

Organizational session / Introduction to quantitative analysis

Lukáš Lehotský & Petr Ocelík

Outline

- Course plan
- Background: quantitative research and statistics
- Data, variable types and levels of measurement

Methodology/methods courses

- Logic built around gradual expansion of method knowledge.
- Extension courses
 - MEB433 – Introduction to quantitative data analysis (Fall 2017)
 - MEB434 – Social network analysis (Fall 2017)
 - HEN645 – Experiment v sociálních vědách (Fall 2017)
 - MEB421 – Vybrané metody výzkumu v mezinárodních vztazích (Fall 2017)
- Support courses
 - ESS412 – Academic skills review (Spring 2017)
- Methods not particularly entertaining but absolutely crucial.

Course requirements

- Read required readings
- Attend lectures

- 2 groups
- Write research proposal
- Write feedback report
- Deliver presentation of the research proposal
- Pass a written exam

Course requirements

- Deadlines
 - April 20 – submission of proposal (group 1)
 - April 22 – submission of feedback report (group 2)
 - April 27 or 28 (TBC) – optional consultation with lecturers (group 1)
 - April 30 – submission of final version of proposal (group 1)

 - May 11 – submission of proposal (group 2)
 - May 13 – submission of feedback report (group 1)
 - May 18 or 19 (TBC) – optional consultation with lecturers (group 2)
 - May 21 – submission of final version of proposal (group 2)

Grading

- Proposal: 20
- Final exam: 20
- Grading
 - A 40 - 38 points
 - B 37 - 36 points
 - C 35 - 34 points
 - D 33 - 32 points
 - E 31 - 30 points
 - F less than 30 points

(Social) research

- SR aims to **advance knowledge** in its field by using **scientific method**.
- The procedures are public.
- The conclusions are uncertain.
- The content of science is its method.

Two meanings of methodology

- 1) Methodology as a **scientific discipline** that studies instruments and practices of scientific research.
- 2) Methodology as a **system of rules of scientific conduct** which conditions specific research design.
- There is **plurality of scientific methodologies**.
- Methodology connects **theory and investigated phenomena**.
- **Method**: an instrument for collection and/or analysis of data that follows specific theoretical-methodological trajectory.

Quantitative and qualitative methodology

Quantitative	Qualitative
Use of mathematical apparatus Numerical data	Mathematical apparatus not necessary Numerical data not necessary
Large(r) number of cases Low(er) number of variables	Low(er) number of cases Large(r) number of variables
Standardization and reduction of info Fixed research designs	Idiosyncrasy and richness of info Flexible research designs
Higher reliability Lower validity Higher generalizability	Lower reliability Higher validity Lower generalizability

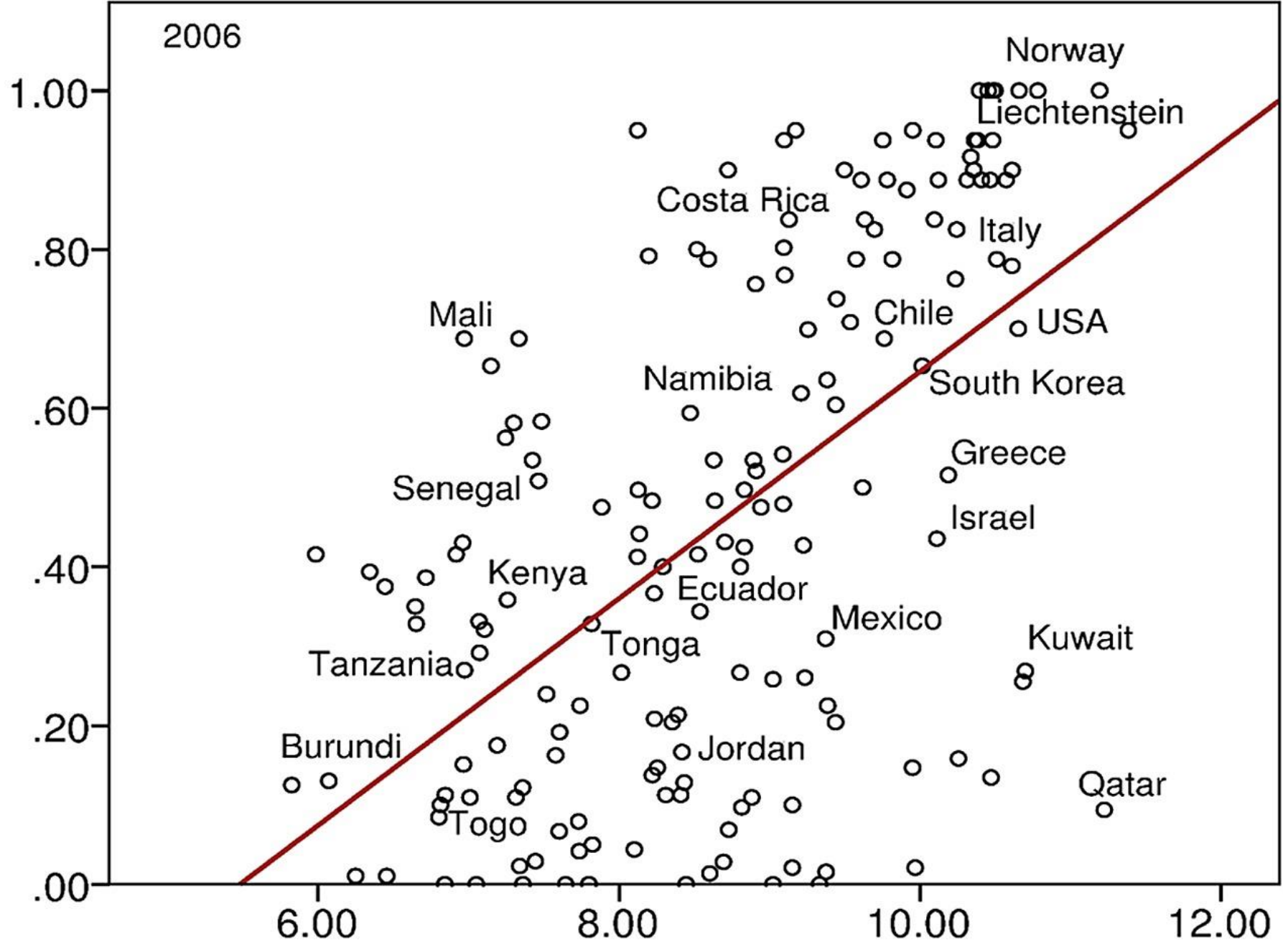
- Just a didactic tool! (Borderline cases such as QCA)
- Outdated dichotomy

