Advanced Econometrics: Inference

Huan Deng

April 29, 2025

Outline

Introduction

Overview

Today, we will discuss inference

1. After getting estimates, we need to ask two questions
2. First, whether the estimates converge to our estimands if we have infinite data. This speaks to identification.
3. Second, how to quantify the uncertainty of our estimates.
4. We will consider two types of inference methods: population-based inference (sampling-based uncertainty), and Fisherian Approach (designed-based uncertainty)

Overview

We must answer two questions:

- **Identification**: suppose we have infinite data, can we uncover our estimand β
- Equivalently, does our estimate $\hat{\beta}$ converge to our estimand β?
- **Quantifying Uncertainty of our estimate**: we need to characterize the variation of our estimate if we can resample our data or reassign the treatment
- I won't go into the Frequentist V.S. Bayesian fight; instead, I will introduce mainstream methods we use in empirical research

Large Sample Distribution of OLS

Large Sample Distribution of OLS

A Brief Review of Large Sample Asymptotics

Overview

- We first need to review basic large sample asymptotics
- I follow the Chapter 6 of Bruce Hansen's textbook
- Key concepts include convergence in probability and in distribution
- We want our estimates to have some nice properties: our estimates will be indistinguishable with our estimand with infinite data; our estimates will be well approximated by a normal distribution with infinite data
- Thus we can identify some interesting estimand and do inference

Modes of Convergence

Definition 6.1: Convergence in Probability

$$Z_n \xrightarrow{p} Z \quad \text{if for all } \delta > 0, \quad \lim_{n\to\infty} P[\|Z_n - Z\| \leq \delta] = 1. \tag{1}$$

Definition 6.2: Convergence in Distribution
Let Z_n be a sequence of random vectors with distributions $F_n(u) = P[\ Z_n \leq u\ ]$, we define convergence in distribution as below:

$$Z_n \xrightarrow{d} Z \quad \text{if} \quad F_n(u) \to F(u) \text{ as } n \to \infty. \tag{2}$$

We refer to Z and its distribution $F(u)$ as the asymptotic distribution or large sample distribution of Z_n

Weak Law of Large Numbers

Theorem (WLLN)

If $Y_i \in \mathbb{R}^k$ are i.i.d. and $E\|Y\| < \infty$, then as $n \to \infty$,

$$\bar{Y} = \frac{1}{n}\sum_{i=1}^{n} Y_i \xrightarrow{p} E[Y]. \tag{3}$$

Central Limit Theorem

Theorem (Multivariate Lindeberg-Lévy CLT)

If $Y_i \in \mathbb{R}^k$ are i.i.d. and $E\|Y\|^2 < \infty$, then as $n \to \infty$

$$\sqrt{n}(\bar{Y} - \mu) \xrightarrow{d} N(0, V), \tag{4}$$

where $\mu = E[Y]$ and $V = E[(Y - \mu)(Y - \mu)^\top]$.

Continuous Mapping Theorem

- Continuous Mapping Theorem (CMT): Continuous functions are limit-preserving.
- Two types of CMT: for convergence in probability and convergence in distribution.

Theorem (CMT for Convergence in Probability)

If $Z_n \xrightarrow{p} c$ and $g(u)$ is continuous at c, then

$$g(Z_n) \xrightarrow{p} g(c). \tag{5}$$

Theorem (CMT for Convergence in Distribution)

If $Z_n \xrightarrow{d} Z$ and $g : \mathbb{R}^m \to \mathbb{R}^k$ has discontinuity set D_g such that $P[Z \in D_g] = 0$, then

$$g(Z_n) \xrightarrow{d} g(Z). \tag{6}$$

Delta Method

Differentiable functions of asymptotically normal random estimates are asymptotically normal.

Theorem 6.8 Delta Method

Let $\mu \in \mathbb{R}^k$ and $g(u) : \mathbb{R}^k \to \mathbb{R}^q$. If

$$\sqrt{n}(\hat{\mu} - \mu) \xrightarrow{d} \xi, \tag{7}$$

where $g(u)$ is continuously differentiable in a neighborhood of μ, then as $n \to \infty$

$$\sqrt{n}(g(\hat{\mu}) - g(\mu)) \xrightarrow{d} \mathbf{G}'\xi. \tag{8}$$

where $\mathbf{G}(u) = \frac{\partial}{\partial u} g(u)'$ and $\mathbf{G} = \mathbf{G}(\mu)$. In particular, if $\xi \sim N(0, \mathbf{V})$, then as $n \to \infty$

$$\sqrt{n}(g(\hat{\mu}) - g(\mu)) \xrightarrow{d} N(0, \mathbf{G}'\mathbf{V}\mathbf{G}). \tag{9}$$

