

Tarea 01

Jesús Gustavo Hernández Herrerías 142258

1 de Marzo del 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?

Se puede pensar a $D_i = \{0, 1\}$ como una variable binaria random para indicar si el nombre asignado es común entre afroamericanos. La pregunta es si Y_i es afectada por el nombre que recibe común a cierto tipo de población Si los resultados potenciales son:

$$Y_i = \begin{cases} Y_{1i} & \text{if } D_i = 1 \\ Y_{0i} & \text{if } D_i = 0 \end{cases}$$

En otras palabras, Y_{0i} es cuando no hubo discriminación racial (más llamadas) sin importar si recibió el nombre de blancos o afroamericanos, mientras que Y_{1i} es el caso de que si hubo discriminación (menos llamadas).

$$E(Y_i|D_i = 1) - E(Y_i|D_i = 0) = E(Y_{1i}|D_i = 1) - E(Y_{0i}|D_i = 1) + E(Y_{0i}|D_i = 1) - E(Y_{0i}|D_i = 0)$$

El sesgo de selección se puede ver como el último término de la ecuación de arriba, donde este significa la diferencia en promedio Y_{0i} entre aquellos que recibieron un nombre de población afroamericana y aquellos que recibieron nombres de población blanca.

Para no tener un sesgo de selección se aleatoriza, ya que pudiera ser que los CV's de cierto tipo de población tengan cierta tendencia a ser de un mayor o menor nivel de calidad. El sesgo de selección hubiese resultado positivo si es que por ejemplo la calidad de los CV's de los afroamericanos fuera de menor calidad, entonces esta población tendría una mayor probabilidad de recibir discriminación o menores llamadas de vuelta. Como resultado nos queda que cualquier diferencia solo se debe a la manipulación de la raza a través del nombre.

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.

En la tabla 3 se puede observar que la media de años de experiencia por nombre es muy similar entre los tipos de nombres para personas afroamericanas y blancas. De igual forma la dispersión entre años de experiencia entre nombres nos permite afirmar que no hay un patrón que sugiera no se hicieron de forma aleatoria.

En la tabla 4 la media en la calidad de los CV's de acuerdo a la clasificación que se hizo de las variables por parte de los autores nos permite observar que no hay un patrón entre los datos dependiendo los nombres asignados.

En la tabla 5 con medias de cada variable por nombre, nos permite identificar de manera más sencilla que las variables no tienen alguna tendencia mayor para cierto tipo de población dependiendo su nombre.

Table 1: Summary Statistics

	variable	mean	n	0	0.05	0.1	0.25	0.5	0.75	0.9	0.95	1
1	ofjobs	3.661	4,870	1	2	2	3	4	4	5	6	7
2	yearsexp	7.843	4,870	1	2	3	5	6	9	14	19	44
3	honors	0.053	4,870	0	0	0	0	0	0	0	1	1
4	volunteer	0.411	4,870	0	0	0	0	0	1	1	1	1
5	military	0.097	4,870	0	0	0	0	0	0	0	1	1
6	empholes	0.448	4,870	0	0	0	0	0	1	1	1	1
7	workinschool	0.560	4,870	0	0	0	0	1	1	1	1	1
8	email	0.479	4,870	0	0	0	0	0	1	1	1	1
9	computerskills	0.821	4,870	0	0	0	1	1	1	1	1	1
10	specialskills	0.329	4,870	0	0	0	0	0	1	1	1	1
11	eo	0.291	4,870	0	0	0	0	0	1	1	1	1
12	manager	0.152	4,870	0	0	0	0	0	0	1	1	1
13	supervisor	0.077	4,870	0	0	0	0	0	0	0	1	1
14	secretary	0.333	4,870	0	0	0	0	0	1	1	1	1
15	offsupport	0.119	4,870	0	0	0	0	0	0	1	1	1
16	salesrep	0.151	4,870	0	0	0	0	0	0	1	1	1
17	retailsales	0.168	4,870	0	0	0	0	0	0	1	1	1
18	req	0.787	4,870	0	0	0	1	1	1	1	1	1
19	expreq	0.435	4,870	0	0	0	0	0	1	1	1	1
20	comreq	0.125	4,870	0	0	0	0	0	0	1	1	1
21	educreq	0.107	4,870	0	0	0	0	0	0	1	1	1
22	compreq	0.437	4,870	0	0	0	0	0	1	1	1	1
23	orgreq	0.073	4,870	0	0	0	0	0	0	0	1	1
24	manuf	0.083	4,870	0	0	0	0	0	0	0	1	1
25	transcom	0.030	4,870	0	0	0	0	0	0	0	0	1
26	bankreal	0.085	4,870	0	0	0	0	0	0	0	1	1
27	trade	0.214	4,870	0	0	0	0	0	0	1	1	1
28	busservice	0.268	4,870	0	0	0	0	0	1	1	1	1
29	othservice	0.155	4,870	0	0	0	0	0	0	1	1	1
30	missind	0.165	4,870	0	0	0	0	0	0	1	1	1
31	black	0.500	4,870	0	0	0	0	0.500	1	1	1	1
32	chicago	0.555	4,870	0	0	0	0	1	1	1	1	1
33	high	0.502	4,870	0	0	0	0	1	1	1	1	1
34	female	0.769	4,870	0	0	0	1	1	1	1	1	1
35	college	0.720	4,870	0	0	0	0	1	1	1	1	1
36	call_back	0.080	4,870	0	0	0	0	0	0	0	1	1

