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

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).

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 uno de los comentarios en 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 Forums, 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.

Evidencia fotográfica

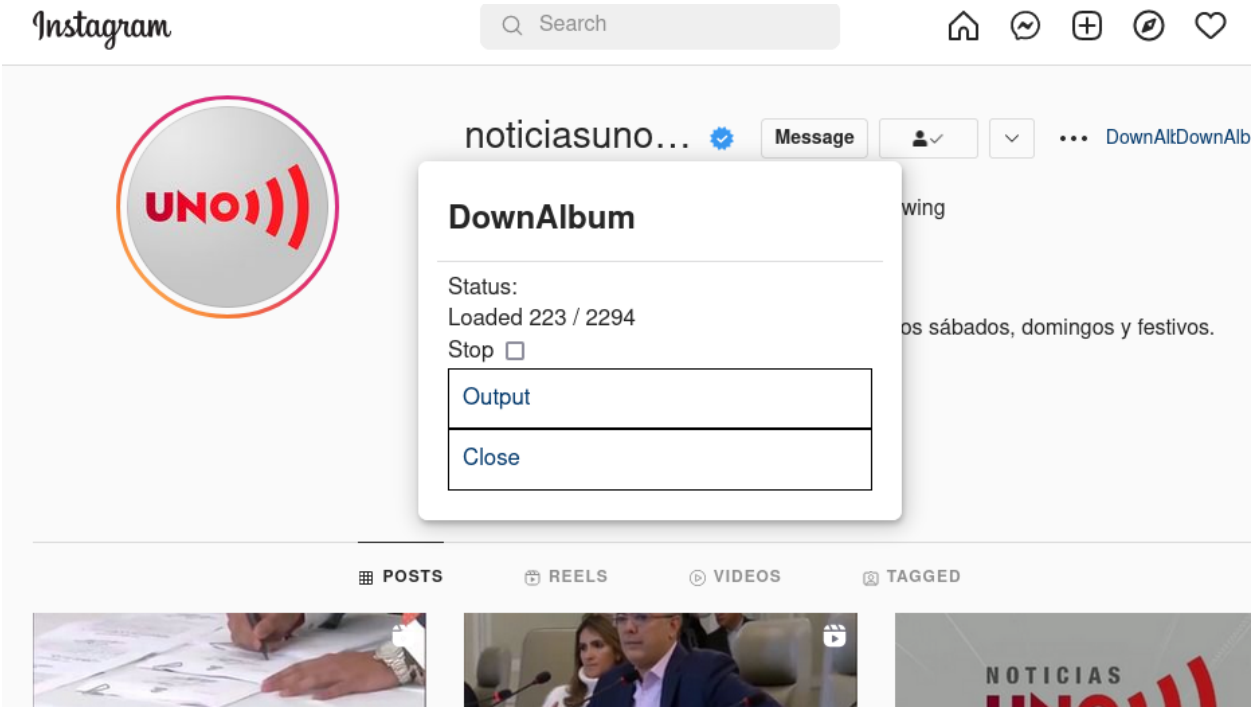


Figure 1: Recolección de publicaciones con la herramienta DownAlbum.

Estudio piloto

`Bla, bla, bla`

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 consideras, 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 siguientes:

