

Taller de Econometría Aplicada: práctica Final

Anna Alejandra Cortes Ordoñez

2023-04-22

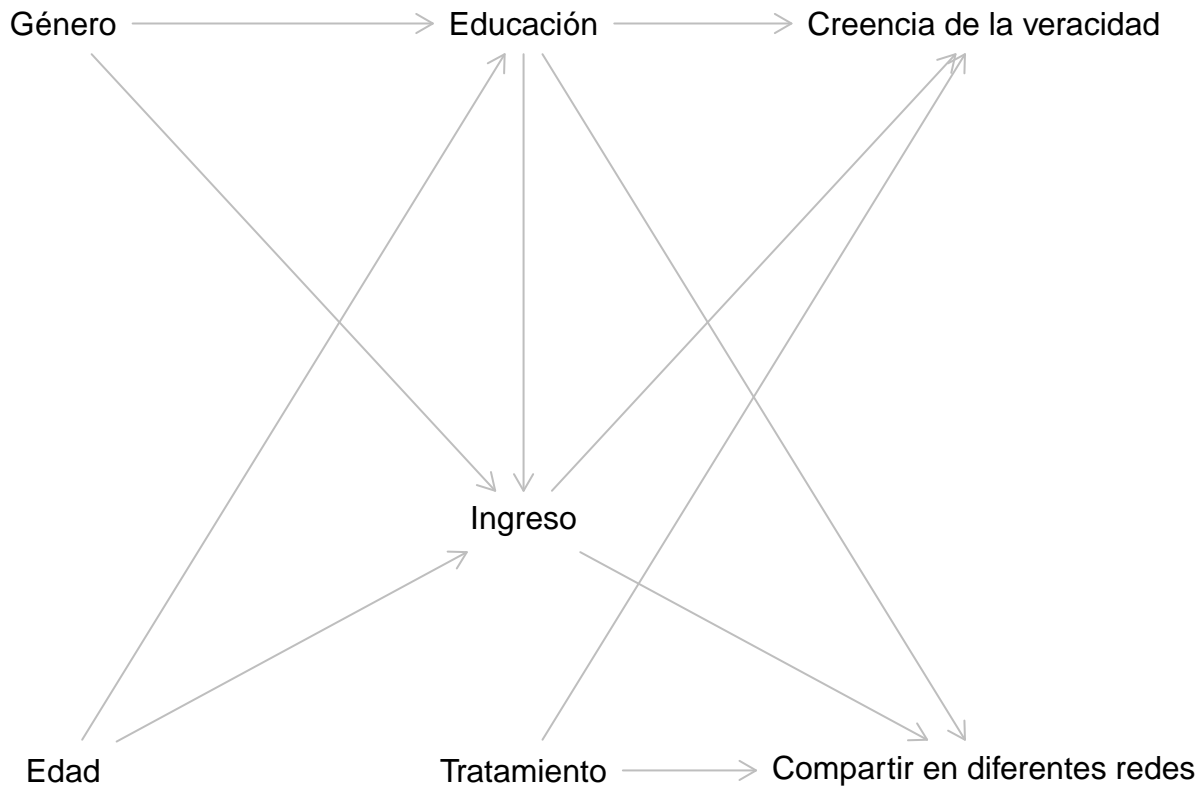
1. Escriba una introducción del análisis que realizará con ayuda de este conjunto de datos de acuerdo al contexto del problema.

A partir de finales del siglo XVII, los diarios se convirtieron en el principal medio para difundir la actividad de los periodistas. En el siglo XX surgieron la radio y la televisión. En la actualidad, el Internet se ha convertido en otro medio importante para la difusión de información. Con el paso de los años se busca que el número de noticias a disposición de los ciudadanos aumente, junto con la velocidad de transmisión de las mismas. Con ello, a pesar de que se ha hecho más eficaz la difusión de las noticias y que cualquier persona puede escribir un texto y publicarlo han surgido retos frente a un mayor número de noticias falsas *fake news*. Con ayuda de un subconjunto de datos obtenido de *Believing and Sharing Information by Fake Sources: An Experiment* de Bauer y Clemm von Hohenberg (2020) basados en un experimento de una 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) se pretende determinar si las fuentes de información falsas se creen sustancialmente menos que las fuentes reales. Así, con ayuda de las variables:

- `d_treatment_source`: reporta si la fuente es falsa o no
- `y_belief_report`: mide si se cree en la veracidad del informe en una escala del 0 al 6.
- `y_share_report`: mide si alguien compartiría el reporte en distintas plataformas (email, facebook, twitter y whatsapp), dummy que toma valores 0 y 1
- `x_age`: edad
- `x_sex`: género
- `x_income`: nivel de ingreso en una escala del 1 al 11
- `x_education`: nivel de educación en una escala del 0 al 5

Dadas las variables se pretende caracterizar a la población y comprender si las personas tienden a creer más en noticias reales o falsas, si comparten más noticias de fuentes reales o falsas, si 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; por último, 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 verdadera. Con la finalidad de comprender el comportamiento de las personas ante *fake news* y su posible propagación a través de medios digitales, pues en la actualidad a parte de poder redactar una noticia falsa también es posible crear imágenes falsas utilizando inteligencia artificial.

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.



Con base en las variables otorgadas realice el gráfico directo acíclico para mostrar la teoría de cambio detrás de la relación causal que se busca analizar. Específicamente, la relación causal entre la edad y el ingreso, ya que dependiendo de la edad la persona puede tener más años de experiencia laboral y tener un mejor ingreso. Así mismo, entre edad y educación, ya que a mayor edad habrá más oportunidades para que una persona haya concluido más años de estudio. Por otro lado, el género resulta un efecto causal del ingreso, ya que en el mercado laboral puede existir mayores/menores oportunidades según el género de la persona y con un ingreso mayor o menor con base en ello. Con base en la ONU las causas de la brecha salarial (la diferencia entre el salario de los hombres y las mujeres, dividida entre el salario de los hombres) son: empleo a tiempo parcial, peores empleos, trabajos socialmente menos valorados, autominusvaloración e ignorancia de los empleadores. También se encuentra la relación entre género y educación, pues en muchos casos las mujeres tienen menos oportunidades para estudiar.

Con ello sumado el tratamiento es que el ingreso y la educación tienen un efecto causal sobre las variables de interés “creencia de la veracidad” y “compartir la noticia en distintas redes”, pues frente a un mayor ingreso y educación la persona podría tener bases más sólidas y permitir cuestionarse sobre la veracidad de noticias que lee en internet.

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 para probar las hipótesis de trabajo planteadas

El experimento realizado por Paul C. Bauer (MZES) & Bernhard Clemm von Hohenberg (EUI) en *Believing and Sharing Information by Fake Sources: An Experiment* toma en cuenta una encuesta en línea realizada entre el 14 y 29 de marzo de 2019, con preregistro el 12 de marzo de 2019. Los **participantes** eran alemanes reclutados por la empresa de encuestas Respondi AG. Con el fin de cumplir con validez externa (la capacidad de generalizar los resultados del experimento en una población distinta a la estudiada) se seleccionaron encuestados según género, edad y estado de residencia. Se excluyeron a 125 encuestados que no habían completado el cuestionario. También, se excluyeron 308 respuestas de aquellos que completaron el cuestionario en menos de la mitad del tiempo promedio. Así la muestra final incluyó 1980 participantes.

A lo largo de los años diversos estudios psicológicos se han preguntado cómo se establece la transmisión con base en la fuente por la cual es publicada y cómo afectan la opinión de los lectores; sin embargo, todos estos estudios fueron realizados antes de que las fake news fueran un tema tan relevante como lo es actualmente. Así, Bauer y Clemm von Hohenberg continuaron con esta idea asumiendo que las personas actúan de manera subjetiva respecto a las noticias de acuerdo a la fuente y así dictaminan el outcome de creer en la noticia y sus intenciones de compartirla. Con base en lo anterior, escogieron sus **variables de respuesta** al considerar importante el impacto de las fake news en la era digital, pues las personas no saben ex ante si la noticia es verdadera o falsa y no tienen tiempo para revisarlo; con ello, la fuente toma importancia. Así, la pregunta es si las fuentes de información falsas son menos creíbles para el público que las verdaderas. Adaptaron una gran medida de distintos medios para compartir noticias dentro de sus variables de respuesta.

Para ello, la **clasificación de grupos de tratamiento y control** se dio posterior a las preguntas que respondieron los participantes respecto al consentimiento, socio demográficas, conocimiento en fuentes de noticias, sentimiento respecto a inmigrantes, uso de redes sociales, disposición a compartir información en redes y conocimiento político. De manera aleatoria se asignaron los participantes a 4 grupos de tratamiento, en los cuales debían leer 5 reportes con variedad de tipo de noticias asignadas de manera aleatoria entre participantes con el fin de que recibieran 5 reportes cada uno de alguna de las dos fuentes (original y falsa).

