Definition of Biostatistics and the Types of Data and Variables

مفهوم الاحصاء الحيوي وانواع البيانات والمتغيرات

Lecture No. 1

Health Education And Promotion Department

College Of Public Health & Health Informatics

Session content

- Statistics and Biostatistics Definitions. تعريف
- Data Types. انواع البيانات
- Variable types انواع المتغيرات
- Data Presentation تمثيل البيانات

Statistics and Biostatistics

What is statistics

Statistics as a discipline is the development and application of methods to collect, analyze and interpret data. Modern statistical methods involve the design and analysis of experiments and surveys, the quantification of biological, social and scientific phenomenon and the application of statistical principles to understand more about the world around us.

What is biostatistics?

Biostatistics is the branch of statistics responsible for the proper interpretation of scientific data generated in the biology, public health and other health sciences (i.E., The biomedical sciences).

Statistical glossary link:

https://hbiostat.org/glossary/

تطبيق علم الاحصاء في المجالات الحيوية

Biostatistics is a word made from biology and statistics, the application of statistics to a wide range of topics in biology.

فائده علم الاحصاء

It is the science which deals with development and application of the most appropriate methods for the:

- Collection of data. جمع البيانات
- Presentation of the collected data. تمثيل البيانات
- Analysis and interpretation of the results. تحليل وشرح النتائج
- Making decisions on the basis of such analysis اتخاذ القرارات

Role of Statisticians دور الاحصائي

- To guide the design of an experiment or survey prior to data collection يوجه تصميم التجربة قبل جمع البيانات
- To analyze data using proper statistical procedures and techniques تحليل البيانات باستخدام طرق احصائية تحليل مناسبة
- To present and interpret the results to researchers and other decision makers عرض وشرح النتائج للباحثين واصحاب القرار

Data Types

Variables and Constants

In many introductory courses you will come across characteristics or elements such as rates, outputs, income, etc., measured by numerical values. Some of these will always remain the same, and some will change. The characteristic or element that remains the same is called a ***constant***. For example, the number of donuts in a dozen is always 12. That means the number of donuts in a dozen is a constant.

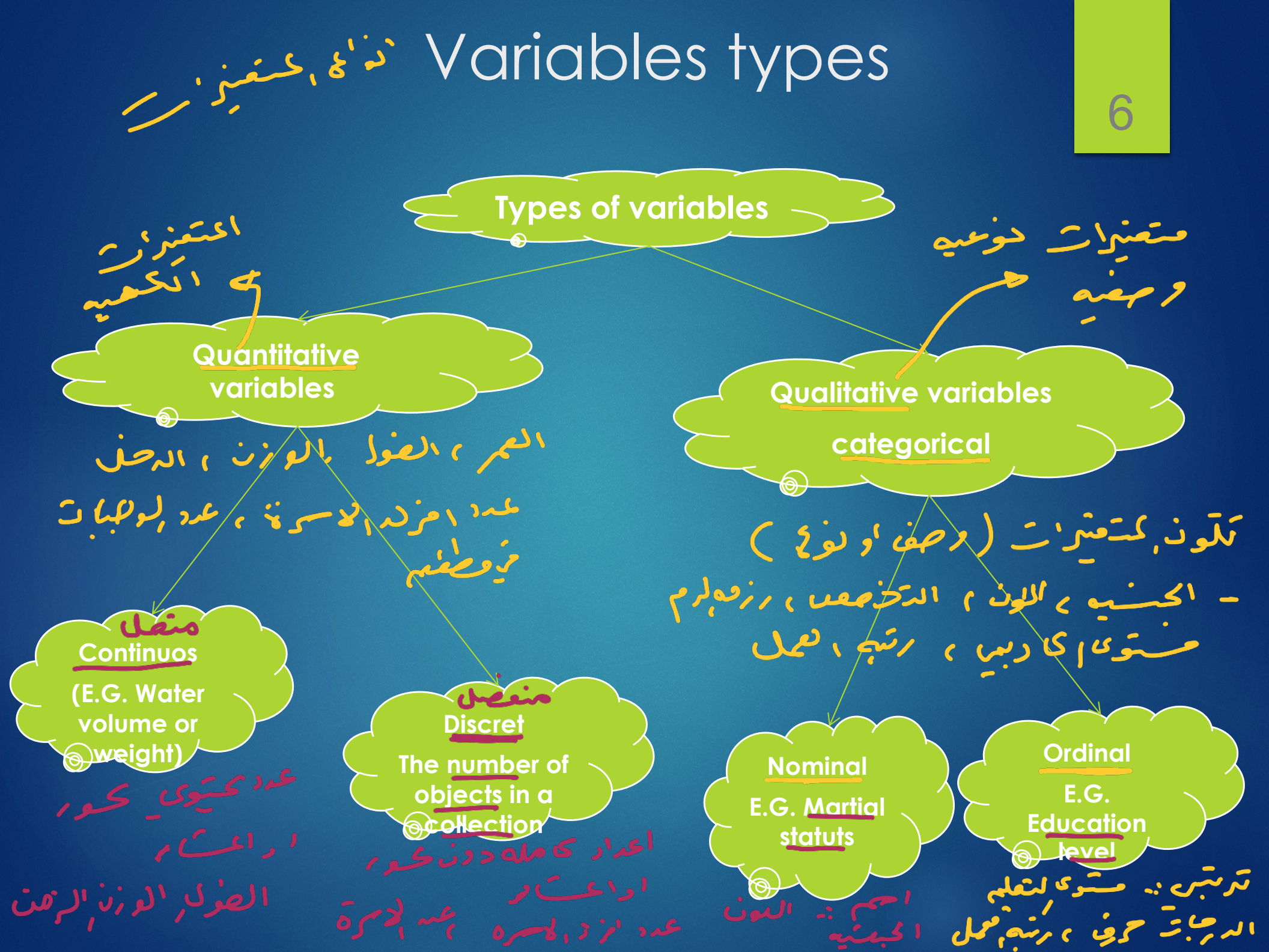
While some of these characteristics or elements remain the same, some of these values can vary (e.g., the price of a dozen donuts can change from $2.50 to $3.00), we call these characteristics or elements ***variables***. **Variable** is the generic term for any characteristic or element that changes. You should be able to determine which characteristics or elements are constants and which are variables.

