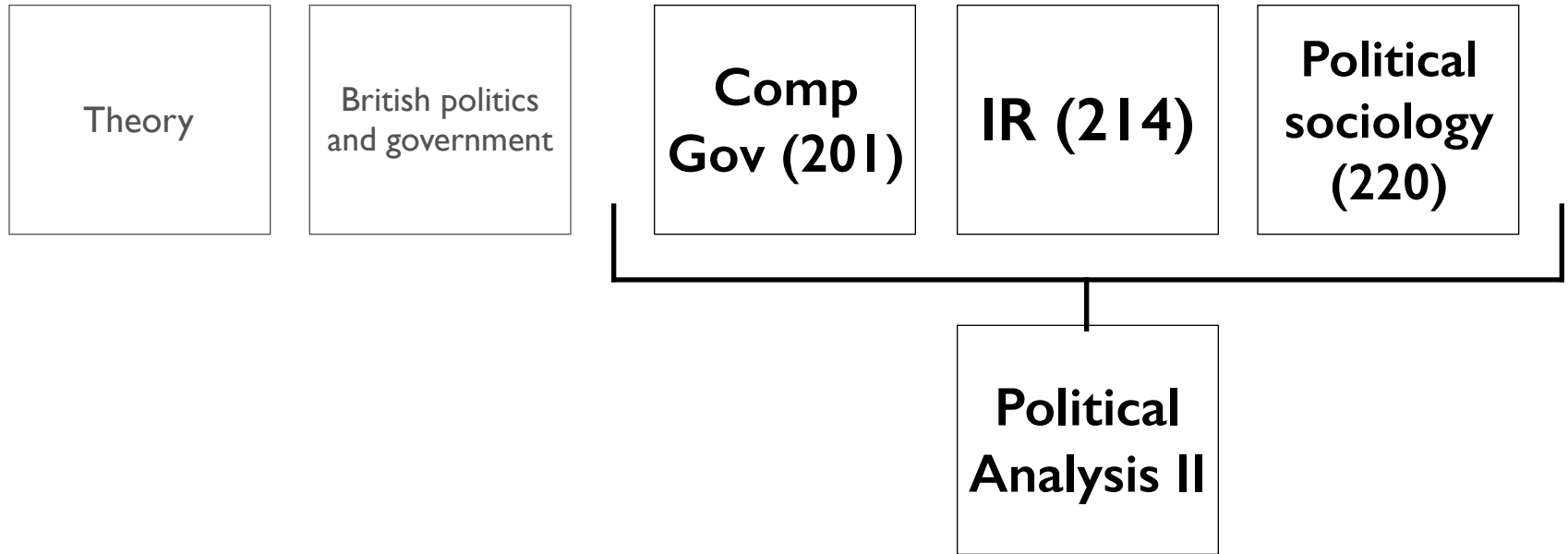# Review and Preview

Political Analysis 2: Week 1

14 October, 2015

Prof. Andrew Eggers

Political Analysis 2 is a new component of 3 core papers in Politics

Theory	British politics and government	**Comp Gov (201)**	**IR (214)**	**Political sociology (220)**

Political Analysis II

Before:
1. Uncoordinated efforts among tutors/lecturers to teach skills (e.g. how to read a regression table)
2. Varied student outcomes; many simply ignoring numbers in papers

After:
1. A single set of lectures and labs to provide the key skills
2. Step-change in students' ability (e.g. in finals) to critically assess quantitative evidence

Our goals for you

**At Oxford:** In tutorial essays and exams, you should be able to

1. Explain briefly the basis for empirical claims: *What is the comparison being made? What type of design is it (e.g. cross-country regression, field experiment)?*

2. Highlight potential shortcomings and weigh evidence: *Under what assumptions does a given study tell us what we want to know? Which claims are more credible and why?*
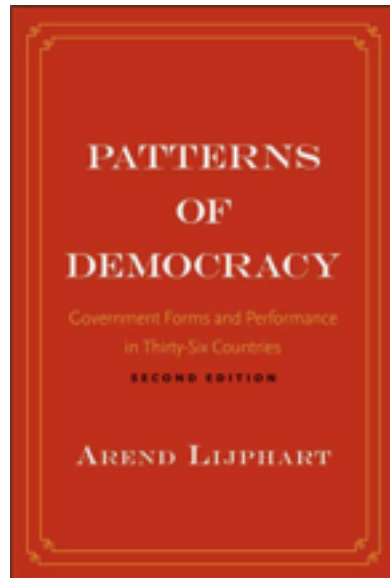
Not just reporting who said what.