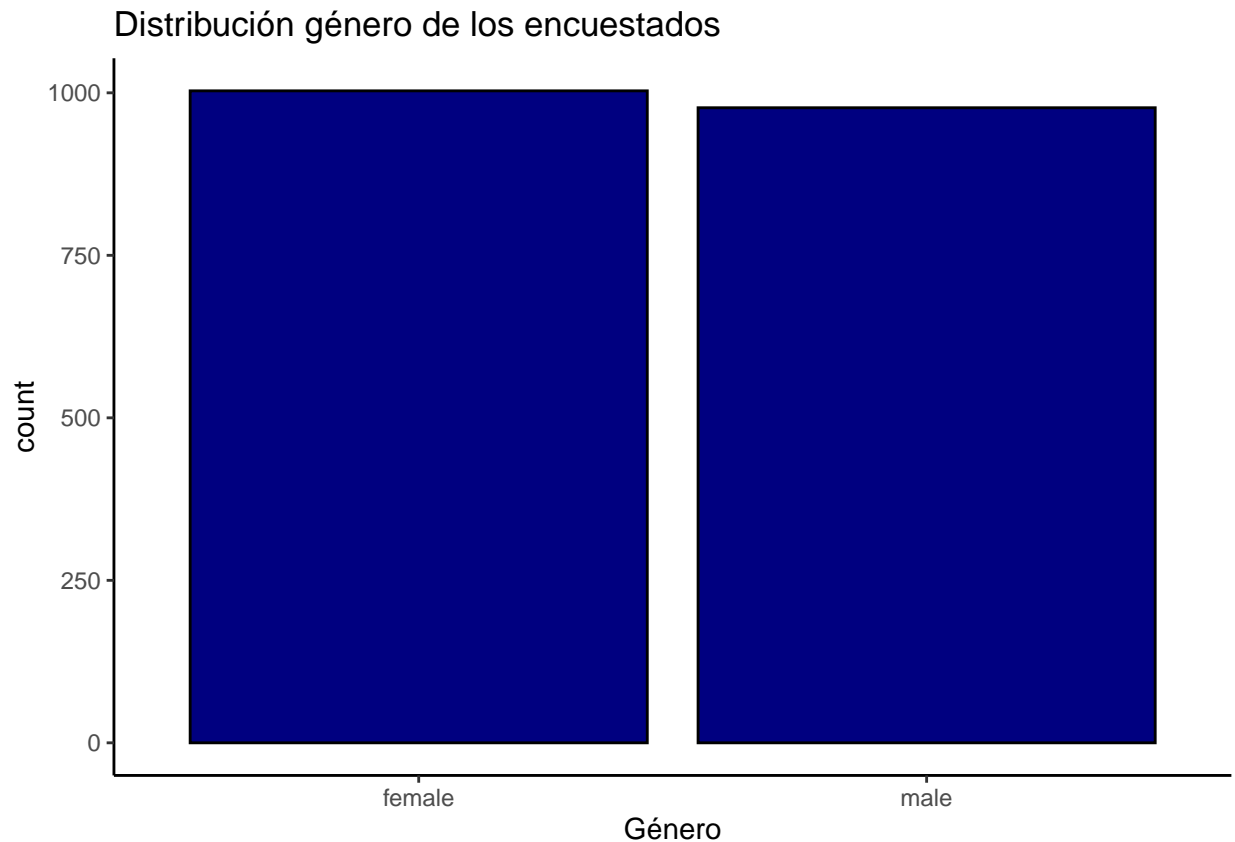
Para construir la fuente real utilizaron un nombre que asemeja la apariencia de Tagesschau, la sección de noticias más grande que se publica en alemán y para la fuente falsa se creó el nombre Nachrichten 360, ya que suena neutral y no tiene ninguna tendencia hacia alguna ideología y un logo neutral para que pareciera una fuente creíble. De los 5 reportes: el primero era verdadero e idéntico para todos los participantes. Los reportes 2, 3 y 4 fueron asignados aleatoriamente sobre a favor de la inmigración y anti inmigración (a dos grupos les tocaban anti y a dos a favor). Así, una versión sería conveniente para los que están a favor y la otra para aquellos que no. El quinto reporte era información verdadera e igual para todos los participantes. De los cuatro grupos: dos recibían información de la fuente Tagesschau (real) y dos de Nachrichten 360 (falsa).

La lista de **covariables** incluye: edad, género, en cuál estado viven, si conocen ciertas fuentes de información, si han leído de alguna de esas fuentes, si confían en la información de las fuentes, opinión respecto a inmigración y qué implicación tiene sobre economía, seguridad, vida y mayor control fronterizo. A su vez, se preguntó qué plataformas utilizan, frecuencia con la que comparten información, conocimiento político, "overconfidence", nivel de educación, si votaron en las elecciones de 2017, por quién votaron y su ingreso neto.

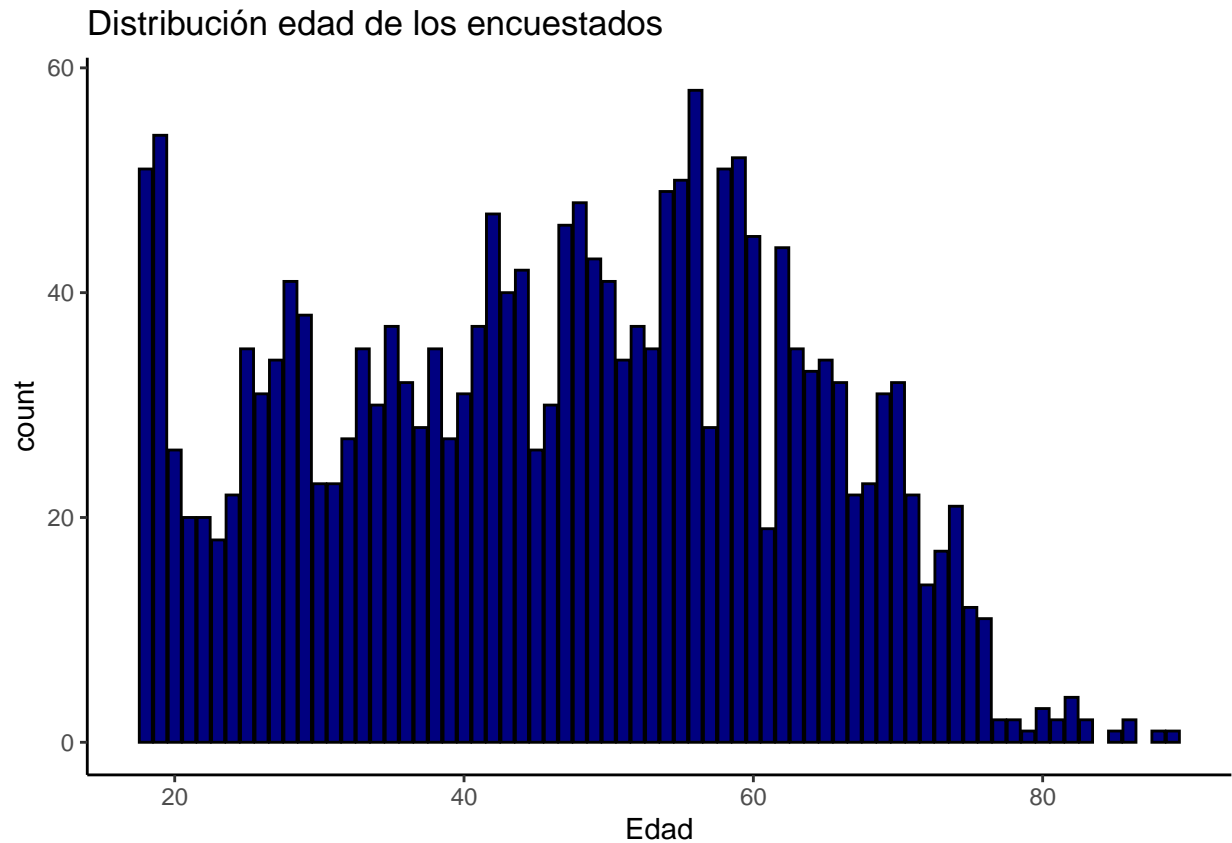
Como se mencionó anteriormente los participantes debían leer 5 reportes, el 1 y 5 eran verdaderos para todos los participantes y del 2 al 4 fueron manipulados entre información a favor o en contra de la inmigración. Así **el experimento se implementó** en la división en tratamiento que se expuso en la clasificación y los reportes trataban de temas diversos acerca de inmigración. El primer reporte era acerca de el número de reuniones familiares entre refugiados en Alemania, el segundo reporte (manipulado) era acerca de estadísticas de crímenes por parte de inmigrantes, el tercer reporte (manipulado) tenía datos de los resultados de refugiados en exámenes de lenguaje, el cuarto reporte (manipulado) tenía datos de rescates marítimos en el mediterráneo y el quinto reporte (verdadero) era de subsidios a los inmigrantes. Para los reportes 1 y 5 fueron tomados de fuentes poco compartidas para evitar que muchos de los encuestados hayan leído dichos reportes.

