

Introduction to Quantitative Methods

(Master 1 EEI)

Maxime Chabriel

2024-2025

Who I am/What I do

- PhD student in Economic history at ENS de Lyon (CERGIC)
- ENSAE + Master in International Development (ScPo)
- **Contact:**
 - Office: D4-tbd
 - Mail: maxime.chabriel@proton.me

Global Organization

- 12 \times 2h class, every Thursdays
- Discover/Recall the basic quantitative tools used in social and political sciences
- Provide **intuition** on how to deal with quantitative data, and we will cover a few ways to infer results from it
- **Preparation course**, so no strict curriculum
→ The main objective is to adapt to your needs

Global Outline

- 1 General Introduction and Conceptual Tools (2h)
- 2 Univariate Analysis and Descriptive Statistics (6h)
- 3 Multivariate Analysis (6h)
- 4 Introduction to linear regressions and causal analysis (4h)
- 5 Introduction to R (4h)
- 6 Final written exam (2h)

Evaluation

- Continuous assessment (20%)
- A 5-10 pages at home report on a chosen topic (40%)
I will give the instructions at the end of the Univariate analysis section
- A final written theoretical exam (40%)

Survey time!

What are Quantitative Methods?

- Measuring
 - surveys, lexicometry, accounting...
 - building indicators (GDP, HDI, debt ratio...)
- Creating statistics
 - mean, median, quantile...
- Regression analysis (causal inference, effect size)
- Machine Learning (clustering, prediction, natural language processing)

Why use Quantitative Methods?

Pros and Cons of Quantitative Methods

Pros:

- Synthesis power
- More "formal" and persuasive (especially visually)
- Data can help to look at the big picture
- We can use mathematical theory to build confidence intervals, causal inference etc...

Cons:

- Not everything can be (relevantly) quantified
- May lack nuance or context
- We cannot ignore the production process of the data (a database is a social construct)

- → In social sciences, one cannot rely only on quantitative or qualitative data, but needs to make both sources dialogue
- In practice, the quantitative and qualitative methods of today are so advanced that no researcher is specialised in both
- → There are communication issues! Or even conflicts, as results can clash
- Famous recent example : *Peut-on faire l'économie de l'histoire ? Verdun, Vichy, et les conditions d'un dialogue entre disciplines* 2021
<https://devhist.hypotheses.org/3921>
- Hence the necessity to be literate in both methods

The necessary steps of a good quantitative analysis

- **Define** your research question / the purpose of your analysis
 - Check the literature. Has your question been answered somewhere else? If so, were you convinced by their method?
 - Do you really need to do a quantitative analysis?
- **Choose** your data
 - Compare the sources
 - Compare the details/reliability of their documentations
- **Explore, Know** your dataset, describe its variables
 - Is your data incoherent? Are there missing values?
 - New variables / peculiar observations can give you new ideas
- **Choose** your method of analysis.
 - Visualisations?
 - Indicators to build?
 - Causal model?
- **Robustness checks**
 - Everything you (or someone else) could have done differently, do it and compare your results

Good practices

- Take good notes of your data source, results, choices etc.
- Take some time cleaning and formatting your data
- This includes choosing intelligent file names and folder structure
- Save frequently!
- Do not hesitate to test things
- Be honest. If some results do not go your way, do not hide it and find possible explanations

Definitions

- **Population** : The group of individuals / objects / events of which we want to know some characteristics or implement an analysis.
- **Sample** : The subset of the population on which information has been gathered.
- **Observation** : One unit of the sample.
- **Variable** : A characteristic of the population that has been measured on the sample.
- **Sample representativity / Sample bias** : The extent to which the sample represents / does not represent the population.
- **Dataframe / Dataset** : Computer object used to store the variables and the observations.

