3.1: Measures of Center

Now, we will begin studying some numerical measures that describe data sets. There are two basic types:

- Measures of central tendency (this section)
- Measures of dispersion (next section)

Summation Notation:

Summation notation is a compact way to write "add up n numbers" of "do something to n numbers first, and then add them up." The numbers are represented as $x_1, x_2, \ldots, x_n$ "

$$\sum_{i=1}^{n} x_i = x_1 + x_2 + \ldots + x_n$$

$$\sum_{i=1}^{n} x_i = x_1 + x_2 + x_3 + \ldots + x_n$$

Σ: Greek letter
Capital Sigma
Stands for Sum

Example 1: Consider the numbers 8, 2, 6, 10, 4, 9. Find $\sum_{i=1}^{6} x_i$ and $\sum_{i=1}^{6} x_i^2$.

$$\sum_{i=1}^{6} x_i = 8+2+6+10+4+9 = \boxed{39}$$

$$\sum_{i=1}^{n} x_i^2 = 8^2+2^2+6^2+10^2+4^2+9^2$$

$$= 64+4+36+100+16+81$$

The Mean: Ungrouped Data:

The *mean* of a set of quantitative data is equal to the sum of all the measurements in the data set divided by the total number of measurements in the set.

If $x_1, x_2, \ldots, x_n$ is a set of n measurements, then the *mean*, or *average*, is given by

$$[\text{mean}] = \frac{\sum_{i=1}^{n} x_i}{n} = \frac{x_1 + x_2 + \ldots + x_n}{n} \quad \text{where}$$

$\bar{x} =$ [mean] if data set is a sample

$\mu =$ [mean] if data set is the population

use $\bar{x}$ ("x-bar") if data are a sample from a larger population

use μ ("mu") if data are the whole population under consideration

Range: max - min = 99 - 7 = 92

The median:

Sometimes the mean can be misleading for a data set. Suppose that a math class had 7 students with test scores (out of a possible 100) of 88, 99, 7, 78, 89, 94, and 75.

mean: $\mu = \frac{88+99+7+78+89+94+75}{7} = \frac{530}{7} \approx 75.7$

All but 1 person scored at least 75.

The low score 7 is pulling the mean down

The *median* is unaffected by extreme values (outliers). Essentially it is the "middle" of the data set.

To find the median, you'll need to sort the data in numerical order.

The Median (Ungrouped Data):
• If the number of measurements is odd, the median is the middle measurement when the measurements are arranged in descending or ascending order.
• If the number of measurements is even, the median is the mean of the two middle measurements when the measurements are arranged in descending or ascending order.

Example 2: Find the median of the test scores 88, 99, 7, 78, 89, 94, and 75.

7, 75, 78, 88, 89, 94, 99

$\boxed{\text{Median} = 88}$

Example 3: Find the median of the test scores 88, 85, 99, 7, 78, 89, 94, and 75.

7, 75, 78, 85, 88, 89, 94, 99

median $= \frac{85+88}{2} = \boxed{86.5}$ (halfway between 85 and 88)

Example 4: Provide some everyday examples in which the median is more useful than the mean.

property values

salary data

The mode:

> <u>The Mode</u>:
>
> The *mode* is the most frequently occurring value in a data set, provided it occurs at least twice. There may be a unique mode, several modes, or no mode.
>
> A data set with two modes is called *bimodal*.

<u>Example 5:</u> Find the median and mode for the following data sets.

a. {4, 5, 5, 5, 5, 6, 7, 8, 12}

mode: 5

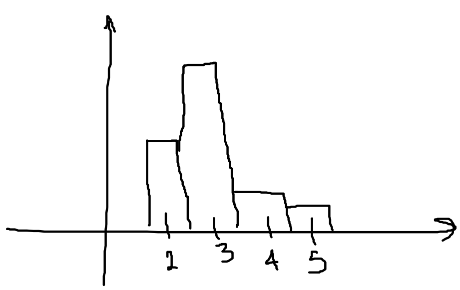
b. {1, 2, 3, 3, 3, 5, 7, 7, 7, 23}

Modes: 3 and 7
This data is bimodal data

c. {1, 3, 5, 6, 7, 9, 11, 15}

No mode.

