Basic Vocabulary and Descriptive Inference

Justin Murphy
jmrphy.net
@jmrphy

Inference?

- The challenge of **inference** is to use available information to make the best possible conclusions about what we don't know but would like to know.
  - **Descriptive inference** seeks to describe the existence of something.
    - Example: The number of people who participate in a riot.
  - **Causal inference** seeks to understand the effect of some variable(s) on some other variable(s)
    - Example: The *causal effect* of unemployment on the probability a riot will occur.
    - Example: The *causal effect* of a riot on next year's government spending.

Some Key Terms

- A **unit of analysis** is simply the object of study.
  - E.g., the individual human being, the constituency, the country, etc.
- A **variable** is a measurement of some concept that varies across a set of **units.**
  - E.g. unemployment rate across EU countries.
  - An **observation** is one realisation of a variable for one unit.
  - E.g., UK unemployment is equal to 6.0% in 2014.
  - Moving from micro-level to macro-level is aggregation (reverse is disaggregation).
  - E.g., from individual-level ("are you unemployed?)" to country-level (unemployment rate)
- Our **sample** is the set of observations we gather to make inferences about the world *outside the sample.*
  - I.e. a quantitative dataset or the cases you select to investigate.
- The **population** is what we call the world outside the sample we want to make generalisations about.

Some Key Terms (2)

- **Cross-sectional data** capture observations at one point in time, across a set of units
  - E.g., one survey of individuals.
- **Time-series data** capture observations over time for one unit.
  - E.g. unemployment rate every month for the UK.
  - **Panel data** capture observations for multiple units over time.
  - E.g., UK unemployment is equal to 6.0% in 2014.
  - Data can be **quantitative** (numerical).
  - Data can be **qualitative** (non-numerical).

TABLE 4.1 Different Types of Data

Type of data	Examples
Unit of analysis	Individuals, households, countries, U.S. states, U.S. or foreign cities, roll call votes of legislators, laws, newspaper articles, etc. . . .
Level of analysis	Micro-, meta-, macro-
Temporality	Cross-sectional, longitudinal, panel
Coverage	Representative sample, non-representative sample, population
Category	Quantitative, qualitative

BSS, pp.89

Descriptive Inference

Let's say we want to know how much of the British population supports the current government.

1. Take a *random, representative sample* of, say, 5,000 Brits.

2. Ask them if they support the government.

3. The sample mean can be used to *infer* the population mean.

4. Statistical theory provides rigorous rules for this inference, accounting for sample size, variance, and random error.

Potential sources of error

in estimating a population distribution using a sample

Sampling error

Because the sample is not the whole population

Non-sampling error

- Poor sampling method
- Questionnaire or measurement error
- Behavioural effects

17 The standard error for interval-level data (such as income) equals the standard deviation (s) divided by the square root of the sample size.

