

Chapter 3 : Central Tendency



Overview

- Definition: Central tendency is a statistical measure to determine a single score that defines the center of a distribution.
 - The goal of central tendency is to find the single score that is most typical or most representative of the entire group.
 - Measures of central tendency are also useful for making comparisons between groups of individuals or between sets of figures.
 - For example, weather data indicate that for Seattle, Washington, the average yearly temperature is 53° and the average annual precipitation is 34 inches.



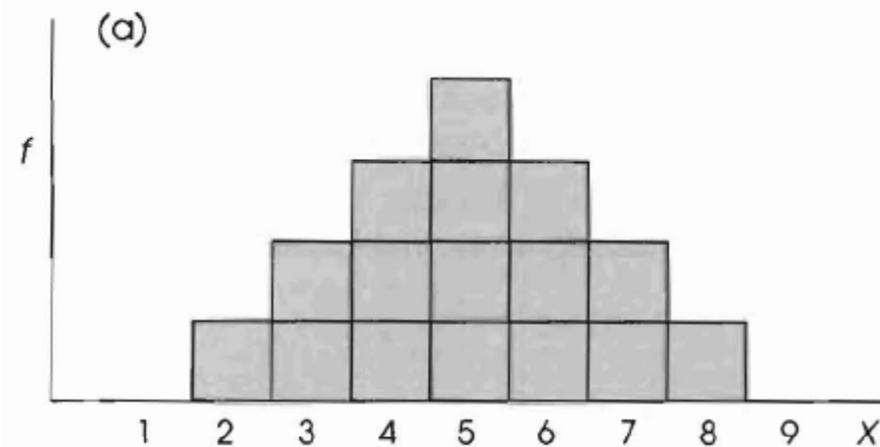
Overview cont.

- By comparison, the average temperature in Phoenix, Arizona, is 71° and the average precipitation is 7.4 inches.
- Clearly, there are problems defining the "center" of a distribution.
- Occasionally, you will find a nice, neat distribution like the one shown in Figure 3.2(a), for which everyone will agree on the center. (See next slide.)
- But you should realize that other distributions are possible and that there may be different opinions concerning the definition of the center. (See in two slides)

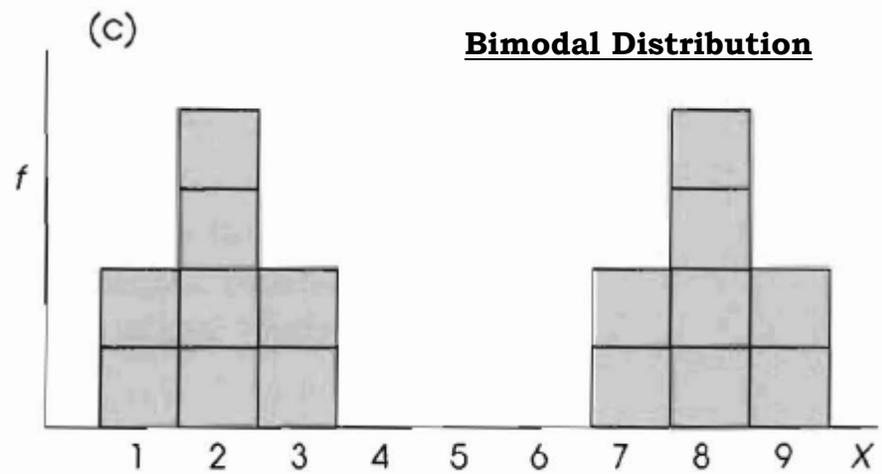
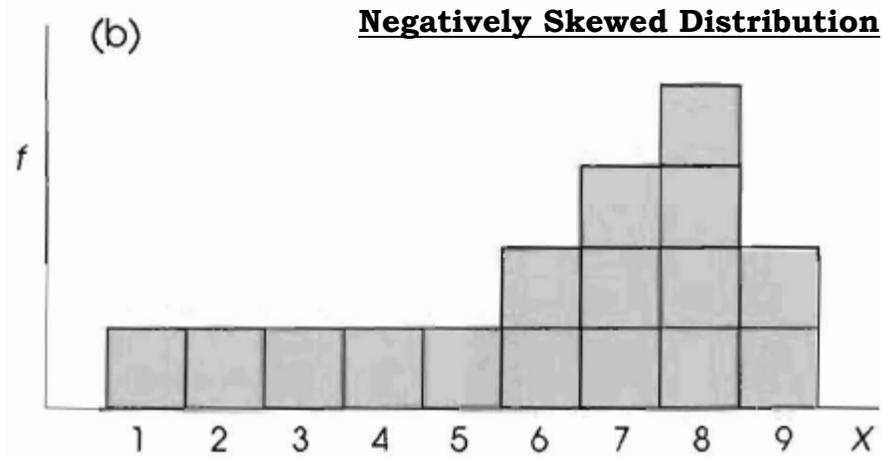


Overview cont.

