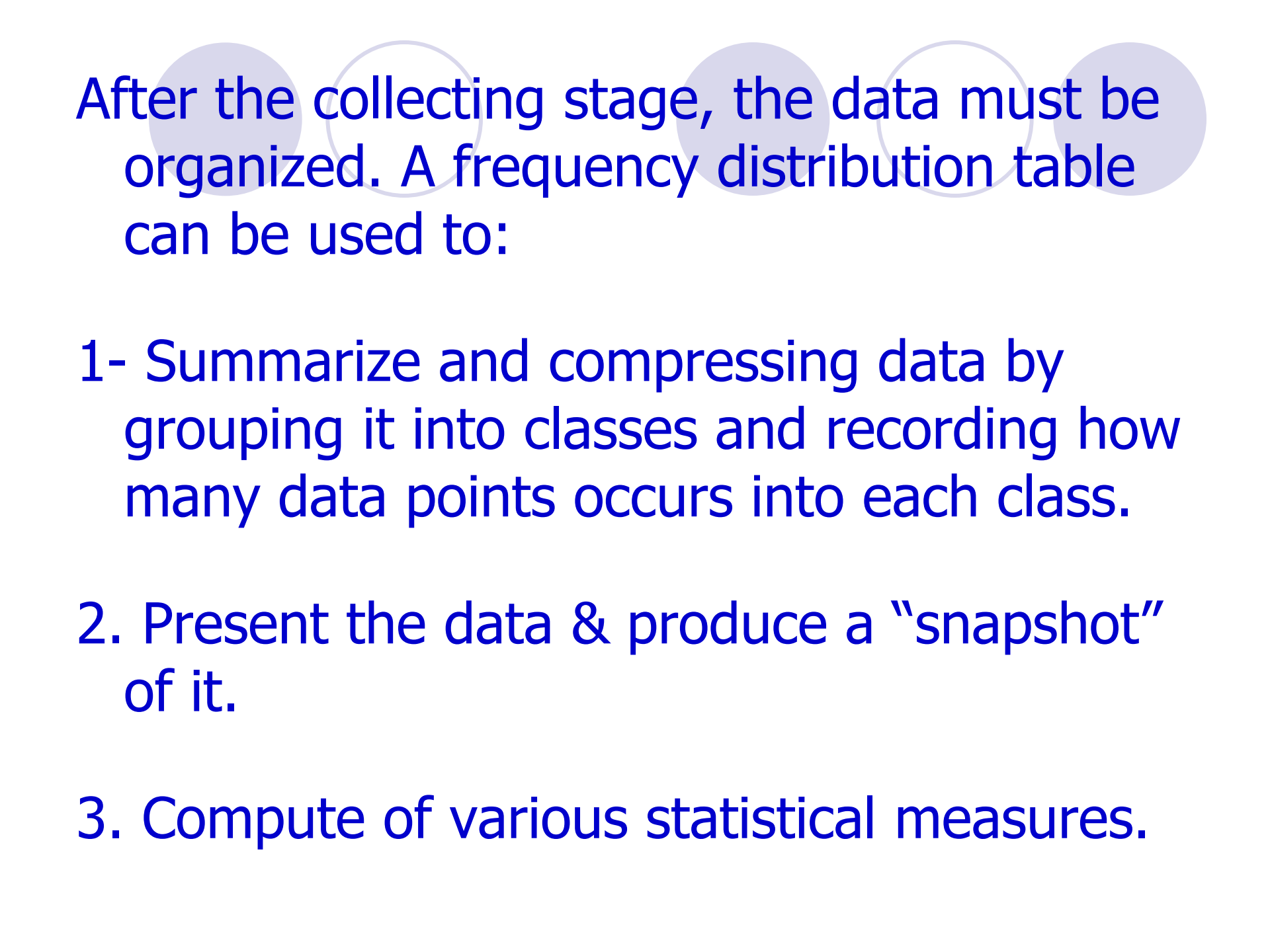




# Biostatistics

## Lecture 2

**Frequency Distribution  
OR  
Organizing The Data**



After the collecting stage, the data must be organized. A frequency distribution table can be used to:

- 1- Summarize and compressing data by grouping it into classes and recording how many data points occurs into each class.
2. Present the data & produce a “snapshot” of it.
3. Compute of various statistical measures.

# There are two types of frequency distributions:

## 1- Ungrouped or simple frequency dist.

used for:

- a- Qualitative variables (nominal and ordinal).
- b- Discrete quantitative v. with a few values.

