References

Walpole, R. E., Myers, R. H., and S. L. Myers (2007), Probability and Statistics for Engineers and Scientists, 8th ed., Prentice-Hall, Inc., Upper Saddle River, New Jersey.

Program

د.محمود هندي، د.خلف سلطان، مفاهيم لطرق التحليل الإحصائي ، مكتبة الرشد٢٠٠٤.

The first lecture

We will examine in this lecture :

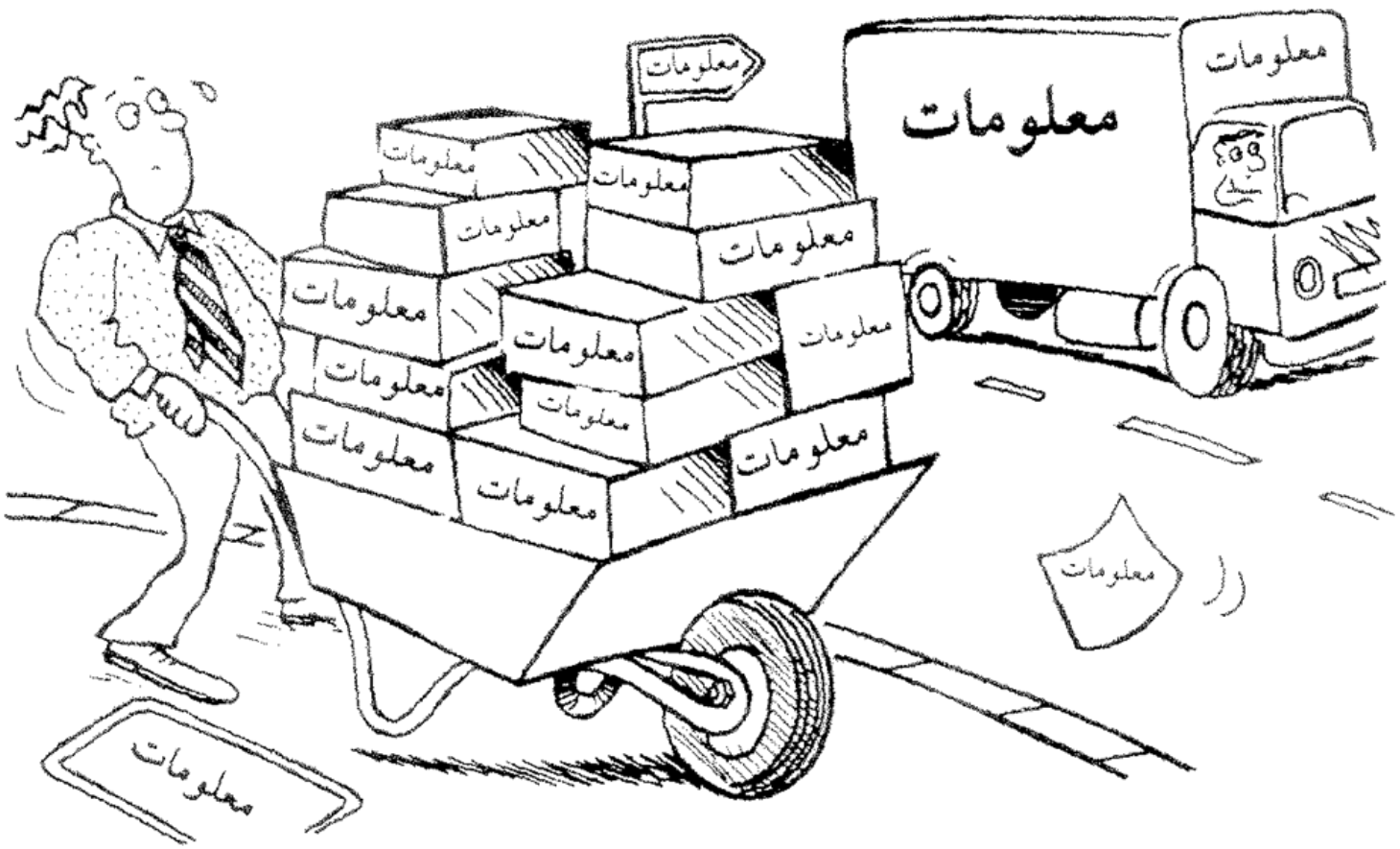
- **Definition of Statistics.**
- **The difference between the population and the sample.**
- **Branches of Statistics.**
- **Displaying data.**