4. Luego de leer el conjunto de datos, realice un 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)

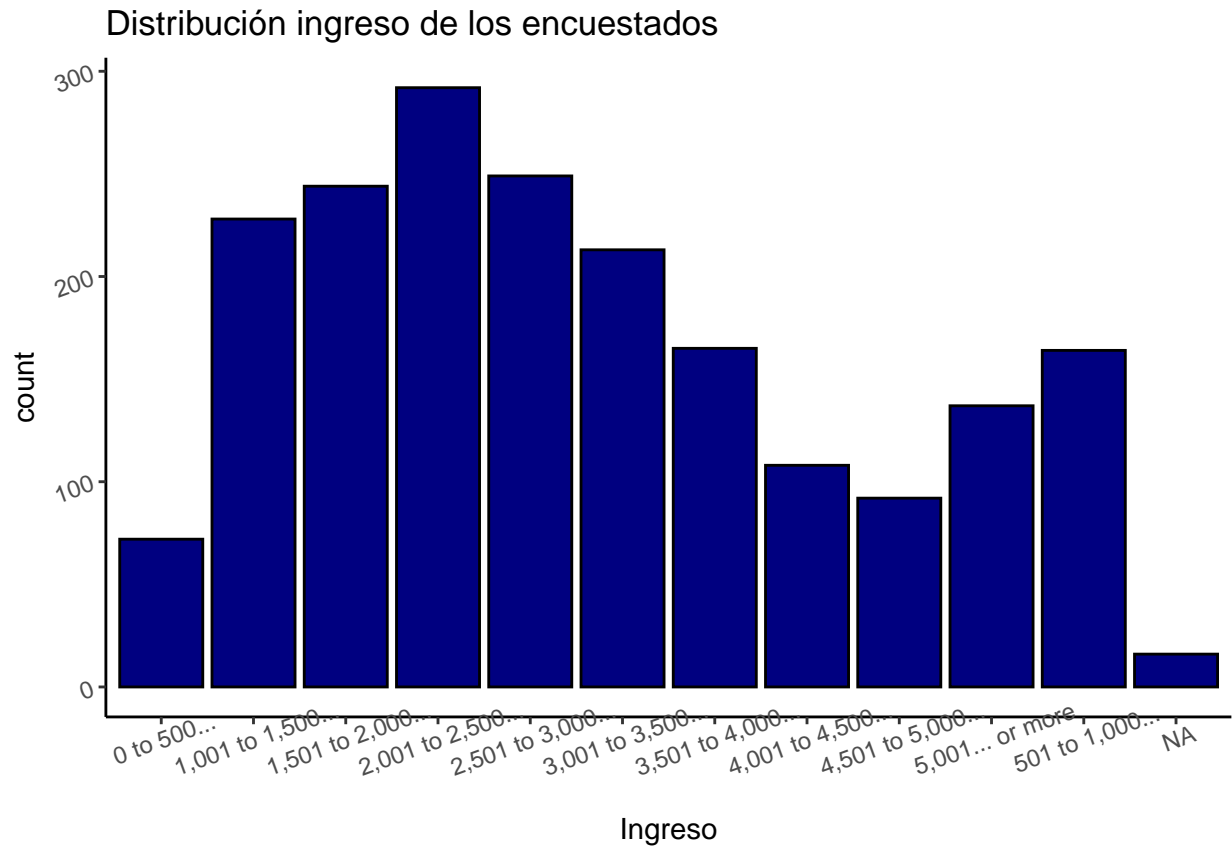
Con base en el conjunto de datos en la siguiente gráfica observamos que la distribución entre hombres y mujeres es similar. Así, es proporcional la cantidad de hombres y mujeres. Con datos del último censo de población en Alemania, solo hay 2 millones más de mujeres que hombres (total de 83.2 millones de habitantes), del mismo modo hay un número mayor de mujeres que hombres en el conjunto de datos pero no es desproporcional.



Posteriormente, observamos como se comporta la edad de los encuestados. Así observamos una mayor concentración por debajo de los 20 años, con un promedio de 46.53 (47 años) y una desviación estándar que cubre de los 31 a los 60 años. Es importante resaltar que el promedio de edad en Alemania es de 35 años. Este se encuentra a menos de una desviación estándar del promedio del conjunto de datos que tenemos.

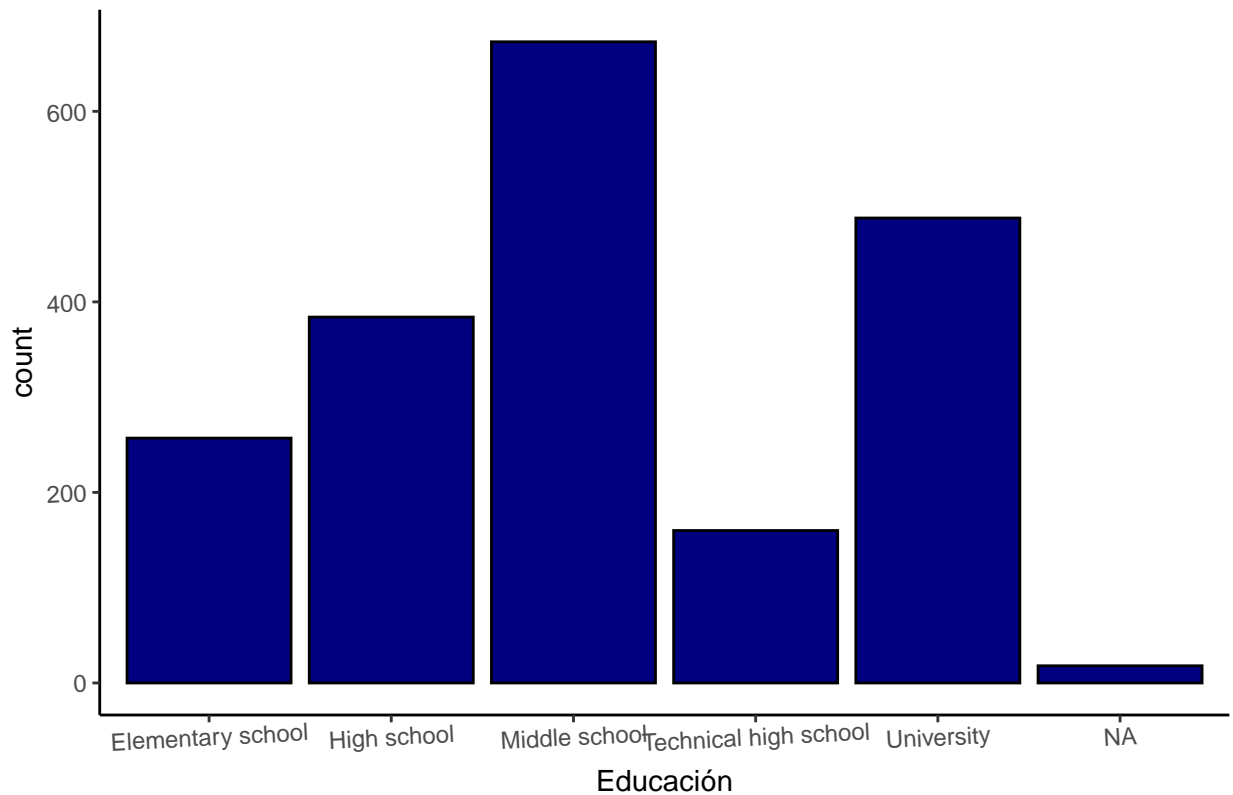


Con base en los datos se observa que una mayor cantidad de población tiene un ingreso entre 2,000 y 2,500 euros. En este caso es importante destacar que como se observa en la gráfica que una menor cantidad de los encuestados gana al rededor de 4,587 euros que es el promedio del ingreso en Alemania. Con base en el artículo y de acuerdo con otros estudios el rol de la educación y el ingreso esta correlacionado con la habilidad de distinguir entre sitios de internet profesionales y no profesionales.