Table 2: Tabla de balance

	variables1	Media_controll	Media_trat1	p_value1
1	call_back	0.097	0.064	0.00004
2	chicago	0.555	0.555	1
3	female	0.764	0.775	0.377
4	high	0.502	0.502	1

Podemos observar que dado que las diferencias en la tabla de balance no son significativas, por lo que podemos asegurar que el experimento ha sido aleatorio.

Table 3: Años de experiencia

	firstname	media	max	min	d.e.
1	Aisha	7.020	20	1	3.440
2	Allison	7.870	26	1	4.780
3	Anne	7.870	26	1	5.100
4	Brad	6.760	26	2	4.600
5	Brendan	8.320	26	2	5.570
6	Brett	9.150	26	2	7.360
7	Carrie	8.070	26	2	4.910
8	Darnell	7.360	26	2	4.730
9	Ebony	7.930	26	1	4.900
10	Emily	7.680	26	1	4.930
11	Geoffrey	8.900	26	2	7.010
12	Greg	7.180	26	1	5.230
13	Hakim	7.070	26	2	4.490
14	Jamal	8	26	2	5.980
15	Jay	7.570	26	1	4.990
16	Jermaine	8	26	2	5.380
17	Jill	8.970	26	1	5.920
18	Kareem	6.580	26	2	3.620
19	Keisha	7.190	26	2	4.490
20	Kenya	8.080	26	1	5.100
21	Kristen	7.760	26	1	4.870
22	Lakisha	8.350	26	1	5.310
23	Latonya	7.990	26	1	4.890
24	Latoya	8.480	44	1	6.180
25	Laurie	7.380	22	1	4.450
26	Leroy	7.420	26	2	5.340
27	Matthew	7.120	20	1	3.930
28	Meredith	7.790	26	1	4.540
29	Neil	8.010	26	2	4.810
30	Rasheed	6.600	26	2	5.150
31	Sarah	7.840	26	2	5.220
32	Tamika	8.250	26	1	5.110
33	Tanisha	8.070	26	1	4.750
34	Todd	6.280	26	2	3.610
35	Tremayne	7.200	26	2	4.770
36	Tyrone	8.120	26	2	5.280

