Describing Data: Numerical Measures

Chapter 3

Learning Objectives

- Calculate the arithmetic mean, weighted mean median, and the mode.
- Explain the characteristics, uses, advantages, and disadvantages of each measure of location.
- Identify the position of the mean, median, and mode for both symmetric and skewed distributions.
- Compute and interpret the range, mean deviation, variance, and standard deviation.
- Understand the characteristics, uses, advantages, and disadvantages of each measure of dispersion.
- Calculate the mean and the standard deviation of grouped data.

The Population Mean

- For ungrouped data, the population mean is the sum of all the population values divided by the total number of population values:
- Population mean: $\mu = \frac{\sum x}{N}$, where:

μ: Represents the population mean.

N: Is the number of values in the population.

x : Represents any particular value.

$\sum x$: Is the sum of the x values in the population.

The Population Mean

- Example:

There are 12 automobile manufacturing companies in the United States. Listed below is the number of patents granted by the United States government to each company in a recent year.

Company	Number of Patents Granted	Company	Number of Patents Granted
General Motors	511	Mazda	210
Nissan	385	Chrysler	97
DaimlerChrysler	275	Porsche	50
Toyota	257	Mitsubishi	36
Honda	249	Volvo	23
Ford	234	BMW	13

Is this information a sample or a population? What is the arithmetic mean number of patents granted?

The Sample Mean

- For ungrouped data, the sample mean is the sum of all the sample values divided by the number of sample values.
- Sample Mean: $\bar{X} = \frac{\sum x}{n}$ where,
- $\bar{X}$: Is the sample Mean. It is read " X bar."
- n: Is the number of values in the sample.

The Sample Mean

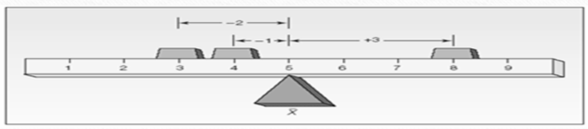
- Example: STC is studying the number of minutes used monthly by clients in the package "Postpaid 30". A random sample of 12 clients showed the following number of minutes used last month. What is the arithmetic mean number of minutes used.

90	77	94	89	119	112
91	110	92	100	113	83

$$\bar{X} = \frac{\sum x}{n} = \frac{90 + 77 + 94 + \cdots + 113 + 83}{12} = \frac{1170}{12} = 97.5\ Minutes$$

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.
- The mean is unique.
- The sum of the deviation of each value from the mean is zero.
- The mean is affected by unusually large or small data values.



The Weighted Mean

- The weighted mean is a special case of the arithmetic mean. It occurs when there are several observations of the same value.

Example: Suppose a near by Herfy Restaurant sold medium, large, and supersize soft drinks for 2.5, 5, 8 Saudi Riyals respectively. Of the last 10 drinks sold, 3 were medium, 4 were large, and 3 were supersize. What is the mean price of the last 10 drinks sold?

$$\bar{X} = \frac{2.5 + 2.5 + 2.5 + 5 + 5 + 5 + 5 + 8 + 8 + 8}{10} = 5.15\ SR$$

The Weighted Mean

- The mean selling price of the last 10 drinks is 5.15 SR.
- An easier way to find the mean selling price is to determine the weighted mean. That is, we multiply each observation by the number of times it happens.
- We will refer to the weighted mean as $\bar{X}_w$ This is read "X bar sub w."

$$\bar{X}_w = \frac{3(2.5) + 4(5) + 3(8)}{10} = \frac{51.5}{10} = 5.15\ SR$$

The Weighted Mean

- In this case, the weights are frequency counts. However, any measure of importance could be used as a weight. In general, the weighted mean of a set of numbers designated $X_1, X_2, X_3, \dots, X_n$ with the corresponding weights $w_1, w_2, w_3, \dots, w_n$ is computed by:

$$\bar{X}_w = \frac{w_1X_1 + w_2X_2 + w_3X_3 + \cdots + w_nX_n}{w_1 + w_2 + w_3 + \cdots + w_n}$$