De acuerdo con datos de la OCDE “el 86% de los adultos de 25 a 64 años han terminado la educación media superior”. Así, con esta base de datos se observa que la mayoría completó la educación media, seguido de la universidad que va de acuerdo con el promedio de la población alemana.

Distribución educación de los encuestados



Con base en la siguiente tabla de las variables numéricas los valores que destacan son que en promedio las personas comparten más noticias a través de twitter seguido de facebook, whatsapp y correo. A su vez, al realizar un comparativo con la población alemana, el promedio de edad está a menos de una desviación estandar de los resultados obtenidos.

	vars	n	mean	sd	median	trimmed	mad	min	max
y_belief_report_num	1	1979	3.46	1.57	4	3.54	1.48	0	6
y_share_report_email_num	2	1930	0.05	0.23	0	0.00	0.00	0	1
y_share_report_fb_num	3	1210	0.11	0.31	0	0.01	0.00	0	1
y_share_report_twitter_num	4	292	0.13	0.33	0	0.03	0.00	0	1
y_share_report_whatsapp_num	5	1639	0.09	0.28	0	0.00	0.00	0	1
x_age	6	1980	46.53	16.07	48	46.61	17.79	18	89
x_sex_num	7	1980	0.49	0.50	0	0.49	0.00	0	1
x_income_num	8	1964	5.71	2.70	5	5.60	2.97	1	11
x_education_num	9	1962	3.09	1.43	3	3.11	1.48	1	5
	range	skew	kurtosis	se					
y_belief_report_num	6	-0.40	-0.35	0.04					
y_share_report_email_num	1	3.90	13.25	0.01					
y_share_report_fb_num	1	2.49	4.21	0.01					
y_share_report_twitter_num	1	2.23	3.00	0.02					
y_share_report_whatsapp_num	1	2.94	6.63	0.01					
x_age	71	-0.05	-0.94	0.36					
x_sex_num	1	0.03	-2.00	0.01					
x_income_num	10	0.31	-0.73	0.06					
x_education_num	4	0.09	-1.47	0.03					

Con base en los siguientes datos de variables categóricas, en el caso de la educación se observa que en un mayor número de encuestados terminó “middle school” seguid de “high school”, la menor proporción de individuos fueron los que únicamente obtuvieron “technical high school”.

```
# A tibble: 6 x 2
# Groups:   x_education_fac [6]
  x_education_fac      n
  <chr>           <int>
1 Elementary school    257
2 High school          384
3 Middle school        673
4 Technical high school 160
5 University           488
6 <NA>                 18
```

Ahora respecto al ingreso, observamos una mayor cantidad de encuestados entre 2,000 y 2,500 euros y una menor proporción debajo de los mil euros y entre 4,500 y 5,000 euros mensuales.

```
# A tibble: 12 x 2
# Groups:   x_income_fac [12]
  x_income_fac      n
  <chr>           <int>
1 0 to 500€        72
2 1,001 to 1,500€  228
3 1,501 to 2,000€  244
4 2,001 to 2,500€  292
5 2,501 to 3,000€  249
6 3,001 to 3,500€  213
7 3,501 to 4,000€  165
8 4,001 to 4,500€  108
9 4,501 to 5,000€   92
10 5,001€ or more  137
11 501 to 1,000€   164
12 <NA>            16
```

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

La tabla de balance de la media y desviación estándar de la edad entre los grupos de tratamiento que recibieron la noticia falsa Nachrichten y la verdadera Tagesschau observamos que al comparar la edad entre ambos grupos no es muy distinta y la desviación estándar tampoco. Al igual que como se ha mencionado anteriormente, es similar a la edad promedio de la población alemana.

```
##           Grupo      Media      SD
## 1 Nachrichten Edad 47.15134 16.03041
## 2 Tagesschau Edad 45.88958 16.10032
```

Con base en el ingreso observamos mayor balance entre ambos grupos tanto del promedio como de la desviación estándar.

```
##           Grupo      Media      SD
## 1 Nachrichten Género 0.5143422 0.5000416
## 2 Tagesschau Género 0.4716202 0.4994517
```


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()`.

En la escala del 0-6 si la persona cree en el reporte con base en la siguiente prueba simple t observamos que el valor p es menor que el nivel de significancia. Así, podemos rechazar la hipótesis nula y concluir que no hay evidencia suficiente que demuestre que la diferencia del promedio de la creencia entre el reporte real y falso es igual a cero. Esto quiere decir que es distinto el promedio de creencia para ambos grupos (recordemos que un grupo recibió una noticia real y el otro una noticia falsa)

Two Sample t-test

```
data: y_belief_report_num by 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
```

Con base en la siguiente prueba simple t observamos que el valor-p es mayor pero muy cercano al nivel de significancia. Así, al aumentar el nivel de significancia podemos rechazar la hipótesis nula y concluir que no hay evidencia suficiente que demuestre que la diferencia del promedio de compartir el reporte via email entre el falso y verdadero es igual a cero. Con ello, hay promedios distintos entre el grupo que recibió la noticia real y el grupo que tuvo la noticia falsa.

Two Sample t-test

```
data: y_share_report_email_num by 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
```

Al igual que en los otros dos casos al tomar en cuenta compartir el reporte en facebook y whatsapp podemos rechazar la hipótesis nula y concluir que no hay evidencia suficiente que demuestre que la diferencia del promedio de compartir el reporte via facebook o whatsapp entre el falso y verdadero es igual a cero.

Two Sample t-test

```
data: y_share_report_fb_num by 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
```

Two Sample t-test

```
data: y_share_report_whatsapp_num by 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
```

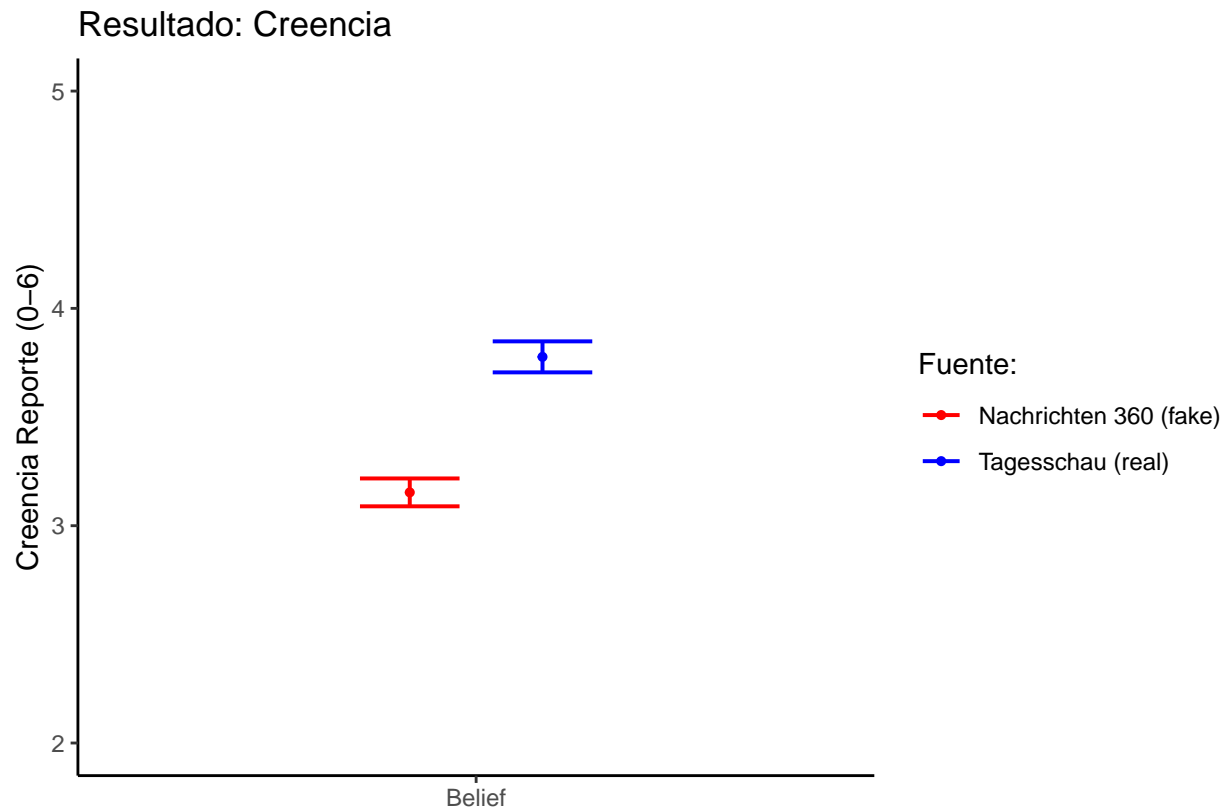
Sin embargo, el valor p de compartir la noticia en twitter es mayor al nivel de significancia. Así no hay evidencia suficiente para rechazar la hipótesis nula de que la diferencia del promedio de ambos grupos es igual a cero.

