

Lec10: Causal Inference: Experiments

Isidoro Garcia Urqueta

2021

Agenda

Inferencia Causal

- ▶ Experimentación: Gold standard en inferencia causal

Diseño de experimentos

- ▶ Estratificación
- ▶ Nivel de aleatorización
- ▶ Pruebas de Poder
- ▶ Balance

Evaluación de impacto

- ▶ ITT
- ▶ Non compliance
- ▶ Efectos heterogéneos
- ▶ Aplicaciones

Introducción a la inferencia causal

Causalidad se refiere al cambio ocasionado por un evento d **Ceteris Paribus**. Esto es, como cambió mi métrica de interés y cuándo el evento d sucedió?

Usemos el modelo de potential outcomes de Rubin para formalizar la idea:

$$Y_i^{obs} = d_i * Y_i(d = 1) + (1 - d_i) * Y_i(d = 0)$$

Donde:

- ▶ Y es la métrica que nos interesa
- ▶ $d_i \in \{0, 1\}$ es el evento sobre el que nos interesa conocer su impacto.
- ▶ $Y_i(d = \{0, 1\})$ Es el valor de la métrica en cada escenario

Noten como, por construcción, **sólo observamos** un escenario de d . Esto genera el problema fundamental de la inferencia causal: **La falta de contrafactual**.

No observamos el contrafactual

Con el modelo de outcomes potenciales podemos definir el impacto del evento (tratamiento):

$$\tau_i \equiv Y_i(1) - Y_i(0)$$

El impacto del tratamiento es como estoy en el caso de que recibiera el tratamiento menos cómo **hubiera** estado en el escenario en que no hubiera recibido el tratamiento.

Noten como este es el impacto de tratamiento a nivel individual.

Típicamente intentamos estimar el **impacto promedio de tratamiento (ATE)**:

$$ATE \equiv E[Y_i(1)] - E[Y_i(0)] = E[\tau]$$

Seguimos ante el problema de como estimar el contrafactual

Estimador Naive

Supongamos que tenemos una población donde algunas personas fueron expuestas al tratamiento $d = 1$ y algunas $d = 0$. Uno se puede tentar a estimar ATE con:

$$E[\tau_{naive}] = E[Y_i(1)|d = 1] - E[Y_i(0)|d = 0]$$

Es decir, simplemente promediar el valor de los que recibieron el tratamiento y restarle el promedio de los que no recibieron el tratamiento.

El problema es que esto asume que $E[Y_i(1)] = E[Y_i(1)|d = 1]$ y que $E[Y_i(0)] = E[Y_i(0)|d = 0]$.

Es decir, el estimador naive ignora el **sesgo de selección**. La asignación a los grupos importa y mucho!

Correlación no es causalidad !



Sesgo de selección

El sesgo de selección significa que los grupos que recibieron el tratamiento y los que no (el grupo control) **son diferentes**. Esto hace que la comparativa entre los grupos sea injusta. No podemos atribuirle las diferencias al tratamiento porque los grupos ya eran distintos.

Hay dos tipos de sesgo de selección:

- ▶ **Selección en observables:** Los grupos son distintos en variables que observamos en los datos
- ▶ **Selección en no observables:** Los grupos de tratamiento son distintos en variables que no observamos.

Esto es $E[Y_i(0)] \neq E[Y_i(0)|d = 0]$ y $E[Y_i(1)] \neq E[Y_i(1)|d = 1]$.

- ▶ Noten como esto es $Y(0, 1) \not\perp d$. Es decir, la asignación de tratamiento no es ortogonal a los potential outcomes.

Por ende, **lo más importante en la inferencia causal es el mecanismo de asignación de tratamientos.**

Mecanismos de asignación

Estamos ante la búsqueda de una asignación de tratamientos que nos permita encontrar el impacto.

- ▶ Esto es, que nos deje usar al grupo que recibe el tratamiento como proxy de cómo se verían los individuos y al grupo que no lo recibió como el contrafactual.
- ▶ Matemáticamente esto es: $Y(0, 1) \perp d \rightarrow E[Y_i(0)] = E[Y_i(0)|d = 0]$

Formalmente, los mecanismos de asignación de tratamientos son una función que, con base en tus características, te asigna a un grupo de tratamiento(s). Esta debe cumplir con 3 requisitos:

1. **Asignación Individual:** El grupo al que pertenece un individuo no afecta al grupo/métricas a las que pertenece otro.
2. **Asignación Probabilística:** Cada individuo tiene una probabilidad positiva de pertenecer a cada grupo de tratamiento.
3. **Unconfoundedness:** La asignación genera que $Y(0, 1) \perp d$

Métodos en Inferencia Causal con base en la asignación

La inferencia causal se basa en intentar estimar el contrafactual con base en el mecanismo de asignación presente:

- 1. Asignación aleatoria (experimentación):** Esta es la única que cumple con 1,2 y 3. El investigador conoce la función de asignación
- 2. Asignación no aleatorio (métodos cuasi experimentales):** Buscamos, entendiendo la función de asignación que no controlamos, estimar el contrafactual.

En esta clase vamos a hablar de 1. Los métodos cuasiexperimentales los cubriremos en la siguiente clase.