Example

Which of the following are variables and which are constants?

- The temperature outside your house. This is a variable.
- The number of square feet in a room that is 12 ft by 12 ft. This is a constant.
- The noise level at a room. This is a variable.

Variables types



Variables Types Cont.

A variable is a characteristic which shows variability or variation they are two types of variables:

1. Quantitative variable is numerical; there are two types of quantitative variables:
   - Discrete variable can only take specific numeric values: e.G. Number of brothers, number of cars in a car park.
   - Continuous variables can take any numerical value: e.G. Height, mass, length

2. Qualitative variable is data that is not given numerically: e.G. Favorite color, place of birth, favorite food, type of car. Qualitative variables may be nominal & ordinal.
   - E.G of nominal variable; sex male or female.
   - E.G of ordinal variable;

     1= very low 2= low 3= medium 4= great 5= very great

Data Presentation

There are two types of data presentation:

1. Numerical presentation: (tabulation)

Tables are the format in which most numerical data are initially stored and analyzed and are likely to be the means uses to organize data collected during experiments and dissertation research

2. Graphical presentation

"The transformation of data through visual methods like graphs, diagrams, maps and charts is called representation of data."

Numerical Presentation

Example no. One simple tabular presentation : table of distribution of 50 patients at the surgical department of EL nor hospital in august 2012 according to their age:

Age (years)	Frequency	%
20 – 29	12	24
30 – 39	18	36
40 – 49	5	10
50+	15	30
Total	50	100

Numerical Presentation Cont.

جدول مفرد

Example no. Two complex tabular presentation : table of distribution of 20 lung cancer patients at the chest department of EL nor hospital and 40 controls in august 2012 according to smoking:

Smoking	Lung cancer				Total	
	yes		no			
	No.	%	No.	%	No.	%
Smoker	15	75%	8	20%	23	38.33
Non smoker	5	25%	32	80%	37	61.67
Total	20	100	40	100	60	100

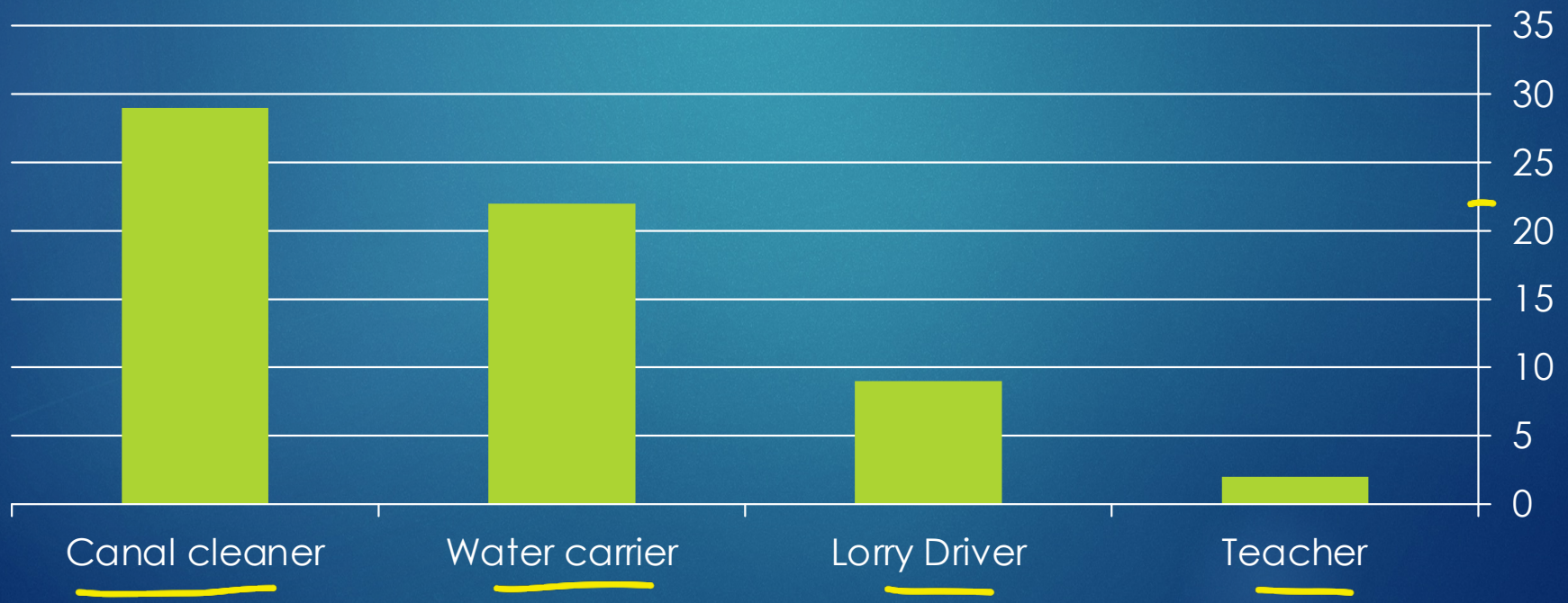
Graphical Presentation

This techniques is to present the data in pictorial form so the readers may readily appreciate and understand the essential features of frequency distribution.

1. Bar graph or bar chart:

This type is a graphic device employed to represent data that are either nominally or ordinally scaled. The area of each bar may be used to represent the frequency for that category, thus the total area of the bars is equal to ***N***.

