

Agresiones en la página de Instagram de Noticias Uno

Sofía Cuartas García

Simón Cuartas Rendón

Julián Alejandro Úsuga Ortiz

Deivid Zhang Figueroa

Febrero de 2022

Contents

Motivación y contexto	1
Objetivos	2
Objetivo general	2
Objetivos específicos	2
Parámetros a estimar	2
Aspectos importantes	2
Levantamiento de la información.	3
Estudio piloto	3
Tamaño de muestra para la media poblacional	5
Tamaño de muestra para el total	6
Tabla resumen de los datos obtenidos en la encuesta.	6
Variable <i>comentarios</i>	6
Variable <i>peyorativos</i>	7
Variable <i>tema</i>	9
Estimadores	9
Conclusiones	10
Anexo. Evidencia fotográfica	10

Motivación y contexto

Durante la última década se ha visto un aumento vertiginoso en la cantidad de usuarios que tienen cuentas en las diferentes redes sociales disponibles en el mercado, permitiendo a las personas interactuar de nuevas maneras mediante estos espacios virtuales, en los cuales les resulta posible compartir momentos de sus vidas

mediante fotografías o videos, así como encontrarse con otros usuarios para discutir puntos de vista sobre algún tema y compartir noticias o contenidos de su interés.

No obstante, uno de los grandes problemas asociados a las redes sociales es la proliferación de agresiones y contenido de odio desde las más diversas fuentes, como lo pueden ser las noticias falsas, interacciones con usuarios con ideologías diferentes y comentarios a influencers (personas destacadas en las redes sociales) que no son de su agrado, entre otras.

Es por ello que nos motivamos a estudiar la cantidad total de comentarios con adjetivos peyorativos, emojis o cualquier otra representación o símbolo que evidencien de forma explícita la intención de agredir a una persona, grupo de personas, empresa, región geográfica o institución pública o privada, entre otras, que ocurren en la sección de comentarios de las publicaciones de Instagram del medio de comunicación Noticias Uno de Colombia. Pero además, se va a estudiar el promedio de comentarios peyorativos por publicación. Adicionalmente, se va a estimar la proporción de comentarios agresivos con el objetivo de mejorar las estimaciones (que son aquellas que carecen de un nombre real o una foto con una persona en ella).

Objetivos

Objetivo general

Realizar un trabajo de **muestreo estadístico** relacionado con la cantidad de comentarios agresivos, violentos, amenazantes o peyorativos que se presentan en las cajas de comentarios de las publicaciones de *Instagram* del perfil de **Noticias Uno**.

Objetivos específicos

- Estimar la cantidad media de comentarios agresivos, violentos, amenazantes o peyorativos en cada publicación de *Instagram* de **Noticias Uno**.
- Estimar el total de comentarios agresivos, violentos, amenazantes o peyorativos en *todas* las publicación de *Instagram* de **Noticias Uno**.

Parámetros a estimar

- **Total** de comentarios peyorativos en la sección de comentarios de las publicaciones de Instagram del medio de comunicación Noticias Uno de Colombia.
- **Promedio** de comentarios que se consideran negativos en la sección de comentarios de las publicaciones de Instagram del medio de comunicación Noticias Uno de Colombia.

Aspectos importantes

- **Elemento.** Se considera como elemento a cada una de las publicaciones de Instagram realizadas por el medio de comunicación radial Noticias Uno de Colombia.
- **Población.** La población está dada por la totalidad de comentarios realizados en las publicaciones de Instagram realizadas por el medio de comunicación radial Noticias Uno de Colombia.
- **Unidad de muestreo.** Para lograr la estimación de los parámetros de interés, se realizará el estudio considerando como unidad de muestreo a cada uno de los comentarios de las publicaciones de Instagram realizadas por el medio de comunicación radial Noticias Uno de Colombia, que con corte a las 4:00 p.m. del viernes 10 de diciembre de 2021 totalizaron 2,290 publicaciones. Se debe tener en cuenta, además,

que para poder listar todas estas publicaciones se usó XXXX, el cual es una extensión de Chrome que se encarga de enlistar de forma automática todas las publicaciones de cualquier página de Instagram junto su enlace y la fecha de publicación de cada publicación, además de asignarle un número, el cual usamos como denominación para una de las publicaciones. Luego, para cada publicación, se tiene que esta exhibe de forma pública el número de comentarios que hay en esta y luego se los enlista en orden de aparición.