Slutsky's Theorem

If $X_n \xrightarrow{d} X$ and $Y_n \xrightarrow{p} c$, where c is a constant, then:

$$X_n Y_n \xrightarrow{d} Xc. \tag{10}$$

$$X_n + Y_n \xrightarrow{d} X + c. \tag{11}$$

This will be useful for deriving the asymptotic normality of our OLS estimate

Smooth Function Model

- Suppose our parameter of interest $\theta = g(\mu)$ is not a population moment so it does not have a direct moment estimator, where $\mu = E[\ h(Y)\ ]$ and $g(\mu)$ is smooth in a suitable sense
- Then use a plug-in estimator
  - We first estimate μ as $\hat{\mu} = \frac{1}{n}\sum h(Y_i)$
  - Then we get $\hat{\theta} = g(\hat{\mu})$
- The consistency of θ requires that $h(Y)$ has finite expectation and $g(u)$ is continuous
- The asymptotic normality of θ requires that $h(Y)$ has finite variance and $g(u)$ is continuously differentiable

Smooth Function Model

Theorem (Consistency)

If $Y_i \in \mathbb{R}^m$ are i.i.d., $h(u) : \mathbb{R}^m \to \mathbb{R}^k$, $E\|h(Y)\| < \infty$, and $g(u) : \mathbb{R}^k \to \mathbb{R}^q$ is continuous at μ, then

$$\theta_n = g(\hat{\mu}) \xrightarrow{p} g(\mu) \quad \text{as } n \to \infty. \tag{12}$$

Theorem (Asymptotic Normality)

If additionally $E\|h(Y)\|^2 < \infty$ and $G(u) = \frac{\partial g'}{\partial u}$ is continuous in a neighborhood of μ, then

$$\sqrt{n}(\theta_n - \theta) \xrightarrow{d} N(0, G'VG), \tag{13}$$

where $V = E[(h(Y) - \mu)(h(Y) - \mu)^T]$.

Large Sample Distribution of OLS

Asymptotics of OLS

Estimand, Estimate, and Estimator (Again)

- Before discussing asymptotics of OLS, we must ask: why do we care about the properties of our estimates?
- As we will show, under some assumptions, we can prove that the OLS estimate $\hat{\beta}_{OLS} = (\frac{1}{N}X'X)^{-1}(\frac{1}{N}X'Y)$ converges to the BLP coefficient $\beta_{BLP} = (E[X_iX_i'])^{-1}(EX_iY_i)$
- But why should we care about the BLP coefficient? Does it deliver things beyond giving us the best linear prediction for Y?
- Prediction is important, but it is rarely the thing we mainly care about.

Why Do We Care About the BLP Coefficient?

- Justification one: BLP offers the best linear prediction of CEF, which carries a causal interpretation under random assignment.
  - Recall that $Y = m(X) + e$ where $m(x) = E(Y|X = x)$ is the CEF function
  - Recall that $ATE = E[Y|x_1] - E[Y|x_0] = m(x_1) - m(x_0)$ when we have random assignment
  - So if CEF is linear or not too far from being linear, then BLP can be a good or even perfect approximation of CEF.

Why Do We Care About the BLP Coefficient?

- Justification two: if the behavioral model takes the same formula as the BLP model
  - Generally, the behavioral model can be written as $Y = g(X, \epsilon)$
  - The BLP model is: $Y = X\beta_{BLP} + \epsilon$, note that $E[X\epsilon] = 0$ holds by definition
  - If $g(X, \epsilon) = X\beta_{BLP} + \epsilon$, then of course the BLP coefficients have causal interpretations.
  - So we often see people **assume** that $E[X\epsilon] = 0$ despite this condition holds automatically for the BLP model
  - We make the assumption for our behavioral model, not for the BLP model
  - A linear, additive functional form + zero correlation between X and ϵ – not a weak assumption

Overview

- We apply WLLN and CMT to establish the consistency of our OLS estimate $\hat{\beta}$
  - $\hat{\beta}$ can be written as a continuous function of a set of sample moments
  - By WLLN, sample moments converge in probability to population moments
  - By CMT: continuous functions preserve convergence in probability
- We apply CLT and Slutsky's Theorem to establish asymptotic normality of $\hat{\beta}$

Assumptions

To ensure consistency of OLS, we assume:

1. **Random Sampling:** The variables $(Y_i, X_i), i = 1, ..., n$, are i.i.d.
2. **Finite Second Moments:** $E[Y]^2 < \infty$, and $E||X||^2 < \infty$
3. **Finite Fourth Moments:** $E[Y]^4 < \infty$, and $E||X||^4 < \infty$
4. **Full Rank:** $Q_{XX} = E[XX']$ is positive definite.