- To deal with these problems, statisticians have developed three different methods for measuring central tendency:
 - Mean
 - Median
 - Mode



Overview cont.



The Mean

- The mean for a population will be identified by the Greek letter *mu*, μ (pronounced "mew"), and the mean for a sample is identified by *M* or \bar{x} (read "x-bar").
- Definition: The mean for a distribution is the sum of the scores divided by the number of scores.
- The formula for the population mean is

$$\mu = \frac{\sum X}{N}$$

- The formula for the sample mean uses symbols that signify sample values:

$$\text{sample mean} = M = \frac{\sum X}{n}$$



The Mean cont.

- In general, we will use Greek letters to identify characteristics of a population and letters of our own alphabet to stand for sample values.
 - Thus, if a mean is identified with the symbol M , you should realize that we are dealing with a sample.
 - Also note that n is used as the symbol for the number of scores in the sample.



The Weighted Mean

- Often it is necessary to combine two sets of scores and then find the overall mean for the combined group.
- In summary, when two samples are combined, the weighted mean is obtained as follows:

$$\text{weighted mean} = \frac{\text{combined sum}}{\text{combined } n} = \frac{\Sigma X_1 + \Sigma X_2}{n_1 + n_2}$$



Computing the Mean from a Frequency Distribution Table

- The formula to calculate the mean from a frequency distribution table is as follows:

$$M = \frac{\sum f X}{n}$$



Characteristics of the Mean

- The mean has many characteristics that will be important in future discussions.
 - Changing a score
 - Changing the value of any score will change the mean.
 - Introducing a new score or removing a score
 - In general, the mean is determined by two values: $\sum X$ and N (or n).
 - Whenever either of these values are changed, the mean also is changed.



Characteristics of the Mean cont.

- Adding or subtracting a constant from each score
 - If a constant value is added to every score in a distribution, the same constant will be added to the mean.
- Multiplying or dividing each score by a constant
 - If every score in a distribution is multiplied by (or divided by) a constant value, the mean will change in the same way.



The Median

- Definition: The median is the score that divides a distribution in half so that 50% of the individuals in a distribution have scores at or below the median.
- Earlier, when we introduced the mean, specific symbols and notation were used to identify the mean and to differentiate a sample mean and a population mean.
- For the median, however, there are no symbols or notation.
- Instead, the median is simply identified by the word median.
- In addition, the definition and the computations for the median are identical for a sample and for a population.



The Median cont.

- The goal of the median is to determine the midpoint of the distribution.
- This commonsense goal is demonstrated in the following two examples which show how the median for most distributions can be found simply by counting scores.

EXAMPLE 3.4

This example demonstrates the calculation of the median when n is an odd number. With an odd number of scores, you list the scores in order (lowest to highest), and the median is the middle score in the list. Consider the following set of $N = 5$ scores, which have been listed in order:

3, 5, 8, 10, 11

The middle score is $X = 8$, so the median is equal to 8. If the scores are measurements of a continuous variable, the median corresponds to the precise point $X = 8.00$. In a frequency distribution graph, this point divides the area of the graph exactly in half (Figure 3.5). The amount of area above the median consists of $2\frac{1}{2}$ "boxes," the same as the area below the median (shaded portion).

