Modern Sampling Methods

Class 2: Randomized Experiments

January 10, 2022

Outline

- Potential Outcomes and Treatment Effects
- Randomized Experiments
- Estimation of ATE
- Testing ATE
- First Discussion of Experimental Design
- Population vs. Superpopulation
- Fisher's Exact Test
- Covariates
- Propensity Scores

Potential Outcomes Framework

Units $i = 1, \ldots, n$.

Treatment: $T_i = 0, 1$.

Potential Outcomes:

$$\begin{aligned} Y_i(0) &= \text{ potential outcome if } T_i = 0 \\ Y_i(1) &= \text{ potential outcome if } T_i = 1. \end{aligned}$$

Observed outcome:

$$Y_i = (1 - T_i)\, Y_i(0) + T_i\, Y_i(1).$$

May also observe covariates X_i (invariant to treatment).

Key Estimands

Average treatment effect:

$$\begin{aligned} ATE &= E\left[Y_i(1) - Y_i(0)\right] \\ &= E\left[Y_i(1)\right] - E\left[Y_i(0)\right]. \end{aligned}$$

Average effect on the treated:

$$\begin{aligned} ATT &= E\left[Y_i(1) - Y_i(0) \mid T_i = 1\right] \\ &= E\left[Y_i(1) \mid T_i = 1\right] - E\left[Y_i(0) \mid T_i = 1\right]. \end{aligned}$$

Assignment Mechanism

The *assignment mechanism* is the procedure that determines T_i given $Y_i(0)$ and $Y_i(1)$ (and given X_i if we also have covariates).

The assignment mechanism is key to:

- Whether it is possible to learn *ATE* or other quantities of interest from the observed data (identification)
- How to estimate and test for treatment effects

Strategies for causal inference are usually based on assumptions or restrictions on the assignment mechanism.

A Randomized Experiment

Suppose that T_i is randomly assigned (as in a randomized control trial, or RCT). More precisely, suppose that

$$T_i \text{ is independent of } Y_i(0), Y_i(1),$$

which we write as

$$T_i \perp Y_i(0), Y_i(1).$$

Then

$$\begin{aligned} E\left[Y_i \mid T_i = 1\right] &= E\left[Y_i(1) \mid T_i = 1\right] = E\left[Y_i(1)\right] \\ E\left[Y_i \mid T_i = 0\right] &= E\left[Y_i(0) \mid T_i = 0\right] = E\left[Y_i(0)\right]. \end{aligned}$$

Identification

So

$$E\left[Y_i \mid T_i = 1\right] - E\left[Y_i \mid T_i = 0\right] = E\left[Y_i(1)\right] - E\left[Y_i(0)\right] = ATE.$$

Recall:

Identification: the quantity of interest can be expressed in terms of the distribution of the observable variables.

Here the ATE is expressed as a function of the joint distribution of (Y_i, T_i).

(For this to be viable, need $0 < \Pr(T_i = 1) < 1$.)

In fact, randomization of treatment implies that

$$F\left(Y_i \leq y \mid T_i = 1\right) = F\left(Y_i(1) \leq y\right)$$

and

$$F\left(Y_i \leq y \mid T_i = 0\right) = F\left(Y_i(0) \leq y\right).$$

Thus the marginal distributions of $Y_i(0)$ and $Y_i(1)$ are identified.

For example, we can estimate the difference in quantiles

$$Q_\theta(Y_i(1)) - Q_\theta(Y_i(0)).$$

This is often called the "quantile treatment effect."

Randomization Mechanism

When we say "randomized treatment assignment," this can refer to different possible mechanisms, e.g.:

1. T_i is independent Bernoulli with $\Pr(T_i = 1) = p$ for some $p \in (0, 1)$.
2. Given a sample of size n, we assign $n_1 < n$ to treatment with equal probability for any configuration of $n_1 > 0$ treated and $n_0 = n - n_1 > 0$ controls
3. Other schemes are possible as well. More on this in later sessions.

Estimation of ATE

Let

$$\alpha_1 = E\left[Y_i \mid T_i = 1\right], \quad \alpha_0 = E\left[Y_i \mid T_i = 0\right],$$

so $ATE = \alpha_1 - \alpha_0$ under randomized assignment of treatment.

