

## 2.8: Measures of Median and Mean on Grouped Data

### Learning Objectives

- Determine the median and mean in grouped discrete data
- Determine the median and mean in grouped relative frequency data
- Determine the median and mean in grouped continuous data
- Extend to weighted mean

▢ [Section 2.8 Excel File](#) (contains all of the data sets for this section)

### Introduction to Grouped Data

In our investigation of descriptive statistics, we worked on a collection of individual data values and then formed appropriate summary measures of that "raw" data. However, we may sometimes be given the data in a summarized frequency distribution format instead of as a "raw" data collection. Can we find our various descriptive statistic measures if we only have the frequency table, which represents the data in grouped form? Although the answer is only "sometimes," the underlying concept of how we can do so is essential for later ideas in the course. We begin with mean and median measures found from frequency tables on grouped quantitative data, but where the grouping was not formed by interval values but only by the same single values.

### Mean and Median of Non-Interval but Grouped Data in a Frequency Table

Look at the frequency distribution in Table 2.8.1 shown below from Section 2.1 about thirty student scores (discrete 10-point scale) for an assignment; we will assume these are only a sample of a larger population of scores. Notice that our table shows no loss of crucial information on the data as each distinct data value is explicitly shown in the table, and no intervals are used to represent the grouped student scores.

Table 2.8.1: Grouped Frequency Distribution

Student Score	Frequency
3	1
4	1
5	3
6	5
7	5
8	7
9	5
10	3

With such a table, we could formally recreate the entire data set  $\{3,4,5,5,5,\dots,9,10,10,10\}$  by recognizing the meaning of the frequency values for each of the various score values in the table. If we had more data values, recreating the data set would be tedious, and we could lose information on the data. Even if using technology to produce our descriptive measures, we must "type in" all individual data values. This process is likely to lead to many data entry errors. Recreating the "raw" data set is unnecessary; we can determine the median and the arithmetic mean by working with data in this frequency distribution format.

First, we examine the median measure by using quantitative reasoning. We note by the sum of the frequency column that there are 30 pieces of data and, by our earlier discussion in Section 2.4, the median is the average of the two data values in position 15 and 16 in the ordered list of all data values. Using our frequency column, we accumulate across our frequency counts to see that the 15<sup>th</sup> data position is within the group of "7" scores and the 16<sup>th</sup> data position is within the group of "8" scores (total accumulation of number of data scores from "3" to "7" includes  $1 + 1 + 3 + 5 + 5 = 15$  scores). So the median value of the data is  $\frac{7+8}{2} = 7.5$ . In general, we can find the median by focusing on our frequency counts to help us determine the center position location, using the location value to determine the median within the grouped variable values.

Next, we examine the mean. Recall that the mean is found in our original discussion by summing our quantitative data values, then dividing by the number of values in the data set:  $\bar{x} = \frac{\sum x_i}{n}$ . Notice in our grouped data, we can find the portion of the entire sum generated by each group by multiplying the group data value by the frequency. For example, the grouped data value 5 will contribute a total of  $3 \cdot 5 = 15$  to the total sum since there are three 5 values in the data set; similarly, the grouped data value 8 will contribute a total of  $8 \cdot 7 = 56$  toward to total sum since there are seven 8 values in the data set. This leads us to the following adjustment of our table to compute the mean of the grouped data (also note the change to general headings on each column).

Table 2.8.2: Computation of arithmetic mean for data from Table 2.8.1

$x_j$	$f_j$	$x_j \cdot f_j$
3	1	3
4	1	4
5	3	15
6	5	30
7	5	35
8	7	56
9	5	45
10	3	30
Totals:	$\sum f_j = 30$	$\sum (x_j \cdot f_j) = 218$
Arithmetic Mean: $\bar{x} = \frac{\sum (x_j \cdot f_j)}{\sum f_j} = \frac{218}{30} \approx 7.2667$		

In conclusion, by summing our  $f_j$  frequency column values, we know the sample size  $n$ . We have accomplished the exact computation by adding our thirty individual data values together by summing our  $x_j \cdot f_j$  column of values. The arithmetic mean of the data is found by our last computation  $\bar{x} = \frac{\sum (x_j \cdot f_j)}{\sum f_j}$ . We divided the total sum of all data values by the number of data values. In grouped data of this form, we can find the mean by the above process, and described symbolically by the given formula:

#### Sample Mean from a Frequency Distribution

$$\bar{x} = \frac{\sum x_i}{n} = \frac{\sum (x_j \cdot f_j)}{\sum f_j}$$

If the data in our table had been population data, we would still perform the same calculation using the same reasoning. And have:

#### Population Mean from a Frequency Distribution

$$\mu = \frac{\sum x_i}{N} = \frac{\sum (x_j \cdot f_j)}{\sum f_j}$$

#### ? Text Exercise 2.8.1

Consider the Quiz 1 data from section 2.6 below in the frequency table format. Determine the mean from the grouped format and compare it with the results obtained in section 2.6 from the "raw" data. We assume the data is population data in this example.