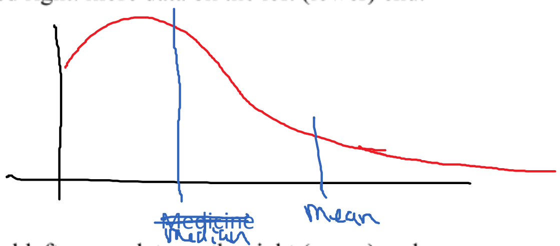
<u>Example 6:</u>



Mode: 3

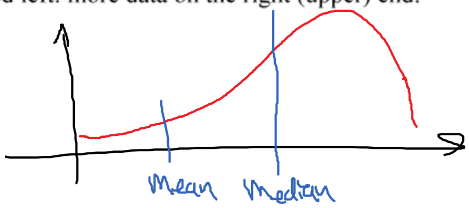
Mean, median, and mode for distributions of different shapes:

Skewed right: more data on the left (lower) end.



Median: half the area to the left; half the area to the right

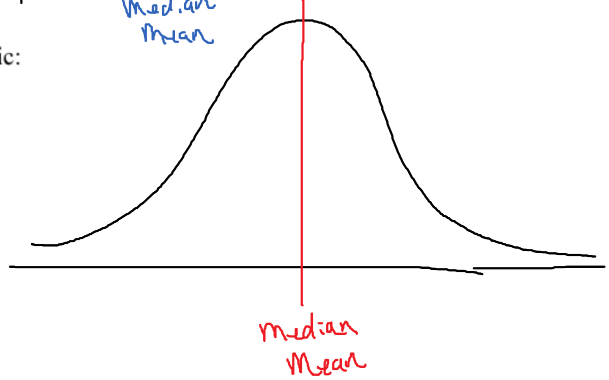
Skewed left: more data on the right (upper) end.

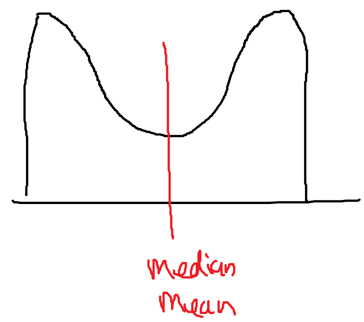


Uniform:

This one is symmetric

Symmetric:





Approximating the mean for grouped data:

Sometimes we only have access to a frequency distribution of the data, and not to the raw data. (This is often called grouped data.) For grouped data, we can approximate the mean, but cannot calculate the mean exactly. Smaller class widths (intervals) generally result in a better approximation.

The Mean: Grouped Data:
A data set of n measurements is grouped into k classes in a frequency table. If x_i is the midpoint of the ith class interval and f_i is the ith class frequency, then the *mean* of the grouped data is

$$[\text{mean}] = \frac{\sum_{i=1}^{k} x_i f_i}{n} = \frac{x_1 f_1 + x_2 f_2 + \ldots + x_k f_k}{n} \quad \text{where}$$

$\bar{x} = [\text{mean}]$ if data set is a sample
$\mu = [\text{mean}]$ if data set is the population

Example 7: Estimate ~~Find~~ the mean for the grouped data.

Height (in inches)	Frequency	midpoints
63.0–65.9	3	64.5
66.0–68.9	6	67.5
69.0–71.9	7	70.5
72.0–74.9	4	73.5
75.0–77.9	3	76.5

$n = 23$

To find the midpoint of a class, subtract the bottom boundary from the bottom boundary of the next class, then divide by 2
(Or, average the lower bound of the current class with the lower bound of next class:
For 63.0 – 65.9: $\frac{63.0 + 66.0}{2} = 64.5$

$66.0 - 63.0 = 3$ units wide
OR
$0.5(3) =$ half the width is 1.5
$63.0 + 1.5 = 64.5$

Assume it's a sample:

$$\bar{x} = \frac{\underbrace{64.5 + 64.5 + 64.5}_{3} + \underbrace{67.5 + \ldots + 67.5}_{6} + \ldots}{23}$$

$$\bar{x} = \frac{64.5(3) + 67.5(6) + 70.5(7) + 73.5(4) + 76.5(3)}{23}$$

$$= \frac{1615.5}{23} \approx \boxed{70.24}$$

Example 8: Find the mean for the grouped data.