## 2- Grouped frequency distributions

used for:

- a- Continuous quantitative variables.
- b- Discrete quantitative variables with large number of values (more than 20).

# 1. Ungrouped (simple) frequency dist.

Example: The following data represent the number of children of 16 Iraqi women:

3, 5, 2, 4, 0, 1, 3, 5, 2, 3, 2, 3, 3, 2, 4, 1

Solution: Use the following steps to present this data in a frequency distribution table:

1- Variable: In this data the type of variable is discrete quantitative variable with a few different values ( $n=16$ ), so we must arrange in ungrouped frequency distribution.

2- Divide the data into intervals *or types of elements* (X), the possible values of the variable are: 0, 1, 2, 3, 4, 5, and then count the number of data in each interval *or number of each element* (F).

3- Create a table with four columns.

- In the first column you would put the intervals of the data.
- The next column is the frequency column (No. of women for each interval).
- The third column is the relative frequency column, find out by:

$$R.F. = f/n \dots\dots\dots (n=\text{sample size})$$

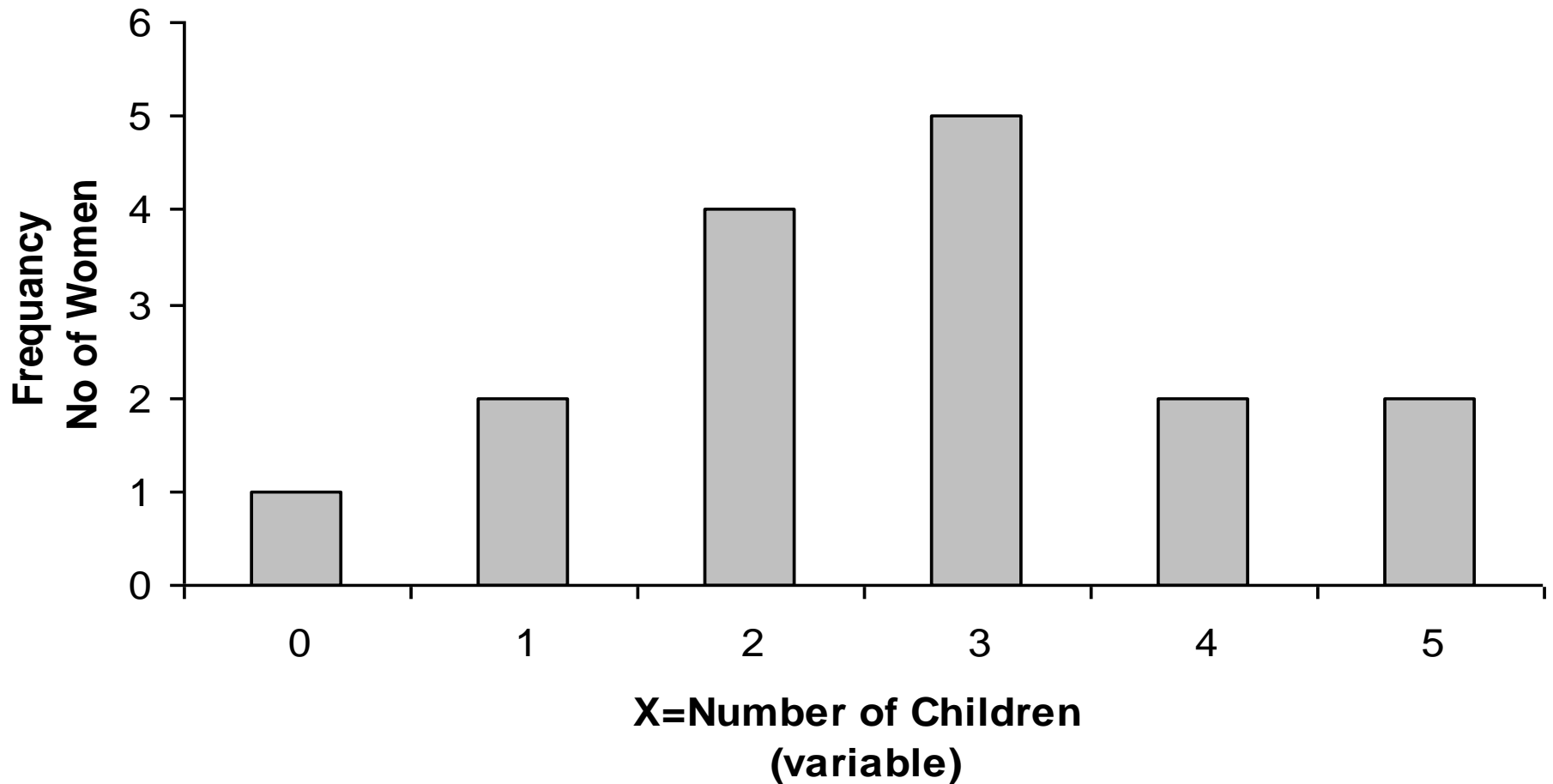
- The final column is the percentage frequency column.

