

Ejercicio N°1: Modelo causal y experimentos

SOL3063 - Inferencia Causal

Matías Deneken

10/04/2023

```
# Cargamos
data <- read.dta("base_ej23.dta")
table(data$condition, data$surveyid)
```

```
##
##      2016 2017
##    1   206  356
##    2   182  336
##    3   204  358
```

1. Formule y explique las hipótesis 1 (H1) del artículo (Puntaje: 1,0).

En el contexto del aumento continuo de los niveles de desigualdad económica en los Estados Unidos, el autor explora cómo la información proporcionada sobre la desigualdad y la prevalencia de los “Working Poor” puede influir en las preferencias y justificaciones de la redistribución económica. Es así que la **Hipótesis 1 planteé que: Aprender sobre la desigualdad en conjunto con información acerca del predominio de la clase obrera pobre (working poor- WP) deteriorará la confianza en el sueño Americano.** Aquí se entiende el sueño americano como un ideal meritocrático que afirma que el trabajo duro puede mejorar la situación social de una persona.

Esta hipótesis busca testear las creencias de la sociedad y legitimidad ideológica sobre la pobreza y la desigualdad económica. El autor no solo tiene en cuenta la información sobre el contexto de desigualdad en los Estados Unidos para los individuos de control, sino que también considera la información sobre el estatus laboral de los pobres (WP). El conocimiento del contexto de desigualdad y las características de los individuos pobres WP podría ayudar a desacreditar ciertas creencias y estereotipos negativos sobre la pobreza, como la idea de que es causada por la inacción y la falta de esfuerzo. Por lo tanto, se podría debilitar la creencia en el “sueño americano”, que sostiene que el trabajo duro y constante permite la movilidad social.

2. Explique el tratamiento, ¿qué se busca manipular? (Puntaje: 0,5).

En este estudio, se implementaron dos tratamientos y un grupo de control:

- El grupo de control recibió información general sobre las características poblacionales y sociodemográficas de los Estados Unidos.
- El primer tratamiento, llamado (1) “Inequality”, agregó información sobre la desigualdad económica en el país, la cual sirvió para caracterizar una clara distinción entre “haves” y “have-nots” (personas que tienen y que no tienen dinero, respectivamente).

- El segundo tratamiento, llamado (2) “InequalityWP”, añadió información adicional a lo anterior referida a que las personas que son pobres y que además poseen trabajo.

En consecuencia, “Inequality” se considera el tratamiento base o “baseline treatment” en comparación con el tratamiento que incluye elementos adicionales, en este caso “InequalityWP”.

El objetivo de este diseño experimental *es la manipulación intencional de la información que se proporciona a las personas encuestadas al momento de evaluar el ideal meritocrático estadounidense*. La información disponible puede variar desde no incluir datos sobre la desigualdad (grupo de control), hasta mencionarla de forma general (tratamiento base), o profundizar en ella con elementos adicionales, como la prevalencia de los trabajadores pobres (tratamiento adicional).

3. Examine el balance de las condiciones experimentales respecto de las covariables. Al respecto (Puntaje: 1,0):

El balance en inferencia causal es esencial para garantizar la validez y confiabilidad de los resultados experimentales en investigación. Proporcionar una descripción estadística del balance entre los grupos de tratamiento y control es una práctica efectiva para evaluar el efecto del tratamiento sobre la variable de interés, utilizando covariables disponibles. Un equilibrio adecuado en los grupos de tratamiento y control ayuda a evitar sesgos en los resultados y, en consecuencia, conclusiones erróneas. En lenguaje estadístico, el balance es fundamental para detectar cualquier posible correlación entre las covariables X y el tratamiento D y así evitar cualquier sesgo en los resultados experimentales.