Introduction to Statistics

The reasons for the appearance of Statistics:

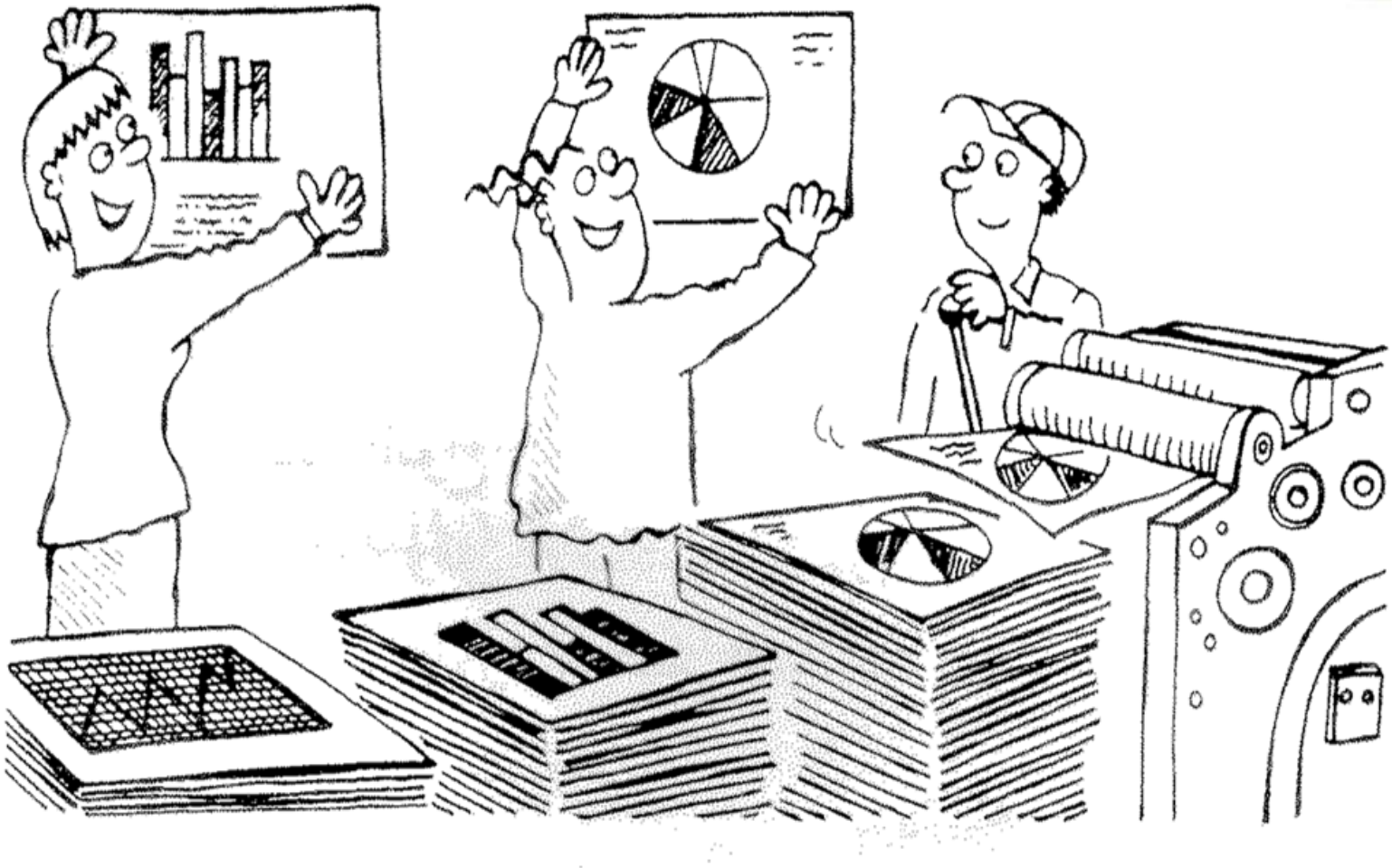
- **Census community.**
- **Inventory of the wealth of individuals.**
- **Data on births, deaths and production and consumption.**

