

# Módulo 1: Inferencia Causal y Resultados Potenciales

---

Santiago Olivella

July 16, 2017

Profesor Asistente de la Universidad de Carolina de Norte, Chapel Hill

# Hoja de ruta día 1

- Resultados potenciales e inferencia causal
- Efectos Promedio del Tratamiento (EPT, ATE en Ingles)
- Efecto Casual del Cumplimiento (ECC, CACE en Ingles)
- Efecto de Intención de Tratar (IT, ITT en Ingles)
- Modos Experimentales: RCT

# Un poco de notación

Para una unidad  $i$ ,

- Defina  $D_i$  como el **indicador de tratamiento**: en caso de tratamientos binarios,  $D_i \in \{0, 1\}$ . Asumimos que  $D_i$  es observable.
  - $D_i = 1$  si participante  $i$  es tratado;  $D_i = 0$  si no.
  - Asumimos que  $D_i$  es observable para todo  $i$ .
- Defina  $\mathbf{D}$  como el régimen asignado a todas las unidades.
- Defina  $\mathcal{D}$  como todas las formas posibles de asignar uno de los regímenes a todas las unidades.

# Un poco de notación

Para una unidad  $i$ ,

- Defina  $Y_i(d)$  como el **resultado potencial**: resultado para participante  $i$  escrito como una función del tratamiento.
- Para tratamientos binarios, consideren los siguientes dos “resultados potenciales”:
  - $Y_i(1)$ , el resultado potencial para  $i$  condicional en  $i$  sea tratado
  - $Y_i(0)$ , el resultado potencial para  $i$  condicional en  $i$  no sea tratado
- Suele concebirse como una función determinística del tratamiento.

# Resultados Potenciales Presupuesto Local

	Porcentaje del presupuesto si el jefe de la villa es hombre	Porcentaje del presupuesto si el jefe de la villa es mujer
Villa 1	10	15
Villa 2	15	15
Villa 3	20	30
Villa 4	20	15
Villa 5	10	20
Villa 6	15	15
Villa 7	15	30

# Un poco de notación

- Defina  $x_i$  como variables observadas pre-asignación, o **covariables observadas**:
  - Una cantidad que no depende de la asignación al régimen de de tratamiento; es decir, no tienen resultados potenciales.
  - No cualquier variable es una covariable; deben *co variar* con el resultado observado  $Y_i^{obs}$ .
- Defina  $u_i$  como una **covariable no observada**.

# Un poco de notación

- 
- Una **asignación aleatoria** consiste en elegir una las posibles asignaciones  $\mathcal{D}$  al azar, i.e. con igual probabilidad de selección.
  - Si las asignaciones han sido aleatorias, ¿depende la asignación de los resultados potenciales?
  - Y las covariables, ¿dependen de dichos resultados?
- En general, si las asignaciones han sido aleatorias, es el caso que  $D_i$  es independiente de  $Y_i(1)$ , de  $Y_i(0)$ , de  $\mathbf{x}_i$  y de  $u_i$ , de modo tal que es imposible predecir la asignación usando cualquiera de esta información.
  - Esto se conoce como **ignorabilidad**.

# Efecto Causal a Nivel Individual

- Para el participante  $i$ , el efecto del tratamiento es convencionalmente definido como la diferencia entre dos resultados potenciales:

$$\delta_i = Y_i(1) - Y_i(0)$$

- Alternativamente:

$$Y_i = Y_i(0) + \delta_i D_i$$

- Generalmente referido como el modelo causal de Rubin; pero es más apropiado referirse como el modelo de causalidad de Neyman-Holland-Rubin
- Dado que solamente uno de los dos resultados potenciales es observado,  $\delta_i$  es generalmente no obtenible sin recurrir a supuestos difíciles de defender.