$$std.error = \frac{s}{\sqrt{n}}$$

BSS, pp. 99

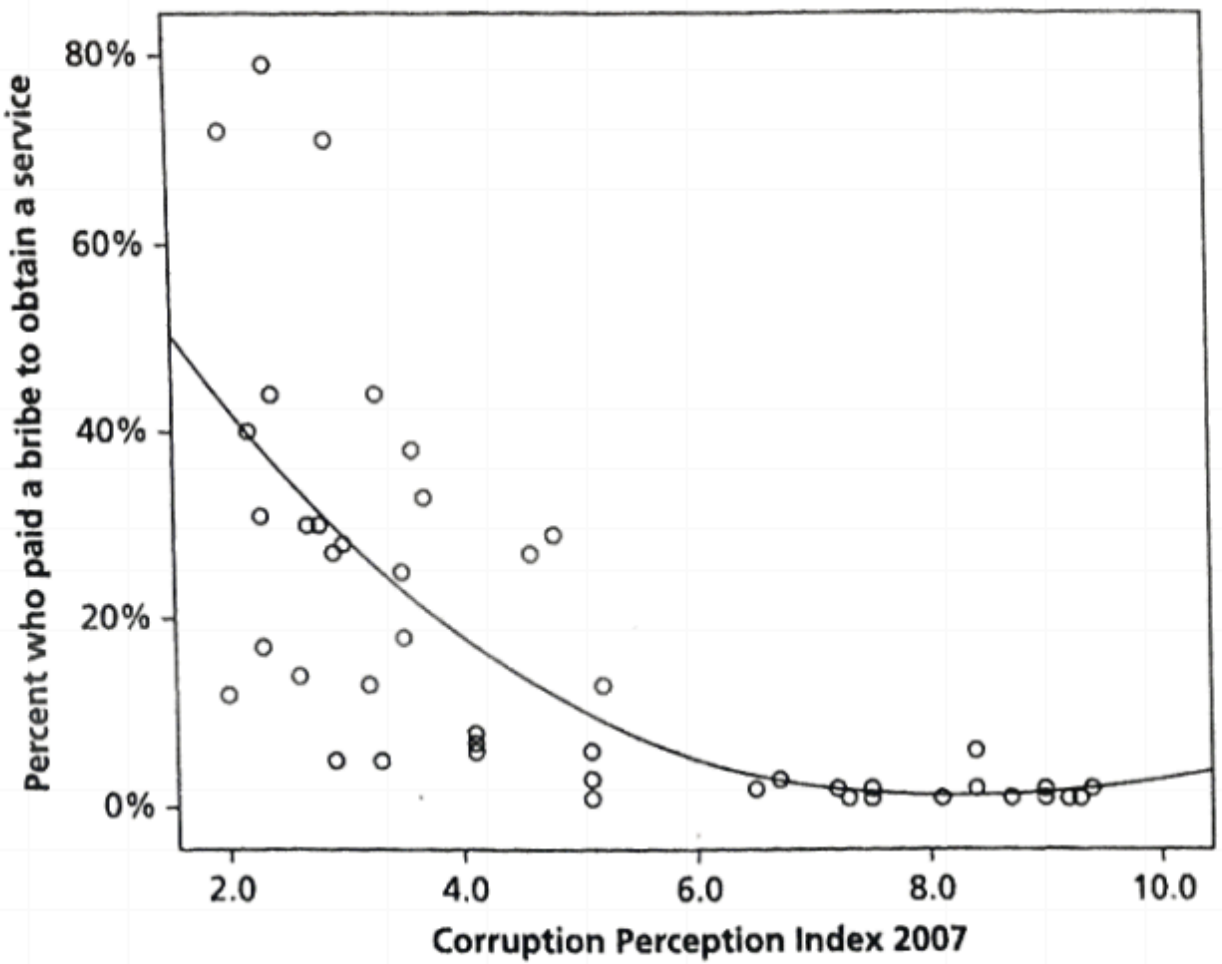

FIGURE 4.1 Scatterplot of the Relationship between Transparency International's CPI and Its Global Corruption Barometer

Source: Author's calculations based on data provided by Transparency International for 2007.

Note: $n = 45$ as the Global Corruption Barometer is only conducted in some countries.