Table 4: Calidad del CV

	firstname	media	max	min	d.e.
1	Aisha	0.440	1	0	0.500
2	Allison	0.500	1	0	0.500
3	Anne	0.480	1	0	0.500
4	Brad	0.490	1	0	0.500
5	Brendan	0.480	1	0	0.500
6	Brett	0.530	1	0	0.500
7	Carrie	0.540	1	0	0.500
8	Darnell	0.620	1	0	0.490
9	Ebony	0.480	1	0	0.500
10	Emily	0.510	1	0	0.500
11	Geoffrey	0.490	1	0	0.500
12	Greg	0.590	1	0	0.500
13	Hakim	0.530	1	0	0.500
14	Jamal	0.380	1	0	0.490
15	Jay	0.510	1	0	0.500
16	Jermaine	0.500	1	0	0.500
17	Jill	0.530	1	0	0.500
18	Kareem	0.360	1	0	0.480
19	Keisha	0.490	1	0	0.500
20	Kenya	0.560	1	0	0.500
21	Kristen	0.490	1	0	0.500
22	Lakisha	0.480	1	0	0.500
23	Latonya	0.490	1	0	0.500
24	Latoya	0.500	1	0	0.500
25	Laurie	0.510	1	0	0.500
26	Leroy	0.500	1	0	0.500
27	Matthew	0.520	1	0	0.500
28	Meredith	0.460	1	0	0.500
29	Neil	0.490	1	0	0.500
30	Rasheed	0.580	1	0	0.500
31	Sarah	0.510	1	0	0.500
32	Tamika	0.550	1	0	0.500
33	Tanisha	0.500	1	0	0.500
34	Todd	0.460	1	0	0.500
35	Tremayne	0.550	1	0	0.500
36	Tyrone	0.560	1	0	0.500

Table 5: Características del grupo por nombre

firstname	Callback	Female	Black	HQresume	Chicago	Yearsexp	College	Email
Aisha	0.02	1	1	0.44	0.69	7.02	0.69	0.45
Allison	0.09	1	0	0.50	0.66	7.87	0.60	0.50
Anne	0.08	1	0	0.48	0.65	7.87	0.67	0.46
Brad	0.16	0	0	0.49	0.25	6.76	0.86	0.46
Brendan	0.08	0	0	0.48	0.40	8.32	0.86	0.42
Brett	0.07	0	0	0.53	0.27	9.15	0.88	0.42
Carrie	0.13	1	0	0.54	0.67	8.07	0.71	0.52
Darnell	0.05	0	1	0.62	0.43	7.36	0.79	0.50
Ebony	0.10	1	1	0.48	0.58	7.93	0.69	0.47
Emily	0.08	1	0	0.51	0.67	7.68	0.70	0.51
Geoffrey	0.07	0	0	0.49	0.29	8.90	0.90	0.44
Greg	0.08	0	0	0.59	0.18	7.18	0.94	0.55
Hakim	0.05	0	1	0.53	0.27	7.07	0.82	0.47
Jamal	0.07	0	1	0.38	0.36	8.00	0.89	0.30
Jay	0.13	0	0	0.51	0.40	7.57	0.85	0.42
Jermaine	0.10	0	1	0.50	0.33	8.00	0.90	0.42
Jill	0.08	1	0	0.53	0.61	8.97	0.69	0.52
Kareem	0.05	0	1	0.36	0.19	6.58	0.94	0.34
Keisha	0.04	1	1	0.49	0.68	7.19	0.74	0.50
Kenya	0.09	1	1	0.56	0.63	8.08	0.65	0.56
Kristen	0.13	1	0	0.49	0.64	7.76	0.67	0.48
Lakisha	0.06	1	1	0.48	0.56	8.35	0.66	0.48
Latonya	0.09	1	1	0.49	0.60	7.99	0.67	0.47
Latoya	0.08	1	1	0.50	0.66	8.48	0.71	0.48
Laurie	0.10	1	0	0.51	0.63	7.38	0.67	0.50
Leroy	0.09	0	1	0.50	0.36	7.42	0.86	0.45
Matthew	0.09	0	0	0.52	0.28	7.12	0.88	0.45
Meredith	0.10	1	0	0.46	0.59	7.79	0.66	0.44
Neil	0.07	0	0	0.49	0.28	8.01	0.93	0.37
Rasheed	0.03	0	1	0.58	0.25	6.60	0.84	0.51
Sarah	0.10	1	0	0.51	0.57	7.84	0.62	0.52
Tamika	0.05	1	1	0.55	0.66	8.25	0.68	0.53
Tanisha	0.06	1	1	0.50	0.59	8.07	0.62	0.49
Todd	0.06	0	0	0.46	0.32	6.28	0.87	0.43
Tremayne	0.04	0	1	0.55	0.32	7.20	0.86	0.49
Tyrone	0.05	0	1	0.56	0.32	8.12	0.95	0.45

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.