Two Sample t-test

```
data: y_share_report_twitter_num by 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
```

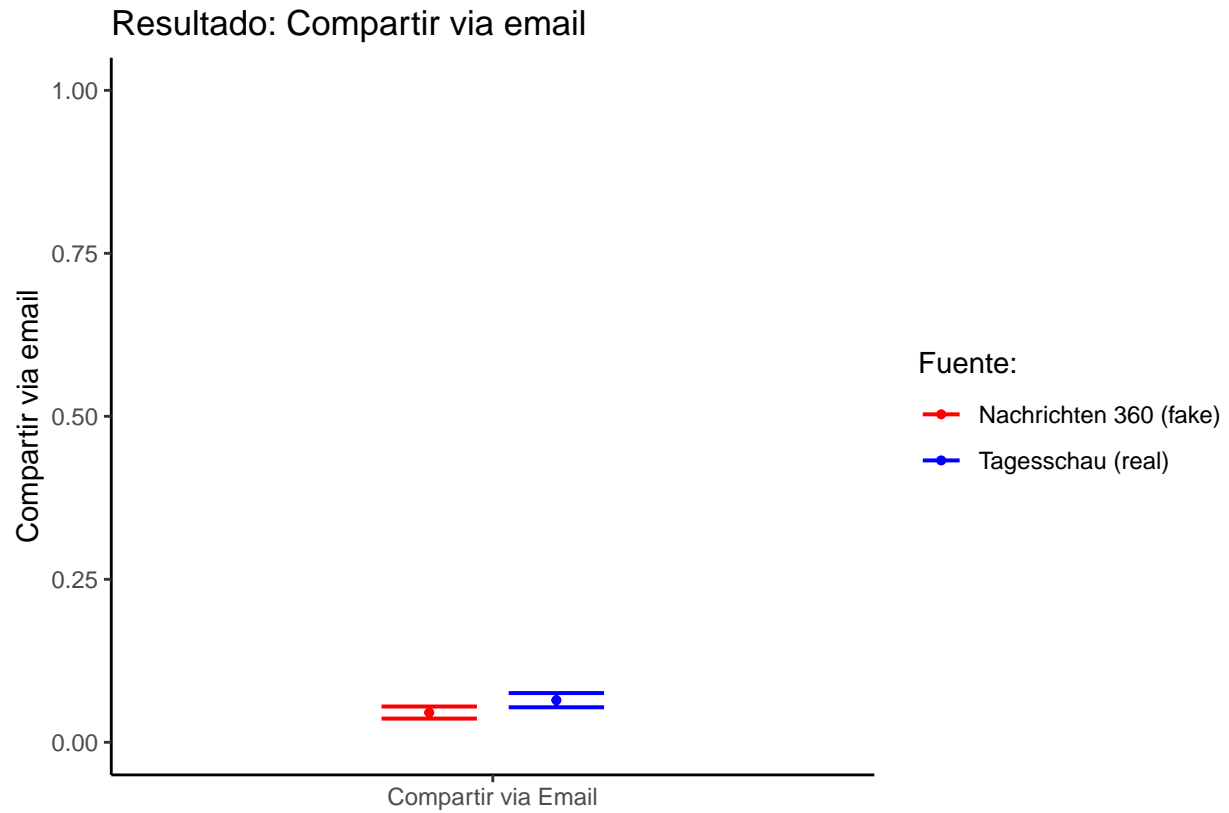
7. Muestre gráficamente los efectos que tiene la fuente en la posibilidad de creer en las noticias y en la propensión a compartirlas. Esto es, replique los gráficos de la figura 3 (pág. 13). Interprete estos resultados y concluya en el contexto del problema de acuerdo a las hipótesis planteadas.

Con base en este primer resultado observamos que la gente cree más en el reporte de la noticia verdadera que el de la noticia falsa.

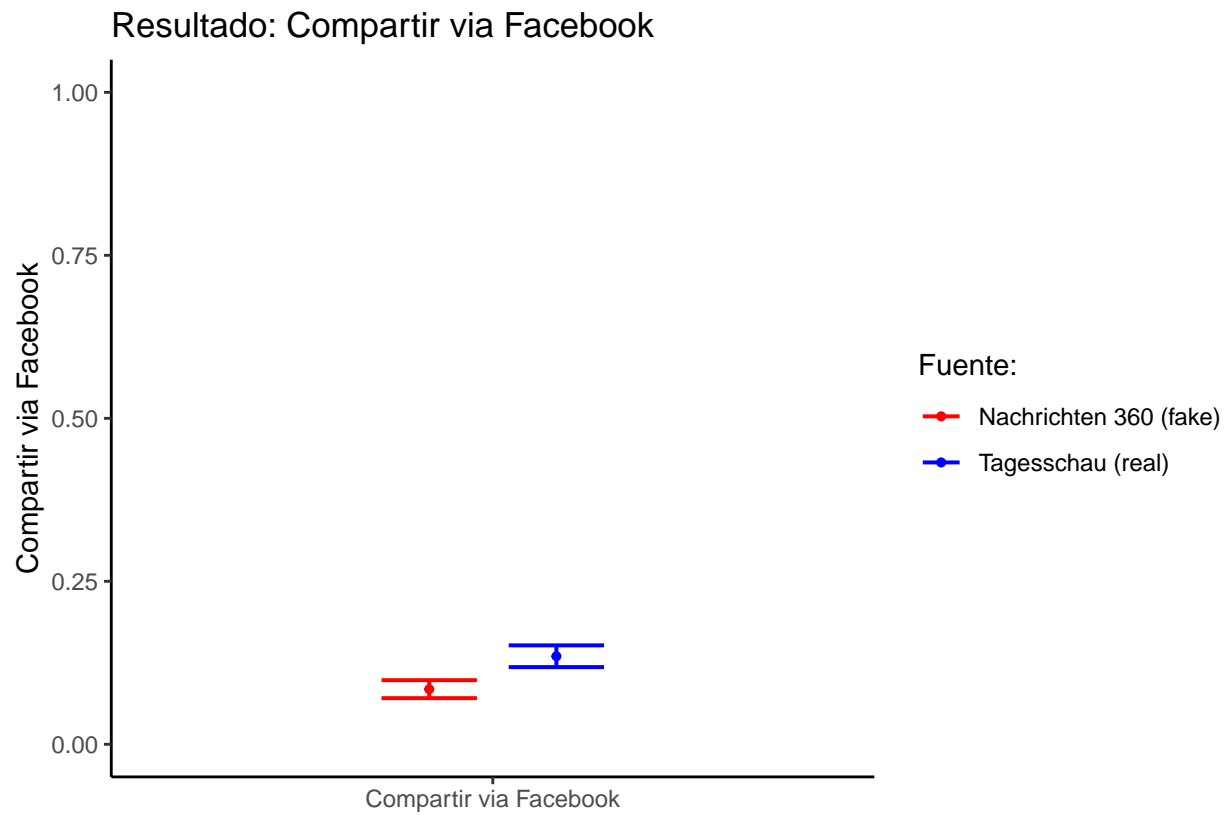


En los siguientes boxplots se respetan los ejes 0 a 1 para que sean comparables entre ellos.

