Gov 51: Hypothesis Tests for Sample Means

Matthew Blackwell

Harvard University

General sample means

- Last time: hypothesis testing for a sample proportion.
    - Binary data $\rightsquigarrow$ easy setting.
    - Distribution of samples just depends on population proportion.
- This time: hypothesis testing for means of any variable.

Hypothesis testing procedure

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

- This procedure is general, but we'll focus on tests of a single population mean today.

Test statistic

Test statistic

A **test statistic** is a function of data and possibly the null hypothesis used to adjudicate between the null and alternative hypotheses.

- Most common form for sample means is the z-statistic:

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

- Put differently:

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

    - How many SEs away from the null guess is the sample mean?

- Usually replace population SD σ with sample SD $\hat{\sigma}$

Example: thermometer scores

- Social scientists often use **thermometer scores** to assess views toward groups.
    - 0-100 scale, where higher is "warmer" feeling toward group.
- You work at advocacy group who got a survey with FT scores for transgender people.
    - $\overline{X} = 52.3$ and SD $\hat{\sigma} = 29.3$
    - Sample size, $n = 912$
- Co-worker Nully is weirdly insistent that these results are consistent with a population mean FT score of 50.
- Hypothesis tests to the rescue!

Calculating the test statistic

- Hypotheses:
  - $H_0 : \mu = 50$, population average is 50.
  - $H_1 : \mu \neq 50$
- Test statistic:

$$Z_{\text{obs}} = \frac{\overline{X} - \mu}{\hat{\sigma}/\sqrt{n}} = \frac{52.3 - 50}{29.3/\sqrt{900}}$$
$$\approx -2.35$$

- Observed average is 2.35 SEs away from the null!
  - Exactly how unlikely is this?

Determining the reference distribution

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

- What is the distribution of Z?
- With sample proportions, we relied on the binomial distribution.
    - Doesn't work if variable is non-binary (age, income, etc)
- Central limit theorem to the rescue! In large samples and under the null:
    - $\overline{X}$ is normal with mean μ and standard deviation $\sigma / \sqrt{n}$.
    - Z will be standard normal (mean 0, SD 1)
- Large samples also justify using sample SD ($\widehat{\sigma}$) in place of population SD (σ).

Finding the p-value

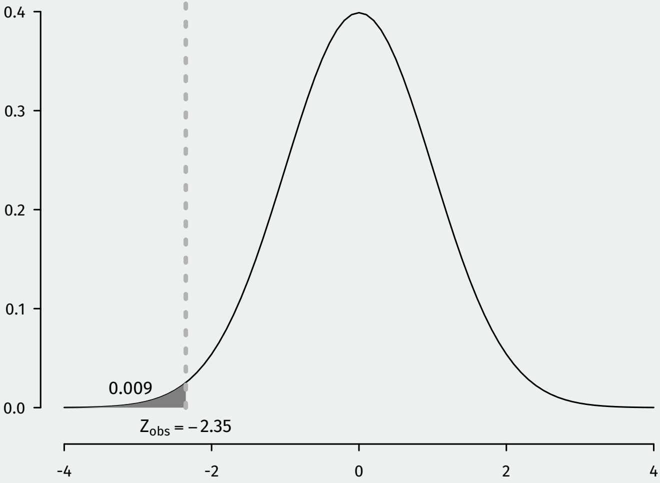
- **Step 4**: determine the p-value.
  - The **p-value** is the probability of observing a test statistic as extreme as Z_{obs}, if the null hypothesis is true.
  - Smaller p-values $\rightsquigarrow$ data less likely under the null $\rightsquigarrow$ null less plausible
- How to calculate?
  - We know Z is distributed standard normal $\rightsquigarrow$ use R!

Standard normal probabilities in R

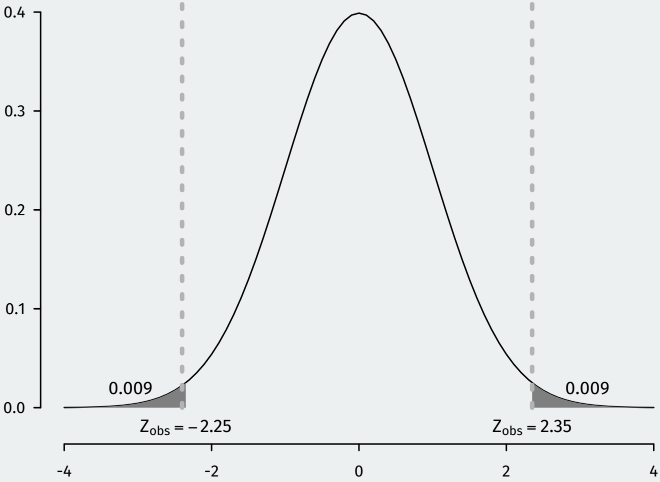
- The `pnorm(x)` function will give the probability of being less x in a standard normal:

```
pnorm(-2.35)
```

```
## [1] 0.00939
```


