Gov 50: 24. More Inference with Mathematical Models

Matthew Blackwell

Harvard University

Roadmap

1. Confidence intervals for experiments
2. Hypothesis testing with the CLT
3. Two-sample tests

1/ Confidence intervals for experiments

Comparison between groups

- More interesting to compare across groups.
    - Differences in public opinion across groups
    - Difference between treatment and control groups.
- Bedrock of causal inference!

Social pressure experiment

- Back to the Social Pressure Mailer GOTV example.
    - Primary election in MI 2006
- Treatment group: postcards showing their own and their neighbors' voting records.
    - Sample size of treated group, $n_T = 360$ (artificially reducing sample size to highlight the math)
- Control group: received nothing.
    - Sample size of the control group, $n_C = 1890$

Outcomes

- Outcome: $Y_i = 1$ if i voted, 0 otherwise.
- Turnout rate (sample mean) in treated group, $\overline{Y}_T = 0.37$
- Turnout rate (sample mean) in control group, $\overline{Y}_C = 0.30$
- Estimated **average treatment effect**

$$\widehat{\text{ATE}} = \overline{Y}_T - \overline{Y}_C = 0.07$$

Inference for the difference

- Parameter: **population ATE** $\mu_T - \mu_C$
  - μ_T: Turnout rate in the population if everyone received treatment.
  - μ_C: Turnout rate in the population if everyone received control.
- Estimator: $\widehat{\text{ATE}} = \overline{Y}_T - \overline{Y}_C$

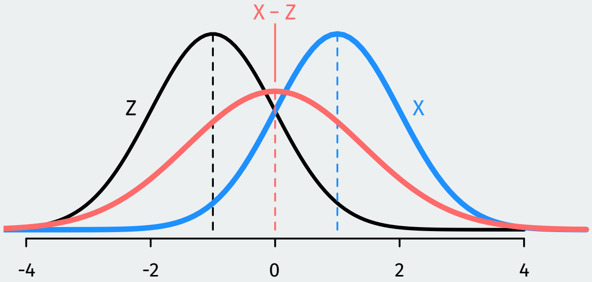
By the CLT in large samples, we know that:

- $\overline{Y}_T \approx N\left(\mu_T, \frac{\mu_T(1-\mu_T)}{n_C}\right)$
- $\overline{Y}_C \approx N\left(\mu_C, \frac{\mu_C(1-\mu_C)}{n_C}\right)$
- $\rightsquigarrow \overline{Y}_T - \overline{Y}_C \approx N(\mu_T - \mu_C, SE^2_{\text{diff}})$

But what is the SE_{diff} in this case?

Spread of a difference in normals

If we take the difference between two independent normal r.v.s, what happens to the spread?



The spread of the difference is **larger** than the spread of the two variables being differenced!

Standard error for the estimated ATE

- SE of a difference in means **adds** the SEs for each group

$$SE_{\text{diff}} = \sqrt{SE_T^2 + SE_C^2}$$

- Using what we know about SEs with binary outcomes:

$$SE_{\text{diff}} = \sqrt{\frac{\mu_T(1-\mu_T)}{n_t} + \frac{\mu_C(1-\mu_C)}{n_C}}$$

- Chance errors $\overline{Y}_T - \overline{Y}_C - (\mu_T - \mu_C) \approx N(0, SE_{\text{diff}})$
  - We can construct a 95% CI with $\widehat{\text{ATE}} \pm 1.96 \times SE_{\text{diff}}$

Confidence intervals

But we don't know μ_T or μ_C! Plug in our sample proportions to estimate the SE:

$$\widehat{\text{SE}}_{\text{diff}} = \sqrt{\frac{\overline{Y}_T(1-\overline{Y}_T)}{n_t} + \frac{\overline{Y}_C(1-\overline{Y}_C)}{n_C}}$$
$$= \sqrt{\frac{0.37 \times 0.63}{360} + \frac{0.3 \times 0.7}{1890}} = 0.028$$

Now we can construct confidence intervals based on the CLT like last time:

$$\begin{aligned} CI_{95} &= \widehat{\text{ATE}} \pm 1.96 \times \widehat{\text{SE}}_{\text{diff}} \\ &= 0.07 \pm 1.96 \times 0.028 \\ &= 0.07 \pm 0.054 \\ &= [0.016, 0.124] \end{aligned}$$

Range of possibilities taking into account plausible chance errors.