- (a) Indica la prueba de hipótesis que estarás contrastando en cada estimación.
- (b) Reporta los resultados de tus 4 estimaciones con una tabla con el formato usual que empleamos el semestre pasado.
- (c) Asegúrate que los resultados reportados en cada columna sean comparables. Es decir, deberán estar reportados en las mismas unidades para poder hacer una comparación a lo largo de las columnas.
- (d) Elige una de las columnas para llevar a cabo una interpretación del coeficiente relevante que estas estimando. Da evidencia como parte de esta interpretación de la importancia del efecto. Es decir, ¿consideras que es un efecto pequeño o grande?

1. Estimador de Neyman

Definimos:

$$\tau = \frac{1}{N} \sum_{i=1}^n (Y_i^T - Y_i^c)$$

Estimador de Neyman:

$$\hat{\tau} = \bar{Y}^1 - \bar{Y}^0$$

Donde \bar{Y}^1 representa el resultado de la variable `call_back` si hubo un efecto de tratamiento (nombre afroamericano) y \bar{Y}^0 representa el resultado de la variables `call_back` si es parte del grupo de control (nombre de blanco).

La prueba de hipótesis relevante es la siguiente:

$$\begin{cases} H_0 : \tau = 0 \\ H_1 : \tau \neq 0 \end{cases}$$

En promedio importa o no el ser afroamericano para ser llamado por los reclutadores.

Table 6: Estimador de Neyman

black	0.032
se	0.008
Estadístico t	4.115
Valor-p	0.00004

Se rechaza H_0 a favor de un efecto significativo de ser población afroamericana. Nota: el valor del estimador de Neyman, está en valor absoluto.

2. Estimación de OLS con errores heterocedásticos

La prueba de hipótesis relevante es la siguiente:

$$\begin{cases} H_0 : \beta_1 = 0 \\ H_1 : \beta_1 \neq 0 \end{cases}$$

Existe un efecto diferenciado para los aplicantes para ser llamados si son afroamericanos contra ser blancos.

3. Estimación de OLS agregando controles

$$\begin{cases} H_0 : \text{efectosdelasvariablesdecontrol} = 0 \\ H_1 : \text{e.o.c} \end{cases}$$

4. Probit sin controles

$$\begin{cases} H_0 : \beta_1 = 0 \\ H_1 : \beta_1 \neq 0 \end{cases}$$

Conclusiones:

En los 3 modelos el hecho de ser hombre afroamericano esta relacionado con una disminucion en la probabilidad de ser llamada por el reclutador. Por otro lado, Neyman nos dice que si existe un efecto diferenciado para la población afroamericana.

Al observar la columna de OLS con errores heterocedásticos (OLS EH) podemos observar que el hecho de ser un hombre afroamericano esta relacionado con una disminución de 3.2 puntos porcentuales en la probabilidad de ser llamado por el reclutador. Además el valor p es significativo en favor de un efecto en favor del tratamiento de ser afroamericano. Cuando agregas controles el efecto de ser afroamericano es un poco menor.

Table 7: Comparación de modelos

black	0.032
se	0.008
Estadístico t	4.115
Valor-p	0.00004

Table 8: Comparación de modelos

	<i>Dependent variable:</i>		
	call_back		
	OLS EH	OLS	probit
	(1)	OLS CONT	Probit
	(1)	(2)	(3)
black	−0.032*** (0.008)	−0.021 (0.018)	−0.217*** (0.053)
female		0.022 (0.013)	
I(black *female)		−0.003 (0.018)	
high		0.023** (0.011)	
I(black *high)		−0.018 (0.016)	
chicago		−0.034*** (0.008)	
Constant	0.097*** (0.006)	0.087*** (0.013)	−1.302*** (0.035)
Observations	4,870	4,870	4,870
R ²	0.003	0.008	
Adjusted R ²	0.003	0.007	
Log Likelihood			−1,354.969
Akaike Inf. Crit.			2,713.938
Residual Std. Error	0.272 (df = 4868)	0.271 (df = 4863)	
F Statistic	16.931*** (df = 1; 4868)	6.758*** (df = 6; 4863)	
<i>Note:</i>		*p<0.1; **p<0.05; ***p<0.01	

Nota: el valor del estimador de Neyman, está en valor absoluto.

	dF/dx	Std. Err.	z	P> z
black	-0.032	0.008	-4.116	0.00004

Figure 1: ProbitMFX

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:

$$H_0 : C\beta_{blancos} = C\beta_{afroam} + 0.01$$

donde CBj es la proporción del grupo $j = \{\text{blancos; afroam}\}$ que reciben una llamada para entrevista (medida entre 0 y 1). Utiliza un Fischer Exact Test para evaluar esta hipótesis. Reporta el valor-p y la conclusión a la que llegas.

