

Aleatorización

Antonella Bandiera

15 December 2021

Aleatorización

Tipos de aleatorización

Buenas prácticas

Aleatorización

Aleatorización en la asignación del tratamiento

- ▶ Los beneficios son conocidos y los discutimos en la clase pasada: balance entre grupos de tratamiento y control en características observadas y no observadas
- ▶ Todas las observaciones tienen una probabilidad de ser asignadas a tratamiento entre 0 y 1
 - ▶ Ninguna unidad se asigna a control o tratamiento con probabilidad 1
- ▶ Sin embargo, las unidades pueden variar en la probabilidad que tengan de ser asignadas a los grupos
 - ▶ Por ejemplo, la probabilidad puede variar por grupo: en un experimento las mujeres pueden tener una probabilidad del 75% de ser asignadas a tratamiento
- ▶ La probabilidad de ser asignados a tratamiento también puede variar entre unidades **siempre y cuando la probabilidad para cada unidad sea conocida**, aunque esto hace que el análisis sea más complicado

Asignación aleatoria \neq muestreo aleatorio

- ▶ Muestro aleatorio de la población: seleccionar a sujetos de una población a la muestra con una probabilidad conocida. No se puede tratar ni observar a toda la población por lo que se selecciona una muestra.
- ▶ Aleatorización del tratamiento: asignar sujetos de un grupo con una probabilidad conocida a tratamiento o control.
 - ▶ Es necesario poder medir outcomes y asignar tratamiento a todos los sujetos de la muestra
 - ▶ La asignación aleatoria se puede implementar con cualquier tipo de muestra: aleatoria, muestra de conveniencia, grupo de unidades, etc., o hasta con toda una población

La aleatorización está estrechamente conectada con el diseño y la pregunta que queremos responder

- ▶ Si queremos saber cuál es el efecto de una intervención sobre un gobierno municipal, no vamos a aleatorizar a nivel del individuo
- ▶ Si sospechamos que hay variables que pueden incidir fuertemente en el outcome porque esperamos que algunas unidades van a responder más al tratamiento, vamos a estratificar o bloquear por esas variables
- ▶ Si queremos saber cuál es el efecto individual de un tratamiento que tiene muchos componentes, hay que tener más de un tratamiento (diseño factorial o de varios brazos)
- ▶ Hay que definir y justificar qué vamos a hacer con el grupo de control
 - ▶ control puro
 - ▶ placebo

Nivel de tratamiento y outcomes

- ▶ El tratamiento puede ser asignado en diferentes niveles: individual, grupos, instituciones, comunidades, períodos de tiempo, etc.
- ▶ Pero a veces hay restricciones en cómo se puede *asignar* el tratamiento y *medir* los outcomes
 - ▶ Ejemplo: tratamiento a nivel de salón de escuela, pero outcomes a nivel de estudiante
 - ▶ Ejemplo: Tratamiento a nivel de distrito, pero outcomes a nivel de comunidad
- ▶ El nivel al que se asigna el tratamiento y se miden los outcomes determina qué pregunta vamos a poder responder o qué aspecto de una intervención vamos a poder evaluar

Tipos de aleatorización

Aleatorización simple

- ▶ Para cada unidad tirar una moneda (o algo análogo) para determinar si la unidad será tratada.
- ▶ No es necesario que sea balanceado (50-50), pero hay que saber la probabilidad de cada unidad.
- ▶ No se pueden garantizar el número exacto de unidades destinadas a control y tratamiento
- ▶ Ejemplo: Si sólo hay 6 unidades y tiramos una moneda balanceada, hay 3% de chances de asignar todas a tratamiento o todas a control

Ejemplo para aleatorización simple

```
# siempre hay que set.seed para poder replicar  
set.seed(12345)  
  
# elegir tamaño de la muestra  
N <- 200  
  
# Generar la asignación simple  
  
simple.ra <- rbinom(n = N,size = 1, prob = .5)  
  
# 112 personas terminaron el grupo de tratamiento  
sum(simple.ra)  
  
## [1] 112
```

Ejemplo para aleatorización simple II

```
# el paquete randomizr es conveniente  
#install.packages("randomizr")  
library(randomizr)  
  
# para poder replicar siempre set.seed, se puede elegir el  
set.seed(23456)  
# para randomización simple podemos usar la función simple_  
# Creamos un objeto que se llama treatment con N personas  
treatment <- simple_ra(N = N, # total sample size  
  prob = .5 # probability of receiving treatment  
)  
sum(treatment)
```

```
## [1] 96
```

Randomización completa

- ▶ Un número de unidades fijas m del total N son asignadas a tratamiento
- ▶ La probabilidad de que una unidad sea asignada a tratamiento es m/N
- ▶ Las asignaciones por lotería usan este método

Ejemplo código para randomización completa

```
# tamaño de la muestra
N <- 200
# número de unidades tratadas
m <- 100

# crear un vector m de 1s, y N-m de 0
complete.ra <- c(rep(1,m), rep(0,N-m))

# y luego mezclarlos usando sample()
# el default es sampling sin reemplazo

set.seed(12345) # para poder replicar
complete.ra <- sample(complete.ra)

sum(complete.ra)

## [1] 100
```

Ejemplo código para aleatorización completa II

```
# también podemos usar randomizr
library(randomizr)

set.seed(23456)

# asignación por aleatorización completa
treatment <- complete_ra(N = 200, # total sample size
                          m = 100) # number to assign to treatment

sum(treatment)
```

```
## [1] 100
```

Problemas: Muestra pequeña

- ▶ Restricciones de recursos o falta de unidades para incluir en el experimento pueden llevar a muestras pequeñas
- ▶ Si la muestra es muy pequeña, los grupos no estarán balanceados
 - ▶ Eso es problemático porque reduce nuestra capacidad para detectar efectos