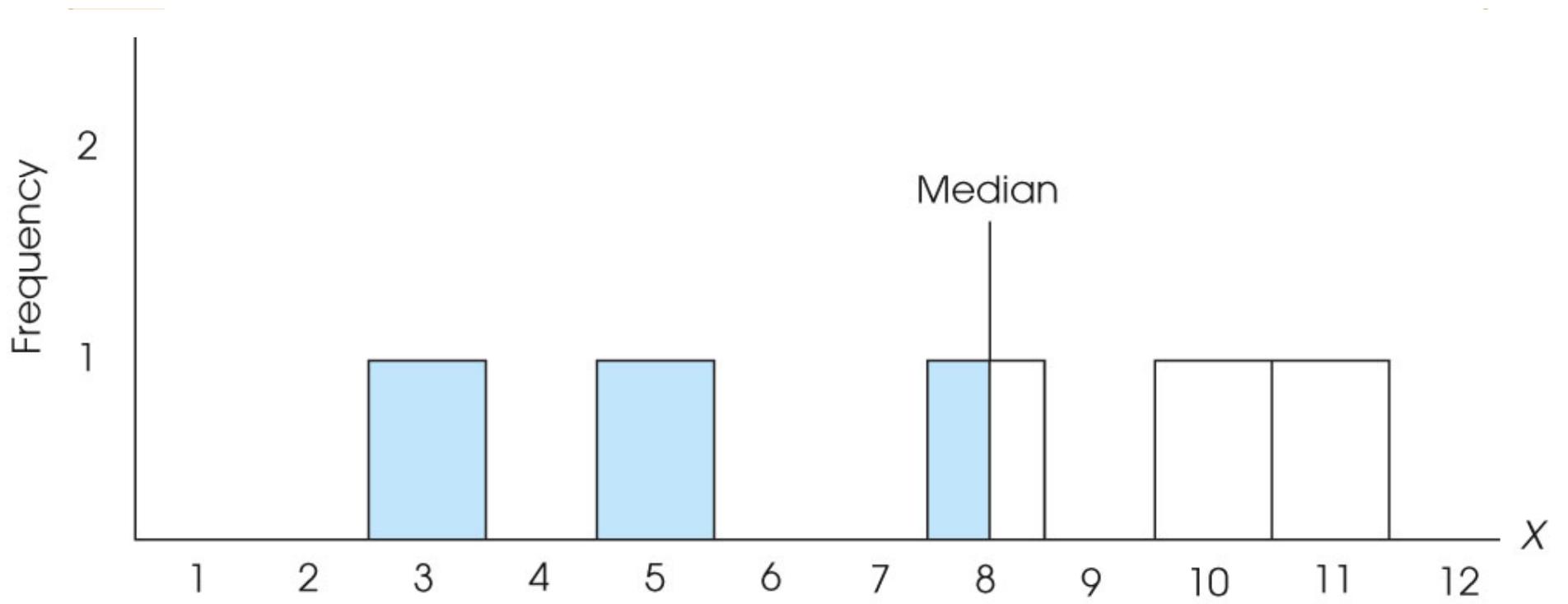


Fig. 3-5, p. 83

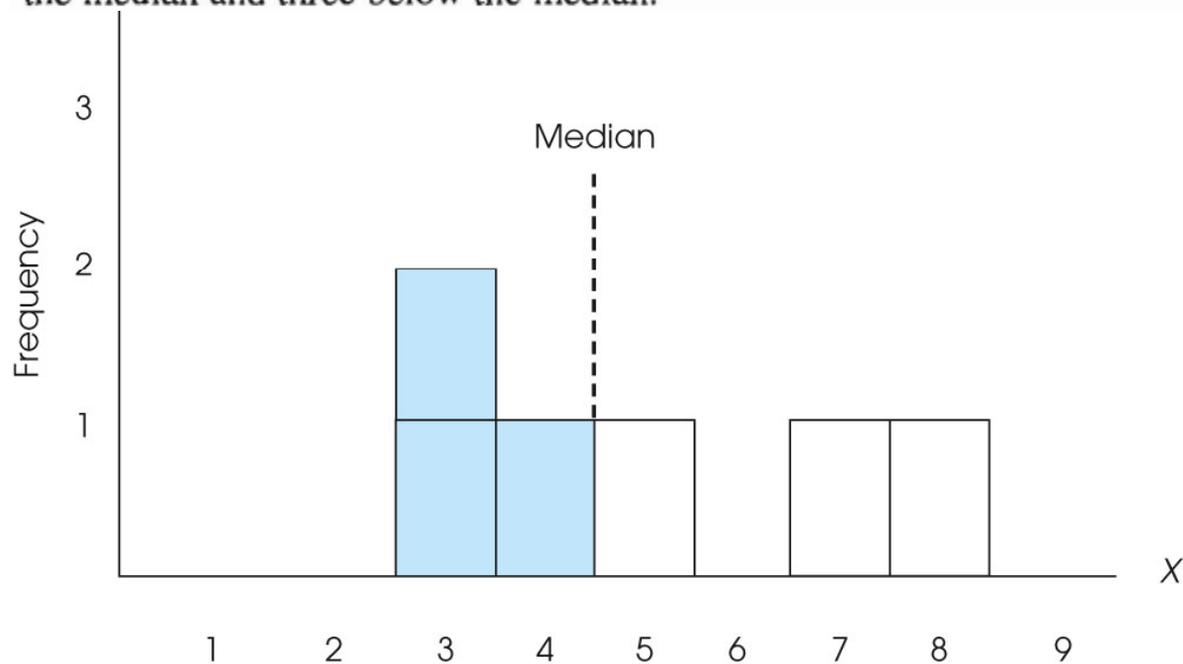
This example demonstrates the calculation of the median when n is an even number. With an even number of scores in the distribution, you list the scores in order (lowest to highest) and then locate the median by finding the average of the middle two scores. Consider the following population:

3, 3, 4, 5, 7, 8

Now we select the middle pair of scores (4 and 5), add them together, and divide by 2:

$$\text{median} = \frac{4 + 5}{2} = \frac{9}{2} = 4.5$$

In terms of a graph, we see again that the median divides the area of the distribution exactly in half (Figure 3.6). There are three scores (or boxes) above the median and three below the median.



The Mode

- Definition: In a frequency distribution, the mode is the score or category that has the greatest frequency.
 - As with the median, there are no symbols or special notation used to identify the mode or to differentiate between a sample mode and a population mode.
 - In addition, the definition of the mode is the same for a population and for a sample distribution.
 - Although a distribution will have only one mean and only one median, it is possible to have more than one mode.
 - Specifically, it is possible to have two or more scores that have the same highest frequency.



The Mode cont.

- In a frequency distribution graph, the different modes will correspond to distinct, equally high peaks.
- A distribution with two modes is said to be *bimodal*, and a distribution with more than two modes is called *multimodal*.



Selecting a Measure of Central Tendency

- How do you decide which measure of central tendency to use?
 - The answer to this question depends on several factors.
 - Before we discuss these factors, however, note that you usually can compute two or even three measures of central tendency for the same set of data.
 - Although the three measures often produce similar results, there are situations in which they are very different (see Section 3.6).



When to use the Median

- Extreme scores or skewed distributions
 - When a distribution has a few extreme scores, scores that are very different in value from most of the others, then the mean may not be a good representative of the majority of the distribution.
 - The problem comes from the fact that one or two extreme values can have a large influence and cause the mean to be displaced.
- Undetermined values
 - Occasionally, you will encounter a situation in which an individual has an unknown or undetermined score.



When to use the Median cont.

- Open-ended distributions
 - A distribution is said to be open-ended when there is no upper limit (or lower limit) for one of the categories.
 - The table at the upper right-hand corner provides an example of an open-ended distribution, showing the number of pizzas eaten during a 1 month period for a sample of $n = 20$ high school students.
- Ordinal scale
 - Many researchers believe that it is not appropriate to use the mean to describe central tendency for ordinal data.

Number of Pizzas (X)	<i>f</i>
5 or more	3
4	2
3	2
2	3
1	6
0	4



When to use the Median cont.

- When scores are measured on an ordinal scale, the median is always appropriate and is usually the preferred measure of central tendency.



When to use the Mode

- Nominal scales
 - The primary advantage of the mode is that it can be used to measure and describe the central tendency for data that are measured on a nominal scale.
- Discrete variables
 - Recall that discrete variables are those that exist only in whole, indivisible categories.
 - Often, discrete variables are numerical values, such as the number of children in a family or the number of rooms in a house.
- Describing shape
 - Because the mode requires little or no calculation, it is often included as a supplementary measure along with the mean or median as a no-cost extra.



Central Tendency and the Shape of the Distribution

- We have identified three different measures of central tendency, and often a researcher calculates all three for a single set of data.
- Because the mean, the median, and the mode are all trying to measure the same thing (central tendency), it is reasonable to expect that these three values should be related.
- In fact, there are some consistent and predictable relationships among the three measures of central tendency.
- Specifically, there are situations in which all three measures will have exactly the same value.



Central Tendency and the Shape of the Distribution cont.

