✓ **Introduction:** Statistics which is a synonym for "numerical facts" is a relatively new branch of mathematics that aims at:

1) Collect data.
2) Classifying, summarizing and organizing data.
3) Reading data in a more efficient way.
4) Representing data in different forms (pie graph, histogram, bar graph …).
5) Interpret and find relation among data.
6) Draw out conclusions from a given set of data.

✓ **Usages:** almost every field of studies benefit from organizing statistical surveys.

- *Educational field*: Teachers evaluating students over a specified set of objectives (how to: factorize, develop…).
- *Entertainment*: A TV show analyzing a football match (goals scored, shoots on target percentage of ball possession…)
- *Industry*: A company introducing a new product to the market (study the profit per year)

✓ **Statistical vocabulary:**

- ***Population:*** is the set of observed elements having a common property.
  - The set of students in a class, The set of teams in a certain league.
- ***Size:*** is the total number of elements in a population.
- ***Character (values) or variable:*** is the common property of the population under study.
  - The height, weight, grades…. of an individual in a set of population.
  - The color of eyes, gender (male, female), behavior of an individual in a population.
    - Types of characters:
      1) ***Quantitative:*** a character is said to be quantitative if it can be measured.
         - Length, number of children, number of books read ….
      2) ***Qualitative:*** a character is said to be qualitative if it cannot be measured.
         - The color of eyes, gender (male, female), behavior ….
- ***Frequency(n):*** is the number of times a character is observed.

  Note that: The total frequency or size is the sum of all frequencies and it is denoted by ***N***.
- ***Relative frequency(R.f):*** is the ratio of the frequency(n) to the size (N) of an object.

  In symbols: $R.f = \frac{n}{N}$

  Note that: The relative frequency is a number strictly included between 0 and 1

  In symbols: $0 < R.f < 1$

Application: Upon studying the number of daily hours spent by each of the ***25*** students of Grade 9 on the internet, we obtained the following results organized in the table below:

Number of daily hours (x_i)	1	2	3	4	5
Number of students (n_i)	2	y	9	x	3

1) Indicate: Population: ……
Individual: ……...
2) Determine the character (variable) under study: ……
   a. Specify its nature. ……
   b. Give three modalities of the given variable: ……
   c. Give examples about qualitative (categorical) data:……
3) Specify the size of the population: ……
   a. Explain what x and y represent in the above table and interpret one of them.
   ……...
   ……...
   b. Deduce a relation between x and y. ……
4) The mean of a numerical set of data is given by the formula: $\overline{X} = \frac{\sum_{i=1}^{k} n_i x_i}{N}$, knowing that the mean number of daily hours spent on the internet is 3.2.
   a. Interpret the value $\overline{X} = 3.2$: ……
   b. Calculate x & y: ……..
5) **For the remaining parts, let x = 7 and y = 4**
   a. Set up the table of increasing cumulative frequency in percentage:

Number of daily hours (x_i)	1	2	3	4	5	Total
Number of students (n_i)	2	y	9	x	3	
$\%R.f$						
$\%I.C.f$						
Central angles (α_i)						

   b. Interpret any value: ……
   c. Is it true that 76% of the students use the internet at least 3 hours daily? **Justify**.
   ……..
6) Calculate the central angles and draw the circular diagram for this statistical distribution.

Note that: *in a statistical table, frequencies, relative frequencies, frequencies in percent and central angles are proportional.*

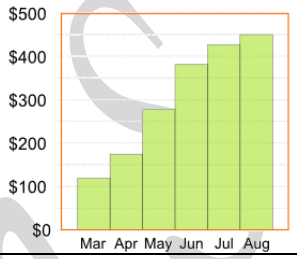
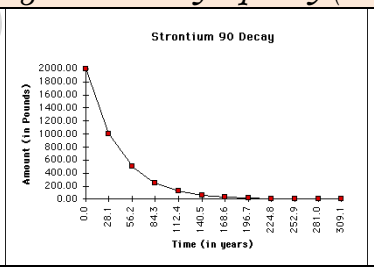
✓ Measure of central tendency:

- ***Range(R):*** is the difference between the ***highest*** and ***lowest*** observed values for a quantitative character.
- ***Mode:*** is the variable that admits the ***highest frequency***.
  - Having two modes is called **"bimodal"**.
  - Having more than two modes is called **"multimodal"**.
- ***Mean (Average):*** is of two main types

Type	*Definition*	*Formula*
Arithmetic Mean	is the ratio of the sum of values to the size of the population	$\overline{X} = \frac{\sum x_i}{N}$
Weighted mean	is the ratio of the sum of product of values by their frequencies to the size	$\overline{X} = \frac{\sum x_i \cdot n_i}{N}$

✓ Cumulative frequency:

- Cumulative means "***how much so far***".
- Think of the word *"**accumulate**"* which means to gather together.