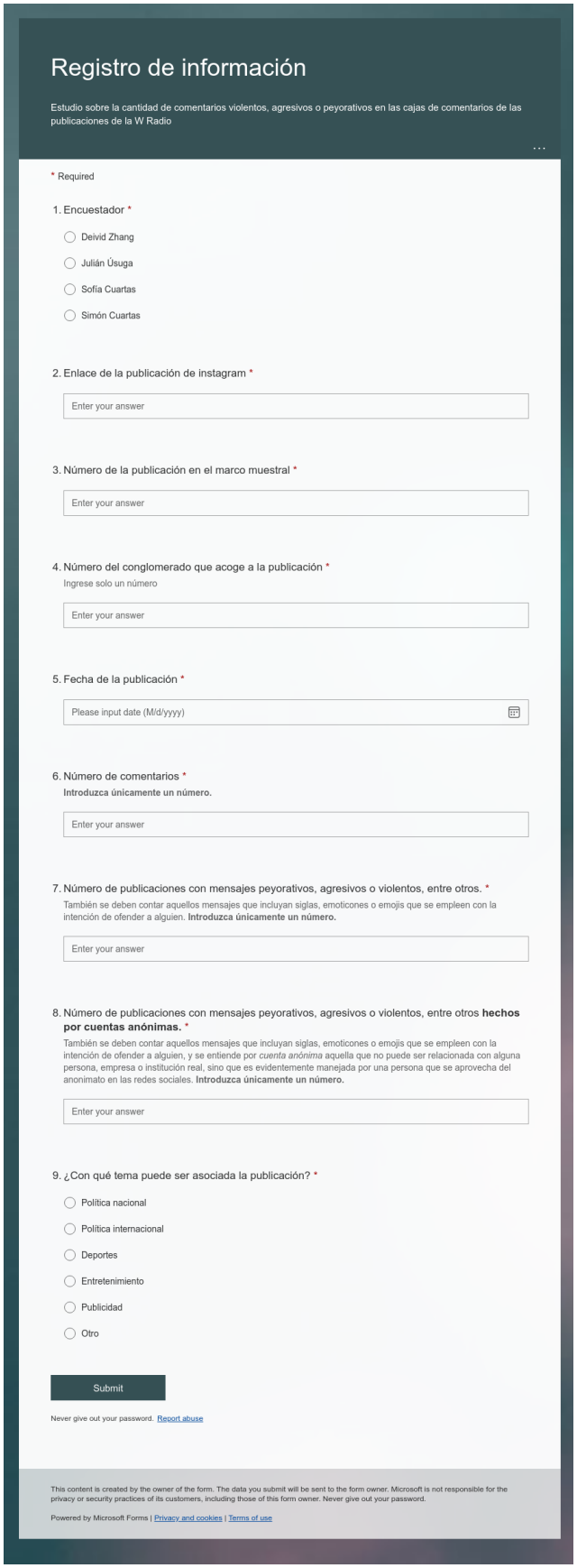
`Insertar lista de conglomerados a estudiar (db conglomerados)`

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-ésimo* publicación del *j-ésimo* conglomerado. Se sabe que los parámetros anteriores pueden ser calculados como sigue:

$$\tau = \sum_{i=1}^{N} \tau_i = \sum_i \sum_j y_{ij}$$

$$\mu_c = \frac{\tau}{N}$$

$$\mu = \frac{\tau}{NM} = \frac{\tau}{M_0}$$



Registro de información

Estudio sobre la cantidad de comentarios violentos, agresivos o peyorativos en las cajas de comentarios de las publicaciones de la W Radio

* Required

1. Encuestador *
   - Deivid Zhang
   - Julián Úsuga
   - Sofía Cuartas
   - Simón Cuartas

2. Enlace de la publicación de instagram *

   Enter your answer

3. Número de la publicación en el marco muestral *

   Enter your answer

4. Número del conglomerado que acoge a la publicación *

   Ingrese solo un número

   Enter your answer

5. Fecha de la publicación *

   Please input date (M/d/yyyy)

6. Número de comentarios *

   Introduzca únicamente un número.

   Enter your answer

7. Número de publicaciones con mensajes peyorativos, agresivos o violentos, entre otros. *

   También se deben contar aquellos mensajes que incluyan siglas, emoticones o emojis que se empleen con la intención de ofender a alguien. **Introduzca únicamente un número.**

   Enter your answer

8. Número de publicaciones con mensajes peyorativos, agresivos o violentos, entre otros **hechos por cuentas anónimas.** *

   También se deben contar aquellos mensajes que incluyan siglas, emoticones o emojis que se empleen con la intención de ofender a alguien, y se entiende por *cuenta anónima* aquella que no puede ser relacionada con alguna persona, empresa o institución real, sino que es evidentemente manejada por una persona que se aprovecha del anonimato en las redes sociales. **Introduzca únicamente un número.**

   Enter your answer

9. ¿Con qué tema puede ser asociada la publicación? *
   - Política nacional
   - Política internacional
   - Deportes
   - Entretenimiento
   - Publicidad
   - Otro

