Basic Mathematics and Statistics

CHAPTER 11: NUMERICAL METHODS: MEASURES OF LOCATION

Dr. Khaing S. Htun

11.1 Introduction

Tabular and graphical methods (chapter 10)

- to summarize data
- effective in written reports and as visual aids for presentations to individuals or groups

several numerical methods of descriptive statistics that provide additional alternatives for summarizing data

11.2 Measure of Location of Ungrouped Data

Mean

raw data
not organize

Measure of location is the measures that describe the center of a distribution.

Let *x* is an arbitrary measurement of a data set.

The **population mean**: $$\mu = \frac{x_1 + x_2 + x_3 + \cdots + x_N}{N} = \frac{\sum_{i=1}^{N} x_i}{N}$$

where N is the population size.

The **sample mean**: $$\bar{x} = \frac{x_1 + x_2 + x_3 + \cdots + x_n}{n} = \frac{\sum_{i=1}^{n} x_i}{n}$$

where n is the sample size.

11.2 Measure of Location of Ungrouped Data

Mean

The Properties of the Arithmetic Mean

1. Every set of interval-level and ratio-level data has a mean.

2. A set of data has **only one mean**.

3. The mean is a useful measure for **comparing two or more populations**.

4. The sum of the deviations of each value from the mean will always be zero, that is $\boldsymbol{\sum(x - \overline{x}) = 0}$

11.2 Measure of Location of Ungrouped Data

Mean

12 8 17 6 11 14 8 17 10 8

$$\bar{x} = \frac{\sum x_i}{n} = \frac{12+8+17+\ldots+10+8}{10}$$

$$= \frac{111}{10} = 11.1$$

11.2 Measure of Location of Ungrouped Data

Median

The median is the **midpoint in the data set** that has been **ranked in increasing order**.

The calculation of the median:-

1. **Rank** the given data set **in increasing order**.
2. **Find the position of the middle term =** $\frac{n+1}{2}$. The value of this term is the median. If the given data set represents a population, replace *n* by *N*.

(a) For an **odd number** of observations, the **median is the middle value**.

(b) For an **even number** of observations, the **median is the average of the two middle value**.

11.2 Measure of Location of Ungrouped Data

Median

even

$n = 10$

12 8 17 6 11 14 8 17 10 8

ranked data	:	6	8	8	8	10	11	12	14	17	17
order	:	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)

even number

the position of median $= \dfrac{n+1}{2} = \dfrac{10+1}{2} = 5.5$

the median is $\dfrac{10+11}{2} = 10.5$

11.2 Measure of Location of Ungrouped Data

Mode

The mode is the value that **occurs with the highest frequency** in a data set.

12 8 17 6 11 14 8 17 10 8

ranked data	:	6	8	8	8	10	11	12	14	17	17
order	:	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)

3 times

the mode is 8 because it appears more often (3 times) than any other numbers

11.2 Measure of Location of Ungrouped Data

Trimmed Mean

k% 10%   80% left   k% 10%

increasing order

The **k% trimmed mean** is obtained by **dropping k% of the smallest values** and **k% of the largest values** from the given data and then **calculating the mean of the remaining (100 – 2k)%** of the values.

12 8 17 6 11 14 8 17 10 8

ranked data	: ~~6~~	8	8	8	10	11	12	14	17	~~17~~
order	: (1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)

the 10% trimmed mean - *exclude 10% of largest & 10% smallest values, therefore compute mean for remaining 80%*

First, find the number of data is deleted to be equal (2k)% of n = 2(10/100)(10) = 2 values which is the value 6 and 17.

Thus the 10% trimmed mean is $\frac{\sum x - 6 - 17}{10 - 2} = \frac{111 - 6 - 17}{8} = \frac{88}{8} = 11$

11.3 Measure of Location of Ungrouped Data

Weighted Mean

The weighted mean is founded by multiplying each observation by its corresponding weight is computed by :

weighted mean $$\bar{x}_w = \frac{w_1x_1 + w_2x_2 + w_3x_3 + \cdots + w_kx_k}{w_1 + w_2 + w_3 + \ldots + w_k} = \frac{\sum_{i=1}^{k} x_i w_i}{\sum_{i=1}^{k} w_i}$$

where x_i is the value of the observation i

w_i is the weight of the observation i

k is the number of category

11.3 Measure of Location of Ungrouped Data

Weighted Mean

The Carter Construction Company pays its hourly employees \$6.50, \$7.50, or \$8.50 per hour. There are 26 hourly employees,

- 14 are paid at the \$6.50 rate,
- 10 at the \$7.50 rate, and
- 2 at the \$8.50 rate.