Considerando lo anterior, los datos de la Tabla 1 indican que no hay diferencias significativas entre el grupo de tratamiento y control para las covariables `age` y `male`, pero sí para `income5`. Además, se utiliza un test conjunto para evaluar la correlación entre covariables y tratamiento, y se concluye que no hay evidencia estadística para asumir tal correlación en las variables `age`, `income5` y `male`. Es decir, es posible concluir que existe balance respecto de las covariables y que la asignación aleatoria ha funcionado.

```
## Balance

# covariables age, income5, and male

data_16 = data %>%
  filter(surveyid == 2016)

# estimation
balance1 <- lm(age ~ factor(condition), data = data_16)
balance2 <- lm(income5 ~ factor(condition), data = data_16)
balance3 <- lm(male ~ factor(condition), data = data_16)

labels <- c('Age', 'InequalityWP', 'Constant')
stargazer::stargazer(balance1, balance2, balance3,
  covariate.labels = labels,
  title = "Balance del experimento, Control vs Tratamientos",
  colnames = FALSE,
  model.numbers = FALSE,
  type = "latex",
  table.placement = "H",
  header=FALSE)
```

Table 1: Balance del experimento, Control vs Tratamientos

	<i>Dependent variable:</i>		
	age	income5	male
Age	0.569 (1.727)	-0.055 (0.151)	-0.063 (0.051)
InequalityWP	-1.840 (1.677)	-0.268* (0.145)	-0.044 (0.049)
Constant	48.777*** (1.183)	2.880*** (0.102)	0.519*** (0.035)
Observations	592	520	592
R ²	0.004	0.007	0.003
Adjusted R ²	0.0003	0.003	-0.001
Residual Std. Error	16.978 (df = 589)	1.383 (df = 517)	0.500 (df = 589)
F Statistic	1.084 (df = 2; 589)	1.890 (df = 2; 517)	0.829 (df = 2; 589)

Note:

*p<0.1; **p<0.05; ***p<0.01

4. Vamos a replicar algunos de los resultados reportados en Figura 1 del artículo que evalúan la hipótesis 1 (Puntaje: 1,5).

Los participantes que recibieron los tratamientos Inequality e InequalityWP en el Experimento 1 reportaron niveles más bajos de confianza en el sueño americano en comparación con el grupo de control, aunque los efectos no fueron significativos al 95% de confianza. En el Experimento 2, los efectos de los tratamientos fueron menores en comparación con el Experimento 1 pero tampoco fueron significativos al 95% de confianza. Al combinar ambos experimentos, se encontró que el efecto del tratamiento InequalityWP fue significativamente diferente del grupo de control y del grupo Inequality al 95% de confianza, proporcionando información de que tener un mayor conocimiento sobre la desigualdad y la clase trabajadora pobre disminuye en un -.040 la confianza en el sueño americano, respaldando la hipótesis 1.

Esto nos quiere decir que en los datos se observa un efecto mayor y en la misma dirección en el grupo con **tratamiento adicional** en comparación con el grupo con **tratamiento basal**.

En relación a lo anterior -atendiendo a los datos arrojado del experimento combinado- se puede concluir que la **hipótesis 1** del autor en el artículo es reafirmada por los datos presentados. De hecho, al aplicar un tratamiento que incorpora información sobre la desigualdad y la prevalencia de los **Working Poor**, se observa una disminución en la creencia en el ideal meritocrático siendo estadísticamente significativa al 95% de confianza.

```
# filter by surveyid == 2017. Dataset for Experiment 2
data_17 = data %>%
  filter(surveyid == 2017)

# Variable dependiente el indicador de apoyo a la meritocracia.
# Se estiman los 6 ATE.

# ----- ATE 2016
ate1 <-lm_robust(meritocracy_01~factor(condition), data=data_16)
#summary(ate1)
# ----- ATE 2017
```

```

ate2 <-lm_robust(meritocracy_01~factor(condition), data=data_17)
# ----- ATE Ambos
ate3 <-lm_robust(meritocracy_01~factor(condition), data=data)

texreg(list(ate1, ate2, ate3),
  custom.coef.names = c("Intercept", "Inequality", "InequalityWP"),
  custom.model.names = c("Exp. 2016", "Exp. 2017", "Exp. Combinado"),
  caption = "ATEs en distintos experimentos",
  digits = 3,
  include.ci = FALSE,
  float.pos = "h",
  caption.above = TRUE)