2/ Hypothesis testing with the CLT

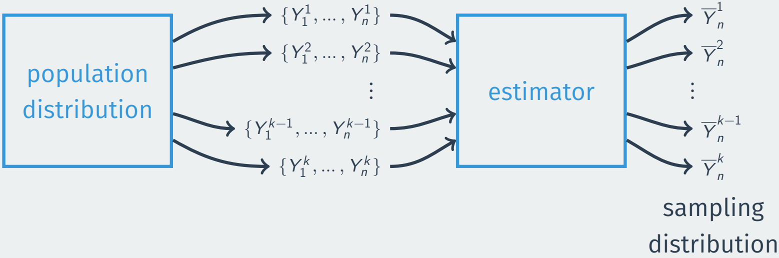
Statistical hypothesis testing

- Statistical hypothesis testing is a **thought experiment**.
- What would the world look like **if we knew the truth**?
- Conducted with several steps:
    1. Specify your **null** and **alternative hypotheses**
    2. Choose an appropriate **test statistic** and level of test α
    3. Derive the **reference distribution** of the test statistic under the null.
    4. Use this distribution to calculate the **p-value**.
    5. Use p-value to decide whether to reject the null hypothesis or not

How popular is Joe Biden?

- What proportion of the public approves of Biden's job as president?
- Latest Gallup poll: $\overline{Y} = 0.42$ with $n = 812$
- Could we reject the null that Biden's national support is 50%?
    - Null: $H_0 : p = 0.5$
    - Alternative: $H_1 : p \neq 0.5$

Sampling distribution, in pictures



CLT for hypothesis testing

Under the null, we know the distribution of $\overline{Y}$:

$$\overline{Y} \approx N\left(p, \frac{p(1-p)}{n}\right) = N\left(0.5, \frac{0.5 \times 0.5}{812}\right)$$

Using the rules of normal transformations if $X \sim N(\mu, \sigma^2)$:

$$\frac{X - \mu}{\sigma} \sim N(0, 1)$$

Then under the null, know the distribution of the following test statistic:

$$Z = \frac{\overline{Y} - 0.5}{0.5/\sqrt{812}} \approx N(0, 1)$$

p-values

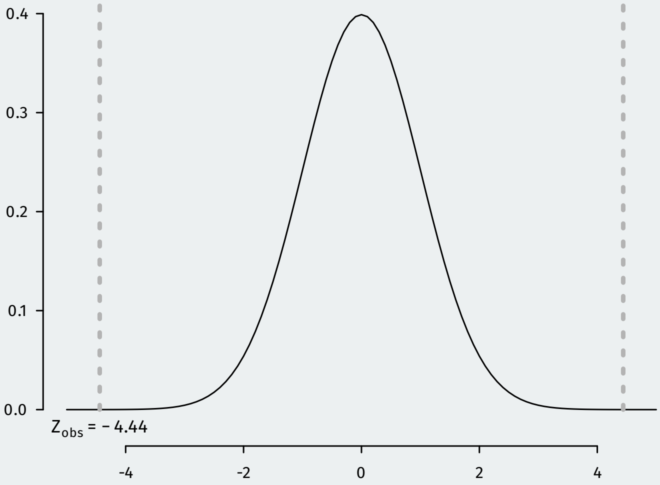
What we observe:

$$
\begin{aligned}
Z_{\text{obs}} = \frac{\overline{Y} - 0.5}{0.5/\sqrt{812}} &= \frac{0.42 - 0.5}{0.5/\sqrt{812}} \\
&= -\frac{0.08}{0.018} = -4.44
\end{aligned}
$$

Our observed sample proportion is 4.44 SEs away from 0.5 under the null. What's the probability of being that far away? (**p-value**)

```r
pnorm(-4.44, mean = 0, sd = 1) + ## prob being below -4.44
  (1 - pnorm(4.44, mean = 0, sd = 1)) ## prob being above 4.44
```

```
## [1] 0.000009
```


0.4
0.3
0.2
0.1
0.0
$Z_{obs} = -4.44$
-4
-2
0
2
4

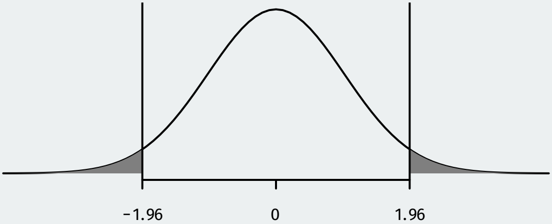
Generalizing hypothesis tests

- Hypothesis testing using the CLT pretty much takes this general form no matter what the estimator of interest is.
- Hypotheses: $H_0 : \mu = \mu_0$ (null guess), $H_1 : \mu \neq \mu_0$
- Test statistic:

$$Z = \frac{\text{observed value} - \text{null guess}}{\widehat{SE}} = \frac{\overline{Y} - \mu_0}{\widehat{SE}}$$