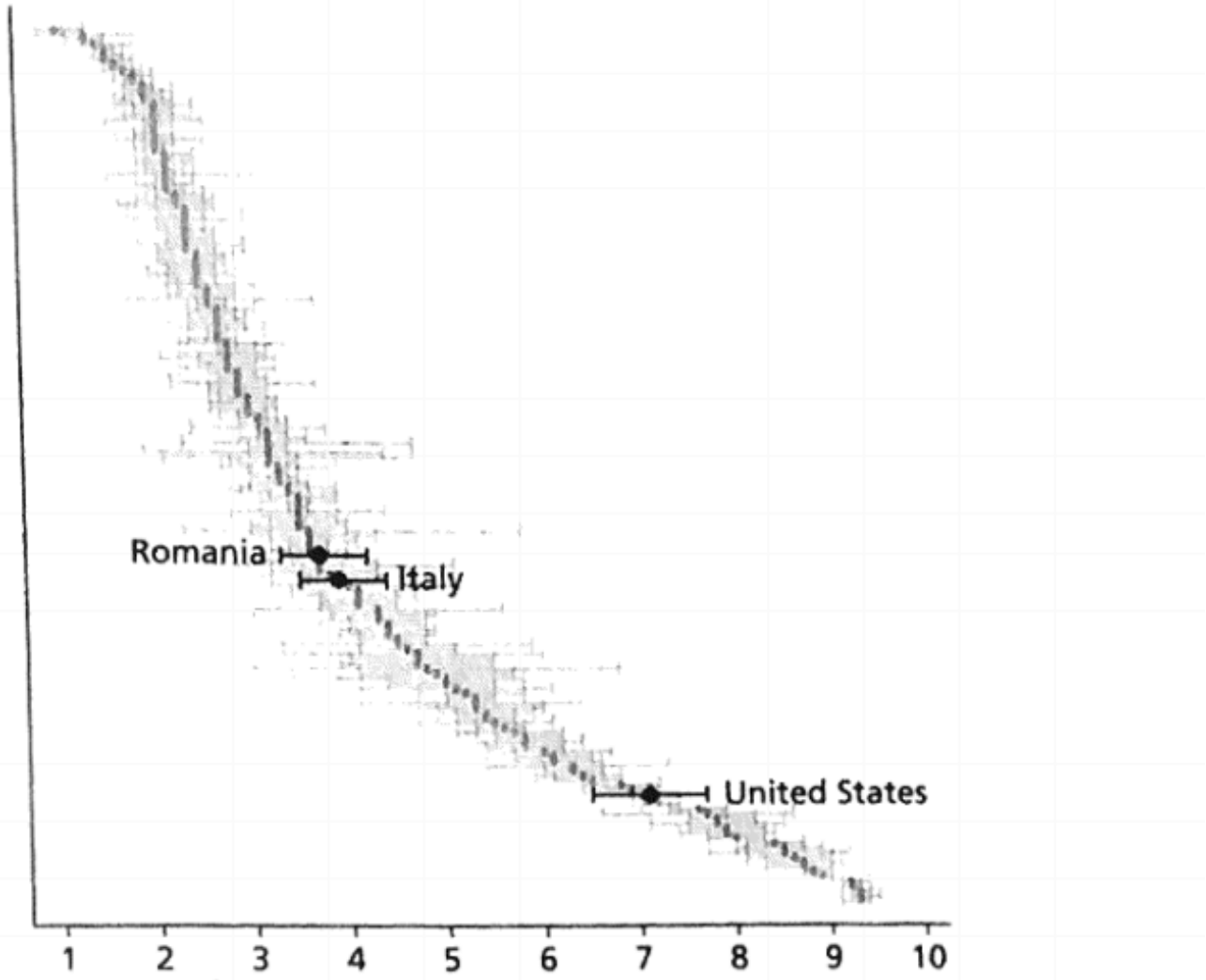

FIGURE 4.3 Plotting Estimates and Uncertainty for the Perception of Corruption Index in 178 Countries

Source: Created by the authors using data from Transparency International.

Note: Bars represent 90 percent confidence intervals.