- **Tipo de muestreo.** Para abordar este trabajo se va a emplear el **muestreo por conglomerados**, para lo cual se toman los índices o ID de cada uno de los elementos muestrales, esto es, las fotos publicadas en la página de Instagram de *Noticias Uno* y se les organiza aleatoriamente para proseguir con una agrupación en grupos de quince publicaciones para poder construir los conglomerados, que serán quince en total, y a continuación se seleccionan de forma aleatoria veinte conglomerados para poder proceder con el muestreo. Todo lo anterior se realiza con ayuda de [R](#).

Levantamiento de la información.

Para poder realizar el levantamiento de la información usando muestreo por conglomerados se apeló a una herramienta llamada *DownAlbum*, que se encarga de enumerar cada una de las publicaciones de la página de *Instagram* de *Noticias Uno*, incluyendo el enlace que dirige a cada publicación, la fecha en la que fue subida a la red social y un índice, dándole el primer número a la publicación más reciente al momento de generar el listado, y el último a la primera publicación compartida por Noticias Uno en *Instagram* (la mas antigua). De esta forma, teniendo en cuenta que se consideraron todas las publicaciones que se hicieron hasta el sábado 11 de diciembre de 2021 a las 16:00 (4:00 p.m. GMT -5, es decir, hora de Bogotá), se tiene que el tamaño de la población, osea, el número total de publicaciones en Instagram de Noticias Uno hasta dicha fecha es de 2,290 y se procede con la construcción de los conglomerados como se explicó antes

Con lo anterior presente, se generan con [R](#) cuarenta números aleatorios para determinar cuáles van a ser los conglomerados a censar para recoger la información necesaria, los cuales se observan en la siguiente tabla

Para su levantamiento, entonces, cada uno de los cuatro integrantes asumió la recogida de datos de cinco conglomerados. Ahora bien, para el registro de los datos se creó un formulario de Microsoft Forms, el cual puede ser visto dando clic aquí. A continuación se muestran y se justifican cada uno de los elementos de este formulario.

- **Encuestador.** Registra cuál es la persona que está llenando la información.
- **Enlace.** Dirección web de la publicación.
- **ID.** Número de la publicación en el marco muestral.
- **Conglomerado.** ID del conglomerado asociado a la publicación censada.
- **Fecha de publicación.** Fecha en la que fue subida la publicación a Instagram.
- **Número de comentarios.** Cantidad total de comentarios.
- **Número de comentarios con mensajes peyorativos.** Cantidad de comentarios en la publicación que tenían mensajes peyorativos, agresivos o violentos, entre otros.
- **Tema.** Temática con la que se asocia la publicación. Para esta pregunta existen cuatro posibilidades:
 - Política nacional.
 - Deportes.
 - Entretenimiento.
 - Otros.

Y una vez recogidos los datos para las cuarenta publicaciones asignadas aleatoriamente por [R](#), se genera automáticamente un cuadro en Excel con las respuestas, que puede ser visualizado haciendo clic aquí.

Estudio piloto

Como se manifestó previamente, en este estudio se están considerando $N = 153$, cada uno de $M = 15$ unidades básicas. Así, se tiene que el número de unidades elementales en la población, es decir, el número

publicaciones total a considerar es $M_0 = NM = 2295$, que como se nota, son cinco unidades elementales más de las que realmente fueron consideradas, esto porque se aleatorizaron los conglomerados con cinco unidades elementales imaginarias extras para poder asegurar que todos tuvieran tamaño uno. Ahora bien, para este estudio piloto se están considerando $n = 20$ conglomerados diferentes, seleccionados de manera aleatoria con ayuda de R, obteniendo que los que deben ser considerados son los que se observan haciendo clic aquí.

