Análisis bajo asignación aleatoria

Diseño e implementación de experimentos en ciencias sociales

Departamento de Economía (UdelaR)

Análisis bajo asignación aleatoria simple

- *Inferencia de aleatorización (p-valores exactos para hipótesis nulas nítidas)*
  - FEDAI, capítulo 3
  - CISSBS capítulos 5 y 6, capítulo 9 (secciones 9.3 y 9.8), capítulo 10 (sección 10.3)
  - Duflo, E., Glennerster, R., & Kremer, M. (2007). Using randomization in development economics research: A toolkit. Handbook of development economics, 4, 3895-3962. (capítulos 4 y 7)
  - EGAP, 10 Things to Know About Randomization Inference
  - Peng Ding, Avi Feller, and Luke Miratrix (2016), "Randomization Inference for Treatment Effect Variation," Journal of the Royal Statistical Society, Series B 78: 655–671
  - Samii, C., & Aronow, P. M. (2012). On equivalencies between design-based and regression-based variance estimators for randomized experiments. Statistics & Probability Letters, 82(2), 365–370.
- *Regresión y ajuste por covariables*
  - FEDAI, secciones 4.1, 4.2, y 4.3.
  - Freedman, David A. 2008. "On Regression Adjustments in Experiments with Several Treatments." *Annals of Applied Statistics* 2(1): 176–96.
  - Lin, Winston. 2013. "Agnostic Notes on Regression Adjustments to Experimental Data: Reexamining Freedman's Critique." *The Annals of Applied Statistics* 7(1): 295–318.
  - Wager, Stefan, Wenfei Du, Jonathan Taylor, and Robert J. Tibshirani. 2016. "High-Dimensional Regression Adjustments in Randomized Experiments." *Proceedings of the National Academy of Sciences of the United States of America* 113(45): 12673–78.

Prueba de permutación

La señora que prueba el té (Fisher 1935. The Design of Experiments. Oliver and Boyd)

- ¿Tiene el té un sabor diferente según si el té se vierte en la leche o si leche se vierte en el té?

La señora que prueba el té (Fisher 1935. The Design of Experiments. Oliver and Boyd)

- ¿Tiene el té un sabor diferente según si el té se vierte en la leche o si leche se vierte en el té?
- Experimento:
  - Unidades: 8 tazas idénticas
  - Aleatorización: 4 tazas al azar en las que el té se sirve primero, otras 4 tazas primero la leche.
  - Hipótesis nula: la señora no puede notar la diferencia
  - Estadístico: número de tazas clasificadas correctamente

La señora que prueba el té (Fisher 1935. The Design of Experiments. Oliver and Boyd)

- ¿Tiene el té un sabor diferente según si el té se vierte en la leche o si leche se vierte en el té?
- Experimento:
  - Unidades: 8 tazas idénticas
  - Aleatorización: 4 tazas al azar en las que el té se sirve primero, otras 4 tazas primero la leche.
  - Hipótesis nula: la señora no puede notar la diferencia
  - Estadístico: número de tazas clasificadas correctamente
- Resultado: La señora clasificó correctamente las 8 tazas.

La señora que prueba el té (Fisher 1935. The Design of Experiments. Oliver and Boyd)

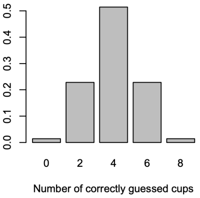
- ¿Tiene el té un sabor diferente según si el té se vierte en la leche o si leche se vierte en el té?
- Experimento:
  - Unidades: 8 tazas idénticas
  - Aleatorización: 4 tazas al azar en las que el té se sirve primero, otras 4 tazas primero la leche.
  - Hipótesis nula: la señora no puede notar la diferencia
  - Estadístico: número de tazas clasificadas correctamente
- Resultado: La señora clasificó correctamente las 8 tazas.
- ¿Ha ocurrido por casualidad?

