

Section 2.3, Measures of Central Tendency

In this section, we will learn several ways to describe the “center” of a quantitative data set.

1 Mean, Median, and Mode

The **mean** is the sum of the data entries divided by the number of entries. With data from all N members of the population, we can calculate the population mean μ :

$$\mu = \frac{\sum x}{N},$$

where $\sum x$ represents the sum of all the values taken for the variable, and μ is the Greek letter pronounced “mu.”

With data from only a sample of size n , we can calculate the sample mean \bar{x} :

$$\bar{x} = \frac{\sum x}{n}.$$

Normally we won’t have data from the entire population, so we will be calculating \bar{x} .

The **median** is the value that lies in the middle of an ordered data set. If the data set has an odd number of entries, the median is the data entry in the middle. If the data set has an even number of entries, the median is the mean of the two middle data entries.

The **mode** of a data set is the entry that occurs most often. If no entry is repeated, we say that there is no mode. It is possible for there to be more than one mode. If there are exactly two modes, we say that the data set is **bimodal**.

Example

Find the mean, median, and mode of the following data set:

5 10 7 19 25 12 15 7 6 8
17 17 22 21 7 7 24 5 6 5

There are 20 numbers, so $n = 20$. Then,

$$\begin{aligned}\bar{x} &= \frac{5 + 10 + 7 + 19 + 25 + 12 + 15 + 7 + 6 + 8 + 17 + 17 + 22 + 21 + 7 + 7 + 24 + 5 + 6 + 5}{20} \\ &= \frac{245}{20} = 12.25\end{aligned}$$

In order to find the median, we need to sort the data, so:

5 5 5 6 6 7 7 7 7 8
10 12 15 17 17 19 21 22 24 25

To get the median, we need to average the 10th and 11th data values, so the median is $(8+10)/2 = 9$. Also, with the data ordered, it is easy to see that 7 is the most common value, so the mode is 7.

An **outlier** is a data entry that is far from the other entries in the data set. Outliers greatly affect the mean, but have little impact on the median and mode. For example, a midterm exam had scores of 9, 31, 47, ..., 93, 108. The overall mean was a 64.72. When the 9 was removed, the mean increased to 68, but the median only increased by 0.5.

When calculating a mean, sometimes different values are more influential. To find the **weighted mean**, we calculate:

$$\bar{x} = \frac{\sum(x \cdot w)}{\sum w},$$

where w is the weight of each value of x .

Example

At the end of the semester, your highest 4 quiz scores, which are worth 5% each, were 76%, 82%, 95%, and 92%. Your highest 2 midterm exam scores, which are worth 20% each, were 78% and 82%. Your final exam score, worth 40%, was an 80%. What is your overall score for this course?

$$\bar{x} = \frac{76 \cdot 0.05 + 82 \cdot 0.05 + 95 \cdot 0.05 + 92 \cdot 0.05 + 78 \cdot 0.2 + 82 \cdot 0.2 + 80 \cdot 0.4}{0.05 + 0.05 + 0.05 + 0.05 + 0.2 + 0.2 + 0.4} = 81.25$$

Using the weighted mean formula, we can see that the mean of a frequency distribution is approximately:

$$\bar{x} \approx \frac{\sum(x \cdot f)}{n},$$

where x and f are the midpoints and frequencies of each class, respectively. Recall that $\sum f = n$. (Note: If the original data is available, you should use the original data to get the exact sample mean.)

Example

Approximate the mean of the following frequency distribution:

Class	Frequency
5–9	10
10–14	2
15–19	4
20–24	3
25–29	1

$$\bar{x} \approx \frac{7 \cdot 10 + 12 \cdot 2 + 17 \cdot 4 + 22 \cdot 3 + 27 \cdot 1}{20} = \frac{255}{20} = 12.75$$

2 Shape of the distribution

A distribution is called **symmetric** if the histogram is symmetric. In this case, the mean (approximately) equals the median.

A distribution is called **uniform** if it is symmetric and the heights of all the bars in the histogram are the same (it will make the shape of a rectangle).

A distribution is **skewed left** if it has a “tail” that extends to the left. The mean will be to the left of the median.

A distribution is **skewed right** if it has a “tail” that extends to the right. In this case, the mean is to the right of the median.