

## STATISTICS

"Statistics is the art of learning from data"

>Statistical thinking is fundamental to all areas of science and it is becoming a mandatory requirement to understand and model real world phenomena, from economics and finance to biology and industry.

>every day we make decisions under conditions of uncertainty

>many times, the situations or problems we face in real world have no precise or definite solution

>Statistical methods help us make scientific and informed decisions

**What is randomness?**

→ Deterministic phenomena: a phenomenon that follows a mathematical law, with "no" error (this is actually not true.

ex: the eclipse> it is predictable with quote no error and it follows a physical law

→ Random phenomena: cannot be predicted without uncertainty

ex: tossing a coin: you cannot predict with certainty the outcome

Is this all?

Could it be that we are not able to predict the coin toss because of our ignorance?

Model for atmospheric dynamics:

we know exactly the equations governing it, still it is no possible to predict the weather in the long run, any short predictions are feasible

why?

moreover it is deterministic!



A **chaotic system** is a deterministic system that acts as a random system (i.e.: unpredictable)

It acts like a random system because it has the "butterfly effect"= sensitive dependence on initial conditions

The tiny error that you have when you observe will be exponentially amplified, leading to unpredictability

Unpredictability is not a byproduct of our ignorance regarding the equation

chaotic=unpredictable

deterministic

unchaotic=predictable

But we cannot distinguish chaos (deterministic) from true randomness> actually it is not possible to prove that a given phenomenon is random (we cannot know if something is random)

Spurious correlations= correlations made by chance

If you want to learn something from data, you shall follow 5 steps:

1) collect data

2)describe/summarize data

3)analyse data

4)diagnostics=check if the assumptions are true

5)interpret results

6)draw conclusions

>this is not an algorithm and each step requires specific knowledge and expertise

## DESCRIPTIVE STATISTICS

>Statistical unit: subject or object (for instance, a person, firm, bank, country) about which the information is collected

>Variable: characteristic (or phenomenon) under investigation  
we indicate a variable by using capital letters, such as:  $X, Y, X_1, X_n$   
the variable assumes different values for different units!

>Observed sample (data set): set of the observed realizations of the variable  
we indicate the realization of a variable by using lowercase letters with a subscript, such as:

$x_1, x_2, x_3, \dots, x_n$

$y_1, y_2, y_3, \dots, y_n$

>Sample size: number of observations ( $n$ )

Examples