Problemas: Muestra pequeña

- ▶ Restricciones de recursos o falta de unidades para incluir en el experimento pueden llevar a muestras pequeñas
- ▶ Si la muestra es muy pequeña, los grupos no estarán balanceados
 - ▶ Eso es problemático porque reduce nuestra capacidad para detectar efectos
- ▶ Una solución posible es estratificar o bloquear
 - ▶ Intuición: formar grupos de unidades similares y aleatorizar **dentro** del grupo
 - ▶ Ejemplo: Dividir entre municipalidades con y sin organizaciones vecinales y aleatorizar dentro de esos grupos

Aleatorización por estratificación o bloques I

- ▶ Creamos bloques de unidades y aleatorizamos dentro de cada bloque. Es como hacer mini-experimentos dentro de cada bloque.
 - ▶ Ejemplo: bloque = departamento, unidades = municipios. Aleatorizamos el tratamiento al nivel de los municipios **dentro** de los departamentos y medimos los outcomes al nivel de los municipios.
- ▶ Bloques que representan un grupo que es sustantivamente importante o distintivo nos pueden ayudar a entender cómo los efectos varían para los subgrupos.
 - ▶ Esto es especialmente útil cuando tenemos algunos grupos que son muy distintos: por ejemplo si un departamento y sus municipios son muy diferentes a los demás nos aseguramos que algunos van a estar en control y otros en tratamiento y que los comparamos entre sí

Aleatorización por estratificación o bloques II

- ▶ Cuando los bloques son homogéneos para cierto outcome aumenta mucho la precisión de las estimaciones comparado con un análisis que no tiene bloques.
 - ▶ Imagine que ciertas características de las unidades puede afectar cómo impacta el tratamiento
 - ▶ Por ejemplo, la presencia de una instancia participativa puede incrementar la efectividad de una intervención destinada a mejorar la provisión de un bien público
 - ▶ En estos casos, en los que estas variables pueden afectar el efecto del tratamiento sobre el outcome es deseable bloquear por estas características dado que aumenta la precisión de los efectos

Aleatorización por estratificación o bloques II

- ▶ La idea es elegir una (o un conjunto, más sobre esto luego) que describe a los grupos que se quiere balancear
- ▶ Hay que implementar la aleatorización de manera tal que exactamente una proporción p de las unidades sea tratada dentro de cada grupo
- ▶ Este experimento va a estar balanceado por definición
- ▶ Hay que elegir variables que sean *predictores* del outcome, si no la mejora en eficiencia será mínima
- ▶ Un caso especial de bloqueo es el caso de block paired matching, o bloques de tamaño 2
- ▶ El análisis de este tipo de experimentos **siempre** debe incluir efectos fijos por bloque

Cómo pensar en la aleatorización

- ▶ Kernan et al. (1999) resumen las ventajas de la estratificación
 - ▶ Balancear (bloquear, estratificar) sobre variables que estén correlacionadas con el outcome
 - ▶ Protegerse del error Tipo I (estamos reduciendo la probabilidad de desbalance por definición)
 - ▶ Estamos facilitando el análisis de subgrupos
 - ▶ Hay que protegerse del *drop-out* si tenemos parejas que conforman bloques, si una de ella se sale del estudio y no podemos medir el outcome, perdemos esa pareja y se achica el n
 - ▶ Aumentar el poder y la eficiencia reduciendo la varianza residual

Re-randomization

- ▶ A veces llamado “Big Stick”
- ▶ Escribir un loop que itera la randomización muchas veces y elige la mejor
- ▶ Testear el balance e iterar hasta que todos los p-values se vean bien,
- ▶ o hacer la randomización muchas veces y elegir la que tiene mejor balance

Aleatorización con blocking: Ejemplo

Pasemos a R

Aleatorización con clusters

- ▶ Un cluster es un grupo de unidades. En un estudio que tiene randomización con clusters, todas las unidades dentro de un cluster tienen el mismo tratamiento
- ▶ Hay que usar clusters cuando la intervención tiene que funcionar en ese nivel
 - ▶ Por ejemplo, si la intervención es sobre escuelas, hay que usar ese nivel de clustering aunque los outcomes los midamos al nivel de estudiante
- ▶ Problema: Tener pocos clusters es un problema serio, puede llevar a estimar de manera errónea los errores estándar. Qué tan grave es este problema depende de la correlación intra-cluster (a veces ICC o ρ)

Cluster randomization II

- ▶ Más correlación es peor
 - ▶ Cuando $\rho = 0$ entonces la escuela no importa para el comportamiento de los individuos
 - ▶ Cuando $\rho = 1$ todos los estudiantes de la escuela se comportan igual.
- ▶ Para el mismo número de unidades, tener **más clusters** con menos individuos puede ayudar.
- ▶ Trade off entre spillover y poder
- ▶ Se pueden tener clusters dentro de bloques, pero no bloques dentro de clusters

Buenas prácticas

Balance

- Chequear balance, con una regresión o usando un test, la función `xBalance` en el paquete `RIttools` es conveniente:

```
xBalance(treatment~x1+x2,data=dat,report='chisquare')
```

- Usar un F-test para una regresión con el tratamiento y variables independientes como outcome:

```
anova(lm(treatment~1,data = dat),  
      lm(treatment~x1+x2+x3,data=dat), test = 'F')
```

Balance

- ▶ La asignación aleatoria nos da balance, en promedio, en muchas variables pero no garantiza que la relación entre el tratamiento y todas las variables será cero. En experimentos con muestras muy pequeñas el desbalance puede ser grande.
- ▶ Se va a mirar la relación con muchas variables y, por azar, es posible que algunas sean significativas.