PROBABILITY & STATISTICS

BS 1402

Contents

Measures of Central Tendency – The Mean

The **mean** is the sum of the values, divided by the total number of values. The symbol $\overline{X}$ represents the sample mean.

$$\overline{X} = \frac{X_1 + X_2 + X_3 + \cdot \cdot \cdot + X_n}{n} = \frac{\Sigma X}{n}$$

where n represents the total number of values in the sample.

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

$$\mu = \frac{X_1 + X_2 + X_3 + \cdot \cdot \cdot + X_N}{N} = \frac{\Sigma X}{N}$$

where N represents the total number of values in the population.

Example 1:

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

Solution

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

Hence, the mean of the number of days off is 30.7 days.

Example 2:

The data shown represent the number of boat registrations for six counties in southwestern Pennsylvania. Find the mean.

3782 6367 9002 4208 6843 11,008

Example 2 (cont.):

Solution

$$\overline{X} = \frac{\Sigma X}{n} = \frac{3782 + 6367 + 9002 + 4208 + 6843 + 11{,}008}{6} = \frac{41{,}210}{6} = 6868.3$$

The mean for the six county boat registrations is 6868.3.

The mean, in most cases, is not an actual data value.

Rounding Rule for the Mean The mean should be rounded to one more decimal place than occurs in the raw data. For example, if the raw data are given in whole numbers, the mean should be rounded to the nearest tenth. If the data are given in tenths, the mean should be rounded to the nearest hundredth, and so on.

The procedure for finding the mean for grouped data uses the midpoints of the classes. This procedure is shown next.

Example 3:

Find the mean for the miles run per week for the data given below:

Class boundaries	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
	20

Example 3: (cont.)

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.

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

Example 3: (cont.)

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

$1 \cdot 8 = 8 \qquad 2 \cdot 13 = 26 \qquad$ 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.

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

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 the frequencies obtained in column B.

The formula for the mean is

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

[*Note:* The symbols $\Sigma f \cdot X_m$ mean to find the sum of the product of the frequency (f) and the midpoint (X_m) for each class.]

Measures of Central Tendency – The Median

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

Steps in computing the median of a data array

Step 1 Arrange the data in order.

Step 2 Select the middle point.

Example 4:

The number of rooms in the seven hotels in downtown Pittsburgh is 713, 300, 618, 595, 311, 401, and 292. Find the median.

Solution

Step 1 Arrange the data in order.

292, 300, 311, 401, 595, 618, 713

Step 2 Select the middle value.

$$\underset{\substack{\uparrow \\ \text{Median}}}{292, 300, 311, 401, 595, 618, 713}$$

Hence, the median is 401 rooms.

Example 5:

Find the median for the daily vehicle pass charge for five U.S. National Parks. The costs are $25, $15, $15, $20, and $15.

Example 5 (cont.):

Solution

\$15	\$15	\$15	\$20	\$25
		↑		
		Median		

Examples 4 and 5 each had an odd number of values in the data set; hence, the median was an actual data value. When there are an even number of values in the data set, the median will fall between two given values.

Example 6:

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

Solution

$$\begin{array}{c} 656, 684, 702, 764, \; 856, 1132, 1133, 1303 \\ \uparrow \\ \text{Median} \end{array}$$

Since the middle point falls halfway between 764 and 856, find the median MD by adding the two values and dividing by 2.

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

The median number of tornadoes is 810.

Example 7:

The number of cloudy days for the top 10 cloudiest cities is shown. Find the median.

209, 223, 211, 227, 213, 240, 240, 211, 229, 212

Example 7 (cont.):

Solution

Arrange the data in order.

$$\underset{\substack{\uparrow \\ \text{Median}}}{209, 211, 211, 212, 213,\ 223, 227, 229, 240, 240}$$

$$\text{MD} = \frac{213 + 223}{2} = 218$$

Hence, the median is 218 days.

Example 8:

Six customers purchased these numbers of magazines: 1, 7, 3, 2, 3, 4. Find the median.

Solution

$$1, 2, 3, 3, 4, 7 \qquad \text{MD} = \frac{3 + 3}{2} = 3$$

Median

Hence, the median number of magazines purchased is 3.

Measures of Central Tendency – The Mode

The value that occurs most often in a data set is called the **mode.**

A data set that has only one value that occurs with the greatest frequency is said to be **unimodal.**

If a data set has two values that occur with the same greatest frequency, both values are considered to be the mode and the data set is said to be **bimodal.** If a data set has more than two values that occur with the same greatest frequency, each value is used as the mode, and the data set is said to be **multimodal.** When no data value occurs more than once, the data set is said to have *no mode.* A data set can have more than one mode or no mode at all.

