بسم الله الرحمن الرحيم

السلام عليكم ورحمة الله وبركاته

Biostatistics

L II

14th -7-2025

PROF. DR. WAQAR AL-KUBAISY

Descriptive Statistics

organizing and summarizing data *and* ***bringing into a focus their essential characteristics***

Descriptive statis.

reduce the information to a manageable size

Presentation of Data

Data that collected from any source, are inadequate for planning .

Data need to be transformed into information

- **by reducing them,**
- **by summarization and**
- **Arrange it in a simple and useful way to**
- **bring out the *important point clearly &* واضح وموجز *concise***

This mean that

display the important feature of the sample .

This is Descriptive Statistics

Descriptive Statistics

This one serve as devices for
organizing **and** **summarizing** **data**
and
bringing into a focus their essential characteristics so;
Descriptive statistics .
reduce the information to a manageable size.

This include:

- **Table**
- **Graph, Chart** or
- **Numerical Description**

An important thing is the type of the variable concerned.

Table

It is first step in data presentation .

Is the simplest and often most useful summary of data

An important thing is the type of the variable concerned.

Nominal :Simple Frequency Table

example

Blood group of 95 children with leukemia shows as (22)A, (25)B, (18)AB, & O(30)

count the No. of observation in each category, these count are called Frequencies .

- frequency
- Relative frequency
- percentage frequency

} **of Bl. group distribution for of 95 children with leukemia**

An important thing is the type of the variable concerned.

Simple Frequency Table

Blood group	Frequency N=95	Relative Frequency	Percentage %
A	22	???	??
B	25		
AB	18		
O	30		
Total	95		

Nominal Simple Frequency Table continuo..

Relative frequency

Frequency each category divided by the total frequency.
No. of children of each category (BG) divided by the total no of children.

Percentage frequency

Frequency of each category divided by the total frequency X 100

Relative or percentage frequency are often more useful than the actual number of individuals in each category.
???????

Simple Frequency Table

Blood group	Frequency N=95	Relative Frequency	Percentage %
A	22	22/95=0.231	23.15
B	25	0.26315	26.315
AB	18	0.18947	18.947
O	30	0.3157	31.5789
Total	95	?????	????

Relative or percentage frequency are often more useful than the actual number of individuals in each category. Why ????????

Type of feeding

Infants 600

Breast 478

Bottle 65

Mixed 57

Simple Frequency Table

Type of feeding	No. of cases (F)	R.F.	percentage %
Breast	478	0.79	79. 7
Bottle	65	0.108	10.8
Mixed	57	0.095	9.8
Total	600	????	100

b- Ordinal Variable

120 individuals were asked about their level of satisfaction toward the health care given by Hospital X. The response as follows

29 very satisfied, 39 satisfied, 20 neutral

18 unsatisfied, 14 highly unsatisfied

level of satisfaction	Frequency N=120	Relative Frequency	Percentage %
very satisfied	29	0.24166??	24.166
satisfied	39	0.325 ??	32.5
neutral	20	0.1666	16.66
unsatisfied	18	0.15	15
highly unsatisfied	14	0.11666	11.66
Total	120	????	99.929

Ex.

The mathematic marks of 26 secondary school students at Amman in 2025

15.2 31.3 14.9 16.3 19.3 18.2 20.2 12.8 14.7
29.4 21.1 20.4 13.6 22.4 14.0 14.3 22.8 26.7
18.9 13.7 17.7 27.2 19.3 16.1 13.5 11.2

? ? ? ?

Metric variable

Continuous variable

? ? ? ?

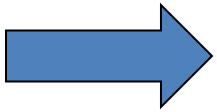
Continuous Metric variable(CMV)

The most useful way for presenting data of CMV to produce grouped frequency distribution

❖ **grouping data first These group of data we call it class interval**

✓ **Each group of data (class interval)consist of values within certain range**

mathematic marks	Frequency	Cumulative frequency
10.0- 14.9	9	9
15.0- 19.9	8	17
20.0- 24.9	5	22
25.0- 29.9	3	25
30.0- 34.9	1	26
	26	



Continuous Metric variable

to produce grouped frequency distribution table

- Grouping data into groups of equal width
- then construct frequency distribution table for grouped data
- Counting the frequency of observation within the groups(class interval)
- Each group of data contain No. of observation

mathematic marks	Frequency	Cumulative frequency
10.0- 14.9	9	9
15.0- 19.9	8	17
20.0- 24.9	5	22
25.0- 29.9	3	25
30.0- 34.9	1	26
	26	

Use sturges rule :

$$K = 1 + 3.322 \ (\log \ N)$$

The mathematic marks of 26 secondary school students at Amman in 2025

15.2 31.3 14.9 16.3 19.3 18.2 20.2
12.8 14.7 29.4 21.1 20.4
13.6 22.4 14.0 14.3 22.8 26.7
18.9 13.7 17.7 27.2 19.3 16.1
13.5 11.2

K= No. of class intervals.

N= sample size .

Width of class intervals :

$$W = \frac{R}{K}$$

W= width .

R= Range = highest – lowest .

K= No. of class intervals .

Continuous Metric variable

Frequency distribution of mathematic marks of 26 secondary school students at Amman in 2025

mathematic marks	Frequency	Cumulative frequency
10.0- 14.9	9	9
15.0- 19.9	8	17
20.0- 24.9	5	22
25.0- 29.9	3	25
30.0- 34.9	1	26
Total	26	? ? ? ?

frequency distribution table

Example

The following data representing age (years) of 50 patients with diabetes Mellitus collected from Al Karak Hospital during march 2024

68, 62, 62, 66, 68, 65, 64, 71,77, 74, 20, 33, 38. 42, 47. 50,55, 56, 60 72, 80 74, 75, 74, 77,80, 81, 89, 86, 85, 83,72, 70, 71, 79, 76, 77, 80, 90, 97, 94, 90,65, .60, 67, 63 88, 84, 84, 87

? ? ? ?
????????/

An important thing is the type of the variable concerned.

Age(year) of 50 patients with diabetes Mellitus attending Al Karak Hospital during march 2025

AGE year	Freq	Cumulative frequency	Relative frequency	% R.F.	Cumulative R.F.	%cum Freq.
20-29	1		??	??		
30-39	2		???	??		
40-49	2		??			
50-59	3	?				
60-69	12					
70-79	14	?				
80-89	12					
90-99	4					
total	50	?	?	?	?	?

Relative Frequency (proportion)

Dividing the No. of values (observation, frequency) in a particular class interval by the total No. of values (observation frequency) in whole data

$$\frac{1}{50} \quad \frac{2}{50} \quad \frac{3}{50} \quad \frac{12}{50} \quad \frac{14}{50} \quad \frac{12}{50} \quad \frac{4}{50}$$

Percentage of Frequency

Dividing frequency of each class interval by the total No. of observation and then multiply by 100 .

$$\frac{1}{50}\times 100 \quad \frac{2}{50}\times 100 \quad \frac{3}{50}\times 100 \quad \frac{12}{50}\times 100 \quad \frac{14}{50}\times 100 \quad \frac{12}{50}\times 100 \quad \frac{4}{50}\times 100$$

❑ Cumulative Freq. Dist.

- That is to convert the frequencies distribution into less than and more than .

❖ This is done by simply

- Adding two or more classes frequency
- Starting either at the top or at the bottom of the distribution .

1+2+ +2+3+12+14+12+4= 50

❑ Cumulative Relative and Percentage Dist.

- add two or more Relative frequencies together .

0.02+0.04+0.04+0.06+0.24+0.28+0.24+0.08=1

- Add the % instead of the frequencies, starting either at the top or at the bottom .