Quantitative methodology

- QM is a systematic investigation of phenomena using mathematical techniques.
- QM explains phenomena by collecting numerical data that are analyzed using mathematically based methods (Cresswell 1994).
- **Always:** numerical data, mathematical framework.
- **Typically:** empirical focus, large(r) N.

2006

Democracy



log GDP per capita

Quantitative research = statistics?

- Not really.
- **Formal modeling:** e.g. game theory
- **Generative methods:** e.g. agent-based modeling.
- **But:** statistics most widespread and influential.

Table 1.2. Mentions of “scientific” terms in various disciplines, 1990–9 (%)

	Causal/causality	Hypothesis	Explanation	Any of these words
Ecology & evolutionary biology	7.9	50.0	31.7	62.3
Sociology	25.1	37.3	40.6	61.7
Economics	10.0	44.6	36.1	61.2
Political science	17.2	27.3	38.0	52.2
Anthropology	12.5	22.8	39.1	52.0
History	8.1	14.3	36.4	44.6
Film studies	5.8	5.3	19.3	25.4
Performing arts	3.6	4.8	18.5	23.8

What do we mean by statistics?

- Statistics consists of a body of methods for obtaining and analyzing data (Agresti & Finlay 2009: 3).
- Statistics is a branch of mathematics dealing with the collection, analysis, interpretation, and organization of data (Wikipedia 2017).
- ...especially analysis of population characteristics by inference from sampling (American Heritage Dictionary 2017).

What is statistics good for?

- **Data description/exploration:** quantifications, relationships, patterns, trends, outliers...
- **Theory testing:** is the relationship (+ GDP → + democracy) visible enough, or is it just random “noise”?
- **Generalization to population:** based on sample estimations we can learn about a broader set of cases (a population).
- Specific kind of an “**open-source thinking**”.

What is statistics good for?

- **Data description/exploration:**
 - What is mean value of energy efficiency in the EU countries?
 - Can we group the EU member states based on their energy efficiency **and** liberalization of their energy market?
- **Theory testing:**
 - Does increase in oil production decrease level of democracy?
(Resource curse theory)
- **Generalization to population:**
 - What are attitudes of Czech adult population towards climate change?

What is statistics good for?

- Not only academic indulgence, **statistics are everywhere** – (near to) every public as well as private organization uses numerical data.
- With rapidly progressing digitalization, there will be (much) more...
- High demand for data analysts → increases your chances on job market.

How to learn statistics?

- **Gradually:** build your skills step-by-step.
- **Consecutively:** more advanced concepts and techniques are dependent on simpler ones → do not skip!
- **Via multiple sources:** freely available textbooks, lectures on YT, community forums, etc.

Data

- Concept of data usually taken for granted.
- Data is **information that has been collected and recorded.**
- What we understand as data depends on our philosophical position.

Types of data

- Qualitative vs. quantitative data.
- Typically described as words vs. numbers.



