Quantitative Policy Evaluation
Randomized Experiments (I).

Eduardo Fé

Introduction.

Part 1. Randomized experiments in general.

- What are they
- What's the point?
- Are these use outside medicine?
- Randomized experiments elicit strong views...

Part 2. Analysis of randomized experiments: Neyman's repeated sample approach.

Part I

Randomized Experiments: General notes.

Randomized experiments.

Recall that, in accordance to our classification, a classical randomized experiment is an assignment mechanism with a known form which is controlled by the experimenter and which is

- individualistic
- probabilistic
- unconfounded

In other words, in a randomized experiment a unit can be allocated to the active treatment with non-0 probability and irrespectively of the status of other units or the potential benefits of the treatment to the units.

Randomized experiments.

Example 1. Bernoulli Trials.
In this type of assignment, a unit is allocated to treatment with probability[1] 1/2,

$$Pr(T_i|\mathbf{Y}(0), \mathbf{Y}(1), \mathbf{X}) = Pr(T_i) = 1/2 \text{ for all } i = 1, \ldots, N.$$

The disadvantage of Bernoulli Trials is that there is always a non-zero probability of extreme assignments (e.g. all units are allocated to the active treatment)

[1] Note that, strictly speaking, $Pr(T_i|\mathbf{Y}(0), \mathbf{Y}(1), \mathbf{X}) = Pr(T_i|Y_i(0), Y_i(1), \mathbf{x}_i)$ since randomized experiments are individualistic (where $\mathbf{x}_i$ is the vector of values of the covariates for unit i only)

Randomized experiments.

Example 2. Completely Randomized Experiments.
In this type of assignment, a N_t units are allocated to the active treatment and $N_c = N - N_t$ units are allocated to the control treatment with probability.

This design, however, does not take into account important covariates. For example, if the units include men and women, in a completely randomized experiment one can have all men allocated to the active treatment with positive probability.

Randomized experiments.

Example 3. Stratified Randomized Experiments.
In this type of assignment, units are partitioned into J blocks or strata (e.g. men/woman; north/south/east/west), and then we conduct a completely randomized experiment within each strata, allocating $N_t(j)$ units to treatment in strata $j = 1, \ldots, J$.

A extreme case of stratification is a **paired randomized experiment**, in which there are exactly two units in each strata.

Randomized experiments: Advantages.

The strength of randomized experiments, comes from their design: the active and control treatments are allocated at random, free of the influence of observable and unobservable characteristics of the units.

As a result, any variation in observable outcomes can be attributed to the treatment.

It also follows that we can estimate the effect of the active treatment using trivial statistical methods.

Furthermore, as we will see in a later section, many of the estimation techniques available for randomized experiments can also be successfully used in observational studies.

Randomized experiments: Examples

Discrimination in the Job Market in the United States, by Bertrand and Mullainathan (2004).

Researchers examined the level of racial discrimination in the United States labor market by randomly assigning identical CVs black-sounding or white-sounding names and observing the impact on requests for interviews from employers. Results found that CVs with white-sounding names received 50 percent more callbacks than those with black names,

Randomized experiments: Examples

Endogenous strategic thinking, Fé and Gill (2019): Do people care about others' intentions?
Two types of players: Allocators and receivers;
Allocators split a pie of 10 between themselves and the receivers.

- Treatment A. Either (keep 5, give 5) or (keep 8, give 2)
- Treatment B. Either (keep 5, give 5) or (keep 2, give 8)

Seeing the split, the receiver decides whether or not to give back 1 piece back to the allocator.

Randomized experiments: Examples

Endogenous strategic thinking, Fé and Gill (2019): Do people care about others' intentions?
In treatment A, the alternative is favourable to the allocator, and choosing an equal split is unambiguously generous; in the second treatment, the alternative is favourable to the receiver.

We find that theory-of-mind and age strongly predict whether receivers respond to the allocator's intentions in our gift-exchange game, while cognitive ability has no influence.

Part II

Samples, populations and superpopulations.

Statistical analyses of causality: Samples, Populations and Superpopulations.

An important subtlety of any causal inference: Which of this is true in a specific study?

1. **Population approach**: The finite sample is the set of units we are interested in; it does not matter how these data were selected and all conclusions are thus conditional on this sample (we don't attempt to extrapolate our conclusions to other populations).
2. **Super-population approach**: Units are drawn at random from a larger population (the **superpopulation**); the characteristic of the latter interest us. The units we observe are the **finite sample**.

Even though each of these approaches often rely on similar estimation methods, the statistical properties of estimators vary depending on which of the two scenarios are considered in practice.