# Algunos estimandos causales de interés

- *Efecto causal al nivel individual*

$$\delta_i = Y_i(1) - Y_i(0); \quad \text{pero también } \frac{Y_i(1)}{Y_i(0)}$$

- *Efecto promedio de tratamiento en la muestra finita (SATE)*

$$\begin{aligned}\delta_{fs} &= \frac{1}{N} \sum_{i=1}^N (Y_i(1) - Y_i(0)) \\ &= \mathbb{E}[Y_i(1) - Y_i(0)]\end{aligned}$$

- *Efecto promedio de tratamiento en la (super-)población (PATE)*

$$\delta_{sp} = \mathbb{E}[\mathbb{E}[Y_i(1) - Y_i(0)]]$$

donde la expectativa exterior se toma sobre la distribución

# Inferencia Causal

## Problema Fundamental de la Inferencia Causal

A fin de calcular el efecto causal de una intervención  $D_i$  sobre el resultado  $Y_i$ , es necesario evaluar con certeza la veracidad de un contrafáctico. Por definición, los contrafácticos no se pueden decidir empíricamente de forma definitiva. ¿Cómo podemos establecer entonces si una intervención tiene un efecto causal?

**Table 1:** Tabla 2.1, p35 Morgan y Winship, *Counterfactuals and Causal Inference*

Grupo	$Y_i(1)$	$Y_i(0)$
Tratamiento ( $D_i = 1$ )	<b>Observable</b>	Contrafactual
Tratamiento ( $D_i = 0$ )	Contrafactual	<b>Observable</b>

# Resultados Observables Presupuesto Local

	Porcentaje del presupuesto si el jefe de la villa es hombre	Porcentaje del presupuesto si el jefe de la villa es mujer	$D_i$
Villa 1	?	15	1
Villa 2	15	?	0
Villa 3	20	?	0
Villa 4	20	?	0
Villa 5	10	?	0
Villa 6	15	?	0
Villa 7	?	30	1

# Resultados Observables

- La conexión entre el resultado observable  $Y_i^{obs}$  y el valor subyacente del resultado potencial esta dada por

$$\begin{aligned}Y_i^{obs} &= D_i Y_i(1) + (1 - D_i) Y_i(0) \\ &= Y_i(0) + (Y_i(1) - Y_i(0)) D_i\end{aligned}$$

- $Y_i^{obs}$  es una variable aleatoria (es función de  $D_i$ ). Para cualquier participante, observamos ya sea  $Y_i(1)$  o  $Y_i(0)$ , pero no ambos (cuál observamos es aleatorio; sus valores no).
- Podemos reexpresar los estimandos causales en función de  $Y_i^{obs}$ . E.g.:

$$\delta_{fs} = \mathbb{E}[Y_i^{obs} | D_i = 1] - \mathbb{E}[Y_i^{obs} | D_i = 0]$$

- ¿Cómo usar evidencia empírica para evaluar los estimandos causales de interés?

# Estimación del efecto causal

- Consideren el siguiente estimador del SATE:

$$\hat{\delta}_{fs} = \bar{Y}_t^{obs} - \bar{Y}_c^{obs}$$

donde

$$\bar{Y}_t^{obs} = \frac{1}{N_t} \sum_{i:D_i=1} Y_i^{obs} \quad (\text{y de forma similar para } \bar{Y}_c^{obs})$$

donde  $N_t$  y  $N_c$  son los números de tratados y controles en la muestra, respectivamente.

- ¿Bajo qué condiciones es este un “buen” estimador?
  - ¿Inssegado?

# Estimación del efecto causal

$$\begin{aligned}\mathbb{E}[\hat{\delta}_{fp}] &= \mathbb{E} \left[ \frac{1}{N} \sum_{i=1}^N \left( \frac{D_i Y_i(1)}{N_t/N} - \frac{(1 - D_i) Y_i(0)}{N_c/N} \right) \right] \\&= \frac{1}{N} \sum_{i=1}^N \left( \frac{\mathbb{E}[D_i] Y_i(1)}{N_t/N} - \frac{\mathbb{E}[1 - D_i] Y_i(0)}{N_t/N} \right) \\&= \frac{1}{N} \sum_{i=1}^N (Y_i(1) - Y_i(0)) \\&= \delta_{fs} \square\end{aligned}$$

- ¿Qué asumimos acerca del mecanismo de asignación de tratamiento?
- ¿Importa la proporción de tratados?

# ¿Insesgado?

- Recuerden de la insesgadez es un propiedad solo del “promedio”
- “Tendemos a estimar la cantidad correcta, pero [aleatorización] apenas si resuelve el problema de estimar el efecto causal típico.” (Rubin 1976)
- ¿Por qué? ¿A qué es lo que Rubin se está refiriendo?  
¿Cómo puede “fallar” la aleotarización?