What is the mean hourly rate paid the 26 employees?

$$\bar{x}_w = \frac{\sum_{i=1}^{3} x_i w_i}{\sum_{i=1}^{3} w_i} = \frac{14(\$6.50)+10(\$7.50)+2(\$8.50)}{14+10+2} = \frac{\$183}{26} = \$7.038$$

The **weighted mean** hourly wage is rounded to **\$7.04**.

11.3 Measure of Location of Grouped Data

organized into frequency distribution

Mean

The population mean:

$$\mu = \frac{\sum_{i=1}^{k} f_i M_i}{\sum_{i=1}^{k} f_i}, \quad N = \sum_{i=1}^{k} f_i$$

where M_i is the midpoint of each class.
f_i is the frequency in each class.
N is the total number of frequencies or the population size.
k is the number of class.

The sample mean:

$$\bar{x} = \frac{\sum_{i=1}^{k} f_i M_i}{\sum_{i=1}^{k} f_i}, \quad n = \sum_{i=1}^{k} f_i$$

where M_i is the midpoint of each class.
f_i is the frequency in each class.
n is the total number of frequencies or the sample size.
k is the number of class.

11.3 Measure of Location of Grouped Data

Median

The sample median:

$$\text{median} = L + \frac{(i)\left[\frac{n}{2} - CF\right]}{f}$$

where L is the lower class boundaries of the class containing the median.
f is the frequency of the class containing the median.
n is the total number of frequencies or sample size.
CF is the cumulative frequency in all the classes preceding the class containing the median.
i is the width of the class.

11.3 Measure of Location of Grouped Data

Mode

The sample mode:

$$\text{mode} = L + i\left[\frac{d_1}{d_1 + d_2}\right]$$

where L is the lower class boundaries of the class containing the mode.

d_1 is the difference of the frequency between the mode class and the connecting class which has the smaller scores.

d_2 is the difference of the frequency between the mode class and the connecting class which has the higher scores.

i is the width of the class.

11.3 Measure of Location of Grouped Data

The following data is a sample data of the number of hours for reading per student in a week. Estimate the mean, the median, and the modal of the data and interpret its meaning.

Class	Number of hours	Number of Students
1	70 - 79	3
2	80 - 89	7
3	90 - 99	18
4	100 - 109	20
5	110 - 119	12
Total		60

11.3 Measure of Location of Grouped Data

Mean

$$\bar{x} = \frac{\sum_{i=1}^{k} f_i M_i}{\sum_{i=1}^{k} f_i}$$

number of hours

number of students

Class (i)	Class Limit	Midpoint (M_i)	f_i	$f_i M_i$
1	70 - 79	74.5	3	223.5
2	80 - 89	84.5	7	591.5
3	90 – 99	94.5	18	1,701.0
4	100 – 109	104.5	20	2,090.0
5	110 – 119	114.5	12	1,374.0
Total			60	5,980

$$= \frac{5{,}980}{60} = 99.67$$

So, the mean number of hours for reading per student in a week is about 99.67 hours

11.3 Measure of Location of Grouped Data

Median

$$\text{median} = L + \frac{(i)\left[\frac{n}{2} - CF\right]}{f}$$

± 0.5

Class (i)	Class Limit	Class Boundaries	f_i	CF
1	70 - 79	69.5 – 79.5	3	3
2	80 - 89	79.5 – 89.5	7	10
3	90 – 99	89.5 – 99.5	18	28
4	100 – 109	**99.5 – 109.5**	**20**	**48 (has the median)**
5	110 – 119	109.5 – 119.5	12	60
Total			60	

10

30th & 31st

$$= 99.5 + \frac{10\left(\frac{60}{2} - 28\right)}{20}$$

$$= 99.5 + 1 = 100.5$$

So, the median number of hours for reading per student in a week is about 100.5 hours.

