

# Introduction: the statistical thinking; basic definitions

Denis Allard

*Biostatistique et Processus Spatiaux (BioSP), INRA, Avignon*

Ecole Doctorale UAPV "Sciences et Agrosociétés", Statistiques paramétriques,  
7 juin 2016



## Introductory dialog

- What? You're going to take a 3 day stat class? How boring!
- Uhhhh... Thanks
- Why?
- Because I've been told so
- By who?
- My advisor
- ?????? Why ?
- Duno ... He seems to be believe it's important
- ...

## Introductory dialog

- What? You're going to take a 3 day stat class? How boring!
- Uhhhh... Thanks
- Why?
- Because I've been told so
- By who?
- My advisor
- ?????? Why ?
- Duno ... He seems to be believe it's important
- ...

## Introductory dialog

- What? You're going to take a 3 day stat class? How boring!
- Uhhhh... Thanks
- Why?
- Because I've been told so
- By who?
- My advisor
- ?????? Why ?
- Duno ... He seems to be believe it's important
- ...

## Introductory dialog

- What? You're going to take a 3 day stat class? How boring!
- Uhhhh... Thanks
- Why?
- Because I've been told so
- By who?
- My advisor
- ?????? Why ?
- Duno ... He seems to be believe it's important
- ...

# History

Some scientific fields cannot go without statistics:



R.A Fisher 1890–1962



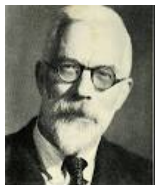
C.E. Spearman, 1863–1945

- ▶ Agronomy (field trials, genetics, seed selection, ...)
- ▶ Psychology (tests, ...)
- ▶ Medical trials
- ▶ Economics, political sciences (polls, surveys, ...)
- ▶ And Geosciences (any idea ?)

Historically, statistics was founded by non mathematicians

## History

Some scientific fields cannot go without statistics:



R.A. Fisher 1890–1962



C.E. Spearman, 1863–1945

- ▶ Agronomy (field trials, genetics, seed selection, ...)
- ▶ Psychology (tests, ... )
- ▶ Medical trials
- ▶ Economics, political sciences (polls, surveys, ...)
- ▶ And Geosciences (any idea ?)

Historically, statistics was founded by non mathematicians

# History

Some scientific fields cannot go without statistics:



R.A Fisher 1890–1962



C.E. Spearman, 1863–1945

- ▶ Agronomy (field trials, genetics, seed selection, ...)
- ▶ Psychology (tests, ... )
- ▶ Medical trials
- ▶ Economics, political sciences (polls, surveys, ...)
- ▶ And Geosciences (any idea ?)

Historically, statistics was founded by non mathematicians



# History

Some scientific fields cannot go without statistics:



R.A Fisher 1890–1962



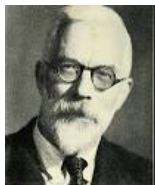
C.E. Spearman, 1863–1945

- ▶ Agronomy (field trials, genetics, seed selection, ...)
- ▶ Psychology (tests, ... )
- ▶ Medical trials
- ▶ Economics, political sciences (polls, surveys, ...)
- ▶ And Geosciences (any idea ?)

Historically, statistics was founded by non mathematicians

# History

Some scientific fields cannot go without statistics:



R.A Fisher 1890–1962



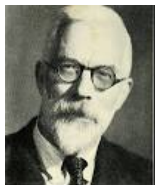
C.E. Spearman, 1863–1945

- ▶ Agronomy (field trials, genetics, seed selection, ...)
- ▶ Psychology (tests, ... )
- ▶ Medical trials
- ▶ Economics, political sciences (polls, surveys, ...)
- ▶ And Geosciences (any idea ?)

Historically, statistics was founded by non mathematicians

# History

Some scientific fields cannot go without statistics:



R.A Fisher 1890–1962



C.E. Spearman, 1863–1945

- ▶ Agronomy (field trials, genetics, seed selection, ...)
- ▶ Psychology (tests, ... )
- ▶ Medical trials
- ▶ Economics, political sciences (polls, surveys, ...)
- ▶ And Geosciences (any idea ?)

Historically, statistics was founded by non mathematicians

# History

Some scientific fields cannot go without statistics:



R.A. Fisher 1890–1962



C.E. Spearman, 1863–1945

- ▶ Agronomy (field trials, genetics, seed selection, ...)
- ▶ Psychology (tests, ... )
- ▶ Medical trials
- ▶ Economics, political sciences (polls, surveys, ...)
- ▶ And Geosciences (any idea ?)