After Oxford: You should be able to

1. Examine quantitative evidence and make a decision

2. Produce your own analysis of a new problem (or learn what you need to know to do it)
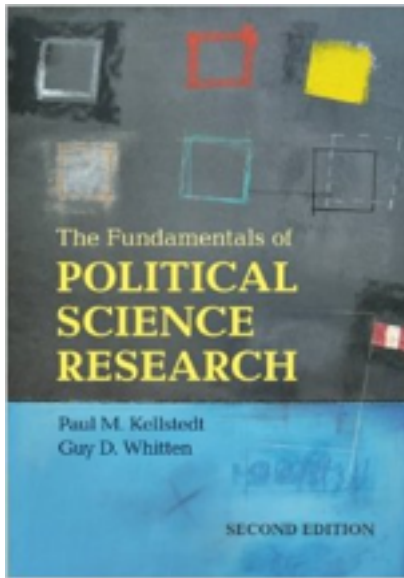
Main text for content

Main context for application

A table from Lijphart:

270   EFFECTIVE GOVERNMENT AND POLICY-MAKING

TABLE 15.2

Multivariate regression analyses of the effect of consensus democracy (executives-parties dimension) on five indicators of violence, with controls for the effects of the level of economic development, logged population size, and degree of societal division, and with extreme outliers removed

Performance variables	Estimated regression coefficient	Absolute t-value	Countries (N)
Political stability and absence of violence (1996–2009)	0.189***	3.360	34
Internal conflict risk (1990–2004)	0.346**	2.097	32
Weighted domestic conflict index (1981–2009)	−105.0*	1.611	30
Weighted domestic conflict index (1990–2009)	−119.7**	2.177	33
Deaths from domestic terrorism (1985–2010)	−2.357**	1.728	33

* Statistically significant at the 10 percent level (one-tailed test)
** Statistically significant at the 5 percent level (one-tailed test)
*** Statistically significant at the 1 percent level (one-tailed test)
Source: Based on data in Kaufmann, Kraay, and Mastruzzi 2010; PRS Group 2004; Banks, 2010; and GTD Team 2010

Performance variables	Estimated regression coefficient	Absolute t-value	Countries (N)
Political stability and absence of violence (1996–2009)	0.189***	3.360	34

What is a **regression coefficient?**

We assume a linear relationship between the **dependent variable** (outcome) and one or more **independent variables**, and use **ordinary least squares** regression (OLS) to find what the best-fitting **coefficients** (slopes) are.

Bivariate regression: week 6

$$\text{PolStab}_i = \beta_0 + \beta_1 \text{ExecParties}_i$$

$$\text{PolStab}_i = \beta_0 + \beta_1 \text{ExecParties}_i + \beta_2 \text{EconDev}_i + \beta_3 \text{LogPop}_i + \beta_4 \text{SocietalDiv}_i$$

Multivariate regression: week 7

Bivariate regression equation to be estimated:

$$\mathrm{PolStab}_i = \beta_0 + \beta_1 \mathrm{ExecParties}_i$$

Command in R:

```
lm(polstab ~ execparties, data = lijphart)
```
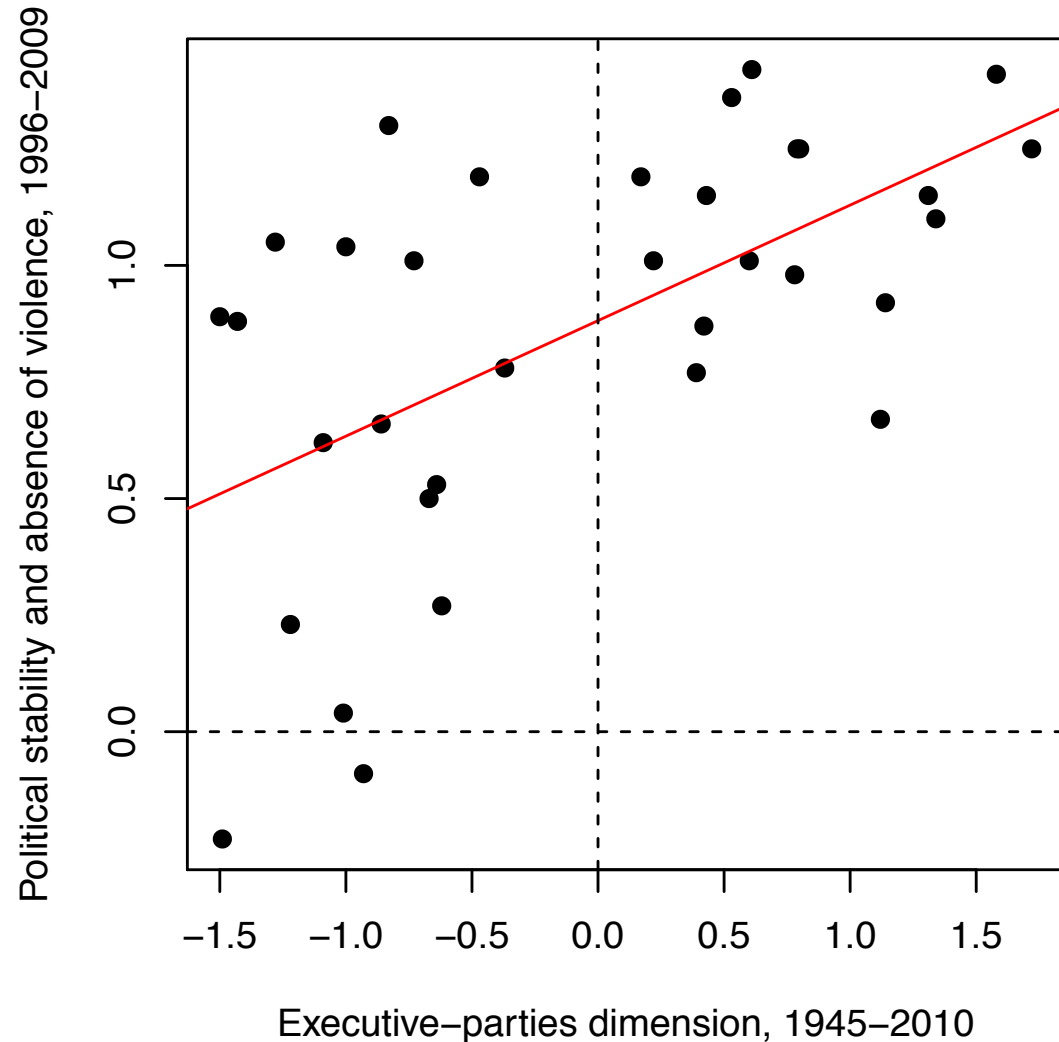
Resulting equation:

$$\mathrm{PolStab}_i = 0.88 + 0.25 \mathrm{ExecParties}_i$$

In figure at right:

- black dots are countries
- red line is the regression line

Which coefficient is the intercept, which is slope?

Political stability and absence of violence, 1996–2009 (y-axis)

Executive–parties dimension, 1945–2010 (x-axis)

Multivariate regression equation to be estimated:

$$\text{PolStab}_i = \beta_0 + \beta_1 \text{ExecParties}_i + \beta_2 \text{EconDev}_i + \beta_3 \text{LogPop}_i + \beta_4 \text{SocietalDiv}_i$$

Command in R:

```
lm(polstab ~ execparties + econdev + logpop + societaldiv, data = lijphart)
```

Resulting equation:

$$\text{PolStab}_i = -0.28 + 0.20 \text{ExecParties}_i + 2.80 \text{EconDev}_i - 0.13 \text{LogPop}_i - 0.08 \text{SocietalDiv}_i$$