AGE year	frequency	Commutative frequency	Relative frequency	% R.F.	Cumulativ R.F.	%cum Freq.
20-29	1	1	0.02	2	0.02	2
30-39	2	3	0.04	4	0.06	6
40-49	2	5	0.04	4	0.1	10
50-59	3	8	0.06	6	0.16	16
60-69	12	20	0.24	24	0.4	40
70-79	14	34	0.28	28	0.68	68
80-89	12	46	0.24	24	0.92	92
90-99	4	50	0.08	8	1.00	100
total	50	---	1	100	---	---

Points should be keep in mind

1-No. of class intervals (5-20) .

2-Classes interval should not overlapping .

3-All classes interval should have the same width across all data (constant width) .

4-There should be no gaps between class interval .

5-Every observation will be uniquely classifiable into one and only one class interval .

❑Class Marks

- ❖ It is the midpoint of the class interval .
- ❖ *It could be obtain by adding the lower and upper limits of a class interval and divided by two then*

Presentation of Data
table
graph, chart or
Numerical Description

Graphical Techniques

- some times table presentation will give some difficulties to the reader, especially to non numerical readers

- Picture speaks lauder than thousand words .

- Graph have powerful impact on the imagination of population .

- Relationships, Trends, and Contrasts are often more
- readily appreciated from diagram than table ..

An important thing is the type of the variable concerned.

types

Nominal and Ordinal Data

Charting

Pie Chart

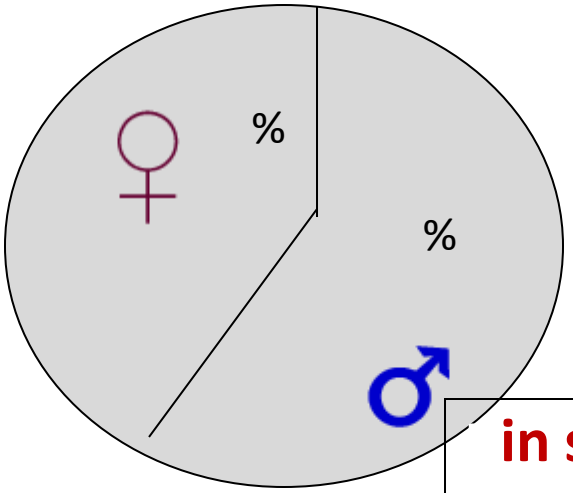
Here the circular is divided into sectors, pie shaped pieces

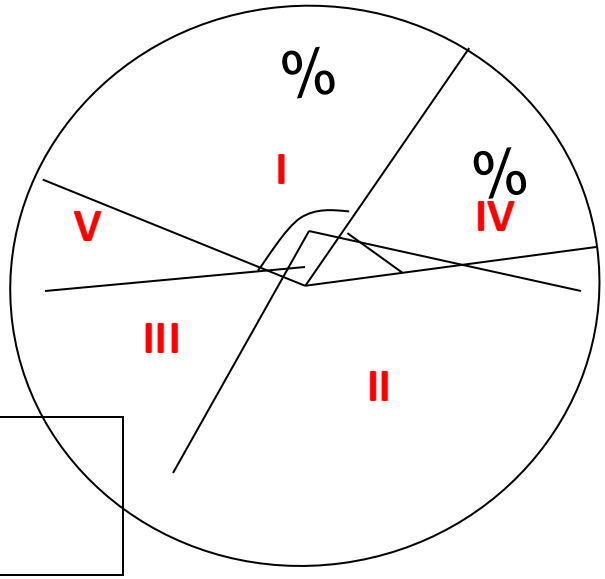
Size of pie proportional to frequency, percentage of that variable.

Disadvantage of pie chart

it can only represented one variable

(sex of children





in showing comparison

Pie Charts

- Displays data in percentages.
- Certain Research Data:

  27 from 1st year, **57.4%**

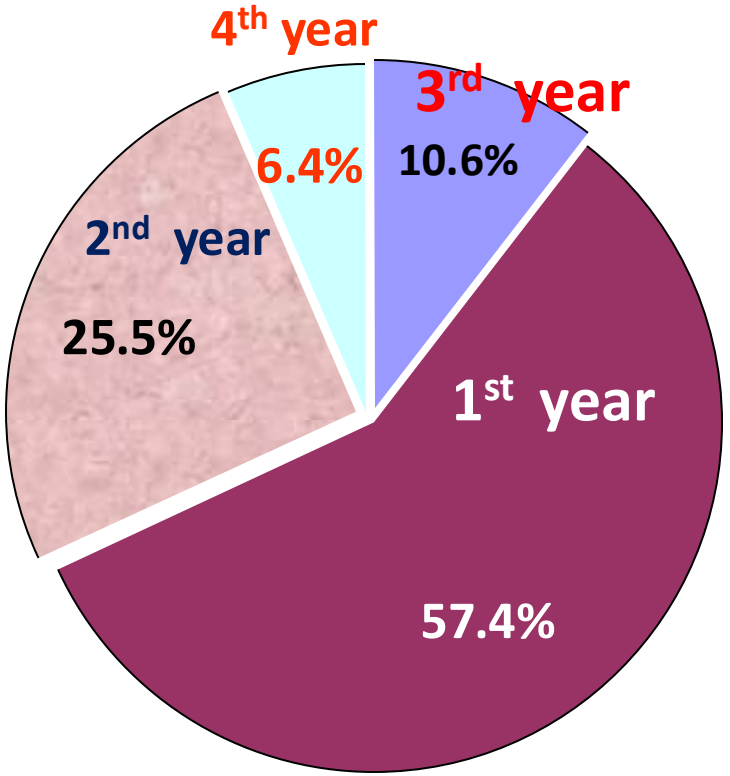
  12 from 2nd year, 25.5%

  5 from 3rd year, 10.6%

  3 from 4th year, 6.4%.

  Should add to 100%, adds to 99.9% due to round-off error

Percentage of (47) medical students in each class level certain research



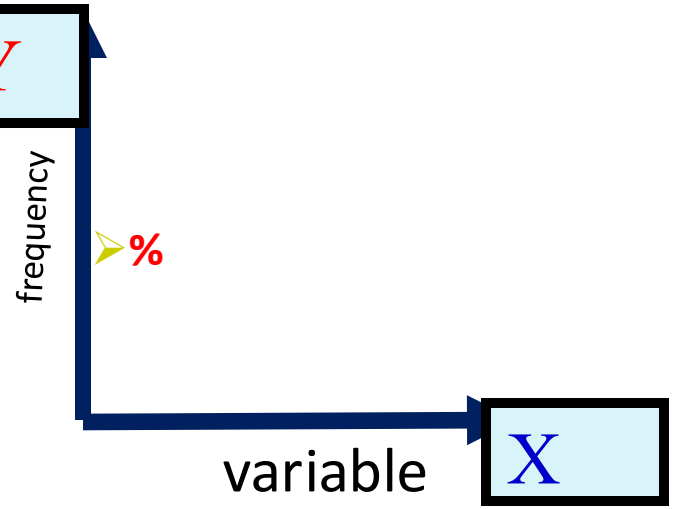
Excellent in showing part vs. whole comparisons

2- THE BAR CHART:

- This type of graph is suitable to represent data of the two subtypes of qualitative, and quantitative discrete type.
- Each category in the table is represented by a bar or column or rectangle,
- So the height of the bar is opposite to the corresponding frequency on the Y axis.
- All bars must have the same width and a space must be left between every two consecutive bars,
- the width of that space is about same or half the width of the bar.

nominal and ordinal data

Bar Chart



Two axis

- **Horizontal, X**
- **plotting the variable .**

- **Vertical, Y**
- **plotting the**
- **frequency, Relative frequency or %**
- **Then draw a Rectangles (bar) .**

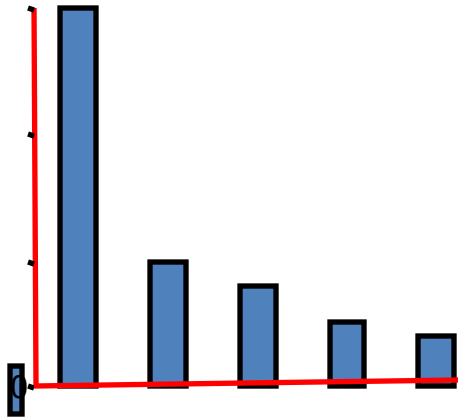
  The length of rectangle (bar) corresponding to the frequency of the variable