# Sesgos: atribución incorrecta de un efecto

- Cuando la asignación no es aleatoria, diferencias en resultados observados puede reflejar diferencias con respecto a algo diferente a la intervención.
  - Más específicamente, pueden reflejar diferencias sistemáticas en los grupos de participantes y de no participantes: heterogeneidad de covariables, o heterogeneidad en los resultados potenciales.
- Cuando las diferencias en resultados observados capturas diferencias tanto en asignación como en covariables y resultados potenciales, y estas fuentes no se pueden distinguir, decimos que hay un **sesgo** en el efecto causal estimado.
  - Heterogeneidad en covariables: sesgo de confusión (o espureidad)



# Estimación ingenua

$$\underbrace{E(Y_i|D_i = 1) - E(Y_i|D_i = 0)}_{\text{Obs diferencia en prom. } Y} = \underbrace{E(Y_i(1)|D_i = 1) - E(Y_i(0)|D_i = 1)}_{\text{ATT: Efecto prom. del tratamiento (tratados)}} + \underbrace{E(Y_i(0)|D_i = 1) - E(Y_i(0)|D_i = 0)}_{\text{Sesgo de selección}}$$

- La diferencia observable en los resultados promedios es generalmente un estimador sesgado del efecto promedio del tratamiento
- **n.b.**, el sesgo de selección es simplemente la diferencia en el resultado potencial de control entre los tratados y los no tratados.

# Resultado Potenciales Grupos

- Algunas veces es útil evaluar los resultados potenciales para un grupo de los participantes
- Expresiones en la forma  $Y_i(d)|X = x$  denotan los resultados potenciales cuando la condición  $X = x$  se mantiene
- Por ejemplo,  $Y_i(0)|D_i = 1$  se refiere a los resultados potenciales de control para un participante que de hecho recibe el tratamiento.

# Propiedades de la Asignación Aleatoria?

- Ante la igual probabilidad de asignación aleatoria, la esperanza condicional de SATE entre los tratados es la misma que la esperanza condicional de SATE del grupo de control, que por consiguiente es el mismo SATE
- El valor esperado de  $Y_i(0)$  en el grupo de tratamiento es el mismo valor esperado de  $Y_i(0)$  en el grupo control
- Cuando no utilizamos asignación aleatoria (Por ejemplo, investigación observacional), el insesgo de la diferencia entre las medias se convierte en una cuestión de conjetura.
  - Tanto en términos de balance como en términos de selección.

# Independencia condicional

- La aleatorización asegura la independencia de los resultados potenciales y el tratamiento (o de la asignación al tratamiento). Notación de independencia condicional:

$$(Y(1), Y(0)) \perp D$$

- Cuando esto se cumple:
  - $\mathbb{E}[Y_i(1)|D_i = 1] = \mathbb{E}[Y_i(1)|D_i = 0] = \mathbb{E}[Y_i(1)]$
  - Del mismo modo para  $\mathbb{E}[Y_i(0)]$
  - $(Y(1), Y(0)) \perp D$ , estimador de la intención de tratar (IT) puede ser escrito como

$$\begin{aligned}\mathbb{E}[Y_i(1)|D_i = 1] - \mathbb{E}[Y_i(0)|D_i = 1] \\&= \mathbb{E}[Y_i(1)|D_i = 1] - \mathbb{E}[Y_i(0)|D_i = 0] \\&= \mathbb{E}[Y_i(1)|D_i = 1] - \mathbb{E}[Y_i(0)|D_i = 1] \\&= \mathbb{E}[Y_i(1) - Y_i(0)] = \text{SATE}\end{aligned}$$

# Resultados Potenciales: Supuestos Básicos

- Cada resultado potencial depende **solamente** de si el **mismo** participante: No hay interferencia entre unidades experimentales.
- Que los resultados potenciales responden solamente al tratamiento y no a otro aspecto del experimento, tales como la asignación o la forma de medición: hay homogeneidad en el régimen de tratamiento
- Cuando no hay interferencia y existe homogeneidad del régimen, decimos que el contexto satisface el supuesto de estabilidad en la tripleta unidad-tratamiento-resultado, o **SUTVA** por sus siglas en inglés.