Experimentos de Control Aleatorio (RCTs)

Los experimentos de control aleatorio son el resultado de una función de asignación de tratamientos:

1. **Asignación individual (SUTVA):** Stable Unit Treatment Value Assumptions (SUTVA). Esto significa no contaminación entre grupos (non-interference)
2. **Asignación Probabilística (Overlap):** Cada individuo tiene una probabilidad de pertenecer a los grupos de tratamiento conocida y $\in (0, 1)$
3. **Unconfoundedness (Excludability):** La asignación aleatoria es por definición independiente a las características de los usuarios.

Son el **Gold Standard** para estimar contrafactuales en medicina, criminología, negocios, etc.

Experimento: Conjuntos de grupos de tratamiento y probabilidades de pertenecer a cada uno.

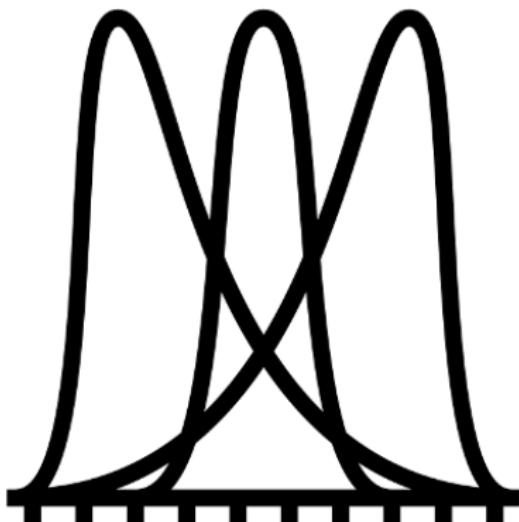
SUTVA

SUTVA puede resumirse como contaminación entre grupos o que las personas cambian su potential outcome dependiendo de a qué grupo pertenezcan.



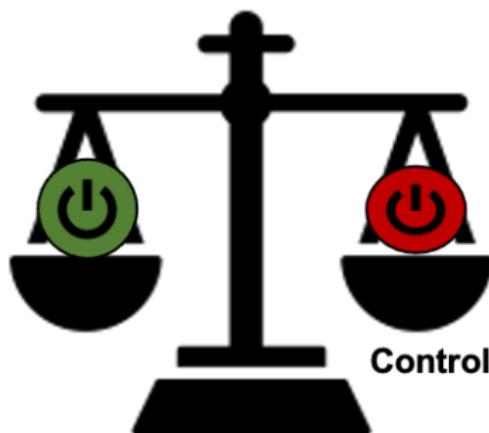
Overlap

La probabilidad de pertenecer a algún grupo siempre está entre cero y uno (sin incluirlos). De otra forma, tus características ($P(\text{treat}|x) = \{0, 1\}$) te seleccionaron para estar en algún grupo.



Unconfoundness

La asignación aleatoria es independiente a las características **observables y no observables** de los individuos. Esto es equivalente a grupos balanceados.



**GRUPOS
COMPARABLES**

Un poco de historia

La primera noción experimental nació por James Lind en una expedición naval británica. De 2,000 marineros que partieron hacia tierras españolas, 1,300 se contagieron de escorbuto.

Típicamente les daban malta, ácido sulfúrico con alcohol, limones y naranjas, entre otros. Nadie tenía idea de lo que realmente funcionaba.

Lind tomó un grupo de 12 enfermos y los dividió en 6 pares, uno para cada tratamiento:

1. Sidra.
2. Ácido sulfúrico con alcohol
3. Agua de mar
4. Mostaza con ajo y alfalfa
5. Vinagre
6. 2 naranjas y un limón al día

James Lind

En menos de una semana, Lind observó que los tratados con cítricos (treat 6) ya estaban lo suficientemente bien para incluso cuidar a los otros enfermos.

A partir de sus descubrimientos, publicó “Treatise of Scurvy”, donde documentaba que podían acabar con la enfermedad mortal muy común entre marineros.

A la marina inglesa le tomó 42 años!!!! hacer caso a los aprendizajes de su experimento y ordenar que cargaran de limones cada barco. Los experimentos siempre sufrieron push back!

Lady Tasting Tea

Por esta historia, Ronald Fisher inventó el concepto de pruebas de hipótesis/p-values. También nació la *prueba exacta de Fisher*.

Muriel Bristol (parlamentaria) se quejó cuando le trajeron su té negro con leche. Argumentó que le pusieron la leche antes del té y que eso era incorrecto. Los otros parlamentarios la retaron a probar que podía distinguir el orden de los ingredientes a probar.

Como diseñarían el experimento?

- ▶ Cuantas tazas?
- ▶ Cuantas de cada tipo?
- ▶ A partir de cuántas buenas declaras que Muriel podía identificar la diferencia?

Lady Tasting Tea

Fisher decidió diseñar el primer experimento así:

- ▶ 8 tazas de te (4 primero leche, 4 primero tea).
- ▶ Se le comunicó a Muriel Bristol el diseño.
- ▶ La hipótesis nula es que **ella no distingue la diferencia**
- ▶ Ella tenía que dividir las 8 tazas en dos grupos.

