

# **Power and Bootstrap**

Gov 2001: Quantitative Social Science Methods I

Scott Cunningham

Harvard University

Spring 2026

# Today's Reading

## Required

- **Aronow & Miller**, §3.3.3: Power (pp. 138–142)
- **Aronow & Miller**, §3.4.3: Bootstrap (pp. 145–150)
- **Blackwell**, Ch. 4 (finish)

**Last probability lecture before the midterm!**

## Two Types of Errors

When we make a decision, we might be wrong:

	$H_0$ True	$H_0$ False
Reject $H_0$	Type I Error	Correct!
Fail to Reject	Correct!	Type II Error

- **Type I Error:** False positive. Convicting an innocent person.
- **Type II Error:** False negative. Letting a guilty person go free.

## Type I Error Rate = $\alpha$

### Type I Error

$$\alpha = \Pr(\text{Reject } H_0 \mid H_0 \text{ true})$$

### **This is our significance level!**

When we set  $\alpha = 0.05$ , we're accepting a 5% chance of Type I error.

### **Why 5%?**

- Tradition (thanks, Fisher)
- Balances false positives against power
- Other fields use different conventions (particle physics:  $5\sigma$ )

## Type II Error and Power

### Type II Error

$$\beta = \Pr(\text{Fail to reject } H_0 \mid H_0 \text{ false})$$

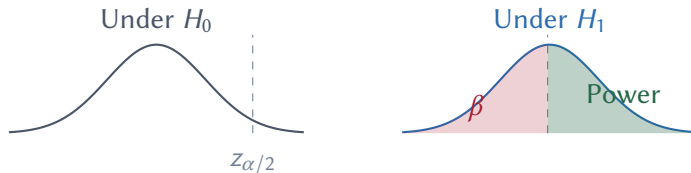
### Power

$$\text{Power} = 1 - \beta = \Pr(\text{Reject } H_0 \mid H_0 \text{ false})$$

**Power** = probability of detecting a real effect when one exists.

Higher power is better. We want to find effects that are really there.

# Visualizing Power



**Left:** Distribution under  $H_0$ . **Right:** Distribution under  $H_1$ .

Power = green area.  $\beta$  = red area.

# What Affects Power?

**Power increases when:**

1. **Effect size is larger:** Easier to detect big effects
2. **Sample size is larger:** More precise estimates, smaller SE
3. **Variance is smaller:** Less noise, clearer signal
4.  **$\alpha$  is larger:** More willing to reject  $\Rightarrow$  more rejections

**The tradeoff:** Increasing  $\alpha$  increases power but also Type I error.  
We typically fix  $\alpha = 0.05$  and increase  $n$  to get power.

## Power Calculation Example

**Setup:** Testing  $H_0 : \mu = 0$  vs.  $H_1 : \mu \neq 0$

True effect:  $\mu = 0.5$ , Standard deviation:  $\sigma = 2$ , Sample size:  $n = 64$

**Standard error:**  $SE = \sigma/\sqrt{n} = 2/8 = 0.25$

**Under  $H_0$ :** Reject if  $|\bar{Y}| > 1.96 \times 0.25 = 0.49$

**Under  $H_1$**  (true  $\mu = 0.5$ ):

$$\begin{aligned}\text{Power} &= \Pr(|\bar{Y}| > 0.49 \mid \mu = 0.5) \\ &\approx \Pr(\bar{Y} > 0.49) \quad (\text{ignoring left tail}) \\ &= \Pr\left(Z > \frac{0.49 - 0.5}{0.25}\right) = \Pr(Z > -0.04) \approx 0.52\end{aligned}$$

Only 52% power—we'd miss this effect half the time!



## Power and Sample Size Planning

**Before running a study:** Calculate required sample size for adequate power.

**Convention:** Target power = 0.80 (80%)

**Formula** (for two-sided test of mean):

$$n = \left( \frac{(z_{\alpha/2} + z_{\beta}) \cdot \sigma}{\mu_1 - \mu_0} \right)^2$$

where  $z_{\beta}$  is the z-value for desired power (e.g.,  $z_{0.20} = 0.84$  for 80% power).

**Example:**  $\sigma = 2$ ,  $\mu_1 - \mu_0 = 0.5$ , 80% power:

$$n = \left( \frac{(1.96 + 0.84) \times 2}{0.5} \right)^2 = (11.2)^2 \approx 126$$

# Power in Political Science Research

## Many studies are underpowered:

- Median power in social science: ~35% (Button et al., 2013)
- Small effects + limited samples = low power

## Political science examples:

- GOTV effects (~2–3 pp) need  $n \approx 5,000+$  for 80% power
- Survey experiments with many conditions: power drops rapidly
- Cross-national studies: 30 countries  $\Rightarrow$  low power for small effects

**Best practice:** Power analysis before collecting data.

# Pre-Study Power Analysis: The Workflow

Before you collect data, specify:

1. **Expected effect size:** Based on prior literature or minimum meaningful effect
2. **Expected variability:** From prior studies or pilot data
3. **Target power:** Usually 80% (sometimes 90% for expensive studies)
4. **Alpha level:** Usually 0.05
5. **Calculate required  $n$ :** Using formulas or simulation

**Key point:** You must specify the effect size *before* seeing data.

This is why pre-registration matters—it forces you to commit to these choices.