Unidad  $\times$  Régimen  $\rightarrow$  Resultado

# Restricción de Exclusión

- Sea  $Y_i(z, d)$  el resultado potencial cuando  $Z_i = z$  y  $D_i = d$  para  $z \in \{0, 1\}$  y para  $d \in \{0, 1\}$
- Por ejemplo, si  $Z_i = 1$  y  $D_i = 1$ , el participante es asignado al grupo de tratamiento y recibe el tratamiento
- O bien  $z_i = 1$  y  $d_i = 0$  - el participante es asignado al tratamiento pero no recibe el tratamiento
- La restricción de exclusión es que  $Y_i(1, d) = Y_i(0, d)$  - los participantes solo responden a la introducción de  $d_i$
- El supuesto de exclusividad no puede ser verificado empíricamente porque nunca observamos ambos resultados y para el mismo participante

# Ejemplo Clásico de Experimento con Fármacos

- El grupo de tratamiento recibe un nuevo fármaco
- El grupo de control no recibe nada
- El participante confunde el tratamiento con la recepción de una píldora
- Si los pacientes responden a la píldora en lugar de los ingredientes de la píldora, el supuesto de exclusión no se cumple
- Afecta la imparcialidad del estimador de diferencias entre medias

# No Interferencia

- Nos permite ignorar los resultados potenciales que surgieran si el participante  $i$  fuera afectado por el tratamiento de otro participante
- Formalmente, reducimos el listado de resultados potenciales  $Y_i(\mathbf{d})$ , donde  $\mathbf{d}$  describe todos los tratamientos administrados a todos los participantes, a un listado más simple  $Y_i(d)$ , donde  $d$  refiere al tratamiento asignado al participante  $i$ .
- Esto implica que siempre y cuando el estatus del tratamiento de un participante sea constante, el resultado de ese participante no se vería afectado en algún modo



# Violación del Supuesto de No Interferencia

- Las patrullas policiales desplazan el crimen de las áreas tratadas a las no tratadas
- la no interferencia no se cumple si estimación es lo siguiente:
  - El resultado promedio potencial de una cuadra es tratada es tratado, menos cuando otra cuadra no es tratado
- Si la policía desplaza el crimen desde las áreas tratadas a las no tratadas, los resultados observables en el grupo de control no serán resultados potenciales cuando el tratamiento no es administrado en ninguna de las cuadras
- La estimación del SATE tenderá a exagerar el verdadero valor de SATE

# Supuestos Básicos no Cumplidos?

- Asignación ciega: Los investigadores saben a que grupo los participantes han sido asignados al momento de analizar los datos
- Atrición: Algunos de los participantes del grupo de tratamiento se desaniman y abandonan el estudio
- Comportamiento compensatorio: Al notar que algunos de los participantes son excluidos del grupo de tratamiento de un programa de ayuda a la superación de la pobreza ya que han sido asignados al grupo de control, otra organización termina tratando a aquellos que han sido asignados al grupo de control

# Supuestos básicos no cumplidos?

- Múltiples resultados: Una intervención de pérdida de peso asigna al azar a los estudiantes que almuerzan en una cafetería donde el tratamiento que consiste en dar pequeños almuerzos y donde los resultados se miden en el total de calorías consumidas en la cafetería a la hora de almuerzo

# Chattopadhyay & Duflo 2004

- Experimento de campo en India
- 1990s, un tercio de los jefes de las villas son reservados para mujeres
- 'women.csv' en GitHub contiene un subgrupo de datos de West Bengal
- Gram Panchayat (GP) = Nivel de gobierno
- ¿Análisis?
  - ¿Fue la asignación aleatoria implementada apropiadamente?
  - Conjetura: más lugares para beber agua potable cuando mujeres son jefas de la villa
  - Conjetura: ningún efecto sobre en el riego (agricultura)

Nombre	Descripción
GP	Un identificador para Gram Panchayat (GP)
village	identificador para cada villa
reserved	variable binaria indicando si el GP fue reservado para mujeres o no
female	variable binaria indicando si el GP tuvo una jefa de consejo o no
irrigation	variable midiendo el numero de instalaciones nuevas o reparadas en la villa desde que la política fue implementada
water	numero de instalaciones de acceso a agua potable desde que la política fue implementada