$$\text{R.F. \%} = \text{R.F.} \times 100$$

Simple frequency distribution of the no. of children

No. of children (X)	No. of women (f)	Relative Freq. (= f/n)	Percentage Freq. (= R.F. $\times$ 100)
0	1	0.0625	6.25%
1	2	0.125	12.5%
2	4	0.25	25%
3	5	0.312	31.25%
4	2	0.125	12.5%
5	2	0.125	12.5%
Total	n=16	1.00	100%

**Frequency bar chart is a graphical representation for the simple frequency distribution**



## 2. Grouped frequency distribution

**Example:** The following are the hemoglobin values (g/dl) of 30 children receiving treatment for hemolytic anemia

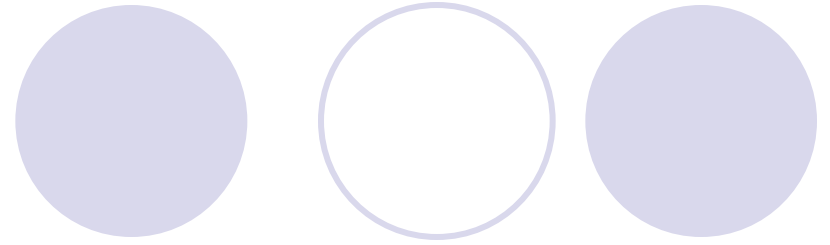
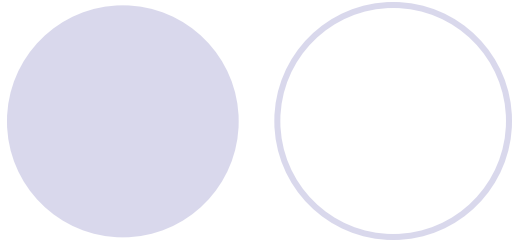
10.0	8.7	6.7	7.8	8.9	10.8
9.7	9.9	8.5	7.5	9.0	10.0
9.1	9.1	8.4	10.6	10.2	8.5
8.6	9.7	9.7	9.6	10.2	11.4
12.2	9.4	9.3	8.4	8.2	9.2





**Solution: Use the following steps to present this data in a frequency distribution table:**

**1- Variable:** In this data the type of variable is continuous quantitative with large number of different values ( $n=30$ ), so we must arrange in grouped frequency distribution.



**2- Classes: To summarize data into classes, the approximate number of classes can be determined by this formula:**

$$\text{No. of classes} = 2.5\left(\sqrt[4]{n}\right) \approx 6$$



### 3- Class interval:

is the different between the lowest and highest possible values in each class. And to compute it finds the different between the highest and lowest value of the data, and divide on number of classes.

$$\text{Class interval} = \frac{\text{highestvalue} - \text{lowestvalue}}{\text{No.of classes}}$$

$$= \frac{12.2 - 6.7}{6} \approx 1$$

#### **4- To arrange the data into classes as follow:**

- First class lower value = lowest values of the data..... (Approximately)**

**Upper value = first lower value + Class interval**

- Second class lower value = Upper value for first class**

**Upper value = lower value + Class interval  
..... (etc)**

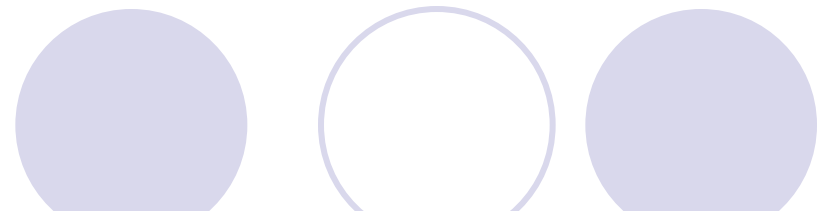
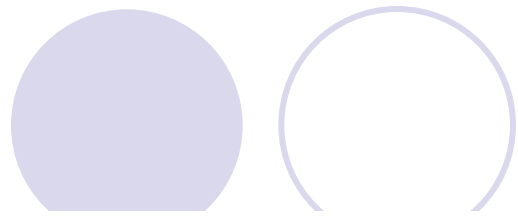
**5- Midpoint (mi): is the number of each center class interval.**