Collection

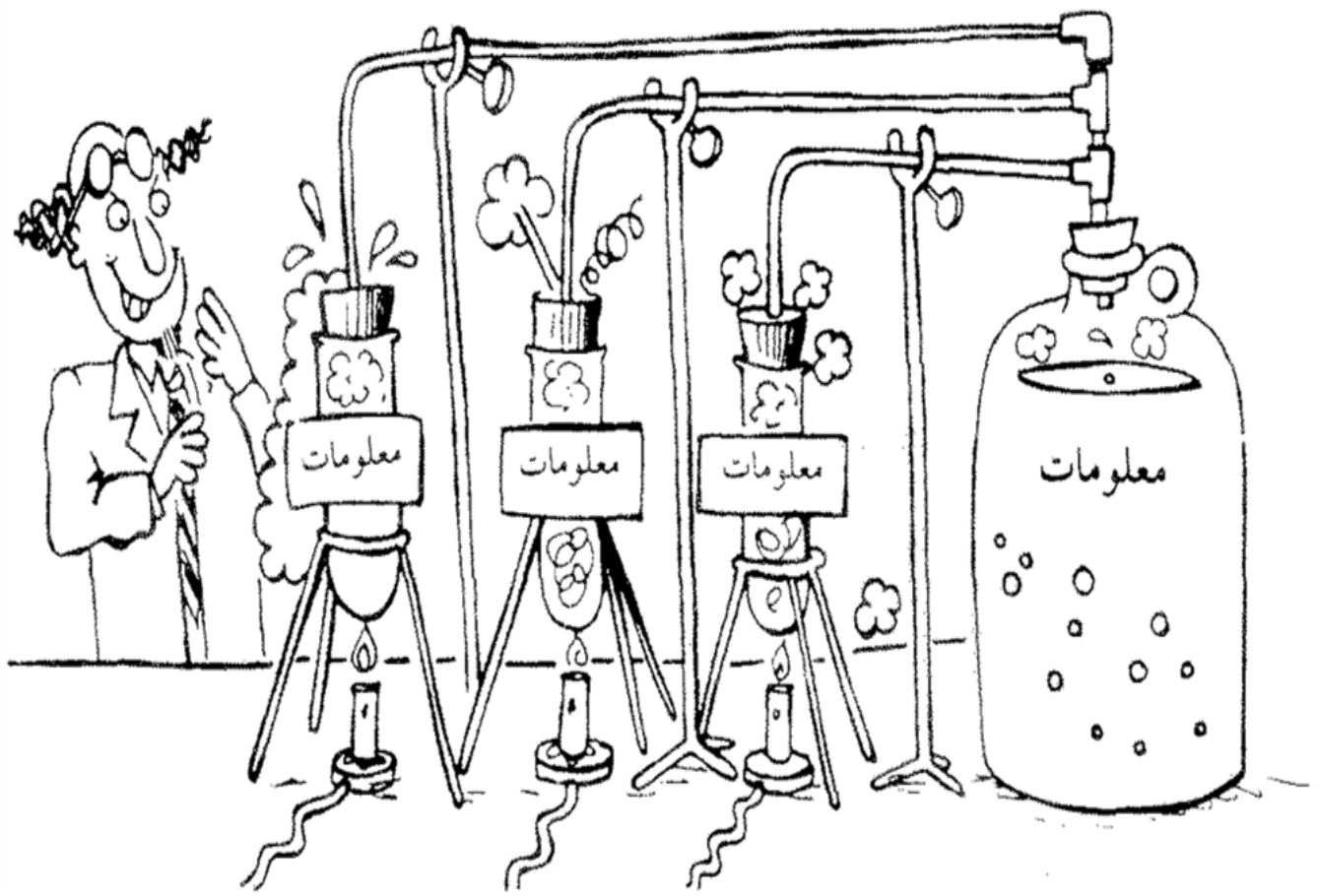

معلومات
معلومات
معلومات
معلومات
معلومات
معلومات
معلومات
معلومات
معلومات
معلومات
معلومات
معلومات
معلومات
معلومات
معلومات
معلومات
معلومات
معلومات
معلومات