Used for

- **frequency or**
- **Relative frequency or**
- **% .**

Charting

nominal and ordinal data



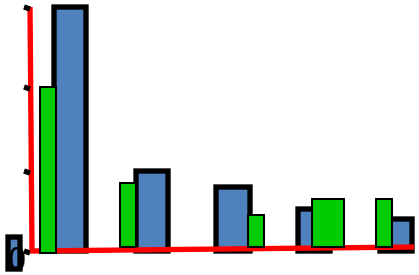
Bar chart

I. Simple bar chart used

-when we have one variable (sex of child)

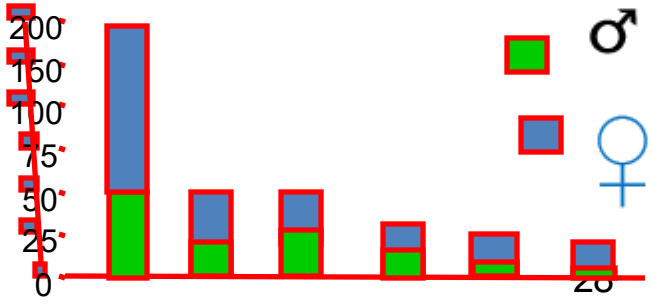
-width of bares should be equal and

-space between bars be the same



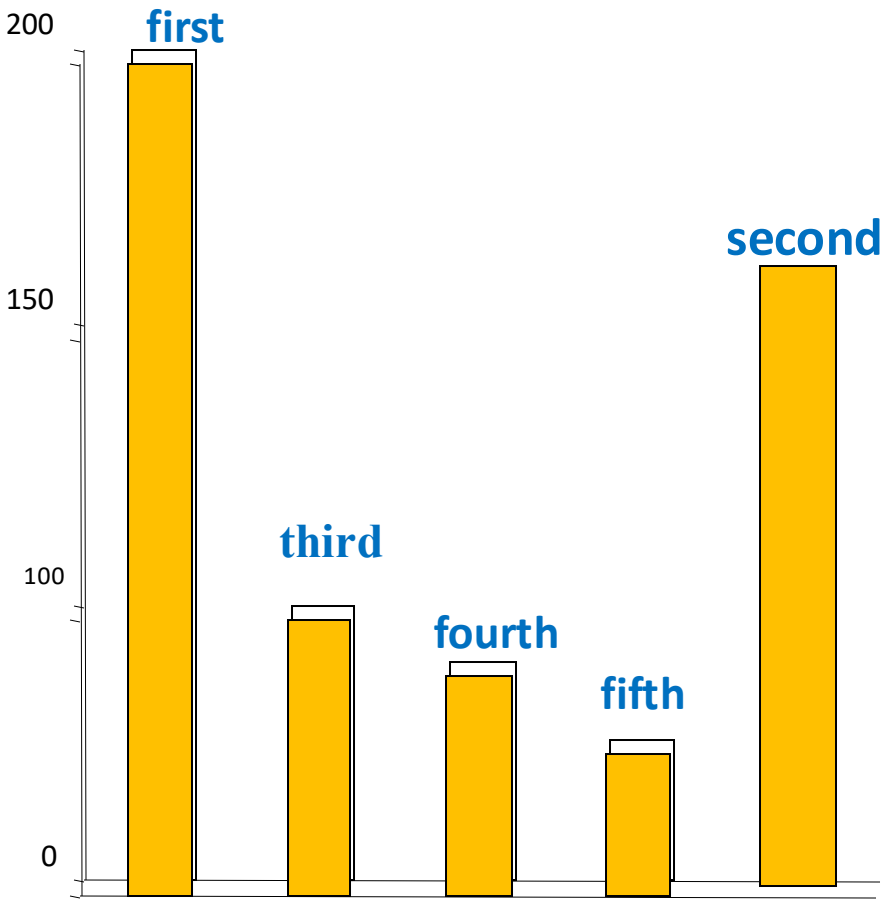
II Clustered bar chart

Used when more than one variable example sex with different class year



III Stacked bar chart

nominal and ordinal data



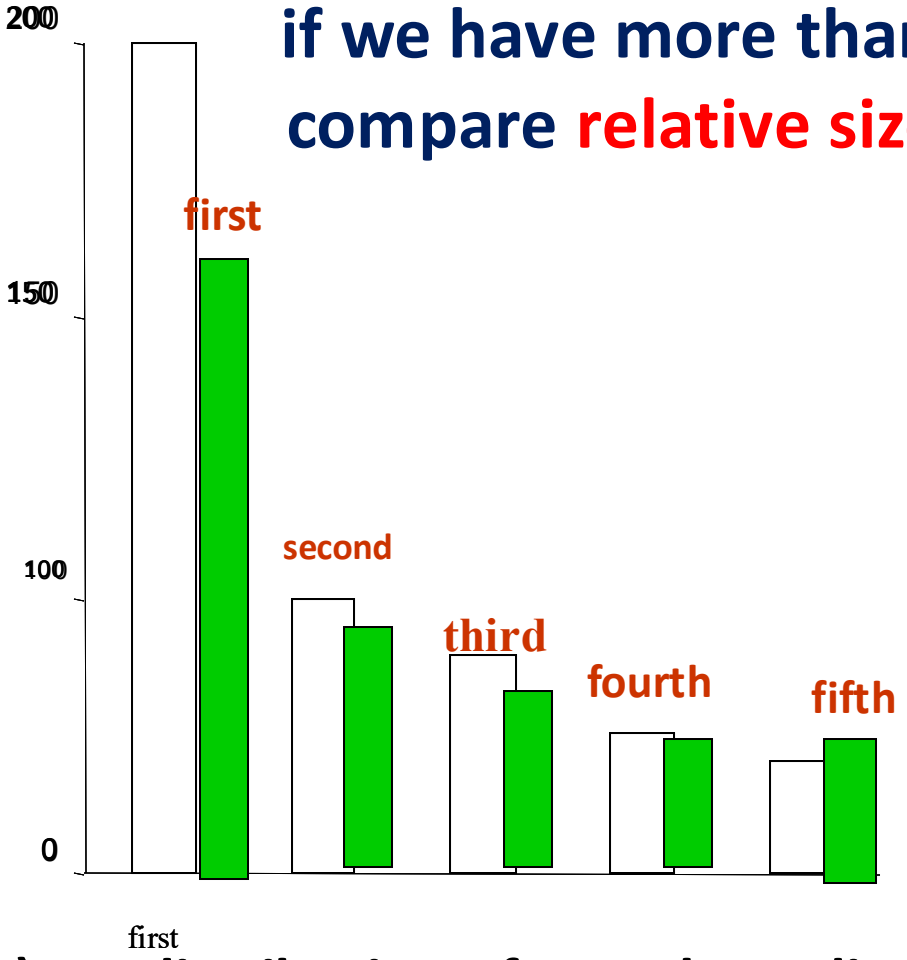
Excellent for showing

Magnitude differences

(I)Mutah medical student according to their year level 2023

nominal and ordinal data

Allows easier comparisons between data sets of different sizes.