Also let

$$\overline{Y}_1 = \frac{\sum_{i=1}^{n} T_i Y_i}{\sum_{i=1}^{n} T_i} = \frac{\sum_{T_i=1} Y_i}{n_1},$$

and

$$\overline{Y}_0 = \frac{\sum_{i=1}^{n} (1 - T_i) Y_i}{\sum_{i=1}^{n} (1 - T_i)} = \frac{\sum_{T_i=0} Y_i}{n_0},$$

where $n_1 = \sum_{i=1}^{n} T_i$ and $n_0 = n - n_1$. Then

$$\hat{\beta} = \overline{Y}_1 - \overline{Y}_0$$

is an unbiased, consistent, and asymptotically normal estimator of $\beta = \alpha_1 - \alpha_0$ by standard arguments.

We may wish to impose further parametric assumptions for estimation and inference purposes. For example if we assume lognormality:

$$Y_i(0) \sim LN(\mu_0, \tau_0)$$
$$Y_i(1) \sim LN(\mu_1, \tau_1)$$

Then (under randomized treatment)

$$E\left[Y_i \mid T_i = 0\right] = \exp\left(\mu_0 + \tau_0/2\right)$$
$$E\left[Y_i \mid T_i = 1\right] = \exp\left(\mu_1 + \tau_1/2\right).$$

The parameters $(\mu_0, \tau_0, \mu_1, \tau_1)$ could be estimated by maximum likelihood and we could plug them in to estimate ATE.

Testing: Large Samples

Suppose we want to test the null hypothesis

$$H_0 : ATE = 0 \quad \Leftrightarrow \quad \alpha_1 - \alpha_0 = 0,$$

Suppose that $(T_i, Y_i(0), Y_i(1))$ are i.i.d. (this is a strong assumption.)

Let

$$\sigma_1^2 = V\left[Y_i \mid T_i = 1\right],$$
$$\sigma_0^2 = V\left[Y_i \mid T_i = 0\right].$$

Assume that these variances are finite, and also assume that

$$n_1/n \to p \in (0, 1),$$

Then

$$\sqrt{n}\left(\hat{\beta} - (\alpha_1 - \alpha_0)\right) \xrightarrow{d} N\left(0, \frac{\sigma_1^2}{p} + \frac{\sigma_0^2}{1-p}\right).$$

Therefore,

$$\frac{\hat{\beta} - (\alpha_1 - \alpha_0)}{\sqrt{\frac{\hat{\sigma}_1^2}{n_1} + \frac{\hat{\sigma}_0^2}{n_0}}} \xrightarrow{d} N(0,1),$$

where

$$\hat{\sigma}_1^2 = \frac{1}{n_1 - 1} \sum_{T_i=1} (Y_i - \bar{Y}_1)^2,$$

and analogously for $\hat{\sigma}_0^2$.

We could also carry out tests under parametric assumptions like the example above.

Experimental Design: Relative Sample Sizes

Suppose we have n subjects and we need to choose n_1 and n_0. The variance of our usual estimator is

$$\begin{aligned} V\left[\hat{\beta}\right] &= \frac{\sigma_1^2}{n_1} + \frac{\sigma_0^2}{n_0} \\ &= \frac{\sigma_1^2}{pn} + \frac{\sigma_0^2}{(1-p)n}, \end{aligned}$$

where

$$p = n_1/n.$$

We could choose p to minimize the variance of $\hat{\beta}$:

$$\min_{p \in (0,1)} \quad \frac{\sigma_1^2}{pn} + \frac{\sigma_0^2}{(1-p)n}.$$

The first order condition for a minimum is

$$-\frac{\sigma_1^2}{p^2 n} + \frac{\sigma_0^2}{(1-p)^2 n} = 0.$$

Solving for p:

$$\frac{1-p}{p} = \frac{\sigma_0}{\sigma_1}.$$

$$\Rightarrow p^* = \frac{\sigma_1}{\sigma_0 + \sigma_1}.$$

So if $\sigma_0 = \sigma_1$, then $p = 1/2$ minimizes variance, which is quite intuitive.

Experimental Design: Total Sample Size

Suppose we want to choose the total sample size n, taking p as given (e.g. $p = 1/2$).

A large sample size is better, but usually it is costly to obtain more experimental units.

One way to operationalize: suppose we want to be able to detect an effect $(\alpha_1 - \alpha_0)$ of magnitude $\kappa \neq 0$ with a given probability.