These are "partial regression coefficients". They indicate how much the predicted outcome changes for a one-unit change in the independent variable, holding everything else constant.
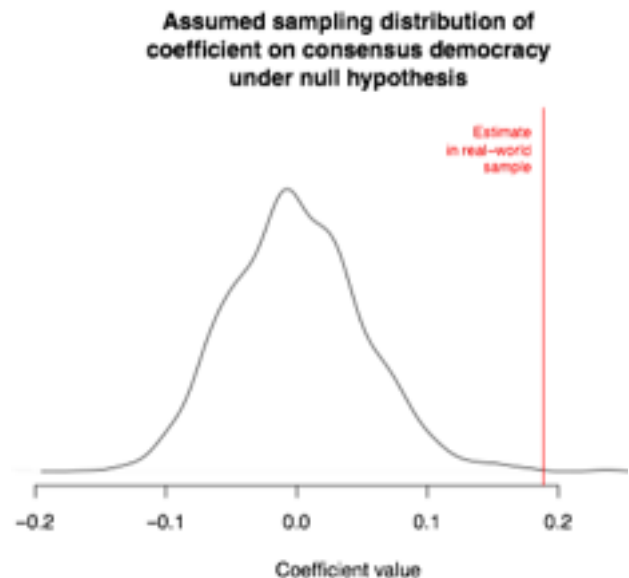
The **regression coefficient** tells us our best estimate of the linear relationship between two variables (possibly controlling for other variables).

The **p-value** tells us how likely it is that we would get a coefficient this large if the true coefficient was 0 (no relationship).

In regression tables, stars mean low p-values, which mean *statistically* strong relationships.

From week 8 of last year's lectures

Performance variables	Estimated regression coefficient	Absolute t-value	Countries (N)
Political stability and absence of violence (1996–2009)	0.189***	3.360	34

Assumed sampling distribution of coefficient on consensus democracy under null hypothesis

Estimate in real–world sample

-0.2   -0.1   0.0   0.1   0.2

Coefficient value
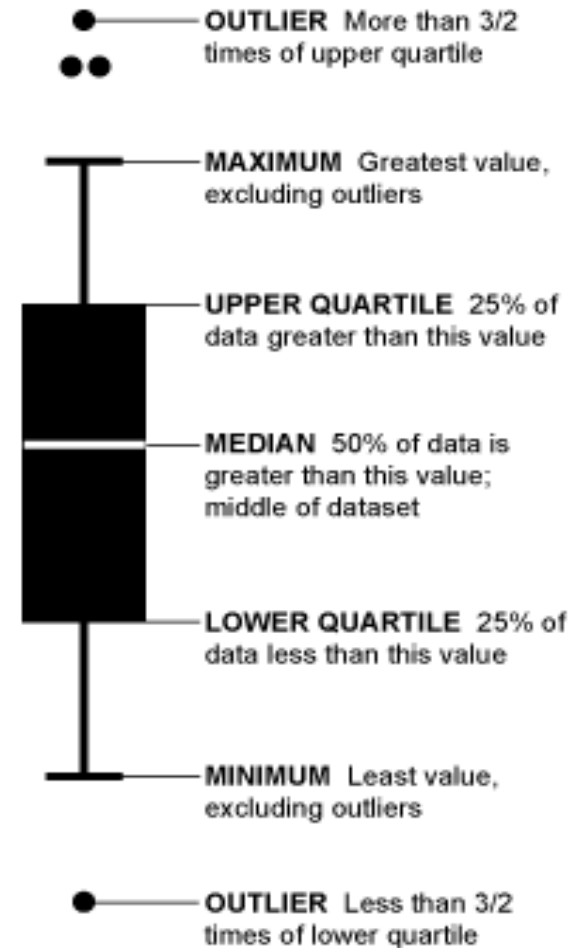
Review: summarizing information

Regressions, correlations, etc, focus on relationships among variables.