```

Table 2: ATEs en distintos experimentos

	Exp. 2016	Exp. 2017	Exp. Combinado
Intercept	0.593*** (0.019)	0.621*** (0.014)	0.611*** (0.011)
Inequality	-0.027 (0.029)	-0.003 (0.021)	-0.011 (0.017)
InequalityWP	-0.047 (0.026)	-0.036 (0.020)	-0.040* (0.016)
R ²	0.005	0.004	0.004
Adj. R ²	0.002	0.002	0.003
Num. obs.	591	1050	1641
RMSE	0.274	0.270	0.272

*** $p < 0.001$; ** $p < 0.01$; * $p < 0.05$

5. Evalúe los resultados en punto 4 incluyendo covariables (Puntaje: 1,0).

A partir de los datos de la Tabla 3, se puede concluir que la inclusión de covariables sociodemográficas no afecta de manera significativa las conclusiones obtenidas en los modelos previos sin covariables. Los efectos de los tratamientos se mantienen no significativos incluso añadiendo las covariables. Sin embargo, el efecto combinado del tratamiento InequalityWP sigue siendo significativo y tiene el mismo signo en comparación con la estimación sin covariables, lo que apoya **la hipótesis 1 (H1)**, referida a que la exposición a información sobre la desigualdad y la prevalencia de la clase trabajadora pobre (working poor - WP) conduce a una menor confianza en el sueño americano. La inclusión de covariables no reduce el error estándar del tratamiento, lo que sugiere que las covariables no están fuertemente correlacionadas con la variable dependiente.

En definitiva, la inclusión de covariables no modifica los resultados de la hipótesis 1 (H1).

```

ate4 <-lm_robust(meritocracy_01~factor(condition) +age+income5+male, data=data, subset=(surveyid==2016))
#summary(ate1)
# ----- ATE 2017
ate5 <-lm_robust(meritocracy_01~factor(condition) +age+income5+male, data=data, subset=(surveyid==2017))
# ----- ATE Ambos
ate6 <-lm_robust(meritocracy_01~factor(condition) +age+income5+male, data=data)

```

```

texreg(list(ate1, ate2, ate3, ate4, ate5, ate6),
  custom.header = list("Sin covariables" = 1:3, "Con covariables" = 4:6),
  custom.coef.names = c("Intercepto", "Inequality", "InequalityWP", "Age",
    "Income", "Male"),
  custom.model.names = c("Exp. 1", "Exp. 2", "Exp. Combinado", "Exp. 1", "Exp. 2", "Exp. Combinado"),
  caption = "ATEs en distintos experimentos con covariables",
  digits = 3,
  include.ci = FALSE,
  float.pos = "h",
  caption.above = TRUE)

```

Table 3: ATEs en distintos experimentos con covariables

	Sin covariables			Con covariables		
	Exp. 1	Exp. 2	Exp. Combinado	Exp. 1	Exp. 2	Exp. Combinado
Intercepto	0.593*** (0.019)	0.621*** (0.014)	0.611*** (0.011)	0.321*** (0.050)	0.392*** (0.030)	0.371*** (0.026)
Inequality	-0.027 (0.029)	-0.003 (0.021)	-0.011 (0.017)	-0.014 (0.029)	0.001 (0.020)	-0.003 (0.017)
InequalityWP	-0.047 (0.026)	-0.036 (0.020)	-0.040* (0.016)	-0.038 (0.027)	-0.031 (0.019)	-0.033* (0.016)
Age				0.004*** (0.001)	0.002*** (0.000)	0.002*** (0.000)
Income				0.030*** (0.008)	0.043*** (0.006)	0.040*** (0.005)
Male				0.023 (0.023)	0.032 (0.018)	0.028* (0.014)
R ²	0.005	0.004	0.004	0.077	0.085	0.079
Adj. R ²	0.002	0.002	0.003	0.068	0.081	0.076
Num. obs.	591	1050	1641	519	1050	1569
RMSE	0.274	0.270	0.272	0.263	0.259	0.261

