

Introduction: Data & measurement

Johan A. Elkink

University College Dublin

9 September 2013

- 1 Introduction to course
- 2 Data
- 3 Measurement
- 4 Introduction to SPSS

Outline

- 1 Introduction to course
- 2 Data
- 3 Measurement
- 4 Introduction to SPSS

Definition: “ N ”

“ N ” refers to the number of cases being studied, at the unit of analysis level.

Qualitative	Case studies	$N = 1$
Comparative methods	small N	
Quantitative	Large N	large N

Descriptive vs inferential

Descriptive statistics: numerically or graphically summarizing a specific set of data.

Inferential statistics: drawing conclusions about a population on the basis of numerical or graphical information on a subset of the population.

End goal

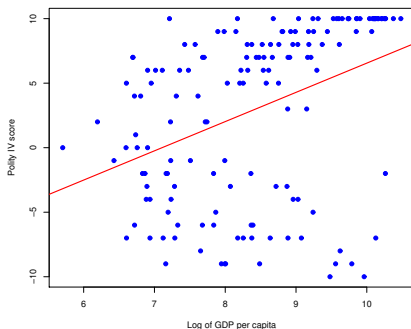


Figure : Level of democracy by GDP per capita, 2000

Introductory comments

- Objectives

Introductory comments

- Objectives
- Syllabus

Introductory comments

- Objectives
- Syllabus
- Grading and homework

Introductory comments

- Objectives
- Syllabus
- Grading and homework
- Plagiarism

Introductory comments

- Objectives
- Syllabus
- Grading and homework
- Plagiarism
- Textbook

Software

- SPSS
- Stata
- R
- etc.

Software

- SPSS
- Stata
- R
- etc.

Graphical user interfaces vs typed commands

Software

- SPSS .sps files
- Stata .do files
- R .R files
- etc.

Graphical user interfaces vs typed commands

Outline

- 1 Introduction to course
- 2 Data**
- 3 Measurement
- 4 Introduction to SPSS

Unit of analysis

The **unit of analysis** refers to the level of the observations at which you are drawing conclusions.

Are older people more likely to vote?

Are richer countries more likely to be democratic?

Does district magnitude affect proportionality?

Do rural areas have lower turnout?

Unit of analysis

The **unit of analysis** refers to the level of the observations at which you are drawing conclusions.

Are older people more likely to vote?

Are richer countries more likely to be democratic?

Does district magnitude affect proportionality?

Do rural areas have lower turnout?

individual

Unit of analysis

The **unit of analysis** refers to the level of the observations at which you are drawing conclusions.

Are older people more likely to vote?

Are richer countries more likely to be democratic?

Does district magnitude affect proportionality?

Do rural areas have lower turnout?

individual
country

Unit of analysis

The **unit of analysis** refers to the level of the observations at which you are drawing conclusions.

Are older people more likely to vote?

Are richer countries more likely to be democratic?

Does district magnitude affect proportionality?

Do rural areas have lower turnout?

individual
country

country

Unit of analysis

The **unit of analysis** refers to the level of the observations at which you are drawing conclusions.

Are older people more likely to vote?

Are richer countries more likely to be democratic?

Does district magnitude affect proportionality?

Do rural areas have lower turnout?

individual
country

country
electoral districts

Definition: “ N ”

“ N ” refers to the number of cases being studied, at the **unit of analysis** level.

Qualitative	Case studies	$N = 1$
Comparative methods	small N	
Quantitative	Large N	large N

Example data set

	Age	Vote	Party	Education	Sex
1	21	Yes	FF	4	Male
2	30	No		3	Female
3	80	Yes	FG	3	Male
4	50	Yes	Lab	2	Male
5	33	No		5	Female
6	20	No		2	Female
7	43	Yes	FF	5	Female
8	42	Yes	FF	2	Male

Example data set

	System	D.Magnitude	Seats	Prop
1	PR	10	80	0.8
2	PR	150	150	0.9
3	STV	9	100	0.8
4	FPTP	1	300	0.4
5	FPTP	1	600	0.5
6	PR	3	200	0.7

Missing values

In observed data, there are often missing values - particular data that is not available for particular cases.

Generally, these need to be excluded from statistical analysis and thus identified in the data set.