Types of cumulative frequencies	*Increasing cumulative frequency (ICf)*	*Decreasing cumulative frequency (DCf)*
Graphing		
Meaning	I.C.F, Histogram	D.C.F, broken line diagram
Note	To have cumulative totals, just **add up the values as you go**.	

Histograms vs Bar Graphs

- Bar Graphs are good when your data is in **categories** (such as "Comedy", "Drama", etc).
- When you have ***continuous data*** (such as a person's height) then use a ***Histogram***.

It is best to leave gaps between the bars of a Bar Graph, so it doesn't look like a Histogram

👀 How can we use the calculator to find some statistical indicators?

CASIO $fx-95ES$	CASIO $fx-991ES$	
1- Mode2 2- Shift/clr/1/= 3- Enter data: i) Variable/shift/,/frequency/m+ 4- To find: a) **Mean**: press shift/2/1/= b) **Standard deviation**: press shift/2/2/=	1- Mode/3/stat/1:1-var 2- On 3- Shift/mode 4- Down 5- 4:stat/1:on 6- Same as step 1. 7- Fill data. 8- Ac 9- Shift/1	10- Press: 4: var Choose one you want to cal. i. Mean: $\overline{x}$ ii. Standard deviation.

✓ Median:

- Def: is the middle value of a set of ordered data.
- Symbol: M_e
- Interpretation: median is the value at which 50% of data is below it and 50% of data above it.
- Determination:

Ungrouped data:

- If N is odd, then the median is the value whose rank is $\frac{N+1}{2}$
  - Eg: Consider the data: $3,5,8,9,11,14,17$

    Since $N=7$ is odd, then M_e is the value whose rank is $\frac{7+1}{2}=4$

    Thus, $M_e=9$

- If N is even, then the median is the average of the values whose ranks are $\frac{N}{2}$ & $\frac{N}{2}+1$
  - Eg: Consider the data: $5,7,8,10,11,13,14,16$

    Since $N=8$ is even, then M_e is the average of values whose ranks are $4\,\&\,5$

    Thus, $M_e=\frac{10+11}{2}=10.5$

Grouped data: To find median using a statistical table:

- 1st: Find $I.C.F$ of the data.
- 2nd: Determine $\frac{N}{2}$
- 3rd: $If\ \frac{N}{2}\ is:\begin{cases}\text{Enclosed between two } I.C.F \text{ corresponding to } x_i\ \&\ x_{i+1}, then\, M_e = x_{i+1}\\ \text{Equals one of the } I.C.F \text{ corresponding to } x_i, then\, M_e=\frac{x_i+x_{i+1}}{2}\end{cases}$

Eg-1: (M_e indicated above the column 11)

x_i	8	9	10	11	12		13	Total
n_i	10	7	2	5	10		6	40
$I.C.F$	10	17	19	24	34		40	

$\frac{N}{2}=20$

Since $\frac{N}{2}=20$, then rank(M_e)lies between the $I.C.F$ of the values $x_3=10\,\&\,x_4=11$.

Thus, $M_e=x_4=11$

Eg-2:

x_i	8	9	10	11	12	13	Total
n_i	10	7	3	5	9	6	40
$I.C.F$	10	17	20	25	34	40	

$\frac{N}{2}=20$

Since, $\frac{N}{2}=20$ then rank(M_e)equals the $I.C.F$ of the value $x_3=10$.

Thus, $M_e=\frac{x_3+x_4}{2}=\frac{10+11}{2}=10,5$

✓ Standard Deviation:

The Standard Deviation: is the measure of how spread out numbers is?
Its symbol is **σ** (the Greek letter sigma)
The formula is easy: it is the **square root** of the **Variance.** $\sigma = \sqrt{v}$
So now you ask, "What is the Variance?"

