

# Probability and Statistics (PHM111s)-Lecture 3

## **Part I: Introduction to Statistical Methods.**

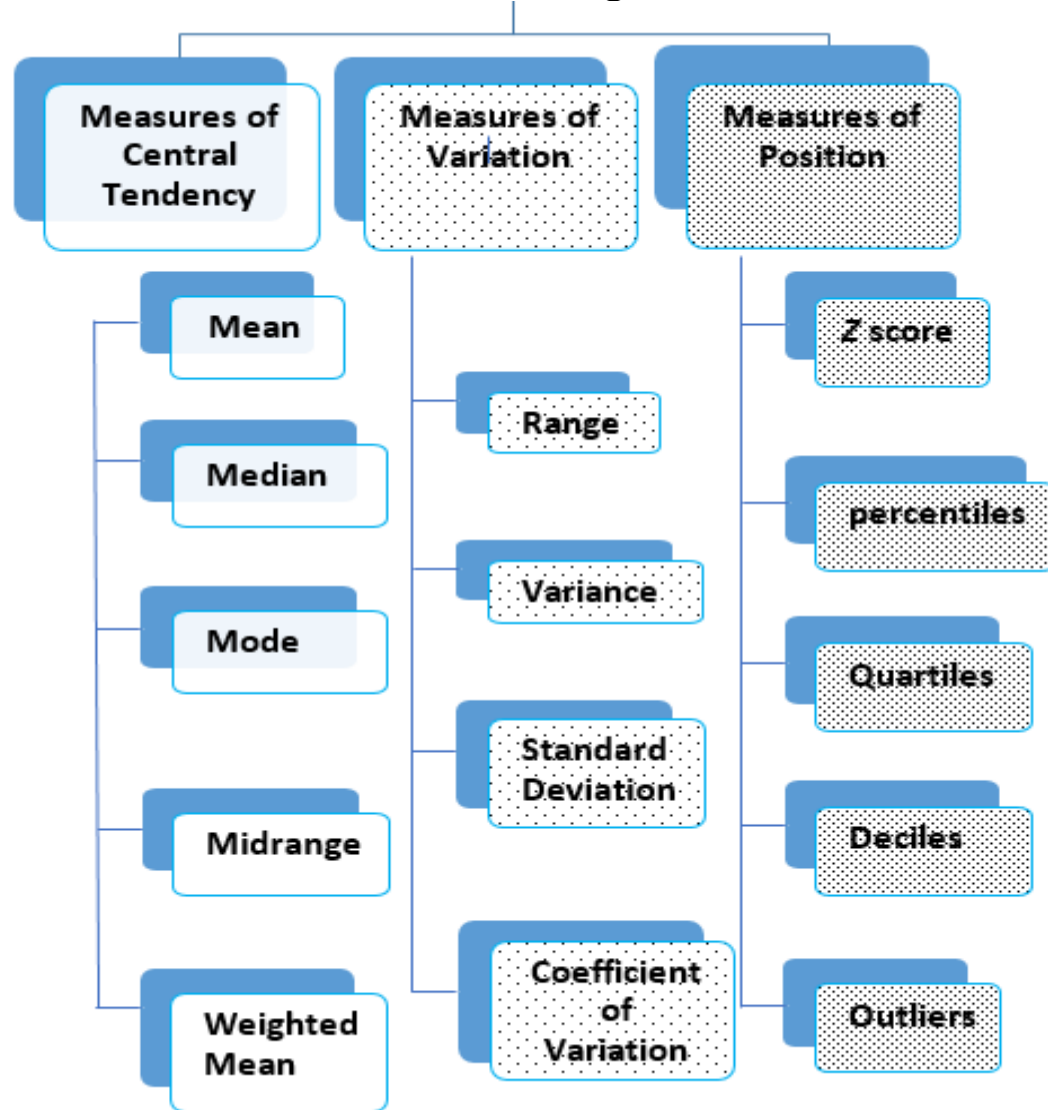
## **Part II: Methods of Descriptive Statistics.**

- 1-Collecting Data.
- 2-Organizing Data.
- 3-Presenting Data.
- 4-Summarizing Data.