Proof of Consistency

By the Weak Law of Large Numbers (WLLN):

$$\hat{Q}_{XX} = \frac{1}{n}\sum_{i=1}^{n} X_i X_i' \xrightarrow{p} E[XX'] = Q_{XX}, \tag{14}$$

$$\hat{Q}_{XY} = \frac{1}{n}\sum_{i=1}^{n} X_i Y_i \xrightarrow{p} E[XY] = Q_{XY}. \tag{15}$$

Applying the Continuous Mapping Theorem:

$$\hat{\beta} = \hat{Q}_{XX}^{-1}\hat{Q}_{XY} \xrightarrow{p} Q_{XX}^{-1} Q_{XY} = \beta. \tag{16}$$

Thus, $\hat{\beta}$ is a consistent estimate of β_{BLP}.

Proof of Asymptotic Normality

We derive the below expression:

$$
\begin{aligned}
\hat{\beta}_{OLS} &= (\frac{1}{n}\sum_{i=1}^{n} X_i X_i')^{-1}(\frac{1}{n}\sum_{i=1}^{n} X_i Y_i) \\
&= (\frac{1}{n}\sum_{i=1}^{n} X_i X_i')^{-1}(\frac{1}{n}\sum_{i=1}^{n} X_i (X_i'\beta_{BLP} + \epsilon_i)) \\
&= \beta_{BLP} + (\frac{1}{n}\sum_{i=1}^{n} X_i X_i')^{-1}(\frac{1}{n}\sum_{i=1}^{n} X_i \epsilon_i) \qquad (17)
\end{aligned}
$$

Thus, we can write out the below expression that facilitates the application of CLT:

$$
\sqrt{n}(\hat{\beta}_{OLS} - \beta_{BLP}) = \left(\frac{1}{n}\sum_{i=1}^{n} X_i X_i'\right)^{-1} \sqrt{n}\left(\frac{1}{n}\sum_{i=1}^{n} X_i \epsilon_i\right) \qquad (18)
$$

Proof of Asymptotic Normality

Using the Central Limit Theorem (CLT), we obtain:

$$\sqrt{n}\left(\frac{1}{n}\sum_{i=1}^{n} X_i\epsilon_i\right) \xrightarrow{d} N(0,\Omega), \tag{19}$$

where:

$$\Omega = E[XX'\epsilon^2]. \tag{20}$$

applying the Slutsky's Theorem, we get:

$$\sqrt{n}(\hat{\beta}-\beta) \xrightarrow{d} N(0, Q_{XX}^{-1}\Omega Q_{XX}^{-1}). \tag{21}$$

We denote $V_\beta = Q_{XX}^{-1}\Omega Q_{XX}^{-1}$, "Sandwich form"

Heteroskedastic Covariance Matrix Estimation

We don't know V_β, so we need to estimate it. A heteroskedasticity-robust estimator of V_β is given by:

$$\hat{V}_{HC} = \hat{Q}_{XX}^{-1}\hat{\Omega}\hat{Q}_{XX}^{-1}, \tag{22}$$

where:

$$\hat{\Omega} = \frac{1}{n}\sum_{i=1}^{n} X_i X_i' \hat{\epsilon}_i{}^2. \tag{23}$$

It is straightforward (but not necessarily easy) to show that $\hat{V}_{HC} \xrightarrow{p} V_\beta$. Check Hansen's textbook for a proof if you are interested.

Function of Parameters

In econometrics, we often analyze functions of estimated parameters, such as nonlinear transformations:

$$\theta = g(\beta), \tag{24}$$

where $g(\cdot)$ is a differentiable function of the OLS estimate $\hat{\beta}_{OLS}$.

Delta Method for Functions of Parameters

Using the Delta Method, we approximate the asymptotic distribution of $g(\hat{\beta})$:

$$\sqrt{n}(g(\hat{\beta}) - g(\beta)) \xrightarrow{d} N(0, G' V_{\beta} G), \tag{25}$$

where:

$$G = \frac{\partial g(\beta)}{\partial \beta}. \tag{26}$$

This provides a basis for inference on transformed parameters.