**Tabla 4.6:** *Nombre de las variables y descripciones obtenida de la base de datos "Women as Policy Makers"*

```

>setwd("~/Dropbox/Experimental_Methodology/
  DPIR_2017/qss-master/PREDICTION")
>women <- read.csv("women.csv")
>##proportion of female politicians in
  reserved GP vs. unreserved GP
>mean(women$female[women$reserved] ==1)
[1] 1
>mean(women$female[women$reserved == 0])
[1] 0.07476636

```

```
>##drinking-water facilities  
>mean(women$water[women$reserved == 1]) -  
mean(women$water[women$reserved ==0])
```

```
[1] 9.25223
```

```
>##irrigation facilities  
>mean(women4irrigation[women$reserved == 1])  
-  
mean(women$irrigation[women$reserved == 0])
```

```
[1] -0.3693319
```

# Intención de Tratar

$$IT = \frac{1}{N} \sum_{i=1}^N (Y_i(Z = 1) - Y_i(Z = 0))$$

- Donde  $Z$  es un indicador para la asignación de tratamiento
- Con 100% de cumplimiento  $SATE=IT$
- $IT$  captura el efecto medio de ser asignado al grupo de tratamiento independientemente de la proporción del grupo de tratamiento que realmente fue tratado



# Efecto Causal de Cumplimiento (ECC)

$$ECC = \frac{IT}{\sigma}$$

- donde  $\sigma$  es el porcentaje de aquellos que fueron asignados al grupo de tratamiento y que recibieron el tratamiento
- ECC es también referido como a el Efecto Promedio de Tratamiento Local (EPTL) y Tratamiento de los tratados (TT)
- SATE entre los “Compliers”

## IT, EPT : Resultados Potenciales

Obs	$Y_i(0)$	$Y_i(1)$	$D_i(0)$	$D_i(1)$	Tipo
1	4	6	0	1	"Complier"
2	2	8	0	0	"Never-Taker"
3	1	5	0	1	"Complier"
4	5	7	0	1	"Complier"
5	6	10	0	1	"Complier"
6	2	10	0	0	"Never-Taker"
7	6	9	0	1	"Complier"
8	2	5	0	1	"Complier"
9	5	9	0	0	"Never-Taker"

# Comparación IT, EPT, y ECC

- SATE no considera incumplimiento:

$$\text{EPT} = \frac{2 + 6 + 4 + 2 + 4 + 8 + 3 + 3 + 4}{9} = 4$$

- IT toma en consideración que los “never-takers” no recibirán el tratamiento:

$$\text{ITT} = \frac{2 + 0 + 4 + 2 + 4 + 0 + 3 + 3 + 0}{9} = 2$$

- ECC es en base a un análisis de solamente aquellos que son “Compliers”:

$$\text{ECC} = \frac{2 + 4 + 2 + 4 + 3 + 3}{6} = 3$$

# Campaña Personaliza & Votos

- Gerber y Green Estudio en New Haven APSR 2000
- Asignación aleatoria a votantes de diferentes tipos de campañas del programa "Get Out The Vote"
  - Contacto campaña personaliza?
  - Por correo?
  - Teléfono?
  - Control?

# Movilización de Votantes en New Haven

Tasa de participación	G d tratamiento	G de control
Entre aquellos contactados	54.43 (395)	
Entre aquellos no contactados	36.48 (1050)	37.54 (5645)
Total	41.38 (1445)	37.45 (5645)

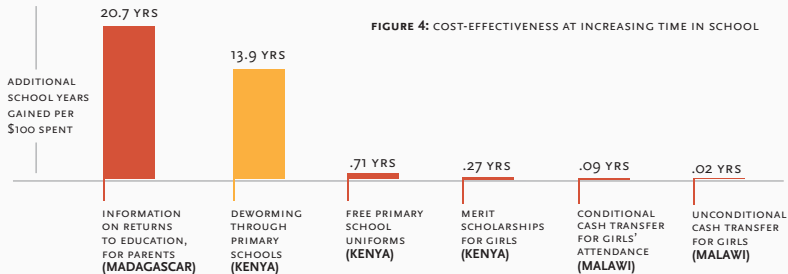
- $IT = 41.38 - 37.54 = 3.84$
- $\sigma = 395/1445 = 0.273$
- $ECC = IT/\sigma = 3.84/0.273 = 14.1$

