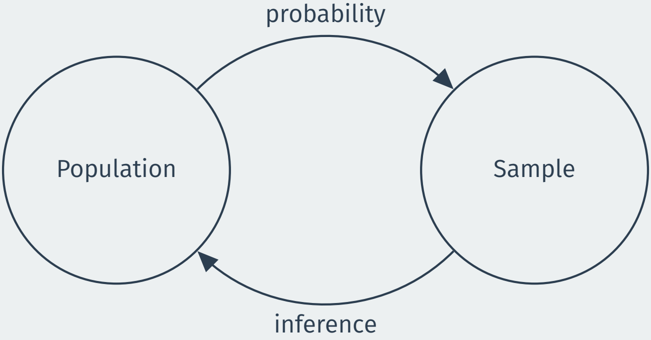
Gov 51: Estimators

Matthew Blackwell

Harvard University

Remember our goal



- We want to learn about the chance process that generated our data.
- Now we switch to **inference**.
    - What can I learn about the population distribution from my sample?

How popular is Donald Trump?

- What proportion of the public approves of Trump's job as president?
- Latest Gallup poll:
    - Oct. 29th–Nov. 4th
    - 1500 adult Americans
    - Telephone interviews
    - Approve (40%), Disapprove (54%)
- What can we learn about Trump approval in the population from this one sample?

Samples from the population

- Simple random sample of size n from some population $Y_1, \ldots, Y_n$
  - $\rightsquigarrow$ i.i.d. random variables
  - e.g.: $Y_i = 1$ if i approves of Trump, $Y_i = 0$ otherwise.
- **Statistical inference**: using data to guess something about the population distribution of Y_i.

Point estimation

- **Quantity of interest**: some feature of the population distribution.
    - Also called: parameters.
    - These are the things we want to learn about.
- **Point estimation**: providing a single "best guess" about this q.o.i.
- Examples of quantities of interest:
    - $\mu = \mathbb{E}[Y_i]$: the population mean (turnout rate in the population).
    - $\sigma^2 = \mathbb{V}[Y_i]$: the population variance.
    - $\mu_1 - \mu_0 = \mathbb{E}[Y(1)] - \mathbb{E}[Y(0)]$: the population ATE.

Estimators

- **Estimator**: function of the data that produces estimates of the q.o.i.
  - An **estimate** is one particular realization of the estimator
- Ideally we'd like to know the **estimation error**, estimator — truth
  - Problem: θ is unknown.
- Solution: figure out the properties of estimator using probability.
  - Estimator is a r.v. because it is a function of r.v.s. (the data)
  - $\rightsquigarrow$ estimator has a distribution has a distribution.

Estimating Trump's support

- Parameter p: **population proportion** of adults who support Trump
- There are many different possible estimators:
  - $\hat{p} = \overline{Y}_n$ the sample proportion of respondents who support Trump.
  - $\hat{p} = Y_1$ just use the first observation
  - $\hat{p} = \max(Y_1, \ldots, Y_n)$
  - $\hat{p} = 0.5$ always guess 50% support
- How good are these different estimators?

Survey

- Assume a simple random sample of n voters: $n = 1500$
- Define r.v. Y_i for Trump approval:
  - $Y_i = 1 \rightsquigarrow$ respondent i approves of Trump
  - $Y_i = 0 \rightsquigarrow$ respondent i disapproves of Trump
- Y_i is **Bernoulli** with probability of success p
  - "success" = "selecting a Trump approver"
  - $p = \mathbb{P}(Y_i = 1)$ the population proportion of Trump approvers.
- Sample proportion is the same as the sample mean:

$$\overline{Y} = \frac{1}{n}\sum_{i=1}^{n} Y_i = \frac{\text{number who support Trump}}{n}$$

Sample mean properties

$$\text{sample proportion} = \text{population proportion} + \text{chance error}$$
$$\overline{Y} = p + \text{chance error}$$

- Remember: the sample mean/proportion is a random variable.
    - Different samples give different sample means.
    - Chance error “bumps” sample mean away from population mean
- $\rightsquigarrow$ $\overline{Y}$ has a distribution across repeated samples.

Central tendency of the sample mean

- Expectation: average of the estimates across repeated samples.
  - From last week, $\mathbb{E}[\overline{Y}] = \mathbb{E}[Y_i] = p$
  - $\rightsquigarrow$ chance error is 0 on average:

$$\mathbb{E}[\overline{Y} - p] = \mathbb{E}[\overline{Y}] - p = 0$$

- **Unbiasedness**: Sample proportion is on average equal to the population proportion.

Spread of the sample mean

- **Standard error**: how big is the chance error on average?
    - This is the standard deviation of the estimator.
- Special rule for sample proportions:

$$\sqrt{\mathbb{V}(\overline{Y})} = \sqrt{\frac{p(1-p)}{n}}$$

- Problem: we don't know p!
- Solution: **estimate** the SE:

$$\sqrt{\widehat{\mathbb{V}}(\overline{Y})} = \sqrt{\frac{\overline{Y}(1-\overline{Y})}{n}} = \sqrt{\frac{0.37 \times (1-0.37)}{1500}} \approx 0.012$$