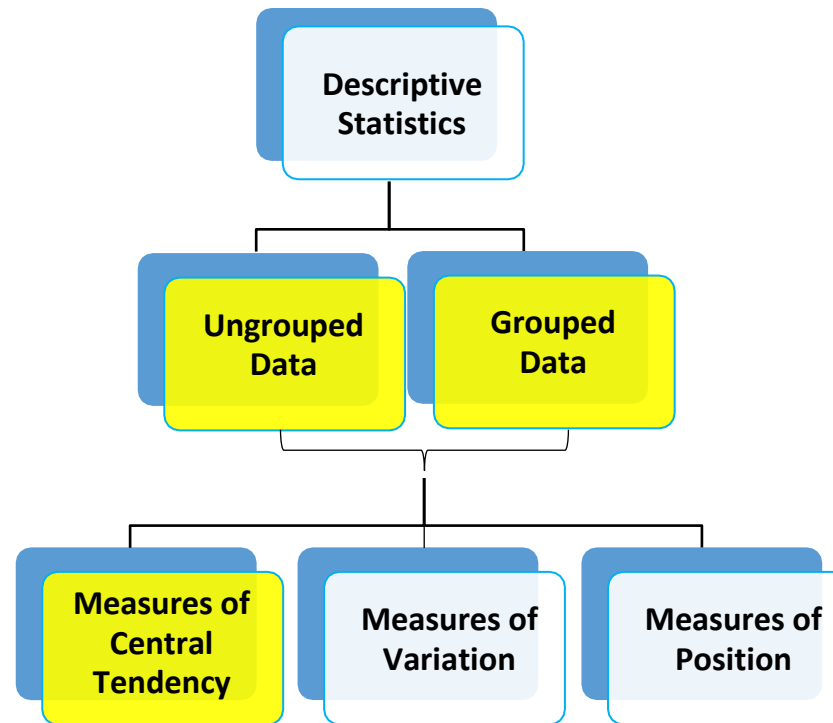
## **Part III: Introduction to Probability.**

## **Part IV: Methods of Inferential Statistics.**

#### 4. Summarizing Data.



Parameter for Population & Statistic for Sample



## 1- Arithmetic Mean

### Case 1: Ungrouped Data

a) The **mean** for **sample** with size  $n$ , is given as:

$$\bar{X} = \frac{X_1 + X_2 + X_3 + \dots + X_n}{n} = \frac{\sum_{i=1}^n X_i}{n}$$

b) For a **population** with size  $N$ , the mean is given as:

$$\mu = \frac{X_1 + X_2 + X_3 + \dots + X_n}{N} = \frac{\sum_{i=1}^N X_i}{N}$$

**Example1:** The data represent the number of days off per year for a sample of individuals selected from nine different countries. Find the mean.

20, 26, 40, 36, 23, 42, 35, 24, 30

**Solution:**

$$\bar{X} = \frac{\sum X}{n} = \frac{20 + 26 + 40 + 36 + 23 + 42 + 35 + 24 + 30}{9} = \frac{276}{9} = 30.7 \text{ days}$$

**Case 2: Grouped Data**

$$\bar{X} = \frac{\sum_{i=1}^n f_i \cdot X_i}{\sum_{i=1}^n f_i} = \frac{1}{n} \sum_{i=1}^n f_i \cdot X_i$$

**Example 2:** The following table gives the frequency distribution of the number of orders received each day during the past 50 days at the office of a mail-order company. Calculate the mean.

**Solution**

Number of orders	<i>f</i>
10 – 12	4
13 – 15	12
16 – 18	20
19 – 21	14

n = 50

Number of orders	<i>f</i>	<i>X</i>	<i>f</i> · <i>X</i>
10 – 12	4	11	44
13 – 15	12	14	168
16 – 18	20	17	340
19 – 21	14	20	280
	n = 50		= 832

$$\bar{X} = \frac{\sum f \cdot X}{n} = \frac{832}{50} = 16.64 \text{ orders}$$

## 2- The Median

### Case 1: Ungrouped Data

**Example 3:** The number of rooms in the seven hotels in downtown Pittsburgh is 713, 300, 618, 595, 311, 401, and 292. Find the median.

**Solution:**

**Step 1** Arrange the data in order.

292, 300, 311, 401, 595, 618, 713

**Step 2** Select the middle value.

292, 300, 311, 401, 595, 618, 713

↑  
Median

Hence, the median is 401 rooms.

