

# Tarea 1: Experimentos aleatorizados

Rafael Sandoval F.

1 de marzo de 2021

1. ¿Por qué era importante para los autores aleatorizar los nombres? Es decir, ¿por qué los investigadores no recopilaban información de postulantes verdaderos a los trabajos y codificaron si los nombres de dichas aplicaciones están más asociados a afroamericanos o blancos? ¿Qué sesgo (positivo o negativo) crees que hubiera resultado de seguir esta estrategia?

Para poder minimizar o incluso neutralizar el sesgo por autoselección. Al aleatorizar los nombres y tener proporciones de características similares en los 2 grupos: de control (blancos) y de tratamiento (afroamericanos) se puede obtener el *Average Treatment Effect* (ATE) .

La ecuación que nos permite identificar el sesgo es la siguiente:

$$ATE = E(Call_i^B|Black) - E(Call_i^W|Black) + [E(Call_i^W|Black) - E(Call_i^W|White)] \quad (1)$$

El componente en negro es el *Treatment on the Treated* (TOT), mientras que el componente azul es el sesgo por autoselección.

De seguir la estrategia de recopilar información de postulantes verdaderos, y no por aleatorización, se podría tener un sesgo de autoselección positivo pensando en aquellos candidatos blancos podrían tener una mayor tasa de respuesta de los empleadores por diversos factores además de la raza. Un ejemplo (únicamente para ilustrar el sesgo) podría ser que los candidatos blancos tenían *college* o habilidades especiales, similar a cuando veíamos variables omitidas en el análisis de regresión múltiple. Entonces tendríamos que:

$$E(Call_i^W|Black) - E(Call_i^W|White) > 0$$

y esto nos llevaría a sobreestimar el efecto de ser blanco en la variable *call\_back*.

Para evitar este problema se recurre a la aleatorización del tratamiento y ahora

$$E(Call_i^W|Black) - E(Call_i^W|White) = 0$$

Entonces, ahora tendríamos que el TOT y ATE son iguales en valor esperado.

**2. Utiliza la base de datos para dar evidencia que la asignación de nombres parece haber sido aleatoria. Deberías incluir la(s) tabla(s) relevante(s) que te haya(n) permitido llegar a esta conclusión.**

Para detectar si la asignación de nombres fue hecha de manera aleatoria se puede crear una tabla de balance, similar a la tabla 3 del paper. Deberíamos observar proporciones de propiedades de CVs muy similares entre los grupos de interés: nombres blancos, nombres afroamericanos, alta calidad y baja calidad

**Tabla 1. Tabla de balance**

Variable.x	Mean_total	Mean_white	Mean_afram	Test
college	0.72	0.716	0.723	F=0.26
volunteer	0.411	0.409	0.414	F=0.166
military	0.097	0.092	0.102	F=1.239
email	0.479	0.479	0.48	F=0.003
empholes	0.448	0.45	0.446	F=0.083
workinschool	0.56	0.558	0.561	F=0.041
honors	0.053	0.054	0.051	F=0.201
computerskills	0.821	0.809	0.832	F=4.693**
specialskills	0.329	0.33	0.327	F=0.046
high	0.502	0.502	0.502	F=0
ofjobs	3.661	3.664	3.658	F=0.031
yearsexp	7.843	7.856	7.83	F=0.034
N	4870	2435	2435	

Con excepción de ofjobs y yearsexp, todas las variables son categóricas

En la tabla 1 se puede observar que las proporciones de todas las características dentro de ambos grupos (white y afroamericans) son muy similares. Además, la columna de Test prueba la hipótesis nula de que la diferencia entre las proporciones de cada grupo son iguales. Vemos que dicha Hipótesis solo se rechaza para computerskills, mientras que para todas las demás variables no se rechaza. (En esta tabla se usa una distribución Chi cuadrada para llevar a cabo la hipótesis. Únicamente para comprobar que se obtienen los mismos resultados asumiendo una distribución t de Student, a continuación se usa el comando `balance_table` de la librería RCT)

