Statistics 1
Chapter 3

Describing Data 2

Chapter Three

Describing Data: Measures of Central Tendency

GOALS

When you have completed this chapter, you will be able to:

ONE

Calculate the arithmetic mean, median, mode, weighted mean, and the geometric mean.

TWO

Explain the characteristics, uses, advantages, and disadvantages of each measure of location.

THREE

Identify the position of the arithmetic mean, median, and mode for both a symmetrical and a skewed distribution.

Population Mean

- For ungrouped data, the population mean is the sum of all the population values divided by the total number of population values:

$$\mu = \Sigma X / N$$

- where μ stands for the population mean.
- N is the total number of observations in the population.
- X stands for a particular value.
- Σ indicates the operation of adding.

EXAMPLE 1

- Parameter: a measurable characteristic of a population.
- The Kiers family owns four cars. The following is the mileage attained by each car: 56,000, 23,000, 42,000, and 73,000. Find the average miles covered by each car.
- The mean is (56,000 + 23,000 + 42,000 + 73,000)/4 = 48,500

Sample Mean

- For ungrouped data, the sample mean is the sum of all the sample values divided by the number of sample values:

$$\overline{X} = \Sigma X / n$$

- where $\overline{X}$ stands for the sample mean
- n is the total number of values in the sample

EXAMPLE 2

- Statistic: a measurable characteristic of a sample.
- A sample of five executives received the following amounts of bonus last year: \$14,000, \$15,000, \$17,000, \$16,000, and \$15,000. Find the average bonus for these five executives.
- Since these values represent a sample size of 5, the sample mean is (14,000 + 15,000 +17,000 + 16,000 +15,000)/5 = \$15,400.

Properties of the Arithmetic Mean

- Every set of interval-level and ratio-level data has a mean.
- All the values are included in computing the mean.
- A set of data has a unique mean.
- The mean is affected by unusually large or small data values.
- The arithmetic mean is the only measure of central tendency where the sum of the deviations of each value from the mean is zero.

EXAMPLE 3

- Consider the set of values: 3, 8, and 4. The mean is 5. Illustrating the fifth property, (3-5) + (8-5) + (4-5) = -2 +3 - 1 = 0. In other words,

$$\Sigma(X - \overline{X}) = 0$$

Weighted Mean

- The weighted mean of a set of numbers $X_1, X_2, ..., X_n$, with corresponding weights $w_1, w_2, ...,w_n$, is computed from the following formula:

$$\overline{X_w} = (w_1 X_1 + w_2 X_2 + \ldots + w_n X_n) / (w_1 + w_2 + \ldots w_n)$$

$$\overline{X}_w = \Sigma(w * X) / \Sigma w$$

EXAMPLE 6

- During a one hour period on a hot Saturday afternoon cabana boy Chris served fifty drinks. Compute the weighted mean of the price of the drinks. (*Price ($), Number sold*): (.50,5), (.75,15), (.90,15), (1.10,15).
- The weighted mean is: $(.50x5 + .75x15 + .90x15 + 1.10x15)/(5 + 15 + 15 + 15) = $43.75/50 = $0.875

The Median

- Median: The midpoint of the values after they have been ordered from the smallest to the largest, or the largest to the smallest. There are as many values above the median as below it in the data array.
- *Note:* For an even set of numbers, the median will be the arithmetic average of the two middle numbers.

EXAMPLE 4

- Compute the median for the following data.
- The age of a sample of five college students is: 21, 25, 19, 20, and 22.
- Arranging the data in ascending order gives: 19, 20, 21, 22, 25. Thus the median is 21.
- the height of four basketball players, in inches, is 76, 73, 80, and 75.
- Arranging the data in ascending order gives: 73, 75, 76, 80. Thus the median is 75.5

Properties of the Median

- There is a unique median for each data set.
- It is not affected by extremely large or small values and is therefore a valuable measure of central tendency when such values occur.
- It can be computed for ratio-level, interval-level, and ordinal-level data.
- It can be computed for an open-ended frequency distribution if the median does not lie in an open-ended

The Mode

- The mode is the value of the observation that appears most frequently.
- EXAMPLE 5: The exam scores for ten students are: 81, 93, 84, 75, 68, 87, 81, 75, 81, 87. Since the score of 81 occurs the most, the modal score is 81.

