 7, Median = 7, Mean = 7.13.
   2004: Mode = 7, Median = 7.5, Mean = 7.57.

1984 Weight (lbs.) X	f	fX	2004 Weight (lbs.) X	f	fX
10	1	10	10	1	10
9	1	9	9	3	27

8	2	16	8	3	24
7	7	49	7	4	28
6	3	18	6	2	12
5	1	5	5	1	5
	Σf = N = 15	Σf X = 107		Σf = N = 14	Σf X = 106

For 1984

$$\mu = \frac{\sum fX}{N} = \frac{107}{15} = 7.13$$

For 2004

$$\mu = \frac{\sum fX}{N} = \frac{106}{14} = 7.57$$

7. Mode = 13, Median = 13, Mean = 13.00. Skew: There is no skew

Age (months) X	f	fX
16	4	64
15	4	60
14	6	84
13	10	130
12	6	72
11	5	55
10	2	20
9	1	9
	Σf = N = 38	ΣfX = 494

$$\mu = \frac{\sum fX}{N} = \frac{494}{38} = 13.00$$

8. Mode: 29, Median: 28, Mean = 27.94. Skew: Negative

X	f	fX
30	2	60
29	6	174
28	4	112
27	3	81
26	2	52
25	0	0
24	1	24
	Σf = 18	ΣfX = 503

$$\bar{X} = \frac{\sum fX}{N} = \frac{503}{18} = 27.94$$

The skew is negative because the mean is smaller than the median and the mean is pulled in the direction of the skew.

References

Shaffer, D. R., & Kipp, K. (2010). *Developmental Psychology: Childhood and Adolescence* (8th ed.). Belmont, CA: Wadsworth, Cengage Learning.