Con esto, se va a considerar a μ_c como la media de comentarios peyorativos por conglomerados, μ la media de comentarios peyorativos en toda la población, es decir, en todas las publicaciones de *Noticias Uno* consideradas hasta la fecha de corte y τ el total. Además, sea y_{ij} el número de comentarios peyorativos registrados en la i -ésima publicación del j -ésimo conglomerado. Se sabe que los parámetros anteriores pueden ser calculados como sigue:

$$\begin{aligned}\tau &= \sum_{i=1}^N \tau_i = \sum_i \sum_j y_{ij} \\ \mu_c &= \frac{\tau}{N} \\ \mu &= \frac{\tau}{NM} = \frac{\tau}{M_0}\end{aligned}$$

Así, vale la pena comenzar observando el resultado obtenido en cada uno de los conglomerados en la siguiente tabla:

Table 1: Resultados en los conglomerados del estudio piloto

Conglomerado	Comentarios	Peyorativos	Anonimo
107	254	44	0
113	226	16	0
125	75	14	0
126	727	108	18
135	1722	230	32
136	53	3	1
139	324	56	6
153	143	20	1
19	242	39	2
2	210	32	8
23	180	5	0
28	590	88	20
34	159	26	7
35	87	18	1
39	481	65	8
45	93	9	0
50	128	18	6
60	95	15	0
70	84	7	0
91	90	10	0

Definido lo anterior, se puede comenzar estimando la media poblacional por conglomerado μ_c mediante la media muestral por conglomerado $\hat{\mu}_c$, el cual, teniendo en cuenta que los conglomerados tienen igual tamaño, puede ser calculada como sigue:

$$\hat{\mu}_c = \frac{1}{n} \sum_{i=1}^n \tau_i = \bar{y}_c$$

Así que al emplear la información de la tabla uno, se tiene que la media muestral por conglomerado es:

$$\hat{\mu}_c = \frac{4 + 16 + \dots + 7 + 10}{20} = \frac{823}{20} = 41.15$$

mo

Lo que quiere decir que en cada conglomerado hay un promedio de 41.15 comentarios peyorativos. Y con esto se puede estimar la cantidad promedio de comentarios peyorativos por unidad elemental como sigue:

$$\hat{\mu} = \frac{\hat{\mu}_c}{M} = 2.7433$$

Es decir, existe un promedio de 2.74 comentarios peyorativos, agresivos o irrespetuosos en cada publicación de *Instagram* de *Noticias Uno*. Además, el total de comentarios peyorativos puede ser estimado como sigue:

$$\hat{\tau}_c = N\hat{\mu}_c = 2295 \times 41.15 \approx 6295.95$$

De manera que se estima que se han dejado un total de 6296 comentarios peyorativos en todas las publicaciones de *Instagram* de *Noticias Uno* que habían si subidas a la red social hasta el once de diciembre de 2021.

Ahora bien, con el objetivo de determinar cuál es el número de conglomerados que deben ser estudiados en el muestreo final, es necesario obtener la varianza estimada de los estimadores expuestos previamente, lo cual se consigue con las siguientes ecuaciones:

$$\widehat{Var}[\hat{\tau}_c] = \frac{N(N-n)}{N} \sum_{i=1}^n \frac{(\tau_i - \hat{\mu}_c)^2}{n-1}$$

$$\widehat{Var}[\hat{\mu}] = \frac{1}{M^2} \widehat{Var}[\hat{\tau}_c]$$

Y al realizar los respectivos cálculos se tiene que los estimadores de dichas varianzas son:

$$\widehat{Var}[\hat{\tau}_c] = 2835394.8525$$

$$\widehat{Var}[\hat{\mu}] = 0.5383$$

Y con estos cálculos se tienen la mayoría de elementos necesarios para poder determinar el número de conglomerados que deben ser finalmente considerados.