## Example: Planning a GOTV Experiment

**Research question:** Does canvassing increase voter turnout?

### Step 1: Expected effect size

- Prior GOTV studies show effects of 2–5 percentage points
- We'll target detecting a **3 pp effect** (conservative)

### Step 2: Expected variability

- Baseline turnout  $\approx 40\%$
- For binary outcome:  $\sigma \approx \sqrt{0.4 \times 0.6} \approx 0.49$

### Step 3: Calculate required $n$

$$n = 2 \times \left( \frac{(1.96 + 0.84) \times 0.49}{0.03} \right)^2 \approx 4,180 \text{ total}$$

Need ~2,000 per treatment arm for 80% power.

# When CLT Doesn't Apply

## The CLT requires:

- I.I.D. observations
- Finite variance
- “Large enough”  $n$

## What if:

- Sample size is small?
- Distribution is highly skewed?
- We want inference for a complicated estimator (median, ratio, etc.)?

## Solution: The Bootstrap

# The Bootstrap Idea

**The problem:** We want to know the sampling distribution of  $\hat{\theta}$ , but we only have one sample.

**The insight:** Treat the sample as a “stand-in” for the population.

**The procedure:**

1. Resample *with replacement* from your data
2. Compute  $\hat{\theta}$  on the resample
3. Repeat many times (e.g., 10,000)
4. Use the distribution of resampled  $\hat{\theta}$ s as the sampling distribution

# Bootstrap Procedure

**Original sample:**  $Y_1, Y_2, \dots, Y_n$

**For**  $b = 1, 2, \dots, B$ :

1. Draw a sample of size  $n$  **with replacement** from  $(Y_1, \dots, Y_n)$
2. Call this  $Y_1^{*b}, Y_2^{*b}, \dots, Y_n^{*b}$
3. Compute  $\hat{\theta}^{*b}$  on this bootstrap sample

**Result:**  $\hat{\theta}^{*1}, \hat{\theta}^{*2}, \dots, \hat{\theta}^{*B}$

**Use this distribution to:**

- Estimate SE:  $\widehat{SE} = (\hat{\theta}^{*1}, \dots, \hat{\theta}^{*B})$
- Construct CI: Use percentiles (e.g., 2.5th and 97.5th)

## Bootstrap Example: Median Income

**Data:** 50 income observations. Median = \$52,000.

**Problem:** No simple formula for SE of the median!

**Bootstrap:**

1. Resample 50 incomes with replacement
2. Compute median of resample
3. Repeat 10,000 times

**Result:** 10,000 bootstrap medians

- Bootstrap SE: \$3,200
- 95% CI: [\$46,000, \$58,500] (2.5th and 97.5th percentiles)



# Bootstrap Confidence Intervals

**Two common methods:**

**1. Percentile method** (simplest):

$$CI = \left[ \hat{\theta}_{(\alpha/2)}^*, \hat{\theta}_{(1-\alpha/2)}^* \right]$$

Use the  $\alpha/2$  and  $(1 - \alpha/2)$  quantiles of bootstrap distribution.

**2. Normal approximation:**

$$CI = \hat{\theta} \pm z_{\alpha/2} \times \widehat{SE}_{boot}$$

Use bootstrap SE with normal critical values.

The percentile method is more robust to skewness.

# Why Does Bootstrap Work?

**Key insight:** The relationship between

Sample  $\leftrightarrow$  Population

is similar to the relationship between

Bootstrap sample  $\leftrightarrow$  Original sample

**For large  $n$ :**

- The sample distribution approximates the population distribution
- Resampling from the sample mimics resampling from the population
- The bootstrap distribution approximates the true sampling distribution

This is the “plug-in principle” applied to distributions.

# When Bootstrap Works (and Doesn't)

## Bootstrap works well for:

- Means, medians, quantiles
- Regression coefficients
- Most “smooth” functions of the data

## Bootstrap can fail for:

- Extremes (max, min)
- Very small samples
- Non-I.I.D. data (need modified versions)
- Parameters on the boundary (e.g., variance = 0)

**Rule of thumb:** If the estimator is consistent and asymptotically normal, bootstrap usually works.

# Bootstrap in R

## Simple implementation:

```
# Original statistic
theta_hat <- median(data)
# Bootstrap
B <- 10000
theta_boot <- numeric(B)
for (b in 1:B) {
  boot_sample <- sample(data, replace = TRUE)
  theta_boot[b] <- median(boot_sample)
}
# SE and CI
se_boot <- sd(theta_boot)
ci_boot <- quantile(theta_boot, c(0.025, 0.975))
```

Or use the boot package for more features.