Variance

The Variance is defined as: The average of the **squared** differences from the Mean.
To calculate the variance: follow these steps:

- Work out the ***Mean***.
- Then for each number: subtract the Mean and square the result.
- Then work out the average of those squared differences.
- Formula: $v = \frac{\sum_{i=1}^{k} n_i \left(x_i - \bar{x}\right)^2}{N}$

Application:

A math teacher is asked to name the best student in his class, but there are two students Sadek and Sara, that he thinks deserve to be named, so he decides to analyse their grades per 20.

Sara's grades	14	9	12	16	13	15
Sadek's grades	12	10	8	16	15	18

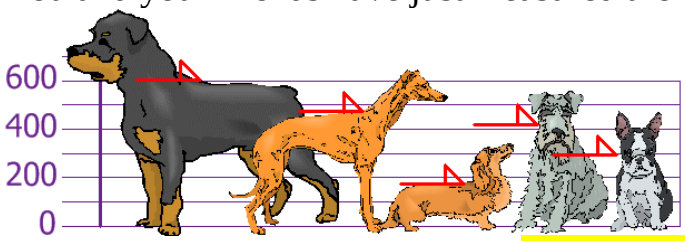
1) Propose a way to decide which deserves to be picked. ……
……...
……...
……...
   a. Did your method work to compare between the students? ..……..
   b. If not, then why do you think that happened? ……
2) Did you find the average? …….
   a. If no find it. ……..……...
   b. If yes then find the mean deviation of each grade from the mean and complete the table:

Sara's grades	14	11	9	16	7	15
Mean Deviation	2	-1	-3			
Absolute deviation	2	1	3			
Sadek's grades	12	8	2	16	15	19
Mean Deviation						
Absolute deviation						

3) Do you think that this will work out? Why? ……

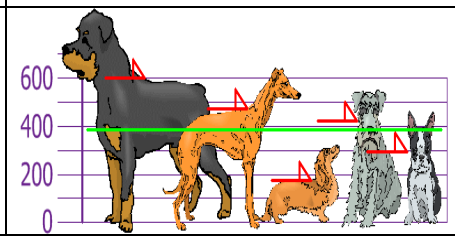
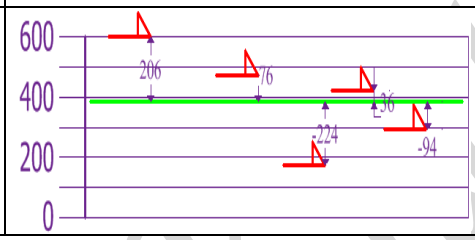
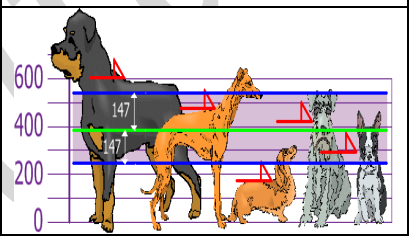
Application:

You and your friends have just measured the heights of your dogs (in millimetres):



The heights (at the shoulders) are: 600mm, 470mm, 170mm, 430mm and 300mm.

the:

Find out	Mean:	Difference from the Mean:	Standard Deviation: **σ**
Graphical meaning			

And the good thing about the Standard Deviation is that it is useful. Now we can show which heights are within one Standard Deviation (147mm) of the Mean.

So, using the Standard Deviation we have a "standard" way of knowing what is normal, and what is extra-large or extra small.

Rottweilers **are** tall dogs. And Dachshunds **are** a bit short ... but don't tell them!

Choose one of the following surveys, and then answer the related questions given at the end.

S-1: Study the distribution of tourists among five Lebanese cities.

S-2: Study the number of goals scored in the last ten games for your best five football teams.

S-3: Study the number of points scored by five players in a basketball match between your two favorite teams.

S-4: Study the number of hours you spend preparing for each of five of your school assignments.

For each statistical survey specify (on your own):

1) The population and the size of the chosen sample space.
2) The variable under study (character) and its type (Qualitative or Quantitative)
3) The range of the data if possible.
4) The highest and the lowest values among the specified data.
5) The average (mean) of your data.