This may be shortened to:

$$\bar{X}_w = \frac{\sum(wX)}{\sum w}$$

The Geometric Mean

- Useful in finding the average change of percentages, ratios, indexes, or growth rates over time.
- It has a wide application in business and economics because we are often interested in finding the percentage changes in sales, salaries, or economic figures, such as the GDP, which compound or build on each other.
- The geometric mean will always be less than or equal to the arithmetic mean.
- The geometric mean of a set of *n* positive numbers is defined as the *n*th root of the product of *n* values.
- $GM = \sqrt[n]{(x_1)(x_2)(x_3)\dots(x_n)}$

The Geometric Mean

- Example 1: Suppose you receive a 5 percent increase in salary this year and a 15 percent increase next year. The average annual percent increase is 9.886, not 10.0. Why is this so? We begin by calculating the geometric mean.

$$GM = \sqrt[2]{1.05 * 1.15} = \sqrt{1.2075} = 1.09886$$

This means that the average annual percent increase is 9.886.

Let us assume that the monthly earing was $3,000.

The Geometric Mean

- A second application of the geometric mean is to find out an average percentage change over a period of time. For example, if you earned $30,000 in 2000 and $50,000 in 2010, what is your annual rate of increase over the period?
- It is 5.24 percent.
- The rate of increase is determined from the following formula.
- $GM = \sqrt[n]{\frac{Value\ at\ end\ of\ period}{Value\ at\ start\ of\ period}} - 1$

The Geometric Mean

- Example 2: During the decade of the 1990s, and into the 2000s, Las Vegas, Nevada, was the fastest-growing city in the United States. The population increased from 258,295 in 1990 to 607,876 in 2009. This is an increase of 349,581 people, or a 135.3 percent increase over the period. The population has more than doubled. What is the average *annual* percent increase?

The Median

- The Median is the midpoint of the values after they have been ordered from the smallest to the largest (ascending order) or from the largest to the smallest (descending order)

The Median

Example 1: The ages for a sample of five college students are:

21, 25, 19, 20, 22.

Arranging the data in ascending order gives:

19, 20, 21, 22, 25

Thus, the median is …

Example 2: The height of four basketball players, in inches, are:

76, 73, 80, 75.

Arranging the data in descending order gives:

80, 76, 75, 73

Thus, the median is…

Properties of the Median

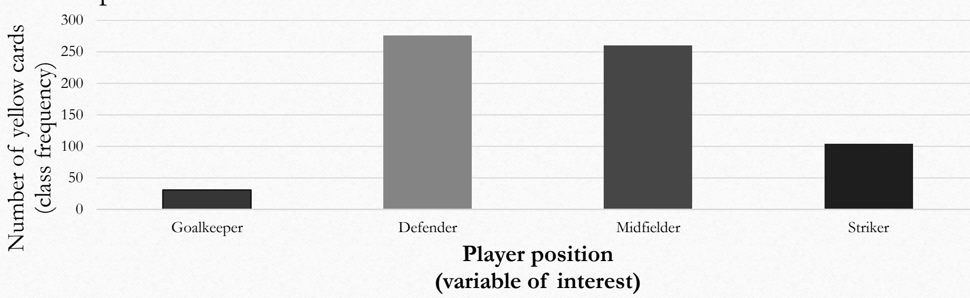
- There is a unique median for each data set.
- It is not affected by extremely large or small values and is therefore a valuable measure of central tendency when such values occur.
- It can be computed for ratio-level, interval-level, and ordinal-level data.
- It can be computed for an open-ended frequency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

The Mode

- The mode is the value of the observation that appears most frequently.

The Mode

- Example1:



The Mode

- Example 2:

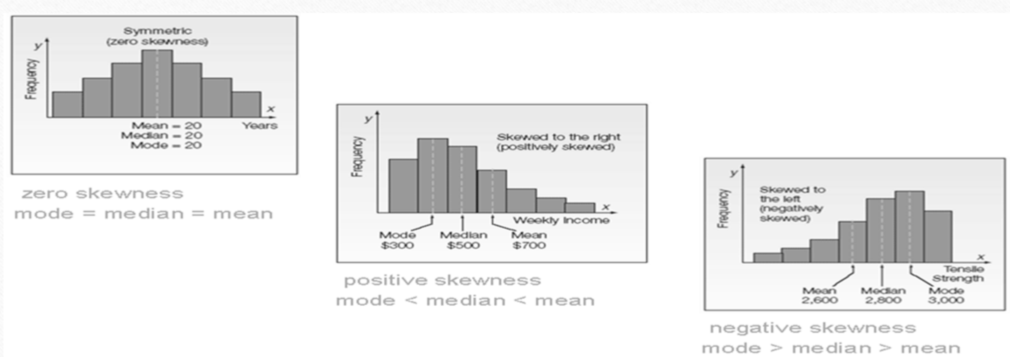
The annual salaries of quality-control managers in selected states are shown below. What is the modal annual salary?

State	Salary	State	Salary	State	Salary
Arizona	$35,000	Illinois	$58,000	Ohio	$50,000
California	49,100	Louisiana	60,000	Tennessee	60,000
Colorado	60,000	Maryland	60,000	Texas	71,400
Florida	60,000	Massachusetts	40,000	West Virginia	60,000
Idaho	40,000	New Jersey	65,000	Wyoming	55,000

Properties of the Mode

- It can be determined for all levels of measurement.
- Not affected by extremely high or low values.
- Might not exist.
- For some data set there are more than one mode.

The Relative Positions of the Mean, Median and the Mode



Dispersion

Why Study Dispersion?

- A measure of location, such as the mean or the median, only describes the center of the data. It is valuable from that standpoint, but it does not tell us anything about the spread of the data.
- For example, if your nature guide told you that the river ahead averaged 3 feet in depth, would you want to wade across on foot without additional information? Probably not. You would want to know something about the variation in the depth.
- A second reason for studying the dispersion in a set of data is to compare the spread in two or more distributions.

The Range

- The range is the difference between the largest value and the smallest values in a data set.

$$Range = Largest\ value\ - Smallest\ value$$

- Very easy to calculate and understand.
- Only two values are considered.

The Range

- Example: The number of cappuccinos sold at the Starbucks location in Sahara Mall between 4 and 7 p.m. for a sample of 5 days last year were 20. 40, 50, 60, and 80. Determine the range for the number of cappuccinos sold.

$$Range = \cdots$$

The Mean Deviation

- The mean deviation is the arithmetic mean of the absolute values of the deviations from the arithmetic mean.

$$MD = \frac{\sum|X - \bar{X}|}{n}$$

- Uses all the values.
- Easy to understand.

The Mean Deviation

- Example: The number of cappuccinos sold at the Starbucks location in Sahara Mall between 4 and 7 p.m. for a sample of 5 days last year were 20, 40, 50, 60, and 80. In the Granada Center, the number of cappuccinos sold at Starbucks between 4 and 7 p.m. for a sample of 5 days last year were 20, 49, 50, 51, and 80. Determine the mean, median, rang, and the mean deviation for the number of cappuccinos sold for each location. Compare the differences.

The Mean Deviation

Sahara Mall	The Granada Center
20, 40, 50, 60, and 80	20, 49, 50, 51, and 80
$\bar{X} = 50$	$\bar{X} = 50$
Median = 50	Median = 50
Range = 60	Range = 60
MD= 16	MD= 12.4

The mean deviation is 16 cappuccinos per day and show that the number of cappuccinos sold deviates, on average, by 16 from the mean of 50 cappuccinos per day.

Variance and Standard deviation

- The variance is the arithmetic mean of the squared deviation from the mean.
- The standard deviation is the square root of the variance.

Variance and Standard deviation

Population:

Variance

$$\sigma^2 = \frac{\sum(X-\mu)^2}{n}$$

Standard deviation

$$\sigma = \sqrt{\frac{\sum(X-\mu)^2}{n}}$$

Sample:

Variance

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

Standard deviation

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

Variance and Standard deviation

- The process of computing the variance:

1. Finding the mean.
2. Finding the difference between each observation and the mean.
3. Squaring the differences.
4. Summing the squared differences.
5. Dividing the sum by the number of items in the population OR the number of items in the sample minus one.