Organization

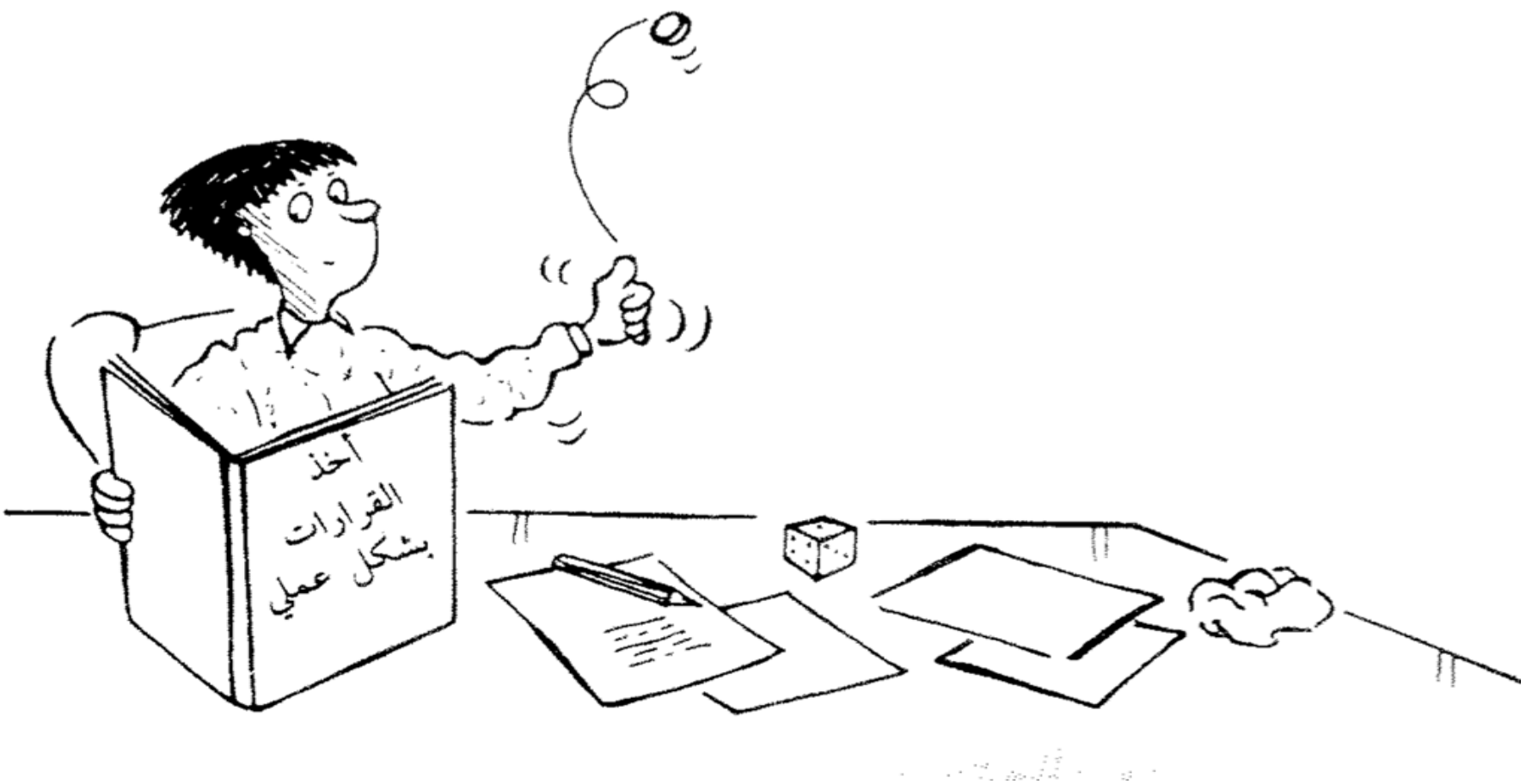
Representation



Analysis of data



Drawing of inferences



Definition statistics

Statistics is a science of collecting , organizing , analyzing and interpreting data in order to make decisions.

Basic Definitions

Definition of a population

A **population** is the collection of all outcomes, responses, measurements, or counts that are of interest.

Types of a population

Finite population: the total number of its observations is a finite number.

- The number of students in computer science college.
- The number of cards in a deck.

Infinite population: the total number of its observations is an infinite number.

- The number of stars in sky.
- The observations obtained by measuring the atmospheric pressure every day from the past on into the future .

Definition of a sample

A **sample** is a subset of a population that is representative of the population.

Reasons draw a sample, rather than study a population

- **We can not study the population :huge, destinations.**
- **Preservation from loss**
- **Less cost.**
- **Save time.**
- **More inclusive.**
- **More accuracy.**

Branches of Statistics

Descriptive statistics is the branch of statistics that involves the organization, summarization , and display of data.

Statistical inference is the branch of statistics that involves using a sample to draw conclusions about a population.

Definition of data

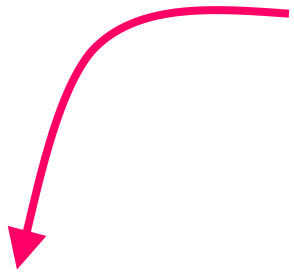