Test de permutación

Taza	Conjetura	Asignación	Escenarios			
1	M	M	T	T	T	
2	T	T	T	T	M	
3	T	T	T	T	T	
4	M	M	T	M	T	...
5	M	M	M	M	M	
6	T	T	M	M	M	
7	T	T	M	T	M	
8	M	M	M	M	T	
# correctas		8	4	6	2	

Test de permutación

Taza	Conjetura	Asignación	Escenarios			
1	M	M	T	T	T	
2	T	T	T	T	M	
3	T	T	T	T	T	
4	M	M	T	M	T	...
5	M	M	M	M	M	
6	T	T	M	M	M	
7	T	T	M	T	M	
8	M	M	M	M	T	
# correctas		8	4	6	2	

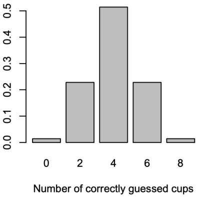


term	estimate	upper_p_value
Z	1	0.0142857

Test de permutación

Taza	Conjetura	Asignación	Escenarios			
1	M	M	T	T	T	
2	T	T	T	T	M	
3	T	T	T	T	T	
4	M	M	T	M	T	...
5	M	M	M	M	M	
6	T	T	M	M	M	
7	T	T	M	T	M	
8	M	M	M	M	T	
# correctas		8	4	6	2	

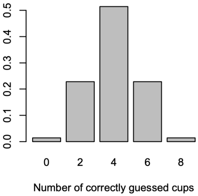
- $C(N,m) = \frac{N!}{m!(N-m)!} = \frac{8!}{4!4!} = 70$



term	estimate	upper_p_value
Z	1	0.0142857

Test de permutación

Taza	Conjetura	Asignación	Escenarios			
1	M	M	T	T	T	
2	T	T	T	T	M	
3	T	T	T	T	T	
4	M	M	T	M	T	...
5	M	M	M	M	M	
6	T	T	M	M	M	
7	T	T	M	T	M	
8	M	M	M	M	T	
# correctas		8	4	6	2	

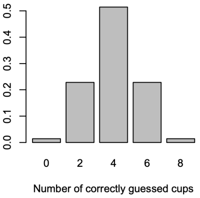


term	estimate	upper_p_value
Z	1	0.0142857

- $C(N, m) = \frac{N!}{m!(N-m)!} = \frac{8!}{4!4!} = 70$
- Bajo la hipótesis nula, la probabilidad de que la señora clasifique todas las tazas correctamente es $1/70 = 0{,}014$

Test de permutación

Taza	Conjetura	Asignación	Escenarios			
1	M	M	T	T	T	
2	T	T	T	T	M	
3	T	T	T	T	T	
4	M	M	T	M	T	...
5	M	M	M	M	M	
6	T	T	M	M	M	
7	T	T	M	T	M	
8	M	M	M	M	T	
# correctas		8	4	6	2	



term	estimate	upper_p_value
Z	1	0.0142857

- $C(N,m) = \frac{N!}{m!(N-m)!} = \frac{8!}{4!4!} = 70$
- Bajo la hipótesis nula, la probabilidad de que la señora clasifique todas las tazas correctamente es $1/70 = 0{,}014$
- Es posible que la señora posea la capacidad de diferenciar

Experimentos aleatorizados (RCTs)

- ¿Por qué aleatorizar la asignación del tratamiento en los experimentos?

Experimentos aleatorizados (RCTs)

- ¿Por qué aleatorizar la asignación del tratamiento en los experimentos?
  - Hace que los grupos de tratamiento y control sean “idénticos”, aparte del tratamiento

Experimentos aleatorizados (RCTs)

- ¿Por qué aleatorizar la asignación del tratamiento en los experimentos?
  - Hace que los grupos de tratamiento y control sean "idénticos", aparte del tratamiento
    - La distribución conjunta de cualquier covariable observada (factor de confusión) **X** y no observado **U** previo al tratamiento es idéntica entre los dos grupos:

$$P(\mathbf{X}, \mathbf{U} | T = 1) = P(\mathbf{X}, \mathbf{U} | T = 0)$$

,

Experimentos aleatorizados (RCTs)

- ¿Por qué aleatorizar la asignación del tratamiento en los experimentos?
  - Hace que los grupos de tratamiento y control sean "idénticos", aparte del tratamiento
    - La distribución conjunta de cualquier covariable observada (factor de confusión) **X** y no observado **U** previo al tratamiento es idéntica entre los dos grupos:

$$P(\mathbf{X}, \mathbf{U} | T = 1) = P(\mathbf{X}, \mathbf{U} | T = 0)$$

, donde **U** incluye los resultados potenciales $(Y(1), Y(0))$.