Example data set

	Age	Vote	Party	Education	Sex
1	21	Yes	FF	4	Male
2	30	No		3	Female
3	80	Yes	FG	3	Male
4	50	Yes	Lab	2	Male
5	33	No		5	Female
6	20	No		2	Female
7	43	Yes	FF	5	Female
8	42	Yes	FF	2	Male

Example data set (missing)

	Age	Vote	Party	Education	Sex
1	21	Yes	FF	4	Male
2	30	...		3	Female
3	80	Yes	FG	3	Male
4	50	Yes	Lab	2	Male
5	33	No	
6	20	No		2	Female
7	43	Yes	FF	5	Female
8	42	Yes	FF	2	...

Example data set (missing)

	Age	Vote	Party	Education	Sex
1	21	Yes	FF	4	Male
2	30	3	Female
3	80	Yes	FG	3	Male
4	50	Yes	Lab	2	Male
5	33	No
6	20	No	...	2	Female
7	43	Yes	FF	5	Female
8	42	Yes	FF	2	...

Variables

“A **variable** is an attribute that has two or more divisions, characteristics, or categories.”

(Argyrous 1997: 3)

Variables

“A **variable** is an attribute that has two or more divisions, characteristics, or categories.”

The opposite is a **constant**, which is “an attribute that does not vary.”

(Argyrous 1997: 3)

Random variables

A **random variable** assigns a particular numerical value to each *possible* outcome of an experiment or random phenomenon.

(Mood, Graybill & Boes 1974: 53; Moore, McCabe & Craig: 245)

Random variables

A **random variable** assigns a particular numerical value to each *possible* outcome of an experiment or random phenomenon.

A **realized** or **observed variable** is the actual value of the variable after the experiment or phenomenon.

(Mood, Graybill & Boes 1974: 53; Moore, McCabe & Craig: 245)

Random variables

A **random variable** assigns a particular numerical value to each *possible* outcome of an experiment or random phenomenon.

A **realized** or **observed variable** is the actual value of the variable after the experiment or phenomenon.

What you see in a data set are thus the observed or measured values on a particular underlying random variable.

(Mood, Graybill & Boes 1974: 53; Moore, McCabe & Craig: 245)

Outline

- 1 Introduction to course
- 2 Data
- 3 Measurement**
- 4 Introduction to SPSS

Definition

Conceptualisation: defining the variable of interest in qualitative or substantive terms.

(Argyrous 1997:5-6)

Definition

Conceptualisation: defining the variable of interest in qualitative or substantive terms.

Operationalisation: defining the variable in terms of “the operations used to measure a variable for individual cases.”

(Argyrous 1997:5-6)

Measurement

“**Measurement** is the process of determining and recording which of the possible traits of a variable an individual case exhibits or possesses.”

(Argyrous 1997: 3-4)

Measurement

“**Measurement** is the process of determining and recording which of the possible traits of a variable an individual case exhibits or possesses.”

“A **case** is an entity that displays or possesses the traits of a given variable.”

(Argyrous 1997: 3-4)

Measurement

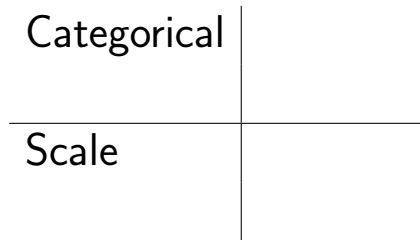
“**Measurement** is the process of determining and recording which of the possible traits of a variable an individual case exhibits or possesses.”

“A **case** is an entity that displays or possesses the traits of a given variable.”

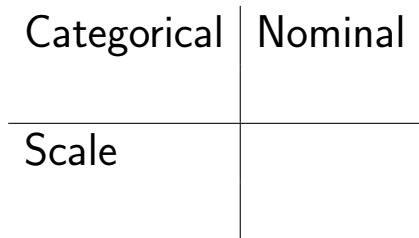
“A **population** is the set of all cases of interest. A **sample** is a subset of the **population**.”

