

Entrega Final. Experimentos aleatorios

Carlos Bautista

2022-12-05

```
library(tidyverse)
library(dagitty)
library(rethinking)
library(stargazer)
library(kableExtra)
```

a. Artículo de Referencia

Bauer, Paul C, and Bernhard Clemm von Hohenberg. 2020. “Believing and Sharing Information by Fake Sources: An Experiment.”

b. Descripción General

En esta nueva era digital, uno de los mayores retos a los que la sociedad se enfrenta es la difusión de información o noticias falsas (fake news), cuyas causas y efectos en todos los ámbitos han sido poco estudiados.

Uno de los aspectos de este fenómeno que falta por entender mejor es por qué y cuándo la gente cree en dicha información a partir de la naturaleza de sus fuentes.

En este trabajo se realiza un experimento aleatorio para investigar el efecto de las fuentes de información (reales vs. falsas) y su contenido (congruentes vs. no congruentes con las actitudes de los individuos) sobre la posibilidad de creer en ellas e incluso compartirlas.

Los autores encuentran que las personas tienen una mayor tendencia a creer y una propensión un poco mayor a compartir noticias cuando se refieren a fuentes reales, pero además, el impacto de la congruencia entre los hechos informados y la visión del mundo de los individuos también es significativo.

Por ello, es mucho más probable que las personas creen en noticias de una fuente que les ha proporcionado previamente información congruente, que si la fuente proviene de una fuente falsa.

c. Análisis

La era digital ha facilitado aún mas la posibilidad de que cualquiera pueda acceder y compartir información, independientemente de su veracidad o la seriedad de la fuente. Además, lo más común es que las personas no sepan ex ante si una noticia es verdadera y, por lo general, pocos se toman el tiempo para verificar los hechos.

La pregunta principal del estudio es si las fuentes de información falsas se creen sustancialmente menos que las fuentes reales. El estudio contrasta esto exponiendo a individuos en un experimento aleatorio a noticias verdaderas y falsas pero buscando cuantificar los efectos de la veracidad de la fuente y la credibilidad del contenido.

Por ello, los autores han elegido las siguientes hipótesis de trabajo:

1. Las personas tienden a creer más en las noticias de una fuente real que de una fuente falsa.
2. Las personas serán más propensas a compartir informes de noticias que provienen de fuentes reales que aquellos que provienen de fuentes falsas.

3. Las personas que vieron noticias congruentes (versus incongruentes) de una fuente tienen una mayor tendencia a creer y compartir noticias posteriores de la misma fuente.
4. La diferencia en la creencia y en la propensión a compartir noticias posteriores entre aquellos que vieron informes congruentes y aquellos que vieron informes incongruentes es mayor para la fuente falsa que para la fuente real

El artículo de referencia muestra el procedimiento utilizado para verificar estas hipótesis, así como las principales conclusiones de un experimento realizado en Alemania. El anexo del documento (bastante extenso) muestra con detalle todos los elementos de dicho experimento.

d. Datos

El conjunto de datos es `fake.csv`. Dicho archivo contiene un subconjunto de los datos de Bauer y Clemm von Hohenberg (2020) que se utilizará para esta práctica. Los datos se basan en un experimento de encuesta en el que los individuos fueron asignados aleatoriamente a dos versiones de un informe de noticias en el que solo se varió el nombre de la fuente (y el diseño).

La mayoría de las variables tienen dos versiones donde una contiene valores numéricos (*num*) y la otra contiene etiquetas descriptivas (*fac*). Las variables de tratamiento generalmente se denominan `d...`, las variables de resultado `y...` y las covariables `x...`.

e. Instrucciones

1. Escriba una introducción del análisis que realizará con ayuda de este conjunto de datos de acuerdo al contexto del problema.

El presente ejercicio tiene el objetivo de evaluar el efecto causal que tienen las fuentes de información falsas y reales en que la gente lo crea (o no), y su disposición a compartirlas.

Para ello se utilizan datos de un experimento aleatorio conducido en Alemania por Bauer y Clemm (2015).

Se pretende, en primer lugar, validar la aleatorización del experimento y posteriormente medir el efecto causal ATE del tratamiento, según el diseño del experimento.

2. Ilustre correctamente los factores considerados en dicho análisis. Para ello, realice el gráfico directo acíclico (DAG) que muestre la teoría de cambio detrás de la relación causal que busca analizar y cuantificar el artículo.

Se genera DAG para la variable resultado `belief`, dado que la estructura será la misma para los diferentes variables resultado relacionadas con la disposición a compartir (email, facebook, twitter, whatsapp) que se evaluarán en las secciones siguientes.