Experimentos aleatorizados (RCTs)

- ¿Por qué aleatorizar la asignación del tratamiento en los experimentos?
  - Hace que los grupos de tratamiento y control sean "idénticos", aparte del tratamiento
    - La distribución conjunta de cualquier covariable observada (factor de confusión) **X** y no observado **U** previo al tratamiento es idéntica entre los dos grupos:

      $$P(\mathbf{X}, \mathbf{U}|T = 1) = P(\mathbf{X}, \mathbf{U}|T = 0)$$

      , donde **U** incluye los resultados potenciales $(Y(1), Y(0))$.
    - Inconfundibilidad (Unconfoundedness) de la asignación al tratamiento:

      $$(X, U) \perp T, \text{y en particular}, Y(1), Y(0) \perp T$$

Experimentos aleatorizados (RCTs)

- ¿Por qué aleatorizar la asignación del tratamiento en los experimentos?
  - Hace que los grupos de tratamiento y control sean "idénticos", aparte del tratamiento
    - La distribución conjunta de cualquier covariable observada (factor de confusión) **X** y no observado **U** previo al tratamiento es idéntica entre los dos grupos:

      $$P(\mathbf{X}, \mathbf{U}|T = 1) = P(\mathbf{X}, \mathbf{U}|T = 0)$$

      , donde **U** incluye los resultados potenciales $(Y(1), Y(0))$.
    - Inconfundibilidad (Unconfoundedness) de la asignación al tratamiento:

      $$(X, U) \perp T, \text{y en particular}, Y(1), Y(0) \perp T$$

- Nos permite cuantificar formalmente el grado de incertidumbre

Inferencia por aleatorización vs a inferencia basada en modelos

- La aleatorización como "base de razonamiento para la inferencia" (Fisher)
- La aleatoriedad surge del acto físico de aleatorización, que puede utilizarse para realizar inferencias estadísticas
- También denominada *inferencia basada en diseño*
- Ventaja: el diseño justifica el análisis

Inferencia por aleatorización vs a inferencia basada en modelos

- La aleatorización como "base de razonamiento para la inferencia" (Fisher)
- La aleatoriedad surge del acto físico de aleatorización, que puede utilizarse para realizar inferencias estadísticas
- También denominada *inferencia basada en diseño*
- Ventaja: el diseño justifica el análisis

- Contrasta con la inferencia basada en modelos, que asume una distribución de resultados potenciales
- Ventaja de la inferencia basada en el modelo: flexibilidad

Test de hipótesis en la inferencia causal

- ¿Cómo aprender de un efecto contrafactual $(y_{Juan}, T = 1 \neq y_{Juan}, T = 0)$?
- Una solución es la **estimación de efectos causales promedios** (ATE, ITT, LATE).
- Esto es lo que llamamos el enfoque de Neyman.

Test de hipótesis en la inferencia causal

- Otra solución es hacer **afirmaciones** o **conjeturas** sobre los efectos causales.
- Podríamos decir: "*Creo que el efecto en Juan es 5*". o "*Este experimento no tuvo ningún efecto en nadie*". Y luego preguntamos "*¿Cuánta evidencia tiene este experimento sobre esa afirmación?*"
- Esta evidencia se resume en un $p - valor$.
- A esto lo llamamos enfoque de Fisher.

Ingredientes de una prueba de hipótesis

- Una **hipótesis** es una afirmación sobre una relación entre resultados potenciales.
- Una **estadístico** resume la relación entre el tratamiento y los resultados observados.
- El **diseño** nos permite vincular la hipótesis y el estadístico: calcular un estadístico que describa una relación entre resultados potenciales.
- El **diseño** también nos dice cómo generar una *distribución* de posibles estadísticos implícitos en la hipótesis.
- Un p-**valor** describe la relación entre los estadísticos observadas y la distribución de posibles estadísticos hipotéticos.

Inferencia por aleatorización

Inferencia por aleatorización

- Nos permite hacer inferencias exactas, sin uso de distribuciones.