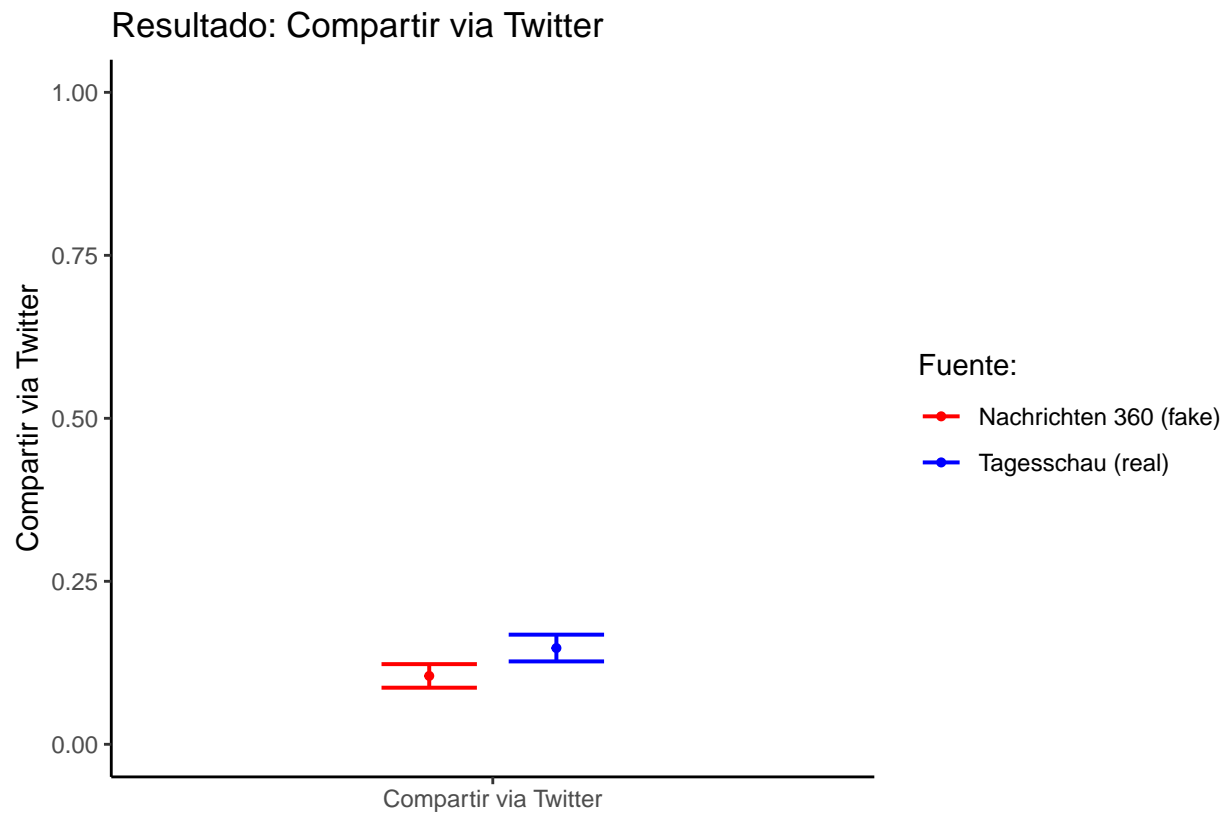
Observamos que un mayor número de encuestados compartirían el reporte real que el fake. Observamos que el rango de los valores que se encuentran fuera del rango intercuartílico superior de la noticia falsa está muy cerca del inferior de la noticia verdadera.



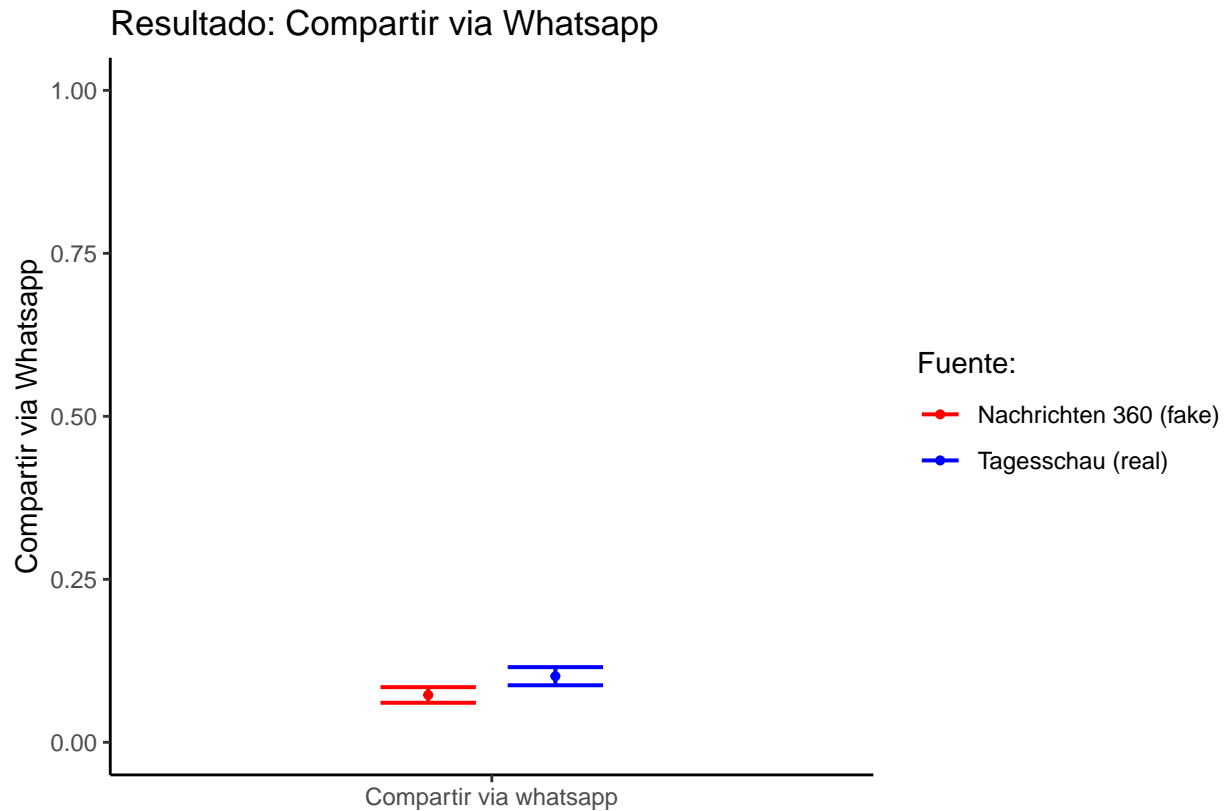
Del mismo modo que el diagrama anterior las personas contestaron que compartirían la noticia verdadera a la falsa. En este caso el rango de los valores que se encuentran fuera del rango intercuartílico está más separado que lo expuesto anteriormente.



Compartir la noticia en twitter también es más factible que sea la verdadera a la falsa.



Obtenemos resultados similares con Whatsapp



En los gráficos anteriores observamos en todos que las personas tienen preferencias por compartir la noticia verdadera más que la falsa a través de todos los medios de distribución (facebook, whatsapp, mail y twitter). Tomando en cuenta resultados anteriores las personas comparten más noticias a través de twitter seguido de facebook, whatsapp y correo y en el gráfico correspondiente a Twitter observamos que por el rango intercuartílico más gente es propensa a compartir información verdadera en Twitter. Esto es relevante, ya que facebook, instagram y twitter son las redes más visitadas en Alemania. En el caso de facebook también más gente compartiría la noticia real; sin embargo, es una menor proporción.

8. Ahora obtenga estos mismos resultados pero implementando sus cálculos mediante la estimación adecuada de mínimos cuadrados ordinarios. Incluya el análisis para las diferentes fuentes, la congruencia de la noticia 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

Observamos que es significativo el tratamiento en todos los casos excepto en twitter. Esto es relevante pues twitter es la red social preferida para compartir noticias de los encuestados. Si tomamos en cuenta que hay menos observaciones en twitter y también las personas prefieren compartir noticias via facebook en este caso el tratamiento si es significativo. A su vez, observamos que el tratamiento de recibir la fuente real a la fake aumenta en mayor medida (.62 puntos en la escala) la creencia en el reporte.

Resultados: Fuente de tratamiento, creencia e intención de compartir

Dependent variable:			
	Creencia (1)	Compartir:Email (2)	Compartir:Facebook (3)
Tratamiento: Fuente	0.62*** (0.07)	0.02* (0.01)	0.05*** (0.02)
Constant	3.15*** (0.05)	0.05*** (0.01)	0.08*** (0.01)
Observations	1,979	1,930	1,210
R2	0.04	0.002	0.01
Adjusted R2	0.04	0.001	0.01
Residual Std. Error	1.54 (df = 1977)	0.23 (df = 1928)	0.31 (df = 1208)
F Statistic	81.38*** (df = 1; 1977)	3.35* (df = 1; 1928)	7.93*** (df = 1; 1208)
Note: *p<0.1; **p<0.05; ***p<0.01			

Resultados: Fuente de tratamiento, creencia e intención de compartir

Dependent variable:			
	Creencia (1)	Compartir:Twitter (2)	Compartir:Whatsapp (3)
Tratamiento: Fuente	0.62*** (0.07)	0.04 (0.04)	0.03** (0.01)
Constant	3.15*** (0.05)	0.10*** (0.03)	0.07*** (0.01)
Observations	1,979	292	1,639
R2	0.04	0.004	0.003
Adjusted R2	0.04	0.001	0.002
Residual Std. Error	1.54 (df = 1977)	0.33 (df = 290)	0.28 (df = 1637)
F Statistic	81.38*** (df = 1; 1977)	1.20 (df = 1; 290)	4.29** (df = 1; 1637)
Note: *p<0.1; **p<0.05; ***p<0.01			