Fisher fijó la hipótesis: A Bristol tiene que irle mejor que la suerte. Esto se hace con $\binom{8}{4} = \frac{8!}{4!(8-4)!} = 70$ formas de ordenar 8 tazas en grupos de 4.

exito	selecciones_posibles	combinaciones
0	oooo	1x1=1
1	ooox, ooxo, oxoo, xooo	4x4=16
2	ooxx, oxox, oxxx, xoxo, xxoo, xoox	6x6=36
3	xxxx, xoxx, xxox, xxxx	4x4=16
4	xxxx	1x1=1

Lady tasting tea

La región crítica que definió Fisher fue:

- ▶ La probabilidad de que Bristol distinga las 4 tazas te-leche es $1/70 = 1.4\%$
- ▶ La probabilidad de que Bristol distinga 3+ de 4 tazas te-leche es $(16+1)/70 = 24\%$
- ▶ La probabilidad de que Bristol distinga 2+ de 4 tazas te-leche es $(1+16+36)/70 = 76\%$
- ▶ La probabilidad de que Bristol distinga 1+ de 4 tazas te-leche es $(1+16+36+16)/70 = 99\%$

Entonces tiene que elegir todas bien para que tengamos un p-value = 1.4% < 5% (la probabilidad de seleccionar todas by chance con 6 (3 y 3) tazas).

Diseño Experimental

Existen 4 tipos de diseños experimentales:

1. **Aleatorización simple:** Es como el que ven en microeconometría
2. **Asignación en bloques/estratificada:** Dentro de cada sub-grupos, realizas un mini experimento.
3. **Asignación por matching de pares:** Se ha probado que tiene problemas de poder serios en muchos contextos
4. **Diseño Factorial:** Se busca entender las interacciones de los tratamientos
5. **Diseños clusterizados:** El nivel de aleatorización cambia

En general, el lema es 'Control for what you can, Randomize what you cannot'

Aleatorización simple

Treatment

Individual 1

Individual 2

...

Individual n_T

Control

Individual 1

Individual 2

...

Individual n_C

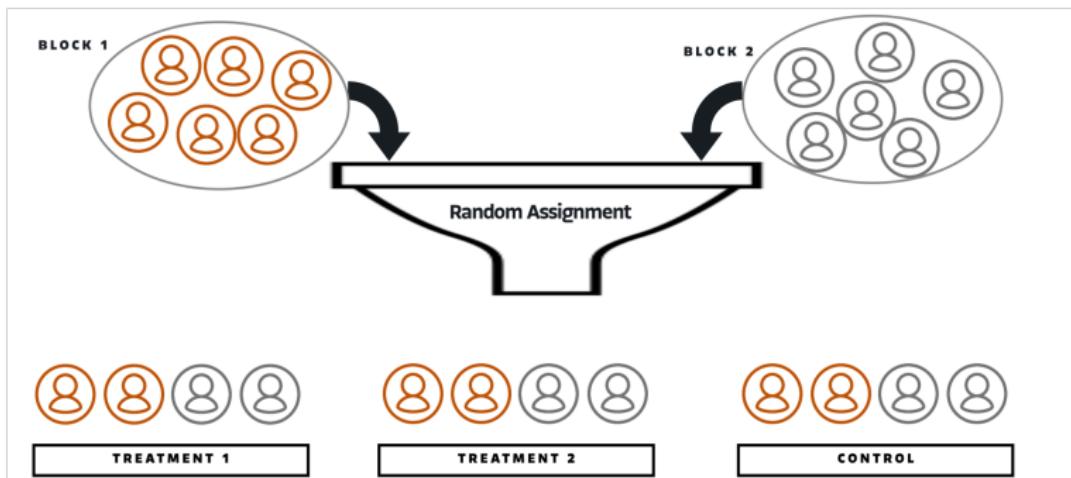
Qué puede salir mal con la aleatorización simple?



Asignación estratificada

	Block 1	...	Block m
Treatment 1	Individual 1 ...Individual n_1	...	Individual 1 ...Individual n_m
Treatment 2	Individual $n_1 + 1$...Individual $2n_1$...	Individual $n_m + 1$...Individual $2n_m$

Asignación estratificada



Con que variables estratificamos?

Debemos elegir las variables sobre las cuales estratificar con base en dos criterios:

1. Variables que creemos están muy **correlacionadas con el impacto de tratamiento τ**
2. Variables sobre las que tengamos interés en conocer efectos heterogéneos.

Mecánica de variables estratificadoras

La idea es construir bloques con **variables categóricas**. Con ellas se crean los estratos a partir de las combinaciones de cada categoría.

Ej: Estratifico por genero y si el paciente fuma o no.

