Quantitative Methods

Weeks 7-8: Statistics

AY 2023-24

Department of Political Economy

Instructor: Daniele Girardi

Our journey

- **W 1-2**: Describing & summarizing data
- **W 3-5**: Probability theory
- **W 7-8** Statistics
- **W 9-11**: Econometrics

Descriptive statistics

Statistical Inference

Application to social science

Weeks 7-8 – Statistics

1. Estimating a population's mean & variance
   - *Ross Ch 8.2 and 8.4; or S&W Ch 3.1*
2. Hypothesis tests
   - *Ross Ch 9; or S&W Ch 3.2*
3. Confidence intervals
   - *Ross Ch 8.5 and 8.6; or S&W Ch 3.3*
4. Testing differences in means
   - *Ross Ch 10; or S&W Ch 3.4*

Write down three things you learned from the reading

(Parts of Ch 8-9 in Ross; or 3.1 to 3.3 in S&W)

If you couldn't do the reading this week:
Write three things that come to your mind when you think about "statistics".

CHAPTER 3

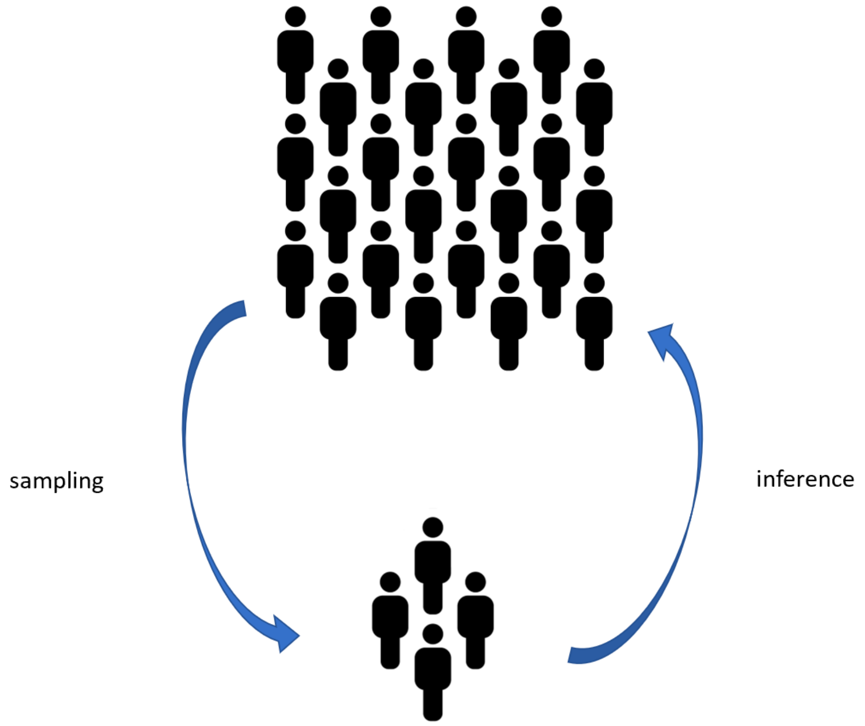
Review of Statistics

Statistics is the science of using data to learn about the world around us. Statistical tools help us answer questions about unknown characteristics of distributions in populations of interest. For example, what is the mean of the distribution of earnings of recent college graduates? Do mean earnings differ for men and women and, if so, by how much?

These questions relate to the distribution of earnings in the population of workers. One way to answer these questions would be to perform an exhaustive survey of the population of workers, measuring the earnings of each worker and thus finding the population distribution of earnings. In practice, however, such a comprehensive survey would be extremely expensive. Comprehensive surveys that do exist, also known as censuses, are often undertaken periodically (for example, every ten years in India, the United States of America and the United Kingdom). This is because the process of conducting a census is an extraordinary commitment, consisting of designing census forms, managing and conducting surveys, and compiling and analyzing data. Censuses across the world have a long history, with accounts of censuses recorded by Babylonians in 4000 BC. According to historians, censuses have been conducted as far back as Ancient Rome; the Romans would track the population by making people return to their birthplace every year in order to be counted.[1] In England and other parts of Wales, a notable census was the Domesday Book, which was compiled in 1086 by William the Conqueror. The U.K. census in its current form dates back to 1801 after essays by economist Thomas Malthus (1798) inspired parliament to want to accurately know the size of the population. Over time the census has evolved from amounting to a mere headcount to the much more ambitious survey of the 2011 U.K. census costing an estimated £482 million. In India, there are accounts of censuses recorded around 300 BC, but the census in its current form has been undertaken since 1872 and every ten years since 1881. In comparison to the U.K. census of 2011, the most recent census of India, also conducted in 2011, approximately cost a mere ₹2200 crore (US$320 million)! Despite the considerable efforts made to ensure that the census records all individuals, many people slip through the cracks and are not surveyed. Thus a different, more practical approach is needed.

The key insight of statistics is that one can learn about a population distribution by selecting a random sample from that population. Rather than survey the entire popu-

What Does Statistics Do?



Learn about a population by analyzing a random sample.

1. Estimation
2. Hypothesis Testing
3. Confidence Intervals

1. Estimating a population mean & variance

Estimators

- **Estimator:** a statistic computed from the sample and used to predict a population parameter.

What makes an estimator "good"?

- **Unbiasedness**: $E(\hat{\mu}_Y) = \mu_Y$
- **Consistency**: $\hat{\mu}_Y \xrightarrow{\text{p}} \mu_Y$
- **Efficiency**: $var(\hat{\mu}_Y)$ smaller rather than larger.