**Tabla 2. Tabla de balance**

Variable	White	Afram	Valor-p
college	0.7162	0.7228	0.6099
computerskills	0.8086	0.8324	0.0303
email	0.4789	0.4797	0.9543
empholes	0.4501	0.4460	0.7733
high	0.5023	0.5023	1.0000
honors	0.0542	0.0513	0.6538
military	0.0924	0.1018	0.2658
ofjobs	3.6645	3.6583	0.8601
specialskills	0.3302	0.3273	0.8310
volunteer	0.4086	0.4144	0.6836
workinschool	0.5581	0.5610	0.8399
yearsexp	7.8563	7.8296	0.8535

<sup>a</sup> Con excepción de ofjobs y yearsexp, todas las variables son categóricas

Efectivamente, se obtienen los mismos resultados, no se rechaza la hipótesis nula de igualdad de medias, salvo para la variable computerskills.

Esto nos permite pensar en que ambos grupos “son iguales” salvo por la raza (usando el nombre como proxy) y así podemos estimar el efecto insesgado (sin sesgo de autoselección) que buscamos: si tener un nombre afroamericano conlleva a tener menos probabilidad de recibir una llamada de empleo. De esta forma se puede encontrar evidencia de la discriminación racial. En otras palabras, podremos obtener el *Average Treatment Effect* insesgado.

**3. La variable *black* es una dummy creada por los investigadores para señalar si el nombre es usual de afroamericanos. Asumiendo que la distribución de nombres fue aleatoria, da evidencia de si existe discriminación racial en el call back utilizando: (i) un estimador de Neyman, (ii) una estimación de OLS con errores heterocedásticos, (iii) una estimación de OLS agregando controles (ustedes deberán decidir cuáles) y (iv) un probit sin controles**

Para el estimador de Neyman tenemos que  $\hat{\tau} = -0.03203$  y que  $\sigma(\hat{\tau}) = 0.00778$ . Al hacer la estimación mediante MCO obtenemos el mismo coeficiente con el mismo error estándar. Al agregar los controles *honors, specialkills, yearsexp, yearsexp*<sup>2</sup> se obtiene el mismo coeficiente para la variable *black*.

En la columna (4) de la tabla 3 se corrió un probit y lo que se reporta en la tabla es el efecto parcial promedio, ya que el coeficiente que se obtiene mediante el comando `lm` no es interpretable, mas que por su signo. Así, la interpretación del coeficiente de *black* es la misma en las 4 columnas: “Ser afroamericano conlleva una reducción de 3.20 puntos porcentuales en la probabilidad de obtener una llamada del empleador”. Primero me pareció un efecto muy pequeño, pero si consideramos que la probabilidad para una persona blanca de recibir una llamada es de 9.65% el efecto sí me parece grave, ya que es aproximadamente 1.5 veces más probable que contacten a un blanco que a un afroamericano.

La prueba de hipótesis relevante para cada una de las 4 columnas (en la primera columna no es el coeficiente de una regresión, sino el estimador de Neyman. Así mismo, en la columna (4) es el efecto marginal promedio) es:

$$H_0 : \beta_{black} = 0 \text{ vs } H_a : \beta_{black} \neq 0 \quad (2)$$

Entonces, al rechazar la hipótesis nula encontraríamos evidencia estadística para decir que hay discriminación racial. Como para cada columna tenemos que  $\beta_{black}$  es estadísticamente significativo, podemos rechazar  $H_0$ . Por lo tanto, en las 4 columnas obtenemos evidencia estadística para afirmar que, en efecto, existe discriminación racial en este caso.