Variance and Standard deviation

- Example: The number of cappuccinos sold at the Starbucks location in Sahara Mall between 4 and 7 p.m. for a sample of 5 days last year were 20. 40, 50, 60, and 80. calculate the variance and the standard deviation of the number of cappuccinos sold.

$$s^2 = 500$$

$$s = 22.36$$

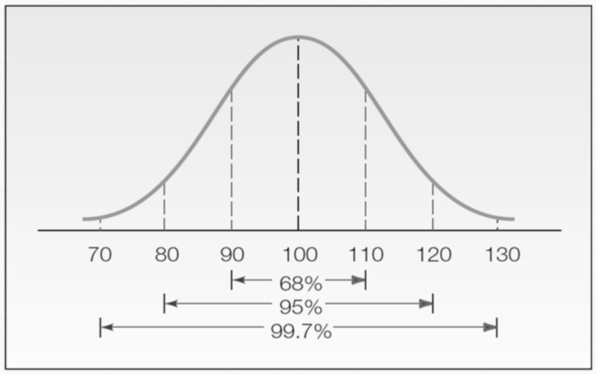
Chebyshev's Theorem

CHEBYSHEV'S THEOREM For any set of observations (sample or population), the proportion of the values that lie within k standard deviations of the mean is at least $1 - 1/k^2$, where k is any constant greater than 1.

Empirical Rule

EMPIRICAL RULE For a symmetrical, bell-shaped frequency distribution, approximately 68 percent of the observations will lie within plus and minus one standard deviation of the mean; about 95 percent of the observations will lie within plus and minus two standard deviations of the mean; and practically all (99.7 percent) will lie within plus and minus three standard deviations of the mean.

Empirical Rule



The arithmetic Mean of Grouped Data

ARITHMETIC MEAN OF GROUPED DATA	$\bar{X} = \frac{\Sigma fM}{n}$	[3–12]

where:

- $\bar{X}$ is the designation for the sample mean.
- M is the midpoint of each class.
- f is the frequency in each class.
- fM is the frequency in each class times the midpoint of the class.
- ΣfM is the sum of these products.
- n is the total number of frequencies.

The arithmetic Mean of Grouped Data

- Example: In an event we asked a sample from the audience about their ages and we construct the following table:

Class	Frequency(f)
5 up to 10	10
10 up to 15	3
15 up to 20	4
20 up to 25	3
25 up to 30	5

Q) Determine the arithmetic mean age.

The arithmetic Mean of Grouped Data

Class	Frequency(f)		
5 up to 10	10		
10 up to 15	3		
15 up to 20	4		
20 up to 25	3		
25 up to 30	5		

$$\bar{X} = \frac{\sum fM}{n} = \frac{387.5}{25} = 15.5$$

Standard Deviation of Grouped Data

STANDARD DEVIATION, GROUPED DATA $$s = \sqrt{\frac{\Sigma f(M - \overline{X})^2}{n - 1}} \qquad [3\text{–}13]$$

where:

s is the symbol for the sample standard deviation.
M is the midpoint of the class.
f is the class frequency.
n is the number of observations in the sample.
$\overline{X}$ is the designation for the sample mean.

Standard Deviation of Grouped Data

Steps to find the standard deviation:

- Step 1: Subtract the mean from the class midpoint. That is, find $(M - \bar{X})$.
- Step 2: Square the difference between the class midpoint and the mean.
- Step 3: Multiply the squared difference between the class midpoint and the mean by the class frequency.
- Step 4: Sum the $f(M - \bar{X})^2$.
- Step 5: Divide the sum by the total number of frequencies minus 1
- Step 6: Take the squared root of step 5.

Standard Deviation of Grouped Data

- Example: In an event we asked a sample from the audience about their ages and we construct the following table:

Class	Frequency(f)
5 up to 10	10
10 up to 15	3
15 up to 20	4
20 up to 25	3
25 up to 30	5

Q) Compute the standard deviation of age.

Standard Deviation of Grouped Data

Class	Frequency(f)
5 up to 10	10
10 up to 15	3
15 up to 20	4
20 up to 25	3
25 up to 30	5
Total	25

$$s = \sqrt{\frac{\sum f(M-\bar{X})^2}{n-1}} = \sqrt{\frac{1550}{24}} = 8.04$$