(Argyrous 1997: 3-4)

Levels of measurement



Levels of measurement



Levels of measurement

Categorical	Nominal
	Ordinal
Scale	

Levels of measurement

Categorical	Nominal
	Ordinal
Scale	Interval

Levels of measurement

Categorical	Nominal
	Ordinal
Scale	Interval
	Ratio

Levels of measurement (examples)

- Age

Levels of measurement (examples)

- Age
- Religious denomination

Levels of measurement (examples)

- Age
- Religious denomination
- Temperature

Levels of measurement (examples)

- Age
- Religious denomination
- Temperature
- Party preference

Levels of measurement (examples)

- Age
- Religious denomination
- Temperature
- Party preference
- Level of education

Levels of measurement (examples)

- Age
- Religious denomination
- Temperature
- Party preference
- Level of education
- GDP per capita

Levels of measurement (examples)

- Age
- Religious denomination
- Temperature
- Party preference
- Level of education
- GDP per capita
- Gender

Levels of measurement (examples)

- Age
- Religious denomination
- Temperature
- Party preference
- Level of education
- GDP per capita
- Gender
- Strongly disagree, disagree, neutral, agree, strongly agree

Levels of measurement (examples)

- Age
- Religious denomination
- Temperature
- Party preference
- Level of education
- GDP per capita
- Gender
- Strongly disagree, disagree, neutral, agree, strongly agree (**Likert scale**)

Discrete vs continuous

“A **discrete variable** is measured by a unit that cannot be subdivided. It has a countable number of values.”

(Argyrous 1997: 11)

Discrete vs continuous

“A **discrete variable** is measured by a unit that cannot be subdivided. It has a countable number of values.”

“A **continuous variable** is measured by units that can be subdivided infinitely. It can take any value in a line interval.”

(Argyrous 1997: 11)

Discrete vs continuous

“A **discrete variable** is measured by a unit that cannot be subdivided. It has a countable number of values.”

“A **continuous variable** is measured by units that can be subdivided infinitely. It can take any value in a line interval.”

(Argyrous 1997: 11)

Therefore, categorical variables are always discrete, scale variables can be either discrete or continuous.

Percentages and proportions

A **proportion** is calculated as the number of cases in a particular category (n) divided by the total number of cases (N): $\frac{n}{N}$.

Percentages and proportions

A **proportion** is calculated as the number of cases in a particular category (n) divided by the total number of cases (N): $\frac{n}{N}$.

A **percentage** is calculated as the proportion times 100%: $\frac{n}{N} \cdot 100\%$.

Percentages and proportions

A **proportion** is calculated as the number of cases in a particular category (n) divided by the total number of cases (N): $\frac{n}{N}$.

A **percentage** is calculated as the proportion times 100%: $\frac{n}{N} \cdot 100\%$.

E.g. 3 out of 20 is $\frac{3}{20} = 0.15 = 15\%$.

Exercise: proportions

- What proportion of crimes in Town A relate to burglary?
- Which town has the highest homicide rate?

	Town A	Town B
Population	20,109	764,213
Homicide	13	78
Robbery	102	617
Auto theft	125	314
Rape	23	79
Burglary	178	537

(Healey 1996: 52)

Outline

- 1 Introduction to course
- 2 Data
- 3 Measurement
- 4 Introduction to SPSS**

SPSS components

- UCD's *Software for U*

SPSS components

- UCD's *Software for U*
- Data screen

SPSS components

- UCD's *Software for U*
- Data screen
- Variables screen

SPSS components

- UCD's *Software for U*
- Data screen
- Variables screen
- Output screen

SPSS components

- UCD's *Software for U*
- Data screen
- Variables screen
- Output screen
- Syntax screen

SPSS components

- UCD's *Software for U*
- Data screen
- Variables screen
- Output screen
- Syntax screen . . . “Paste”

Exercise: create an SPSS data set

Civil wars in 2002 with more than 1000 battle deaths:

Country	Polity IV	Regime	Deaths	ELF	Cleavage
Burundi			1100	0.313	No
Colombia	7	Dem	2417	0.596	Yes
India	9	Dem	2795	0.878	Yes
Nepal	6	Dem	2500	0.659	Yes
Sudan	-7	Aut	2000	0.731	Yes
Uganda	-4	Aut	1032	0.922	

- Variable names and labels
- Definition of categories and missing values
- Definition of measurement level

Cleavage = religion of rebels different from government.