Tabla 3

	Variable dependiente: call_back			
	Neyman	OLS		Probit
	(1)	(2)	(3)	(4)
black <sup>1,2</sup>	-0.032*** (0.008)	-0.032*** (0.008)	-0.032*** (0.008)	-0.032*** (0.008)
honors			0.070*** (0.023)	
specialskills			0.064*** (0.009)	
yearsexp			0.008*** (0.002)	
yearsexp <sup>2</sup>			-0.0002** (0.0001)	
Constant		0.097*** (0.006)	0.027** (0.012)	-1.302*** (0.035)
Observations	4,870	4,870	4,870	4,870
R <sup>2</sup>		0.003	0.024	
Adjusted R <sup>2</sup>		0.003	0.023	
Log Likelihood				-1,354.969
F Statistic		16.931*** (df = 1; 4868)	23.756*** (df = 5; 4864)	

<sup>1</sup>El coeficiente de *black* de la columna (1) corresponde a  $\hat{\tau}$  \*p<0.1; \*\*p<0.05; \*\*\*p<0.01

<sup>2</sup>El coeficiente y error estándar de *black* de la columna (4) corresponde al efecto marginal promedio

4. Imagina que encuentras una noticia que sugiere que la diferencia en el call\_back es de 1 punto porcentual a favor de los blancos. Establecemos dicha prueba de hipótesis como

$$H_0 : CB_{white} = CB_{afram} + 0.01 \quad (3)$$

Utiliza un *Fisher Exact Test* para evaluar esta hipótesis. Reporta el valor-p y la conclusión a la que llegas.

Se quiere probar la siguiente hipótesis:

$$H_0 : CB_{afram} - CB_{white} = -0.01 \text{ vs } H_a : CB_{afram} - CB_{white} \neq -0.01$$

Para responder a esta pregunta se hicieron los siguientes pasos con ayuda de un loop:

1. Se realizan  $J = 1000$  simulaciones de la asignación del tratamiento (2,435 personas reciben el tratamiento en cada simulación)
2. Se calcula un estadístico

$$W^j(Y, T) = \left( \frac{1}{N_T} \sum_{i|T_i^j=1} Y_i^T \right) - \left( \frac{1}{N_C} \sum_{i|T_i^j=0} Y_i^C \right) + 0.01$$

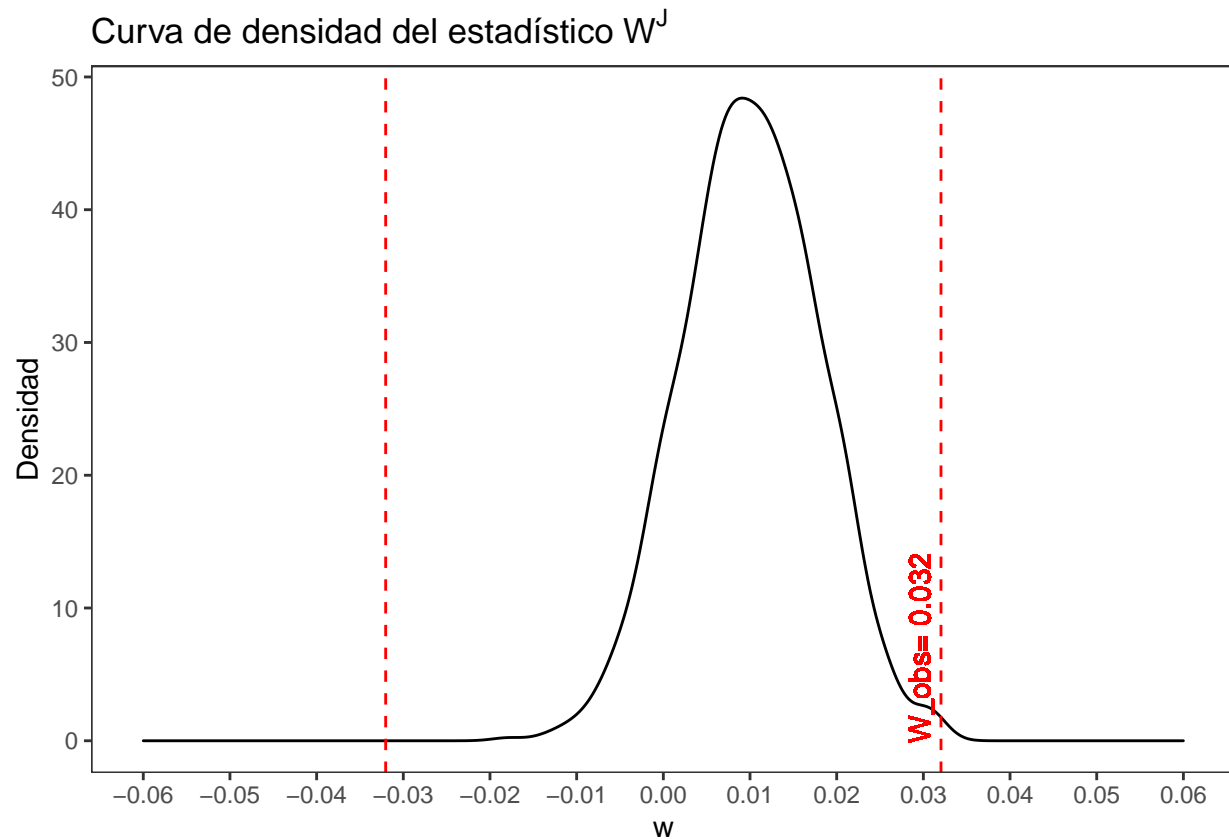
3. Se cuentan las simulaciones que cumplen:

$$pvalue = \sum_{j=1}^{1000} I\{|W^j(Y, T)| \geq |W^{obs}(Y, T)|\}$$

,donde  $W^{obs} = -0.03203285$  y se divide entre  $J = 1000$  para obtener el  $p - value$ .

Se obtuvo un  $p - value$  igual a 0.001. Entonces, con un nivel de significancia del 5% sí hay evidencia estadística para rechazar la hipótesis nula  $H_0 : CB_{afram} = CB_{white} - 0.01$

A continuación se muestra la distribución del estadístico  $W^j$  bajo la Hipótesis nula:



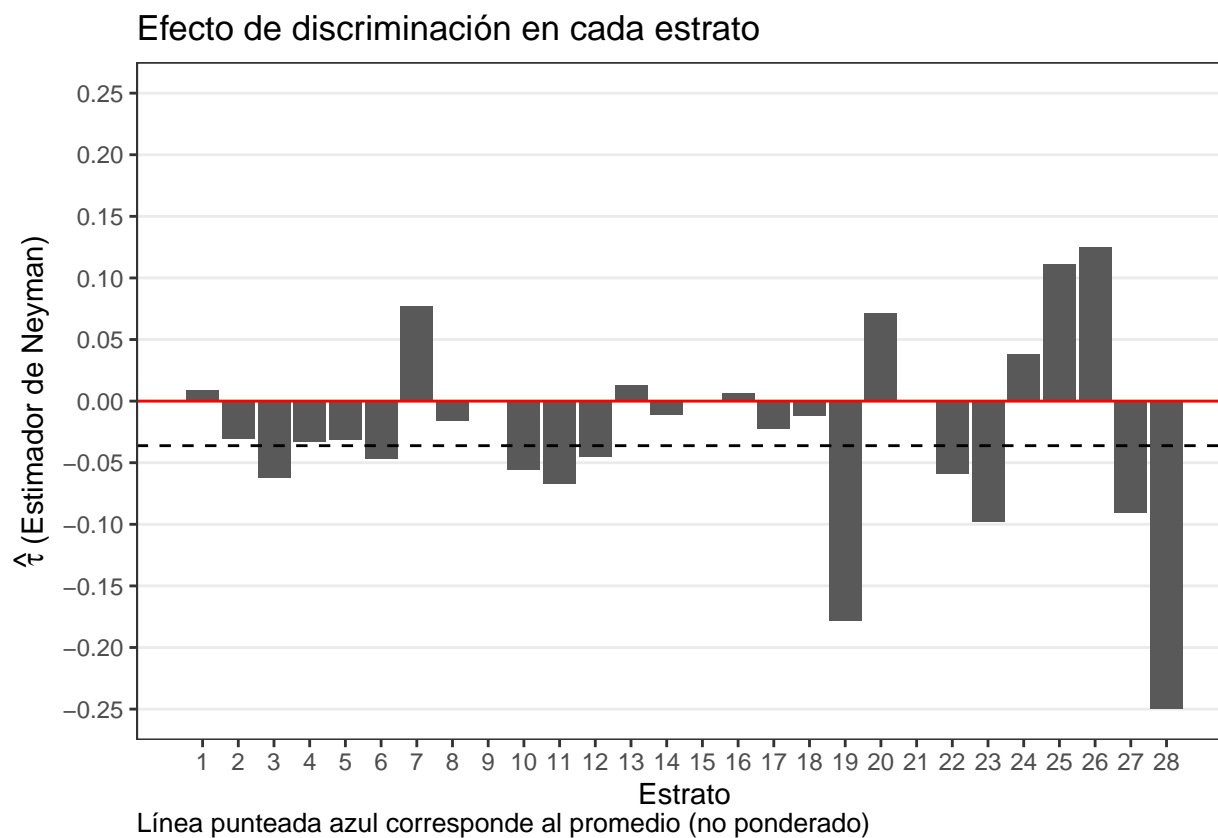
De forma alternativa, se realizó el Test con la función de R *twoSamplePermutationTestLocation*.

```
# con fórmula de R
attach(data)
x <- call_back[black==1]
y <- call_back[black==0]
twoSamplePermutationTestLocation(x,y, fcn="mean", mu1.minus.mu2=-0.01,
                                  seed=666, n=1000)
```

Se obtiene un  $p\text{-Value} = 0.004$ . La conclusión a la que se llega es la misma; con un nivel de significancia del 5% sí hay evidencia estadística para rechazar la hipótesis nula  $H_0 : CB_{afram} = CB_{white} - 0.01$ .

5. Imagina que estratificas por: (i) sexo del aplicante (hombre o mujer), (ii) ciudad donde se postula al trabajo (Chicago o Boston) e (iii) industria de la empresa que publico el puesto (ver el pdf que indica las industrias disponibles) [Ejemplo: un posible estrato seríaa hombres aplicantes a trabajos en Chicago en la industria manufacturera]. Empleando todas las combinaciones posibles de las variables (i)-(iii), utiliza el método de Neyman para calcular el efecto de discriminación en cada estrato (elige el formato que quieras para reportar este resultado, tabla o gráfica). Utilizando los efectos por estrato, calcula el efecto promedio de tratamiento. Compara este estimador promedio y la varianza con el resultado que obtuviste en la pregunta (3).

A continuación se presenta de manera gráfica el efecto de discriminación en cada estrato, la línea punteada azul corresponde al promedio simple de todos los estratos.