Wooclap : wooclap.com + event code PTHGYF

- **Population ? :**

Dataset example

Figure: Screenshot from Rstudio

	Country.Name	Country.Code	year	Access to clean fuels and technologies for cooking (% of population)	Access to clean fuels and technologies for cooking, rural (% of rural population)	Access to clean fuels and technologies for cooking, urban (% of urban population)	Access to electricity (% of population)	Access to electricity, rural (% of rural population)	Access to electricity, urban (% of urban population)	Adjusted net national income (annual % growth)	Adjusted net national income (constant 2015 US\$)	Adjusted net national income (current US\$)
1	Albania	ALB	2020	83.10	66.7	93.8	100.0	100.0	100.0	-3.6930944	1.002937e+10	1.226674e+10
2	Andorra	AND	2020	100.00	100.0	100.0	100.0	100.0	100.0	NA	NA	NA
3	Austria	AUT	2020	100.00	100.0	100.0	100.0	100.0	100.0	-6.5959246	3.119575e+11	3.523549e+11
4	Belarus	BLR	2020	99.70	99.1	99.9	100.0	100.0	100.0	-1.6409753	4.726317e+10	4.878461e+10
5	Belgium	BEL	2020	100.00	100.0	100.0	100.0	100.0	100.0	-6.6003963	3.810269e+11	4.246354e+11
6	Bosnia and Herzegovina	BIH	2020	40.85	12.7	60.5	100.0	100.0	100.0	-3.3496070	1.575126e+10	1.682432e+10
7	Bulgaria	BGR	2020	NA	NA	NA	99.7	99.4	99.8	-2.1805601	4.854830e+10	5.763036e+10
8	Croatia	HRV	2020	100.00	100.0	100.0	100.0	100.0	100.0	-10.5838314	4.428972e+10	4.787618e+10
9	Cyprus	CYP	2020	100.00	100.0	100.0	100.0	100.0	100.0	-8.5636630	1.913057e+10	2.005980e+10
10	Czechia	CZE	2020	100.00	100.0	100.0	100.0	100.0	100.0	-5.4607542	1.498732e+11	1.801972e+11
11	Denmark	DNK	2020	100.00	100.0	100.0	100.0	100.0	100.0	-0.5379666	2.810831e+11	3.046444e+11
12	Estonia	EST	2020	100.00	100.0	100.0	100.0	100.0	100.0	-0.4462136	2.215057e+10	2.543788e+10
13	Faroe Islands	FRO	2020	NA	NA	NA	100.0	100.0	100.0	NA	NA	NA
14	Finland	FIN	2020	100.00	100.0	100.0	100.0	100.0	100.0	-1.0599569	2.051778e+11	2.234231e+11
15	France	FRA	2020	100.00	100.0	100.0	100.0	100.0	100.0	-9.6924857	1.974283e+12	2.157023e+12
16	Georgia	GEO	2020	89.60	79.7	97.2	100.0	100.0	100.0	-3.9469748	1.447900e+10	1.302810e+10
17	Germany	DEU	2020	100.00	100.0	100.0	100.0	100.0	100.0	-4.7255906	2.909988e+12	3.246858e+12
18	Gibraltar	GIB	2020	NA	NA	NA	100.0	100.0	100.0	NA	NA	NA
19	Greece	GRC	2020	100.00	100.0	100.0	100.0	100.0	100.0	-9.7802387	1.536829e+11	1.578551e+11

Data extracted from the World Bank's World Development Indicators

- **Population** : Countries
- **Sample ?** :