Inferencia por aleatorización

- Nos permite hacer inferencias exactas, sin uso de distribuciones.
- Sin depender de supuestos de normalidad, etc.

Inferencia por aleatorización

- Nos permite hacer inferencias exactas, sin uso de distribuciones.
- Sin depender de supuestos de normalidad, etc.
- Sin depender de aproximaciones con muestras grandes.

Inferencia por aleatorización

- Nos permite hacer inferencias exactas, sin uso de distribuciones.
- Sin depender de supuestos de normalidad, etc.
- Sin depender de aproximaciones con muestras grandes.
- Verdaderamente no paramétrico, pero menos flexible.

(Sharp null) Hipótesis nula nítida de ausencia de efecto

- Hipótesis nula nítida:

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

(Sharp null) Hipótesis nula nítida de ausencia de efecto

- Hipótesis nula nítida:

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

- Qué sucede si el tratamiento no afectara a nadie en absoluto?

(Sharp null) Hipótesis nula nítida de ausencia de efecto

- Hipótesis nula nítida:

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

- Qué sucede si el tratamiento no afectara a nadie en absoluto?
- Implica que no hay efecto de tratamiento **medio**, pero no hay ATE (*sharp null*)

(Sharp null) Hipótesis nula nítida de ausencia de efecto

- Hipótesis nula nítida:

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

- Qué sucede si el tratamiento no afectara a nadie en absoluto?
- Implica que no hay efecto de tratamiento **medio**, pero no hay ATE (*sharp null*)
- Ejemplo sencillo con dos unidades: $\tau_1 = 1$ y $\tau_2 = -1$

(Sharp null) Hipótesis nula nítida de ausencia de efecto

- Hipótesis nula nítida:

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

- Qué sucede si el tratamiento no afectara a nadie en absoluto?
- Implica que no hay efecto de tratamiento **medio**, pero no hay ATE (*sharp null*)
- Ejemplo sencillo con dos unidades: $\tau_1 = 1$ y $\tau_2 = -1$
- Aquí, $\tau = 0$ pero se viola la hipótesis nula nítida.

(Sharp null) Hipótesis nula nítida de ausencia de efecto

- Hipótesis nula nítida:

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

- Qué sucede si el tratamiento no afectara a nadie en absoluto?
- Implica que no hay efecto de tratamiento **medio**, pero no hay ATE (*sharp null*)
- Ejemplo sencillo con dos unidades: $\tau_1 = 1$ y $\tau_2 = -1$
- Aquí, $\tau = 0$ pero se viola la hipótesis nula nítida.
- Si la *sharp null* es cierta, conocemos todos los resultados potenciales:

$$Y_i(1) = Y_i(0) = Yi$$

La vida bajo la Hipótesis nula nítida

Village	D_i	Y_i	$Y_i(0)$	$Y_i(1)$
1	1	0	?	0
2	1	0	?	0
3	0	1	1	?
4	1	0	?	0
5	1	1	?	1
6	0	1	1	?
7	0	0	0	?
8	1	1	?	1
9	0	1	1	?
10	0	0	0	?

La vida bajo la Hipótesis nula nítida

Village	D_i	Y_i	$Y_i(0)$	$Y_i(1)$
1	1	0	0	0
2	1	0	0	0
3	0	1	1	1
4	1	0	0	0
5	1	1	1	1
6	0	1	1	1
7	0	0	0	0
8	1	1	1	1
9	0	1	1	1
10	0	0	0	0

p-valor

- ¿Con qué frecuencia obtendríamos un estadístico de prueba tan grande o más si se cumpliera la hipótesis nula nítida?
- $\tau^{obs} = \tau(\mathbf{D}, \mathbf{Y}, \mathbf{X})$
- Ω es el conjunto de 2^N vectores de asignación

p-valor

- ¿Con qué frecuencia obtendríamos un estadístico de prueba tan grande o más si se cumpliera la hipótesis nula nítida?
- $\tau^{obs} = \tau(\mathbf{D}, \mathbf{Y}, \mathbf{X})$
- Ω es el conjunto de 2^N vectores de asignación
- **p-valor exacto**