Important to also understand different types of variables:

- **interval variables** (aka continuous): values are placed on real numerical scale. *Examples:* GDP, Polity score, number of casualties

- **ordinal variables** (aka categorical): values can be ordered but are not put on a scale. *Example:* social class as salariat, routine non-manual, manual, working class

- **nominal variables:** values are not ordered. *Possible example:* degree of support for party (strongly support, weakly support, do not support, do not know)

And ways we summarize information about them:

Frequency distributions, histograms, measures of central tendency (mean, median, mode), percentiles, variance/standard deviation, box plot (see figure at right)

OUTLIER More than 3/2 times of upper quartile

MAXIMUM Greatest value, excluding outliers

UPPER QUARTILE 25% of data greater than this value

MEDIAN 50% of data is greater than this value; middle of dataset

LOWER QUARTILE 25% of data less than this value

MINIMUM Least value, excluding outliers

OUTLIER Less than 3/2 times of lower quartile

9

Before we analyze data, we must collect it.

Before we collect data, we must
- define a concept, i.e. the thing we want to measure (**conceptualization**)
- determine a procedure for measuring the concept (**operationalization**)

Behind every measure is
- lots of hard work, difficult choices
- (usually) a theory or research question.

How do we measure democracy? Bureaucracy?

Two kinds of critiques of measures:
- A measure is not **valid** if it is fundamentally aiming at the wrong thing: as designed, it does not correspond to the concept.
- A measure is not **reliable** if it is error-prone: it may be right for some units, but not for others.

What is a critique of Lijphart's measure of consensus democracy from the standpoint of
- validity?
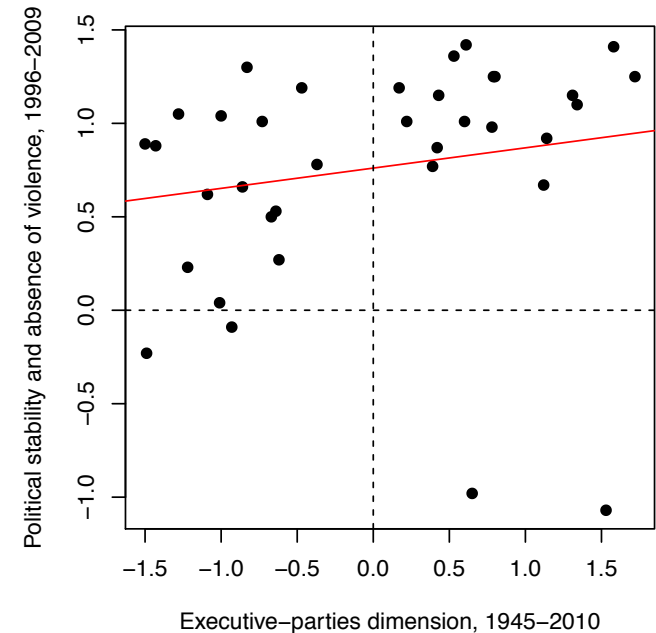- reliability?

All empirical claims

- have a **scope** of application: set of units (e.g. countries) within which claim is supposed to hold *(population)*
- are based on a **sample**: set of units (e.g. countries) used to establish/test the claim.

For Lijphart, the scope is *advanced democracies*; the sample is "36 countries continuously democratic since 1989 or earlier" (excluding outliers in particular cases).

Useful critiques of case selection:

- Sample is chosen based on the outcome *(selecting on the dependent variable)*
- Sample is chosen to support a point of view, not test a conjecture *(cherry picking)*
- Claim may hold true in the sample examined, but not more broadly *(over-generalization)*

All variations on one point: "The sample analyzed is not representative of the population of interest."

<figure>
Scatter plot with x-axis "Executive–parties dimension, 1945–2010" ranging from −1.5 to 1.5, and y-axis "Political stability and absence of violence, 1996–2009" ranging from −1.0 to 1.5, with a red regression line.
</figure>

Same relationship as above, with Israel and India included
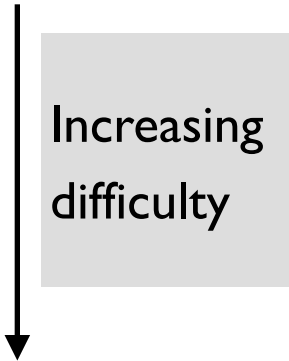
Some research is purely *descriptive*: it asks **"what is going on?"**

Some is *treatment-based*: it asks **"what is the effect of X?"**

Most is *explanatory* or *puzzle-based*: it asks **"why Y?"**

Often a combination.