This is a condition on the *power* of the test of $\alpha_1 - \alpha_0 = 0$ against the alternative $\alpha_1 - \alpha_0 = \kappa$ given sample size n.

As before let $\hat{\beta} = \overline{Y}_1 - \overline{Y}_0$, and recall that

$$\frac{\hat{\beta} - (\alpha_1 - \alpha_0)}{\sqrt{\frac{\hat{\sigma}_1^2}{n_1} + \frac{\hat{\sigma}_0^2}{n_0}}} \xrightarrow{d} N(0,1),$$

Let t be the t-statistic for testing $\alpha_1 - \alpha_0 = 0$ and let

$$\sigma_\beta^2 = \frac{\sigma_1^2}{pn} + \frac{\sigma_0^2}{(1-p)n}.$$

Then

$$t = \frac{\hat{\beta}}{\sqrt{\frac{\hat{\sigma}_1^2}{n_1} + \frac{\hat{\sigma}_0^2}{n_0}}} \approx \frac{\hat{\beta}}{\sigma_\beta} \approx N\left(\frac{\alpha_1 - \alpha_0}{\sigma_\beta}, 1\right).$$

Then

$$\Pr(|t| > 1.96) = \Pr(t < -1.96) + \Pr(t > 1.96)$$

$$\approx \Phi\left(-1.96 - \frac{\alpha_1 - \alpha_0}{\sigma_\beta}\right) + \left[1 - \Phi\left(1.96 - \frac{\alpha_1 - \alpha_0}{\sigma_\beta}\right)\right].$$

Note that σ_β depends on σ_1, σ_0, n, and p.

If we want to choose n to ensure that the rejection probability is larger than some value (e.g. 0.80) for some $\kappa = \alpha_1 - \alpha_0$, we need to find n s.t.

$$\Phi\left(-1.96 - \frac{\kappa}{\sigma_\beta}\right) + \left[1 - \Phi\left(1.96 - \frac{\kappa}{\sigma_\beta}\right)\right] \geq 0.80$$

Population vs. Superpopulation

So far, we have viewed the experimental subjects $i = 1, \ldots, n$ as draws from a large population. The average treatment effect

$$\text{ATE} = E[Y_i(1) - Y_i(0)]$$

is the average effect in this larger population.

The randomness comes from drawing a sample of subjects.

If the population is infinitely large, or if the individuals are drawn "with replacment" from a finite population, then the subjects could be viewed as i.i.d. draws from the population.

If the population is finite and individuals are drawn *without* replacement, then $Y_i \not\perp\!\!\!\perp Y_j$ in general.

A different perspective is to consider the individuals in the study as the "population." So

$$Y_1(0), Y_1(1), Y_2(0), Y_2(1), \ldots, Y_n(0), Y_n(1)$$

are fixed numbers.

(The i's may be draws from some "superpopulation," but we focus on the treatment effect at the level of the actual set of subjects.)

The "finite-population" or in-sample ATE is:

$$ATE_n = \frac{1}{n} \sum_{i=1}^{n} [Y_i(1) - Y_i(0)] .$$

There is still randomness in the data (T_i, Y_i) because T_i is random.

So, for example $\hat{\beta}$ is random because which observations get treated and not treated are random, making $\overline{Y}_1$ and $\overline{Y}_0$ random.

The choice of targets can make a difference for inference:

$$\text{Var}\left(\hat{\beta} - ATE\right) \quad \neq \quad \text{Var}\left(\hat{\beta} - ATE_n\right).$$

Fisher's Exact Test

Suppose we want to test the *sharp null hypothesis* of no treatment effect

$$H_0 : Y_i(1) = Y_i(0) \quad \forall i.$$

This is stronger than $ATE = 0$ or $ATE_n = 0$. Under H_0,

$$Y_i = Y_i(1) = Y_i(0) \quad \forall i.$$

So we know all the missing potential outcomes if H_0 holds.

For a statistic like $\hat{\beta}$, we can calculate the value we would have obtained under any realization of the treatments $(T_1, \ldots, T_n)$.

If we also know the distribution of $(T_1, \ldots, T_n)$, we would be able to obtain the exact distribution of $\hat{\beta}$ under the randomization of treatment.