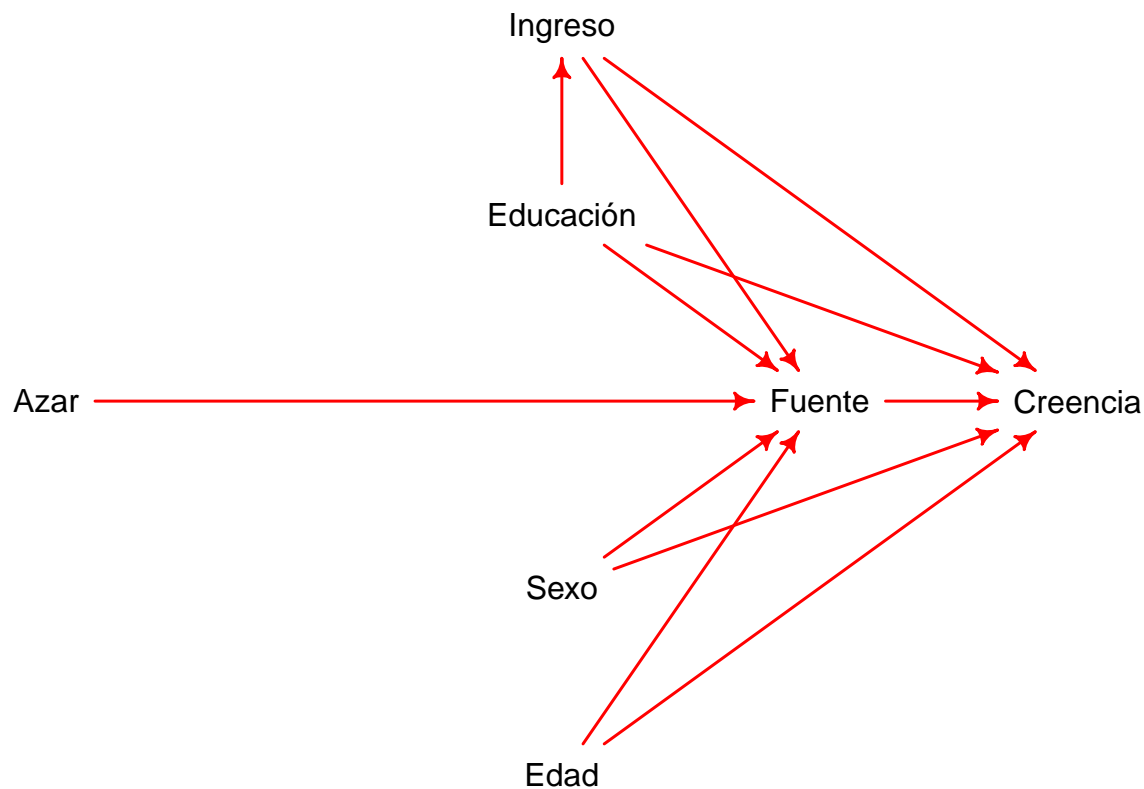
```
dag_rct <- dagitty('dag{
  "Fuente" [pos="0.5,0"]
  "Creencia" [pos="1,0"]
  "Azar" [pos="-1,0"]
  "Edad" [pos="0,1"]
  "Sexo" [pos="0,0.5"]
  "Ingreso" [pos="0,-1"]
  "Educación" [pos="0,-0.5"]
  "Fuente" -> "Creencia"
  "Azar" -> "Fuente"
  "Edad" -> "Fuente"
  "Edad" -> "Creencia"
  "Sexo" -> "Fuente"
  "Sexo" -> "Creencia"
  "Ingreso" -> "Fuente"
  "Ingreso" -> "Creencia"
})
```

```

    "Educación" -> "Fuente"
    "Educación" -> "Creencia"
    "Educación" -> "Ingreso"}')

drawdag(dag_rct, col_arrow="red")

```



3. Describa de manera general el diseño del experimento aleatorizado: participantes, variables de respuesta, clasificación de grupos de tratamiento y control y covariables. Mencione cómo se implementó el experimento en la práctica.

Los datos del experimento se obtuvieron a través de una encuesta en línea realizada entre el 14 y 29 de marzo de 2019 a ciudadanos alemanes mayores de 18 años. Se obtuvieron un total de 1980 respuestas, excluyendo a quienes no completaron todo el cuestionario.

El experimento consistió en exponer a los participantes a una lista de 11 nombres de medios de información que podían ser reales o inventados, asignados de manera aleatoria. Posteriormente se les preguntó si conocían la fuente (medios) y partir de esto se les aplicó una serie de preguntas sobre que tanto confiaban en dichas fuentes. Además, se complementó la información con preguntas sobre si contaban con redes sociales (facebook, twitter, whatsapp) y su propensión a compartir los medios y noticias que observaron a través de dichas redes sociales.

Asimismo, se incluyeron una serie de covariables relacionadas con las características del individuo (edad, sexo, ingresos, nivel educativo, etc.) y sobre su postura ante temas como migración y política.

4. Luego de leer el conjunto de datos, empiece por realizar el análisis exploratorio de las variables y describa su comportamiento. Tome en cuenta que hay un grupo de variables numéricas y otras categóricas que deben ser descritas de manera particular de acuerdo a su naturaleza (ver apéndice 2).

Resumen estadístico de variables numéricas:

```
dta <- read_csv("data/fake.csv")
attach(dta)
```

```
dta %>% select_if(is.numeric) %>% summary()
```

```
## y_belief_report_num y_share_report_email_num y_share_report_fb_num
## Min. :0.000 Min. :0.00000 Min. :0.0000
## 1st Qu.:3.000 1st Qu.:0.00000 1st Qu.:0.0000
## Median :4.000 Median :0.00000 Median :0.0000
## Mean :3.458 Mean :0.05492 Mean :0.1099
## 3rd Qu.:5.000 3rd Qu.:0.00000 3rd Qu.:0.0000
## Max. :6.000 Max. :1.00000 Max. :1.0000
## NA's :1 NA's :50 NA's :770
## y_share_report_twitter_num y_share_report_whatsapp_num d_treatment_source
## Min. :0.0000 Min. :0.0000 Min. :0.0000
## 1st Qu.:0.0000 1st Qu.:0.0000 1st Qu.:0.0000
## Median :0.0000 Median :0.0000 Median :0.0000
## Mean :0.1267 Mean :0.0866 Mean :0.4894
## 3rd Qu.:0.0000 3rd Qu.:0.0000 3rd Qu.:1.0000
## Max. :1.0000 Max. :1.0000 Max. :1.0000
## NA's :1688 NA's :341
## x_age x_sex_num x_income_num x_education_num
## Min. :18.00 Min. :0.0000 Min. : 1.000 Min. :1.000
## 1st Qu.:33.00 1st Qu.:0.0000 1st Qu.: 4.000 1st Qu.:2.000
## Median :48.00 Median :0.0000 Median : 5.000 Median :3.000
## Mean :46.53 Mean :0.4934 Mean : 5.715 Mean :3.088
## 3rd Qu.:59.00 3rd Qu.:1.0000 3rd Qu.: 8.000 3rd Qu.:4.000
## Max. :89.00 Max. :1.0000 Max. :11.000 Max. :5.000
## NA's :16 NA's :18
```

