

Taller con R

Norman Simón Rodríguez

21 de julio de 2016

Ejercicio en R: evaluando una política pública

En este taller estaremos trabajando con los datos de una evaluación de política pública realizada en Colombia hace unos años por el DNP (Departamento Nacional de Planeación) y la firma Econometría. El artículo puede ser descargado en este link (el archivo se llama **sinergia_educacion.pdf**) y los datos pueden ser encontrados en este link: https://anda.dnp.gov.co/index.php/catalog/29/get_microdata.

Contexto de la evaluación

Entre 2009 y 2011 el Ministerio de Educación abrió el Fondo de Fomento a la Educación Media (FEM) para financiar programas de capacitación laboral a los estudiantes de la media vocacional. Estos programas fueron operados por instituciones de educación superior (IES), entre las cuales hubo varias universidades. También hubo programas de capacitación laboral por parte del SENA, mas estos no fueron financiados con el fondo FEM. Esta evaluación estima el impacto de la Estrategia de Articulación de la Educación Media, que incluye tanto las capacitaciones dadas a través del FEM como las que dio el SENA.

La muestra se compone de 162 instituciones de educación media, de las cuales 82 hicieron parte de la Estrategia (grupo de tratamiento). Las restantes 80 no hicieron parte de la Estrategia (grupo de control). Las instituciones están radicadas en 55 municipios. De entre todas las instituciones se seleccionaron 2600 egresados, una parte para el grupo de control y otra para el grupo de tratamiento. A estos egresados se les aplicó una encuesta y a partir de los datos de la encuesta se hizo la evaluación. El texto de la encuesta puede ser encontrado en este link link (archivo: **encuesta.pdf**). La población es todos los municipios colombianos con 100 o más beneficiarios (un total de 318.591 jóvenes en 518 municipios).

La evaluación mira si el programa afectó positivamente los resultados del SABER 11, y la situación académica/laboral general de los estudiantes.

La evaluación se realizó en 2012, su informe final fue entregado en 2013 y la última actualización de microdatos fue en agosto de 2015.

Cómo se calificará este taller

En este taller haremos varias gráficas a las cuales usted debe añadir su nombre para que puedan ser calificadas. También se pedirán pantallazos del código.

Cargando y organizando los datos en R

Vamos a cargar los datos de la encuesta principal. El archivo de datos se llama **encuestaA.csv** y se encuentra en la carpeta “data” (esta carpeta no se encuentra publicada en la versión GitHub de este ejercicio por motivos de propiedad intelectual).

```
encuestaA <- read.csv("./data/encuestaA.csv")
```

Vamos a mirar cómo mejoraron las capacidades de estos estudiantes. Por ejemplo, la capacidad de aplicar los conocimientos adquiridos en la práctica (pertinencia educativa). Se les preguntó a los encuestados que calificaran de 1 a 5 su capacidad, siendo 5 la máxima. El número de la pregunta es M5_501B.

Vamos a invocar esta pregunta. Como son muchos datos (2600) sólo vamos a mirar los primeros. Para ello utilizamos la función `head()`:

```
head(encuestaA$M5_501B)
```

```
## [1] 4 4 4 4 5 5
```

Este vector incluye tanto a los egresados del grupo de control como a los del tratamiento. Para poder compararlos entre ellos debemos separarlos.

La pregunta M2_201 es la que nos dice si los estudiantes participaron en la Estrategia de Articulación o no. Entonces vamos a crear una nueva tabla que sólo incluya a los de tratamiento y otra que sólo incluya a los de control (vamos a partir en dos el objeto `encuestaA`).