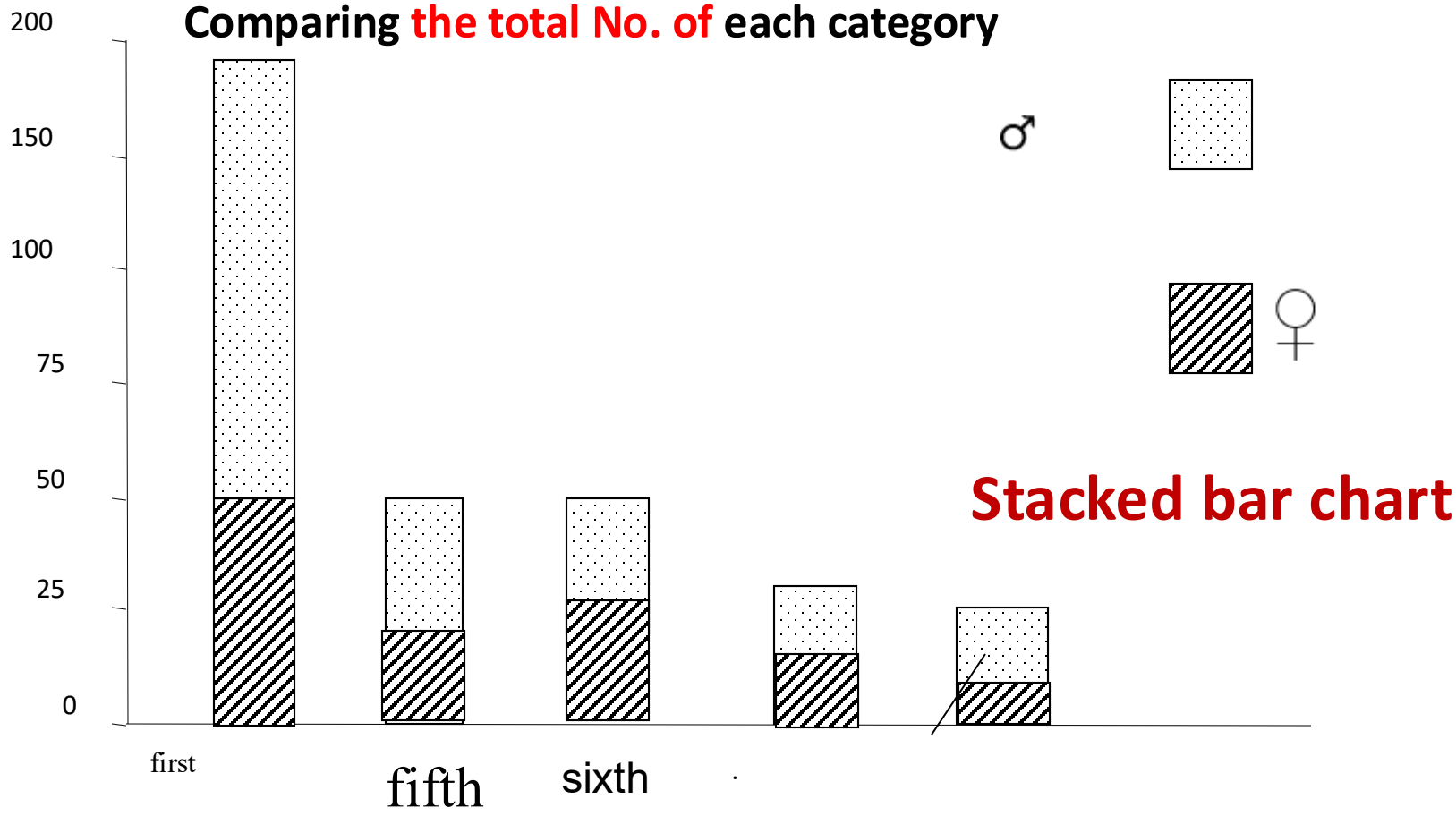
if we have more than one group

compare relative size of each group

♂ ♀

Clustered bar chart

(II)Sex distribution of Mutah medical student according to their year level 2023



Sex distribution of Mutah medical student according to their year level 2023

Charting

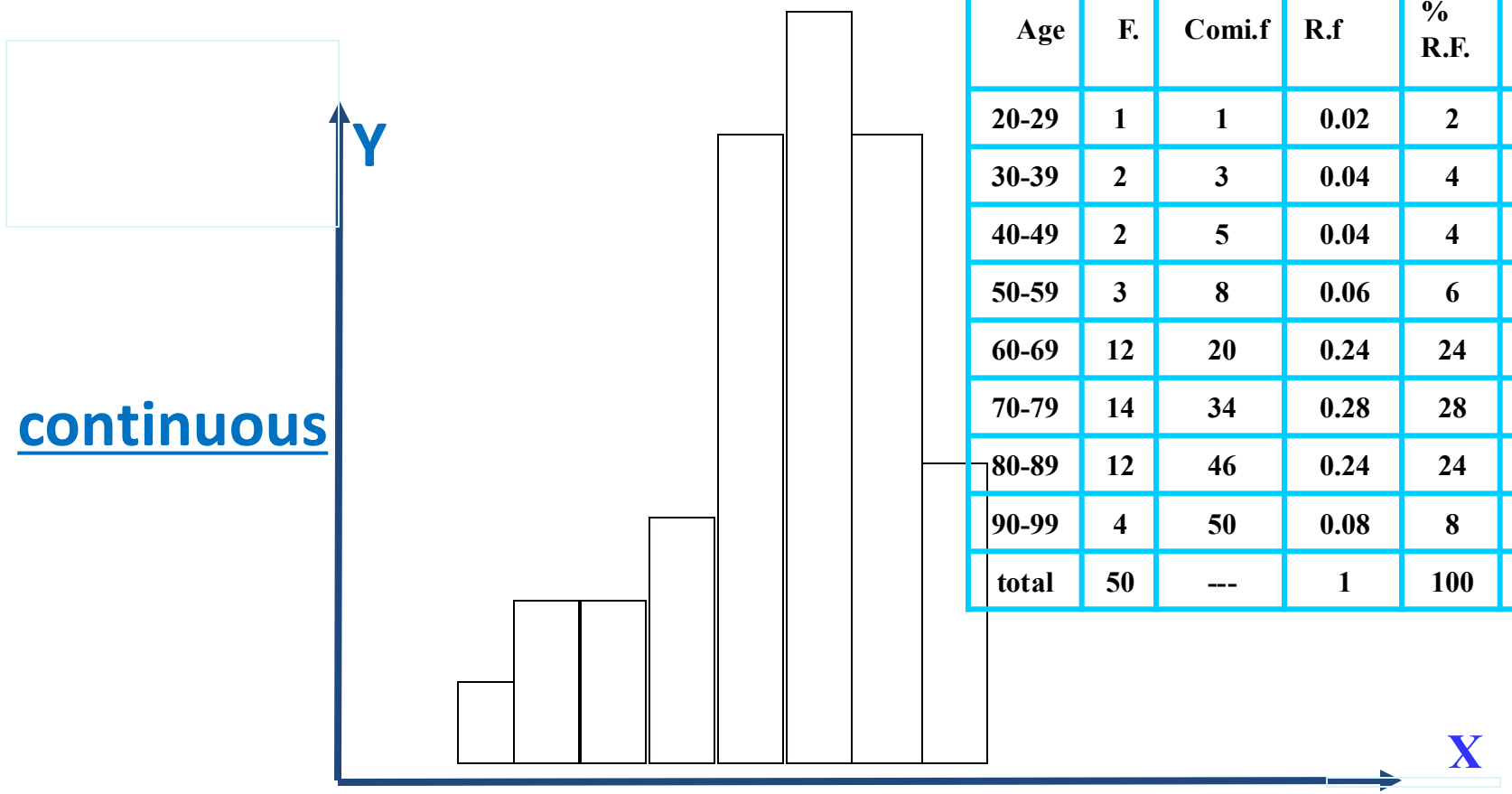
Continuous Metric Variable by Histogram

Age (year)	F.	Cumulative Frequency	Relative frequency	% R.F.	Cumulated R.F.	%cum Freq.
20-29	**1**	1	0.02	**2**	0.02	2
30-39	**2**	3	0.04	**4**	0.06	6
40-49	**2**	5	0.04	**4**	0.1	10
50-59	**3**	8	0.06	**6**	0.16	16
60-69	**12**	20	0.24	**24**	0.4	40
70-79	**14**	34	0.28	**28**	0.68	68
80-89	**12**	46	0.24	**24**	0.92	92
90-99	**4**	50	0.08	**8**	1.00	100
total	**50**	---	1	**100**	---	---

Histogram

The group frequency distribution table usually represented graphically or diagrammatically by histogram .