Dataset example

Figure: Screenshot from Rstudio

	Country.Name	Country.Code	year	Access to clean fuels and technologies for cooking (% of population)	Access to clean fuels and technologies for cooking, rural (% of rural population)	Access to clean fuels and technologies for cooking, urban (% of urban population)	Access to electricity (% of population)	Access to electricity, rural (% of rural population)	Access to electricity, urban (% of urban population)	Adjusted net national income (annual % growth)	Adjusted net national income (constant 2015 US\$)	Adjusted net national income (current US\$)
1	Albania	ALB	2020	83.10	66.7	93.8	100.0	100.0	100.0	-3.6930944	1.002937e+10	1.226674e+10
2	Andorra	AND	2020	100.00	100.0	100.0	100.0	100.0	100.0	NA	NA	NA
3	Austria	AUT	2020	100.00	100.0	100.0	100.0	100.0	100.0	-6.5959246	3.119575e+11	3.523549e+11
4	Belarus	BLR	2020	99.70	99.1	99.9	100.0	100.0	100.0	-1.6409753	4.726317e+10	4.878461e+10
5	Belgium	BEL	2020	100.00	100.0	100.0	100.0	100.0	100.0	-6.6003963	3.810269e+11	4.246354e+11
6	Bosnia and Herzegovina	BIH	2020	40.85	12.7	60.5	100.0	100.0	100.0	-3.3496070	1.575126e+10	1.682432e+10
7	Bulgaria	BGR	2020	NA	NA	NA	99.7	99.4	99.8	-2.1805601	4.854830e+10	5.763036e+10
8	Croatia	HRV	2020	100.00	100.0	100.0	100.0	100.0	100.0	-10.5838314	4.428972e+10	4.787618e+10
9	Cyprus	CYP	2020	100.00	100.0	100.0	100.0	100.0	100.0	-8.5636630	1.913057e+10	2.005980e+10
10	Czechia	CZE	2020	100.00	100.0	100.0	100.0	100.0	100.0	-5.4607542	1.498732e+11	1.801972e+11
11	Denmark	DNK	2020	100.00	100.0	100.0	100.0	100.0	100.0	-0.5379666	2.810831e+11	3.046444e+11
12	Estonia	EST	2020	100.00	100.0	100.0	100.0	100.0	100.0	-0.4462136	2.215057e+10	2.543788e+10
13	Faroe Islands	FRO	2020	NA	NA	NA	100.0	100.0	100.0	NA	NA	NA
14	Finland	FIN	2020	100.00	100.0	100.0	100.0	100.0	100.0	-1.0599569	2.051778e+11	2.234231e+11
15	France	FRA	2020	100.00	100.0	100.0	100.0	100.0	100.0	-9.6924857	1.974283e+12	2.157023e+12
16	Georgia	GEO	2020	89.60	79.7	97.2	100.0	100.0	100.0	-3.9469748	1.447900e+10	1.302810e+10
17	Germany	DEU	2020	100.00	100.0	100.0	100.0	100.0	100.0	-4.7255906	2.909988e+12	3.246858e+12
18	Gibraltar	GIB	2020	NA	NA	NA	100.0	100.0	100.0	NA	NA	NA
19	Greece	GRC	2020	100.00	100.0	100.0	100.0	100.0	100.0	-9.7802387	1.536829e+11	1.578551e+11

Data extracted from the World Bank's World Development Indicators

- **Population** : Countries
- **Sample** : European Countries
- **Observation ?** :
- **Variable ?** :