Variance of Function of Parameters

Denote $V_\theta = G' V_\beta G$, we can also use a plug-in estimator:

$$\hat{V}_\theta = \hat{G}' \hat{V}_\beta \hat{G}, \tag{27}$$

where $\hat{G} = \frac{\partial g(\hat{\beta}_{OLS})}{\partial \beta}$ and $\hat{V}_\beta$ can be the HC estimate we just introduced.

Definition of t-Statistic

- Suppose $\theta = g(\beta) : R^k \to R$, consider the below statistic:

$$T(\theta) = \frac{\hat{\theta} - \theta}{SE(\hat{\theta})}, \tag{28}$$

where $SE(\hat{\theta}) = \sqrt{\frac{1}{n}\hat{V}_\theta}$

- $T(\theta)$ converges to the standard normal distribution:

$$\begin{aligned} T(\theta) &= \frac{\hat{\theta} - \theta}{SE(\hat{\theta})} \\ &= \frac{\sqrt{n}(\hat{\theta} - \theta)}{\sqrt{\hat{V}_\theta}} \\ &\xrightarrow{d} \frac{N(0, V_\theta)}{\sqrt{V_\theta}} \\ &= Z \sim N(0, 1) \end{aligned} \tag{29}$$

- Sometimes, we care about the absolute value of $T(\theta)$, by CMT, we also have $|T(\theta)| \xrightarrow{d} |Z|$

Confidence Interval

- $\hat{C} = [\hat{L}, \hat{U}]$ is called an interval estimator
- The coverage probability of the interval $\hat{C}$ is $P[\theta \in \hat{C}]$
- $\hat{C}$ is called a **confidence interval** when the goal is to set the coverage probability to be a pre-specified target such as 95%.
- $\hat{C}$ is called a $1 - \alpha$ confidence interval if $inf_{\theta} P[\theta \in \hat{C}] = 1 - \alpha$
- When $\hat{\theta}$ is asymptotically normal, the confidence interval takes the form:

$$\hat{C} = [\hat{\theta} - c \times SE(\hat{\theta}), \hat{\theta} + c \times SE(\hat{\theta})], \tag{30}$$

  where $SE(\hat{\theta})$ is the standard error of $\hat{\theta}$, and c is the $1 - \frac{\alpha}{2}$ quantile of the standard normal distribution.
- It is easy to show that the coverage probability of $\hat{C}$ converges to $1 - \alpha$

Definition

- The Wald statistic is used to conduct inference about multiple parameters in econometric models. It is given by:

$$W(\theta) = (\hat{\theta} - \theta)'(\frac{1}{n}\hat{V}_{\theta})^{-1}(\hat{\theta} - \theta), \tag{31}$$

  where $\theta \in R^q$
- It is easy to show that $W(\theta)$ converges to the sum of squared standard normal random variables, which means $W(\theta) \xrightarrow{d} \chi^2_q$
- Like for t-statistic and confidence interval, we will use the Wald-statistic to construct confidence regions.

Definition of Confidence Regions

- A confidence region $\hat{C}$ is a set in R^q that intended to cover the true parameter value θ with a pre-selected probability $1-\alpha$.
- Thus an ideal confidence region has the coverage probability $P[\theta \in \hat{C}] = 1-\alpha$
- A good choice for a confidence region is the ellipse:

$$\hat{C} = \{\theta : W(\theta) \leq c_{1-\alpha}\}, \tag{32}$$

  where $c_{1-\alpha}$ is the $1-\alpha$ quantile of the χ^2_q
- Again, it is easy to show that the coverage probability of $\hat{C}$ converges to $1-\alpha$.

Bootstrap

Bootstrap

Bootstrap as an Algorithm

Overview

- Large sample asymptotics prodives us with a method to derive asymptotic distribution of our estimates, with which we can conduct the two important statistical inference: constructing confidence intervals and doing hypothesis testing
- But in many applications, it is hard to derive the asymptotic distribution of our estimates. For example, multi-step estimation.
- Then we can use the Bootstrap method to conduct inference.
- In addition, Bootstrap can handle non i.i.d. sampling and provide asymptotic refinement.
- In practice, when it is easy to derive the asymptotic distribution, do it. Bootstrap is computationally expensive as it involves repeating the estimation many times.