Historically, statistics was founded by non mathematicians

# History

Some scientific fields cannot go without statistics:



R.A Fisher 1890–1962



C.E. Spearman, 1863–1945

- ▶ Agronomy (field trials, genetics, seed selection, ...)
- ▶ Psychology (tests, ... )
- ▶ Medical trials
- ▶ Economics, political sciences (polls, surveys, ...)
- ▶ And Geosciences (any idea ?)

Historically, statistics was founded by non mathematicians

# Statistical Triangle

What is statistics ?

- ▶ Statistics is about describing and analyzing data (samples)
- ▶ Using mathematic methods derived from probability theory
- ▶ In view of testing scientific hypothesis

Statistical Triangle

Data

Mathematics

Scientific hypothesis

# Statistical Triangle

What is statistics ?

- ▶ Statistics is about describing and analyzing data (samples)
- ▶ Using mathematic methods derived from probability theory
- ▶ In view of testing scientific hypothesis

Statistical Triangle

Data

Mathematics

Scientific hypothesis

# Statistical Triangle

What is statistics ?

- ▶ Statistics is about describing and analyzing data (samples)
- ▶ Using mathematic methods derived from probability theory
- ▶ In view of testing scientific hypothesis

Statistical Triangle

Data

Mathematics

Scientific hypothesis



# Statistical Triangle

What is statistics ?

- ▶ Statistics is about describing and analyzing data (samples)
- ▶ Using mathematic methods derived from probability theory
- ▶ In view of testing scientific hypothesis

## Statistical Triangle

Data

Mathematics

Scientific hypothesis

# Objectives

- ▶ Estimating characteristics of a **population**, based on **samples**
- ▶ **Testing** scientific hypothesis
- ▶ Quantifying causes of **variations**
- ▶ Conducting a statistical study
- ▶ Using R

## Some definitions

### Population



"In statistics, a population is a set of similar items or events which is of interest for some question or experiment.

A statistical population can be a group of actually existing objects (e.g. the set of all stars within the Milky Way galaxy) or a hypothetical and potentially infinite group of objects conceived as a generalization from experience (e.g. the set of all possible hands in a game of poker).

A common aim of statistical analysis is to produce information about some chosen population."

## Some definitions

A population has to be very clearly defined. For example, population of Switzerland

Catégories de personnes	Notion		
	Population résidente permanente jusqu'/au 31.12.2009	dès le 31.12.2010	Population résidente jusqu'/au 31.12.2009
<b>Personnes de nationalité suisse</b>			
Domicile civil en Suisse	inclus	...	inclus
Domicile principal en Suisse	...	inclus	...
Domicile (permanent) à l'étranger			
Domicile économique en Suisse	pas inclus	...	inclus
Domicile secondaire en Suisse	...	pas inclus	...
<b>Personnes de nationalité étrangère</b>			
Titulaires d'une autorisation d'établissement (C), d'une autorisation de séjour (B), d'une autorisation de séjour de courte durée de ≥ 12 mois (L)	inclus	inclus	inclus
Saisonniers (A) <sup>1</sup>	pas inclus	...	inclus
Titulaires d'une autorisation de séjour de courte durée de moins de 12 mois (L)	pas inclus	pas inclus	inclus
Requérants d'asile (Q) et personnes admises à titre provisoire (F)	pas inclus	partiellement inclus <sup>2</sup>	inclus
Diplomates, fonctionnaires internationaux (autorisation du DFAE)	inclus	inclus	inclus
Frontaliers (G)	pas inclus	pas inclus	pas inclus

From <http://www.media-stat.admin.ch>

## Some definitions

### Sample

A sample,  $X_1, X_2, \dots, X_n$  is a subset of a population

### Random Sample

A sample is **random** if each individual in the sample is drawn randomly

- ▶ randomly
- ▶ independently to each other

### Sampling bias

A random sample is **biased** when samples are collected in such a way that some members of the intended population are less likely to be included than others.

Examples:

- ▶ Internet surveys
- ▶ Survivorship bias
- ▶ Sampling in specific area or in "interesting areas"

# Program

1. **The statistical thinking; basic definitions**
2. Univariate statistics: inference and testing
3. Simple regression, linear model and ANOVA
4. Applications to environmental statistics: time series and geostatistics