Statistical analyses of causality: Samples, Populations and Superpopulations.

The super-population approach is dominant in the literature and virtually any textbook in statistics (on any topic) implicitly operates within that framework.

The population approach, however, will be particular important appropriate when discussing randomization inference later on this course.

Specific properties of the superpopulation approach

In the superpopulation approach,

- For convenience, it is assumed that the superpopulation is of size infinite.
- In the superpopulation, each unit has fixed potential outcomes, $Y_i(1), Y_i(0)$.
- The average $Y_i(1)$ in the superpopulation is $E\big[Y(1)\big]$; the average $Y_i(0)$ is $E\big[Y(0)\big]$.
- When we draw a finite sample of $i = 1...N$ observations, the units allocated to the active treatment $(T_i = 1)$ reveal $Y_i = Y_i(1)$; similarly, the units allocated to the control treatment $(T_i = 0)$ reveal $Y_i = Y_i(0)$;

Specific properties of the superpopulation approach

In the discussion that will follow, which takes a superpopulation approach, we will be interested in specific superpopulation quantities:

- $E(Y_i(t)|T_i = t)$ is the average potential outcome under treatment $t \in \{0, 1\}$ for those units that receive treatment t. These units, when sampled, reveal $Y_i(t)$, which implies $E(Y_i(t)|T_i = t) = E(Y_i|T_i = t)$.
- $E(Y_i|T_i = t)$, in turn, can be estimated as the sample mean of Y for units with $T_i = t$ -we say $E(Y_i|T_i = t)$ is **nonparametrically identified**
- $E(Y_i(t)|T_i \neq t)$ is the average potential outcome under treatment t of units that do NOT receive treatment t (they receive the alternative treatment). This is a **counterfactual** quantity.
- Counterfactuals are not nonparametrically identified. We will have to introduce **assumptions** on counterfactuals to be able to proceed with any causal analysis.

Part III

Randomized Experiments: Neyman's repeated sample approach (Superpopulation perspective).

Neyman's superpopulation approach.

1. SUTVA is assumed throughout, so that $Y_i(\mathbf{T}) = Y_i(T_i)$
2. We focus on completely randomized experiments. Specifically, note that

Neyman's superpopulation approach.

We are interested in the effect of a treatment T on an outcome Y.

- There is an infinite population of units.
- Each unit has two fixed potential outcomes $Y_i(1), Y_i(0)$. Under SUTVA, these potential outcomes do not depend on other unit's treatment status
- A finite number N of units are randomly drawn from the superpopulation. Each unit has potential outcomes $Y_i(T_i)$ for $T_i = 1$ if a unit is allocated to the active treatment (otherwise $T_i = 0$).
- Treatment is assigned in a completely randomized experiment[2]

[2] The results can be extended to stratified assignment mechanisms.

Neyman's superpopulation approach.

Recall that $Y_i(1) - Y_i(0)$ is the unobservable unit-level effect of the treatment.

In Neyman's causal framework, the focus is typically on the average effect of T on the superpopulation or the **Average Treatment Effect** in the superpopulation,

$$\tau_{ATE} = E\big[Y_i(1) - Y_i(0)\big] = E\big[Y_i(1)\big] - E\big[Y_i(0)\big] \tag{1}$$

That is, the average unit level treatment effect for the superpopulation.

It is implicit in the definition that an *ineffective* active treatment might have positive or negative effects for some units in the population; on average, however, the negative and positive effects cancel out for ineffective treatments.

Neyman's superpopulation approach: Critical result.

It can be shown[3] that for any $t \in \{0, 1\}$

$$E\big[Y_i(t)\big] = E\big[Y_i(t)|T = 1\big]P(T_i = t) + E\big[Y_i(t)|T \neq t\big]P(T_i \neq t) \quad (2)$$

[3] This is a consequence of the theorem of total probabilities

Neyman's superpopulation approach: Critical result.

It can be shown[3] that for any $t \in \{0, 1\}$

$$E\big[Y_i(t)\big] = E\big[Y_i(t)|T = 1\big]P(T_i = t) + E\big[Y_i(t)|T \neq t\big]P(T_i \neq t) \quad (2)$$

$E\big[Y_i(t)|T = t\big]$ is the average level of Y under treatment t in the superpopulation for those units that receive treatment t.