Tamaño de muestra para la media poblacional

Para calcular el número de conglomerados n_μ necesarios para estimar con la muestra a la media de comentarios peyorativos en todas las publicaciones es necesario emplear la siguiente ecuación:

$$n = \frac{n_0}{1 + \frac{n_0}{N}}, \quad n_0 = \frac{Z_{\alpha/2}^2 S_{con}^2}{B_\mu^2 M^2}$$

Así, es necesario calcular el valor de S_{con}^2 , el cual puede ser calculado como sigue:

$$S_{con}^2 = \sum_{i=1}^n \frac{(\tau_i - \hat{\mu}_c)^2}{n-1}$$

Para el cual, al realizar los cálculos con R, se obtiene que $S_{con}^2 = 2786.7658$. Además, considerando que se va a emplear un nivel de confianza del 90 % ($\alpha = 0.10$) (con lo que $Z_{0.05/2} = 1.6449$) y un límite en el error de la estimación relativo del 10%, es decir, $\epsilon = 0.15$, de manera que el límite en el error de estimación B_μ es: $B_\mu = 0.10 \times 2.7433 = 0.4115$. Así, se tiene que el número de conglomerados a muestrear si estos fuesen infinitos es:

De manera que, al realizar la corrección por población finita y realizar las cuentas, se tiene que el número de conglomerados a muestrear es 87

Tamaño de muestra para el total

Para encontrar el tamaño de muestra necesario para estimar el total poblacional es necesario emplear la siguiente ecuación:

$$n = \frac{NS_{con}^2}{ND + S_{con}^2}, \quad D = \frac{B_\tau^2}{Z_{\alpha/2}^2 N^2}$$

De manera que para poder emplear esta ecuación es necesario comenzar encontrando el límite en el error de estimación, que, considerando que este está dado por $B_\tau = \epsilon \times \hat{\tau}_c$, de manera que se obtiene que $B_\tau = 944.3925$. Luego, realizando las cuentas indicadas por las dos ecuaciones anteriores se tiene que:

$$D = 14.0821, \quad n_\tau \approx 86.2876$$

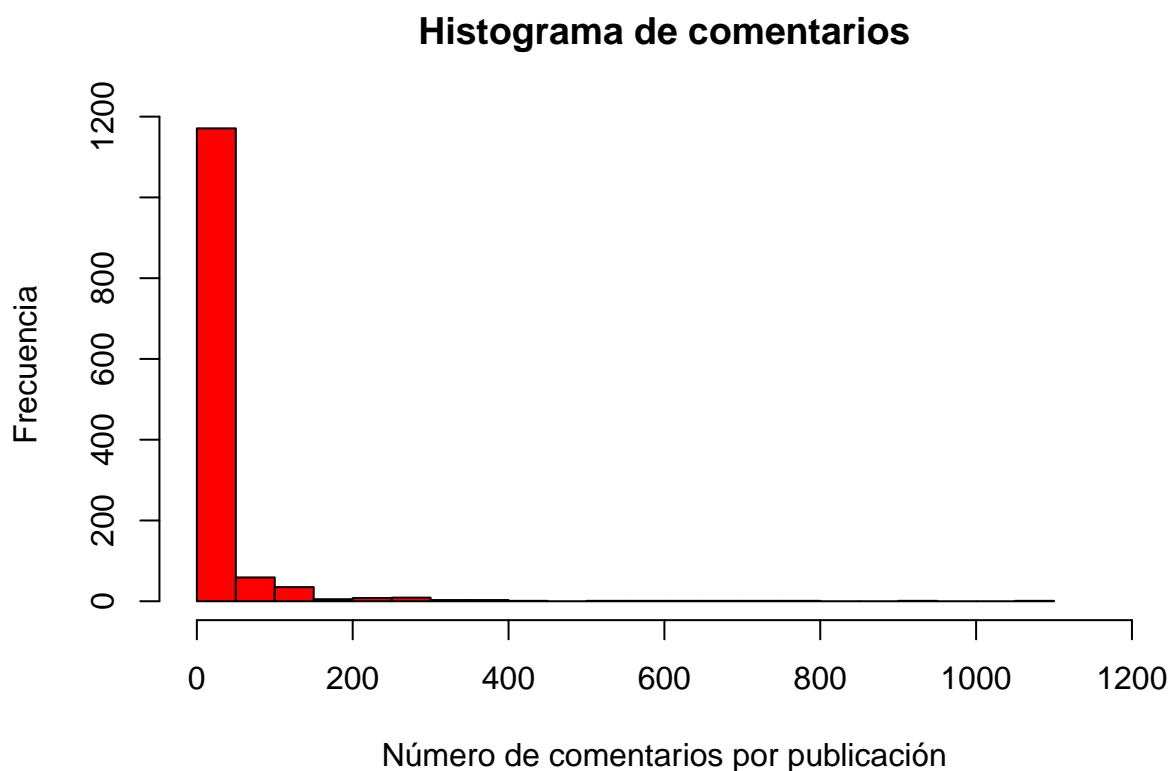
Y como se ve el resultado es el mismo que para el tamaño de la muestra de la media poblacional, es decir, que se deben muestrear un total de 87 conglomerados.