Table 2.8.3: Grouped Frequency Distribution of Quiz 1 Data

Quiz Scores	Frequency
5	2
6	6

Quiz Scores	Frequency
7	5
8	4
9	3

### Answer

To find the mean from this frequency table, we follow these steps with thoughts of what we are doing with the data:

1. Sum the frequency column  $f_j$  to determine the size of the data set.
2. Compute the column of values  $x_j \cdot f_j$  to weight each quiz score with their occurrence frequency.
3. Sum the column of  $x_j \cdot f_j$  values to produce the total sum as if summing the individual data values.
4. Produce the mean by dividing the sum of the  $x_j \cdot f_j$  column by the sum of the frequency column  $f_j$ .

Table 2.8.4 Computation of arithmetic mean

$x_j$	$f_j$	$x_j \cdot f_j$
5	2	10
6	6	36
7	5	35
8	4	32
9	3	27
		$\mu = \frac{\sum x_j \cdot f_j}{\sum f_j} = \frac{140}{20} = 7$

We notice this is the exact arithmetic mean value computed when working with the twenty individual quiz scores.

## Mean and Median of Grouped Data in a Relative Frequency Table

What would happen if we had a relative frequency distribution of this data instead of a frequency distribution table? Recall that relative frequency in this situation measures the proportion of the data set that has a specific data value. We will be using  $P(x_j)$  to represent the relative frequency or proportion measure as tied to specific data value  $x_j$ .

Table 2.8.5: Relative frequency table of student scores from Table 2.8.1

Student Score $x_j$	Relative Frequency $P(x_j)$
3	$\frac{1}{30} \approx 0.0333 = 3.33\%$
4	$\frac{1}{30} \approx 0.0333 = 3.33\%$
5	$\frac{3}{30} = 0.1000 = 10.00\%$
6	$0.1667 = 16.67\%$
7	$0.1667 = 16.67\%$
8	$0.2333 = 23.33\%$
9	$0.1667 = 16.67\%$
10	$0.1000 = 10.00\%$
Totals:	$\sum P(x_j) = 1.0000 = 100\%$

With a relative frequency table, we could not formally recreate the entire data set unless we first knew the number of data values in the data set (i.e., the sample or population size). However, we do not need such information to determine the distribution's median or mean. We proceed as above, working with relative frequency measures instead of counted frequency measures.

We first examine the median measure. All data is accounted for by the sum of the frequency column that 100%; we should always sum our relative frequency measures to see if we have a total of  $1.0000 = 100\%$ . As discussed previously, the median is at the 50<sup>th</sup> percentile position in the ordered list of our data set. Using our relative frequency column, we can accumulate our relative percentages to see that the 50% data position is right on the border between the group of "7" scores and the group of "8" scores; total relative frequency accumulation of number of data scores from "3" to "7" includes  $3.33\% + 3.33\% + 10.00\% + 16.67\% + 16.67\% = 50\%$ . The median value of the data is  $\frac{7+8}{2} = 7.5$ , just as above. We can find the median by focusing on our relative frequency measures to help us determine the location of 50% of the data set from the smallest value and the 50% location value to determine the median within the grouped variable values.

Next, we examine the mean measure. In forming the relative frequency measures, we divided each frequency count by the sample size to form the relative frequency measures. This earlier division and some algebraic reasoning show how we can adjust our standard arithmetic mean formula to fit this situation.

$$\begin{aligned}\bar{x} &= \frac{\sum x_i}{n} = \frac{x_1 + x_2 + \dots + x_n}{n} \\ &= \frac{x_1}{n} + \frac{x_2}{n} + \dots + \frac{x_n}{n} = \sum \frac{x_j \cdot f_j}{n} \\ &= \sum \left( x_j \cdot \frac{f_j}{n} \right) = \sum [x_j \cdot P(x_j)]\end{aligned}$$

$P(x_j)$  stands for the relative frequency of data value  $x_j$ ; it is the proportion of the data set with that specific data value,  $x_j$ . In a sense, each distinct data value is being "weighted" by the relative frequency of occurrence. For example, the fact that the data value 8 occurs with 23.33% relative frequency should make this data value "weigh-in" more heavily to the average than does the data value 5 that only occurs with 10% relative frequency. Our relative frequency gives us this "weighting" of the data in a relative sense instead of the above, in which the actual frequency measures give us a weighted "count" sense. This leads us to the following adjustment of our table to compute the mean of the grouped data (note the change to general headings on each column).