If a unit is allocated to treatment $T_i = t$, this unit reveals $Y_i(t)$ in the data, that is $Y_i = Y_i(t)$ in the data whenever $T_i = t$ and so, $E\big[Y_i(t)|T = t\big] = E\big[Y_i|T = t\big]$ and

$$E\big[Y_i(t)\big] = E\big[Y_i|T = t\big]P(T_i = t) + E\big[Y_i(t)|T \neq t\big]P(T_i \neq t) \quad (3)$$

[3] This is a consequence of the theorem of total probabilities

Neyman's superpopulation approach: Critical result.

The previous result matters, because $E\left[Y_i|T=t\right]$, $P(T_i=t)$ and $P(T_i \neq t)$ can be unbiasedly estimated from data:

$$\bar{Y}_1 = \frac{1}{N_t}\sum_{i=1}^{N} Y_i \cdot T_i, \text{ and } \bar{Y}_0 = \frac{1}{N-N_t}\sum_{i=1}^{N} Y_i \cdot (1-T_i)$$

$$\hat{P}(T_i=1) = \frac{1}{N_t}\sum_{i=1}^{N} T_i$$

$$\hat{P}(T_i=0) = \frac{1}{N-N_t}\sum_{i=1}^{N} (1-T_i)$$

this is useful, but not the critical aspect of this result...

Neyman's superpopulation approach: Critical result.

...the critical aspect is that in

$$E\big[Y_i(t)\big] = E\big[Y_i|T=t\big]P(T_i=t) + E\big[Y_i(t)|T\neq t\big]P(T_i\neq t)$$

$E\big[Y_i(t)|T\neq t\big]$ is the average level of Y under treatment t in the superpopulation for those units which did not receive treatment t.

There is no way we can estimate $E\big[Y_i(t)|T\neq 0\big]$ because if a unit i receives the treatment $(T\neq t)$ then it reveals $Y(\neq t)$, not $Y(t)$.

This is where the **selection problem** kicks in:

We cannot estimate $E\big[Y_i(t)\big]$ from data alone and, therefore, the $ATE = E\big[Y_i(1)\big] - E\big[Y_i(0)\big]$ cannot be estimated from data alone

Neyman's superpopulation approach: Critical result.

The assumption of **unconfoundedness**, however implies

$$E[Y_i(t)|T = s] = E[Y_i(t)] \text{ for any } t, s \in \{0, 1\}. \tag{4}$$

that is, in the (infinite) superpopulation, the average value of a potential outcome does not depend on the treatment to which units are assigned.

This allows us to write

$$E[Y_i(t)|T \neq t] = E[Y_i(t)|T = t]... \tag{5}$$

Neyman's superpopulation approach: Critical result.

...and so, under unconfoundedness:

$$\begin{aligned} E\big[Y_i(t)\big] &= E\big[Y_i(t)|T=t\big]\,P(T_i=t) + E\big[Y_i(t)|T=t\big]\,P(T_i\neq t) \\ &= E\big[Y_i(t)|T=t\big] = E\big[Y_i|T=t\big] \end{aligned} \tag{6}$$

In other words, under unconfoundedness we can estimate $E\big[Y_i(t)\big]$ with the estimators of $E\big[Y_i|T=t\big]$ (for $t \in \{0,1\}$), namely

$$\bar{Y}_1 = \frac{1}{N_t}\sum_{i=1}^{N} Y_i \cdot T_i, \text{ and } \bar{Y}_0 = \frac{1}{N-N_t}\sum_{i=1}^{N} Y_i \cdot (1-T_i)$$

Neyman's superpopulation approach: Inference.

Overall, under the assumptions of SUTVA, individualistic, probabilistic assignment and, critically, unconfounded assignment, we can estimate

$$\tau_{ATE} = E\big[Y_i(1) - Y_i(0)\big] = E\big[Y_i(1)\big] - E\big[Y_i(0)\big] \tag{7}$$

with

$$\begin{aligned}\hat{\tau}_{ATE} &= \bar{Y}_1 - \bar{Y}_0 \\ &= \frac{1}{N_t}\sum_{i=1}^{N} Y_i \cdot T_i - \frac{1}{N - N_t}\sum_{i=1}^{N} Y_i \cdot (1 - T_i)\end{aligned} \tag{8}$$

Neyman's superpopulation approach: Inference.

Under the superpopulation framework, the variance of τ_{ATE} equals

$$V(\hat{\tau}_{ATE}) = \frac{\sigma_1}{N_t} + \frac{\sigma_0}{N - N_t} \tag{9}$$

where

$$\sigma_t = V\left[Y_i(t)\right] = E\Big[Y_i(t) - E\left[Y_i(t)\right]\Big]^2 \tag{10}$$

$V(\hat{\tau}_{ATE})$ can be consistently estimated with,

$$\hat{V}(\hat{\tau}_{ATE}) = \frac{s_1}{N_t} + \frac{s_0}{N - N_t} \tag{11}$$

where s_1 and s_0 are the sample variances of Y_i calculated in the active treatment group ($T_i = 1$) and the control treatment group ($T_i = 0$) respectively.