$\bar{Y}$ as an Estimator of μ_Y

- Sample average: $\bar{Y} = \frac{1}{n}(Y_1 + Y_2 + \cdots + Y_n) = \frac{1}{n}\sum_{i=1}^{n} Y_i$

1. $E(\bar{Y}) = \mu_Y$
   - because $Y_1, Y_{...}, Y_N$ are i.i.d.

2. $\bar{Y} \xrightarrow{p} \mu_Y$
   - because Law of Large Numbers

3. $var(\bar{Y}) < var(\hat{\mu}_Y)$
   - where $\hat{\mu}_Y$=every other unbiased estimator of μ_Y
   - because it can be proved (although we won't here)

$\bar{\boldsymbol{Y}}$ is BLUE

Importance of random sampling

- We are assuming $Y_1, \dots, Y_n$ are i.i.d., as in random sampling.
- If sampling is not random, $\bar{Y}$ might be a *biased* estimator of μ_Y!
    - $E(\bar{Y}) \neq \mu_Y$
- Is this why pollsters were wrong about Brexit and Trump in 2016?

The standard deviation of $\bar{Y}$

- What is the spread (variance & SD) of the sampling distribution of $\bar{Y}$?
- We know from Probability Section that if the sample is random (i.i.d.)

$$SD(\bar{Y}) = \frac{\sigma_Y}{\sqrt{n}}$$

  - (σ_Y = standard error of Y in the population)
  - (n = sample size)

Estimating the population variance

What sample statistics would you use as estimator of σ_Y^2?

The variance of Y in the sample!

Using s^2 to estimate σ_Y^2

- If you knew μ_Y, the correct estimator would be

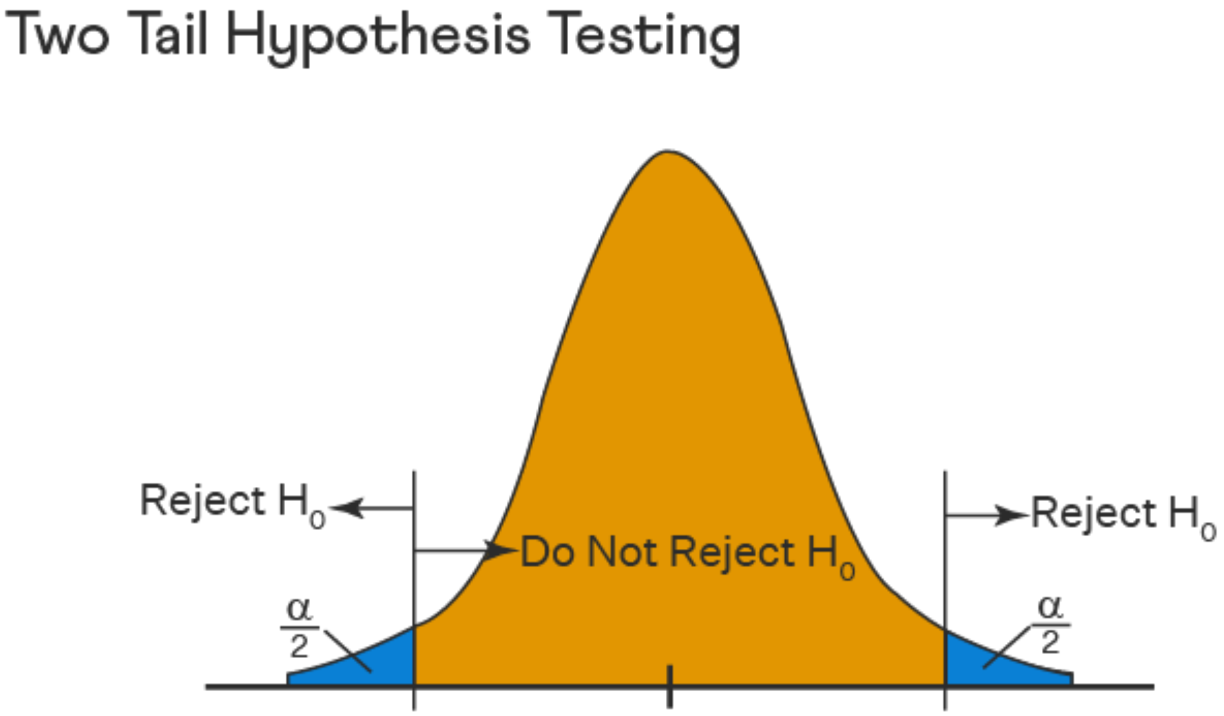
$$\frac{\sum_{i=1}^{n}(Y_i - \mu_Y)^2}{n}$$

- But if you don't know μ_Y, then the estimator is

$$s^2 = \frac{\sum_{i=1}^{n}(Y_i - \bar{Y})^2}{n-1}$$

- *Unbiased*: $\mathrm{E}(s^2) = \sigma_Y^2$ and $E\left(\sqrt{s^2}\right) = \sigma_Y$

2. Hypothesis tests



Hypothesis

- A *statement about the nature of a population, often stated in terms of a population parameter*.
    - Do average earnings of recent graduates equal 20£/hour?
    - Does more than 50% of UK voters plan to vote Labour in 2024 GE?
    - Did the average hourly wage increase in the last year?

- Null hypothesis:

$$H_0: E(Y) = \mu_{Y,0}$$

- Alternative hypothesis:

$$H_1: E(Y) \neq \mu_{Y,0}$$

Thank you for your attention