Bloques:

1. Mujer, fuma
2. Mujer, no fuma
3. Hombre, fuma
4. Hombre, no fuma

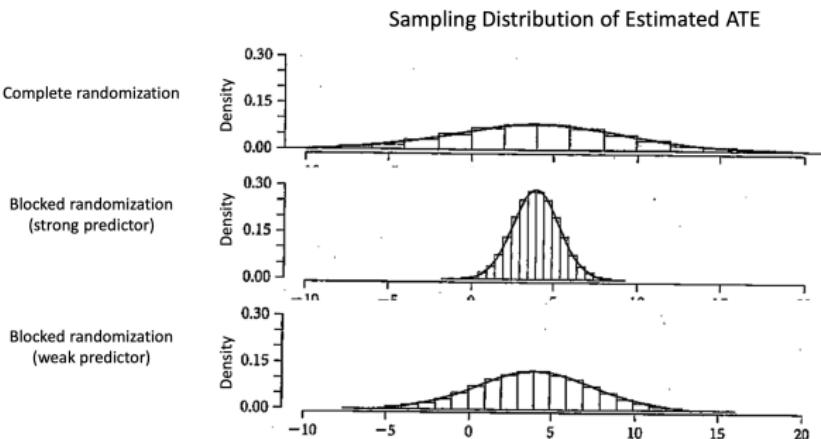
Posteriormente, aleatorizas dentro de cada bloque (con los mismos tratamientos y probabilidades).

Si las variables que nos interesa estratificar son continuas, debemos categorizarlas de alguna manera (i.e. deciles, cuartiles, etc).

Impacto de la estratificación sobre la medición de impacto

En general, si elegimos variables que están correlacionadas con el impacto de tratamiento, vamos a ganar mucha eficiencia en los estimadores (menos varianza)

Si elegimos variables no correlacionadas, no hay daño. Simplemente el estimador se parece más al estimador de un experimento de aleatorización simple



Cosas a considerar

Hasta donde estratificar: Si elegimos demasiadas variables, pulverizamos mucho las observaciones en cada estrato. Esto genera que se pierda el overlap

- ▶ Debemos estratificar por las variables más correlacionadas al impacto/que nos interese observar efectos heterogeneos **jerarquizando**, sujeto a un mínimo de observaciones por estrato (i.e. $n_{strata} > grupos$)

Misfits: Estas son observaciones que no son asignables porque:

1. El tamaño del estrato es menor a los grupos a asignar
 2. El residual de las divisiones por grupo (i.e. dividir 10 observaciones en 3 grupos (1/3,1/3,1/3)).
- ▶ Debemos escoger fracciones de tratamiento que sean, de preferencia, números racionales.
 - ▶ Posteriormente, debemos asignar los misfits de 3 maneras (global, strata, NA)

Misfits

Hay tres maneras de asignar los misfits:

1. Global: Juntar todos los misfits en un sólo grupo y asignar de manera aleatoria a cada grupo (siguiendo las mismas probabilidad de tratamiento). Esto genera que las probabilidades agregadas por grupo sean exactas.
2. Por Estrato: En cada estrato, asignar al grupo de misfits a algún grupo de tratamiento. Se favorece mejor el balance pero no sale la probabilidad por grupo exacta.
3. NA: Simplemente descartar a los misfits del experimento.

Diseño de Experimentos en R

treatment_assign(RCT)

R Documenter

treatment_assign() carries out robust treatment assignment by strata/blocks

Description

`treatment_assign()` carries out robust treatment assignment by strata/blocks

Usage

```
treatment_assign(
  data,
  share_control,
  n_t = 2,
  strata_varlist,
  misfits = c("global", "NA", "strata"),
  seed = 1990,
  share_t_i = rep(i/n_t - share_control/n_t, times = n_t),
  key
)
```

Arguments

<code>data</code>	A data frame, tibble or data.table
<code>share_control</code>	share of the observations assigned to control group
<code>n_t</code>	Number of treatments groups
<code>strata_varlist</code>	vector of categorical variables to form the strata/blocks for random assignment. Should be in the form of vars(var1, var2, ...)
<code>misfits</code>	How to handle the misfits. Default is "global". See Carril (2016) for details.
<code>seed</code>	A number used to set.seed().
<code>share_t_i</code>	The share of each treatment group. If NULL, (Default), each treatment group will have equal share.
<code>key</code>	The key identifier column of data.

This function creates a variable that indicates the treatment status. The random assignment is made by strata/blocks. It can handle equal or unequal treatment shares. Finally, it has three methods available to handle misfits (same as randtreat in STATA): "global": assigning the observations that couldn't be randomly assigned globally; "strata": assigning the observations that couldn't be randomly assigned by strata; "NA": set the treat observations that couldn't be randomly assigned to NA.

Value

A tibble "data" = the data with key, treat, strata, misfit columns; "summary_strata" = A summary tibble with the membership of each strata and its size.

Examples

```
data<-data.frame(key = c(1:1000),
  age_quartile = rep(c("Q1", "Q2", "Q3", "Q4"), each = 250),
  ing_quartile = rep(c("Q1", "Q2", "Q3", "Q4"), times = 250))
assignment<-treatment_assign(data = data, share_control = 0.5, n_t = 3,
  strata_varlist = c("age_quartile", "ing_quartile",
  "key"), misfits = "strata",
  seed = 1990, key = "key")
table(data$treat, sumdata = T)
prop.table(table(data$treat, sumdata = T))
```

Diseños factoriales

En este diseño se busca la interacción entre los distintos tratamientos. Noten como el diseño estratificado cabe en los diseños factoriales.