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

Primero observamos que el valor p es mayor al nivel de significancia en el caso del ingreso con base en haber recibido la noticia verdadera o falsa. Así, no rechazamos la hipótesis nula de que la diferencia del promedio sea igual a cero.

```
##
## Welch Two Sample t-test
##
## data: i1 and i0
## t = -0.39354, df = 1957.1, p-value = 0.694
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
## -0.2872427 0.1912306
## sample estimates:
## mean of x mean of y
## 5.690302 5.738308
```

Con base en la educación del mismo modo observamos un valor p mayor al nivel de significancia y no rechazamos la hipótesis nula de que la diferencia del promedio sea igual a cero.

```
##
## Welch Two Sample t-test
##
## data: e1 and e0
## t = -0.023224, df = 1958.1, p-value = 0.9815
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
## -0.1283497 0.1253454
## sample estimates:
## mean of x mean of y
## 3.087409 3.088911
```

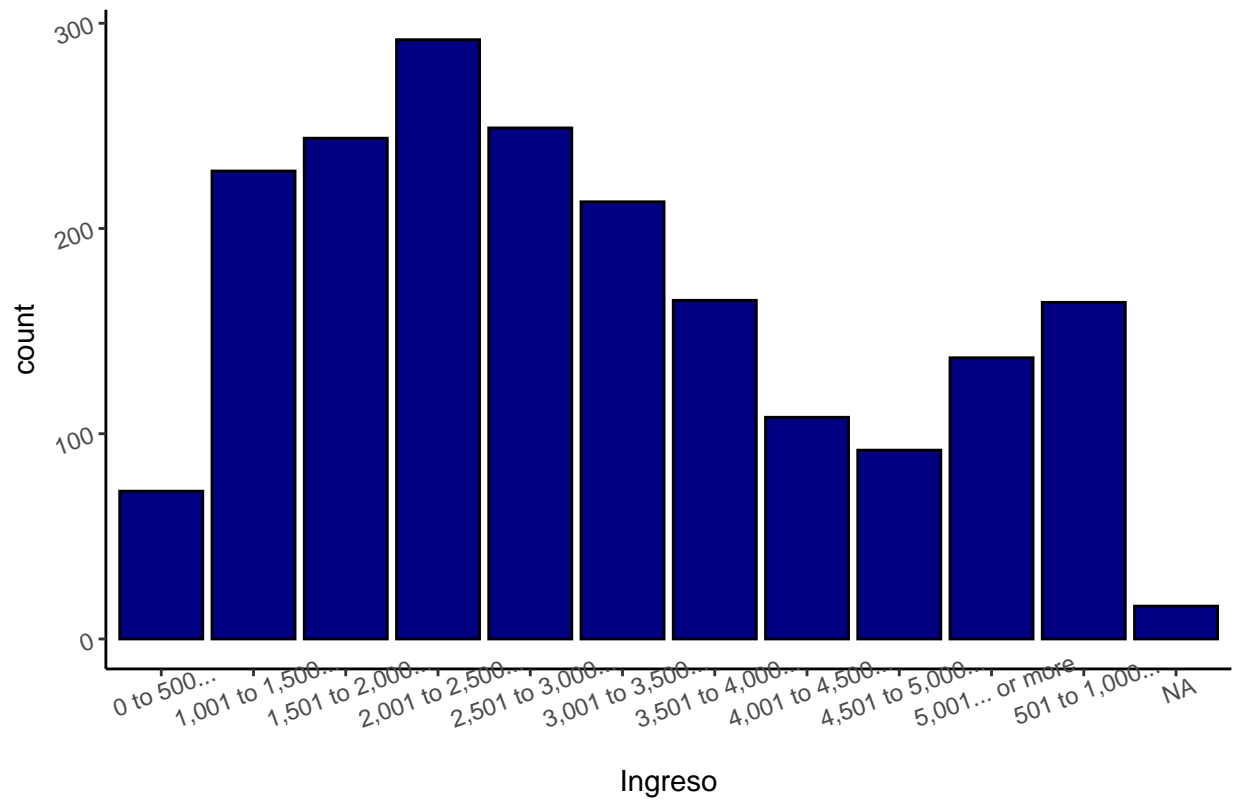
Two Sample t-test

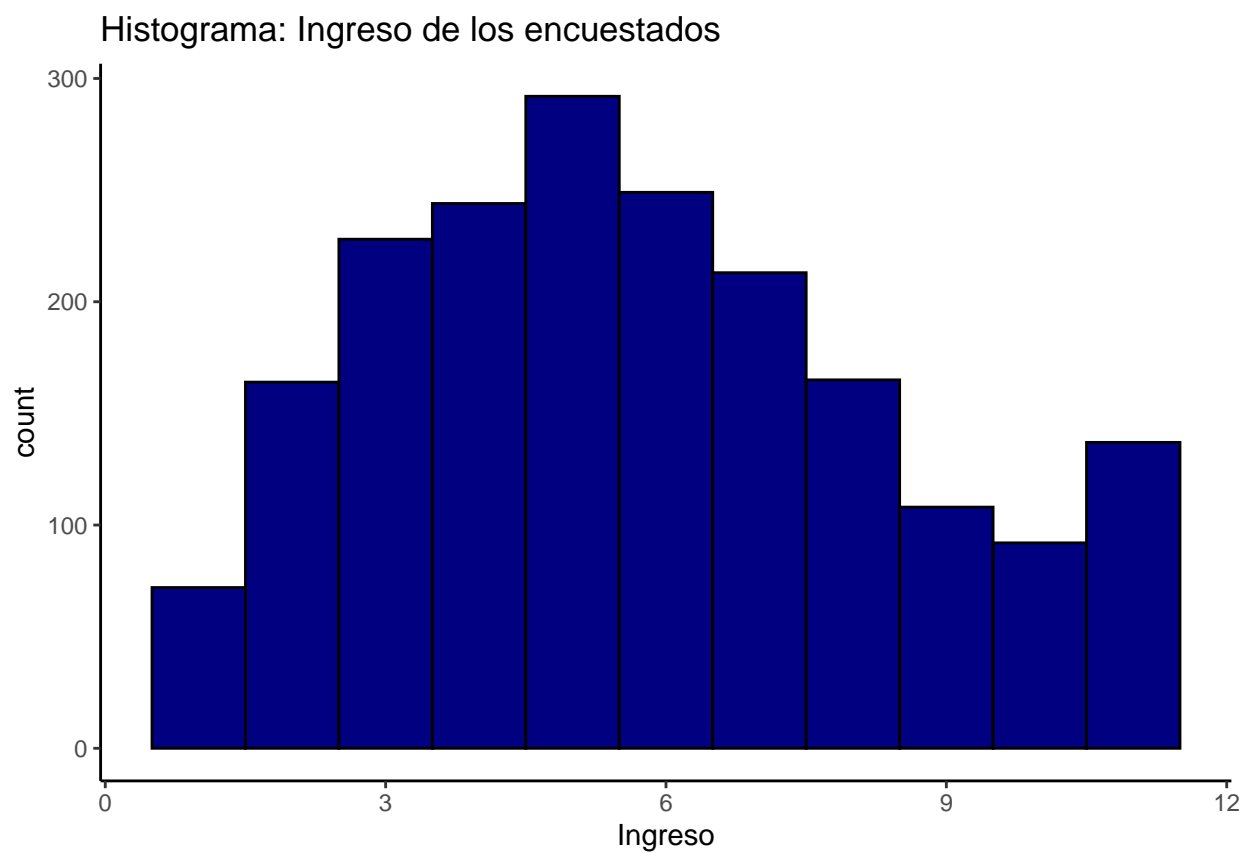
```
data: y_share_report_fb_num by d_treatment_source
t = -2.8157, df = 1208, p-value = 0.004947
alternative hypothesis: true difference in means between group 0 and group 1 is not equal to 0
95 percent confidence interval:
-0.08571086 -0.01531613
sample estimates:
mean in group 0 mean in group 1
0.08457711 0.13509061
```

En ambos casos no se observa una diferencia entre el grupo de tratamiento y control en el promedio de ingreso y educación. Lo cual mantiene balance.