Geometric Mean

- The geometric mean (GM) of a set of n numbers is defined as the *nth* root of the product of the *n* numbers. The formula is:

$$GM = \sqrt[n]{(X_1)(X_2)(X_3)\ldots(X_n)}$$

-
  - Here the geometric mean is used to average percents, indexes, and relatives.

EXAMPLE 7

- The interest rate on three bonds were 5, 7, and 4 percent.
- The geometric mean is $GM = \sqrt[3]{(7)(5)(4)}$ =5.192.
- The arithmetic mean is (6 + 3 + 2)/3 =5.333.
- The GM gives a more conservative profit figure because it is not heavily weighted by the rate of 7 percent.

Geometric Mean *continued*

- Another use of the geometric mean is to determine the average percent increase in sales, production or other business or economic series from one time period to another. The formula for this type of problem is:

$$GM = \sqrt[n]{(\text{value at end of period}) / (\text{value at beginning of period})} - 1$$

EXAMPLE 8

- The total number of females enrolled in American colleges increased from 755,000 in 1986 to 835,000 in 1995.
- Here n = 10, so (n-1) = 9.

$$GM = \sqrt[8]{835,000 / 755,000} - 1 = .0127 .$$

- That is, the geometric mean rate of increase is 1.27%.

The Mean of Grouped Data

- The mean of a sample of data organized in a frequency distribution is computed by the following formula:

$$\overline{X} = \frac{\Sigma Xf}{\Sigma f} = \frac{\Sigma Xf}{n}$$

EXAMPLE 9

- A sample of ten movie theaters in a large metropolitan area tallied the total number of movies showing last week. Compute the mean number of movies showing.
- $$\overline{X} = \frac{\Sigma Xf}{\Sigma f} = \frac{\Sigma Xf}{n}$$

EXAMPLE 9 *continued*

Movies showing	frequency f	class midpoint X	(f)(x)
1-2	1	1.5	1.5
3-4	2	3.5	7.0
5-6	3	5.5	16.5
7-8	1	7.5	7.5
9-10	3	9.5	28.5
Total	10		61

61/10 = 6.1 movies

The Median of Grouped Data

- The median of a sample of data organized in a frequency distribution is computed by the following formula:
- Median = L + [(n/2 - CF)/f] (i)
- where L is the lower limit of the median class, CF is the cumulative frequency preceding the median class, f is the frequency of the median class, and i is the median class interval.

Finding the Median Class

- To determine the median class for grouped data:
- Construct a cumulative frequency distribution.
- Divide the total number of data values by 2.
- Determine which class will contain this value. For example, if n=50, 50/2 = 25, then determine which class will contain the 25th value - the median class.

EXAMPLE 10

Movies showing	Frequency	Cumulative Frequency
1-2	1	1
3-4	2	3
5-6	3	6
7-8	1	7
9-10	3	10

- The median class is 5-6, since it contains the 5th value (n/2 =5)