Increasing difficulty

In your reading (and writing), take note of:
- the research question
- the *type* of question
- the *research design,* i.e. the comparisons and/or measurements being made to answer the question. For example,
    - cross-country regression
    - case study
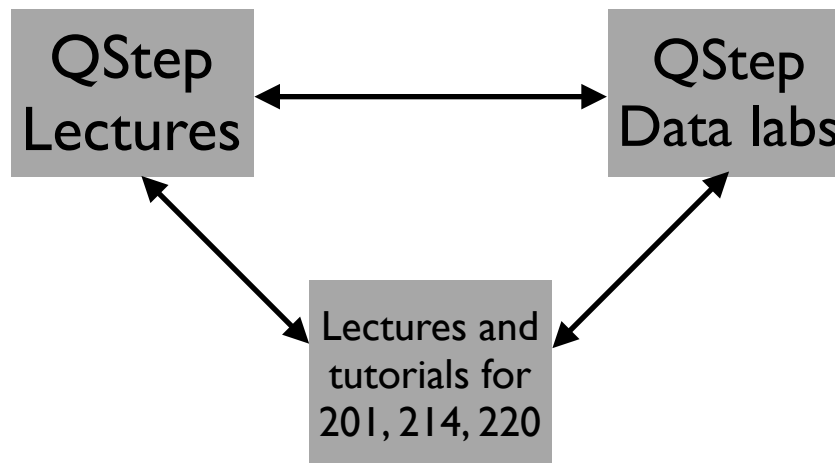   - survey experiment

Some key distinctions among designs:
- Is it an *experiment* (i.e. researcher manipulates something) or an *observational study*?
- What are the units? (Countries, individuals, cities, country-years)

Preview: Political Analysis 2

Themes we will emphasize:

- How to interpret statistical results beyond a simple multivariate regression
  - Interactions
  - Non-linearities
  - Logistic regression
- How to assess what control variables should be included
- Identifying and comparing research designs
- "Internal and external validity" as strategies of assessing evidence

Three key components:

```
 ┌──────────┐                          ┌──────────┐
 │  QStep   │  ←──────────────────→    │  QStep   │
 │ Lectures │                          │ Data labs│
 └──────────┘                          └──────────┘
         ↘                            ↗
          ┌────────────────┐
          │  Lectures and  │
          │  tutorials for │
          │ 201, 214, 220  │
          └────────────────┘
```

Political Analysis 2: roadmap

Lectures by week:

1. Review and preview (AE)
2. Presenting results and interactions (AR)
3. Modeling and interpreting non-linearity (AR)
4. What should we control for? (AR)
5. Randomized experiments (DK)
6. Alternatives to randomized experiments (DK)
7. Threats to internal and external validity (DK)
8. Evaluating research and evidence (DK)

Data labs by week:

3. Comparative Government (201) Replication 1
Political Sociology (220) Replication 1

4. International Relations (214) Replication 1

6. Comparative Government (201) Replication 2
Political Sociology (220) Replication 2

7. International Relations (214) Replication 2

For the time & location of lab sessions, see email from PPE office.

Lecturers:
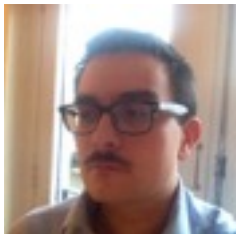
Andrea Ruggeri

Dave Kirk

Assessment: 2000-word essay replicating and extending analysis in a reading from your core paper (201, 214, 220) to be submitted by noon on Friday 29 January 2015

Review/preview: data analysis in RStudio

Each lab session is specific to a core option paper, e.g. IR.

In each lab we will replicate and extend a reading from the relevant core paper.

Comp Gov	**IR**	**Political Sociology**
Fish (2002), "Islam and authoritarianism" Brambor et al (2007), "Are African party systems different?"	Oneal & Russett (2005) "Rule of three, let it be? When more really is better." Gartzke (2007), "The capitalist peace"	Campbell et al (2010) "Do women need women representatives?" Best (2011), "The declining relevance of traditional cleavage groups"

Before first lab:

- Refresh your memory about using R. We will post resources on Weblearn.
- Read the article (specific to your option paper) and do the pre-lab assignment.

See Weblearn for announcements about drop-in sessions if you want help.