Observamos el ingreso de los encuestados:

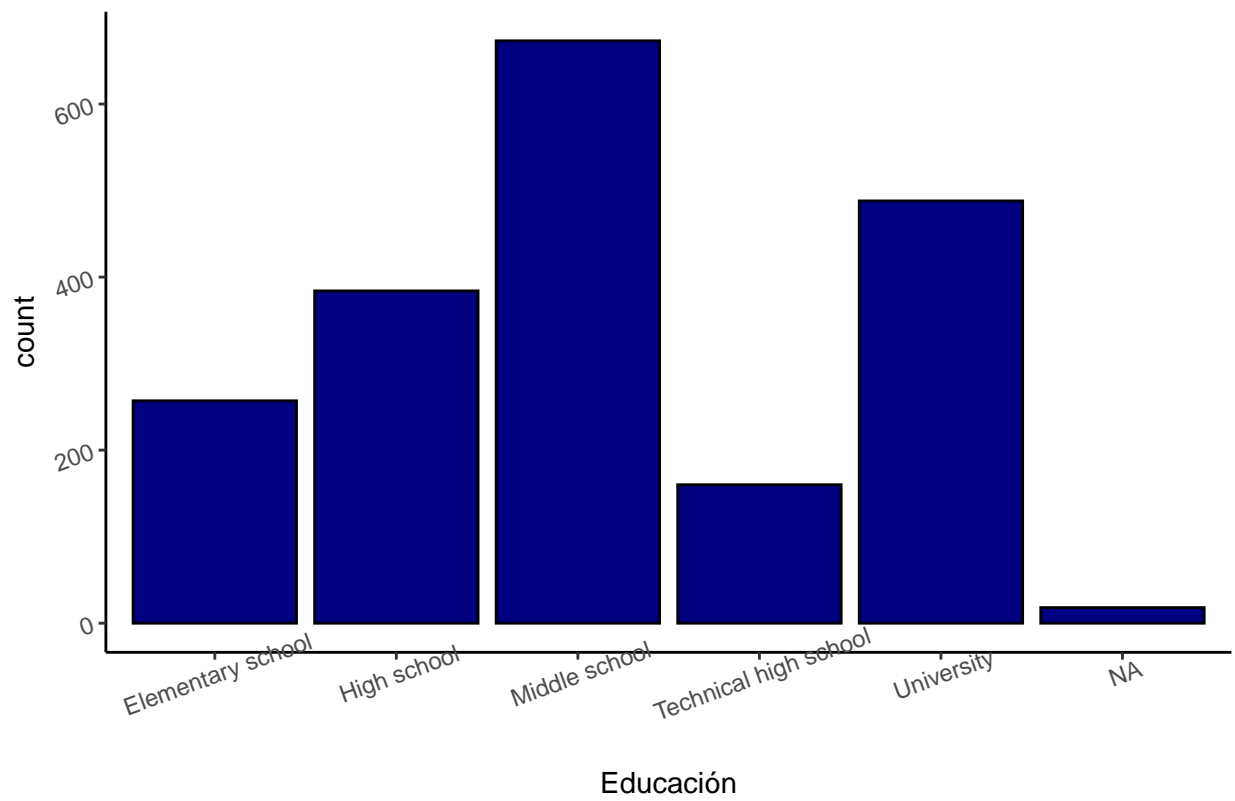
Gráfico: Ingreso de los encuestados



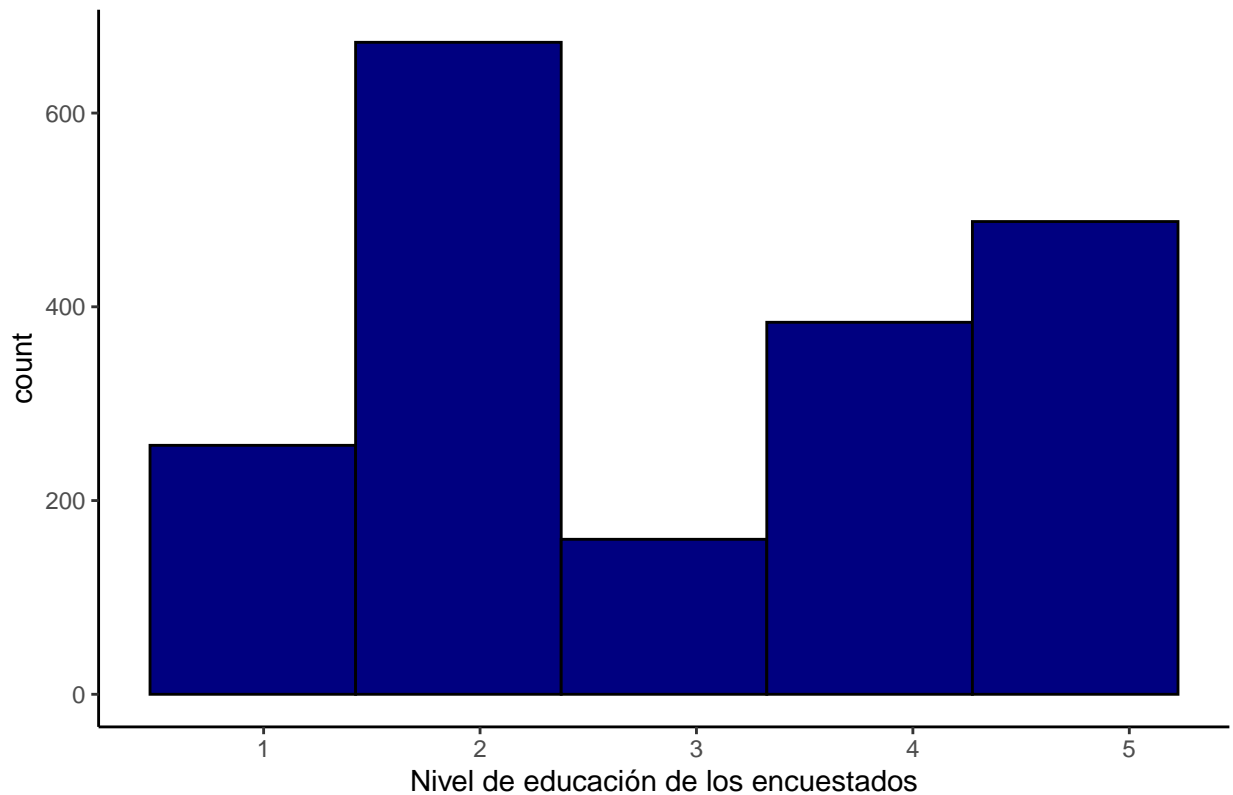


Nivel de educación de los encuestados:

Gráfico: Educación de los encuestados



Histograma: Nivel de educación de los encuestados



Incluir las en el modelo debería implicar relevancia teórica la cual sí tiene, pues como se comentó anteriormente tanto la educación como el ingreso pueden determinar si una persona va a compartir una noticia falsa o no. A su vez, es relevante para el objeto de estudio que es la difusión de las noticias falsas el nivel de ingresos y educación de las personas que interactúan en distintas redes sociales y comparten información. Con lo visto al inicio de esta pregunta observamos que no hay diferencias entre los datos de tratamiento y control por lo que se encuentran balanceados. Ahora bajo un modelo lineal simple observamos lo siguiente

Observamos que agregar el ingreso no es significativo y no hay una mejoría significativa del tratamiento. Así, no las agregaría en el caso de creencia en el reporte y con base en ello tampoco en el resto.

Call:

```
lm(formula = y_belief_report_num ~ d_treatment_source, data = datos)
```

Residuals:

Min	1Q	Median	3Q	Max
-3.7769	-0.7769	0.2231	1.2231	2.8467

Coefficients:

	Estimate	Std. Error	t value	Pr(> t)
(Intercept)	3.15331	0.04834	65.229	<2e-16 ***
d_treatment_source	0.62355	0.06912	9.021	<2e-16 ***

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 1.537 on 1977 degrees of freedom

(1 observation deleted due to missingness)

Multiple R-squared: 0.03954, Adjusted R-squared: 0.03905
F-statistic: 81.38 on 1 and 1977 DF, p-value: < 2.2e-16

Call:

```
lm(formula = y_belief_report_num ~ d_treatment_source + x_income_num +  
    x_education_num, data = datos)
```

Residuals:

Min	1Q	Median	3Q	Max
-4.2700	-0.8783	0.0937	1.0360	3.3355

Coefficients:

	Estimate	Std. Error	t value	Pr(> t)
(Intercept)	2.45047	0.10514	23.306	< 2e-16 ***
d_treatment_source	0.62128	0.06827	9.100	< 2e-16 ***
x_income_num	0.02139	0.01302	1.643	0.101
x_education_num	0.19260	0.02447	7.872	5.75e-15 ***

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 1.505 on 1941 degrees of freedom
(35 observations deleted due to missingness)

Multiple R-squared: 0.07459, Adjusted R-squared: 0.07316
F-statistic: 52.15 on 3 and 1941 DF, p-value: < 2.2e-16