Bars chart: Prevalence Of Schistosoma Mansoni



2. PIE CHART:

This type is a graphic device employed to represent data that are either nominally or ordinaly scaled. The number of individual in each category are represented by the size of slice of the pie, thus the total area of the pie is equal to ***N.***

Example:

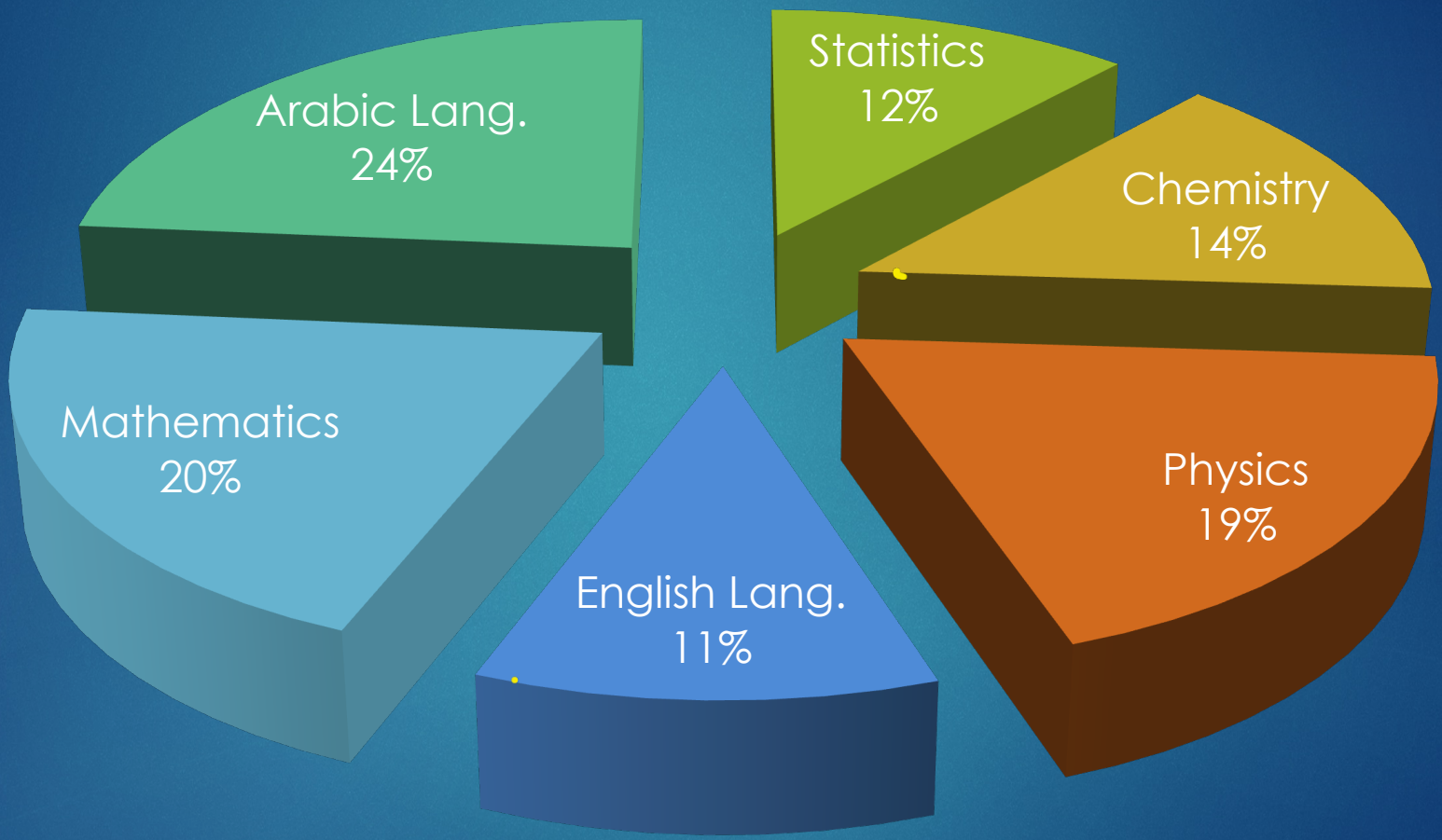
The table below show the result of a student in different subjects.

Subjects	Arabic Lang.	Mathematics	English Lang.	Physics	Chemistry	Statistics
Result	30	26	14	24	18	15

Total 127

23.6% ≈ 24%

STUDENT RESULTS



3. Histogram:

This type is a graphic device employed to represent scale data. The frequency may be represented either by the area of the bar or by it's height. The histogram for the grouped age data can be drawn either showing the class interval or by using the midpoint.