Data consist of information coming from observations, counts, measurement, or responses. The singular for data is datum.

Types of data

1. **Quantitative data:** it can be measured in the usual sense like length, weight , and age.

2. **Qualitative data:** it cannot be measured in the usual sense but it can be ordered or ranked. for example:marital status,blood group and eye color

Descriptive statistics



Class Intervals	Frequency
50-59	3
60-69	5
70-79	18
80-89	16
90-99	8
Total	50

51	95	70	74	73	90	71	74	90	67
91	72	83	89	50	80	72	84	85	69
62	82	87	76	91	76	87	75	78	79
71	96	81	88	64	82	73	57	86	70
80	81	75	85	74	90	83	66	77	91

Organization Data

We will learn how to creat:

- **Frequency table.**
- **Relative frequency table.**
- **Percentage frequency table.**

Example (1):
the following data represent the level of 60 students in a course:

D	B	E	C	D	B	D	C	E	A
B	E	C	D	B	D	D	A	E	C
C	D	A	C	E	D	C	C	D	B
D	E	D	D	A	D	D	C	D	C
D	A	B	D	B	D	C	D	C	E
D	B	C	C	E	D	C	C	D	A

Example (2):

the following data represent the marks of 50 students in a course:

51	95	70	74	73	90	71	74	90	67
91	72	83	89	50	80	72	84	85	69
62	82	87	76	91	76	87	75	78	79
71	96	81	88	64	82	73	57	86	70
80	81	75	85	74	90	83	66	77	91

Definition of frequency table (frequency distribution)

A **frequency distribution** is a table that shows **classes** or **intervals** of data entries with a count of the number of entries in each class. The **frequency f** of a class is a number of data entries in the class.