Exploración de variables categóricas:

```
t1 <- table(dta$y_belief_report_fac)
t2 <- table(dta$y_share_report_email_fac)
t3 <- table(dta$y_share_report_fb_fac)
t4 <- table(dta$y_share_report_twitter_fac)
t5 <- table(dta$y_share_report_whatsapp_fac)
t6 <- table(dta$d_treatment_source)
t7 <- table(dta$x_sex_fac)
t8 <- table(dta$x_income_fac)
t9 <- table(dta$x_education_fac)
```

```
print(c("creencia", t1))
```

```
##          0 - Not at all          1          2          3
## "creencia"      "119"      "111"      "237"      "515"
##          4          5 6 - Completely
##      "447"      "374"      "176"
```

```
print(c("email", t2))
```

```
##          No          Yes
## "email"  "1824"  "106"
```

```
print(c("facebook", t3))
```

```
##           No      Yes
## "facebook" "1077" "133"
```

```
print(c("twitter", t4))
```

```
##           No      Yes
## "twitter"  "255"  "37"
```

```
print(c("whatsapp", t5))
```

```
##           No      Yes
## "whatsapp" "1497" "142"
```

```
print(c("tratamiento", t6))
```

```
##           0      1
## "tratamiento" "1011" "969"
```

```
print(c("sexo", t7))
```

```
##      female  male
## "sexo" "1003" "977"
```

```
print(c("ingresos", t8))
```

```
##           0 to 500€ 1,001 to 1,500€ 1,501 to 2,000€ 2,001 to 2,500€
## "ingresos"      "72"      "228"      "244"      "292"
## 2,501 to 3,000€ 3,001 to 3,500€ 3,501 to 4,000€ 4,001 to 4,500€ 4,501 to 5,000€
## "249"      "213"      "165"      "108"      "92"
## 5,001€ or more 501 to 1,000€
## "137"      "164"
```

```
print(c("educación", t9))
```

```
##           Elementary school      High school
## "educación"      "257"      "384"
## Middle school Technical high school      University
## "673"      "160"      "488"
```

5. Realice los cuadros de balance para verificar la aleatorización del experimento. Replique, en la medida de lo posible los cuadros del apéndice 3. Interprete sus resultados.

La tabla de balance para las covariables de interés, que compara las medias según su pertenencia a grupo de tratamiento o control, muestran que sí existe un equilibrio. No obstante, hay relativamente mayores diferencias en las variables de edad y sexo.

Para asegurarnos del balance podemos hacer las correspondientes pruebas T, que se muestran en el siguiente apartado

```
dta_cov <- c('x_age', 'x_sex_num', 'x_income_num', 'x_education_num')
dta %>% group_by(d_treatment_source) %>%
  select(one_of(dta_cov)) %>%
  summarise_all(funs(mean(., na.rm = TRUE))) %>%
  kable()
```

d_treatment_source	x_age	x_sex_num	x_income_num	x_education_num
0	47.15134	0.5143422	5.738308	3.088911
1	45.88958	0.4716202	5.690302	3.087409

T-test para covariables:

Con base en los valores p de la prueba no podemos rechazar la hipótesis nula de que la diferencia en las medias entre grupos (tratamiento y control) es igual a cero.

```
lapply(dta[, dta_cov],
  function(v) {t.test(v ~ dta$d_treatment_source, var.equal=TRUE)})

## $x_age
##
## Two Sample t-test
##
## data: v by dta$d_treatment_source
## t = 1.7471, df = 1978, p-value = 0.08078
## alternative hypothesis: true difference in means between group 0 and group 1 is not equal to 0
## 95 percent confidence interval:
## -0.1546242 2.6781411
## sample estimates:
## mean in group 0 mean in group 1
## 47.15134 45.88958
##
##
## $x_sex_num
##
## Two Sample t-test
##
## data: v by dta$d_treatment_source
## t = 1.9015, df = 1978, p-value = 0.05738
## alternative hypothesis: true difference in means between group 0 and group 1 is not equal to 0
## 95 percent confidence interval:
## -0.001340019 0.086784036
## sample estimates:
## mean in group 0 mean in group 1
## 0.5143422 0.4716202
##
##
## $x_income_num
##
## Two Sample t-test
##
## data: v by dta$d_treatment_source
## t = 0.39357, df = 1962, p-value = 0.6939
## alternative hypothesis: true difference in means between group 0 and group 1 is not equal to 0
## 95 percent confidence interval:
```

```
## -0.1912127 0.2872248
## sample estimates:
## mean in group 0 mean in group 1
##      5.738308      5.690302
##
##
## $x_education_num
##
## Two Sample t-test
##
## data: v by dta$d_treatment_source
## t = 0.02322, df = 1960, p-value = 0.9815
## alternative hypothesis: true difference in means between group 0 and group 1 is not equal to 0
## 95 percent confidence interval:
## -0.1253700 0.1283743
## sample estimates:
## mean in group 0 mean in group 1
##      3.088911      3.087409
```

6. Dado que los supuestos SUTVA e independencia son válidos, compare el resultado promedio en el tratamiento y el control mediante pruebas simples t. Para ello, en R puede utilizar la función `t.test()`.

Evaluamos el resultado promedio en los grupos tratamiento y control, considerando las diferentes variables de respuesta “y”.

Para el caso de la creencia, el efecto del tratamiento es significativo, con un valor p pequeño (prácticamente cero) que nos permite rechazar la hipótesis de que no hay diferencia en el valor promedio del grupo de tratamiento y control.

En el caso de compartir las noticias por facebook, el efecto del tratamiento también es significativo.

Si elegimos un umbral de significancia de 95%, con whatsapp también es significativo el efecto. Por su parte, el efecto del tratamiento en compartir con email o twitter no es significativo.

```
dta_resp <- c('y_belief_report_num', 'y_share_report_email_num',
             'y_share_report_fb_num', 'y_share_report_twitter_num',
             'y_share_report_whatsapp_num')

lapply(dta[, dta_resp],
       function(v) {t.test(v ~ dta$d_treatment_source, var.equal = TRUE)})
```

```
## $y_belief_report_num
##
## Two Sample t-test
##
## data: v by dta$d_treatment_source
## t = -9.021, df = 1977, p-value < 2.2e-16
## alternative hypothesis: true difference in means between group 0 and group 1 is not equal to 0
## 95 percent confidence interval:
## -0.7591045 -0.4879874
## sample estimates:
## mean in group 0 mean in group 1
##      3.153314      3.776860
##
##
```

```

## $y_share_report_email_num
##
## Two Sample t-test
##
## data: v by dta$d_treatment_source
## t = -1.8301, df = 1928, p-value = 0.06739
## alternative hypothesis: true difference in means between group 0 and group 1 is not equal to 0
## 95 percent confidence interval:
## -0.039318778 0.001359381
## sample estimates:
## mean in group 0 mean in group 1
## 0.04563895 0.06461864
##
##
## $y_share_report_fb_num
##
## Two Sample t-test
##
## data: v by dta$d_treatment_source
## t = -2.8157, df = 1208, p-value = 0.004947
## alternative hypothesis: true difference in means between group 0 and group 1 is not equal to 0
## 95 percent confidence interval:
## -0.08571086 -0.01531613
## sample estimates:
## mean in group 0 mean in group 1
## 0.08457711 0.13509061
##
##
## $y_share_report_twitter_num
##
## Two Sample t-test
##
## data: v by dta$d_treatment_source
## t = -1.0964, df = 290, p-value = 0.2738
## alternative hypothesis: true difference in means between group 0 and group 1 is not equal to 0
## 95 percent confidence interval:
## -0.11950563 0.03399383
## sample estimates:
## mean in group 0 mean in group 1
## 0.1048951 0.1476510
##
##
## $y_share_report_whatsapp_num
##
## Two Sample t-test
##
## data: v by dta$d_treatment_source
## t = -2.0702, df = 1637, p-value = 0.03859
## alternative hypothesis: true difference in means between group 0 and group 1 is not equal to 0
## 95 percent confidence interval:
## -0.056004710 -0.001510636
## sample estimates:
## mean in group 0 mean in group 1
## 0.07261905 0.10137672