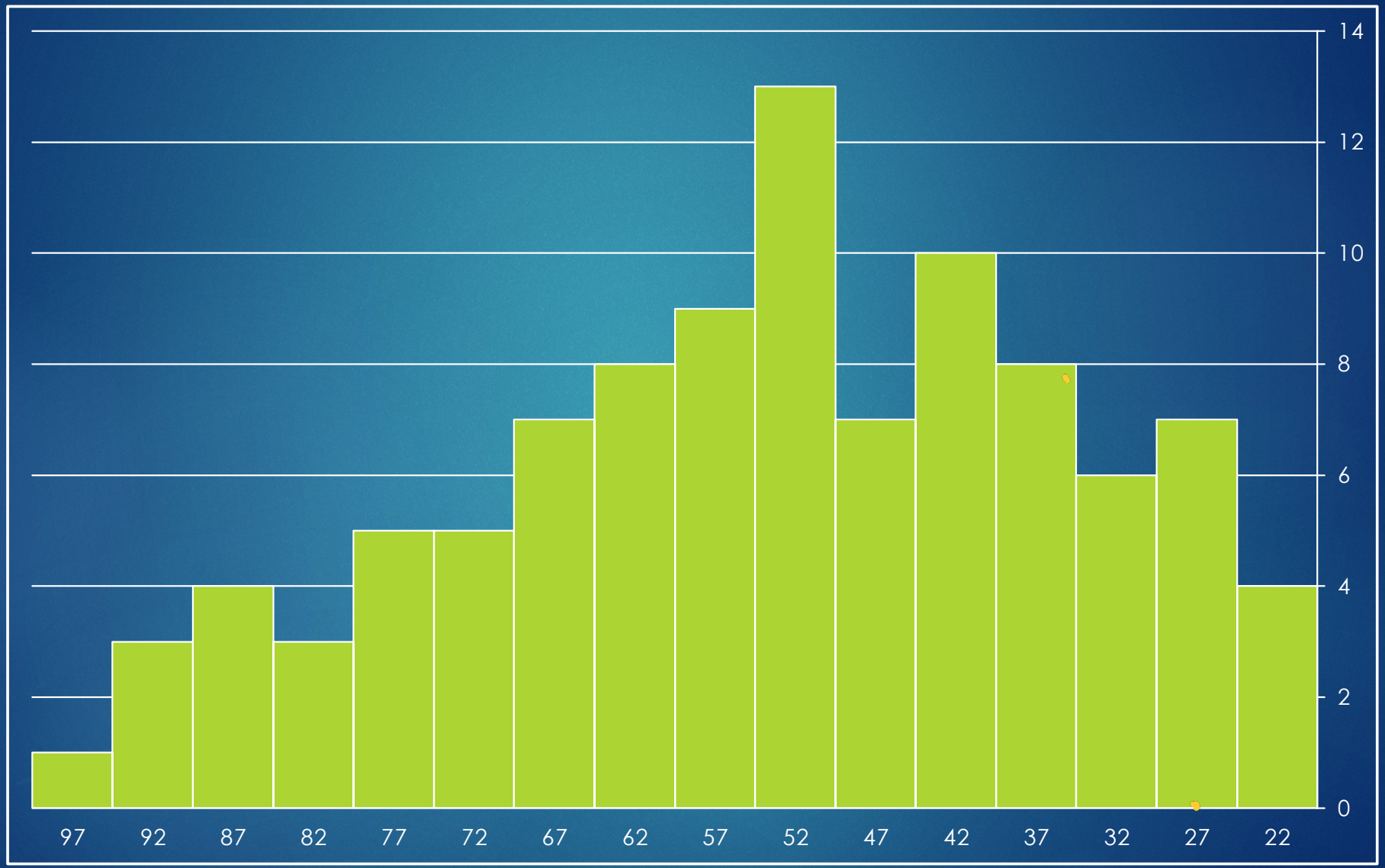
The advantage of a histogram is that one can readily see how the data are organized. The disadvantage is that we pay a price for grouping the data, thus the wider the interval the more information is lost.

Example:

The table below show the ungrouped data for the age of the children in months.

Midpoint	22	27	32	37	42	47	52	57	62	67	72	77	82	87	92	97
Frequency	4	7	6	8	10	7	13	9	8	7	5	5	3	4	3	1

HISTOGRAM OF AGE OF CHILDREN



4. FREQUENCY CURVE:

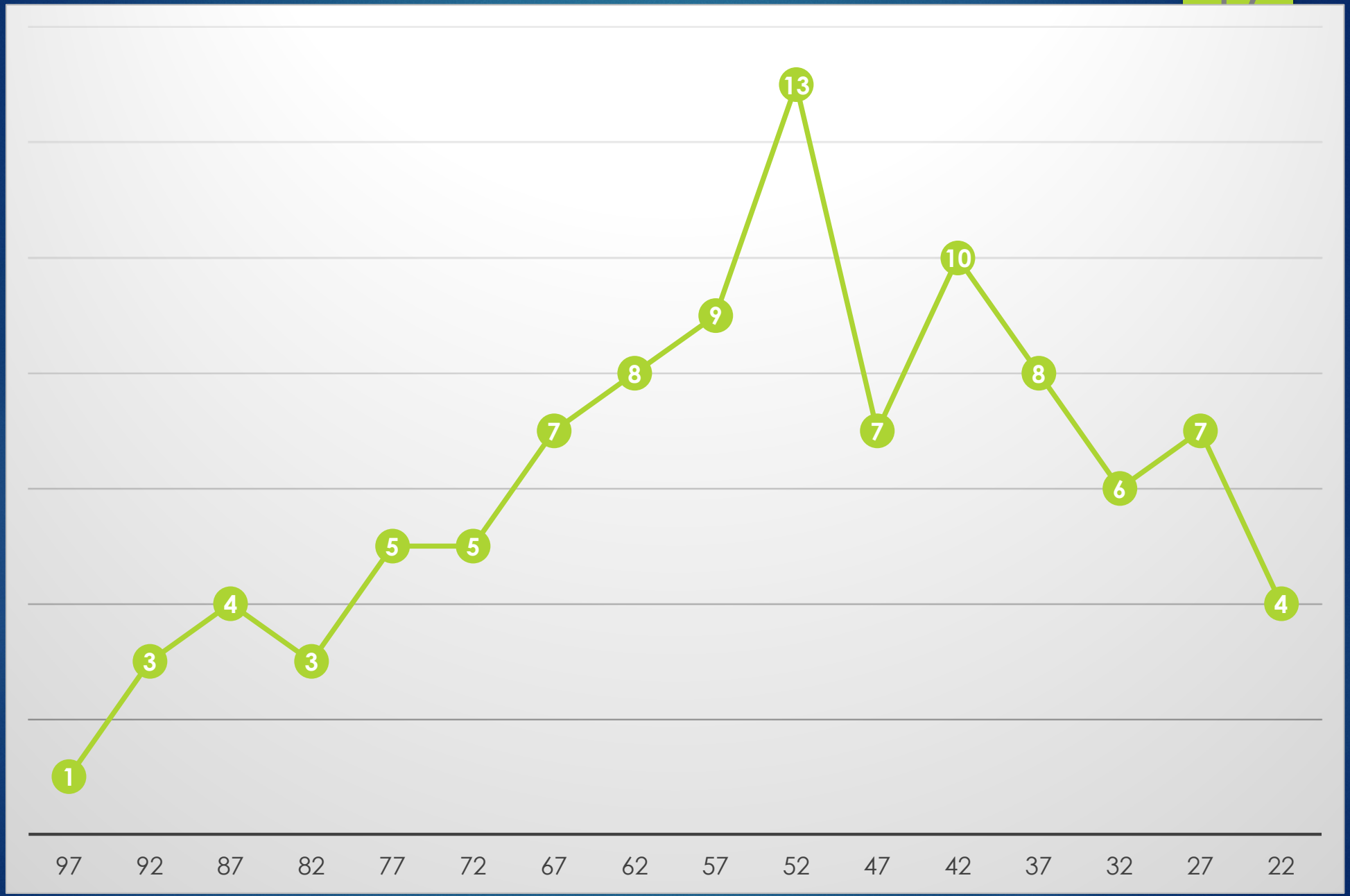
This type is a graphic device employed to represent scale data. The frequency can be represented by the frequency curve by joining the midpoint of the bar (in histogram) with straight lines

Example:

The table below show the ungrouped data for the age of the children months.