Valor-p [1] 0.005 Se rechaza H0 de que la diferencia en el call_back es de 1pp a favor de los blancos.

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 sería 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).

Table 9: Estratos

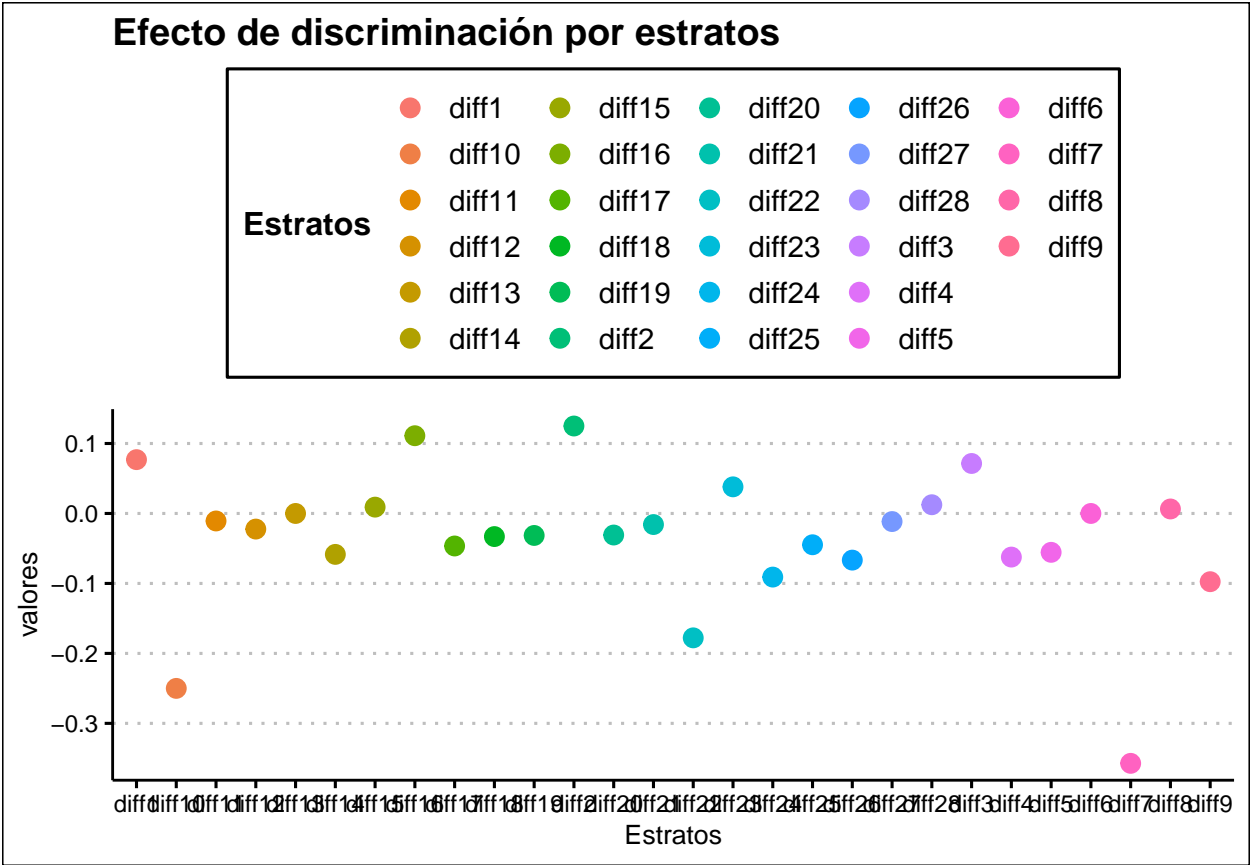
Estratos
"HChiMa"
"HChiTrans"
"HChiBan"
"HChiTra"
"HChiBus"
"HChiOth"
"HChiMiss"
"HBosMa"
"HBosTrans"
"HBosBan"
"HBosTra"
"HBosBus"
"HBosth"
"HBosMiss"
"MChiMa"
"MChiTrans"
"MChiBan"
"MChiTra"
"MChiBus"
"MChiOth"
"MChiMiss"
"MBosMa"
"MBosTrans"
"MBosBan"
"MBosTra"
"MBosBus"
"MBosth"
"MBosMiss"