```


7. Muestre gráficamente los efectos que tiene la fuente en la posibilidad de creer en las noticias y en la propensión a compartirlas. Esto es, replique los gráficos de la figura 3 (pág. 13). Interprete estos resultados y concluya en el contexto del problema de acuerdo a las hipótesis planteadas.

Podemos visualizar los resultados anteriores y verificamos los resultados. En las variables que son significativas los intervalos de confianza no se cruzan (rechazamos hipótesis nula), mientras que en las que no resultaron significativas sí se cruzan (no podemos rechazar la H0).

Nota. En el caso de twitter, observamos que el número de datos disminuye considerablemente (290 observaciones), por lo que puede estar afectando los resultados y la gráfica.

```
#belief
dta_graf_belief <- dta %>%
  group_by(d_treatment_source) %>%
  select(y_belief_report_num) %>%
  group_by(n_tot = n(), add = TRUE) %>%
  summarize_all(funs(mean, var, sd, n_na = sum(is.na(.))), na.rm = TRUE) %>%
  mutate(n = n_tot - n_na) %>%
  mutate(se = sqrt(var/n),
         ci.error = qt(0.975,df=n-1)*sd/sqrt(n),
         conf.low = mean - ci.error,
         conf.high = mean + ci.error) %>%
  mutate(grupo = ifelse(d_treatment_source == 1, "Tratamiento", "Control")) %>%
  mutate(respuesta = "Belief") %>%
  mutate(respuesta = str_to_title(respuesta))

g_belief <- ggplot(dta_graf_belief, aes(x=respuesta,
                                       y=mean,
                                       group=as.factor(d_treatment_source),
                                       color=as.factor(d_treatment_source))) +
  geom_point(position = position_dodge(0.4), size = 1.1) +
  geom_errorbar(aes(ymin = conf.low, ymax = conf.high), width=.3,
               position = position_dodge(0.4), size = 0.7) +
  labs(title = "Belief") +
  scale_color_manual(values=c('red','green'),
                    labels = c("Fake", "Real"),
                    name = "Tratamiento:") +
  theme_light()

# email
dta_graf_email <- dta %>%
  group_by(d_treatment_source) %>%
  select(y_share_report_email_num) %>%
  group_by(n_tot = n(), add = TRUE) %>%
  summarize_all(funs(mean, var, sd, n_na = sum(is.na(.))), na.rm = TRUE) %>%
  mutate(n = n_tot - n_na) %>%
  mutate(se = sqrt(var/n),
         ci.error = qt(0.975,df=n-1)*sd/sqrt(n),
         conf.low = mean - ci.error,
         conf.high = mean + ci.error) %>%
  mutate(grupo = ifelse(d_treatment_source == 1, "Tratamiento", "Control")) %>%
  mutate(respuesta = "Email") %>%
  mutate(respuesta = str_to_title(respuesta))
```

```

g_email <- ggplot(dta_graf_email, aes(x=respuesta,
                                     y=mean,
                                     group=as.factor(d_treatment_source),
                                     color=as.factor(d_treatment_source))) +
  geom_point(position = position_dodge(0.4), size = 1.1) +
  geom_errorbar(aes(ymin = conf.low, ymax = conf.high), width=.3,
               position = position_dodge(0.4), size = 0.7) +
  labs(title = "Email") +
  scale_color_manual(values=c('red','green'),
                    labels = c("Fake", "Real"),
                    name = "Tratamiento:") +
  theme_light()

# fb
dta_graf_fb <- dta %>%
  group_by(d_treatment_source) %>%
  select(y_share_report_fb_num) %>%
  group_by(n_tot = n(), add = TRUE) %>%
  summarize_all(funs(mean, var, sd, n_na = sum(is.na(.))), na.rm = TRUE) %>%
  mutate(n = n_tot - n_na) %>%
  mutate(se = sqrt(var/n),
         ci.error = qt(0.975,df=n-1)*sd/sqrt(n),
         conf.low = mean - ci.error,
         conf.high = mean + ci.error) %>%
  mutate(grupo = ifelse(d_treatment_source == 1, "Tratamiento", "Control")) %>%
  mutate(respuesta = "Facebook") %>%
  mutate(respuesta = str_to_title(respuesta))

g_fb <- ggplot(dta_graf_fb, aes(x=respuesta,
                                y=mean,
                                group=as.factor(d_treatment_source),
                                color=as.factor(d_treatment_source))) +
  geom_point(position = position_dodge(0.4), size = 1.1) +
  geom_errorbar(aes(ymin = conf.low, ymax = conf.high), width=.3,
               position = position_dodge(0.4), size = 0.7) +
  labs(title = "Facebook") +
  scale_color_manual(values=c('red','green'),
                    labels = c("Fake", "Real"),
                    name = "Tratamiento:") +
  theme_light()

# twitter
dta_graf_tw <- dta %>%
  group_by(d_treatment_source) %>%
  select(y_share_report_twitter_num) %>%
  group_by(n_tot = n(), add = TRUE) %>%
  summarize_all(funs(mean, var, sd, n_na = sum(is.na(.))), na.rm = TRUE) %>%
  mutate(n = n_tot - n_na) %>%
  mutate(se = sqrt(var/n),
         ci.error = qt(0.975,df=n-1)*sd/sqrt(n),
         conf.low = mean - ci.error,
         conf.high = mean + ci.error) %>%