Midpoint	22	27	32	37	42	47	52	57	62	67	72	77	82	87	92	97
Frequency	4	7	6	8	10	7	13	9	8	7	5	5	3	4	3	1

FREQUENCY CURVE OF AGE OF CHILDREN



4. CUMULATIVE FREQUENCY CURVE:

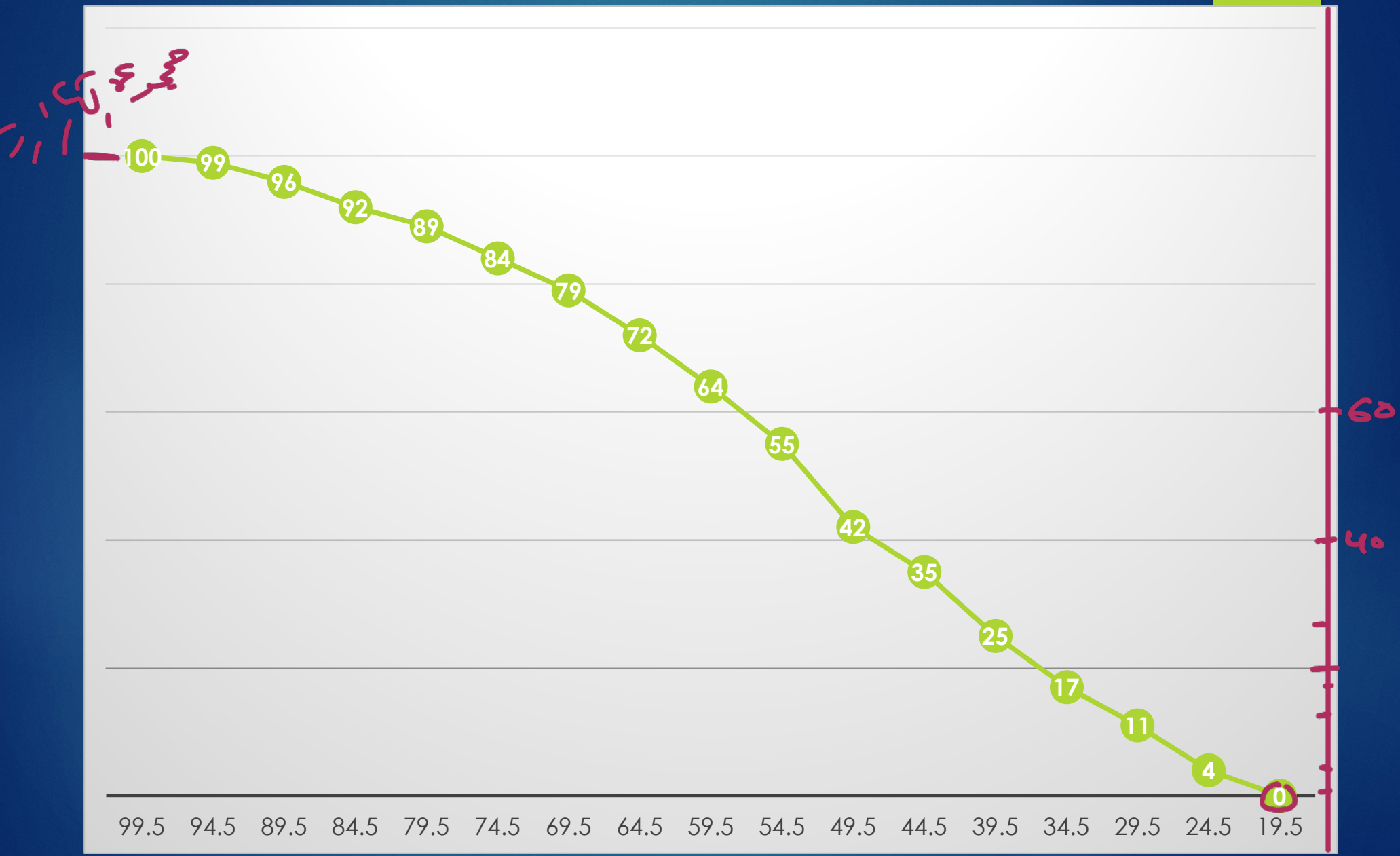
This type is a graphic device employed to represent scale data. The frequency can be represented by the cumulative frequency curve by joining the midpoint of the bar with straight lines

Example:

The table below show the ungrouped data for the age of the children.