Age	F.	Comi.f	R.f	% R.F.	Cumi. RF.	% cum F.
20-29	1	1	0.02	2	0.02	2
30-39	2	3	0.04	4	0.06	6
40-49	2	5	0.04	4	0.1	10
50-59	3	8	0.06	6	0.16	16
60-69	12	20	0.24	24	0.4	40
70-79	14	34	0.28	28	0.68	68
80-89	12	46	0.24	24	0.92	92
90-99	4	50	0.08	8	1.00	100
total	50	---	1	100	---	---



IV. Age(year) of 50 patients with diabetes Mellitus attending Al Karak Hospital during march 2025

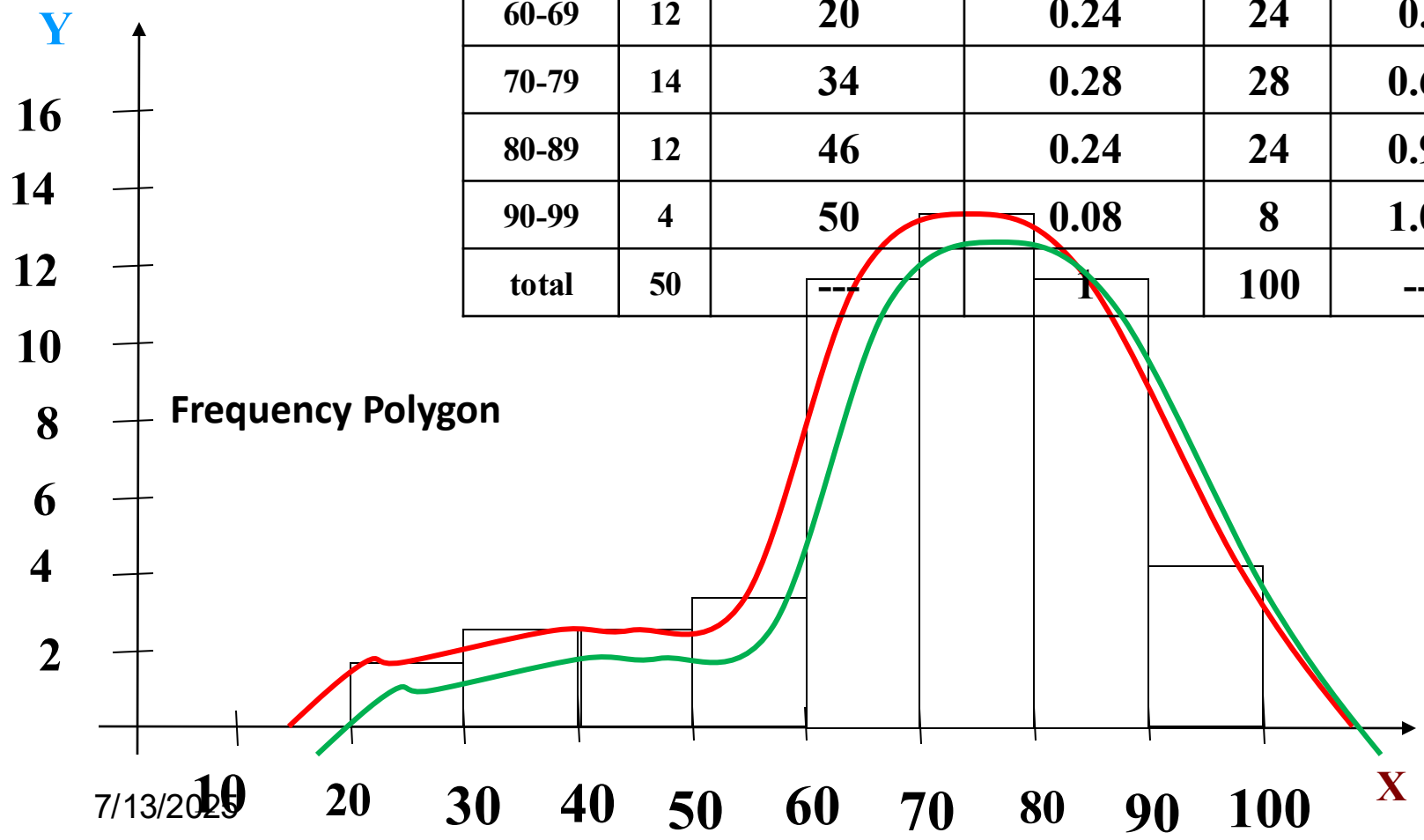
THE FREQUENCY POLYGON:

This type is used when the variable is of **continuous quantitative type and the table is of simple or complex type.**

Each category on the table represented by single point opposite its frequency on Y axis and **the mid-point of the interval on X axis.**

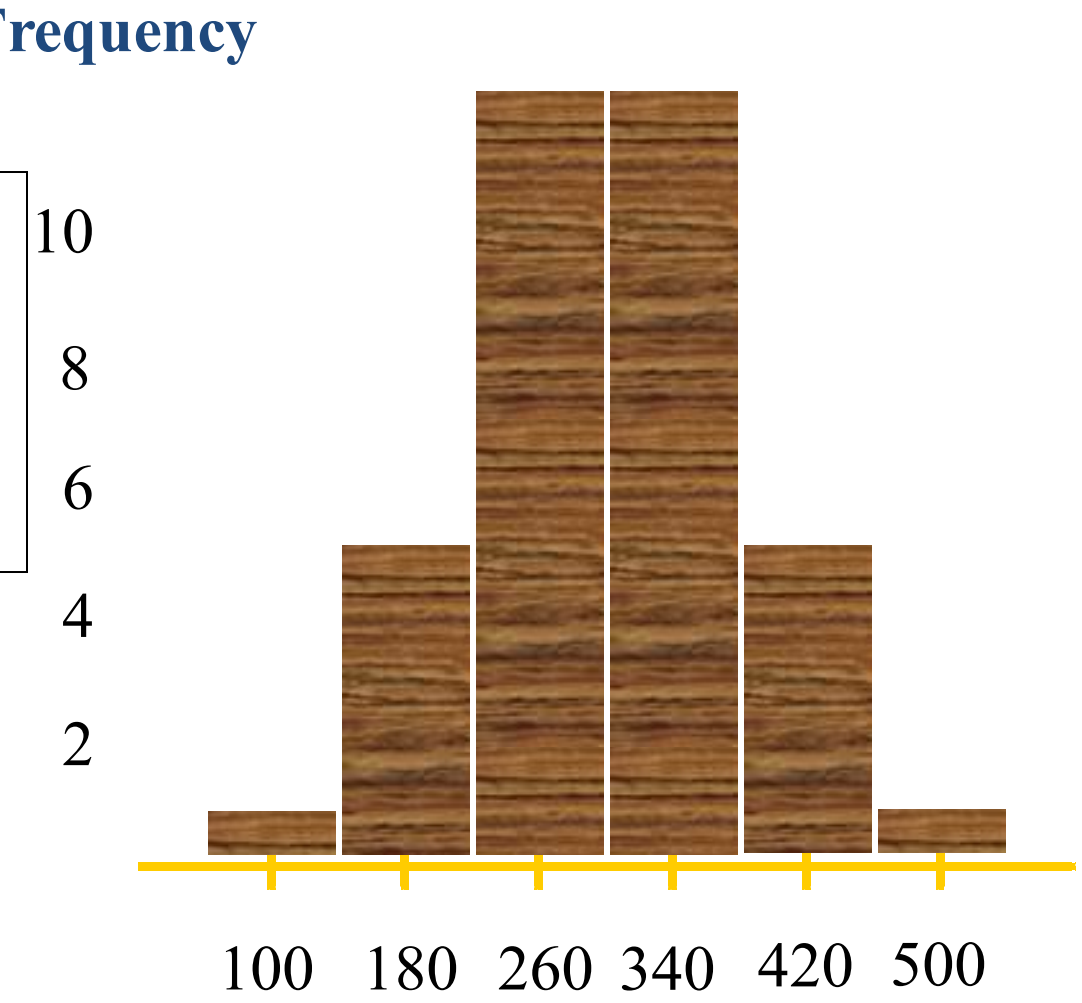
Then every two consecutive points are joined together by a straight line.