# Ensayos Controlados Aleatorios (RCTs)

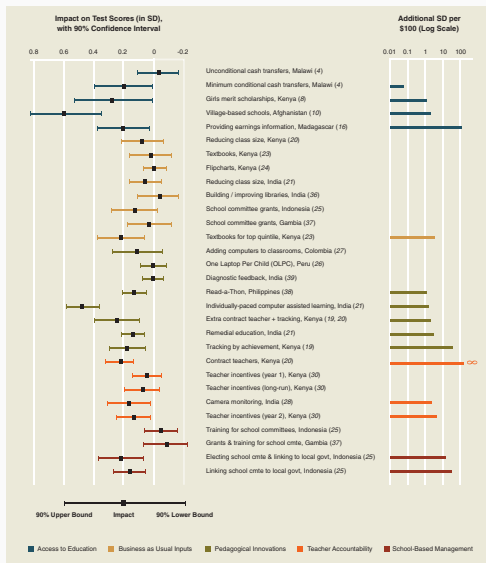
- RCTs Se llevan a cabo en un entorno real (=experimentos de campo)
  - Intentan de implementar el tratamiento de forma discreta
- Pueden minimizar cualquier artefacto del diseño experimental - validez interna
  - Por ejemplo. "experimenter demand effects" de los estudios de laboratorio
  - Permite hacer generalizaciones de los resultados y menos dependientes de supuestos

# Ensayos controlados aleatorios RCTs

- Cuatro criterios principales
  - Tratamiento = intervención en el mundo real?
  - Participantes = aquellos que normalmente lo recibirían?
  - Contexto = contexto real de interés?
  - Medida del resultado = resultado real del interés?







**Fig. 1. Impacts on test scores from randomized trials of primary school programs in the developing world.** Results drawn from 18 randomized studies that report test-score outcomes. The charts also rely on detailed cost data provided

by the authors of these studies and discussions with them about the most appropriate way to calculate cost effectiveness. The underlying data and calculations are available at [www.povertyactionlab.org/policy-lessons/education/student-learning](http://www.povertyactionlab.org/policy-lessons/education/student-learning).

# Prueba de Control Aleatorio

- Bertrand y Mullainathan (2004)
- ¿El prejuicio racial afecta la contratación de negros?
- ¿Cómo controlar factores tales como logros educativos, etc.?
- Los investigadores controlan la asignación aleatoria
- Asignación aleatoria de curriculum vitae a empresas contratando empleados

# B&M: Race in the Labor Market

- Por qué hay desigualdad racial en el mercado de trabajo de los Estados Unidos?
- ¿Es el resultado del tratamiento de contratación diferencial / discriminatorio?
- Los datos observacionales (encuestas) no pueden responder a esta pregunta debido a los no observables – Los datos de la encuesta no contienen todas las características que los empleadores observan
- Solución: Un experimento de campo
- Solución: Asignado al azar a un currículums ficticios
- Solución: curriculum vitae en papel elimina los "demand effects"

- Recolectaron curriculums vitae de sitios web; Los clasificaron en bajo y alta calidad; aumenta
- Identificaron nombres que son distintivamente asociados personas blancas y negras
- Cuatro currículos seleccionados para cada trabajo anunciado - Alto / Bajo a Blanco; Alto / bajo a negro
- 1,300 anuncios – 5,000 curriculums enviados
- Medida de respuesta: Empleador llamaba o contactaba por e-mail

```
>setwd("~/Dropbox/Experimental_Methodology/  
DPIR_2017/qss-master/CAUSALITY")
```

```
>read.csv("resume.csv")
```

	firstname	sex	race	call	
1	Allison		female	white	0
2	Kristen		female	white	0
3	Lakisha		female	black	0
4	Latonya		female	black	0
5	Carrie		female	white	0
6	Jay			male	white 0
7	Jill		female	white	0
8	Kenya		female	black	0
9	Latonya		female	black	0
10		Tyrone		male	black 0
11		Aisha		female	black 0
12		Allison		female	white 0

```

>setwd("~/Dropbox/Experimental_Methodology/DPIR_2017/qss-master/
  CAUSALITY")
>india <- read.csv("resume.csv")
>summary(india)