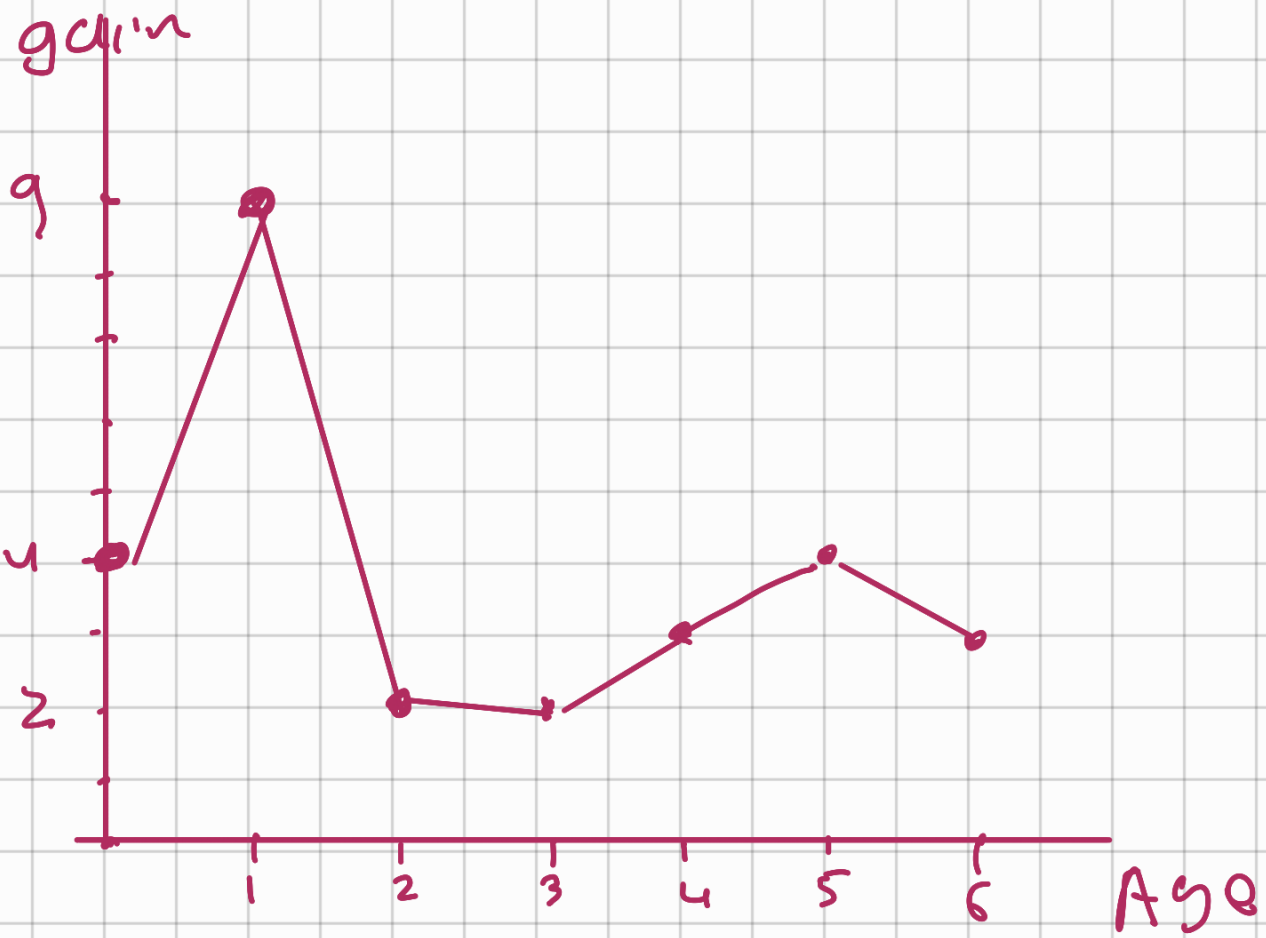
Midpoint	19.5	24.5	29.5	34.5	39.5	44.5	49.5	54.5	59.5	64.5	69.5	74.5	79.5	84.5	89.5	94.5	99.5
Frequency	0	4	11	17	25	35	42	55	64	72	79	84	89	92	96	99	100