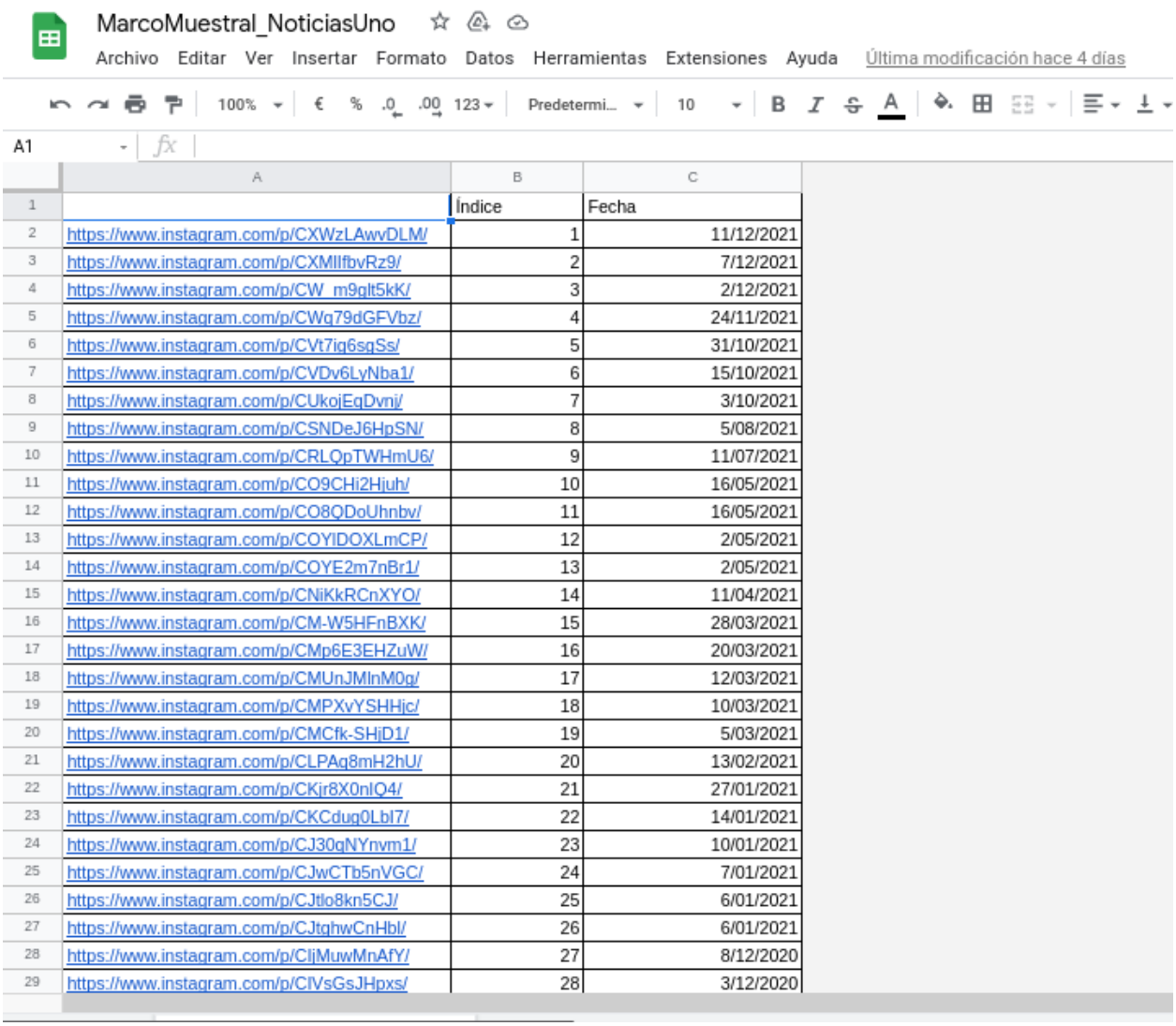
Submit

Never give out your password. Report abuse

This content is created by the owner of the form. The data you submit will be sent to the form owner. Microsoft is not responsible for the privacy or security practices of its customers, including those of this form owner. Never give out your password.

Powered by Microsoft Forms | Privacy and cookies | Terms of use

Figure 2: Herramienta usada para poder recopilar los datos facilmente (Office Forms).