0.4
0.3
0.2
0.1
0.0
0.009
$Z_{obs} = -2.35$
-4
-2
0
2
4

One-sided vs. two-sided tests



- two-sided p-value: 0.018

Problem of small samples

- Central limit theorem justifies the z-test we've been doing.
    - "Sums and means of random variables tend to be normally distributed as sample sizes get big."
- What if our sample sizes are low?
    - Distribution of $\overline{X}$ will be unknown
    - $\rightsquigarrow$ can't determine p-values
    - $\rightsquigarrow$ can't get z values for confidence intervals
- Very difficult to get around this problem without more information.

Solution to small samples?

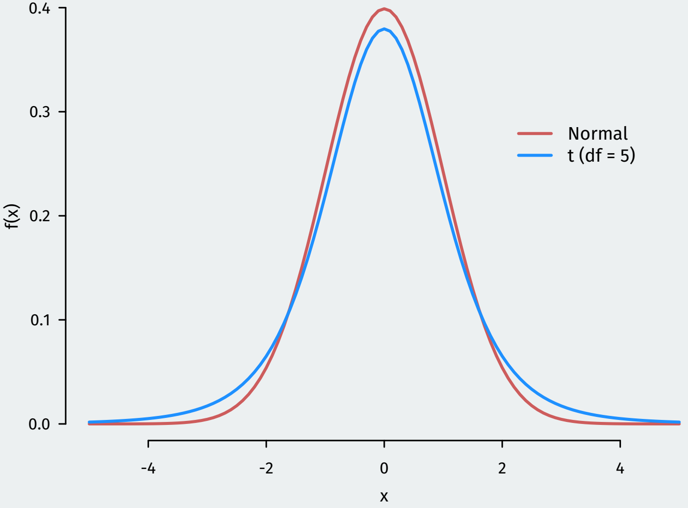
- Common approach: assume data X_i are **normally distributed**
  - THIS IS AN ASSUMPTION, PROBABLY IS WRONG.
  - For instance, if X_i is binary, then it is very wrong.
- If true, then we can determine the distribution of the following test statistic:

$$T = \frac{\overline{X} - \mu}{\widehat{\mathsf{SE}}} \sim t_{n-1}$$

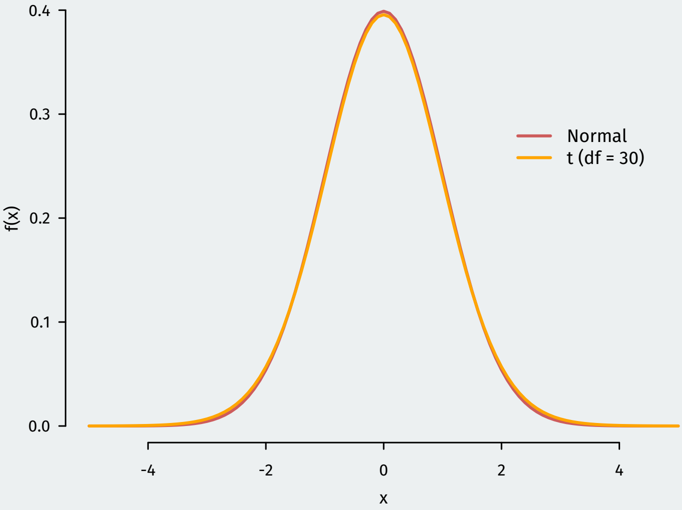
- T follows a Student's t distribution with $n - 1$ degrees of freedom.
  - Degrees of freedom determines the spread of the distribution.
  - Centered around 0
  - Similar to normal with fatter tails $\rightsquigarrow$ higher likelihood of extreme events.

Who was Student?

Student's t distribution



Student's t distribution



z-test vs t-test

- z-tests are what we have seen: relies on the normal distribution.
    - Justified in large samples (roughly $n > 30$) by CLT
- t-tests rely on the the t-distribution for calculating p-values.
    - Justified in small samples if data is normally distributed.
- Common practice is to use t-tests all the time because t is "conservative"
    - $\rightsquigarrow$ p-values will always be larger under t-test
    - $\rightsquigarrow$ always less likely to reject null under t
    - t-distribution converges to standard normal as $n \to \infty$
- R will almost always calculate p-values for you, so details of t-distribution aren't massively important.