## Summary: Errors and Power

Concept	Definition	Typical Value
Type I Error ( $\alpha$ )	$\Pr(\text{reject } H_0 \mid H_0 \text{ true})$	0.05
Type II Error ( $\beta$ )	$\Pr(\text{fail to reject} \mid H_0 \text{ false})$	0.20
Power	$1 - \beta$	0.80

**Power depends on:** Effect size, sample size, variance,  $\alpha$

# Key Takeaways

1. **Type I error** = false positive; controlled by  $\alpha$
2. **Type II error** = false negative; related to power
3. **Power** = probability of detecting a real effect
4. **Plan sample size** to achieve adequate power (usually 80%)
5. **Bootstrap** provides inference when CLT is questionable
6. **Bootstrap CI**: Resample, compute statistic, use percentiles

# Midterm Preview

**Midterm Exam:** Covers Weeks 1–7

## Topics:

- Probability: axioms, conditional probability, Bayes' Rule
- Random variables: PMF, PDF, CDF, expectation, variance
- Joint distributions, conditional expectation, CEF
- Sampling distributions, LLN, CLT
- Estimation: bias, variance, MSE, consistency
- Confidence intervals and hypothesis testing

**After spring break:** We start regression!

# R Code: Power Analysis

**Goal:** Understand how power depends on effect size and sample size.

**Topics covered:**

- Computing power for different scenarios
- Power curves across effect sizes
- Sample size determination

Code and figures available in the course repository.



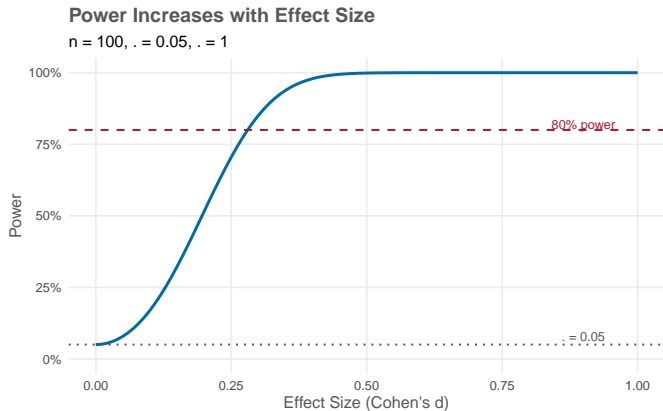
# Power Calculation Function

```
# Power for a z-test
calculate_power <- function(d, n, alpha = 0.05) {
  se <- 1 / sqrt(n) # Assuming sigma = 1
  z_crit <- qnorm(1 - alpha/2)

  # Power = P(reject | H1 true)
  power <- pnorm(-z_crit + d * sqrt(n)) +
    pnorm(-z_crit - d * sqrt(n))
  return(power)
}

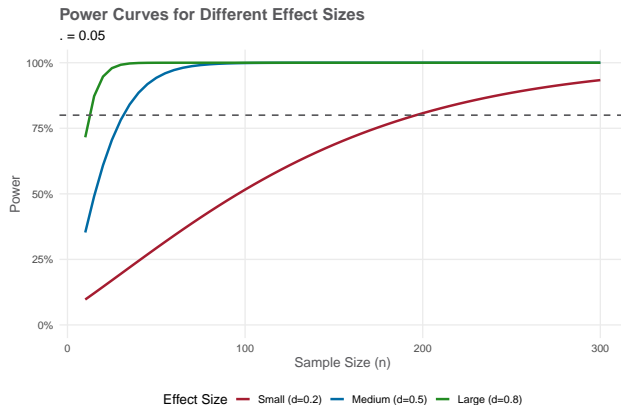
# Example: d = 0.3, n = 100
calculate_power(d = 0.3, n = 100) # About 85%
```

# Power Increases with Effect Size



With  $n = 100$ , we need  $d \approx 0.28$  for 80% power.

# Power Curves for Different Effect Sizes



Small effects need much larger samples to detect reliably.

## Sample Size for 80% Power

```
# Find required n for target power
find_sample_size <- function(target_power, d, alpha = 0.05) {
  for (n in 5:2000) {
    if (calculate_power(d, n, alpha) >= target_power) {
      return(n)
    }
  }
  return(NA)
}
```

```
# Required n for different effect sizes at 80% power
find_sample_size(0.80, d = 0.2) # Small effect
find_sample_size(0.80, d = 0.5) # Medium effect
find_sample_size(0.80, d = 0.8) # Large effect
```

# Looking Ahead

**Spring Break:** March 15–23

**Week 8:** What Is Regression?

- The Best Linear Predictor (BLP)
- OLS as sample BLP
- Connection to CEF

**Reading:**

- Blackwell Ch. 5
- A&M §2.2.4
- Angrist & Pischke Ch. 3.1

The second half of the course: applying what we've learned to regression.