**Tabla 4. Estratos y efecto promedio de tratamiento al interior de cada estrato**

estrato	n	calls	female	chicago	manuf	transcom	bankreal	trade	busservice	othservice	missind	tau_hat	weight
1	181	11	1	1	1	0	0	0	0	0	0	0.0090376	0.0371663
2	469	35	1	1	0	0	0	0	0	1	0	-0.0308067	0.0963039
3	165	16	0	1	0	0	0	1	0	0	0	-0.0623342	0.0338809
4	337	21	1	1	0	0	0	1	0	0	0	-0.0330255	0.0691992
5	678	39	1	1	0	0	0	0	1	0	0	-0.0314327	0.1392197
6	287	19	1	1	0	0	1	0	0	0	0	-0.0464879	0.0589322
7	29	1	0	1	1	0	0	0	0	0	0	0.0769231	0.0059548
8	373	17	1	1	0	0	0	0	0	0	1	-0.0158415	0.0765914
9	31	0	0	1	0	0	0	0	0	1	0	0.0000000	0.0063655
10	66	9	0	1	0	0	0	0	1	0	0	-0.0555556	0.0135524
11	401	49	1	0	0	0	0	0	1	0	0	-0.0666700	0.0823409
12	260	21	1	0	0	0	0	1	0	0	0	-0.0447407	0.0533881
13	242	26	1	0	0	0	0	0	0	0	1	0.0124726	0.0496920
14	280	14	0	0	0	0	0	1	0	0	0	-0.0107280	0.0574949
15	18	0	0	0	0	0	0	0	0	1	0	0.0000000	0.0036961
16	91	2	0	0	1	0	0	0	0	0	0	0.0064103	0.0186858
17	159	12	0	0	0	0	0	0	1	0	0	-0.0223172	0.0326489
18	236	43	1	0	0	0	0	0	0	1	0	-0.0115636	0.0484600
19	103	8	1	0	1	0	0	0	0	0	0	-0.1777778	0.0211499
20	23	1	0	1	0	0	1	0	0	0	0	0.0714286	0.0047228
21	13	4	0	1	0	0	0	0	0	0	1	-0.3571429	0.0026694
22	176	14	0	0	0	0	0	0	0	0	1	-0.0584795	0.0361396
23	41	5	0	0	0	1	0	0	0	0	0	-0.0975000	0.0084189
24	55	6	1	0	0	1	0	0	0	0	0	0.0380435	0.0112936
25	36	6	1	1	0	1	0	0	0	0	0	0.1111111	0.0073922
26	16	3	0	1	0	1	0	0	0	0	0	0.1250000	0.0032854
27	88	8	1	0	0	0	1	0	0	0	0	-0.0909091	0.0180698
28	16	2	0	0	0	0	1	0	0	0	0	-0.2500000	0.0032854

<sup>a</sup> tau\_hat es la diferencia en la probabilidad de recibir una llamada si se tiene un nombre afroamericano vs. si se tiene un nombre blanco

<sup>a</sup> weight es la proporción (del total de la muestra) de individuos en cada estrato

Utilizando los efectos por estrato obtenemos un efecto promedio de tratamiento  $\tau = -0.0328$  y una varianza de: 0.0001. En la siguiente tabla se muestra la comparación de los resultados obtenidos en la pregunta 3 y en esta pregunta:

Pregunta	tau	Varianza
(3): estimador de Neyman	-0.0320329	0.0000606
(5): estratificación	-0.0328092	0.0000603

Llama la atención que se obtienen valores casi idénticos. Si se observa la tabla 1 del *paper* tal vez se podría pensar que los autores llevaron a cabo la aleatorización por estratos.

6. Replica la primera parte de la Tabla 7 del paper. Solo para el renglón de “Total Number of Requirements” da una interpretación lo más específica posible de la columna “marginal effects.” (Ojo: Puedes considerar los errores estándar que arroja por default el software que utilices).

**Tabla 5. Effect of Job Requirement and Employer Characteristics on Racial Differences in Callbacks**

Job requirement	Sample mean (standard deviation)	Marginal effect on callbacks for Afroamerican names
Any requirement	0.79 ( 0.41 )	0.022 ( 0.018 )
Experience	0.44 ( 0.5 )	0.01 ( 0.016 )
Computer skills	0.44 ( 0.5 )	0.001 ( 0.016 )
Communication skills	0.12 ( 0.33 )	0 ( 0.024 )
Organizational skills	0.07 ( 0.26 )	0.028 ( 0.042 )
Education	0.11 ( 0.31 )	-0.036 ( 0.03 )
Total number of requirements	1.18 ( 0.93 )	0.002 ( 0.009 )