Age(year)	Freq	Commu.frequ.	Relat.Freque	% R.F.	Cumul. R.F.	%cum.Freq.
20-29	1	1	0.02	2	0.02	2
30-39	2	3	0.04	4	0.06	6
40-49	2	5	0.04	4	0.1	10
50-59	3	8	0.06	6	0.16	16
60-69	12	20	0.24	24	0.4	40
70-79	14	34	0.28	28	0.68	68
80-89	12	46	0.24	24	0.92	92
90-99	4	50	0.08	8	1.00	100
total	50	---	1	100	---	---



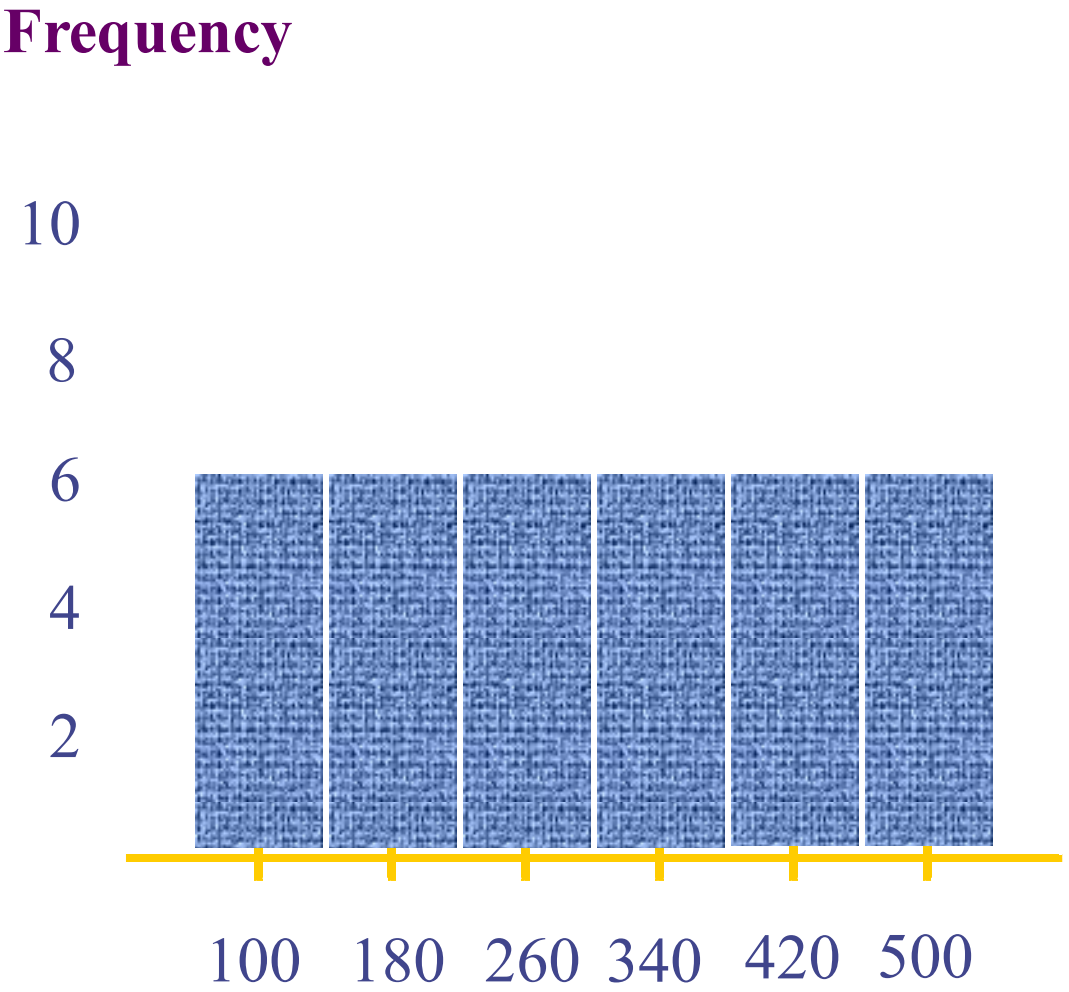
Shapes of Histograms I

Symmetrical, normal, or bell-shaped



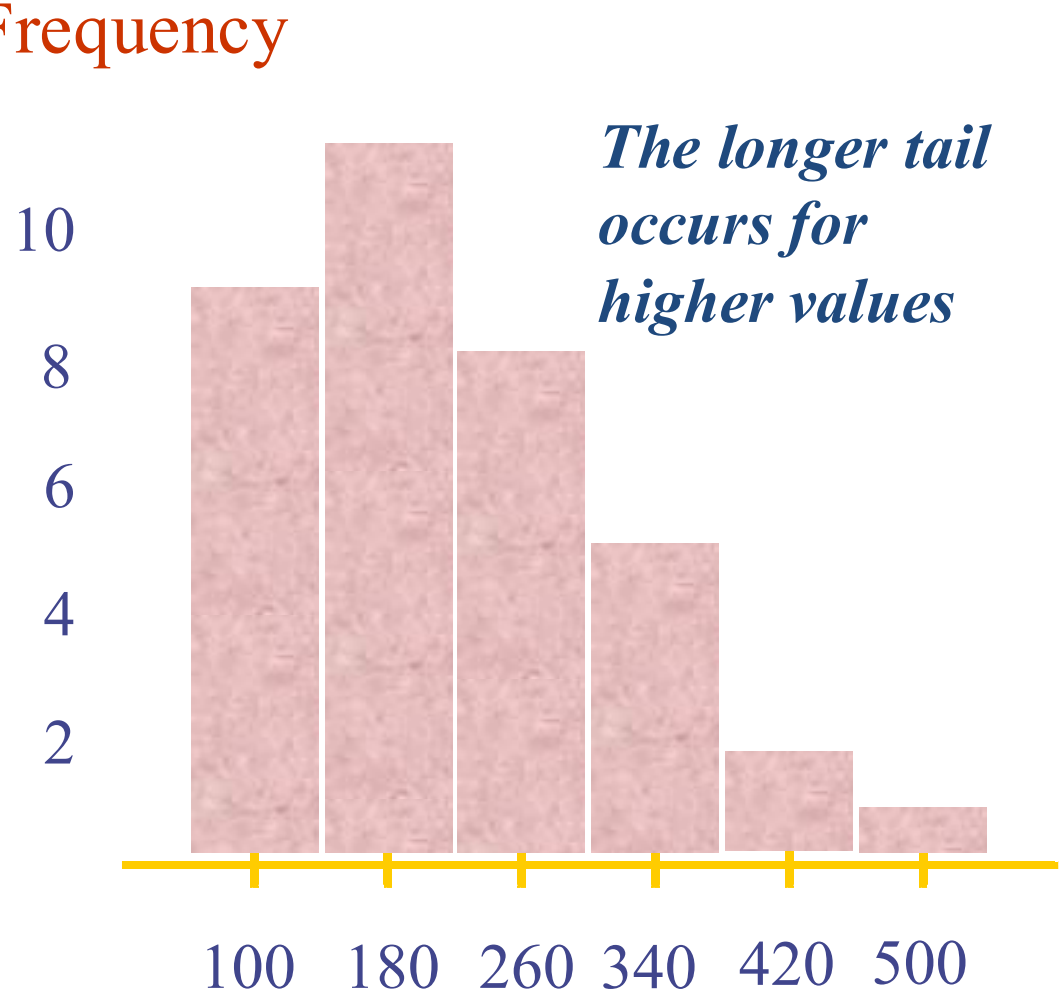
Shapes of Histograms II

Uniform or rectangular



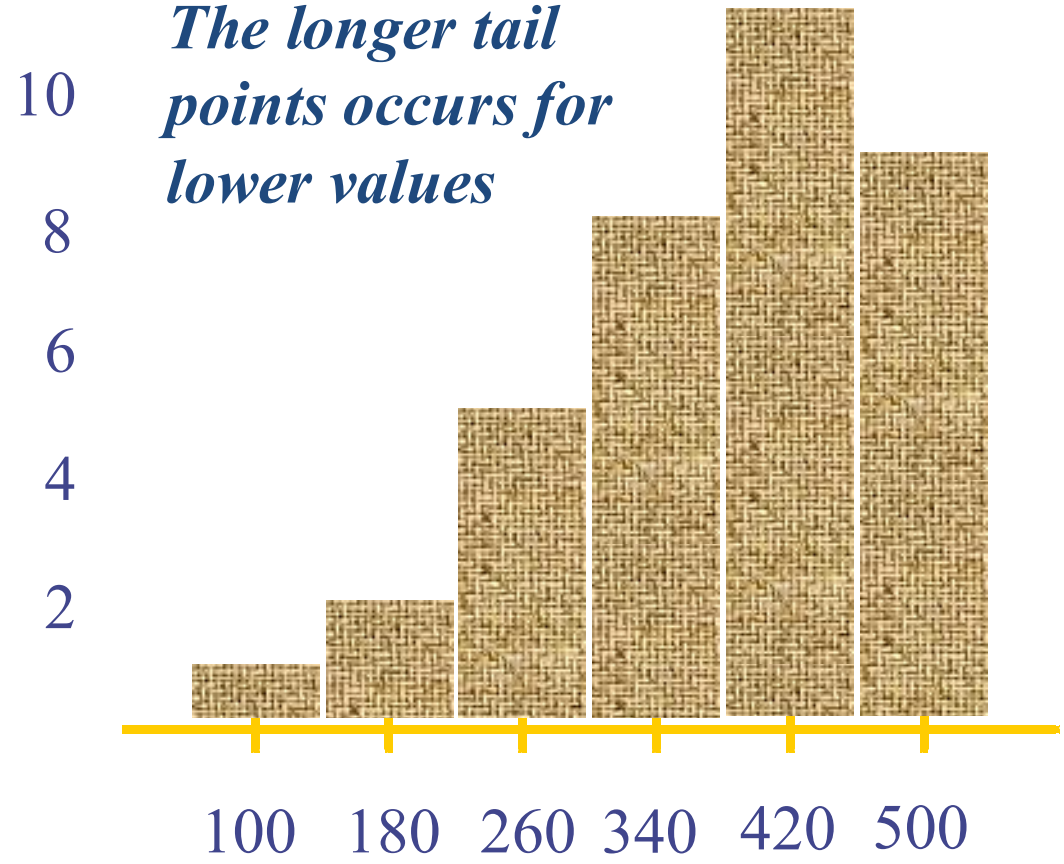
Shapes of Histograms III

Skewed right
or
Positively skewed

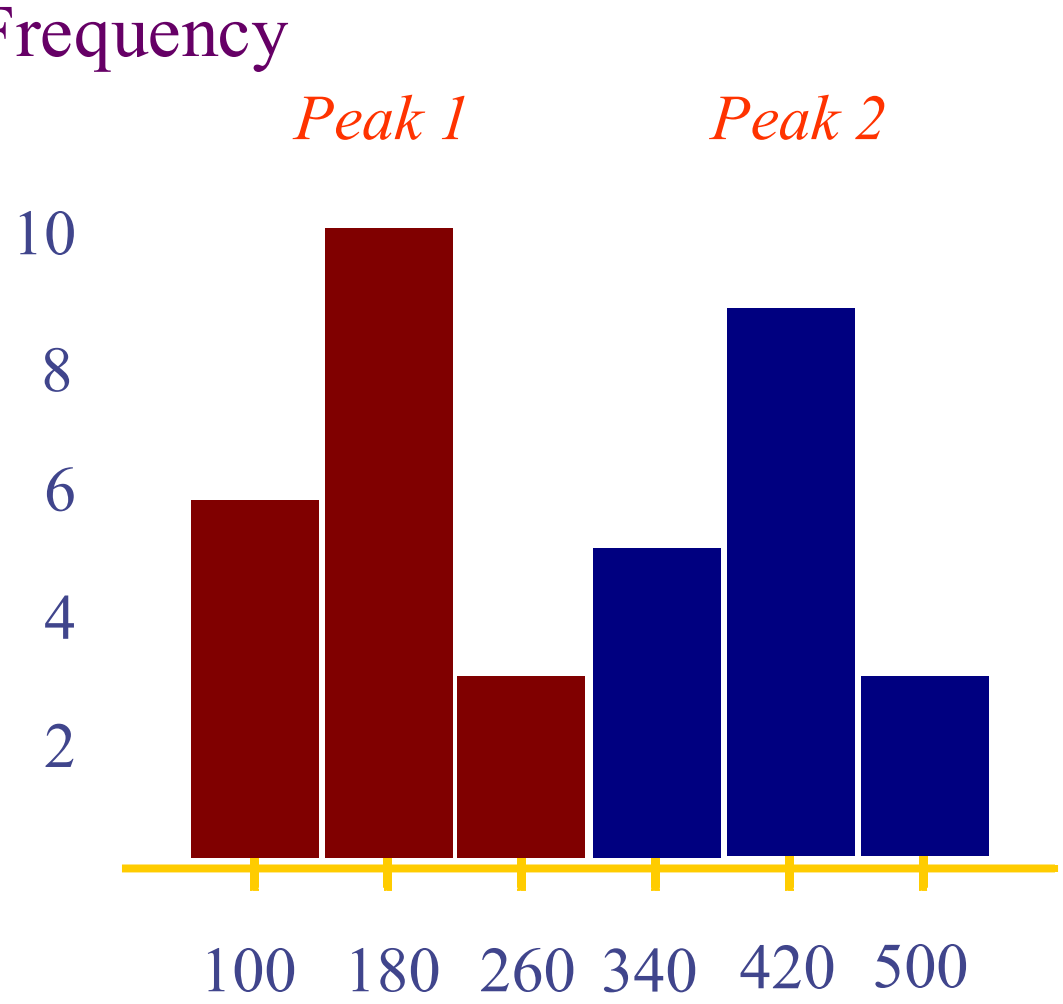


Shapes of Histograms IV

Skewed left
or
Negatively skewed



Shapes of Histograms V



Bimodal

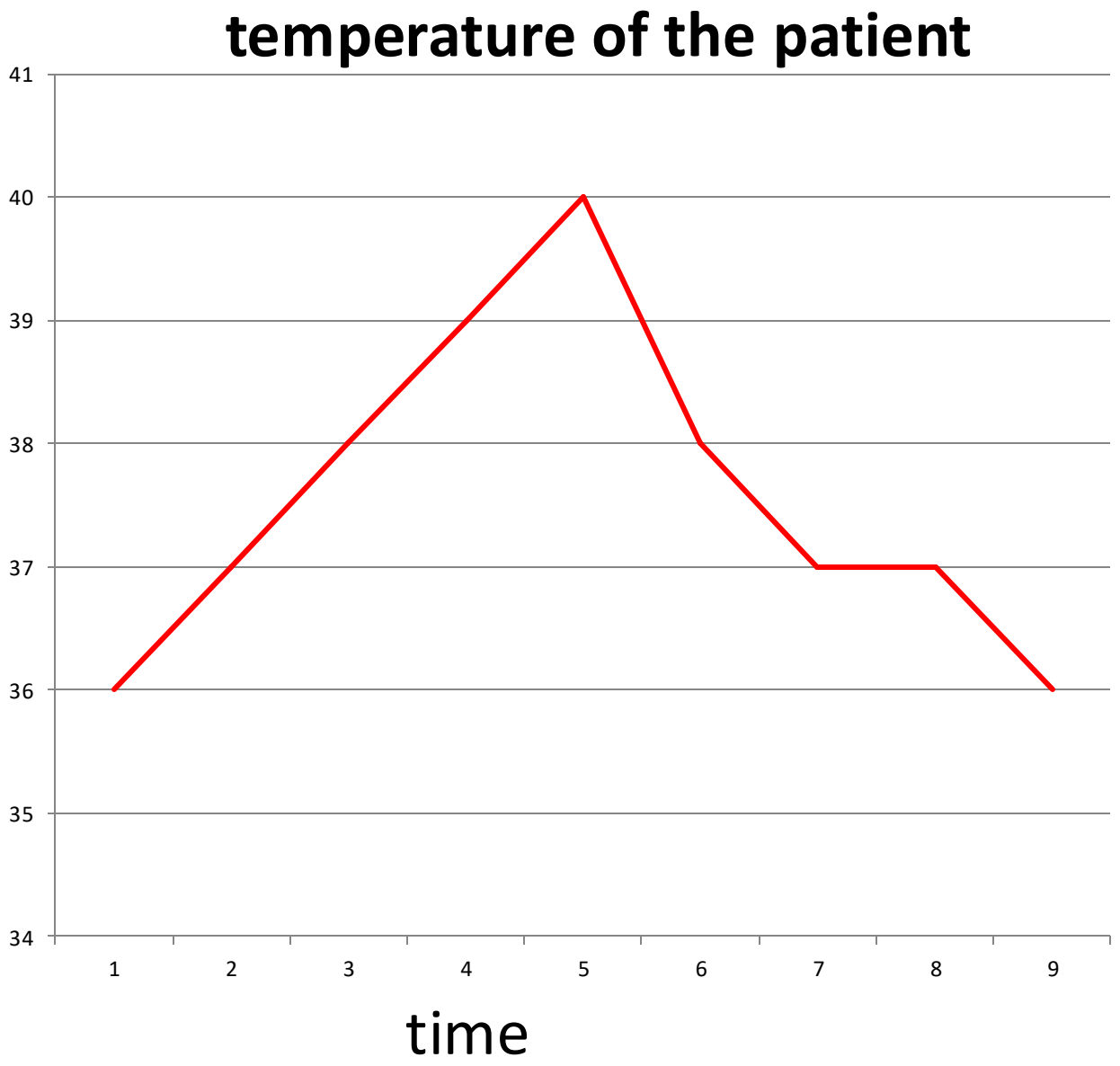
THE LINE GRAPH

- This type is specifically used when we are dealing with a **certain observation** that varies according to ***time***.