**Example 4:** The number of cloudy days for the top 10 cloudiest cities is shown. Find the median.

209, 223, 211, 227, 213, 240, 240, 211, 229, 212

**Solution:**

Arrange the data in order.

209, 211, 211, 212, 213, 223, 227, 229, 240, 240

$$\begin{array}{c} \uparrow \\ \text{Median} \\ MD = \frac{213 + 223}{2} = 218 \end{array}$$

Hence, the median is 218 days.

### Case 2: Grouped Data

$$\text{Median} = MD = L + i \left( \frac{\frac{n}{2} - F_{m-1}}{f_m} \right)$$

**Example 5:** Based on the grouped data below, find the median.

**Solution**

Seconds	$f$
51 – 55	2
56 – 60	7
61 – 65	8
66 – 70	4

Construct the cumulative frequency distribution

Seconds	$f$	$cf$
51 – 55	2	2
56 – 60	7	9
61 – 65	8	17
66 – 70	4	21

$$\frac{n}{2} = \frac{21}{2} = 10.5 \rightarrow \text{class median is the 3}^{rd} \text{ class}$$

So,  $F_{m-1} = 9$ ,  $f_m = 8$ ,  $i = 5$  and  $L = 60.5$

Therefore,

$$\begin{aligned} MD &= L + i \left( \frac{\frac{n}{2} - F_{m-1}}{f_m} \right) \\ &= 60.5 + 5 \left( \frac{10.5 - 9}{8} \right) \\ &= 60.5 + 0.9375 = 61.4375 \end{aligned}$$

## Finding Median Graphically (from Ogive)



### Example (1)

The following table shows the frequency distribution for the scores of 60 students in an exam.

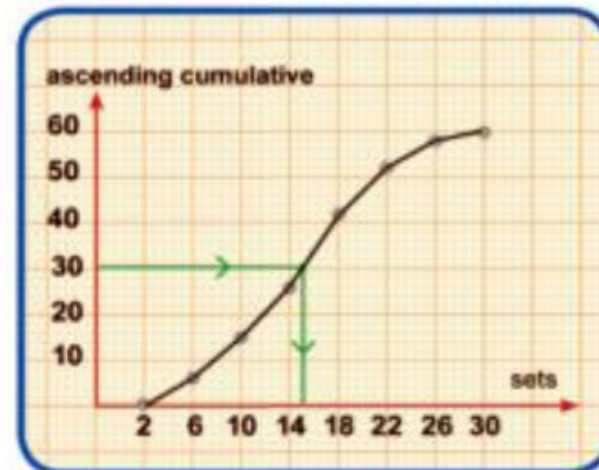
Sets	2–	6–	10–	14–	18–	22–	26–	Total
Frequency	6	9	12	15	10	5	3	60

Find the median of the distribution using the ascending cumulative frequency table.

#### Solution

- 1 Draw an ascending cumulative frequency table.
- 2 Find the order of the median =  $\frac{60}{2} = 30$
- 3 Draw the ascending cumulative frequency curve, and get the median from the graph.

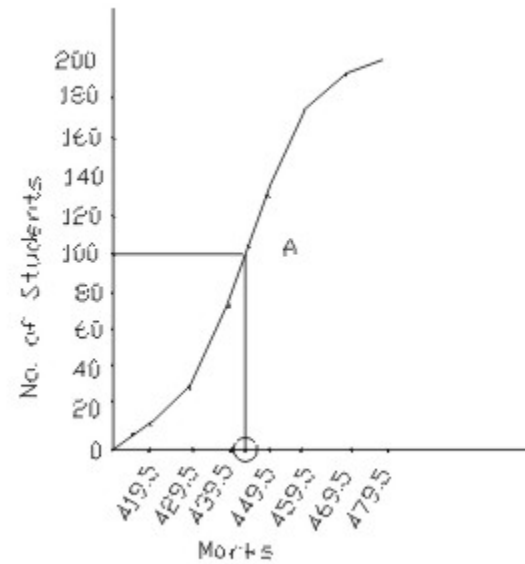
The upper limits of the sets	The ascending cumulative frequency
Less than 2	0
Less than 6	6
Less than 10	15
Less than 14	27
Less than 18	42
Less than 22	52
Less than 26	57
Less than 30	60



