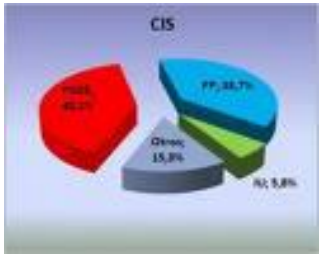
Statistics

Motivation: uses of statistics



Surveys

Estimating unemployment

Economic Predictions

Decision making

Course outline

1. Introduction.
2. Analysis of univariate data.
3. Analysis of bivariate data.
4. Probability and probabilistic models.
5. Introduction to statistical inference.

Bibliography

- D. Huff. *How to Lie with Statistics.* W.W. Norton & Company. 1993.
- D. Levitin. *A Field Guide to Lies and Statistics.* Penguin. 2018
- P. Hudson and M. Ishuzu. *History by Numbers* (2nd ed): Bloomsbury. 2016.
- P. Galderisi. *Understanding Political Science Statistics: Observations and Expectations in Political Analysis: Routledge*. 2015.
- J. Schmuller. *Statistical Analysis with Excel For Dummies*. Wiley. 2016.
- C. Carlberg. *Statistical Analysis: Microsoft Excel 2016*. Que Publishing 2017.

Blography

- Not awful and boring ideas for teaching statistics.
- Bad Statistics. Exposés of bad statistical analyses and screwed-up graphs
- Malaprensa sobre errores estadísticos en la prensa.
- You can find many more resources; notes, videos, examples … searching.

Bases de datos

- Comparative Political Data Set
- Global Health and Human Rights Database
- Political Science Resources
- Inter-university Consortium for Political and Social Research database
- UN data
- World Bank data

Chapter 1: Introduction

1. Concept and use of statistics.
2. Statistical terminology: population, sample, ...
3. Variables.
4. Types of variable.

Recommended reading:

This page explains the difference between a census and a survey.
This page compares the different approaches to sampling.
A video on What is Statistics?
Another video on basic statistical concepts

Concept and use of Statistics

- What is / are statistics?
- Dangers of using statistics.
- What are statistics used for?

What is / are statistics?

Merriam-Webster gives the following definition of statistics:

Definition of STATISTICS

1 : a branch of mathematics dealing with the collection, analysis, interpretation, and presentation of masses of numerical data

2 : a collection of quantitative data

Origin and Etymology of STATISTICS

German *Statistik* study of political facts and figures, from New Latin *statisticus* of politics, from Latin *status* state

The word Statistics is derived from the State:

There are dedicated statistical institutions at regional, national and international levels.

El Instituto de Statistics de la Comunidad de Madrid
El Instituto Nacional de Statistics
Eurostat

In Spain, various specialist organizations publish surveys related to politics.

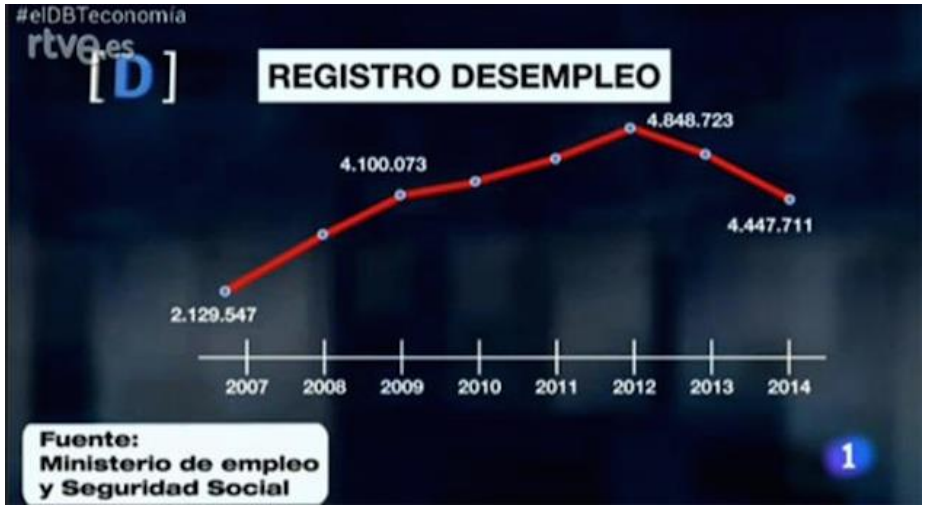
El Centro de Investigaciones Sociológicas
El Real Instituto Elcano
Metroscopia

A statistic is a single measure of some feature of a sample.

- Economic statistics: unemployment, inflation …
- Demographic statistics: birth rate, age distribution, …
- Sport statistics: goals scored, red cards, …
- Meteorological statistics: temperature, rainfall, …

Dangers of using statistics

In many cases it is possible to use statistics to influence or mislead the public.



Is there less unemployment in 2014 than 2009?