- That is when we are dealing with a time variable.
- (The time variable is a special type of continuous quantitative variable)
- Usually the time variable is put on the **horizontal axis** (**X-axis**) and the **other variable** is put on the vertical axis (**Y-axis**),
- then each observation is shown on the graph by means of a point opposite to the **exact time value** on the horizontal axis and opposite the corresponding value on the vertical axis,
- then every two consecutive points are joined by a straight line.

Example of this is a temperature chart of the patient.(It is also used in study of trends of birth and death rate)

Time	temperature
1	36
2	37
3	38
4	39
5	40
6	38
7	37
8	37
9	36

temperature



Evaluation of table or graph

Can this table or graph stand alone ?

It should be self explanatory, Through,

Labeling it properly .

Begin with title and carried on through out table or graph

Title should contain :

what kind of data is this

who were involved .

where it was collected .

when it was done .

I II III …Graph

No. 1 2 3 … Table

above table

Title

Below graph.

Foot note may needed.


Thank you ...


Thank
you


?
Any questions?
© Presentation-Process.com

This include:

Presentation of data by

~~1. Graph and or~~

~~2. Tables~~

3. **Calculation or numerical summaries, such as Frequency, Average, Mean, Median, Mode Percentages**

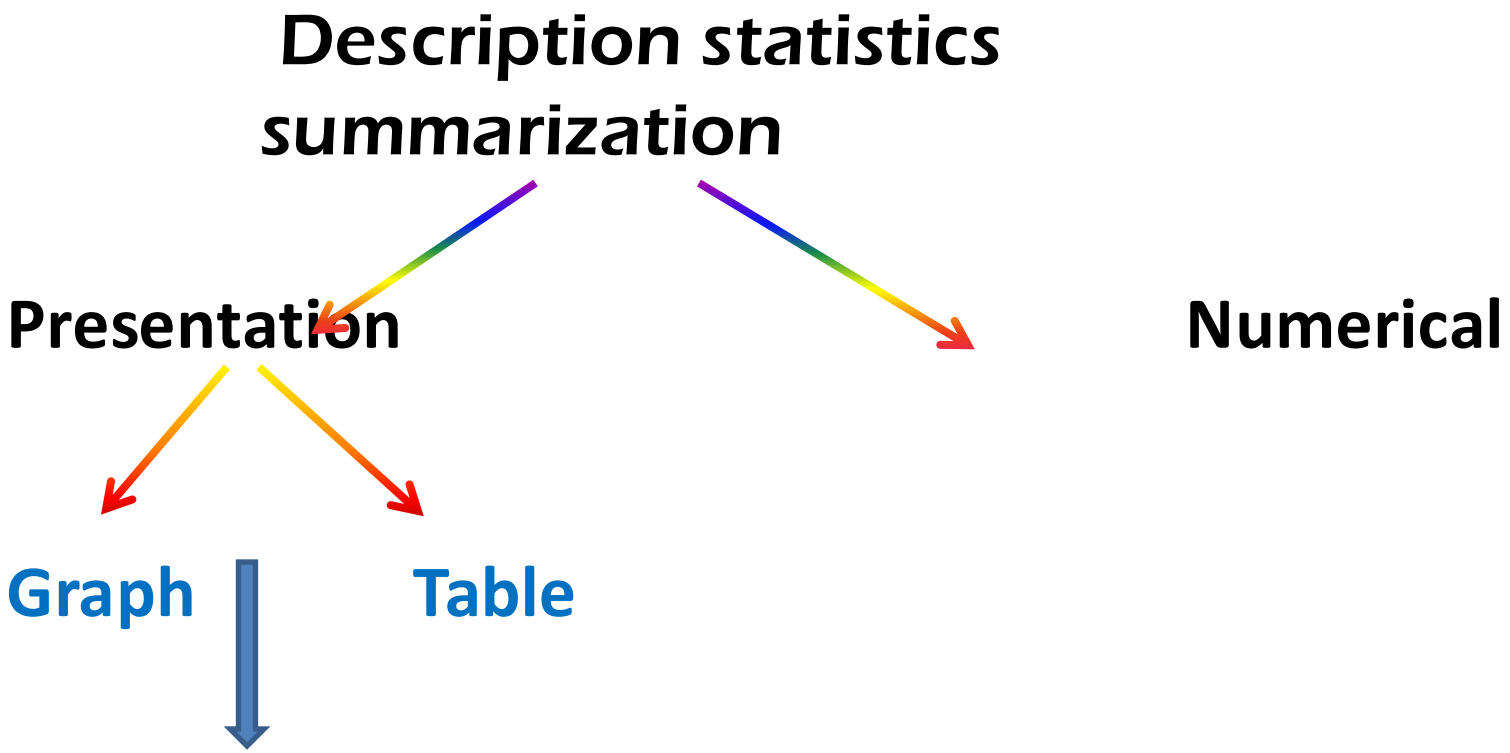
Descriptive statistics

Biostatistics consist of

1-Collection of data .

2-Presentation of data

3-.Estimation of data



-this approach might not be enough,

-comparisons between one set of data & another

-summarize data by one more step further .

-presenting a set of data by a

- single Numerical value