*** $p < 0.001$; ** $p < 0.01$; * $p < 0.05$

6. La covariable de ingreso (income) tiene información perdida (missings). Al respecto, realice lo siguiente (Puntaje: 1,0).

```
data %>% frq (income5)
```

Recodificamos la variable tomando 1 los NA y el resto 0. En segundo paso, le pedimos a R que cree una nueva variable “income5_new” que tomé los mismos valores de income 5. A partir de duplicación, le pedimos a R que en income5_new se imputen los valores perdidos con el promedio de ingresos de todas las personas que pertenecen al estudio.

```
# (1) Recodificación de income a dummy
data$income_dummy=car::recode (data$income5, "1=0; 2=0; 3=0; 4=0; 5=0; NA=1")
# (2) Creación de nueva variable `income5_new`
data <- data %>% mutate(income5_new = case_when(TRUE ~ income5))
## Variable Income5_new toma el valor promedio.
data <- data %>% mutate_at(vars(income5_new), ~replace_na(.,mean(., na.rm = TRUE)))
```

Basándonos en los datos obtenidos al utilizar el método propuesto para tratar los valores faltantes, se puede afirmar que no hay un cambio sustancial en las conclusiones previas relacionadas con la Hipótesis 1.

Al comparar cada una de estas especificaciones, se puede observar que el coeficiente de la variable de interés “InequalityWP” mantiene prácticamente su magnitud, significancia y signo. Asimismo, el coeficiente de la variable “Dummy-income” no resulta significativo, lo que podría indicar que no hay una gran diferencia en términos estadísticos entre los casos perdidos y los no perdidos en el efecto sobre la variable dependiente “meritocracy_01” de un individuo promedio.

En conclusión, el ajuste por valores perdidos en la variable ingreso no tiene una incidencia significativa en las variables de tratamiento, lo que significa que no modifica los resultados obtenidos para la hipótesis 1 (H1) presentada en el punto 4. Es decir, seguimos aceptando que el tratamineto InequalityWP disminuye en 0.04 puntos la escala de creencia sobre el sueño americano.

Es importante señalar que la inclusión de variables de control no debió influir en el diseño de la investigación causal, dado que el diseño del experimento tuvo una adecuado asignación de los individuos participantes en los grupos de control y tratamiento.

```
## 1) ATEs sin covariable

ate7 <-lm_robust(meritocracy_01~factor(condition), data=data)
#summary(ate7)

## 2) ATEs, controlando por ingreso sin ajuste para missings

ate8 <-lm_robust(meritocracy_01~factor(condition) + income5 , data=data)
#summary(ate8)

# ATEs, controlando por ingreso y ajustado por missings

ate9 <-lm_robust(meritocracy_01~factor(condition) + income5_new + income_dummy, data=data)
#summary(ate9)

texreg(list(ate7, ate8, ate9),
        custom.coef.names = c("Intercepto", "Inequality", "Inequality (WP)",
                               "Income no ajustado" , "Ingreso ajustado", "Ingreso Dummy"),
```

```

custom.model.names = c("Sin covariable", "Income no ajustado", "Ingreso Dummy"),
caption = "ATEs con covariable de ingresos",
digits = 3,
include.ci = FALSE,
float.pos = "h",
caption.above = TRUE)

```

Table 4: ATEs con covariable de ingresos

	Sin covariable	Income no ajustado	Ingreso Dummy
Intercepto	0.611*** (0.011)	0.485*** (0.019)	0.485*** (0.018)
Inequality	-0.011 (0.017)	-0.005 (0.017)	-0.010 (0.016)
Inequality (WP)	-0.040* (0.016)	-0.037* (0.016)	-0.035* (0.016)
Income no ajustado		0.042*** (0.005)	
Ingreso ajustado			0.042*** (0.005)
Ingreso Dummy			-0.038 (0.035)
R ²	0.004	0.055	0.052
Adj. R ²	0.003	0.053	0.050
Num. obs.	1641	1569	1641
RMSE	0.272	0.264	0.266

*** $p < 0.001$; ** $p < 0.01$; * $p < 0.05$