How the Bootstrap Works

- Original data sample: N observations from the population.
- Bootstrap sample: N observations resampled with replacement.
- Ensures each draw remains independent and from the same distribution.
- We repeat this process B times. The number of bootstrap draws, B, is often called the "number of bootstrap replications".
- This is nonparametric, we put no parametric assumptions on data distribution.

How the Bootstrap Works

- Consider the estimate $\hat{\theta}$, for each bootstrap sample, we can compute the estimate, which we refer as $\hat{\theta}^*(b)$
- We have thus created a new dataset of bootstrap draws $\{\theta^*(b) : b = 1, ..., B\}$. By construction the draws are independent across b and identically distributed.
- Note that the bootstrap distribution and any statistics derived from it depend on both the original sample we have and the number of bootstrap replications
- Generally speaking, a larger B leads to better approximation. But in practice, we need to consider the computational costs.

Bootstrap Variance

- We can use the Bootstrap method to estimate the variance-covariance matrix of θ:

$$\hat{V}_{\hat{\theta}}^{\text{boot}} = \frac{1}{B-1} \sum_{b=1}^{B} \left(\hat{\theta}^*(b) - \bar{\theta}^* \right) \left(\hat{\theta}^*(b) - \bar{\theta}^* \right)', \tag{33}$$

where $\bar{\theta}^* = \frac{1}{B} \sum_{b=1}^{B} \hat{\theta}^*(b)$

- Then the bootstrap standard error is given by:

$$s_{\hat{\theta}}^{\text{boot}} = \sqrt{\hat{V}_{\hat{\theta}}^{\text{boot}}} \tag{34}$$

- Standard errors are conventionally reported to convey the precision of the estimator.
- Bootstrap standard errors can also be used to create normal-approximation bootstrap confidence interval:

$$C^{\text{nb}} = \left[\hat{\theta} - z_{1-\alpha/2} s_{\hat{\theta}}^{\text{boot}}, \quad \hat{\theta} + z_{1-\alpha/2} s_{\hat{\theta}}^{\text{boot}} \right] \tag{35}$$

Percentile Interval

- The above CI can be seen as replacing the SE of CI derived from the large sample asymptotics with the bootstrap SE
- We can also create CI using the method of Percentile Interval, which is based on the quantiles of the bootstrap distribution.
- The bootstrap percentile interval is formed from the quantiles q^*_α of the distribution of bootstrap estimates
- The percentile bootstrap $100(1-\alpha)\%$ confidence interval is:

$$C^{\mathrm{pc}} = \left[q^*_{\alpha/2}, q^*_{1-\alpha/2} \right], \tag{36}$$

- The percentile interval is transformation-respecting, while the delta-method asymptotic interval and C^{nb} do not share this property.

Bootstrap

A Deeper Dive into Bootstrap

A Deeper Dive into Bootstrap

- For applications, it is often sufficient if one understands the bootstrap as an algorithm.
- The key is that the distribution of any estimator or statistic is determined by the distribution of the data.
- Though we don't know the distribution of the data, we can estimate it by the empirical distribution of the data. This is what the bootstrap does.

A Deeper Dive into Bootstrap

- In general, the distribution of any estimator $\hat{\theta}$ can be written as

$$G_N(u|F) = \text{pr}(\hat{\theta} \leq u|F)$$

- where F is the distribution of the individual observation.
- The notation makes clear that the distribution of the estimator in general depends on the distribution of the data F and the sample size n.

Two Barriers

We want to know G_N but we face two barriers

- We don't know F.
- Even if we know F, we still don't know how the estimator distribution depends on the data distribution F.

How Bootstrap Conquers the Two Barriers

- Estimating F
  - Estimating F by the empirical distribution function (EDF) F_n:

$$F_n(w) = n^{-1} \sum_{i=1}^{n} \mathbf{1}\{W_i \leq w\} \tag{37}$$

  - Replacing F with F_n we obtain the idealized bootstrap estimator of the distribution of θ:

$$G_n^*(u) = G_n(u, F_n) \tag{38}$$

- Estimating $G_n^*(u)$ by simulation
  - Simulation from F_n is sampling with replacement from the original data
  - By making a large number B of such draws we can estimate any feature of G_n^* of interest

An example

- Consider the distribution of a scalar statistic $\hat{\theta}$ formed from a sample of scalar $Y_1, \ldots, Y_N$:

$$\hat{\theta} = h(Y_1, \ldots, Y_N) \tag{39}$$

- We can then write the distribution of the estimator $\hat{\theta}$ as:

$$\begin{aligned} G_n(u, F) &= \Pr(\hat{\theta} \leq u | F) \\ &= \int \mathbf{1}\{h(Y_1, \ldots, Y_n) \leq u\} dF. \end{aligned} \tag{40}$$