From the graph, the median = 14.8 mark

Marks	Conversion into exclusive series	No. of students	Cumulative Frequency
(x)		(f)	(C.M)
410-419	409.5-419.5	14	14
420-429	419.5-429.5	20	34
430-439	429.5-439.5	42	76
440-449	439.5-449.5	54	130
450-459	449.5-459.5	45	175
460-469	459.5-469.5	18	193
470-479	469.5-479.5	7	200

$$\frac{n}{2} = \frac{200}{2} = 100$$





### 3- The Mode

#### Case 1: Ungrouped Data

**Example 6:** The data show the number of licensed nuclear reactors in the United States for a recent 15-year period.  
Find the mode.

104 104 104 104 104  
107 109 109 109 110  
109 111 112 111 109

#### Solution

Since the values 104 and 109 both occur 5 times, the modes are 104 and 109. The data set is said to be bimodal.

#### Case 2: Grouped Data

$$Mode = L + i \left( \frac{f_m - f_{m-1}}{(f_m - f_{m-1}) + (f_m - f_{m+1})} \right)$$

**Example 7:** Based on the grouped data below, find the mode.

#### Solution

Seconds	$f$
51 – 55	2
56 – 60	7
61 – 65	8
66 – 70	4

$$L = 60.5, f_{m-1} = 7, f_m = 8, f_{m+1} = 4, i = 5$$

$$\begin{aligned} Mode &= 60.5 + 5 \left( \frac{8 - 7}{(8 - 7) + (8 - 4)} \right) \\ &= 60.5 + 5 \left( \frac{1}{5} \right) = 61.5 \end{aligned}$$

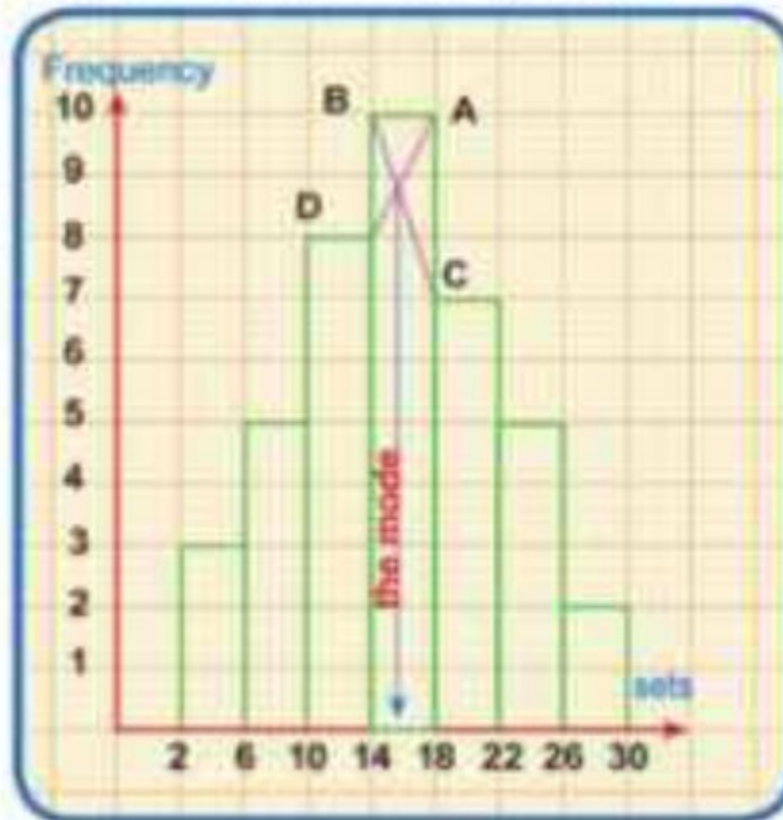
## Finding Mode Graphically (from Histogram)



### Example (1)

The following table shows the frequency distribution for the scores of 60 students in an exam.

Sets	2–	6–	10–	14–	18–	22–	26–	Total
Frequency	6	9	12	15	10	5	3	60



#### 4- The Midrange

$$MR = \frac{\text{lowest value} + \text{highest value}}{2}$$

**Example 8:** Find the midrange of data:

18.0, 14.0, 34.5, 10, 11.3, 10, 12.4, 10

**Solution**

The smallest bonus is \$10 million and the largest bonus is \$34.5 million.

$$MR = \frac{10 + 34.5}{2} = \frac{44.5}{2} = \$22.25 \text{ million}$$

Notice that this amount is larger than seven of the eight amounts and is not typical of the average of the bonuses. The reason is that there is one very high bonus, namely, \$34.5 million.

