


## Tarea 1

**Fecha de entrega: 24 de febrero**

**Nota:** Deberán subir a *Canvas* un archivo de texto con sus respuestas. Pueden utilizar el formato de su preferencia (e.g.  $\text{\LaTeX}$ , Word u hojas escritas a mano y escaneadas). Además, deberán subir otro archivo que genere todos sus resultados de las preguntas prácticas. Puede ser un R-script, Do-File o script de algún otro software similar.

Bertrand y Mullainathan (2004) estaban interesados en determinar el grado de discriminación racial que podía darse en el mercado laboral en Estados Unidos. Para esto utilizaron un experimento aleatorizado. Su experimento consistió en preparar CVs ficticios de diversa calidad. A cada CV se le asignaría de manera aleatoria un nombre. Los nombres utilizados en su experimento fueron nombres que se utilizan mayoritariamente entre afroamericanos (e.g. Tanisha y Hakeem) y otros que se utilizan mayoritariamente entre blancos (e.g. Allison y Todd). Los CVs fueron enviados como respuesta a distintos anuncios que fueron publicados en periódicos. La idea era ver la diferencia en llamadas para entrevista (*call-back*) que podían recibir CVs con nombres afroamericanos versus CVs con nombres de blancos. 

Para esta tarea utilizarás la base de datos `Names.dta` que está disponible en *Canvas*. De igual manera, en *Canvas* encontrarás la descripción de las variables de la base de datos en el archivo `nombres_des.pdf`.

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

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:

$$H_0 : CB_{blancos} = CB_{afroam} + 0.01 \quad (1)$$

donde  $CB_j$  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.

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).
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).
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.
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 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?
- (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 (número de observaciones, efectos reales y significancia).