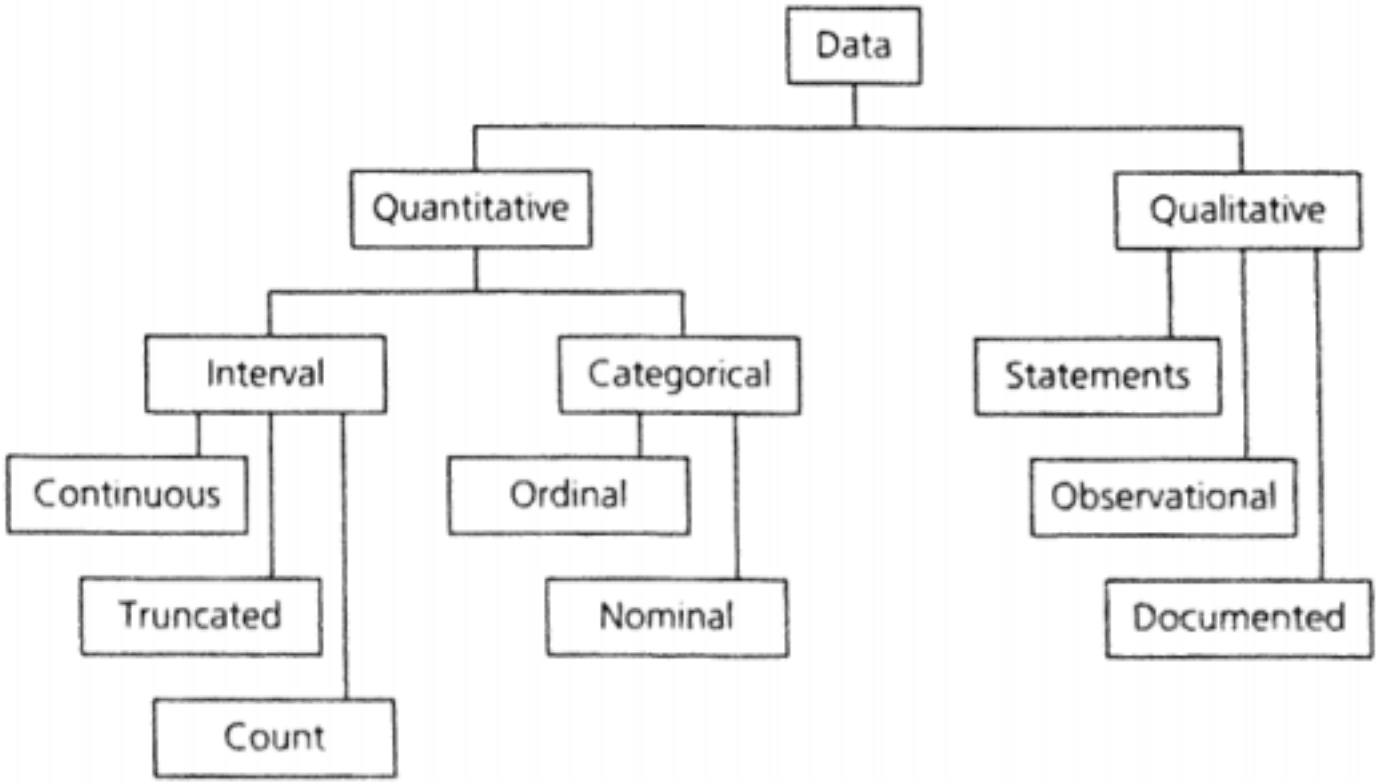

FIGURE 4.4 Different Types of Data

Nominal-Level Data

TABLE 4.2 Regime Classification

Regime type	Frequency	Percentage	Valid percentage
Parliamentary democracy	56	29.3%	29.6%
Mixed democracy	21	11.0%	11.1%
Presidential democracy	37	19.4%	19.6%
Civilian dictatorship	38	19.9%	20.1%
Military dictatorship	24	12.6%	12.7%
Monarchic dictatorship	13	6.8%	6.9%
Missing data	2	1.0%	
Total	191	191 (100%)	189 (100%)

Source: Democracy Cross-National Data.[35]

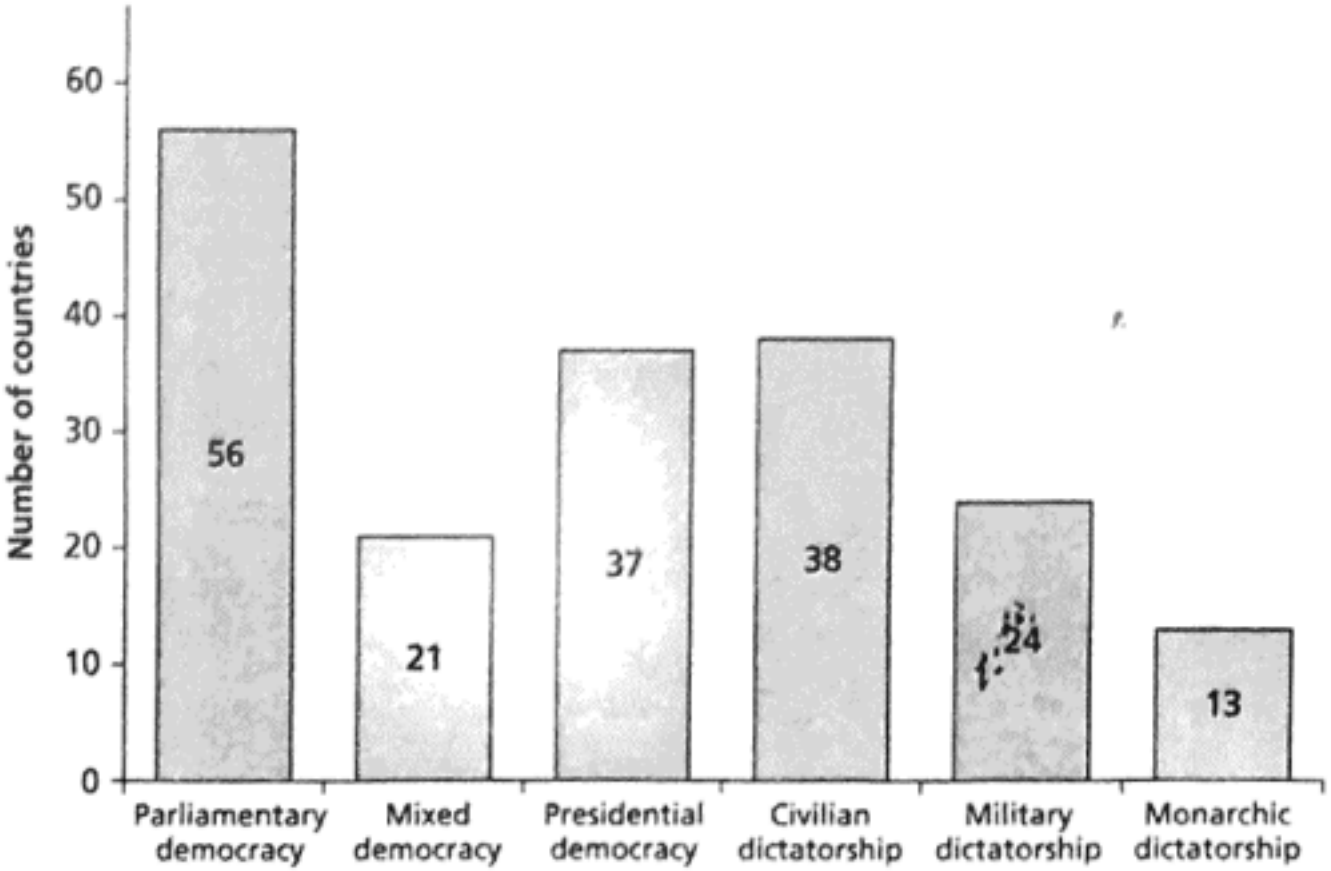

FIGURE 4.5 Nominal-Level Data: Regime Classification across Countries

Source: Democracy Cross-National Data.[27]

Note: $n = 189$ countries.