Lesson 1

Introduction

- Theory
- Pratical

Blog = simple static we page made like a blog or word press site.

Excel sheet with:

- homework 1 = each student put here the link of his homework

Thesis will be given at the end of the course at the end we have:

- thesis
- blog with all the homework

Theory

- Presentation and representation of situation
- Reporting
- Data analysis big data (regression analysis)
- Quantify the change of an event
- Prediction
- drug testing before publications (statistical test)
- market analysis

> Statistic is almost everywhere

Everything's starts to notion of population.

Population

Population is set of statistical unit. (set of elements that are the object of study). Population does not to be something that exists but can be something that is dream.

- Population is not static, he can change during the time (is a dynamic flow).
- Can be a stream of data.
- I have to recognize the statistical unit and the variables.

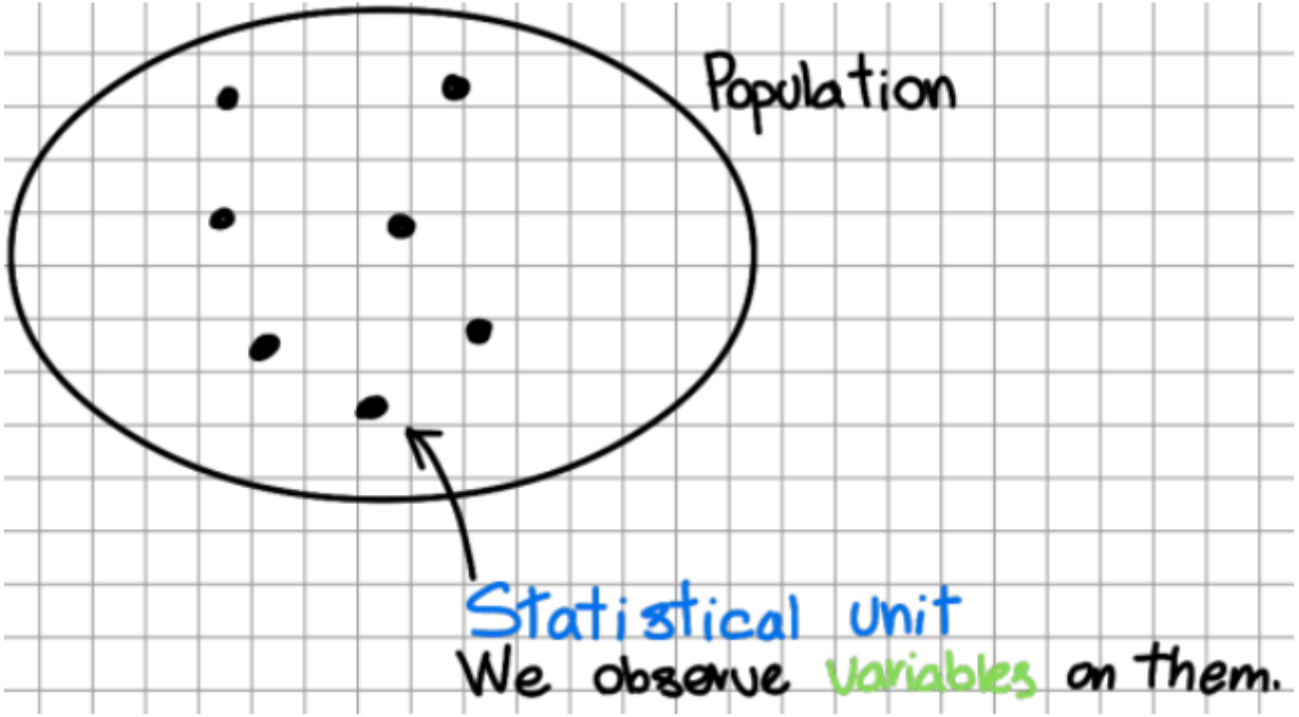
If we are population and i'm interesting to observe something like an activity, in this case i will asking wich is your "hobby". That phenomenal is called the character or variable.

$(x_1, \cdots, x_n)$ this are called the characters (variables)

- character is more general more astract (weight is an example)
- variable = character + set of all possible value (way that i mesure it)

Statistical Units

For statistical terminology, each of us is called statistical unit, what we observe in a statistical unit is called variable, which can take a certain range of values (in this case yes, no, not expressing my opinion).
We always start from a population, set of statistical units (dots in the circle) and on each statistical unit we observe variables.



Another example of variable that could have been observed could have been the background, putting it in a table as such:

ID	Background	Coder/programmer diff
Salvatore	CS	D
Leonardo	CS	S
Alex	CE	D
...	...	...

set number that i observer and for each character i can have an array of variable.

Example

$75kg \rightarrow$ is quantitative measure

three categorical lines:

- low
- medium
- high
  In this case the cale of measure is differentive, now is qualitative

Missing: image on the paper

Looking at a pratical case:
When we have a DB we have a table wit various column. Each row of the table represent a record. each row will be a statistical unit $V_i\ (x_1^{(i)}, \cdots, x_k^{(i)})$.

Population = set of statistical unit. population does not to be something that exists but ca be something that is dream. Population is not static, he can change during the time (is a dynamic flow). Can be a stream of data.
I have to recognize the statistical unit and the variables.

Analisys can be:

- Univariate
- Bivariate
- .....
- Multivariate $(x_1, \cdots, x_n)$

> We need two variables to studying the relationsheep between two cases of study. If we have more than three variables we have a problem to represent. there is a strument that allow us to create advanced graphs.

if we have a DB and we want to create a plot table how we can do?

- with excel

Missing notes in the paper even to Recap

Meaning of Average (Arithmetically)

the arithmetical average is ...
Defining property of average: i set a cut and the number of distance at the left must be the same of the number of distance at the right.

> we have a different definitions of average.

Meaning of Median (Arithmetically)

More general concept then average.

Practical

Welford is recoursive formula to calculate.
This formula is equal/equivalent to the definitions. is somethings that in mathematically is the same but in practical is better, thanks to the floating point representation.All the trick here, is to find the same formula but floating point representation.

Floating point = sing + mantissa + exponent

Homework 1

Research

- Base notion in Statistics (half page to refresh our memory, concept of distribution, variables, character, population, statistical units)
- Notion of average (Mathematically).
- Computational problems with floating problem representation and recoursive formula (refresh our mind/knoledgement, structure of representation of floating point, errors, numerical instability)
- Numerical Solution (Knuth).

Program exercise

- Visual studio (C#, VB, Javascript) use one of this.
  - javascript is webpage
  - $C\#$ and VB is winform

We have n servers (system) with n attackers. The hacker (each) has the probability 0.7 to penetrate the system. Make a graphical representation where each attacker is represented like a line that if it penetrates makes a jump of 1 else stay flat. Each line will show the story of that attacker. In practice we draw something like that:

Missing draw

Parameters: m, n, p
This parameters could be:

- $m = 1000$
- $n = 50$
- $p = 0, 4 \ldots 0, 7 \ldots 0, 5$

At time n we want to count how many reached each level. (frequency distribution) Most of the hacker will go to the min, but they can also go anywhere. To make the jump we can use randomizer such as RANDOM $\rightarrow \{0, 1\}$ if the number is minus 0.7 (we decide this value to do some tests) stay flat else is bigger than 0.7 do a jump.

> Don't use library!

There are these functions for plotting: drawlines, drawrectangle.