Msg / Price	No Price discount	Price discount
No Msg	Control: None	T1: Only Price
Msg	T2: Only Msg	T3: Price & Msg

Nivel de aleatorización: Diseño en cluster

Los diseños en cluster se crearon cuando es fácil violar SUTVA si asignamos a nivel individuo. Por ende, se asigna a nivel cluster para asegurar la no contaminación.

El ejemplo más común de este tipo de diseños son los experimentos escolares:

- ▶ Si aleatorizamos a nivel alumno, el alumno de al lado puede tener otro grupo de tratamiento (i.e. un propedeutico) vs su vecino.
- ▶ El vecino puede recibir parte del tratamiento si platican entre los alumnos.
- ▶ Por ende, se aleatoriza a nivel escuela para distintas escuelas.

Que implicaciones tiene esto sobre el unconfoundedness? (Matching Design!)

Pruebas de poder

Para las pruebas de poder tenemos dos dimensiones que considerar:

1. Dado el tamaño de la muestra (N), cuál es mi efecto mínimo detectable?
 - 1.1 Dado un efecto de tratamiento que nos interesa detectar, de que tamaño debe ser mi experimento?
2. Con que probabilidades asigno cada grupo de tratamiento?

Para 1, mostremos un poco de álgebra. El estadístico T se distribuye así:

$$T = \frac{\bar{Y}_1 - \bar{Y}_0 - \tau}{\sqrt{\sigma^2\left(\frac{1}{N_T} + \frac{1}{N_c}\right)}} \sim N(0, 1)$$

Poder I

La distribución es:

$$T = \mathcal{N}\left(\frac{\tau}{\sqrt{\sigma^2\left(\frac{1}{N_T} + \frac{1}{N_c}\right)}}, 1\right)$$

Ahora, la probabilidad de rechazar H_0 a un nivel de significancia α es:

$$\Pr(|T| > \Phi^{-1}(1 - \frac{\alpha}{2})) = \Phi(-\Phi^{-1}(1 - \frac{\alpha}{2}) + \frac{\tau}{\sqrt{\sigma^2\left(\frac{1}{N_T} + \frac{1}{N_c}\right)}})$$

Queremos que la probabilidad de rechazar la $H_0 : \tau = 0$ cuando es falsa sea igual a β (poder!).

$$\beta = \Phi(-\Phi^{-1}(1 - \frac{\alpha}{2}) + \frac{\tau}{\sqrt{\sigma^2\left(\frac{1}{N_T} + \frac{1}{N_c}\right)}})$$

Poder II

Aplicando Φ^{-1} en ambos lados:

$$\Phi^{-1}(\beta) = -\Phi^{-1}\left(1 - \frac{\alpha}{2}\right) + \frac{\tau}{\sqrt{\sigma^2\left(\frac{1}{N_T} + \frac{1}{N_c}\right)}}$$

Multiplicando por $\frac{\sqrt{N}}{\sqrt{N}}$

Finalmente, $sharecontrol = \frac{N_c}{N}$ y $1 - sharecontrol = \frac{N_T}{N}$:

$$\Phi^{-1}(\beta) = -\Phi^{-1}\left(1 - \frac{\alpha}{2}\right) + \frac{\tau\sqrt{N}}{\sqrt{\sigma^2\left(\frac{1}{1-sharecontrol} + \frac{1}{sharecontrol}\right)}}$$

$$\Phi^{-1}(\beta) = -\Phi^{-1}\left(1 - \frac{\alpha}{2}\right) + \frac{\tau\sqrt{N * sharecontrol(1 - sharecontrol)}}{\sigma}$$

Poder III

Efecto mínimo detectable:

$$\tau = \frac{[\Phi^{-1}(\beta) + \Phi^{-1}(1 - \frac{\alpha}{2})]\sigma}{\sqrt{N * sharecontrol(1 - sharecontrol)}}$$

Población mínima:

$$N = \frac{[\Phi^{-1}(\beta) + \Phi^{-1}(1 - \frac{\alpha}{2})]^2 \sigma^2}{\tau^2 * sharecontrol(1 - sharecontrol)}$$

Población mínima en R

N_min (RCT)

R Documentation

N_min() computes the minimum population needed to detect difference between control group and each treatment, given a target minimum detectable effect

Description

`N_min` computes the minimum population needed to detect difference between control group and each treatment, given a target minimum detectable effect

Usage

```
N_min(
  outcome_var,
  tau_min,
  power = 0.8,
  significance = 0.05,
  share_control,
  n_groups = 2
)
```

Arguments

outcome_var	the variable for which you wish to test the impact of treatment
tau_min	the target detectable effect (in outcome_var units)
power	The level of power of the test ($1 - \text{Pr}(\text{Reject } H_0 H_0 \text{ True})$). Default is 0.8
significance	The level of significance of the test $\text{Pr}(\text{Reject } H_0 H_0 \text{ False})$. Default is 0.05
share_control	The share of observations in <code>N</code> assigned to control. This argument allows for sequences (i.e. <code>seq(0,1)</code>)
n_groups	Number of groups (control + # treatment groups)

Details

This function calculates the minimum experiment's population needed in order to detect at least a difference of `tau_min` statistically significantly. This is between any two given groups (e.g. control vs each treatment), given the outcome variable, power and significance.