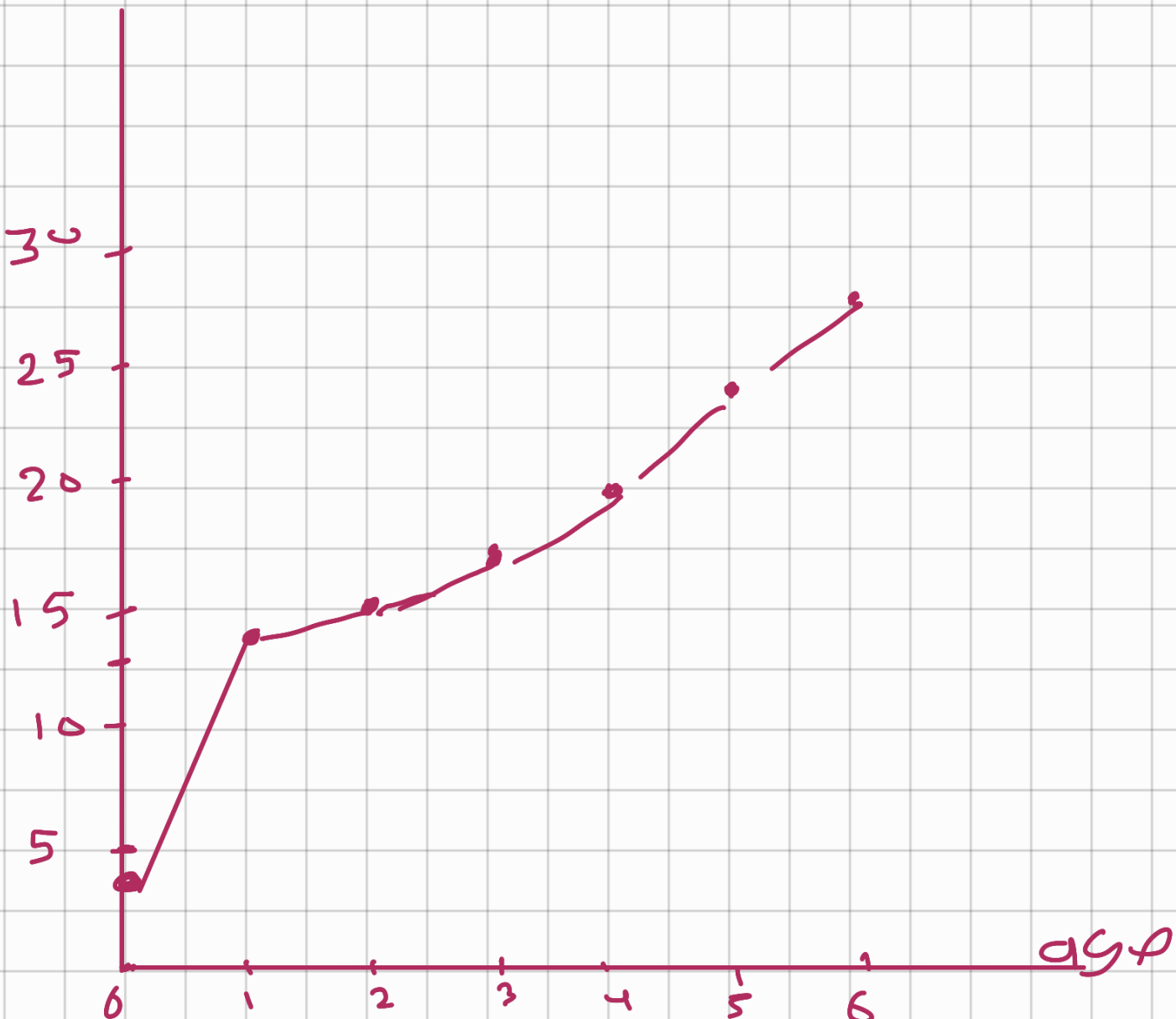
CUMULATIVE FREQUENCY CURVE OF AGE OF CHIL N



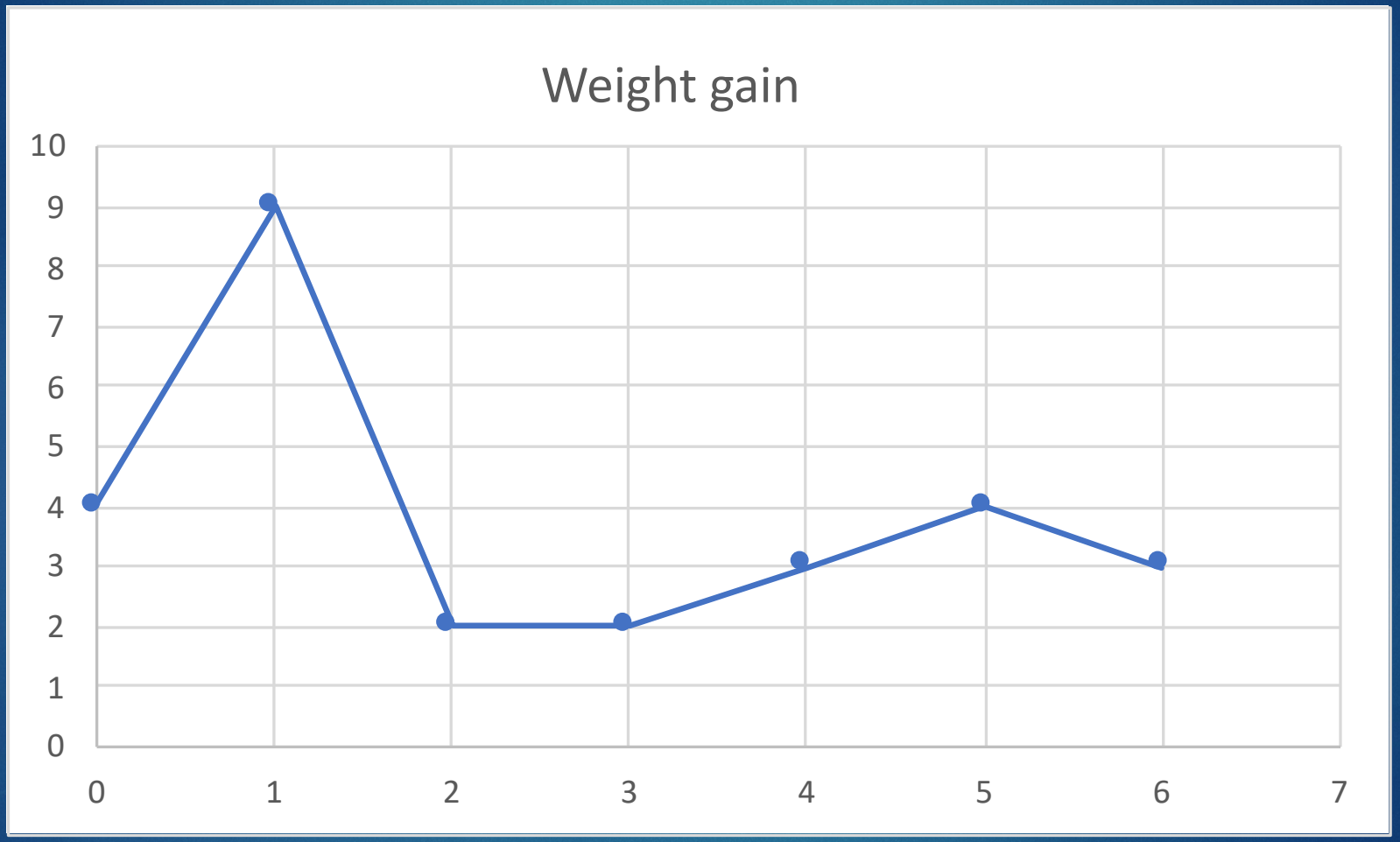
Children Age and Weight gaining

Age	Weight gain	Total weight
0	4	4
1	9	13
2	2	15
3	2	17
4	3	20
5	4	24
6	3	27

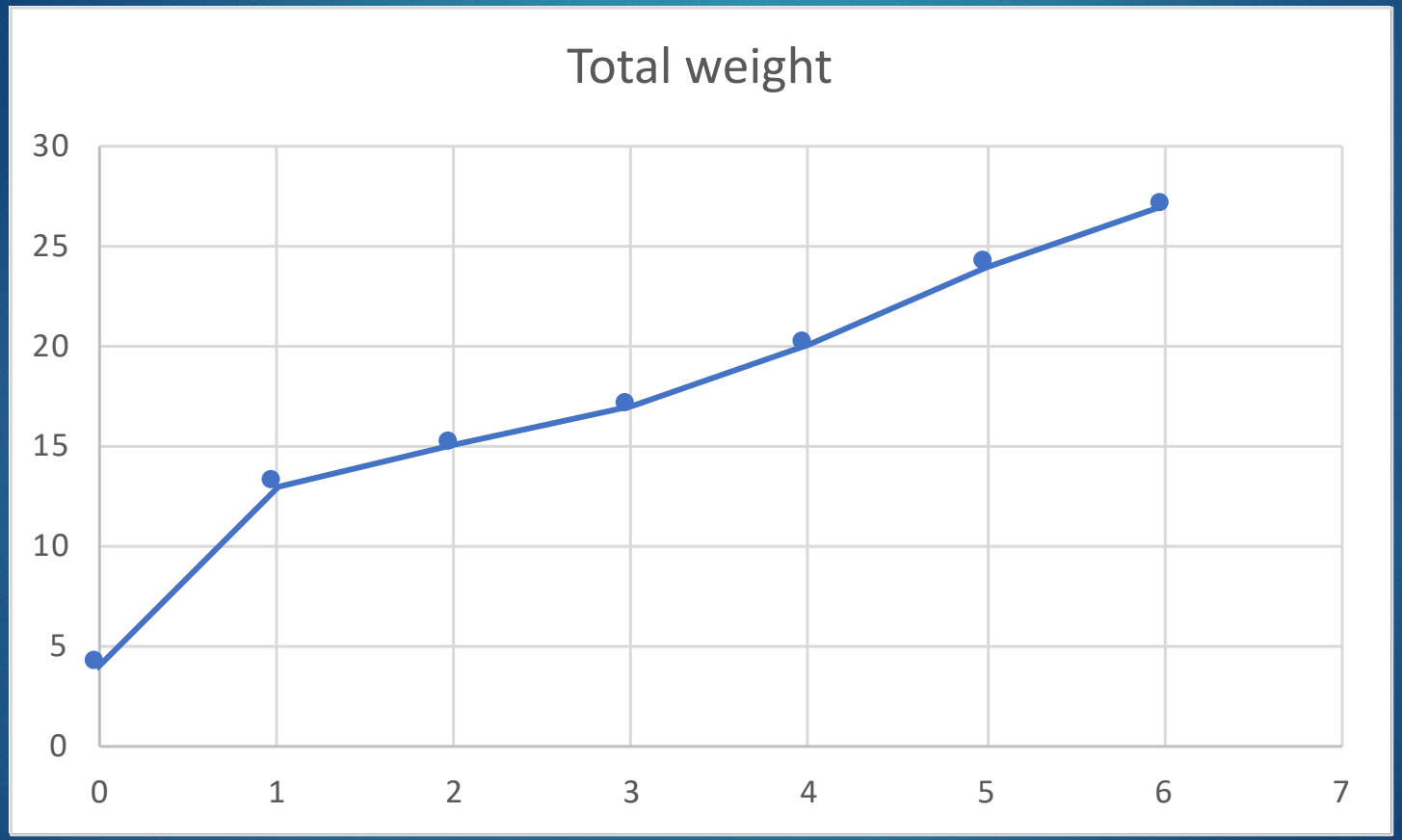

gain
9
4
2
1
2
3
4
5
6
Age


30
25
20
15
10
5
0
1
2
3
4
5
6
age

Frequency Line Chart



Cumulative Frequency Curve



Variable types	Presentation		
	Example	Picture	Statistic
Quantitative (Continues)	• Rate • Age • Hight • Weight • Temperature • Count (number of things)	• Histogram	• Mean • standard Deviation • Median • Interquartile range
Qualitative (Categorical)	• Gender • Education level • Colors • Martial Status • Point scale (very low – very high)	• Bar Chart • Pie Chart	• Frequency (count) • Percentage (%) • Mode

Game Time

Questions

Question 1:

What type of variables?

- height measured in number of feet
- weight measured in number of pounds
- number of days it snowed
- hair color
- gender
- average daily temperature

Questions Cont.

Question 2:

Indicate the best type of graph to use in each of the following: (put a circle)

1) You want to show the trend of maternal mortality rate of (country), 1960-2000

a. pie chart b. bar chart c. line graph d. histogram

2) You want to show the Distribution of 45 patients at (place) , in (time) by age

a. pie chart b. bar chart c. line graph d. histogram

3) You want to show the Distribution of 100 cholera patients at (place) , in (time) by age group.

a. pie chart b. bar chart c. line graph d. histogram