11.3 Measure of Location of Grouped Data

Mode

$$\text{mode} = L + i\left[\frac{d_1}{d_1 + d_2}\right]$$

$$= 99.5 + 10\left[\frac{2}{2+8}\right]$$

$$= 101.5$$

Class(i)	Class Limit	Class Boundaries	f_i
1	70 - 79	69.5 – 79.5	3
2	80 - 89	79.5 – 89.5	7
3	90 – 99	89.5 – 99.5	18
4	100 – 109	**99.5 – 109.5**	**20 (has the mode)**
5	110 – 119	109.5 – 119.5	12
Total			60

$d_1 = 20 - 18 = 2$

$d_2 = 20 - 12 = 8$

So, the mode number of hours for reading per student in a week is about 101.5 hours.

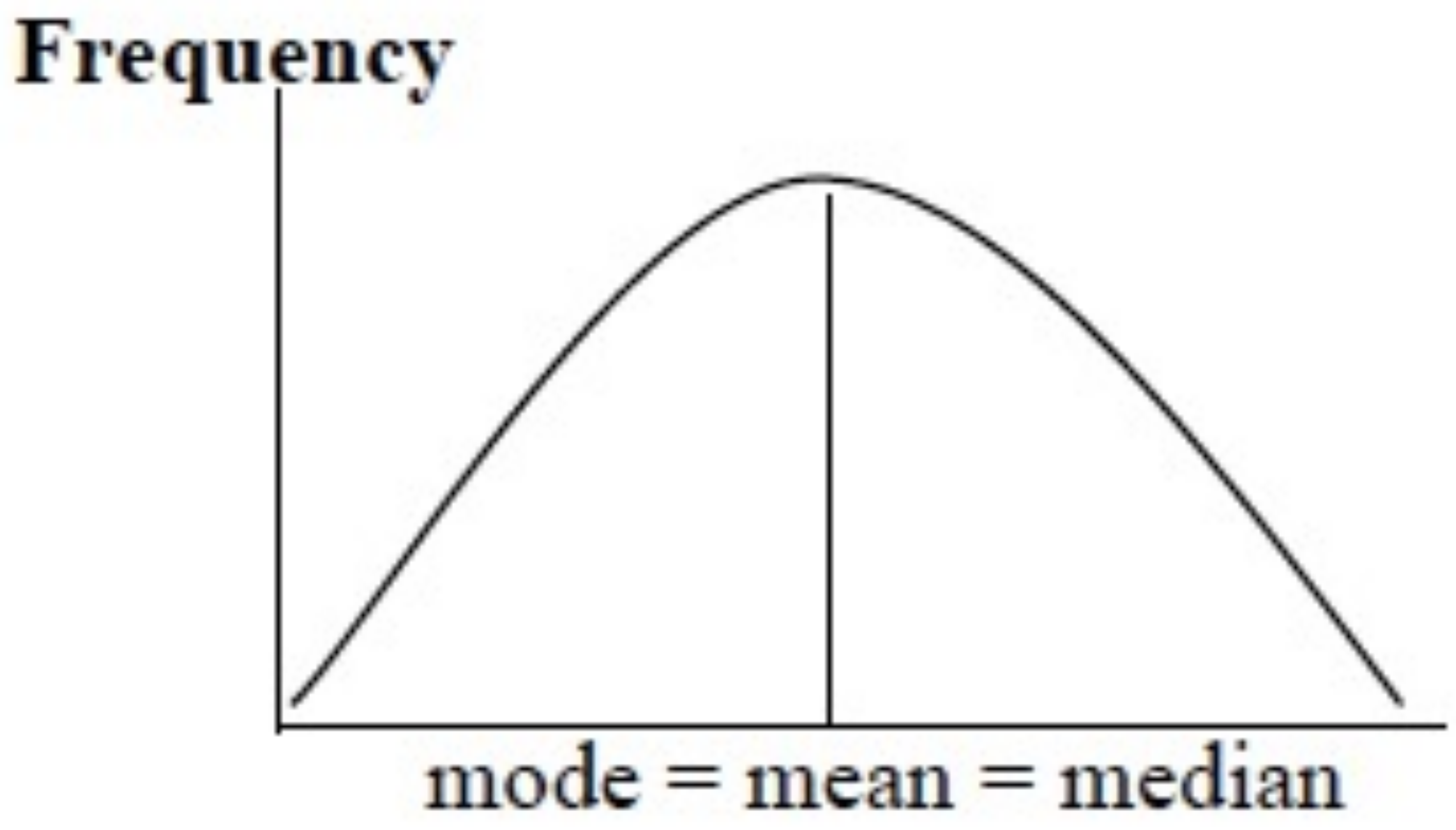
11.4 Relationship between Mean, Median, and Mode

