Gov 51: Two-sample Tests

Matthew Blackwell

Harvard University

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

Last time

- We looked at hypothesis tests for means.
    - Tested null that true population mean was some value: $H_0 : \mu = \mu_0$
- Under the null hypothesis, we can determine the (approximate) distribution of the test statistic:

$$Z = \frac{\overline{X} - \mu_0}{\sigma / \sqrt{n}}$$

- Calculated p-values of this test statistic
- Today: generalizing to differences in means.

Social pressure example

- Back to the Social Pressure Mailer GOTV example.
  - Treatment group: mailers showing voting history of them and neighbors.
  - Control group: received nothing.
- Samples are **independent**
  - Example of dependent comparisons: **paired comparisons**

Two-sample hypotheses

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

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

- Standard error of difference in means of independent samples:

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

- Since turnout is binary, we can use the special proportions rule for the SEs:

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

CLT again and again

- $\overline{X}_T$ is a sample mean and so tends toward normal as $n_T \to \infty$
- $\overline{X}_C$ is a sample mean and so tends toward normal as $n_C \to \infty$
- $\rightsquigarrow \overline{X}_T - \overline{X}_C$ will tend toward normal as sample sizes get big.
- Using the z-transformation/standardization:

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

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

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

The usual null of no difference

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

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

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

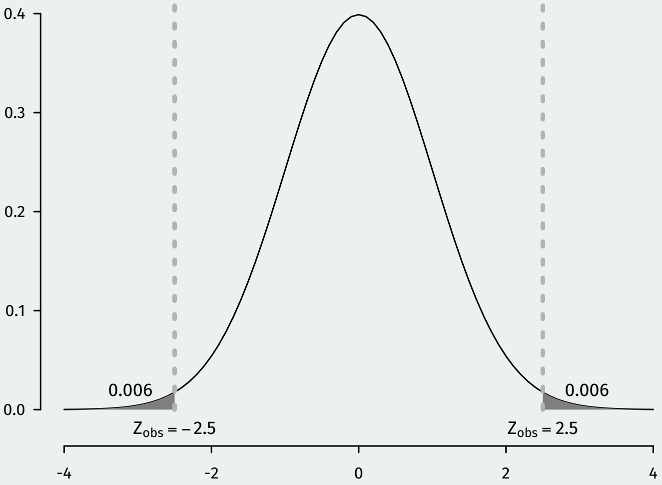
$$\widehat{\text{SE}}_{\text{diff}} = \sqrt{\widehat{\text{SE}}_T^2 + \widehat{\text{SE}}_C^2}$$

Calculating p-values

- Finally! Our test statistic in this sample:

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

- 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.




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

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

Tests and confidence intervals

- Deep connection between confidence intervals and tests.
- A 95% CI contains all null hypotheses with p-values greater than 0.05.
    - All the nulls we couldn't reject if $\alpha = 0.05$
    - 95% CI for social pressure experiment: $[0.016, 0.124]$
    - $\rightsquigarrow$ p-value for $H_0 : \mu_T - \mu_C = 0$ less than 0.05.
- More generally: Any value outside of a $100 \times (1 - \alpha)$% confidence interval would have a p-value less than α if we tested it as the null hypothesis.