**$mi = \text{lower value} + \text{upper value}/2$ .....(For each class interval)**

**6- Cumulative frequency: Here we add the cumulative frequency of the previous row to the frequency of the current row. The cumulative frequency is the same as the frequency.**

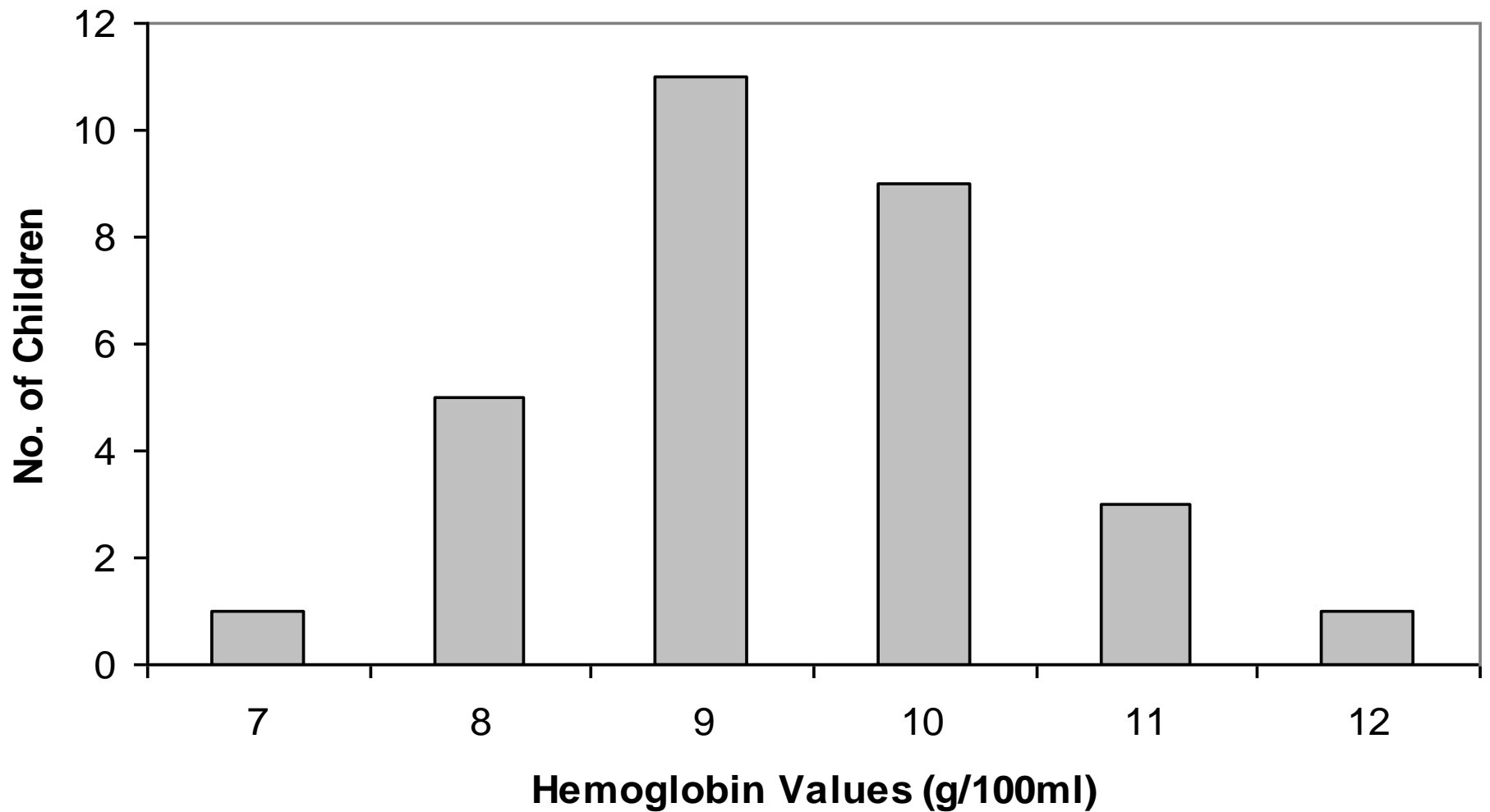
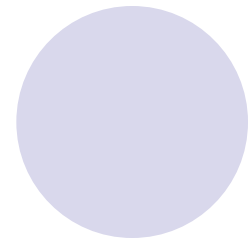
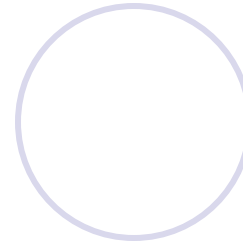
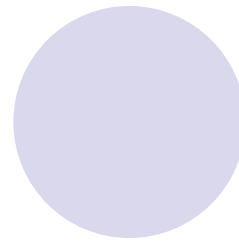
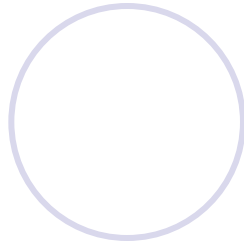
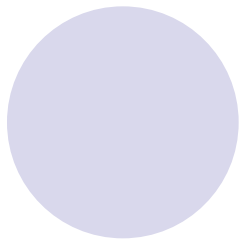
**7- Relative frequency: Is the indicator percentage of frequency.**