Score	Frequency
60–69	12
70–79	7
80–89	14
90–99	10

The weighted mean:

For a weighted mean, or weighted average, each data point is assigned a *weight*. Data points with higher weights are counted more heavily that data points that are weighted less.

Each data point is multiplied by its weight before being totaled. The total is then divided by the sum of all the weights.

The weighted mean:

If $x_1, x_2, ..., x_n$ are data values and $w_1, w_2, ..., w_n$ are their corresponding weights, then their weighted average is

$$\mu = \frac{w_1x_1 + w_2x_2 + \ldots + w_nx_n}{w_1 + w_2 + \ldots + w_n}.$$

When calculating the *mean* of a set of scores, each score carries the same weight. To calculate the mean, we add all the scores, and divide by the number of scores. Therefore, no score carries more weight/importance than any other score.

For a weighted mean, or weighted average, each data point is assigned a *weight*. Data points with higher weights are counted more heavily that data points that are weighted less.

Each data point is multiplied by its weight before being totaled. The total is then divided by the sum of all the weights.

The Weighted Mean:
Suppose a data set has n measurements, and that each measurement x_i is associated with a weight w_i. Then the *weighted mean* of the set of measurements is

$$[\text{mean}] = \frac{\sum_{i=1}^{n} x_i w_i}{\sum_{i=1}^{n} w_i} = \frac{x_1 w_1 + x_2 w_2 + \ldots + x_n w_n}{w_1 + w_2 + \ldots + w_n} \qquad \text{where}$$

$\bar{x} = [\text{mean}]$ if data set is a sample
$\mu = [\text{mean}]$ if data set is the population

Example 1: Suppose the components of a college class are weighted as follows: Homework (15%), Quizzes (15%), Computer Lab (10%), Midterm Exam (30%), and Final Exam (30%). Suppose a student's grades for homework, quizzes, lab, midterm, and final were 92, 80, 100, 75, and 76, respectively. Calculate the student's course average.

$$\text{Course Average} = \frac{0.15(92) + 0.15(80) + 0.10(100) + 0.30(75) + 0.30(76)}{1}$$

Note: $\Sigma w_i = 0.15 + 0.15 + 0.10 + 0.30 + 0.30 = 1$

$$= \boxed{81.1}$$

Example 2: Using the weights in the above example, and using the same homework, quizzes, lab, and midterm grades (92, 80, 100, 75), what grade must the student receive on the final exam to earn an 80% course average and thus a ~~B~~ A in the course?
90%

Let x = final exam grade. We will solve for x.

$$0.15(92) + 0.15(80) + 0.10(100) + 0.30(75) + 0.30x = 90$$

$$58.3 + 0.30x = 90$$
$$-58.3 \qquad\qquad -58.3$$

$$0.30x = 31.7$$

$$x = \frac{31.7}{0.30} = 105.7$$

They need 105.7 on the final! (Need some bonus problems)

Example 3: Suppose a student takes ten courses during the academic year, earning the following grades. Calculate the student's grade point average (GPA).

Calculus I (4 credit hours): A
Physics I (5 credit hours): C
History (3 credit hours): B
English Literature (3 credit hours): B
Racquetball (1 credit hour): B
Calculus II (4 credit hours): B
Physics I (5 credit hours): C
Government (3 credit hours): A
Speech (3 credit hours): D
Psychology (3 credit hours): A

A = 4
B = 3
C = 2
D = 1
F = 0
Pass, NoPass W is neutral

Total Credit Hours

$4+5+3+3+1+4+5+3+3+3$
$= 5(3)+2(4)+1+2(5)$
$= 34$

$GPA = 4(4)+5(2)+3(3)+3(3)+1(3)+4(3)+5(2)+3(4)+3(1)+3(4)$
$= \frac{96}{34} = \boxed{2.8}$

Example 4: Suppose 30% of a car manufacturer's vehicles get 22 miles per gallon (MPG), 27% get 35 MPG, 8% get 49 MPG, and 35% get 29 MPG. What is the average MPG for all the company's cars?