mode = 3 medium – 2 mean

$$\boxed{\text{mode} = \text{mean} - 3(\text{mean} - \text{median})}$$

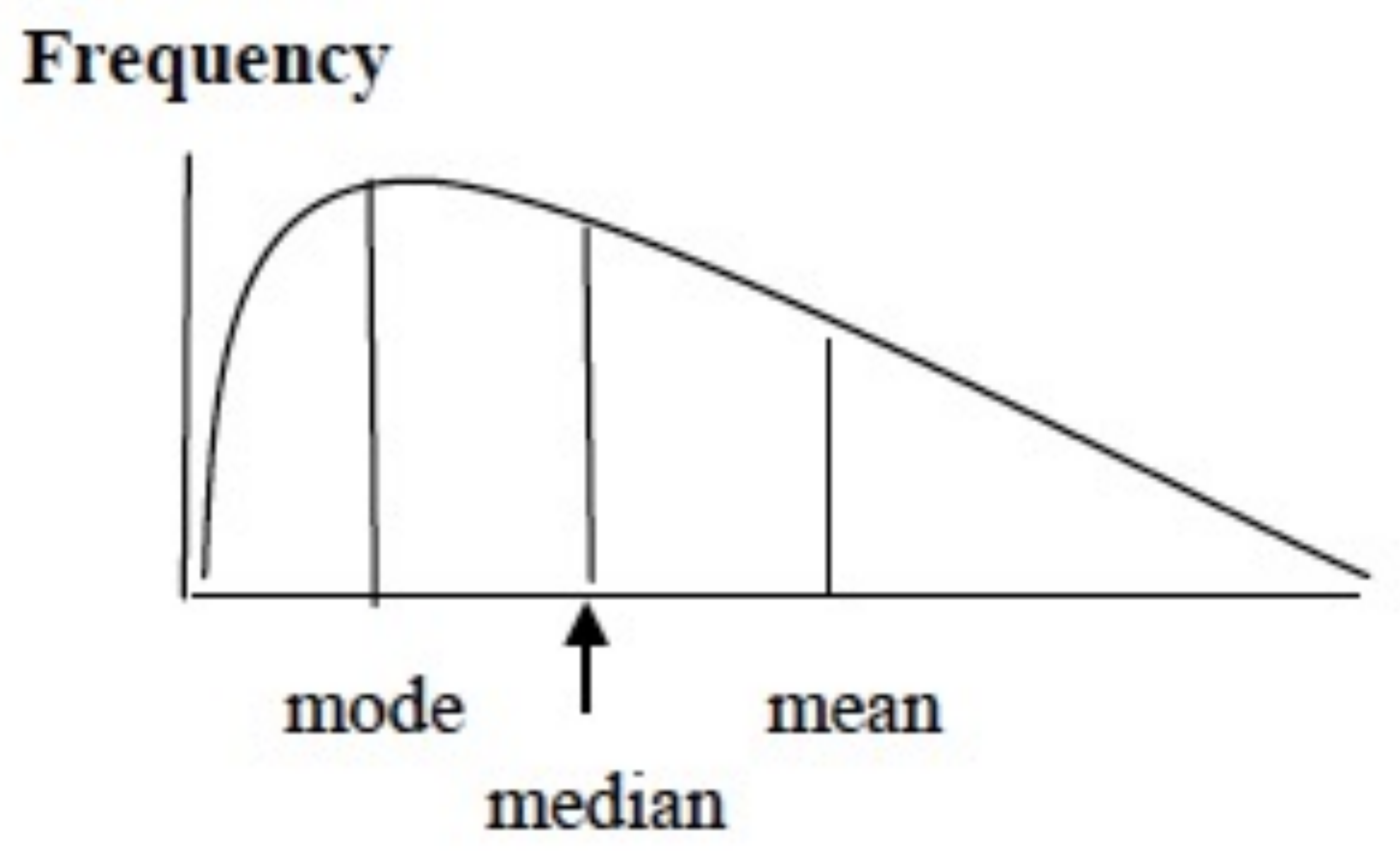
11.5 Selecting an Average for Data in a Frequency Distribution

For **a symmetric distribution (bell-shaped distribution)**, the mode, median, and mean are located at the center and are always equal.