#### 5- The Weighted Mean

$$\bar{X} = \frac{w_1X_1 + w_2X_2 + \dots + w_nX_n}{w_1 + w_2 + \dots + w_n} = \frac{\sum_{i=1}^n w_iX_i}{\sum_{i=1}^n w_i}$$

**Example 9:** A student received an A in English Composition I (3 credits), a C in Introduction to Psychology (3credits), a B in Biology I (4credits), and a D in Physical Education (2credits). Assuming A =4 grade points, B =3 grade points, C =2 grade points, D =1 grade point, and F =0 grade points, find the student's grade point average.

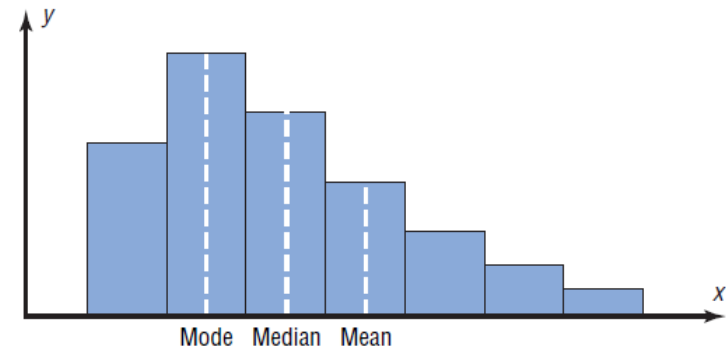
**Solution**

Course	Credits ( $w$ )	Grade ( $X$ )
English Composition I	3	A (4 points)
Introduction to Psychology	3	C (2 points)
Biology I	4	B (3 points)
Physical Education	2	D (1 point)

$$\text{The grade point average (GPA)} = \bar{X} = \frac{\sum_{i=1}^n w_iX_i}{\sum_{i=1}^n w_i} = \frac{(3)(4) + (3)(2) + (4)(3) + (2)(1)}{3 + 3 + 4 + 2} = \frac{32}{12} = 2.7$$

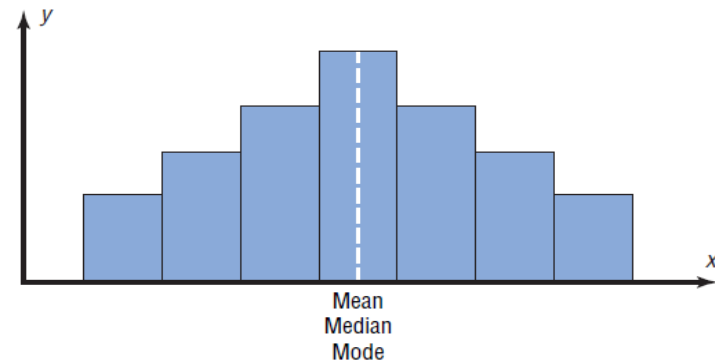
## Distribution Shapes

**positively skewed or right-skewed distribution,**



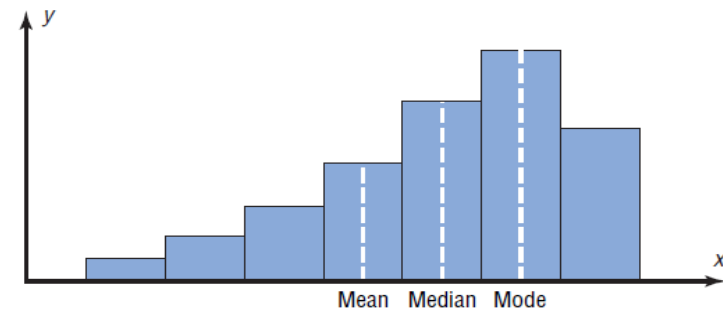
(a) Positively skewed or right-skewed

**symmetric distribution,**



(b) Symmetric

**negatively skewed or left-skewed.**



(c) Negatively skewed or left-skewed