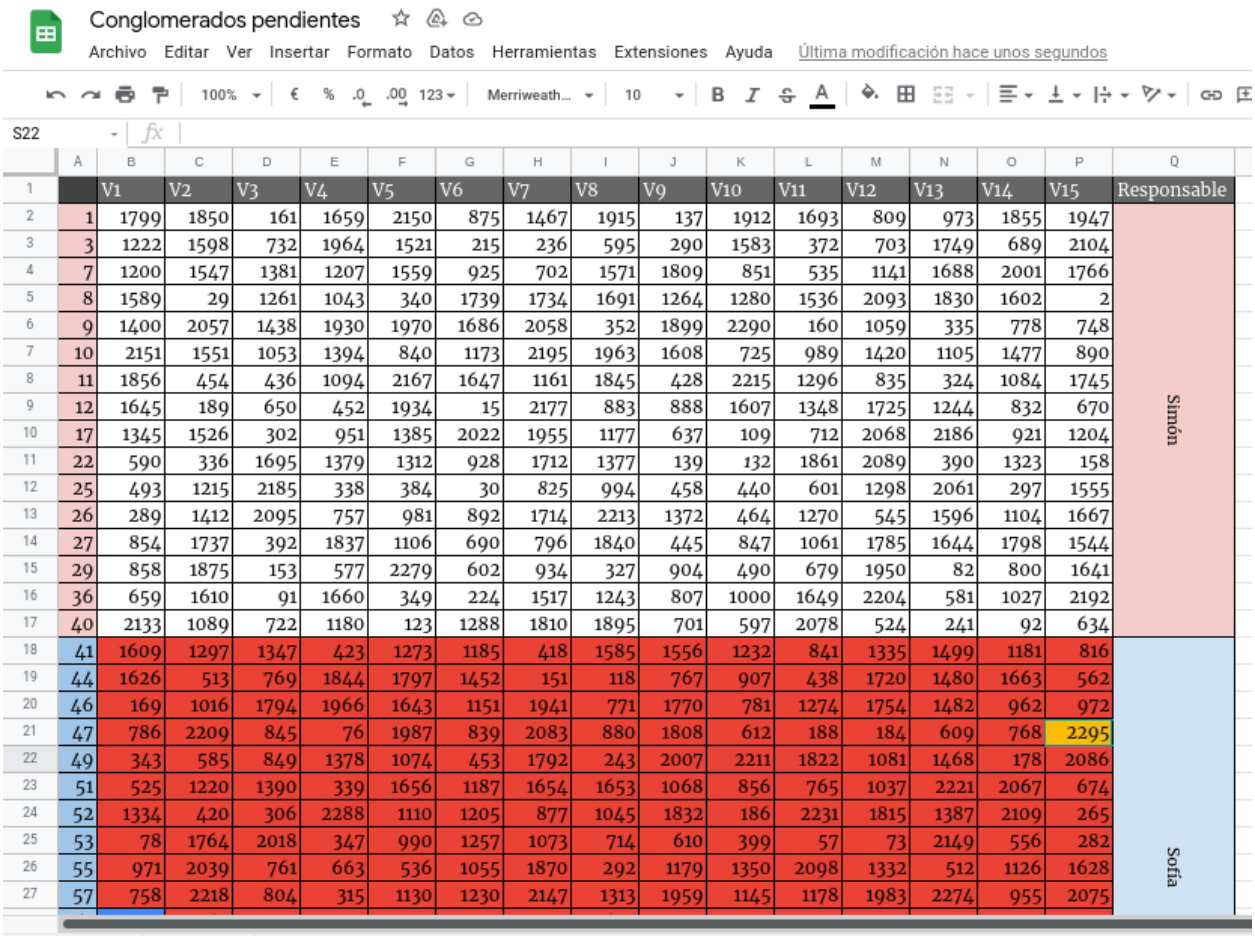
MarcoMuestral_NoticiasUno

Archivo Editar Ver Insertar Formato Datos Herramientas Extensiones Ayuda Última modificación hace 4 días