<sup>a</sup> Tabla 7 del paper

En la fila de *Total number of requirements* de la columna (3) de la tabla 5 se muestra el efecto marginal de la interacción entre ser afroamericano (*black*) y el número total de requerimientos del empleo ( *Total number of requirements* ) de la siguiente ecuación de regresión estimada con un modelo Probit.

$$call\_back_i = \beta_0 + \beta_1 black_i + \beta_2 total\_number\_requirements + \beta_3 (black_i \times total\_number\_requirements) + U_i$$

El efecto marginal de *Total number of requirements* indica qué tan mayor es el *call\_back* ante un aumento de una unidad de *Total number of requirements* para los afroamericanos que para los blancos. En este caso la interpretación es: Ante un aumento de 1 requerimiento del empleo, para una persona afroamericana se espera un aumento en *call\_back* de 0.2 puntos porcentuales más que para una persona blanca. Sin embargo, hay que aclarar que el coeficiente no es estadísticamente significativo, se puede observar directamente en el error estándar de la tabla 5. Por lo tanto no hay un efecto claro del número de requerimientos del empleo sobre la discriminación racial.

7. Quisieras saber si la discriminación racial disminuye conforme aumenta la experiencia laboral de los aplicantes. Elige el método y formato que prefieras para reportar tus resultados. Muestra claramente qué parámetro o combinación de parámetros contestan tu pregunta.

A continuación se presenta la ecuación que se busca estimar con un modelo Probit.

$$call\_back_i = \beta_0 + \beta_1 black_i + \beta_2 yearsexp_i + \beta_3 (black_i \times yearsexp_i) + U_i$$

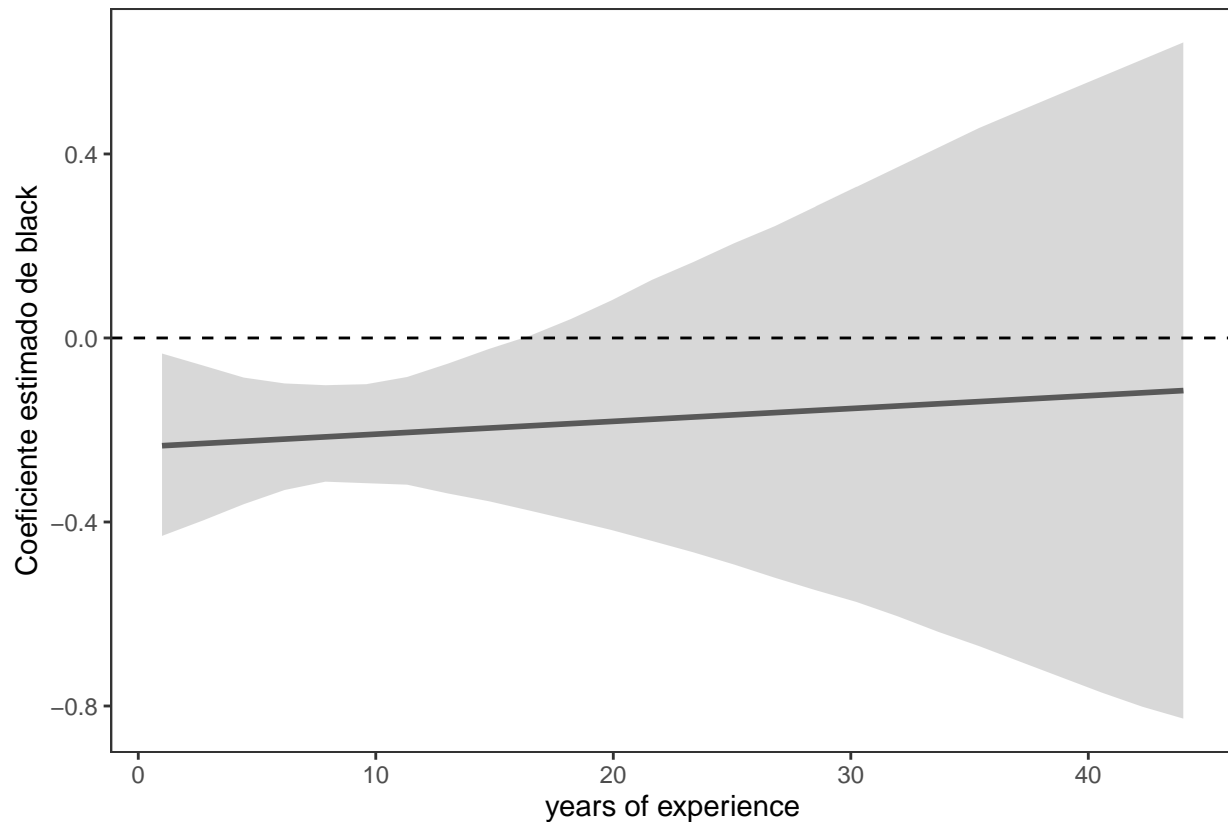
La prueba de hipótesis que permite analizar si la discriminación racial disminuye conforme aumenta la experiencia laboral de los aplicantes es:

$$H_0 : \beta_3 \neq 0$$

**Tabla 6**

<i>Dependent variable:</i>	
	call_back
black	-0.242** (0.097)
yearsexp	0.019*** (0.006)
black:yearsexp	0.003 (0.010)
Constant	-1.457*** (0.064)
Observations	4,870
Log Likelihood	-1,346.386
Akaike Inf. Crit.	2,700.772
<i>Note:</i> *p<0.1; **p<0.05; ***p<0.01	

En la siguiente gráfica se gráfica  $\beta_1$  en función de los años de experiencia.



Se observa que conforme aumentan los años de experiencia no hay un efecto claro (no hay evidencia al 95% de confianza) de si el coeficiente de black aumenta o disminuye. Además, el coeficiente de la interacción  $\beta_3$  no es estadísticamente significativo. Por lo tanto, con el uso del modelo previo, no hay evidencia estadística para determinar si la discriminación racial aumenta o disminuye conforme aumenta la experiencia laboral de los aplicantes.

**8. Por último, imagina que el gobierno está interesado en replicar este estudio en México para ver posible discriminación en contra de indígenas. Te pide que lo asesores para definir el número de CVs ficticios (aplicaciones) que necesita realizar. Realiza cálculos de poder para indicar:**

(a) Cuántos CVs ficticios necesitaría aleatorizar si es que: (i) tu anticipas que los efectos (varianza y efecto real) sean iguales a los obtenidos por Bertrand y Mullainathan, (ii) quieres un poder estadístico de 85%, (iii) asumes una significancia de 1%, y (iv) vas a dividir 50-50 tratamiento y control?

Para calcular el tamaño de la muestra empleamos la fórmula derivada en Athey & Imbens (2017):

$$N = \frac{(\Phi^{-1}(\beta) + \Phi^{-1}(1 - \alpha/2))^2}{(\tau^2/\sigma_Y^2) \cdot \gamma \cdot (1 - \gamma)} \quad (4)$$

Para  $\alpha = 0.01$ ,  $\beta = 0.85$ ,  $\tau = -0.032$ ,  $Var(call\_back) = \sigma^2 = 0.07403$  y  $\gamma = 0.5$  tenemos que

$$N = 3765.5532$$

En la siguiente tabla se muestra el tamaño de muestra requerido si se considerara la varianza del grupo de control y el grupo de tratamiento:

Varianza	N
Sigma = Var(call_back)	3766
Sigma = Var(call_back black=0)	4437
Sigma = Var(call_back black=1)	3069

(b) En R o Stata, produce una gráfica que ilustre el tradeoff entre poder estadístico y proporción de tratamiento y control (similar a lo que hicimos con Optimal Design) fijando los valores que obtuviste en el inciso anterior.

Para  $\alpha = 0.01$ ,  $\tau = -0.032$ ,  $Var(call\_back) = \sigma^2 = 0.07403$  y  $N = 3766$