Grupo de tratamiento:

```
tratamiento <- subset(encuestaA, encuestaA$M2_201=="Si")
```

Grupo de control:

```
control <- subset(encuestaA, encuestaA$M2_201=="No")
control <- rbind(control, subset(encuestaA, is.na(encuestaA$M2_201)))
```

Contemos cuántos jóvenes quedaron en cada grupo:

```
NROW(control)
```

```
## [1] 1268
```

```
NROW(tratamiento)
```

```
## [1] 1351
```

Un primer análisis

Ahora volvamos a mirar la pregunta anterior (capacidad para aplicar lo aprendido):

```
head(tratamiento$M5_501B)
```

```
## [1] 4 4 5 5 3 5
```

```
head(control$M5_501B)
```

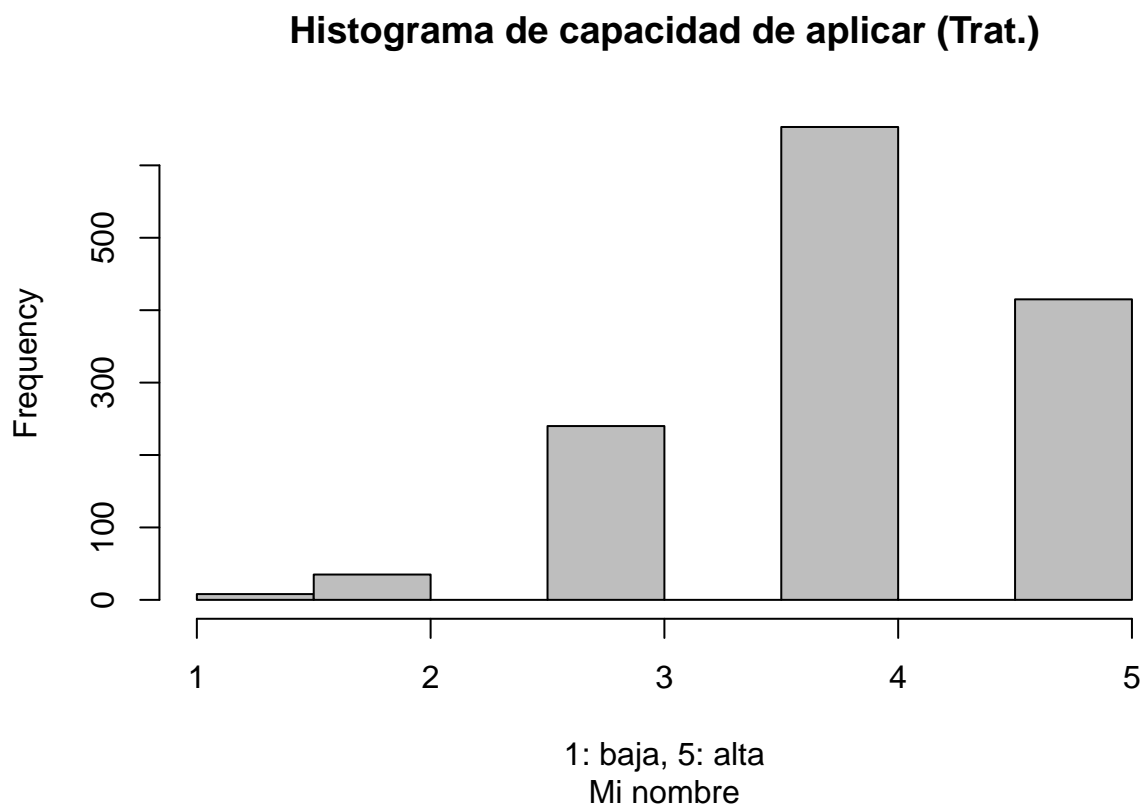
```
## [1] 3 5 4 5 4 5
```

Esta información no es tan útil todavía. Ahora vamos a hacer un histograma para mirar cómo se distribuyen los datos.

Nota: En este ejercicio usted debe poner su nombre como título de los gráficos. Para ello reemplace con su nombre en el argumento `sub="Mi nombre"`. **Si la gráfica no incluye su nombre no será calificada.**

Primero entre los jóvenes del grupo de tratamiento:

```
hist(tratamiento$M5_501B
     , col="gray"
     , main="Histograma de capacidad de aplicar (Trat.)"
     , xlab="1: baja, 5: alta"
     , sub="Mi nombre")
```