A1

	A	B	C
1		Índice	Fecha
2	https://www.instagram.com/p/CXWzLAwvDLM/	1	11/12/2021
3	https://www.instagram.com/p/CXMIIfbvRz9/	2	7/12/2021
4	https://www.instagram.com/p/CW_m9glt5kK/	3	2/12/2021
5	https://www.instagram.com/p/CWq79dGFVbz/	4	24/11/2021
6	https://www.instagram.com/p/CVt7ig6sgSs/	5	31/10/2021
7	https://www.instagram.com/p/CVDv6LyNba1/	6	15/10/2021
8	https://www.instagram.com/p/CUkojEqDvnj/	7	3/10/2021
9	https://www.instagram.com/p/CSNDeJ6HpSN/	8	5/08/2021
10	https://www.instagram.com/p/CRLQpTWHmU6/	9	11/07/2021
11	https://www.instagram.com/p/CO9CHi2Hjuh/	10	16/05/2021
12	https://www.instagram.com/p/CO8QDoUhnbv/	11	16/05/2021
13	https://www.instagram.com/p/COYIDOXLmCP/	12	2/05/2021
14	https://www.instagram.com/p/COYE2m7nBr1/	13	2/05/2021
15	https://www.instagram.com/p/CNiKkRCnXYO/	14	11/04/2021
16	https://www.instagram.com/p/CM-W5HFnBXK/	15	28/03/2021
17	https://www.instagram.com/p/CMp6E3EHZuW/	16	20/03/2021
18	https://www.instagram.com/p/CMUnJMlnM0g/	17	12/03/2021
19	https://www.instagram.com/p/CMPXvYSHHjc/	18	10/03/2021
20	https://www.instagram.com/p/CMCfk-SHjD1/	19	5/03/2021
21	https://www.instagram.com/p/CLPAq8mH2hU/	20	13/02/2021
22	https://www.instagram.com/p/CKjr8X0nIQ4/	21	27/01/2021
23	https://www.instagram.com/p/CKCdug0LbI7/	22	14/01/2021
24	https://www.instagram.com/p/CJ30qNYnvm1/	23	10/01/2021
25	https://www.instagram.com/p/CJwCTb5nVGC/	24	7/01/2021
26	https://www.instagram.com/p/CJtlo8kn5CJ/	25	6/01/2021
27	https://www.instagram.com/p/CJtghwCnHbl/	26	6/01/2021
28	https://www.instagram.com/p/CljMuwMnAfY/	27	8/12/2020
29	https://www.instagram.com/p/ClVsGsJHpxs/	28	3/12/2020

Figure 3: Marco muestral, donde cada publicación representa un conglomerado de comentarios.

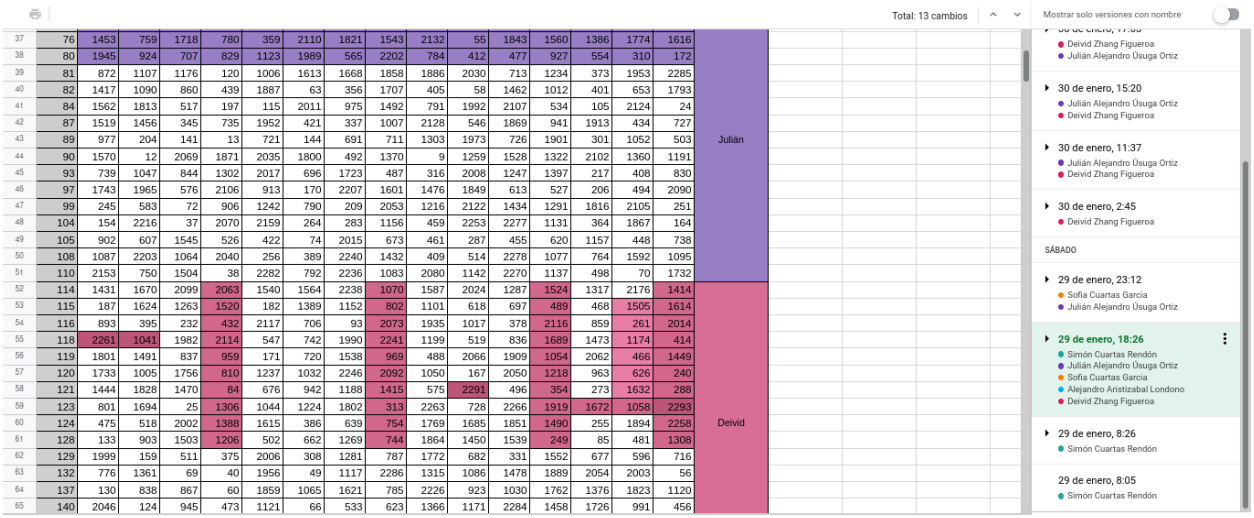


Conglomerados pendientes

Archivo Editar Ver Insertar Formato Datos Herramientas Extensiones Ayuda Última modificación hace unos segundos