Dataset example

Figure: Screenshot from Rstudio

	Country.Name	Country.Code	year	Access to clean fuels and technologies for cooking (% of population)	Access to clean fuels and technologies for cooking, rural (% of rural population)	Access to clean fuels and technologies for cooking, urban (% of urban population)	Access to electricity (% of population)	Access to electricity, rural (% of rural population)	Access to electricity, urban (% of urban population)	Adjusted net national income (annual % growth)	Adjusted net national income (constant 2015 US\$)	Adjusted net national income (current US\$)
1	Albania	ALB	2020	83.10	66.7	93.8	100.0	100.0	100.0	-3.6930944	1.002937e+10	1.226674e+10
2	Andorra	AND	2020	100.00	100.0	100.0	100.0	100.0	100.0	NA	NA	NA
3	Austria	AUT	2020	100.00	100.0	100.0	100.0	100.0	100.0	-6.5959246	3.119575e+11	3.523549e+11
4	Belarus	BLR	2020	99.70	99.1	99.9	100.0	100.0	100.0	-1.6409753	4.726317e+10	4.878461e+10
5	Belgium	BEL	2020	100.00	100.0	100.0	100.0	100.0	100.0	-6.6003963	3.810269e+11	4.246354e+11
6	Bosnia and Herzegovina	BIH	2020	40.85	12.7	60.5	100.0	100.0	100.0	-3.3496070	1.575126e+10	1.682432e+10
7	Bulgaria	BGR	2020	NA	NA	NA	99.7	99.4	99.8	-2.1805601	4.854830e+10	5.763036e+10
8	Croatia	HRV	2020	100.00	100.0	100.0	100.0	100.0	100.0	-10.5838314	4.428972e+10	4.787618e+10
9	Cyprus	CYP	2020	100.00	100.0	100.0	100.0	100.0	100.0	-8.5636630	1.913057e+10	2.005980e+10
10	Czechia	CZE	2020	100.00	100.0	100.0	100.0	100.0	100.0	-5.4607542	1.498732e+11	1.801972e+11
11	Denmark	DNK	2020	100.00	100.0	100.0	100.0	100.0	100.0	-0.5379666	2.810831e+11	3.046444e+11
12	Estonia	EST	2020	100.00	100.0	100.0	100.0	100.0	100.0	-0.4462136	2.215057e+10	2.543788e+10
13	Faroe Islands	FRO	2020	NA	NA	NA	100.0	100.0	100.0	NA	NA	NA
14	Finland	FIN	2020	100.00	100.0	100.0	100.0	100.0	100.0	-1.0599569	2.051778e+11	2.234231e+11
15	France	FRA	2020	100.00	100.0	100.0	100.0	100.0	100.0	-9.6924857	1.974283e+12	2.157023e+12
16	Georgia	GEO	2020	89.60	79.7	97.2	100.0	100.0	100.0	-3.9469748	1.447900e+10	1.302810e+10
17	Germany	DEU	2020	100.00	100.0	100.0	100.0	100.0	100.0	-4.7255906	2.909988e+12	3.246858e+12
18	Gibraltar	GIB	2020	NA	NA	NA	100.0	100.0	100.0	NA	NA	NA
19	Greece	GRC	2020	100.00	100.0	100.0	100.0	100.0	100.0	-9.7802387	1.536829e+11	1.578551e+11

Data extracted from the World Bank's World Development Indicators

- **Population** : Countries
- **Sample** : European Countries
- **Observation** : One line of the dataframe
- **Variable** : One column of the dataframe
- **Sample representativity / Sample bias ?** :