```

firstname	sex	race	call
Tamika : 256	female :3746	black:2435	Min. :0.00000
Anne : 242	male :1124	white:2435	1st Qu.:0.00000
Allison : 232			Median :0.00000
Latonya : 230			Mean :0.08049
Emily : 227			3rd Qu.:0.00000
Latoya : 226			Max. :1.00000
(Other) : 3457			

```
>## overall callback rate: total callbacks divided by  
  the sample size  
>sum(race.call.tab[,2])/nrow(resume)  
[1] 0.08049281  
>## callback rates for each race  
>race.call.tab[1,2] / sum(race.call.tab[1,]) #black  
[1] 0.06447639  
>race.call.tab[2,2]/sum(race.call.tab[2,]) #white  
[1] 0.09650924
```

## B&M: Resultados

- Currículums con nombres asociados a gente blanca tienen un 9.65 por ciento de chance de recibir una respuesta
- Currículums con nombres asociados a gente negra tienen un 6.45 por ciento de chance de recibir una respuesta
- 1 de 10 para blancos – 1 de 15 para negros
- Esto en términos de retorno implica 8 años adicionales de experiencia laboral
- Blancos fueron favorecidos en un 8.4 por ciento comparado con un 3.5 por ciento favoreciendo a negros



TABLE 1—MEAN CALLBACK RATES BY RACIAL SOUNDINGNESS OF NAMES

	Percent callback for White names	Percent callback for African-American names	Ratio	Percent difference ( <i>p</i> -value)
Sample:				
All sent resumes	9.65 [2,435]	6.45 [2,435]	1.50	3.20 (0.0000)
Chicago	8.06 [1,352]	5.40 [1,352]	1.49	2.66 (0.0057)
Boston	11.63 [1,083]	7.76 [1,083]	1.50	4.05 (0.0023)
Females	9.89 [1,860]	6.63 [1,886]	1.49	3.26 (0.0003)
Females in administrative jobs	10.46 [1,358]	6.55 [1,359]	1.60	3.91 (0.0003)
Females in sales jobs	8.37 [502]	6.83 [527]	1.22	1.54 (0.3523)
Males	8.87 [575]	5.83 [549]	1.52	3.04 (0.0513)

TABLE 3—RESUME CHARACTERISTICS: SUMMARY STATISTICS

Sample:	All resumes	White names	African-American	Higher quality	Lower quality
Characteristic:					
College degree (Y = 1)	0.72 (0.45)	0.72 (0.45)	0.72 (0.45)	0.72 (0.45)	0.71 (0.45)
Years of experience	7.84 (5.04)	7.86 (5.07)	7.83 (5.01)	8.29 (5.29)	7.39 (4.75)
Volunteering experience? (Y = 1)	0.41 (0.49)	0.41 (0.49)	0.41 (0.49)	0.79 (0.41)	0.03 (0.16)
Military experience? (Y = 1)	0.10 (0.30)	0.09 (0.29)	0.10 (0.30)	0.19 (0.39)	0.00 (0.06)
E-mail address? (Y = 1)	0.48 (0.50)	0.48 (0.50)	0.48 (0.50)	0.92 (0.27)	0.03 (0.17)
Employment holes? (Y = 1)	0.45 (0.50)	0.45 (0.50)	0.45 (0.50)	0.34 (0.47)	0.56 (0.50)
Work in school? (Y = 1)	0.56 (0.50)	0.56 (0.50)	0.56 (0.50)	0.72 (0.45)	0.40 (0.49)
Honors? (Y = 1)	0.05 (0.22)	0.05 (0.23)	0.05 (0.22)	0.07 (0.25)	0.03 (0.18)
Computer skills? (Y = 1)	0.82 (0.38)	0.81 (0.39)	0.83 (0.37)	0.91 (0.29)	0.73 (0.44)
Special skills? (Y = 1)	0.33 (0.47)	0.33 (0.47)	0.33 (0.47)	0.36 (0.48)	0.30 (0.46)
Fraction high school dropouts in applicant's zip code	0.19 (0.08)	0.19 (0.08)	0.19 (0.08)	0.19 (0.08)	0.18 (0.08)
Fraction college or more in applicant's zip code	0.21 (0.17)	0.21 (0.17)	0.21 (0.17)	0.21 (0.17)	0.22 (0.17)
Fraction Whites in applicant's zip code	0.54 (0.33)	0.55 (0.33)	0.54 (0.33)	0.53 (0.33)	0.55 (0.33)
Fraction African-Americans in applicant's zip code	0.31 (0.33)	0.31 (0.33)	0.31 (0.33)	0.32 (0.33)	0.31 (0.33)
Log(median per capital income) in applicant's zip code	9.55 (0.56)	9.55 (0.56)	9.55 (0.55)	9.54 (0.54)	9.56 (0.57)
Sample size	4,870	2,435	2,435	2,446	2,424