Tabla resumen de los datos obtenidos en la encuesta.

Variable *comentarios*.

Esta variable contiene el números total de comentarios para cada publicación muestreada, unos estadísticos de resumen son los siguientes:

Número de comentarios por publicación	
Minimo	0
Primer Cuantil (Q1)	1
Mediana	2
Tercer cuantil (Q3)	12
Máximo	1052
Moda	0



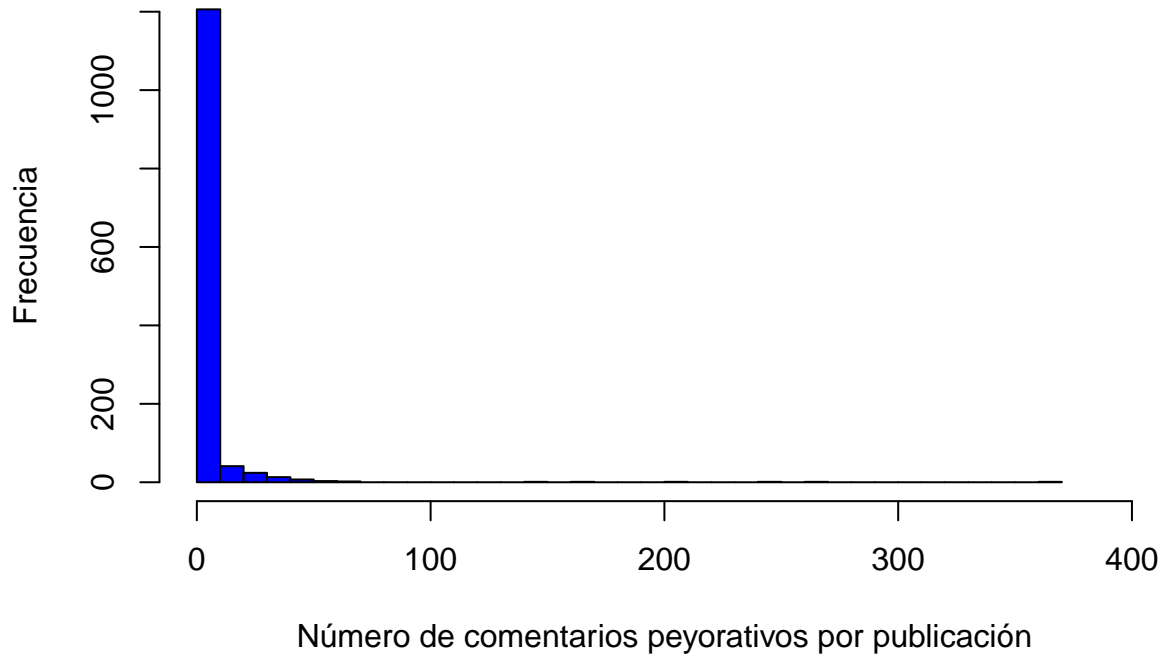
vemos que el número de comentarios que más se repite es cero con una frecuencia de 301, además como se puede ver en el gráfico de barra el número de comentarios varía mucho, y es evidente que un número de comentarios más pequeño es lo más frecuente para una publicación.