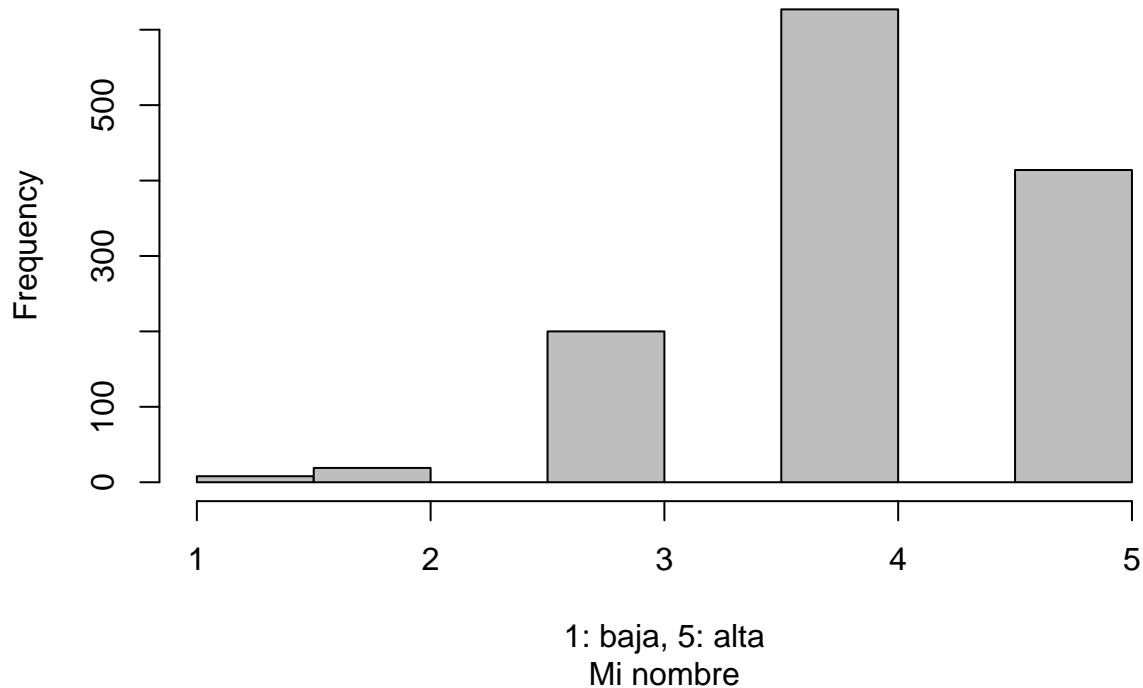


Ahora exporte el gráfico haciendo clic en **Export** en el panel de gráficos. Guárdelo en el computador. Al final de la clase deberá enviar estos gráficos al profesor como prueba de su participación en el taller.

Haremos lo mismo con el grupo de control:

```
hist(control$M5_501B
     , col="gray"
     , main="Histograma de capacidad de aplicar (Cont.)"
     , xlab="1: baja, 5: alta"
     , sub="Mi nombre")
```

Histograma de capacidad de aplicar (Cont.)



Exporte este gráfico también.

Ahora compare ambos histogramas haciendo clic en las flechas de navegación del panel de gráficos en RStudio.

Pregunta: ¿Hay una diferencia sustancial entre el grupo de tratamiento y el grupo de control?

Mirando el histograma no vemos gran diferencia, lo cual significa que en este componente la política no tuvo efectos detectables. Corroboremos esto hallando la media y la desviación estándar. En estos códigos reemplace las palabras *Mi Nombre* con su nombre después del signo #.

Las medias:

```
# Mi nombre  
mean(tratamiento$M5_501B)
```

```
## [1] 4.059956
```

```
mean(control$M5_501B)
```

```
## [1] 4.119874
```

Y las desviaciones estándar:

```
# Mi Nombre  
sd(tratamiento$M5_501B)
```

```
## [1] 0.7991401
```

```
sd(control$M5_501B)
```

```
## [1] 0.7661851
```

Tome un pantallazo del código (incluyendo los resultados) y guárdelo. Más tarde deberá enviarlo al profesor. Como la distribución es asimétrica podríamos buscar más bien la mediana y la desviación mediana absoluta. Las medianas:

```
# Mi Nombre  
median(tratamiento$M5_501B)
```

```
## [1] 4
```

```
median(control$M5_501B)
```

```
## [1] 4
```

Las desviaciones medianas absolutas:

```
# Mi Nombre  
mad(tratamiento$M5_501B)
```

```
## [1] 1.4826
```

```
mad(control$M5_501B)
```

```
## [1] 1.4826
```

Guarde también el pantallazo de este código.