- Replacing F with F_n, we get:

$$G_n^*(u) = \int \mathbf{1}\{h(Y_1, \ldots, Y_n) \leq u\} dF_n \tag{41}$$

Computing an Integral by Simulation

- It is clear that the expression for $G_n^*(u)$ is just an integral
- We compute the integral by simulation:

$$
\begin{aligned}
G_n^*(u) &= \int \mathbf{1}\{h(Y_1, \ldots, Y_n) \leq u\} dF_n \\
* &\simeq \frac{1}{B} \sum_{b}^{B} \mathbf{1}\{h(Y_1(b), \ldots, Y_n(b)) \leq u\}
\end{aligned} \tag{42}
$$

Bootstrap Inference

- We conduct Bootstrap to conduct inference for $\hat{\theta}$
- But as a function of the data, the bootstrap estimate is also random, thus has a distribution as well
- We also have the bootstrap version for WLLN, CLT, etc.
- We won't discuss Bootstrap inference here. Read Hansen's textbook if you feel interested.
- Generally speaking, bootstrap consistency requires $B \to \infty$, not just $n \to \infty$

Design-based Approach

Design-based Approach

Fisher's Approach

Overview

- Large sample asymptotics provides us with a method to derive asymptotic distribution of our estimates
- Bootstrap provides a resampling approach to statistical inference
- Now we discuss a "reassigning" approach
- **Fisherian approach**: permutation testing, reassign the treatment status; null hypotheses that restrict how treatment can affect potential outcomes
- Today we discuss Fisher's Exact P-Values in the context of RCT, following Imbens and Rubin (2015)

Overview

- Consider any test statistic T, which is a function of the stochastic assignment vector W and the observed outcomes Y^{obs}
- Now the randomness comes from the assignment but not a sampling from the infinite data
- Our data only realizes one of $\binom{N_c+N_t}{N_t}$ possible values of the assignment vector
- For each possible assignment vector, the **sharp** null hypothesis imputes the missing unrealized potential outcomes
- Thus we can generate the distribution of T
- The test statistic is stochastic solely through the stochastic nature of the assignment vector.
- We refer to the distribution of the statistic determined by the randomization as the randomization distribution of the test statistic T.

Overview

- We compare the observed value of the statistic T^{obs} against the distribution of T
- If T^{obs} looks like an outlier given the distribution of T, then we see it as the evidence **against** the sharp null hypothesis
- A statistical version of "proof by contradiction"
- We compute the P-value as $Pr(T \geq T^{obs})$ and reject the null hypothesis when P-value is smaller than a pre-specified value, like 5%.
- The statistic should be chosen to have statistical power against a scientifically interesting alternative hypothesis.
- Statistical power: the probability of rejecting the false null hypothesis
- Equivalently, the likelihood of getting a P-value that is smaller than a pre-specified value.

Sharp Null Hypothesis

Definition: The null hypothesis assumes no effect of the treatment on any unit:

$$H_0 : Y_i(0) = Y_i(1) \text{ for all } i. \tag{43}$$

This allows inference of all missing potential outcomes.

Test Statistic

The test statistic T must be a function of the assignments and observed outcomes:

$$T(W, Y^{obs}). \tag{44}$$

The difference-in-means estimator provides a natural test statistic:

$$T_{dif} = \left| \frac{1}{n_t} \sum_{i:W_i=1} Y_i - \frac{1}{n_c} \sum_{i:W_i=0} Y_i \right| . \tag{45}$$

Randomization-Based P-Value

The p-value is computed as:

$$P = \frac{\sum I(T \geq T_{obs})}{\text{total random assignments}}. \tag{46}$$

where $I(\cdot)$ is an indicator function. If P is small, we reject H_0.

Design-based Approach

A Simulation Study

A Simulation

The best way to test our understanding of a method is to implement it in a simulation study:

- Here we conduct a power analysis by varying the true level of the treatment effect
- For each value of treatment effect, we create n_{sim} samples by repeatedly sampling n potential outcomes from a known distribution
- For each sample we draw, we conduct the Fisher's test
- If n is too large, we can't afford to exhaust all possible permutations, so we also sample from the set of permutations
- A loop times a loop: the computational cost is quadratic.
- See my R codes for implementation details

Graphical Results from Simulation