$$Pr(\tau \geq \tau^{obs} | Y(1), Y(0), X, H_0) = \frac{1}{\Omega_0} \sum_{\mathbf{d} \in \Omega} (\tau(\mathbf{D}, \mathbf{Y}, \mathbf{X}) \geq \tau^{obs})$$

p-valor

- ¿Con qué frecuencia obtendríamos un estadístico de prueba tan grande o más si se cumpliera la hipótesis nula nítida?
- $\tau^{obs} = \tau(\mathbf{D}, \mathbf{Y}, \mathbf{X})$
- Ω es el conjunto de 2^N vectores de asignación
- **p-valor exacto**

$$Pr(\tau \geq \tau^{obs} | Y(1), Y(0), X, H_0) = \frac{1}{\Omega_0} \sum_{\mathbf{d} \in \Omega} (\tau(\mathbf{D}, \mathbf{Y}, \mathbf{X}) \geq \tau^{obs})$$

- Frecuencia con la que τ es mayor que el observado dividido por el número total de aleatorizaciones.

Inferencia por aleatorización, paso a paso

1. Determinar una hipótesis nula nítida un estadístico

Inferencia por aleatorización, paso a paso

1. Determinar una hipótesis nula nítida un estadístico
2. Calcular el estadístico observado: $\tau^{obs} = \tau(\mathbf{D_1}, \mathbf{Y}, \mathbf{X})$

Inferencia por aleatorización, paso a paso

1. Determinar una hipótesis nula nítida un estadístico
2. Calcular el estadístico observado: $\tau^{obs} = \tau(\mathbf{D_1}, \mathbf{Y}, \mathbf{X})$
3. Seleccionar aleatoriamente un vector de tratamiento diferente $\tilde{\mathbf{D}}$ tomado de $\mathbf{\Omega}$

Inferencia por aleatorización, paso a paso

1. Determinar una hipótesis nula nítida un estadístico
2. Calcular el estadístico observado: $\tau^{obs} = \tau(\mathbf{D_1}, \mathbf{Y}, \mathbf{X})$
3. Seleccionar aleatoriamente un vector de tratamiento diferente $\tilde{\mathbf{D}}$ tomado de $\boldsymbol{\Omega}$
4. Calcular $\tilde{\tau}_1^{obs} = \tau(\tilde{\mathbf{D}}_\mathbf{1}, \mathbf{Y}, \mathbf{X})$

Inferencia por aleatorización, paso a paso

1. Determinar una hipótesis nula nítida un estadístico
2. Calcular el estadístico observado: $\tau^{obs} = \tau(\mathbf{D_1}, \mathbf{Y}, \mathbf{X})$
3. Seleccionar aleatoriamente un vector de tratamiento diferente $\tilde{\mathbf{D}}$ tomado de $\boldsymbol{\Omega}$
4. Calcular $\tilde{\tau}_1^{obs} = \tau(\tilde{\mathbf{D}}_\mathbf{1}, \mathbf{Y}, \mathbf{X})$
5. Repetir pasos 3-4 para todo el espacio de aleatorización Ω_0 para obtener un vector $\tilde{\tau} = \{\tilde{\tau}_1, \ldots, \tilde{\tau}_k\}$

Inferencia por aleatorización, paso a paso

1. Determinar una hipótesis nula nítida un estadístico
2. Calcular el estadístico observado: $\tau^{obs} = \tau(\mathbf{D_1}, \mathbf{Y}, \mathbf{X})$
3. Seleccionar aleatoriamente un vector de tratamiento diferente $\tilde{\mathbf{D}}$ tomado de $\boldsymbol{\Omega}$
4. Calcular $\tilde{\tau}_1^{obs} = \tau(\tilde{\mathbf{D}}_\mathbf{1}, \mathbf{Y}, \mathbf{X})$
5. Repetir pasos 3-4 para todo el espacio de aleatorización Ω_0 para obtener un vector $\tilde{\tau} = \{\tilde{\tau}_1, \ldots, \tilde{\tau}_k\}$
6. Calcular el p-valor: $p = \frac{1}{k}\sum_{K=1}^{K}(\tilde{\tau}_K \geq \tau^{obs})$

Catadora de té

```r
plot(ri_out, p = "upper")
```