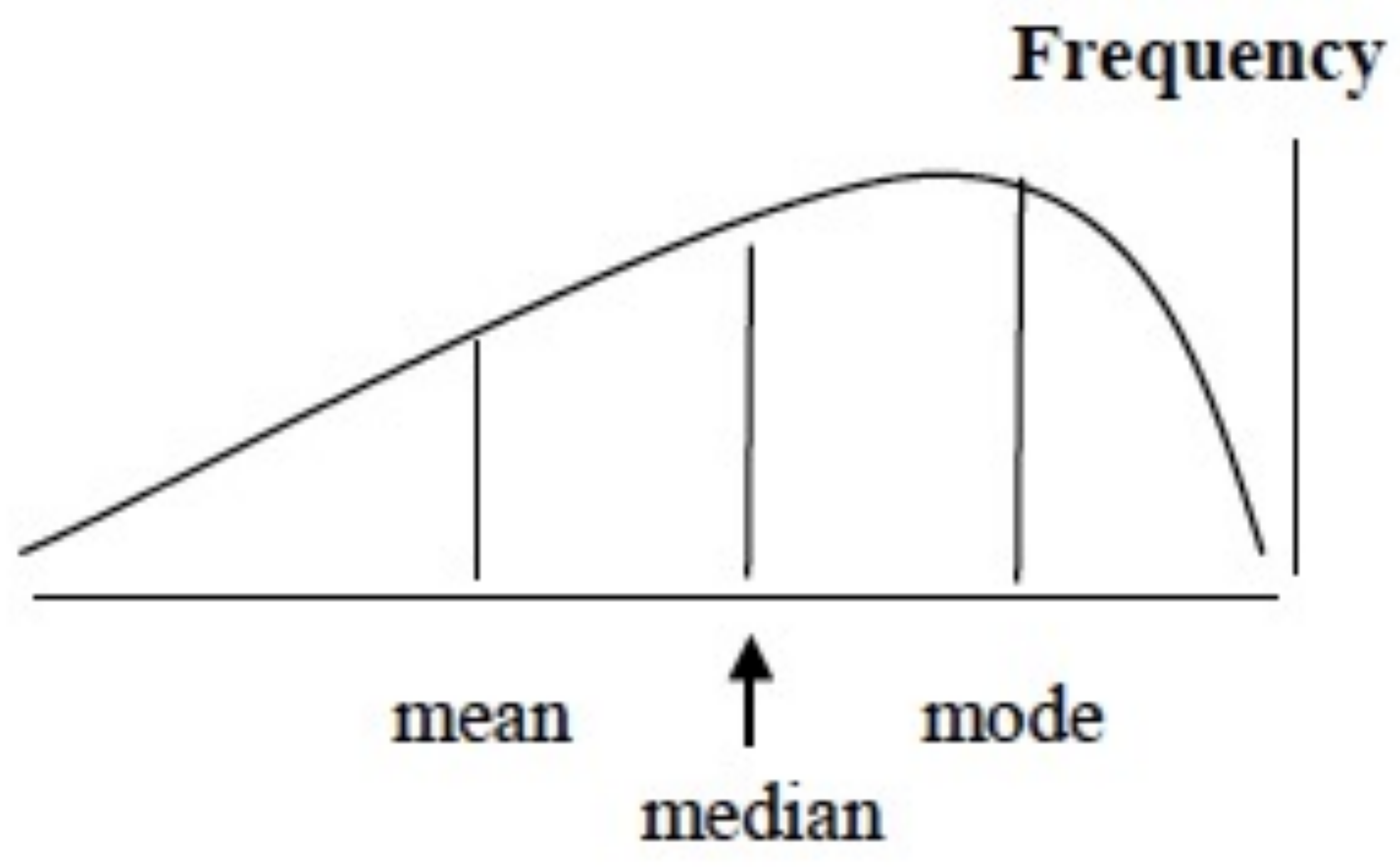
11.5 Selecting an Average for Data in a Frequency Distribution

For **a skewed distribution**, the relationship among the three averages changes.