Three headlines

El Mundo, 21 February 2005, after the approval of the European Constitution in a a referendum by 76% of the voters (32% of the census)
Great victory of the 'yes' to the Constitution with a low but aceptable partcipation

El Mundo, 19 June 2006, after the approval, by 74% of the voters, of the Statute for Cataluña in a referendum (36% of the census):
Most Catalans are not interested in the Statute that defines them as a nation

El Mundo, 19 February 2007, after the approval of the Statute for Andalusia by 87,5% of the voters (31% of the census):
Only 31% of Andalusians vote in their "national reality" referendum

Fuente: Malaprensa

Mortalities due to crimes of passion

Are there more male chauvinists in Andalusia?

Source: Malaprensa

What is statistics used for?

- Collection and summary of information
- Ilustrating the relation between different variables
- Changes in a variable over time
- Estimation of the characteristics of a population using data from a sample.
- Forecasting or prediction

Descriptive statistics **Inferential statistics**

Statistical terminology

POPULATION

SAMPLE

VARIABLE

C's Ciudadanos

PODEMOS

PSOE

PP

SUB-POPULATION

INDIVIDUAL

DATUM

PODEMOS

Censuses and surveys

In a census, data are recompiled from the whole population:

If people don't lie, results are exact

Census are expensive and time consuming. Only the state has the resources to carry them out.

In Spain, a census is carried out every 10 years, to compile demographic information used to calculate the "shopping basket" used for the retail price index.

In a survey, a sample of the population is used.

It is much cheaper and faster than a census.

Results are imprecise.

How to take a sample?

We want the sample to be representative of the population:

- Non probabilistic sampling

Easy to carry out

Results can be heavily biased

- Probabilistic sampling
  - Simple random sampling
  - Stratified sampling
  - Cluster sampling

No inherent biases

A list of the whole population is required

The dangers of surveys

The results of surveys can sometimes be very inaccurate.

This article from the Guardian tells you why surveys can't always be trusted.

Variables

A variable is a characterístic of the population that we wish to study.

- Gender: male / female.
- Political party voted for by Madrileños: Cs, Podemos, PP, PSOE, other.
- The work status of Getafenses: unemployed, part time, full time.
- Number of newspapers bought per week.
- Spending of town halls in Spain.

Different types of data require different types of analysis.

Types of variable

A **qualitative variable** or **attribute** is a non-quantitative characteristic or feature of a population that we want to study.

A **quantitative variable** is one that is naturally measured or expressed numerically.

- **Gender: male / female.**
- **Political party voted for by Madrileños: Cs, Podemos, PP, PSOE, other.**
- **The work status of Getafenses: unemployed, part time, full time.**
- **Number of newspapers bought per week.**
- **Spending of town halls in Spain.**

Qualitative variables or attributes

Qualitative variables can be **dichotomous** or **polytomous**.

- **Gender: male / female.**
- **Political party voted for by Madrileños: Cs, Podemos, PP, PSOE, other.**
- **The work status of Getafenses: unemployed, part time, full time.**

Polytomous variables

Polytomous variables can be **nominal** or **ordinal**.

- **Political party voted for by Madrileños: Cs, Podemos, PP, PSOE, other.**
- **The work status of Getafenses: unemployed, part time, full time.**

Quantitative variables

Quantitative variables can be **discrete** or **continuous** (or mixed)

- **Number of newspapers bought per week.**
- **Spending of town halls in Spain.**
- **Time spent queuing before being served in a supermarket.**

Exercise

Classify the following variables:

- Town where you were born
- Total earnings
- Number of points on a driving licence
- Level of agreement with the Bologna system
- Telephone number
- Level of education
- Post code of your address

Exercise

What do you think?

La Vanguardia (13/12/2009)

“An overwhelming “yes vote” in the sovereignty consultations with 94.9%”

The final recount showed 3.2% against, 1,6% abstentions and 0.3% null votes. The coordinator defined the referéndum as “heroic” …

ABC (13/12/2009)

“There is little interest in the independence vote with less than 30% participation”

According to the data provided by the organisers, with a census of 700.000 adults, only 200000 people voted, just under 30%.

Exercise

In the next municipal elections, we are going to vote for the **party who we want to govern our municipality**. The variable marked in black is:

a) quantitative and nominal

b) qualitative and discrete

c) qualitative and nominal

d) qualitative and ordinal

Exercise (Exam question)

The following questions come from an electoral survey in India.

Z1. What is your age? _____ ***(in completed years)***

Q5. How would you rate the economic situation of the country at the moment. Is it Very Good, Good, So-so, Bad or Very Bad? 1. Very Good 2. Good 3. So-so 4. Bad 5. Very Bad

Q25. Have you heard/read about the Right to Education (RTE) Act? 2. Heard 1. Not heard

Classify the variables in the questions above.

Exercise (Exam question)

The Research Department of the Chamber of Commerce of Madrid periodically produces reports on the regional economy and measures consumer trends. You want to carry out a study to find out the number of full time workers in subcontracting companies with more than 50 workers therefore take a sample of 100 companies.

Define which companies form:

a) The population

b) The sample

c) A unit of information or individual (datum).