```

```

mutate(grupo = ifelse(d_treatment_source == 1, "Tratamiento", "Control")) %>%
mutate(respuesta = "Twitter") %>%
mutate(respuesta = str_to_title(respuesta))

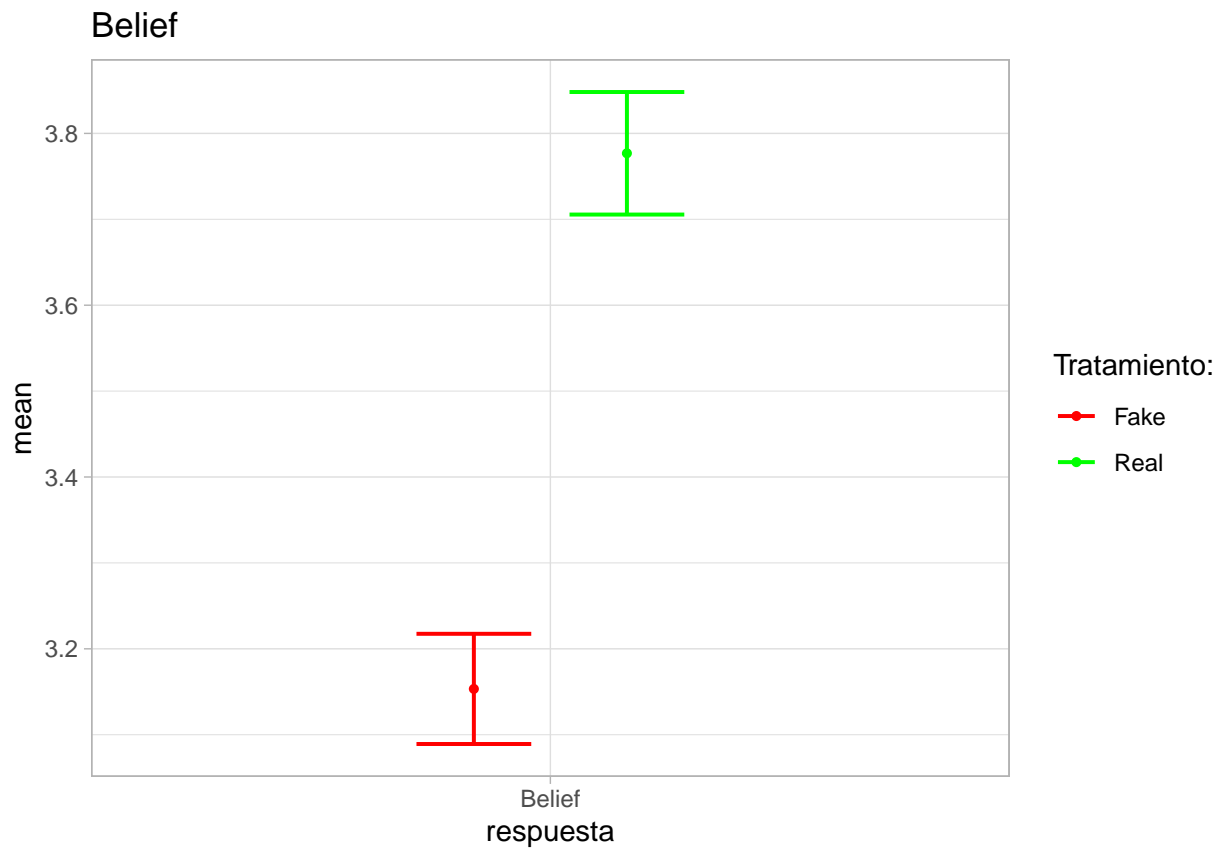
g_tw <- ggplot(dta_graf_tw, aes(x=respuesta,
                               y=mean,
                               group=as.factor(d_treatment_source),
                               color=as.factor(d_treatment_source))) +
  geom_point(position = position_dodge(0.4), size = 1.1) +
  geom_errorbar(aes(ymin = conf.low, ymax = conf.high), width=.3,
                position = position_dodge(0.4), size = 0.7) +
  labs(title = "Twitter") +
  scale_color_manual(values=c('red','green'),
                    labels = c("Fake", "Real"),
                    name = "Tratamiento:") +
  theme_light()

# whatsapp
dta_graf_wa <- dta %>%
  group_by(d_treatment_source) %>%
  select(y_share_report_whatsapp_num) %>%
  group_by(n_tot = n(), add = TRUE) %>%
  summarize_all(funs(mean, var, sd, n_na = sum(is.na(.))), na.rm = TRUE) %>%
