10.2: Frequency Distributions and Measures of Central Tendency

There are two basic types of numerical measures that describe data sets.

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

.

Summation Notation: This is a compact way to write "add up the numbers $x_1, x_2, \ldots, x_n$"

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

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

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

$$\sum_{i=1}^{6} x_i^2 = (8)^2 + (2)^2 + (6)^2 + (10)^2 + (4)^2 + (9)^2$$
$$= 64 + 4 + 36 + 100 + 16 + 81$$

The Mean:

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* is calculated by dividing the sum of the measurements by the number of measurements. The mean is sometimes known as the average.

For a population of size n, the mean is

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

μ = mu

For a sample of size n taken from a larger population, the mean is

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

$\bar{x}$ = "x-bar"

Example 1: A sample has the following measurements. Find the mean.

23.1, 14.6, 21.2, 18.9, 19.3, 17.6

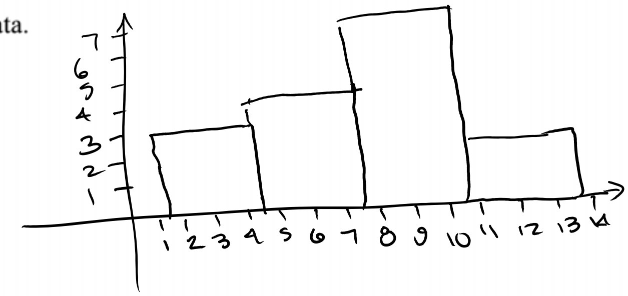
$$\bar{x} = \frac{23.1 + 14.6 + 21.2 + 18.9 + 19.3 + 17.6}{6} = \frac{114.7}{6} \approx 19.1167 \approx \boxed{19.12}$$

> 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 approximated by
>
> $$\mu = \frac{\sum_{i=1}^{k} x_i f_i}{n} = \frac{x_1 f_1 + x_2 f_2 + \ldots + x_k f_k}{n} \text{ for a population, or}$$
>
> $$\bar{x} = \frac{\sum_{i=1}^{k} x_i f_i}{n} = \frac{x_1 f_1 + x_2 f_2 + \ldots + x_k f_k}{n} \text{ for a sample.}$$
>
> In both cases, $n = \sum_{i=1}^{k} x_i f_i$ is the total number of measurements.

Example 2: Approximate the mean for the grouped data.

midpt	Interval	Frequency
3	1.5–4.5	3
6	4.5–7.5	4
9	7.5–10.5	7
12	10.5–13.5	2

$n = 16$



Note: The same data can be represented as

Interval	Frequency
2–4	3
5–7	4
8–10	7
~~12-13~~ 11-13	2

For each category, we assume the data points are in the middle.

Midpoint	Interval	Frequency
3	2-4	3
6	5-7	4
9	8-10	7
12	11-13	2

$n = 16$

To get the midpoint, average the upper and lower boundaries:

$$\frac{2+4}{2} = \frac{6}{2} = 3 \qquad \frac{4.5+7.5}{2} = 6$$

$$\frac{5+7}{2} = \frac{12}{2} = 6$$

See next page

$n = 16$ next page 2 2

$$\bar{x} = \frac{\sum x_i f_i}{n} = \frac{3(3)+6(4)+9(7)+12(2)}{16}$$

$$= \frac{3+3+3+6+6+6+6+9+9+9+9+9+9+9+12+12}{16}$$

$$= \frac{120}{16} = 7.5 \Rightarrow \boxed{\bar{x} = 7.5}$$

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 76.

The average (mean) is $\mu = \frac{88+99+7+78+89+94+76}{7} = \frac{531}{7} \approx 75.86$

6 out of the 7 students are above the average. The 7 pulled the mean way down.

The *median* is unaffected by extreme values. 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 3: Find the median of the test scores 88, 99, 7, 78, 89, 94, and ~~76~~ 76.

7 76 78 88 89 94 99

median

$\boxed{\text{The median is 88.}}$

The mode:

> The Mode:
>
> The *mode* is the most frequently occurring measurement in a data set. There may be a unique mode, several modes, or no mode.

Example 4: Find the median and mode for the following data sets.

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

Median is 5
Mode is 5

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

$\text{median} = \frac{3+5}{2} = \frac{8}{2} = 4$

Modes are 3, 7 (bimodal data)

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

$\text{median} = \frac{6+7}{2} = \frac{13}{2} = 6.5$

No mode

> The Median (Grouped Data):
>
> For grouped data, the median is the value of the variable that divides the area of the histogram into two equal portions.
> (So the area to the left of the median is equal to the area to the right of the median.)

Example 5: Approximate the median of the grouped data.

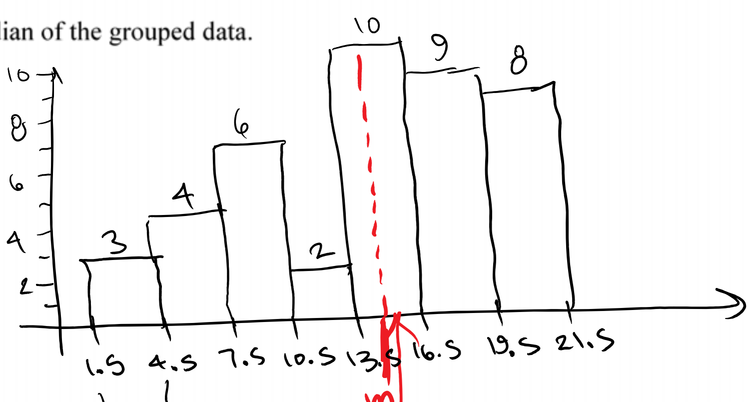
Cumulative Frequency	Interval	Frequency
3	1.5–4.5	3
7	4.5–7.5	4
13	7.5–10.5	6
15	10.5–13.5	2
25	13.5–16.5	10
	16.5–19.5	9
	19.5–21.5	8

15<21

25>21

median must be in this category

$n = 42$



width of 1 bar is $4.5-1.5=3$

median is in this bar

Let m be the median. We want 21 on one side of m, and 21 on other side.
(Half the area on each side)

Right Side

$3(8)+3(9)+10(16.5-m)=63$

$24+27+165-10m=63$

$216-10m=63$

$-10m=-153$

$m=\frac{-153}{-10}=15.3 \Rightarrow \boxed{\text{Median is } 15.3}$

Total Area $= 3(3)+4(3)+6(3)+2(3)$
$10(3)+9(3)+8(3)$
$= 3(42) = 126$

Half the area is $\frac{126}{2} = 63$