Questions Cont.

Question 3:

Indicate the best type of graph to use in each of the following: (put a circle)

1) You want to show how the marital status is divided among single, married, divorced and widowed by sex.

a. pie chart b. bar chart c. line graph d. histogram

2) You want to show how the total number of people living in Saudi Arabia is divided among Saudi, Arab and None-Arab

a. pie chart b. bar chart c. line graph d. histogram

Questions Cont.

Question:

Frequency Distributions, Present the math score in a Bar chart.

Class	Frequency
41-50	1
51-60	2
61-70	6
71-80	8
81-90	14
91-100	9

Questions Cont.

Question:

The areas of the various continents of the world (in millions of square miles) 7 are as follows:11.7 for Africa; 10.4 for Asia; 1.9 for Europe; 9.4 for North America; 3.3 Oceania; 6.9 South America; 7.9 Soviet Union.

Draw a bar chart representing the above data and where the bars are horizontal.

The population (in millions) of the US for the years 1860-1950 is as follows: 31.4 in 1860; 39.8 in 1870; 50.2 in 1880; 62.9 in 1890; 76.0 in 1900; 92.0 in 1910; 105.7 in 1920; 122.8 in 1930; 131.7 in 1940; and 151.1 in 1950.

Make a time plot showing this information using line graph

Questions Cont.

Question:

Problem 30.21) Ten people were surveyed about their favorite pets and the result is shown in the table below.

Pet	Frequency
Dog	2
Cat	5
Hamster	3

Make a bar & Pie chart for the following table of data

Further study – Internet sites

https://www.youtube.com/watch?v=v-jvFrxhjkw