- Blaikie: all primary data start as words (Blaikie 2003).

Data

- What does data consist of?
- **Case:** is a unit of observation / analysis.
- **Variable:** is a concept that can have different mutually exclusive values.
- **Value / score:** particular category or point on a measurement scale.

Example of coding: variable stat.boredom

- Do you agree with following statement?
Statistics is boring.
- Strongly disagree = 1
- Disagree = 2
- Nor disagree nor agree = 3
- Agree = 4
- Strongly agree = 5

Case, variable, value

Ordinary usage	Science	Statistics
Object	Unit of analysis Unit of observation	Case
Property	Attribute	Variable
Specific property	Attribute level	Value

Case-by-variable matrix

The diagram illustrates a case-by-variable matrix. The word "Cases" is on the left with three arrows pointing to the first three rows of the matrix. The word "Variables" is at the top with three arrows pointing to the first three columns of the matrix. The matrix has a header row for variables (V1, V2, V3, V4) and a header column for cases (1, 2, i, ..., n). The cells contain values x_{ij} or ellipses. The word "Values" is centered in the bottom-right area of the matrix.

	V1	V2	V3	V4
1	x_{11}	x_{21}	...	x_{m1}
2	x_{12}	x_{22}	...	x_{m2}
i	x_{1i}	x_{2i}	...	x_{mi}
...	Values	...
n	x_{1n}	x_{2n}	...	x_{mn}

Discrete and continuous variables

- **Discrete variables:** separate values without intermediate values.
 - Categories (single, married, divorced)
 - Whole numbers (0, 1, 2, ...)
- **Continuous variables:** any value over certain interval.
 - Real numbers (0, 1, $1/3$, 0.333, 3.14, ...)

Levels of measurement

- **Categorical measurement:**
 - Assigns entity to a discrete category.
- **Metric measurement:**
 - Assigns entity to a position at a given numerical scale.

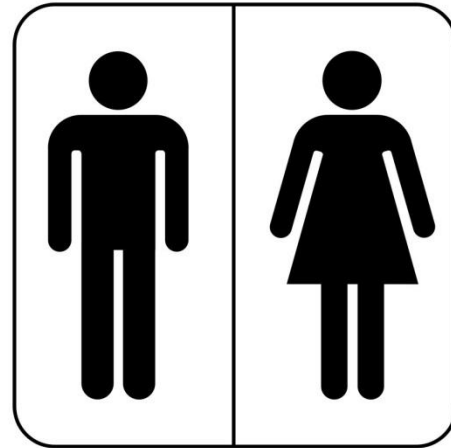
Levels of measurement

- Categorical measurement:
 - Nominal
 - Ordinal
- Metric measurement:
 - Interval
 - Ratio

Nominal measurement

- Construction of categories must be:

- Homogeneous
- Mutually exclusive
- Exhaustive



- We can arbitrarily assign numbers to categories.

Ordinal measurement

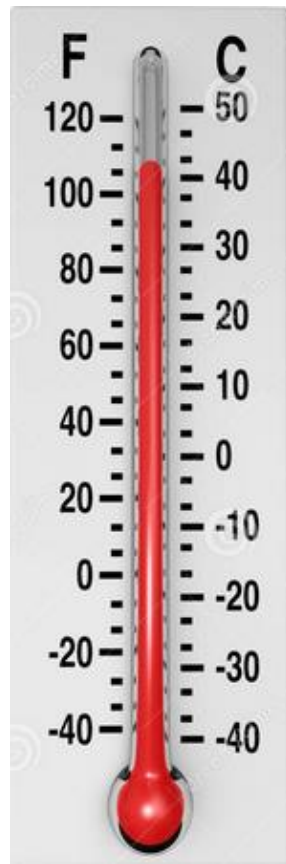
- Defined by same conditions as nominal level plus **orders categories along some dimension.**
- We can order categories, but distance between them is not equal.



- Numbers indicate only order of categories.

Interval measurement

- Defined by same conditions as ordinal level plus **scores on a scale are at the same distance apart.**
- We can measure and compare intervals.
- **No true zero:** position of zero is arbitrary.
- Measures how many (counts), not how much (ratios).



Ratio measurement

- Defined by same conditions as interval level plus it has **true zero**.
- We can measure and compare ratios / proportions.
- Degrees Celsius (interval) vs. degrees Kelvin (ratio).



Levels of measurement: recap

Measurement level	Comparison of characteristics	Comparison of values	Examples
Nominal	same/different	$a = b, a \neq b$	Nationality Energy sources
Ordinal	bigger/smaller	$a < b, a > b, a = b$	Level of democracy Likert scale
Interval	differences	$a - b = c - d$	Years Temperature in °C
Ratio	ratios	$a/b = c/d$	GDP Energy consumption Temperature in °K