Neyman's superpopulation approach: Inference.

For moderate/large samples, we can then construct a confidence interval for τ_{ATE},

$$\left(\hat{\tau}_{ATE} - z_{\alpha/2} \cdot \hat{V}(\hat{\tau}_{ATE})^{1/2},\ \hat{\tau}_{ATE} + z_{\alpha/2} \cdot \hat{V}(\hat{\tau}_{ATE})^{1/2}\right) \tag{12}$$

where $z_{\alpha/2}$ is the $\alpha/2^{th}$ quantile of the standard normal distribution. Finally, tests of hypothesis about τ_{ATE} can be based on a standard t-ratio. Specifically, a test of $H_0 : \tau_{ATE} = \tau^*$ can be based on:

$$\nu = \frac{\hat{\tau}_{ATE} - \tau^*}{\hat{V}(\hat{\tau}_{ATE})^{1/2}} \tag{13}$$

Under the null hypothesis, the distribution of ν follows a standard normal distribution as $N \to \infty$ and so a test of H_0 at $\alpha\%$ nominal level can rely on $|\nu| > z_{\alpha/2}$

Neyman's superpopulation approach: Pre-treatment information.

It is straightforward to extend the preceding framework to situation when researchers believe that the active treatment might have a differential effect across subsets of the population defined by specific values of a relatively small set of discrete **pre-treatment** variables.

Neyman's estimator can be calculated across the different subpopulations. A weighted average of these estimates provides an unbiased estimator of the overall treatment effect.

Neyman's superpopulation approach: Pre-treatment information.

When one suspects that the outcome under consideration might be driven by a large set of continuous and discrete variables, however, one might find a dearth of data within certain sub-categories or that only treated or untreated units have been observed for these sub-categories. In these settings one might want to consider a **regression analysis** of the treatment effect.

If the pre-treatment information is highly predictive of the levels of the outcome, then the regression approach might result in inferences that are more precise than those obtained with Neyman's estimator.

Neyman's superpopulation approach: Regression.

Let $\mathbf{X}_i = (X_{1,i}, X_{2,i}, \ldots, X_{K,i})'$ collect the observations of K **pre-treatment** variables for unit i. Let $E(\mathbf{X}_i) = \mu_X$.

We now, allow the potential outcomes to be a function of $\mathbf{X}_i$, so that $Y_i(t) = Y_i(t; \mathbf{X}_i)$ (SUTVA is implicit in this notation)

In the linear regression model, however, it is assumed that

$$Y_i(t; \mathbf{X}_i) = Y_i(t) + \mathbf{X}_i'\beta \tag{14}$$

Neyman's superpopulation approach: Regression.

With $Y_i(t; \mathbf{X}_i) = Y_i(t) + \mathbf{X}_i'\beta$, under unconfounded assignment and given the pre-treatment $\mathbf{X}_i$,

$$E\big[Y_i(1; \mathbf{X}_i) - Y_i(0; \mathbf{X}_i)\big] = E\big[Y_i(1) - Y_i(0)\big] = \tau_{ATE} \tag{15}$$

It can further be shown that, under linearity in $\mathbf{X}$, SUTVA and unconfoundedness,

$$\begin{aligned} E(Y_i|T_i, \mathbf{X}_i) &= E\left[Y_i(0)\right] + T_i \cdot \tau_{ATE} + \mathbf{X}_i'\beta \\ &= \alpha + T_i \cdot \tau_{ATE} + \mathbf{X}_i'\beta \end{aligned} \tag{16}$$

which is a standard regression model[4]

[4]The model can be estimated via OLS. Unlike with standard linear models, under unconfounded the OLS estimator of τ_{ATE} **has** a causal interpretation. All the other well known results from regression would apply verbatim. Note that we are NOT assuming that Y is normally distributed, nor continuous. Specifically, if Y is discrete we can still use OLS, but the standard errors need amending to obtain test of hypothesis with close to nominal size and confidence intervals with close to nominal coverage.

References I

Bertrand, M. and **Mullainathan, S.** (2004). Are Emily and Greg More Employable Than Lakisha and Jamal? A Field Experiment on Labor Market Discrimination. *American Economic Review*, 94(4): 991–1013

Fé, E. and **Gill, D.** (2019). Cognitive skills and development of strategic sophistication.