Organization qualitative data

Example (1):

the following data represent the level of 60 students in a course:

D	B	E	C	D	B	D	C	E	A
B	E	C	D	B	D	D	A	E	C
C	D	A	C	E	D	C	C	D	B
D	E	D	D	A	D	D	C	D	C
D	A	B	D	B	D	C	D	C	E
D	B	C	C	E	D	C	C	D	A

Level	tally	frequency
A		
B		
C		
D		
E		
Total		

Example (1):

the following data represent the level of 60 students in a course:

D	B	E	C	D	B	D	C	E	A
B	E	C	D	B	D	D	A	E	C
C	D	A	C	E	D	C	C	D	B
D	E	D	D	A	D	D	C	D	C
D	A	B	D	B	D	C	D	C	E
D	B	C	C	E	D	C	C	D	A

Level	tally	frequency
A	𝍸 |	6
B		
C		
D		
E		
Total		

Example (1):

the following data represent the level of 60 students in a course:

D	B	E	C	D	B	D	C	E	A
B	E	C	D	B	D	D	A	E	C
C	D	A	C	E	D	C	C	D	B
D	E	D	D	A	D	D	C	D	C
D	A	B	D	B	D	C	D	C	E
D	B	C	C	E	D	C	C	D	A

Level	tally	frequency
A	𝍸 |	6
B	𝍸 | | |	8
C		
D		
E		
Total		

Level	tally	frequency
A	𝍸 |	6
B	𝍸 | | |	8
C	𝍸 𝍸 𝍸 |	16
D	𝍸 𝍸 𝍸 | |	22
E	𝍸 | | |	8
Total		60

Table(1): frequency table

Level	frequency
A	6
B	8
C	16
D	22
E	8
Total	60

Table(1): frequency table

Level	A	B	C	D	E	Total
frequency	6	8	16	22	8	60

Organization quantitative data

Example (2):

the following data represent the marks of 50 students in a course:

51	95	70	74	73	90	71	74	90	67
91	72	83	89	50	80	72	84	85	69
62	82	87	76	91	76	87	75	78	79
71	96	81	88	64	82	73	57	86	70
80	81	75	85	74	90	83	66	77	91

Frequency distribution for quantitative data

For large samples, we can't use the simple frequency table to represent the data.

We need to divide the data into groups or intervals or classes.

So, we need to determine:

- **First step :the number of intervals (k).**
- **Second step :the range (R).**
- **Third step :the Width of the interval (w).**

The number of intervals (k)

A small number of intervals are not good because information will be lost.

A large number of intervals are not helpful to summarize the data.

A commonly followed rule is that $5 \leq k \leq 20$ or the following formula may be used, $k=1+3.322 (\log n)$.

We select 5 intervals in our example.

The range (R)

It is the difference between the maximum and the minimum observation (entries) in the data set.

R = the maximum entry - the minimum entry

R =96-50

=46

Example (2):

the following data represent the marks of 50 students in a course:

51	95	70	74	73	90	71	74	90	67
91	72	83	89	50	80	72	84	85	69
62	82	87	76	91	76	87	75	78	79
71	96	81	88	64	82	73	57	86	70
80	81	75	85	74	90	83	66	77	91

The range (R)

It is the difference between the maximum and the minimum observation(entries) in the data set.

$$R = X_{max} - X_{min}$$

$$R = 96 - 50$$

$$= 46$$

The Width of the interval (w)

Class intervals generally should be of the same width. Thus, if we want k intervals, then w is chosen such that w $\geq$ R/k.

$$w = \frac{46}{5} = 9.4 \approx 10$$

Forth step:

Choose the minimum observation to be the lower limit of the first interval and add the width of interval to get the lower limit of the second interval and so on

the lower limit of the second interval

50+10=60

the lower limit of the third interval

60+10=70

the lower limit of the fourth interval

70+10=80

the lower limit of the fifth interval

80+10=90

Fifth step:

To find the upper limit of any interval add the following to the lower limit of interval :