Table 2.8.6: Computation of arithmetic from relative frequency found in Table 2.8.5

$x_j$	$P(x_j)$	$x_j \cdot P(x_j)$
3	$\frac{1}{30} \approx 0.0333 = 3.33\%$	$3 \cdot \frac{1}{30} = 0.1000$
4	$\frac{1}{30} \approx 0.0333 = 3.33\%$	$4 \cdot \frac{1}{30} \approx 0.1333$
5	$\frac{3}{30} = 0.1000 = 10.00\%$	$5 \cdot \frac{3}{30} = 0.5000$
6	$0.1667 = 16.67\%$	$1.0000$
7	$0.1667 = 16.67\%$	$1.1667$
8	$0.2333 = 23.33\%$	$1.8667$
9	$0.1667 = 16.67\%$	$1.5000$
10	$0.1000 = 10.00\%$	$1.0000$
Totals:	$\sum P(x_j) = 1.0000 = 100\%$	$\sum [x_j \cdot P(x_j)] \approx 7.2667$

We notice that the results from this relative frequency distribution are the same as those from the previous section of the plain frequency distribution.

In conclusion, by multiplying each unique data value  $x_j$  by its relative frequency measure  $P(x_j)$ , we have used a relative weighting of each value to produce the arithmetic mean; so, computationally, we need only sum these products  $x_j \cdot P(x_j)$  to

produce our arithmetic mean. In grouped data of this relative frequency form, we can find the mean by the above process, as described symbolically by the given formula:

### Sample Mean from a Relative Frequency Distribution

$$\bar{x} = \sum [x_j \cdot P(x_j)]$$

Once again, if the data in our table had been population data, then we would still perform the same calculation work using the same reasoning:

### Population Mean from a Relative Frequency Distribution

$$\mu = \sum [x_j \cdot P(x_j)]$$

#### ? Text Exercise 2.8.2

Consider the Quiz 1 data from section 2.6, this time given in the relative frequency table format below. Determine the mean from the grouped format and compare it with the earlier results.

Table 2.8.7: Grouped Relative Frequency Distribution of Quiz 1 Data

Quiz Scores	Relative Frequency
5	10%
6	30%
7	25%
8	20%
9	15%

#### Answer

To find the mean from this relative frequency table, we follow these steps as established in the discussion above:

1. Sum the relative frequency column  $P(x_j)$  to check that 100% of the data is accounted for in the table.
2. Compute the column of values  $x_j \cdot P(x_j)$  to weight each of the various quiz scores with their relative frequency of occurrence.
3. Sum the column of  $x_j \cdot P(x_j)$  values to produce mean of the data values.

Table 2.8.8 Computation of arithmetic mean using relative frequencies from Table 2.8.7

$x_j$	$P(x_j)$	$x_j \cdot P(x_j)$
5	10%	0.5
6	30%	1.8
7	25%	1.75
8	20%	1.60
9	15	1.35
Totals:	100%	$\mu = 7.00$

Again, this is the exact arithmetic mean value computed previously when working with the raw data set or the grouped frequency table set.

We extend these ideas one more step with the concept of "weighted" averages.

## Weighted Mean Measures

Sometimes, data values are assigned different weights; for example, course averages are often determined through a "weighting" of the various assessment values. This weighting is usually given as a percentage but can be shown in any chosen relative form (such as a "2" weight for those values that carry twice the weight of any values assigned a "1" weight). As such, we can see how the weights play the same role as the frequency or relative frequency values in the above discussion.

As an example, suppose a school, as is commonly done, uses a four-point scale (A = 4 points, B = 3 points, C = 2 points, D = 1 point, and U = 0 points) to determine grade point average (GPA) weighted by the number of credit hours for the class. A randomly chosen student's recent letter grades awarded and number of credits in eight courses were as follows: A with 3 credits, U with 2 credits, C with 4 credits, A with 5 credits, B with 3 credits, B with 3 credits, C with 5 credits, and D with 3 credits. We organize this information in table 2.8.3 to determine this student's GPA.

Table 2.8.9: Grouped Frequency Distribution

Letter Grade	Point Value	Credit Hours (Weight)
A	4	8
B	3	6
C	2	9
D	1	3
U	0	2

Again, we use the above ideas to compute the GPA, a weighted mean.

Table 2.8.10: Computation of GPA as a weighted mean

Letter Grade	Point Value ( $x_j$ )	Credit Hours ( $w_j$ )	$x_j \cdot w_j$
A	4	8	32
B	3	6	18
C	2	9	18
D	1	3	3
U	0	2	0
Totals:		$\sum w_j = 28$	$\sum (x_j \cdot w_j) = 71$
		Weighted Mean:	$\frac{\sum (x_j \cdot w_j)}{\sum w_j} \approx 2.5357$

This student had a GPA of 2.5357 for those courses. In data that carries varied weights, we can determine the mean as described symbolically.