Variable *peyorativos*.

Esta variable contiene el número de comentarios peyorativos de cada publicación; a continuación una tabla con estadísticos de resumen y el respectivo diagrama de barras

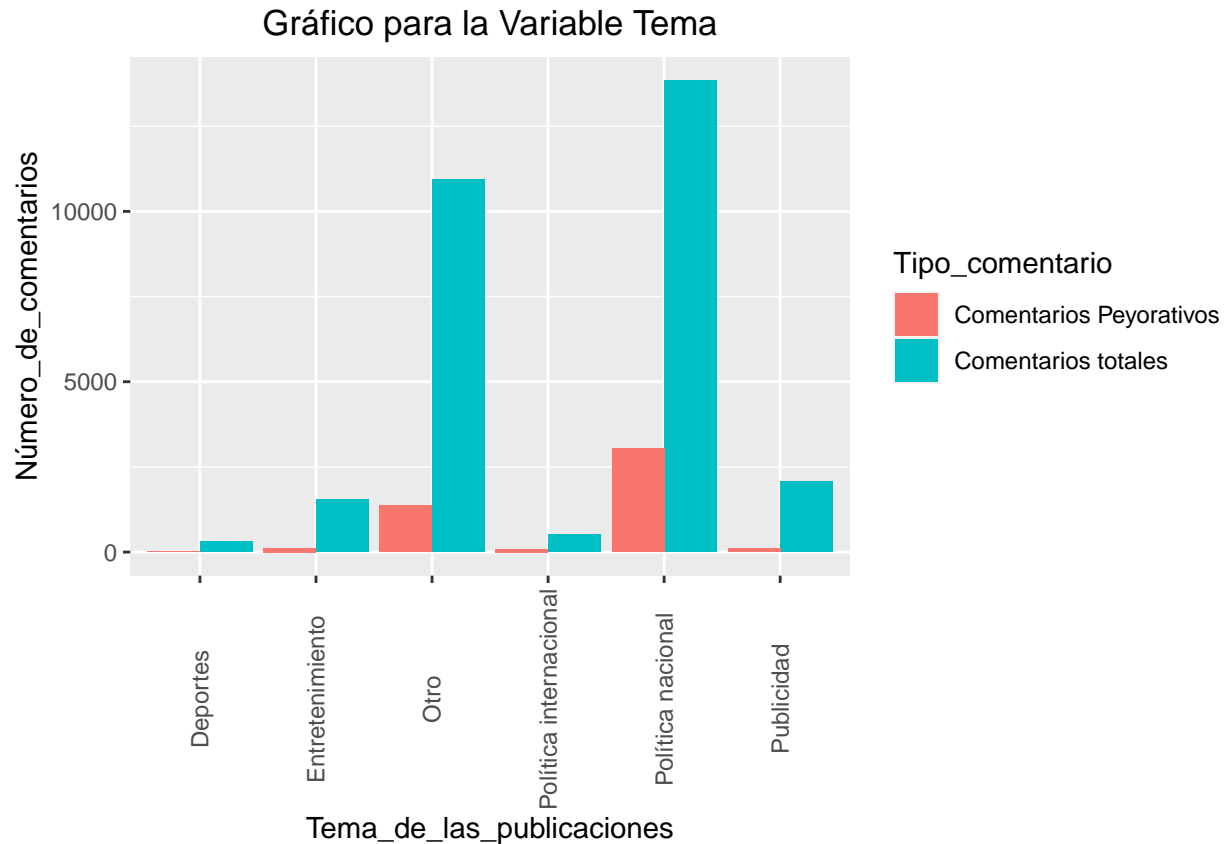
Número de comentarios peyorativos por publicación	
Minimo	0
Primer Cuantil (Q1)	0
Mediana	0
Tercer cuantil (Q3)	1
Máximo	363
Moda	0

Histograma de comentarios peyorativos



Sabemos que el numero que mas se repite es cero con una frecuencia de 841, Nuevamente se observa que los valores que puede alcanzar esta variable varía bastante y que aún sigue siendo más frecuente un número más pequeño de comentarios peyorativos, lo que cambia esta vez es que la frecuencia de estos comentarios es mucho menos en comparación con la frecuencia de comentarios totales.