W-1=10-1

=9

the upper limit of first interval

50+9=59

the upper limit of second interval

60+9=69

the upper limit of third interval

70+9=79

the upper limit of fourth interval

80+9=89

the upper limit of fifth interval

90+9=99

Class interval	tally	frequency
50-59		
60-69		
70-79		
80-89		
90-99		
Total		

Example (2):

the following data represent the marks of 50 students in a course:

51	95	70	74	73	90	71	74	90	67
91	72	83	89	50	80	72	84	85	69
62	82	87	76	91	76	87	75	78	79
71	96	81	88	64	82	73	57	86	70
80	81	75	85	74	90	83	66	77	91

Class interval	tally	frequency
50-59	| | |	3
60-69		
70-79		
80-89		
90-99		
Total		

Example (2):

the following data represent the marks of 50 students in a course:

51	95	70	74	73	90	71	74	90	67
91	72	83	89	50	80	72	84	85	69
62	82	87	76	91	76	87	75	78	79
71	96	81	88	64	82	73	57	86	70
80	81	75	85	74	90	83	66	77	91

Class interval	tally	frequency
50-59	𝍸	3
60-69	𝍸	5
70-79	𝍸 𝍸 𝍸	18
80-89	𝍸 𝍸 𝍸	16
90-99	𝍸	8
Total		50

Table(2): frequency table of students' marks

Class interval	frequency
50-59	3
60-69	5
70-79	18
80-89	16
90-99	8
Total	50

Table(2): frequency table of students' marks

Class interval	50-59	60-69	70-79	80-89	90-99	**Total**
frequency	3	5	18	16	8	50

Definition of the true class intervals

the true lower unit = the lower limit - 0.5

the true upper unit = the upper limit + 0.5

Class interval	True class interval	frequency
50-59	**49.5-59.5**	3
60-69	**59.5-69.5**	5
70-79	**69.5-79.5**	18
80-89	**79.5-89.5**	16
90-99	**89.5-99.5**	8
Total		50

Definition of the Mid-interval (Midpoints)

the Mid-interval(Midpoints)

=(the lower limit+ the upper limit)/2

$$\frac{69.5+59.5}{2}$$

$$\frac{59.5+49.5}{2}$$

True class interval	Midpoints	frequency
49.5-59.5	**54.5**	3
59.5-69.5	**64.5**	5
69.5-79.5	**74.5**	18
79.5-89.5	**84.5**	16
89.5-99.5	**94.5**	8
Total		50

Definition of the relative frequency

the relative frequency of interval=the frequency of interval/the sum of frequencies (n)

the relative frequency table

Class interval	frequency	the relative frequency
50-59	3	**0.06**
60-69	5	**0.10**
70-79	18	**0.36**
80-89	16	**0.32**
90-99	8	**0.16**
Total	50	**1**

$\frac{3}{50}$

$\frac{5}{50}$

Definition of the percentage frequency

the percentage frequency = the relative frequency $\times$ 100

the percentage frequency table

Class interval	the relative frequency	the percentage frequency
50-59	0.06	**6**
60-69	0.10	**10**
70-79	0.36	**36**
80-89	0.32	**32**
90-99	0.16	**16**
Total	1	**100**

0.06×100

0.10×100

Class interval	True class interval	Midpoints	frequency	the relative frequency	the percentage frequency
50-59	49.5-59.5	54.5	3	0.06	6
60-69	59.5-69.5	64.5	5	0.10	10
70-79	69.5-79.5	74.5	18	0.36	36
80-89	79.5-89.5	84.5	16	0.32	32
90-99	89.5-99.5	94.5	8	0.16	16
Total			50	1	100

Example

Class interval	frequency
16-20	100
21-25	122
26-30	900
31-35	207
36-40	795
41-45	568
46-50	322

Find from the tabe:

- **The Width of the interval**
- **The midpoints**
- **True class intervals**
- **The relative frequency of intervals.**
- **The percentage frequency of intervals.**

In these lecture we create:

- frequency table
- the percentage frequency table
- the relative frequency table

Summary of lecture