  mutate(n = n_tot - n_na) %>%
  mutate(se = sqrt(var/n),
         ci.error = qt(0.975,df=n-1)*sd/sqrt(n),
         conf.low = mean - ci.error,
         conf.high = mean + ci.error) %>%
  mutate(grupo = ifelse(d_treatment_source == 1, "Tratamiento", "Control")) %>%
  mutate(respuesta = "Whatsapp") %>%
  mutate(respuesta = str_to_title(respuesta))

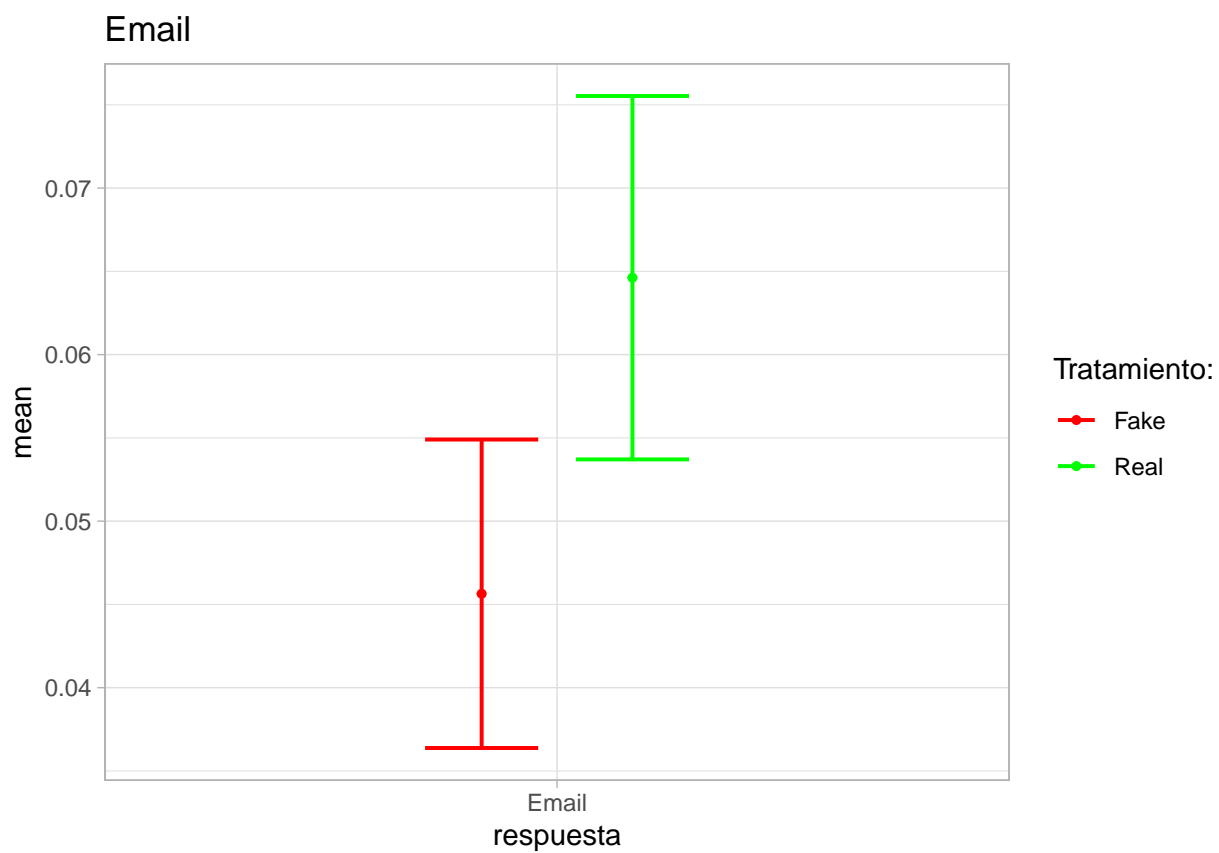
g_wa <- ggplot(dta_graf_wa, aes(x=respuesta,
                                y=mean,
                                group=as.factor(d_treatment_source),
                                color=as.factor(d_treatment_source))) +
  geom_point(position = position_dodge(0.4), size = 1.1) +
  geom_errorbar(aes(ymin = conf.low, ymax = conf.high), width=.3,
                position = position_dodge(0.4), size = 0.7) +
  labs(title = "Whatsapp") +
  scale_color_manual(values=c('red','green'),
                    labels = c("Fake", "Real"),
                    name = "Tratamiento:") +
  theme_light()

g_belief

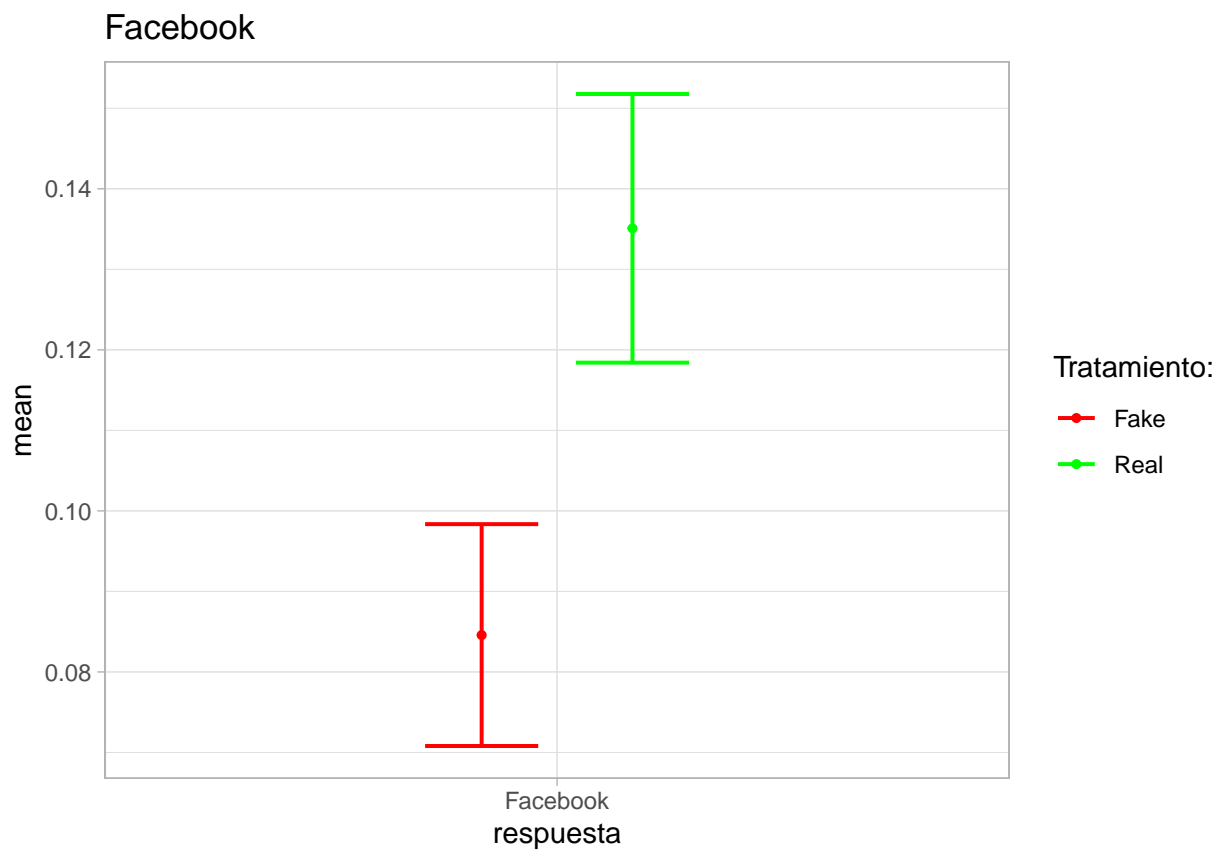
```