EXAMPLE 10 *continued*

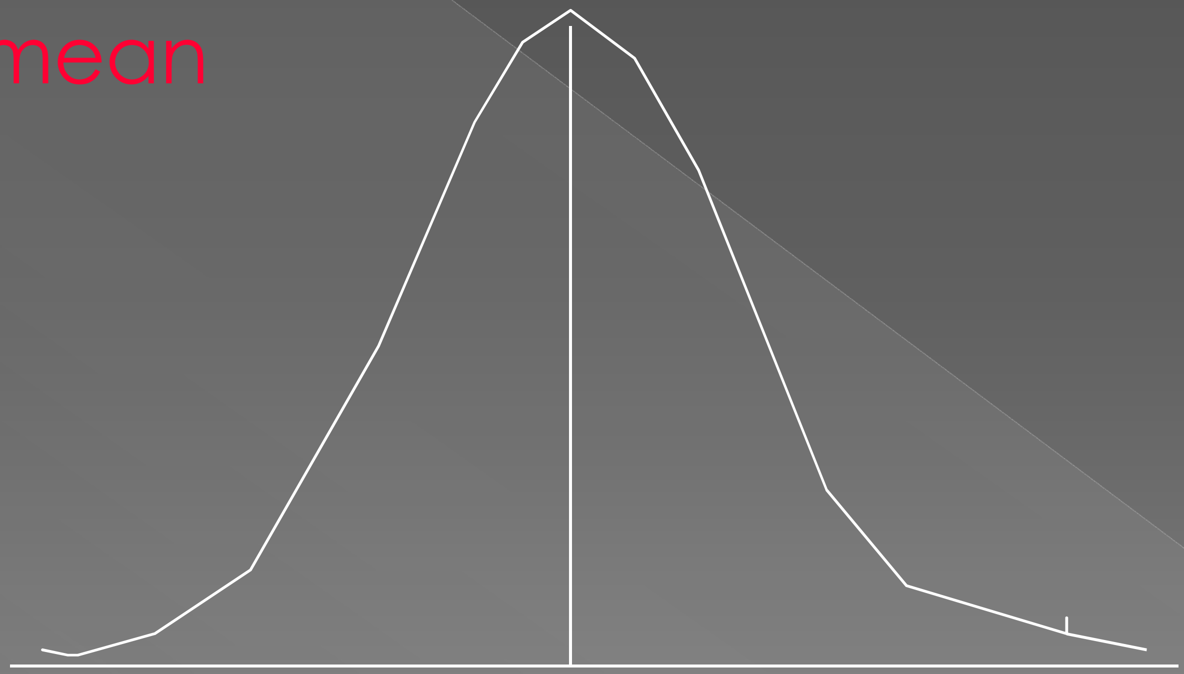
- From the table, L=5, n=10, f=3, i=2, CF=3.
- Thus, Median= 5 + [((10/2) - 4)/3](2)= 6.33

The Mode of Grouped Data

- The mode for grouped data is approximated by the midpoint of the class with the largest class frequency.
- The modes in EXAMPLE 10 are 5.5 and 9.5. When two values occur a large number of times, the distribution is called bimodal, as in example 10.

