

Mean, Variance, and Standard Deviation for Grouped Data

In this section, we will learn how to calculate the mean, variance, and standard deviation for grouped data.

1- Mean for Grouped Data

We learned before the mean is obtained by dividing the sum of all values by the number of values in a data set. However, if the data are given in the form of a frequency table, we no longer know the values of individual observations. Consequently, in such cases, we cannot obtain the sum of individual values. We find an approximation for the sum of these values using the procedure explained in the next paragraph and example. The formulas used to calculate the mean for grouped data follow.

Calculating Mean for Grouped Data

$$\text{Mean for population data: } \mu = \frac{\sum mf}{N}$$

$$\text{Mean for sample data: } \bar{x} = \frac{\sum mf}{n}$$

where m is the midpoint and f is the frequency of a class.

To calculate the mean for grouped data, first find the midpoint of each class and then multiply the midpoints by the frequencies of the corresponding classes. The sum of these products, denoted by $\sum mf$ gives an approximation for the sum of all values. To find the value of

the mean, divide this sum by the total number of observations in the data.

EXAMPLE [1]

Table(1) gives the frequency distribution of the daily commuting times (in minutes) from home to work for *all* 25 employees of a company.

Daily Commuting Time (minutes)	Number of Employees
0 to less than 10	4
10 to less than 20	9
20 to less than 30	6
30 to less than 40	4
40 to less than 50	2

Calculate the mean of the daily commuting times.

Solution: Note that because the data set includes *all* 25 employees of the company, it represents the population. Table(2) shows the calculation of $\sum mf$.

Note that in Table (2), m denotes the midpoints of the classes.

Daily Commuting Time (minutes)	f	m	mf
0 to less than 10	4	5	20
10 to less than 20	9	15	135
20 to less than 30	6	25	150
30 to less than 40	4	35	140
40 to less than 50	2	45	90
	$N = 25$		$\sum mf = 535$

To calculate the mean, we first find the midpoint of each class. The class midpoints are recorded in the third column of Table 3.9. The products of the midpoints and the corresponding frequencies are listed in the fourth column. The sum of the fourth column values, denoted by $\sum mf$ gives the approximate total daily commuting time (in minutes) for all 25 employees. The mean is obtained by dividing this sum by the total frequency. Therefore,

$$\mu = \frac{\sum mf}{N} = \frac{535}{25} = \mathbf{21.40 \text{ minutes}}$$

Thus, the employees of this company spend an average of 21.40 minutes a day commuting from home to work.

EXAMPLE[2]

Table (3) below gives the frequency distribution of the number of orders received each day during the past 50 days at the office of a mail-order company

Number of Orders	Number of Days
10–12	4
13–15	12
16–18	20
19–21	14

Calculate the mean.

Solution Because the data set includes only 50 days, it represents a sample. The value of $\sum mf$ is calculated in Table (4).

Number of Orders	f	m	mf
10–12	4	11	44
13–15	12	14	168
16–18	20	17	340
19–21	14	20	280
	$n = 50$		$\sum mf = 832$

The value of the sample mean is

$$\bar{x} = \frac{\sum mf}{n} = \frac{832}{50} = 16.64 \text{ orders}$$

2- Variance and Standard Deviation for Grouped Data

Following are what we will call the *basic formulas* used to calculate the population and sample variances for grouped data:

Short-Cut Formulas for the Variance and Standard Deviation for Grouped Data

$$\sigma^2 = \frac{\sum m^2f - \frac{(\sum mf)^2}{N}}{N} \quad \text{and} \quad s^2 = \frac{\sum m^2f - \frac{(\sum mf)^2}{n}}{n - 1}$$

where σ^2 is the population variance, s^2 is the sample variance, and m is the midpoint of a class. The standard deviation is obtained by taking the positive square root of the variance.

$$\text{Population standard deviation: } \sigma = \sqrt{\sigma^2}$$

$$\text{Sample standard deviation: } s = \sqrt{s^2}$$

The following examples [3] and [4] illustrate the use of these formulas to calculate the variance and standard deviation.

EXAMPLE [3]

The following data, reproduced from Table 3.8 of Example 3–14, give the frequency distribution of the daily commuting times (in minutes) from home to work for all 25 employees of a company.

Daily Commuting Time (minutes)	Number of Employees
0 to less than 10	4
10 to less than 20	9
20 to less than 30	6
30 to less than 40	4
40 to less than 50	2

Calculate the variance and standard deviation.

Solution: All four steps needed to calculate the variance and standard deviation for grouped data are shown after the following table.

Daily Commuting Time (minutes)	f	m	mf	m^2f
0 to less than 10	4	5	20	100
10 to less than 20	9	15	135	2025
20 to less than 30	6	25	150	3750
30 to less than 40	4	35	140	4900
40 to less than 50	2	45	90	4050
	$N = 25$		$\Sigma mf = 535$	$\Sigma m^2f = 14,825$

Step 1. Calculate the value of $\sum mf$

To calculate the value of first find the midpoint m of each class (see the third column in the above table) and then multiply the corresponding class midpoints and class frequencies (see the fourth column). The value of $\sum mf$ is obtained by adding these products. Thus, $\sum mf = 535$

Step 2. Find the value of $\sum m^2f$

To find the value of $\sum m^2f$ square each m value and multiply this squared value of m by the corresponding frequency (see the fifth column in Table 3.12). The sum of these products (that is, the sum of the fifth column) gives $\sum m^2f$ Hence,

$$\sum m^2f = 14.825.$$

Step 3. Calculate the variance.

Because the data set includes all 25 employees of the company, it represents the population.

Therefore, we use the formula for the population variance:

$$\sigma^2 = \frac{\sum m^2f - \frac{(\sum mf)^2}{N}}{N} = \frac{14,825 - \frac{(535)^2}{25}}{25} = \frac{3376}{25} = 135.04$$

Step 4. *Calculate the standard deviation.*

To obtain the standard deviation, take the (positive) square root of the variance.

$$\sigma = \sqrt{\sigma^2} = \sqrt{135.04} = 11.62 \text{ minutes}$$

Thus, the standard deviation of the daily commuting times for these employees is 11.62 minutes.

Note that the values of the variance and standard deviation calculated in Example 3–16 for grouped data are approximations. The exact values of the variance and standard deviation can be obtained only by using the ungrouped data on the daily commuting times of the 25 employees.