**Relative frequency=  $f/n$  ..... (n=sample size)**



Grouped frequency distribution for the hemoglobin level of the 50 men

Class Interval (Hemoglobin)	Midpoint (mi)	Frequency (f)	Cumulative Frequency	Relative Frequency
6.5 – 7.5	7	1	1	0.033
7.5 – 8.5	8	5	6	0.167
8.5 – 9.5	9	11	17	0.367
9.5 – 10.5	10	9	26	0.300
10.5 – 11.5	11	3	29	0.100
11.5 – 12.5	12	1	30	0.033
Total	---	n=30	---	1.000



# Exercise 2:

Q1: The following are the blood groups of a sample of patients who attend clinic A.

A, B, O, AB, B, A, O, O, AB, B  
B, B, A, O, O, AB, B, O, B, A  
AB, A, O, A, A, B, O, A, A, B

(a) Identify the variable, type of variable, sample and the population in the study?

variable: blood groups

type of variable Qualitative & ordinal

Population: patients

Sample: patients who attend clinic A



The following are the blood groups of a sample of patients who attend clinic A.

A, B, O, AB, B, A, O, O, AB, B  
B, B, A, O, O, AB, B, O, B, A  
AB, A, O, A, A, B, O, A, A, B

(b) What is the sample size?

The sample size is 30

(c) Draw a frequency table for the data?

Blood Groups	Frequency
A	9
B	9
AB	4
O	8
Total	30

**Q2: In a study, the blood glucose levels of a sample of students of school A were measured.**

103	125	120	118	117	118	109	114	118	131
116	119	117	119	110	113	117	124	117	124
127	127	114	129	120	126	105	121	112	115
101	114	128	125	109	123	122	123	130	115

**Q2: In a study, the blood glucose levels of a sample of students of school A were measured.**


**(a) Identify the variable, type of variable, sample and population in this study.**

variable: blood glucose levels

type of variable: Quantitative & continuous

Population: schools

Sample: students of school A



**Q2: In a study, the blood glucose levels of a sample of students of school A were measured.**

**(b) Make a frequency table for the data, using the class intervals**

**100-104, 105-109, 110-114, ... , 130-134.**

**(c) Obtain the class mid points, relative frequency and cumulative frequency of the frequency distribution in part (a).**



Class Interval	Mid Point (x)	Freq. (f)	C.F.	R.F.
100-104	102	2	2	0.050
105-109	107	3	5	0.075
110-114	112	6	11	0.150
115-119	117	12	23	0.300
120-124	122	8	31	0.200
125-129	127	7	38	0.175
130-134	132	2	40	0.050
Total	-	40	-	1.000

**Q3: The following table gives the distribution of the ages of 75 patients who attend clinic A.**

Age (years)	Frequency	R.F.
05 - 14	6	0.08
15 - 24	9	-
25 - 34	-	0.24
35 - 44	24	0.32
45 - 54	15	-
55 - 64	-	-

**Complete the blank cells in the table?**

Q3:

Age (year)	Frequency	R.F.
5-15	6	0.08
15-24	9	0.12
25-34	18	0.24
35-44	24	0.32
45-54	15	0.20
55-64	3	0.04
Total	75	1.000

$$9/75=0.12$$

$$0.24*75=18$$

$$15/75=0.20$$

$$1-(0.08+0.12+0.24+0.32+0.20)=0.04$$

$$75-(6+9+18+24+15)=3$$