Value

A tibble with the `share_control` and `N` observations in control group (`N_control`), the share and `N` of each treatment (`share_treat`, `N_treat`), total share of treatment rows and `N` treated (`share_treat`, `N_treat`), `N`, the minimum detectable difference between control and all treatments together (`tau_min_global`), the minimum detectable difference between control and each treatment (`tau_min_each_treat`)

Examples

```
data <- data.frame(y_1 = rbinom(n = 100, size = 1, prob = 0.3),
                    y_2 = rbinom(n = 100, mean = 0, sd = 2))
N_min(data$y_1, tau_min = 0.3, share_control = seq(0.1, 0.1), n_groups = 2)
```

Efecto mínimo detectable en R

tau_min(RCT)

R Documentation

tau_min() computes the minimum detectable difference between control group and each treatment

Description
`tau_min()` computes the minimum detectable difference between control group and each treatment

Usage

```
tau_min(
  outcome_var,
  N,
  power = 0.8,
  significance = 0.05,
  share_control,
  n_groups = 2
)
```

Arguments

<code>outcome_var</code>	the variable for which you wish to test the impact of treatment
<code>N</code>	number of observations in the RCT, usually <code>nrow(data)</code>
<code>power</code>	The level of power of the test ($1 - \text{Pr}[\text{Reject } H_0 H_0 \text{ True}]$). Default is 0.8
<code>significance</code>	The level of significance of the test $\text{Pr}[\text{Reject } H_0 H_0 \text{ False}]$. Default is 0.05
<code>share_control</code>	The share of observations in <code>N</code> assigned to control. This argument allows for sequences (i.e. <code>seq(0,1,1)</code>)
<code>n_groups</code>	Number of groups (control + # treatment groups)

Details
This function calculates the minimum difference that could show significant $E[Y(1) - Y(0)] = \tau_{\text{min}}$, between any two given groups (e.g. control vs each treatment), given the population size (`N`), the outcome variable, power and significance.

Value
A tibble with the `share_control` and `N` observations in control group (`N_control`), the `share_treat` and `N` of each treatment (`share_treat`, `N_treat`), total share of treatment rows and `N_treated` (`share_treat`, `N_treated`), `N`, the minimum detectable difference between control and all treatments together (`tau_min_global`), the minimum detectable difference between control and each treatment (`tau_min_each_treat`)

Examples

```
data <- data.frame(y_1 = rbinom(n = 100, size = 1, prob = 0.3),
                    y_2 = rnorm(n = 100, mean = 0, sd = 2))
tau_min(data$y_1, N = nrow(data), share_control = seq(0,1,0.1), n_groups = 3)
```

Balance

El balance se realiza para evaluar (3) Unconfoundedness/Excludability. Al menos para las variables observables. Esto es:

$$Y(0, 1) \perp d \rightarrow E[Y_i(0)] = E[Y_i(0)|d = 0]$$

$$X(0, 1) \perp d \rightarrow E[X_i(0)] = E[X_i(0)|d = 0]$$

Existen dos pruebas que deben correr:

- ▶ Pruebas de balance individuales: Realizar pruebas T por variable entre grupos de tratamiento.

Cuando el número de columnas crece mucho, es muy fácil caer en falsos positivos (Bonferroni!)

- ▶ Pruebas conjuntas F: Correr un modelo de probabilidad lineal de pertenecer a cada tratamiento vs el grupo control.

Balance tests: F-test

Cuando el número de variables a evaluar son iguales, entramos en un problema de multiple hypothesis testing (i.e. Bonferroni adjustments).

La manera más sencilla de solucionar esto es hacer una prueba única conjunta:

$$H_0 : \beta_1 = \beta_2 = \dots \beta_k = 0$$

$$H_a : \beta_1 = \beta_2 = \dots \beta_k \neq 0$$

Balance tests: F-test II

Si tenemos i grupos de tratamiento, corremos lo siguiente:

$$treat_i = 0, 1, 2, 3, \dots, n$$

For $i = 1$ to $i = n$

$$\text{each}treat = treat = i > 0 \mid treat = 0$$

$$Pr(\text{each}treat) = X'\beta + \epsilon$$

Finalmente, revisamos si el modelo (con todas las variables) explica la probabilidad de pertenecer al tratamiento mejor que el modelo nulo (i.e. la media de probabilidad)

Balance en R

balance_table (RCT)

R Documentation

Creates balance table for the X variables across treatment status

Description

Creates balance table for the X variables across treatment status

Usage

```
balance_table(data, treatment)
```

Arguments

`data` A data.frame, tibble or data.table

`treatment` a string with treatment status column

Details

balance_table() performs t-test(X-treatment) for each X column in data. Every value of treatment 1,2,3,...N is compared against control value (0) or the first value of the treatment column. For instance, If treatment column has values of {0,1,2,3}, balance_table will return: the mean value of each treatment (or of X), and the p-values of the t-test of (1,2,3) against treatment = 0.

Value

A tibble with Mean_value of each treatment status and p_values

Examples

```
data <- data.frame(x1 = rnorm(n = 100, mean = 100, sd = 15),
                    x2=rnorm(n = 100, mean = 65),
                    treatment = rep(c(0,1,2,3,4), each = 20))
balance_table(data, "treatment")
```

Balance conjunto en R

balance_regression(RCT)

R Documentation

balance_regression() Runs a LPM of treatment status against all covariates (treatment~XB).