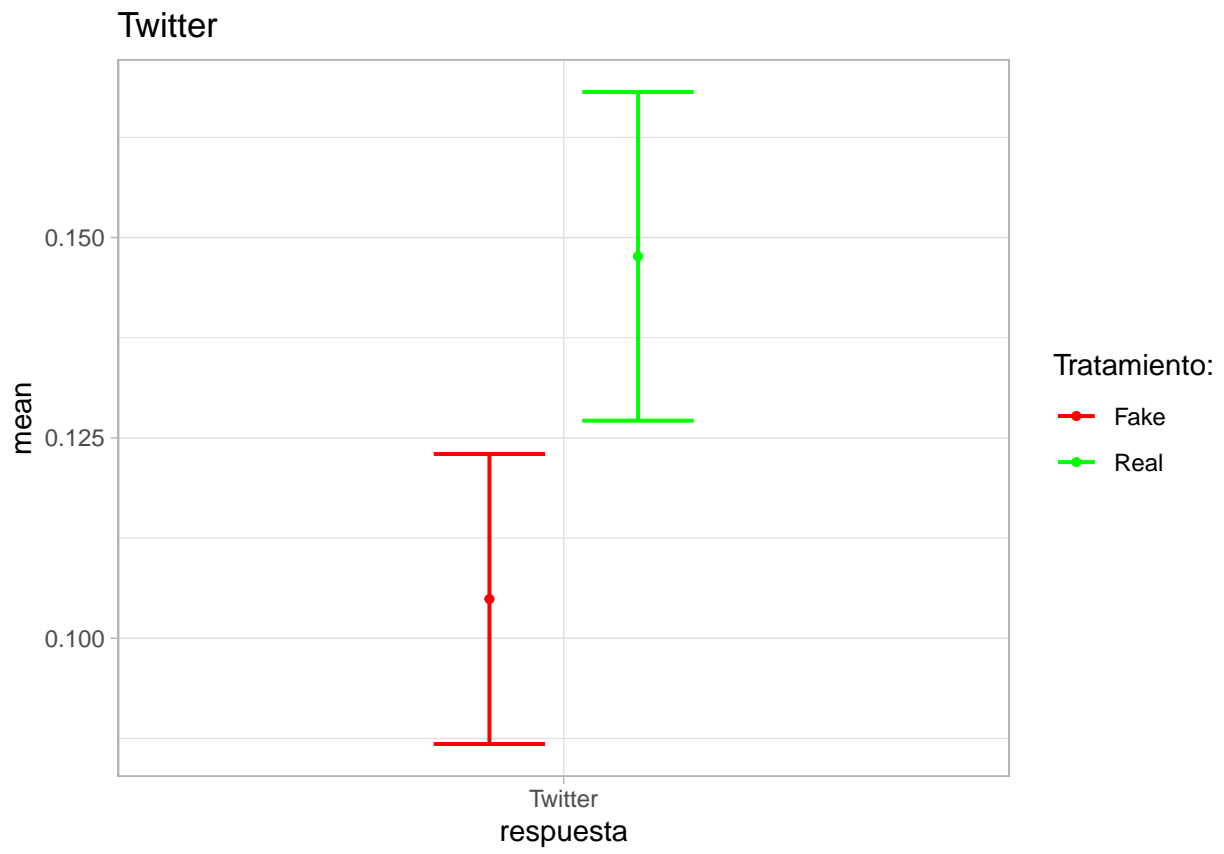
g_email



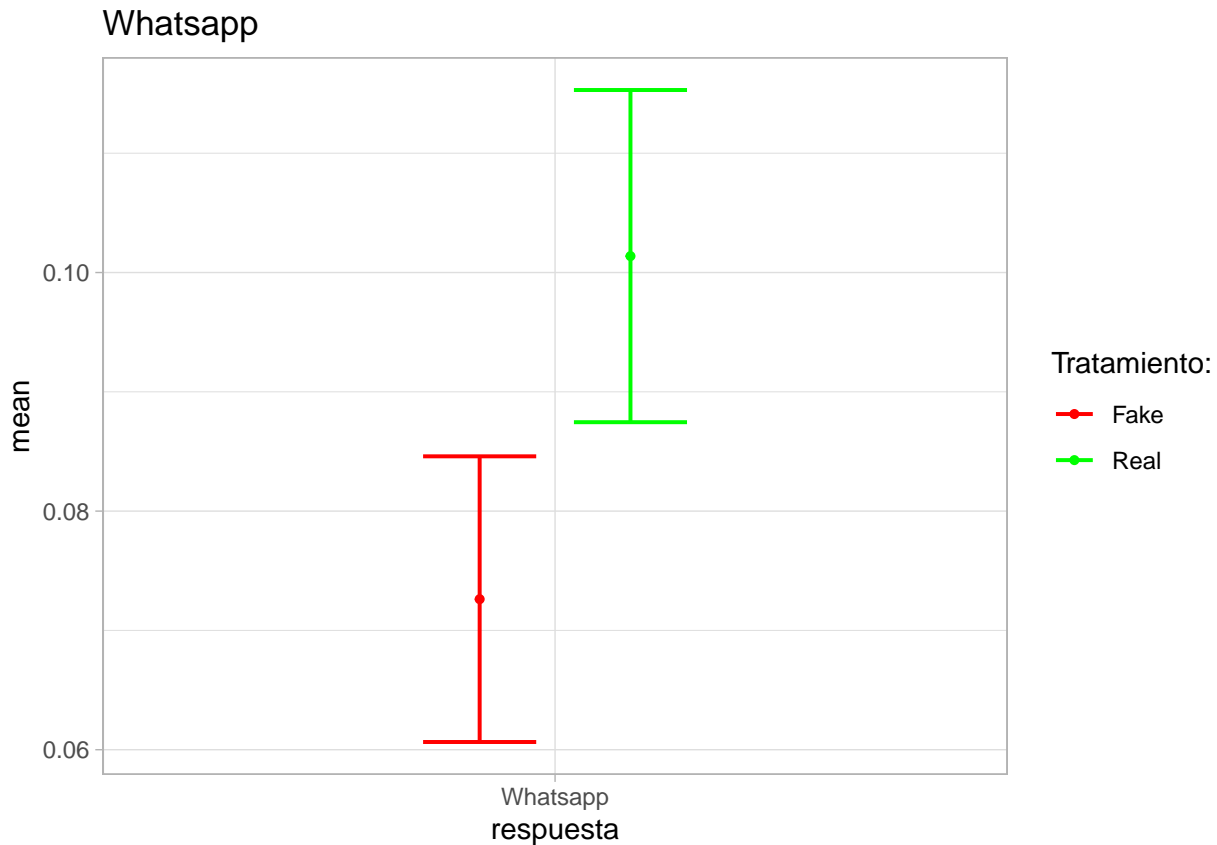
g_fb



g_tw



g_wa



8. Ahora obtenga estos mismos resultados pero implementando sus cálculos mediante la estimación adecuada de mínimos cuadrados ordinarios. Incluya el análisis para la fuente, la congruencia de la misma y la posibilidad de la interacción. Esto es, replique los resultados de las tablas del anexo 9. ¿Qué conclusión obtiene de ellas? Interprete sus resultados en el contexto del problema de acuerdo a las hipótesis del estudio. No olvide comparar mediante pruebas t los efectos para cada una de las redes sociales.