Fisher's Exact Test:

Let Ω denote the set of possible realizations of $(T_1, \ldots, T_n)$ under the treatment assignment mechanism.

For $\omega \in \Omega$, let $\hat{\beta}(\omega)$ denote the value we would have obtained for $\hat{\beta}$ if the treatments were realized as ω, under the sharp null hypothesis.

Can simulate the null distribution of $\hat{\beta}$ by drawing repeatedly for ω under the known treatment assignment mechanism.

Example

Suppose $n = 8$ and the treatment assignment mechanism assigns $n_1 = 4$ to treatment randomly. Then there are

$$\binom{8}{4} = 70 \text{ possible configurations of } (T_1, \ldots, T_8).$$

The randomization distribution of $\hat{\beta}$ is

$$\Pr(\hat{\beta} \leq z) = \frac{1}{70} \sum_{\omega \in \Omega} \mathbf{1}\left\{\hat{\beta}(\omega) \leq z\right\}.$$

We can find

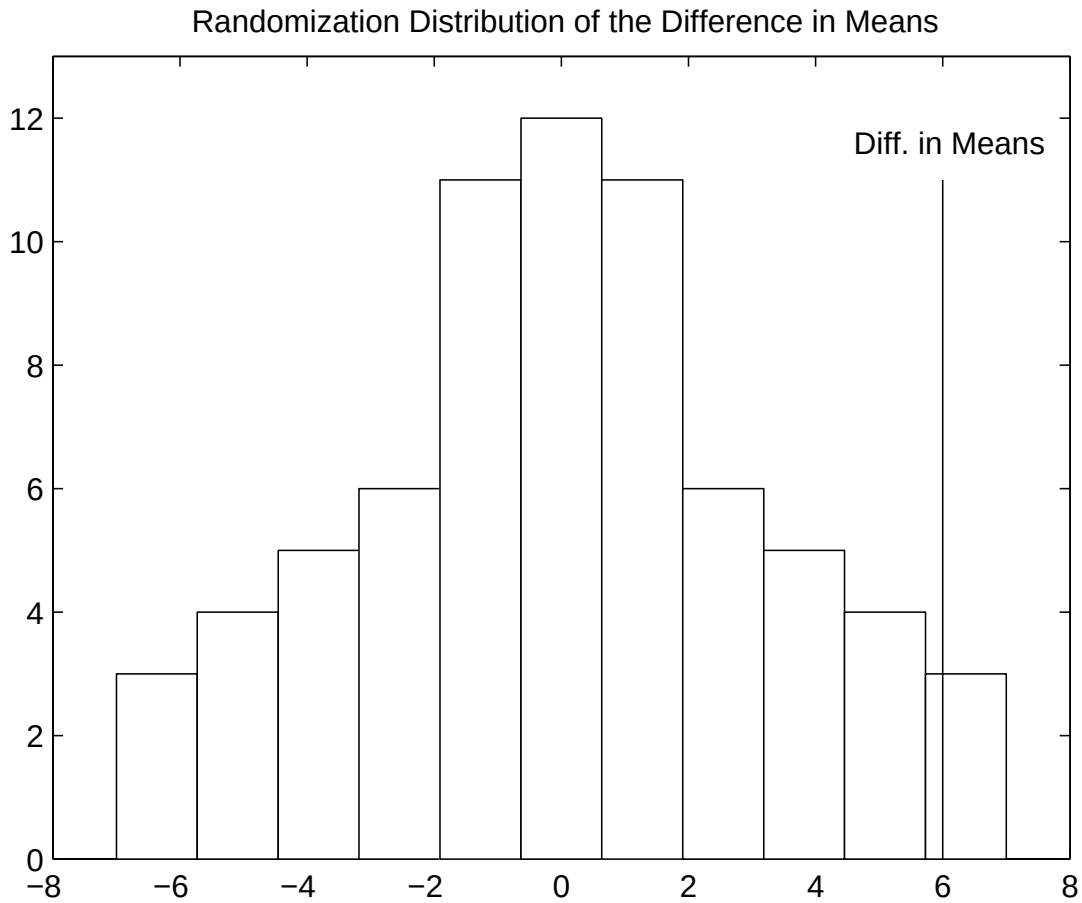
$$\bar{z} = \inf\left\{z : \Pr(|\hat{\beta}(\omega)| > z) \leq 0.05\right\},$$

and reject the sharp null if $|\hat{\beta}| > \bar{z}$.