Description
`balance_regression()` Runs a LPM of treatment status against all covariates (treatment~XB).

Usage
`balance_regression(data, treatment)`

Arguments

- `data` A data.frame, tibble or data.table
- `treatment` a string with treatment status column

Details
This function runs a Linear Probability model of each treatment group & control on all the columns in data. For instance, if treatment column has values of (0,1,2), balance_regression will run two models: 1) LPM(treatment|0,1)-XB; and 2) LPM(treatment|0,2)-XB. The value are the regression tables and details of the F. test of these models.

Value
A list: `regression_tables` = regression output of treatment against all covariates, `F_Test` = table with the F tests of each regression

Examples

```
data <- data.frame(x1 = rnorm(n = 100, mean = 100, sd = 15), x2= rnorm(n = 100, mean = 65),
treat = rep(c(0,1,2,3,4), each = 20))
balance_regression(data = data, treatment = "treat")
```

Ejemplo final: Poder

Natural Transacting rate calculation:

Population shares: 50% New non-organic users, 30% Never active existing users, 20% Inactive existing users

$$\text{Natural transacting rate} = 50\% * 7.94\% + 30\% * 0\% + 20\% * 8\% = 2.394\%$$

tau_min	N_min_A/B	N_min_control_treat_i	N_min_control_treat_i_total	share_control	share_treat_i	share_treat_total	Treated units	Budget (MXN)
0.01	14,978	7,336	44,017	14%	14%	86%	37,729	5,659,317
0.02	3,744	1,834	11,004	14%	14%	86%	9,432	1,414,829
0.03	1,664	815	4,891	14%	14%	86%	4,192	628,813
0.04	936	459	2,751	14%	14%	86%	2,358	353,707
0.05	599	293	1,761	14%	14%	86%	1,509	226,373
0.06	416	204	1,223	14%	14%	86%	1,048	157,203
0.07	306	150	898	14%	14%	86%	770	115,496
0.08	234	115	688	14%	14%	86%	590	88,427
0.09	185	91	543	14%	14%	86%	466	69,868
0.1	150	73	440	14%	14%	86%	377	56,593

Estratos

strata	curp_verified	device_value	n_strata	n_missfits	Inactive: 13,947 users.						
					0	1	2	3	4	5	6
1001	0	[\$0-\$10,000]	3484	87	3431	10	9	9	9	8	8
1002	0	[\$10,001+]	5	5	4	0	0	0	0	1	0
1003	0	No info	2013	13	1982	5	5	5	5	5	6
1004	1	[\$0-\$10,000]	6155	155	6062	15	16	15	15	16	16
1005	1	[\$10,001+]	9	9	8	0	0	0	1	0	0
1006	1	No info	2418	18	2381	6	6	7	6	6	6
			14084	287	13868	36	36	36	36	36	36
				2.04							
				Treated units:	216						

Balance T

Inactive users (13,947)													
variables1	Media_control	Media_trat1	Media_trat2	Media_trat3	Media_trat4	Media_trat5	Media_trat6	p_value1	p_value2	p_value3	p_value4	p_value5	p_value6
age	33.275	32.667	35.806	32.000	38.056	33.972	31.750	0.761	0.294	0.457	0.035	0.752	0.318
balance_usd	7.416	6.225	4.469	8.920	8.050	8.427	6.018	0.653	0.155	0.639	0.775	0.684	0.509
curr_verified	0.609	0.583	0.611	0.611	0.611	0.611	0.611	0.757	0.983	0.983	0.983	0.983	0.983
gender_F	0.346	0.417	0.389	0.389	0.389	0.306	0.333	0.402	0.605	0.605	0.605	0.609	0.876
months_since_register	7.841	6.000	8.139	9.167	7.889	7.972	7.306	0.059	0.851	0.470	0.975	0.936	0.708
price	922.729	868.490	820.320	979.869	1271.406	1554.633	773.188	0.517	0.221	0.711	0.392	0.312	0.087

Balance F

estadistico	Msj1	Msj2	Msj3	Msj4	Msj5	Msj6
F-statistic	1.51	0.82	1.74	0.90	0.88	0.55
k	4	4	4	4	4	4
n-k-1	37	39	32	42	40	32
F_critical	0.17	0.17	0.17	0.18	0.17	0.17
p_value	0.22	0.52	0.17	0.47	0.49	0.70
R cuadrada	14.06%	7.74%	17.84%	7.87%	8.08%	6.39%