- On the other hand, there are situations in which the three measures are guaranteed to be different.
- In part, the relationships among the mean, median, and mode are determined by the shape of the distribution.
- We will consider two general types of distributions.
 - Symmetrical
 - Skewed Distributions



Symmetrical Distributions

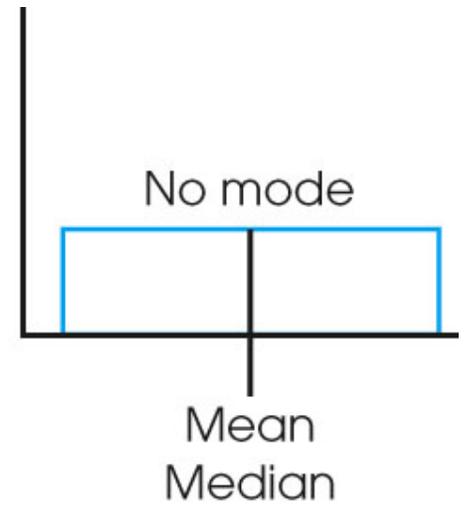
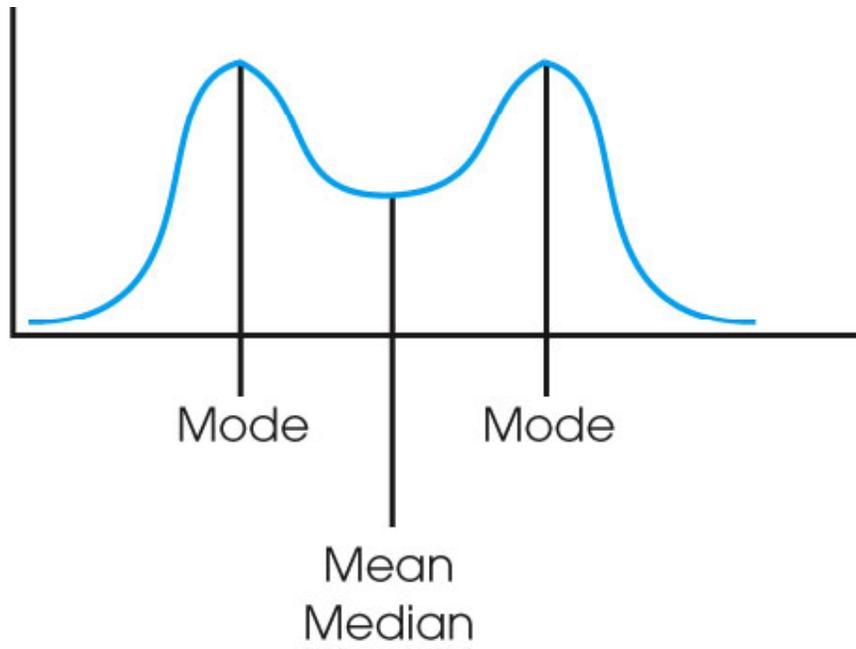
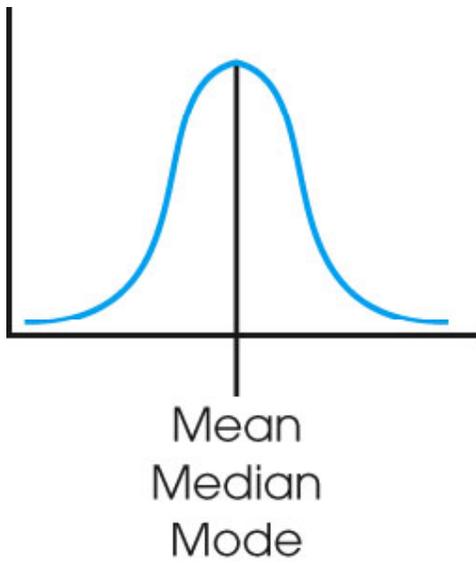
- For a symmetrical distribution, the right-hand side of the graph is a mirror image of the left-hand side.
- If a distribution is perfectly symmetrical, the median is exactly at the center because exactly half of the area in the graph will be on either side of the center.
- For a perfectly symmetrical distribution with one mode, all three measures of central tendency, the mean, the median, and the mode, will have the same value.
- For a roughly symmetrical distribution, the three measures are clustered together in the center of the distribution.



Symmetrical Distributions

- On the other hand, a bimodal distribution that is symmetrical [see Figure 3.13(b)] will have the mean and median together in the center with the modes on each side.
- A rectangular distribution [see Figure 3.13(c)] has no mode because all X values occur with the same frequency.
- Still, the mean and the median are in the center of the distribution.





Skewed Distributions

- Distributions are not always symmetrical; quite often, they are lopsided, or skewed.
 - For example, Figure 3. 14(a) shows a positively skewed distribution.
- In skewed distributions, especially distributions for continuous variables, there is a strong tendency for the mean, median, and mode to be located in predictable positions.