Variable *tema*.



En la anterior gráfica de barras se puede evidenciar que las publicaciones que están asociadas a política nacional son las que reciben más comentarios en general y comentarios peyorativos, le siguen las publicaciones relacionadas con otro tipo de temas, además resalta que las publicaciones relacionadas con deporte son las que menos comentarios reciben y parece ser que los comentarios peyorativos que contienen son mínimos.

Estimadores

Primeramente agrupamos por conglomerado y hallamos los parámetros de interés que son el total y la media.

La media estimada $\hat{\mu}_c$ de comentarios peyorativos por conglomerado está dado por:

```
## [1] 53.89773
```

Es decir, en promedio hay 54 comentarios peyorativos en cada conglomerado.

Ahora, se halla la media estimada de comentarios peyorativos por publicación $\hat{\mu}$, la cual está dada por:

```
## [1] 3.593182
```

Es decir, en promedio hay 4 comentarios peyorativos en cada publicación.

Para el total estimado de comentarios peyorativos para los 153 conglomerados está dado por:

$\hat{\tau}_c = 153 * \hat{\mu}_c = 153 * 53.89773 = 8246.353$. Es decir, hay 8247 comentarios peyorativos en total.

Teniendo estos valores, se halla los intervalos de confianza al 90% para los parámetros de interés μ y τ .

Primeramente, se halla la varianza estimada de $\hat{\tau}_c$ y $\hat{\mu}_c$, y con esta se hallará el error estándar estimado para estos estimadores:

```
## [1] 776.964
```

Así, sabiendo que $S_{con}^2 = 5341.7$, se tiene que el error estándar estimado es $e.e.[\hat{\tau}_c] = 776.964$.

```
## [1] 0.3392856
```

El error estándar estimado es $e.e.[\hat{\mu}] = 0.3392856$.

```
## Intervalo de confianza al 90% para tau: 6968.361 9524.345
```

```
## Intervalo de confianza al 90% para mu: 3.035107 4.151257
```

Luego con un nivel de confianza de 90%, el número total τ de comentarios peyorativos que hay en las 2290 publicaciones se encuentra entre 6969 comentarios y 9524 comentarios, y la cantidad media μ de comentarios peyorativos por publicación se encuentra entre 3 comentarios y 4 comentarios.

Por último, se halla el límite en el error de estimación B usando 87 conglomerados para comprobar si este mejoró.

```
## [1] 1277.992
```

```
## [1] 0.5580751
```

Como se puede ver ambos B aumentaron.

Conclusiones

- Nuestra motivación inicial al realizar este estudio era saber que tanto acoso puede haber en las redes sociales, el cual es muy común y a la vez muy poco cuantificado.
- Como vimos, tan solo en esta red social y en este perfil en particular, que pertenece a un medio de comunicación, obtenemos un numero total de comentarios peyorativos entre 6703 y 9791, los cuales son cientos de personas expresando sus opiniones de una forma no adecuada, ofendiendo comúnmente a las personas que pertenecen al medio de comunicación, a políticos o a otras personas que tienen puntos de vista diferentes a estos.
- También, al ver el valor promedio obtenido de comentarios peyorativos por publicación, que es entre 3 y 5, podemos decir que las personas tienden a expresarse de una forma peyorativa, a veces, sin importar de que contenido se les esté presentando.
- Se puede observar que las varianzas son altas y los intervalos de confianza, sobre todo el del total, amplios, lo cual se debe a que, como se vio en los histogramas, hay una dispersión alta del número de comentarios peyorativos. Esto probablemente contribuyó al degradamiento, esto es, en el aumento en el límite en el error de estimación absoluto tanto del total como de la media.

Anexo. Evidencia fotográfica