  - The exact estimator for the standard error $\widehat{SE}$ will depend on the estimator of interest.
- Null distribution: $Z \approx N(0, 1)$ by the CLT
- p-value: probability of a standard normal being bigger than $|Z_{\text{obs}}|$

Rejecting regions



- Reject if p-value is below α (usually 0.05).
    - We know 5% of the time Z will be bigger than 1.96.
    - If $Z_{\text{obs}} > 1.96$ or $Z_{\text{obs}} < -1.96$, then the p-value must be below 0.05
    - We can reject if $|Z_{\text{obs}}| > 1.96$

3/ Two-sample tests

Two-sample hypotheses

- Parameter: **population ATE** $\mu_T - \mu_C$
- Goal: learn about the population difference in means
- Usual null hypothesis: no difference in population means (ATE = 0)
  - Null: $H_0 : \mu_T - \mu_C = 0$
  - Two-sided alternative: $H_1 : \mu_T - \mu_C \neq 0$
- In words: are the differences in sample means just due to chance?

Difference-in-means review

- Sample turnout rates: $\overline{Y}_T = 0.37$, $\overline{Y}_C = 0.30$
- Sample sizes: $n_T = 360$, $n_C = 1890$
- Estimator is the **sample difference-in-means**:

$$\widehat{\text{ATE}} = \overline{Y}_T - \overline{Y}_C = 0.07$$

- Estimated SE for the difference in means:

$$\widehat{\text{SE}}_{\text{diff}} = \sqrt{\frac{\overline{Y}_T(1 - \overline{Y}_T)}{n_T} + \frac{\overline{Y}_C(1 - \overline{Y}_C)}{n_C}} = 0.028$$

CLT again and again

Earlier we saw that by the CLT we have:

$$\overline{Y}_T - \overline{Y}_C \approx N(\mu_T - \mu_C, \text{SE}^2_{\text{diff}})$$

We can use Z-scores to get a test statistic:

$$Z = \frac{(\overline{Y}_T - \overline{Y}_C) - (\mu_T - \mu_C)}{\text{SE}_{\text{diff}}} \sim N(0, 1)$$

Same general form of the test statistic as with one sample mean/proportion:

$$\frac{\text{observed - null guess}}{\text{SE}}$$

The usual null of no difference

- Null hypothesis: $H_0 : \mu_T - \mu_C = 0$
- Test statistic:

$$Z = \frac{(\overline{Y}_T - \overline{Y}_C) - (\mu_T - \mu_C)}{\text{SE}_{\text{diff}}} = \frac{(\overline{Y}_T - \overline{Y}_C) - 0}{\text{SE}_{\text{diff}}}$$

- In large samples, we can replace true SE with an estimate:

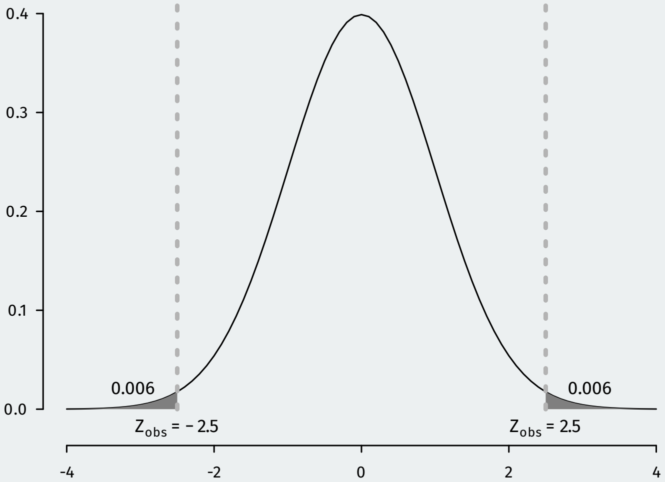
$$\widehat{\text{SE}}_{\text{diff}} = \sqrt{\widehat{\text{SE}}_T^2 + \widehat{\text{SE}}_C^2}$$

Calculating p-values

- Finally! Our test statistic in this sample:

$$Z = \frac{\overline{Y}_T - \overline{Y}_C}{\widehat{\mathsf{SE}}_{\mathsf{diff}}} = \frac{0.07}{0.028} = 2.5$$

- p-value based on a two-sided test: probability of getting a difference in means this big (or bigger) if the null hypothesis were true
    - Lower p-values $\rightsquigarrow$ stronger evidence against the null.




```r
2 * pnorm(2.5, lower.tail = FALSE)
```

```
## [1] 0.0124
```