Al estimar mediante MCO obtenemos los resultados equivalentes, para las variables de resultado de creencia, facebook y whatsapp el efecto del tratamiento es significativo al 95%, mientras que para email y twitter no lo es a dicho nivel de significancia.

Nota. No encontré (o entendí del todo) la variable de congruencia, por lo que no pude completar el modelo con la interacción.

```
mod_belief <- lm(y_belief_report_num ~ d_treatment_source, data = dta)
mod_email <- lm(y_share_report_email_num ~ d_treatment_source, data = dta)
mod_fb <- lm(y_share_report_fb_num ~ d_treatment_source, data = dta)
mod_tw <- lm(y_share_report_twitter_num ~ d_treatment_source, data = dta)
mod_wa <- lm(y_share_report_whatsapp_num ~ d_treatment_source, data = dta)

stargazer(mod_belief, mod_email, mod_fb, mod_tw, mod_wa,
           type = "text")
```

```
##
## =====
##                                     Dependent variable:
## -----
##          y_belief_report_num    y_share_report_email_num    y_share_report_fb_num    y_share_report_twitter_num
##                (1)                (2)                (3)                (4)
```



```
## -----
## d_treatment_source      0.624***          0.019*          0.051***
##                        (0.069)          (0.010)          (0.018)
##
## Constant                3.153***          0.046***          0.085***
##                        (0.048)          (0.007)          (0.013)
##
## -----
## Observations            1,979            1,930            1,210
## R2                      0.040            0.002            0.007
## Adjusted R2             0.039            0.001            0.006
## Residual Std. Error    1.537 (df = 1977)    0.228 (df = 1928)    0.312 (df = 1208)
## F Statistic            81.379*** (df = 1; 1977)  3.349* (df = 1; 1928)  7.928*** (df = 1; 1208)
## =====
## Note:
```

9. Hay dos covariables más aún no utilizadas (educación e ingreso). Realice algún análisis para identificar si tiene alguna significancia el incluirlas como controles.

Se agregan como controles las covariables de educación e ingreso y se calculan nuevamente los 5 modelos en función de la variable resultado.

Para el modelo base, cuya variable de respuesta es la creencia, el coeficiente sí es significativo al 95%, y el del ingreso no.

En el resto de modelos, que miden la propensión a compartir las noticias mediante diversos canales y redes sociales, el efecto de educación e ingreso no es significativo a dicho nivel de 95%.

```
mod_belief_2 <- lm(y_belief_report_num ~ d_treatment_source + x_education_num +
  x_income_num, data = dta)
mod_email_2 <- lm(y_share_report_email_num ~ d_treatment_source + x_education_num +
  x_income_num, data = dta)
mod_fb_2 <- lm(y_share_report_fb_num ~ d_treatment_source + x_education_num +
  x_income_num, data = dta)
mod_tw_2 <- lm(y_share_report_twitter_num ~ d_treatment_source + x_education_num +
  x_income_num, data = dta)
mod_wa_2 <- lm(y_share_report_whatsapp_num ~ d_treatment_source + x_education_num +
  x_income_num, data = dta)

stargazer(mod_belief_2, mod_email_2, mod_fb_2, mod_tw_2, mod_wa_2,
  type = "text")
```

```
##
## =====
##                                     Dependent variable:
## -----
##                y_belief_report_num    y_share_report_email_num y_share_report_fb_num y_share_
##                (1)                (2)                (3)
## -----
## d_treatment_source      0.621***          0.018*          0.049***
##                        (0.068)          (0.010)          (0.018)
##
## x_education_num         0.193***          -0.006*          -0.002
##                        (0.024)          (0.004)          (0.006)
##
```

```

## x_income_num          0.021          0.003          -0.0001
##                   (0.013)          (0.002)          (0.003)
##
## Constant              2.450***          0.050***          0.090***
##                   (0.105)          (0.016)          (0.028)
##
## -----
## Observations          1,945          1,901          1,196
## R2                    0.075          0.004          0.006
## Adjusted R2           0.073          0.002          0.004
## Residual Std. Error    1.505 (df = 1941)    0.228 (df = 1897)    0.311 (df = 1192)    0.3
## F Statistic           52.146*** (df = 3; 1941)  2.309* (df = 3; 1897)  2.512* (df = 3; 1192)  1.024
## =====
## Note:

```

10. Escriba la conclusión general del análisis realizado. Interprete la cuantificación del efecto causal y discuta la utilidad y ventajas de aplicar la estrategia empírica seleccionada para la cuantificación del efecto causal.

El análisis mostró que sí existe un efecto causal en la creencia en noticias dependiendo de la naturaleza de la fuente de información (real o falsa). Además, al agregar controles de educación e ingreso se observa que solo el ingreso tiene un efecto significativo. Es decir, el nivel de educación también tiene un efecto en creer las noticias (reales o falsas) a partir de la fuente.

El procedimiento para poder evaluar este fenómeno de manera consistente y no sesgada fue a través del diseño y la implementación de un experimento aleatorio donde la asignación del tratamiento es justamente aleatorio y se compara sus resultados con un grupo de control (que no recibió el tratamiento).

Esta es una técnica muy poderosa ya que mediante la aleatorización del experimento se asegura que se cumplan los supuestos SUTVA e independencia. En ese sentido, para proceder es importante verificar que los datos cumplan con esos supuestos mediante tablas de balance y pruebas t que aseguren que no existen diferencias significativas entre los grupos de control y tratamiento.

Una vez validado lo anterior, la ventaja de utilizar este método es que la medición del efecto causal (ATE) es relativamente sencilla, ya que consiste en medir las diferencias promedio en la variable de resultado para ambos grupos y probar que efectivamente dicha diferencia sea significativa.