Medición de impacto

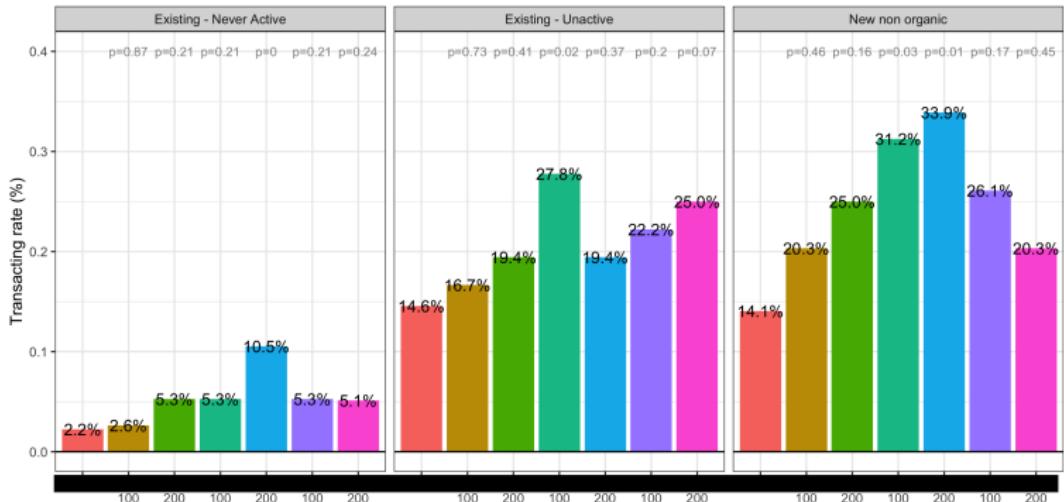
La medición de impacto tiene las siguientes directrices:

- ▶ Se puede hacer comparación de medias.
- ▶ SIEMPRE deben incluir en la medición el diseño. Esto es, incluir efectos fijos por estrato del diseño.

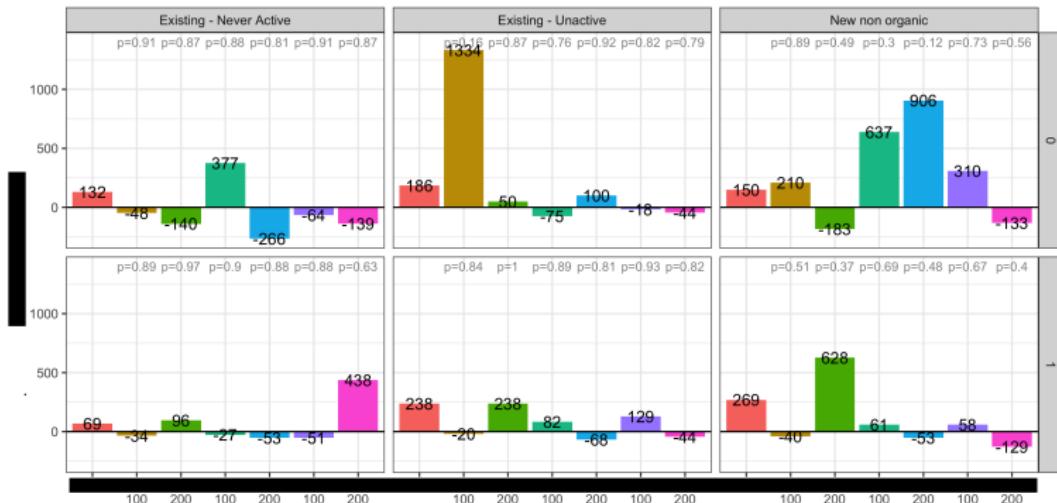
Esto genera que los estimadores pierdan varianza. Adicionalmente, genera que sea más sencillo correr una regresión para medir los impactos.

- ▶ Cuando la variable y métrica de interés es dicotómica, se debe correr un **Modelo de Probabilidad Lineal** y no un logit. La razón es no debemos imponer formas funcionales que generen sesgo. Más aún, es muy poco probable que nos encontremos con $P > 1$, $P < 0$.
- ▶ Podemos agregar controles en la medición. Esto no es necesario pero puede ayudar a la eficiencia. Se recomienda agregar variables que salieron desbalanceadas (para rebalancear).
- ▶ Finalmente, podemos usar Dif in Dif si el experimento corrió varios períodos y tenemos los datos pre-tratamiento.

Ejemplo: HTE (1 sola interacción)



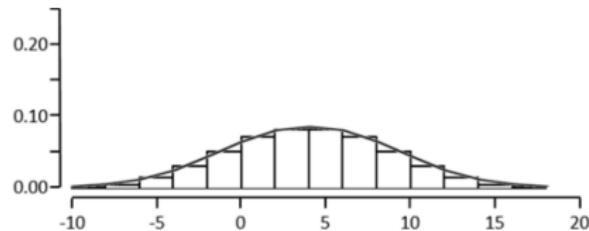
Ejemplo: HTE (doble interaccion)



Medición de impacto II: Dif in Dif

En general, el dif in dif no debería de generar cambios en el estimador de impacto (en principio). Su principal ganancia es la disminución de la varianza de los estimadores:

Difference-in-Means



Difference-in-Differences



Otros temas

Finalmente, nos faltó cubrir algunos otros temas importantes:

- ▶ Non-compliance: One sided y two sided. Se debe usar la asignación como variable instrumental del tratamiento efectivo.
- ▶ Medición de efectos heterogéneos: Siempre debemos reportar los impactos heterogéneos para las variables que utilizamos para crear bloques y NO otras (en principio). Esto refleja cierta honestidad intelectual del experimentador.