Dataset example

Figure: Screenshot from Rstudio

	Country.Name	Country.Code	year	Access to clean fuels and technologies for cooking (% of population)	Access to clean fuels and technologies for cooking, rural (% of rural population)	Access to clean fuels and technologies for cooking, urban (% of urban population)	Access to electricity (% of population)	Access to electricity, rural (% of rural population)	Access to electricity, urban (% of urban population)	Adjusted net national income (annual % growth)	Adjusted net national income (constant 2015 US\$)	Adjusted net national income (current US\$)
1	Albania	ALB	2020	83.10	66.7	93.8	100.0	100.0	100.0	-3.6930944	1.002937e+10	1.226674e+10
2	Andorra	AND	2020	100.00	100.0	100.0	100.0	100.0	100.0	NA	NA	NA
3	Austria	AUT	2020	100.00	100.0	100.0	100.0	100.0	100.0	-6.5959246	3.119575e+11	3.523549e+11
4	Belarus	BLR	2020	99.70	99.1	99.9	100.0	100.0	100.0	-1.6409753	4.726317e+10	4.878461e+10
5	Belgium	BEL	2020	100.00	100.0	100.0	100.0	100.0	100.0	-6.6003963	3.810269e+11	4.246354e+11
6	Bosnia and Herzegovina	BIH	2020	40.85	12.7	60.5	100.0	100.0	100.0	-3.3496070	1.575126e+10	1.682432e+10
7	Bulgaria	BGR	2020	NA	NA	NA	99.7	99.4	99.8	-2.1805601	4.854830e+10	5.763036e+10
8	Croatia	HRV	2020	100.00	100.0	100.0	100.0	100.0	100.0	-10.5838314	4.428972e+10	4.787618e+10
9	Cyprus	CYP	2020	100.00	100.0	100.0	100.0	100.0	100.0	-8.5636630	1.913057e+10	2.005980e+10
10	Czechia	CZE	2020	100.00	100.0	100.0	100.0	100.0	100.0	-5.4607542	1.498732e+11	1.801972e+11
11	Denmark	DNK	2020	100.00	100.0	100.0	100.0	100.0	100.0	-0.5379666	2.810831e+11	3.046444e+11
12	Estonia	EST	2020	100.00	100.0	100.0	100.0	100.0	100.0	-0.4462136	2.215057e+10	2.543788e+10
13	Faroe Islands	FRO	2020	NA	NA	NA	100.0	100.0	100.0	NA	NA	NA
14	Finland	FIN	2020	100.00	100.0	100.0	100.0	100.0	100.0	-1.0599569	2.051778e+11	2.234231e+11
15	France	FRA	2020	100.00	100.0	100.0	100.0	100.0	100.0	-9.6924857	1.974283e+12	2.157023e+12
16	Georgia	GEO	2020	89.60	79.7	97.2	100.0	100.0	100.0	-3.9469748	1.447900e+10	1.302810e+10
17	Germany	DEU	2020	100.00	100.0	100.0	100.0	100.0	100.0	-4.7255906	2.909988e+12	3.246858e+12
18	Gibraltar	GIB	2020	NA	NA	NA	100.0	100.0	100.0	NA	NA	NA
19	Greece	GRC	2020	100.00	100.0	100.0	100.0	100.0	100.0	-9.7802387	1.536829e+11	1.578551e+11

Data extracted from the World Bank's World Development Indicators

- **Population** : Countries
- **Sample** : European Countries
- **Observation** : One line of the dataframe
- **Variable** : One column of the dataframe
- **Sample representativity / Sample bias** : Mostly rich/developped countries
- **Dataframe / Dataset** : World Development Indicators
<https://datacatalog.worldbank.org/search/dataset/0037712/World-Development-Indicators>

Variable types

- **Qualitative variable** : A variable that cannot be represented by a number
 - **Nominal categorical variable** (continent, dominant religion, currency, political regime)
 - **Ordinal categorical variable** : When categories can be hierarchised

Variable types

- **Qualitative variable** : A variable that cannot be represented by a number
 - **Nominal categorical variable** (continent, dominant religion, currency, political regime)
 - **Ordinal categorical variable** : When categories can be hierarchised (income level, political regime)
- **Quantitative variable**
 - **Discrete variable** (rank, number of neighbors, year)
 - **Continuous variable** (income, land size, time)

Variable types

- **Qualitative variable** : A variable that is not represented by numeric values
 - **Nominal categorical variable** (continent, dominant religion, currency, political regime)
 - **Ordinal categorical variable** : When categories are hierarchised (income level, political regime)
- **Quantitative variable**
 - **Discrete variable** (rank, number of neighbors, year)
 - **Continuous variable** (income, land size, time, population(!))
 - **NB** Any Quantitative variable can be converted to an Ordinal categorical variable using intervals (ex: income levels)
- **Other** :
 - **Dummy / Boolean** : Takes the value 1 or 0 to decompose categorical variables. Or to code yes/no questions (landlocked, free elections, war)
 - **Mixed** (year of decolonisation, age of elected leader). Also, some people like to encode their missing values with the numbers -1 or 99