Example cont'd

Y_i	12	4	6	10	6	0	1	1	
T_i	1	1	1	1	0	0	0	0	$\hat{\beta} = 6$
									$\hat{\beta}(\omega)$
$\omega = 1$	1	1	1	1	0	0	0	0	6
$\omega = 2$	1	1	1	0	1	0	0	0	4
$\omega = 3$	1	1	1	0	0	1	0	0	1
$\omega = 4$	1	1	1	0	0	0	1	0	1.5
$\vdots$				$\vdots$					$\vdots$
$\omega = 70$	0	0	0	0	1	1	1	1	-6

Example cont'd



$$\Pr(|\hat{\beta}(\omega)| \geq 6) = 0.0857.$$

Covariates in Randomized Experiments

$X_i \in \mathcal{X} \subset \mathcal{R}^k$

Assume X_i is *not* affected by treatment T.

If T_i is randomly assigned independent of X_i:

- simple difference-in-means $\overline{Y}_1 - \overline{Y}_0$ remains unbiased
- but could gain efficiency by using covariates in estimation/testing.

Could also use X in the treatment assignment mechanism: suppose

Unconfoundedness:

$$T_i \perp Y_i(0), Y_i(1) \mid X_i.$$

An equivalent condition is

$$\Pr(T_i = 1 \mid X_i = x, Y_i(0), Y_i(1)) = \Pr(T_i = 1 \mid X_i = x).$$

Also assume:

Overlap:

$$0 < \Pr(T_i = 1 \mid X_i = x) < 1 \quad \forall x,$$

Identification under Unconfoundedness

By unconfoundedness:

$$\begin{aligned} E[Y_i|\,T_i = 1, X_i = x] &= E[Y_i(1)|\,T_i = 1, X_i = x] \\ &= E[Y_i(1)|X_i = x]. \end{aligned}$$

Since $Pr(\,T_i = 1|X_i = x) > 0$ by the overlap assumption, we can consistently estimate $E[Y_i|\,T_i = 1, X_i = x]$.

Therefore we can identify $E[Y_i(1)|X_i = x]$.

By similar arguments, $E[Y_i|\,T_i = 0, X_i = x] = E[Y_i(0)|X_i = x]$.

Thus, can identify $ATE(x)$:

$$\begin{aligned} ATE(x) &= E\left[Y_i(1) - Y_i(0) \mid X_i = x\right] \\ &= E\left[Y_i \mid T_i = 1, X_i = x\right] - E\left[Y_i \mid T_i = 0, X_i = x\right] \end{aligned}$$

Then

$$ATE = E[ATE(X_i)] = \int ATE(x) dF_X(x),$$

where F_X is the (marginal) distribution of X_i.

So $ATE(x)$ and ATE are identified.

There are many different estimation methods in the literature, some tailored for specific situations (e.g. discrete X, high-dimensional X, parametric restrictions on $E[Y|T, X]$, etc.)

Propensity Score

Continue to assume unconfoundedness and overlap.

Propensity Score:

$$p(x) = \Pr(T_i = 1 \mid X_i = x) = E[T_i \mid X_i = x].$$

The propensity score characterizes the treatment assignment mechanism, and suggests alternative ways to estimate ATE.

Two useful results:

Result 1: (Rosenbaum and Rubin, 1983) Under unconfoundedness and overlap,

$$T_i \perp Y_i(0), Y_i(1) \mid X_i.$$

This implies that conditional on the *scalar* $p(X_i)$, treatment is unconfounded.

Can match on $p(X_i)$, or use $p(X_i)$ as a control variable in a regression-based estimation method.

Result 2: Under unconfoundedness and overlap,

$$E\left[\frac{T_i Y_i}{p(X_i)} - \frac{(1-T_i)Y_i}{1-p(X_i)}\right] = E\left[Y_i(1) - Y_i(0)\right] = ATE.$$

Then, for example, a sample analog-type estimator could be used

$$\hat{\tau} = \frac{1}{n}\sum_{i=1}^{n}\left\{\frac{T_i Y_i}{p(X_i)} - \frac{(1-T_i)Y_i}{1-p(X_i)}\right\}.$$

This estimator is not necessarily efficient, but efficient estimators have been devised and many variations of this idea are possible.