TABLE 4—AVERAGE CALLBACK RATES BY RACIAL SOUNDINGNESS OF NAMES AND RESUME QUALITY

Panel A: Subjective Measure of Quality (Percent Callback)				
	Low	High	Ratio	Difference ( <i>p</i> -value)
White names	8.50 [1,212]	10.79 [1,223]	1.27	2.29 (0.0557)
African-American names	6.19 [1,212]	6.70 [1,223]	1.08	0.51 (0.6084)
Panel B: Predicted Measure of Quality (Percent Callback)				
	Low	High	Ratio	Difference ( <i>p</i> -value)
White names	7.18 [822]	13.60 [816]	1.89	6.42 (0.0000)
African-American names	5.37 [819]	8.60 [814]	1.60	3.23 (0.0104)

# Marco de Resultados Potenciales

CV I	Nombres "Negros"	Respuesta			
	Nombre $T_i$	$Y_i(1)$	$Y_i(0)$	Edad	Educación
1	1	1	?	20	Universidad
2	0	?	0	55	Secundaria
3	0	?	1	40	Postgrado
.	.	.	.	.	.
.	.	.	.	.	.
.	.	.	.	.	.
n	1	0	?	62	Universidad

# Presion Social y Votación

- Presión social dentro un vecindario aumenta participación en votaciones?
- Asignación aleatoria de distintos mensajes GOVT
  - ¿Fueron las tarjetas postales las que aumentaron la participación de los votantes?
  - Mensajes GOVT tipo "Naming-and-shaming"?
  - Efecto Hawthorne

# Datos de Participación

```
>social <- read.csv("social.csv") #load the data
>summary(Social) #summarize the data
```

sex	yearofbirth	primary2004	messages
female	:152702	Min. :1900	Min. :0.0000
	:38218		Civic Duty
male	:153164	1st Qu. :1947	1st Qu. :0.0000
	:191243		Control
Median	:1956	Median :0.0000	Hawthorne :38204
Mean	:1956	Mean :0.4014	Neighbors :38201
3rd Qu.	:1965	3rd Qu. :1.0000	
Max	:1986	Max. :1.0000	

primary 2006	hhsizes
Min. :0.0000	Min. :1.000
1st Qu. :0.0000	1st Qu. :2.000
Median :0.0000	Median :2.000
3rd Qu. :1.0000	3rd Qu. :2.000
Max. :1.0000	Max. :8.0000

Nombre	Descripción
hhsz	Tamaño del hogar del votante
messages	Mensajes de GOTV recibidos por el votante (Civic, Control, Neighbours, Hawthorne)
sex	sexo del votante (female or male)
yearofbirth	Año de nacimiento del votante
primary2004	Si un votante participo en la elección primaria de 2004 (1=voted, 0=abstained)
primary2008	Si un votante participo en la elección primaria de 2008 (1=voted, 0=abstained)

**Table 2.3:** *Los Nombres y Descripciones de las Variables en los Datos del Experimento de Presión Social.*

# Efectos del Tratamiento

```
>## turnout for each group
>tapply(social$primary2006, social$messages, mean)
Civic Duty      Control      Hawthorne
Neighbors
0.3145377 0.2966383      0.3223746      0.3779482
```

```
>## subtract control group turnout for each group
>tapply(social$primary2006, social$messages, mean) -
  mean(social$primary2006[social$messages == "
    Control"])
Civic Duty      Control      Hawthorne
Neighbors
0.01789934      0.00000000      0.0.02573631
0.08130991
```



# Balance

```
>social$age <- 2006 - social$yearofbirth # create age
  variable
>tapply(social$age, social$messages, mean)
Civic Duty      Control      Hawthorne
Neighbors
49.65904 49.81355      49.70480      49.85294
>tapply(social$primary2004, social$messages, mean)
Civic Duty      Control      Hawthorne
Neighbors
0.3994453 0.4003388      0.4032300      0.4066647
>tapply(social$hhsz, social$messages, mean)
Civic Duty      Control      Hawthorne
Neighbors
2.189126 2.183667      2.180138      2.187770
```