Estos resultados son consistentes con el informe presentado por el DNP y Econometría, que nos dicen que el impacto en este componente fue nulo.

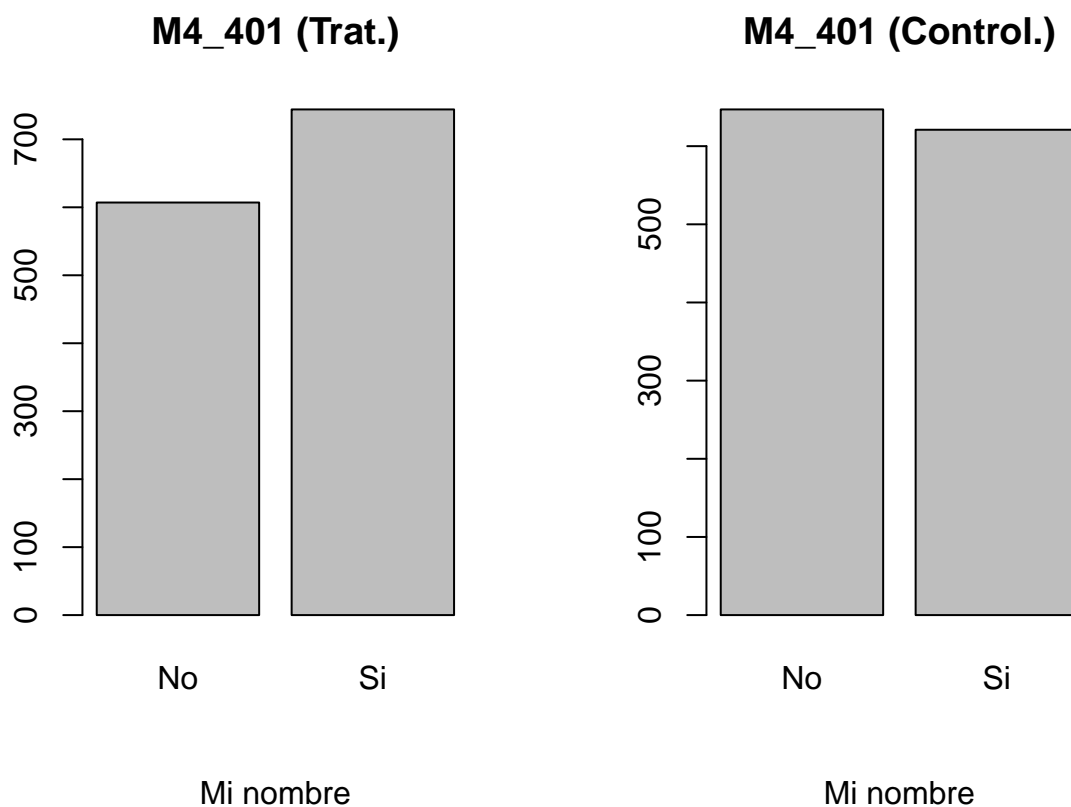
Segundo análisis

Según el informe final, hubo un impacto positivo en la continuidad educativa, es decir, el porcentaje de egresados que siguió estudiando. En otras palabras, los egresados del grupo de tratamiento tuvieron una mayor tendencia a seguir sus estudios. Confirmemos esta información.

La pregunta “¿En la actualidad estudia?” es la número M4_401.

Esta vez haremos diagramas de barras, ambos en la misma ventana (no olvide ingresar su nombre):

```
par(mfrow=c(1, 2))  
barplot(table(tratamiento$M4_401)  
        , main="M4_401 (Trat.)"  
        , sub="Mi nombre")  
barplot(table(control$M4_401)  
        , main="M4_401 (Control.)"  
        , sub="Mi nombre")
```



Vemos que comparativamente los del grupo de tratamiento tuvieron mayor propensión a seguir estudiando. Hay una diferencia entre ambos grupos, lo cual confirma la afirmación del informe final.