Table 10: Efecto de discriminación por estratos

Estratos	valores
diff1	0.0769231
diff2	0.1250000
diff3	0.0714286
diff4	-0.0623342
diff5	-0.0555556
diff6	0.0000000
diff7	-0.3571429
diff8	0.0064103
diff9	-0.0975000
diff10	-0.2500000
diff11	-0.0107280
diff12	-0.0223172
diff13	0.0000000
diff14	-0.0584795
diff15	0.0090376
diff16	0.1111111
diff17	-0.0464879
diff18	-0.0330255
diff19	-0.0314327
diff20	-0.0308067
diff21	-0.0158415
diff22	-0.1777778
diff23	0.0380435
diff24	-0.0909091
diff25	-0.0447407
diff26	-0.0666700
diff27	-0.0115636
diff28	0.0124726

Table 11: Diferencia entre el estimador considerando los estratos vs el estimador de Neyman(3)

Estimador de Neyman
-0.0161384
-0.0320329



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).

	dF/dx	Std. Err.	z	P> z
black	-0.034	0.012	-2.746	0.006
total_req	-0.014	0.006	-2.416	0.016
I(black * total_req)	0.002	0.009	0.175	0.861

Figure 2: Replica tabla 7 paper

Observamos que el dato que buscamos es el que pertenece a la interacción de black con total requirement y es el mismo dato que en la tabla 7.

Un aumento en 1 unidad de requerimientos totales, está relacionado con un aumento de 0.2 pts porcentuales en la probabilidad de recibir una llamada de vuelta por parte de los reclutadores para la población afroamericana en comparación con la población blanca. Sin embargo, el dato resulta ser no significativo.

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

Table 12:

	<i>Dependent variable:</i>	
	call_back	
	<i>OLS</i>	<i>probit</i>
	(1)	(2)
black	0.002 (0.025)	-0.065 (0.173)
yearsexp	0.011*** (0.004)	0.066*** (0.024)
I(black *yearsexp)	-0.008 (0.005)	-0.038 (0.033)
I(yearsexp *yearsexp)	-0.0003** (0.0001)	-0.002** (0.001)
I(black *yearsexp *yearsexp)	0.0003 (0.0002)	0.002 (0.001)
Constant	0.038** (0.018)	-1.663*** (0.121)
Observations	4,870	4,870
R ²	0.008	
Adjusted R ²	0.007	
Log Likelihood		-1,344.218
Akaike Inf. Crit.		2,700.437
Residual Std. Error	0.271 (df = 4864)	
F Statistic	7.998*** (df = 5; 4864)	
<i>Note:</i> *p<0.1; **p<0.05; ***p<0.01		

Se puede observar como para la interacción de la variable black con years (ATE|exp), tiene un signo negativo por lo que no aumenta la probabilidad de ser llamado de vuelta por los reclutadores para los afroamericanos, sin embargo, en el caso de la variable year, si hay un efecto positivo en la probabilidad de ser llamado de vuelta, ya sea para el caso del modelo de probabilidad lineal donde la interpretación es directa o el probit.

8. Por último, imagina que el gobierno esta 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 necesitará 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?

(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) Dejando los valores que obtuviste en el inciso anterior (número de observaciones, efectos reales y significancia).

Número de CV's 3768

Two-sample t test power calculation

```
n = 1884.436
delta = 0.03203285
sd = 0.2720826
sig.level = 0.01
power = 0.85
alternative = two.sided
```

NOTE: n is number in *each* group