Positively skewed distribution

Skewed right distribution



Negatively skewed distribution

Skewed left distribution

If the distribution is highly skewed, the mean would not be a good average to use because the mean is influenced more than the median and mode. Hence, the median or mode would be more representative

$n=21$

59, 65, 61, 62, 53, 55, 60, 70, 64, 56, 58, 58, 62, 62, 68, 65, 56, 59, 68, 61, 67

Ungrouped

mean

$$\frac{59 + 65 + 61 + 62 + 53 + 55 + 60 + 70 + 64 + 56 + 58 + 58 + 62 + 62 + 68 + 65 + 56 + 59 + 68 + 61 + 67}{21}$$

=61.38095

median location of median $= \frac{n+1}{2} = \frac{21+1}{2} = 11$

53, 55, 56, 56, 58, 58, 59, 59, 60, 61, 61, 62, 62, 62, 64, 65, 65, 67, 68, 68, 70

11^{th}

mode

53, 55, 56, 56, 58, 58, 59, 59, 60, 61, 61, 62, 62, 62, 64, 65, 65, 67, 68, 68, 70

Grouped

Estimate mean

$$\frac{1288}{21} = \mathbf{61.333...}$$

Seconds	Frequency	Midpoint	Midpoint × Frequency fx
51 - 55	2	53	106
56 - 60	7	58	406
61 - 65	8	63	504
66 - 70	4	68	272
			1288

Estimate median

$$= 60.5 + \frac{(21/2) - 9}{8} \times 5$$

$$= 60.5 + 0.9375$$

$$= \mathbf{61.4375}$$

Seconds	Frequency	CF	Calss bound
51 - 55	2	2	50.5 - 55.5
56 - 60	7	9	55.5 – 60.5
61 - 65	8	17	60.5 – 65.5
66 - 70	4	21	65.5 – 70.5

+5

Estimate mode

$$= 60.5 + \frac{8-7}{(8-7)+(8-4)} \times 5$$

$$= 60.5 + (1/5) \times 5$$

$$= \mathbf{61.5}$$

d_1 d_2

Seconds	Frequency	Calss bound
51 - 55	2	50.5 - 55.5
56 - 60	7	55.5 – 60.5
61 - 65	8	60.5 – 65.5
66 - 70	4	65.5 – 70.5

+5

Grouped Data

Example:

The ages of the 112 people who live in a small village are grouped as follows:

Age	Number	Midpoint x	fx
0 - 9	20	5	100
10 - 19	21	15	315
20 - 29	23	25	575
30 - 39	16	35	560
40 - 49	11	45	495
50 - 59	10	55	550
60 - 69	7	65	455
70 - 79	3	75	225
80 - 89	1	85	85
		Totals:	**3360**

$$\textbf{Mean} = \frac{3360}{112} = \mathbf{30}$$

$$\textbf{Median} = 20 + \frac{(112/2) - 41}{23} \times 10$$

$$= 20 + 6.52...$$

$$= \mathbf{26.5} \text{ (to 1 decimal)}$$

$$\textbf{Mode} = 20 + \frac{23 - 21}{(23 - 21) + (23 - 16)} \times 10$$

$$= 20 + 2.22...$$

$$= \mathbf{22.2} \text{ (to 1 decimal)}$$

Assignment

Deadline for submission: Monday, September 20, 2021

- Exercises 11
- Example – Obtain mean, median and mode of both ungrouped and grouped data

12	14	19	18	15	15	18	17	20	27
22	23	22	21	33	28	14	18	16	13

Class	Class Limit	Class Boundary	Midpoint	Frequency (f)	Relative Frequency (%)
1	12 – 16	11.5 – 16.5	14	7	35.0
2	17 – 21	16.5 – 21.5	19	7	35.0
3	22 – 26	21.5 – 26.5	24	3	15.0
4	27 – 31	26.5 – 31.5	29	2	10.0
5	32 – 36	31.5 – 36.5	34	1	5.0
Total				**20**	**100.0**