CHAPTER THREE

Data Description

3-1   Measures of Central Tendency

3-2   Measures of Variation

Measures of Central Tendency: Summarize data, using measures of central tendency, such as the mean, median, mode, and midrange.

Measures of Variation: Describe data, using measures of variation, such as the range, variance, and standard deviation.

Measures of Central Tendency	
A **statistic** is a characteristic or measure obtained by using the data values from a <u>sample</u>.	A **parameter** is a characteristic or measure obtained by using all the data values for a specific <u>population</u>.

Mean

- The mean is the quotient of the sum of the values and the total number of values.

- The symbol $\overline{X}$ is used for sample mean.

$$\overline{X} = \frac{X_1 + X_2 + X_3 + \cdots + X_n}{n} = \frac{\sum X}{n}$$

- For a population, the Greek letter μ (mu) is used for the mean.

$$\mu = \frac{X_1 + X_2 + X_3 + \cdots + X_N}{N} = \frac{\sum X}{N}$$

- The mean is the sum of the values, divided by the total number of values.

$\bar{x}$	μ
The symbol for the sample mean:	The symbol for the population mean:
$\overline{X} = \dfrac{\sum X}{n} = \dfrac{X_1 + X_2 + X_3 + \ldots\ldots\ldots + X_n}{n}$	$\mu = \dfrac{\sum X}{N} = \dfrac{X_1 + X_2 + X_3 + \ldots\ldots\ldots + X_n}{N}$
Where n: no. of val. In sample.	Where N: no. of val. In population.

Examples

Days off per Year

The data represent the number of days off per year for a sample of individuals selected from nine different countries. Find the mean.

20, 26, 40, 36, 23, 42, 35, 24, 30

$$\overline{X} = \frac{X_1 + X_2 + X_3 + \cdots + X_n}{n} = \frac{\sum X}{n}$$

$$\overline{X} = \frac{20 + 26 + 40 + 36 + 23 + 42 + 35 + 24 + 30}{9} = \frac{276}{9} = 30.7$$

The mean number of days off is 30.7 years.

Police Incidents

The number of calls that a local police department responded to for a sample of 9 months is shown. Find the mean.

475, 447, 440, 761, 993, 1052, 783, 671, 621

$$\overline{X} = \frac{\Sigma x}{n} = \frac{475 + 447 + 440 + 761 + 993 + 1052 + 783 + 671 + 621}{9}$$

$$= \frac{6243}{9} \approx 693.7$$

Finding the Mean for Grouped Data

Step 1 Make a table as shown

A	B	C	D
Class	Frequency f	Midpoint X_m	$f \cdot X_m$

Step 2 Find the midpoints of each class and place them in column C.

Step 3 multiply the frequency by the midpoint for each class, and place the product in column D.

Step 4 Find the sum of column D.

Step 5 Divide the sum obtained in column D by the sum of frequencies obtained in column B.

The formula for the mean is

$$\overline{X} = \frac{\Sigma f \cdot X_m}{n}$$

Miles Run per Week

Using the frequency distribution for Example below, find the mean. The data represent the number of miles run during one week for a sample of 20 runners.

Solution

The procedure for finding the mean for grouped data is given here.

Step 1 Make a table as shown.

A Class	B Frequency f	C Midpoint X_m	D $f \cdot X_m$
5.5–10.5	1		
10.5–15.5	2		
15.5–20.5	3		
20.5–25.5	5		
25.5–30.5	4		
30.5–35.5	3		
35.5–40.5	2		
	$n = 20$		

Step 2 Find the midpoints of each class and enter them in column C

$$X_m = \frac{5.5 + 10.5}{2} = 8 \qquad \frac{10.5 + 15.5}{2} = 13 \qquad \text{etc.}$$

Step 3 For each class, multiply the frequency by the midpoint, as shown, and place the product in column D.
$1 * 8 = 8 \quad 2 * 13 = 26 \qquad$ etc. The completed table is shown here.

A Class	B Frequency f	C Midpoint X_m	D $f \cdot X_m$
5.5–10.5	1	8	8
10.5–15.5	2	13	26
15.5–20.5	3	18	54
20.5–25.5	5	23	115
25.5–30.5	4	28	112
30.5–35.5	3	33	99
35.5–40.5	2	38	76
	$n = 20$		$\Sigma f \cdot X_m = 490$

Step 4 Find the sum of column D.
Step 5 Divide the sum by n to get the mean.

$$\overline{X} = \frac{\Sigma f \cdot X_m}{n} = \frac{490}{20} = 24.5 \text{ miles}$$

MEDIAN

The **median** is the midpoint of the data array. The symbol for the median is MD.

Finding the median

Step 1 Arrange the data values in ascending order.

Step 2 determine the number of values in the data set.

Step 3 a. If n is odd, select the middle data value as the median.

b. If n is even, find the mean of the two middle values. That is, add them and divide the sum by 2.

Examples

Police Officers Killed

The number of police officers killed in the line of duty over the last 11 years is shown. Find the median.

177 153 122 141 189 155 162 165 149 157 240

Sort in ascending order

122, 141, 149, 153, 155, 157, 162, 165, 177, 189, 240

Select the middle value.

MD = 157

The median is 157 rooms.

Tornadoes in the U.S.

The number of tornadoes that have occurred in the United States over an 8-year period follows. Find the median.

684, 764, 656, 702, 856, 1133, 1132, 1303

Find the average of the two middle values.

656, 684, 702, 764, 856, 1132, 1133, 1303

$$MD = \frac{764 + 856}{2} = \frac{1620}{2} = 810$$

The median number of tornadoes is 810.

The Mode

- The **mode** is the value that occurs most often in a data set.
- It is sometimes said to be the most typical case.
- There may be no mode, one mode (unimodal), two modes (bimodal), or many modes (multimodal).

Example

NFL Signing Bonuses

Find the mode of the signing bonuses of eight NFL players for a specific year. The bonuses in millions of dollars are

18.0, 14.0, 34.5, 10, 11.3, 10, 12.4, 10

You may find it easier to sort first.

10, 10, 10, 11.3, 12.4, 14.0, 18.0, 34.5

Select the value that occurs the most.

> The mode is 10 million dollars.

Licensed Nuclear Reactors

The data show the number of licensed nuclear reactors in the United States for a recent 15-year period. Find the mode.

104 104 104 104 104 107 109 109 109 110
109 111 112 111 109

104 and 109 both occur the most. The data set is said to be bimodal.

> The modes are 104 and 109.

Miles Run per Week

Find the modal class for the frequency distribution of miles that 20 runners ran in one week.

Class	Frequency
5.5 – 10.5	1
10.5 – 15.5	2
15.5 – 20.5	3
20.5 – 25.5	5
25.5 – 30.5	4
30.5 – 35.5	3
35.5 – 40.5	2

The modal class is 20.5 – 25.5.

The mode, the midpoint of the modal class, is 23 miles per week.

Area Boat Registrations

The data show the number of boats registered for six counties in southwestern Pennsylvania. Find the mode.

Westmoreland	11,008
Butler	9,002
Washington	6,843
Beaver	6,367
Fayette	4,208
Armstrong	3,782

Since the category with the highest frequency is Westmoreland, the most typical case is Westmoreland. Hence, the mode is Westmoreland.