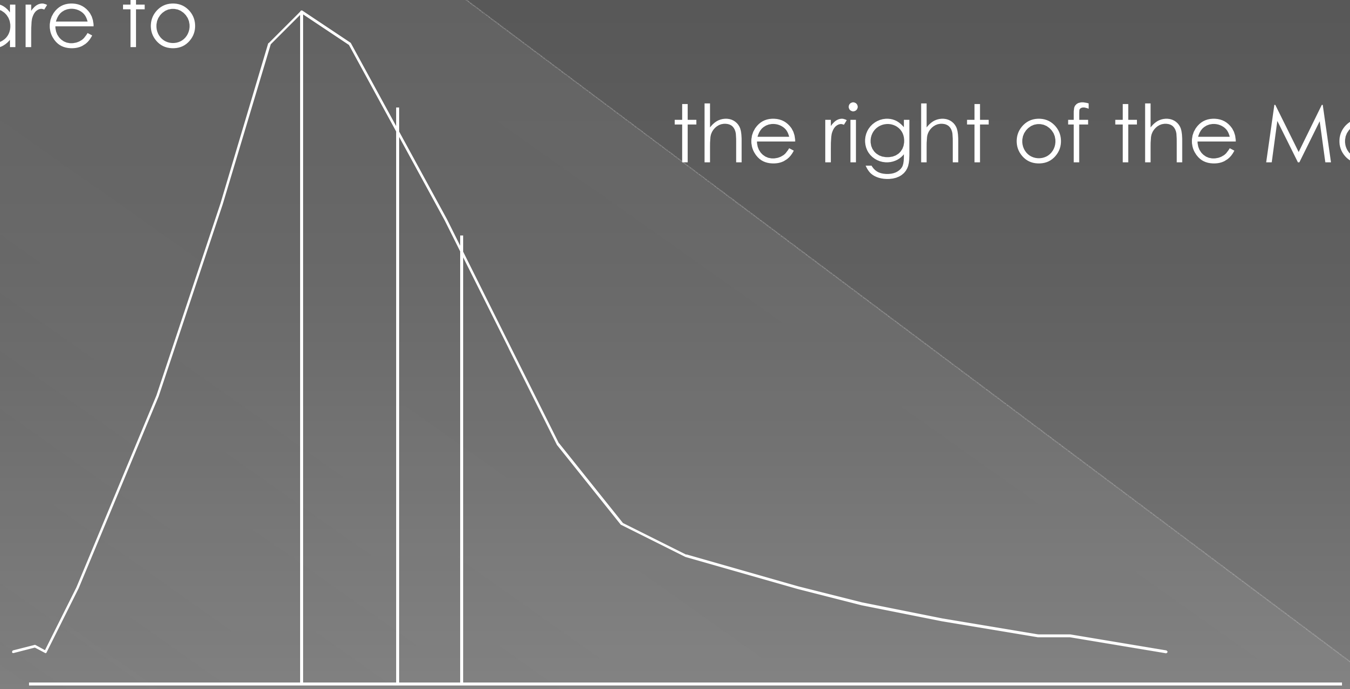
Symmetric Distribution

- zero skewness mode = median = mean



Right Skewed Distribution

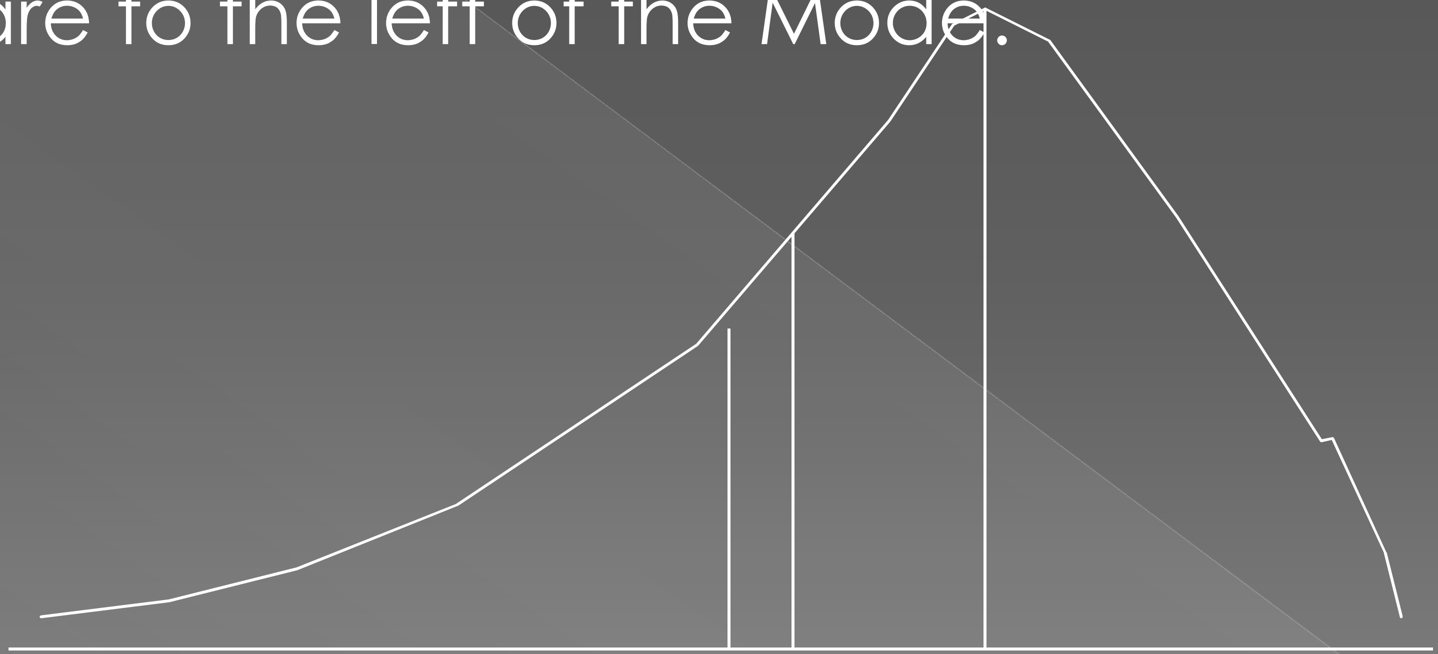
- positively skewed: Mean and Median are to
- the right of the Mode.



- Mode<Median<Mean

Left Skewed Distribution

- Negatively Skewed: Mean and Median are to the left of the Mode.



- Mean<Median<Mode

NOTE

- If two averages of a moderately skewed frequency distribution are known, the third can be *approximated*.
- Mode = mean - 3(mean - median)
- Mean = [3(median) - mode]/2
- Median = [2(mean) + mode]/3