Example 9:

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

Solution

It is helpful to arrange the data in order although it is not necessary.

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

Since $10 million occurred 3 times—a frequency larger than any other number—the mode is $10 million.

Example 10:

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

Example 10 (cont.):

Solution

Since the values 104 and 109 both occur 5 times, the modes are 104 and 109. The data set is said to be bimodal.

Example 11:

Find the mode for the number of coal employees per county for 10 selected counties in southwestern Pennsylvania.

110, 731, 1031, 84, 20, 118, 1162, 1977, 103, 752

Solution

Since each value occurs only once, there is no mode.

Note: Do not say that the mode is zero. That would be incorrect, because in some data, such as temperature, zero can be an actual value.

The mode for grouped data is the modal class. The **modal class** is the class with the largest frequency.

Example 12:

Find the modal class for the frequency distribution of miles that 20 runners ran in one week (see Example 3)

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

Solution

The modal class is 20.5–25.5, since it has the largest frequency. Sometimes the midpoint of the class is used rather than the boundaries; hence, the mode could also be given as 23 miles per week.

Example 13:

A survey showed this distribution for the number of students enrolled in each field. Find the mode.

Business	1425
Liberal arts	878
Computer science	632
Education	471
General studies	95

Solution

Since the category with the highest frequency is business, the most typical case is a business major.

Example 14:

A small company consists of the owner, the manager, the salesperson, and two technicians, all of whose annual salaries are listed here. (Assume that this is the entire population.)

Staff	**Salary**
Owner	$50,000
Manager	20,000
Salesperson	12,000
Technician	9,000
Technician	9,000

Find the mean, median, and mode.

Example 14 (cont.):

Solution

$$\mu = \frac{\Sigma X}{N} = \frac{50{,}000 + 20{,}000 + 12{,}000 + 9000 + 9000}{5} = \$20{,}000$$

Hence, the mean is \$20,000, the median is \$12,000, and the mode is \$9,000.

Midrange

The **midrange** is defined as the sum of the lowest and highest values in the data set, divided by 2. The symbol MR is used for the midrange.

$$MR = \frac{\text{lowest value} + \text{highest value}}{2}$$

The *midrange* is a rough estimate of the middle. It is found by adding the lowest and highest values in the data set and dividing by 2. It is a very rough estimate of the average and can be affected by one extremely high or low value.

Example 15:

In the last two winter seasons, the city of Brownsville, Minnesota, reported these numbers of water-line breaks per month. Find the midrange.

2, 3, 6, 8, 4, 1

Solution

$$MR = \frac{1 + 8}{2} = \frac{9}{2} = 4.5$$

Hence, the midrange is 4.5.

If the data set contains one extremely large value or one extremely small value, a higher or lower midrange value will result and may not be a typical description of the middle.

Example 16:

Find the midrange of data for the NFL signing bonuses in Example 3–9. The bonuses in millions of dollars are

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

Solution

The smallest bonus is $10 million and the largest bonus is $34.5 million.

$$\text{MR} = \frac{10 + 34.5}{2} = \frac{44.5}{2} = \$22.25 \text{ million}$$

Notice that this amount is larger than seven of the eight amounts and is not typical of the average of the bonuses. The reason is that there is one very high bonus, namely, $34.5 million.

The Weighted Mean

Find the **weighted mean** of a variable X by multiplying each value by its corresponding weight and dividing the sum of the products by the sum of the weights.

$$\overline{X} = \frac{w_1X_1 + w_2X_2 + \cdots + w_nX_n}{w_1 + w_2 + \cdots + w_n} = \frac{\Sigma wX}{\Sigma w}$$

where $w_1, w_2, \ldots, w_n$ are the weights and $X_1, X_2, \ldots, X_n$ are the values.

Example 17:

Grade Point Average

A student received an A in English Composition I (3 credits), a C in Introduction to Psychology (3 credits), a B in Biology I (4 credits), and a D in Physical Education (2 credits). Assuming A = 4 grade points, B = 3 grade points, C = 2 grade points, D = 1 grade point, and F = 0 grade points, find the student's grade point average.

Solution

Course	Credits (w)	Grade (X)
English Composition I	3	A (4 points)
Introduction to Psychology	3	C (2 points)
Biology I	4	B (3 points)
Physical Education	2	D (1 point)

$$\overline{X} = \frac{\Sigma wX}{\Sigma w} = \frac{3 \cdot 4 + 3 \cdot 2 + 4 \cdot 3 + 2 \cdot 1}{3 + 3 + 4 + 2} = \frac{32}{12} = 2.7$$

The grade point average is 2.7.

Thank You.