### Sample Mean from Weighted Data

$$\frac{\sum (x_j \cdot w_j)}{\sum w_j}$$

As we have seen, we do not always need "raw" data, especially with huge data sets, to formulate many of our descriptive statistics for the data. Grouped data conserves space required to represent data and can often be used to produce many summary statistic measures with minor adjustments to our computational thinking. However, we must know if we have any "loss" in the data representation due to the grouping. All the above data sets were discrete, and each grouping was done on single values, not over interval values. When we group data over interval values, we lose some information in the data. The following optional subsection examines this issue of continuous data.

**Note: Level of measurement and careful consideration of results**

An astute reader will notice that the four-point grading scale, common in many academic institutions, takes values on an ordinal scale; the arithmetic differences in values do not provide any information other than the underlying ordering of letter grades. If one student earns a 99% while a second student earns 90.1%, both students would be awarded the same letter grade of an A, despite having achieved different levels of performance in the course.

When we look at a semester's average GPA, as we did above, how are we to interpret two students in the same courses having the same average? Just like with the racing example at the end of [section 1.6](#), we cannot say that they performed (earned points), on average, the same. One student could have outperformed the other student on all assessments in each class, yet still be awarded the same letter grades in each class thus earning an equivalent GPA. All we can say is that the students earned, on average, the same letter grades.

**Text Exercise 2.8.3**

Consider two physics majors, Aaron and Elise, who took Engineering Physics I (five credit hours), Calculus I (five credit hours), and Elements of Statistics (three credit hours) last semester. Aaron earned 85%, 96%, and 98%, respectively, and Elise earned 90%, 91%, and 98%, respectively.

1. Convert each student's semester grades to the four-point grading scale and then compute the weighted average using the number of credit hours as the weight. This is the standard way four-point scale averages are computed.

**Answer**

Aaron would receive a 3 for his physics course, and then 4 for each of his math courses. Since physics and calculus were five credit hour courses, those two grades will be weighted by 5, and statistics will be weighted by 3. We thus have the following computation.

$$\text{GPA}_{\text{Aaron}} = \frac{3 \cdot 5 + 4 \cdot 5 + 4 \cdot 3}{5 + 5 + 3} = \frac{15 + 20 + 12}{13} = \frac{47}{13} \approx 3.6154$$

Elise earned an A in each course thus earning 4 in each of her courses. Since physics and calculus were five credit hour courses, those two grades will be weighted by 5, and statistics will be weighted by 3. We thus have the following computation.

$$\text{GPA}_{\text{Elise}} = \frac{4 \cdot 5 + 4 \cdot 5 + 4 \cdot 3}{5 + 5 + 3} = \frac{20 + 20 + 12}{13} = \frac{52}{13} = 4$$

We thus have that Aaron earned a 3.6154 and Elise earned a 4.0 last semester.

2. Compute each student's weighted average percentage using the number of credit hours as the weight and then convert the averages to the four-point scale. This is a nonstandard way to compute four-point scale averages.

**Answer**

We compute the weighted averages similarly.

$$\text{GPA}_{\text{Aaron}} = \frac{85 \cdot 5 + 96 \cdot 5 + 98 \cdot 3}{5 + 5 + 3} = \frac{425 + 480 + 294}{13} = \frac{1199}{13} \approx 92.2308\%$$

$$\text{GPA}_{\text{Elise}} = \frac{90 \cdot 5 + 91 \cdot 5 + 98 \cdot 3}{5 + 5 + 3} = \frac{450 + 455 + 294}{13} = \frac{1199}{13} \approx 92.2308\%$$

In converting the two weighted averages to the four-point scale, both Aaron and Elise would receive a 4.0 for the semester. Despite having the same average percentages, the standard way of computation distinguishes between a 4.0 student, Elise, and Aaron, a student who did not get straight A's. There is only one way to get a 4.0. There are many ways to get a lower GPA. The four-point scale emphasizes the distinction between straight A students and everyone else.

2.8: Measures of Median and Mean on Grouped Data is shared under a [Public Domain](#) license and was authored, remixed, and/or curated by The Math Department at Fort Hays State University.

- **3.3: Measures of Central Tendency** by [David Lane](#) is licensed [Public Domain](#). Original source: <https://onlinestatbook.com>.
- **1.10: Distributions** by [David Lane](#) is licensed [Public Domain](#). Original source: <https://onlinestatbook.com>.