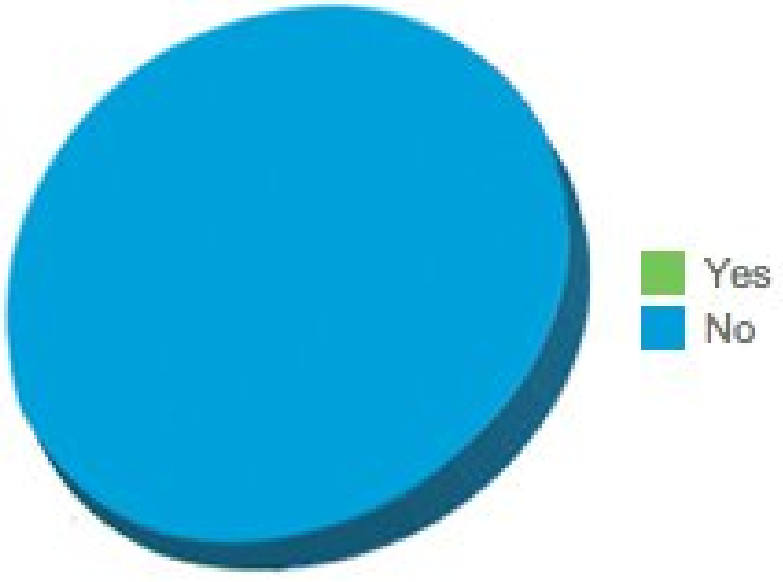
Variables



Statistical Reasoning
and Quantitative Methods

François Briatte & Ivaylo Petev

Session 3

Outline

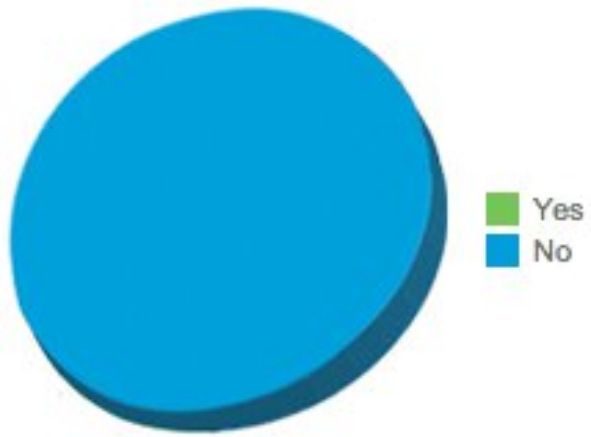
A variable is something that varies over $k > 1$ states. Quantitative data is stored as variables of different types. Each type corresponds to a scale of measurement and obeys a particular numeric coding.

Before we go further, a word on how the class will work from now on.

Routines

Variables

Recoding



$k = 1$ (constant)

Work routines

This class works through three routines:

- The **class** routine includes a 'slides session' and a 'do-file session'.
- The **homework** routine includes readings and replication.
- The **group** routine includes two draft papers and the final paper.

The last routine is the only one directly considered for grading, and within it, the **final paper** itself is the primary benchmark.

Sections 1–4 of the Stata Guide cover the course requirements, including **computer skills** (Sec. 2) and **Stata** (Sec. 3).

URL: `http://f.briatte.org/teaching/quanti/#guide`

Class routine

From Session 4 onwards, all classes include:

- **A slide session**, to review statistical theory, learn Stata commands and go over course logistics;
- **A practice session**, to work on an empirical example, familiarize with a set of Stata commands and build analytical skills; and
- **A decent break**, because we are humans.

Check the course **syllabus** or the course **website** for the topics.

The logic of this course requires you to

1. attend all sessions,
2. replicate them at home, and
3. read all assigned material (in that order).

Class routine: Details

When you get to class:

- **Open Stata and set the working directory** to the `SRQM` folder, as covered in `week1.do` with other Stata settings.
- **Get the do-file for this session** from the course website. Copy and paste it to a new-do-file window, or download it.

Check the Stata Guide for computer skill requirements (Sec. 2) and further help with Stata (Sec. 3).

If you attend all sessions and replicate them once at home, you will get enough training before starting to work on your research project.

We naturally recommend that you catch up any class that you might miss, since every session introduces new content.

Homework routine

Each course session should be replicated at home:

- **Fetch the last course email**: this is where we send most information on course logistics.
- **Read the Stata Guide and handbook chapters**: both readings go hand-in-hand.
- **Replicate the course session**: make sure that you have understood the last session in full.

Check the course **syllabus** or the course **website** for the readings.

Each week generally has one chapter from Feinstein and Thomas' *Making History Count* and at least one section of the Stata Guide.

Homework routine: Details

All do-files are archived online:

`http://f.briatte.org/teaching/quanti/`

- Download each do-file to your `Replication` folder, and make sure that all course datasets are stored in the `Datasets` folder.
- Open Stata, set your working directory to the **SRQM** folder and adjust other settings as shown in `week1.do`.
- To replicate, type e.g. `doedit "Replication/week2.do"` and run it while reading the comments.

Check the **Stata tutorials** for help with setup and commands.

Replicating takes 20 minutes at most if you have all files ready and are familiar enough with Stata. Practice does it all.

Group routine

You are required to work on a research project:

- **Form a pair**: find someone to work with on a common theme.
- **Play around**: open the course datasets and inspect their content.
- **Code it**: start writing a do-file to analyse a set of variables.

Check the **Stata Guide**, Sec. 1–4 and Sec. 13–16, for instructions.

- Your **do-file** should draw from the ones that we cover in class, which is why attendance and replication count so much.
- Your **paper** will use the same terminology, reasoning and concepts as those that we practice in each course session.
- Your **final paper and do-file** will be graded twice as drafts during the semester and once at the end.

Additional helpers

To help you work on all routines:

- **Read the course planning**: `http://goo.gl/BJHkQ`
- **Discuss your project**: `http://goo.gl/brYmB`
- **Look at a template**: `http://goo.gl/7u8oa`

All links point to course material hosted by Google Documents, which you can comment and/or edit.

Check the **course syllabus** and **Stata Guide** for additional details, which we will also introduce gradually in class.

The **course slides** also contain some information on how we organized the course: make sure to save them for later reviewing.

If completely lost, email us—but look in the course material first.

Variables

The essential element of quantitative analysis is called a variable. Datasets are series of variables for a given set of units.

The most important properties of variables are:

- Variables have **values** such as -1, 0, 1, 27 and so on.
- Variables have **labels** when their values require explanation.
- Variables have **missing values**; Stata codes them as 'dots' (.).
- Variables are **manipulable**: you can modify their values and labels.

Check the Stata Guide, Sec. 5–6, for more information on variables.

Make sure, in particular, that you understand **continuous** and **categorical** variable types. Basically, anything that can be understood on an **ordinal scale** is continuous for our purposes.

Recoding

Manipulating variables is essential to analysis for several reasons:

- **Missing values** should be properly encoded in '`.`' Stata format.
- **Groups** can be formed out of existing measures.
- **Measurements** can be simplified, combined, expanded...

Turn to the `replace` and `recode` commands to manipulate variables.

In class, pay attention to terms like 'continuous data', 'dummies' or 'nominal variables', which all point to different variable types.

At home, make sure that you understand the types of the variables that you are handling, and that missing values are correctly encoded.