	V1	V2	V3	V4	V5	V6	V7	V8	V9	V10	V11	V12	V13	V14	V15	Responsable
1	1799	1850	161	1659	2150	875	1467	1915	137	1912	1693	809	973	1855	1947	Simón
3	1222	1598	732	1964	1521	215	236	595	290	1583	372	703	1749	689	2104	
7	1200	1547	1381	1207	1559	925	702	1571	1809	851	535	1141	1688	2001	1766	
8	1589	29	1261	1043	340	1739	1734	1691	1264	1280	1536	2093	1830	1602	2	
9	1400	2057	1438	1930	1970	1686	2058	352	1899	2290	160	1059	335	778	748	
10	2151	1551	1053	1394	840	1173	2195	1963	1608	725	989	1420	1105	1477	890	
11	1856	454	436	1094	2167	1647	1161	1845	428	2215	1296	835	324	1084	1745	
12	1645	189	650	452	1934	15	2177	883	888	1607	1348	1725	1244	832	670	
17	1345	1526	302	951	1385	2022	1955	1177	637	109	712	2068	2186	921	1204	
22	590	336	1695	1379	1312	928	1712	1377	139	132	1861	2089	390	1323	158	
25	493	1215	2185	338	384	30	825	994	458	440	601	1298	2061	297	1555	
26	289	1412	2095	757	981	892	1714	2213	1372	464	1270	545	1596	1104	1667	
27	854	1737	392	1837	1106	690	796	1840	445	847	1061	1785	1644	1798	1544	
29	858	1875	153	577	2279	602	934	327	904	490	679	1950	82	800	1641	
36	659	1610	91	1660	349	224	1517	1243	807	1000	1649	2204	581	1027	2192	
40	2133	1089	722	1180	123	1288	1810	1895	701	597	2078	524	241	92	634	
41	1609	1297	1347	423	1273	1185	418	1585	1556	1232	841	1335	1499	1181	816	Sofía
44	1626	513	769	1844	1797	1452	151	118	767	907	438	1720	1480	1663	562	
46	169	1016	1794	1966	1643	1151	1941	771	1770	781	1274	1754	1482	962	972	
47	786	2209	845	76	1987	839	2083	880	1808	612	188	184	609	768	2295	
49	343	585	849	1378	1074	453	1792	243	2007	2211	1822	1081	1468	178	2086	
51	525	1220	1390	339	1656	1187	1654	1653	1068	856	765	1037	2221	2067	674	
52	1334	420	306	2288	1110	1205	877	1045	1832	186	2231	1815	1387	2109	265	
53	78	1764	2018	347	990	1257	1073	714	610	399	57	73	2149	556	282	
55	971	2039	761	663	536	1055	1870	292	1179	1350	2098	1332	512	1126	1628	
57	758	2218	804	315	1130	1230	2147	1313	1959	1145	1178	1983	2274	955	2075	

Figure 4: Asignación de las publicaciones y conglomerados a muestrear y progreso de este.



76	1453	759	1718	780	359	2110	1821	1543	2132	55	1843	1560	1386	1774	1616	Julián
80	1945	924	707	829	1123	1989	565	2202	784	412	477	927	554	310	172	
81	872	1107	1176	120	1006	1613	1668	1858	1886	2030	713	1234	373	1953	2285	
82	1417	1090	860	439	1887	63	356	1707	405	58	1462	1012	401	653	1793	
84	1562	1813	517	197	115	2011	975	1492	791	1992	2107	534	105	2124	24	
87	1519	1456	345	735	1952	421	337	1007	2128	546	1869	941	1913	434	727	
89	977	204	141	13	721	144	691	711	1303	1973	726	1901	301	1052	503	
90	1570	12	2069	1871	2035	1800	492	1370	9	1259	1528	1322	2102	1360	1191	
93	739	1047	844	1302	2017	696	1723	487	316	2008	1247	1397	217	408	830	
97	1743	1965	576	2106	913	170	2207	1601	1476	1849	613	527	206	494	2090	
99	245	583	72	906	1242	790	209	2053	1216	2122	1434	1291	1816	2105	251	
104	154	2216	37	2070	2159	264	283	1156	459	2253	2277	1131	364	1867	164	
105	902	607	1545	526	422	74	2015	673	461	287	455	620	1157	448	738	
108	1087	2203	1064	2040	256	389	2240	1432	409	514	2278	1077	764	1592	1095	
110	2153	750	1504	38	2282	792	2236	1083	2080	1142	2270	1137	498	70	1732	
114	1431	1670	2099	2063	1540	1564	2238	1070	1587	2024	1287	1524	1317	2176	1414	Deivid
115	187	1624	1263	1520	182	1389	1152	802	1101	618	697	489	468	1505	1614	
116	893	395	232	432	2117	706	93	2073	1935	1017	378	2116	859	261	2014	
118	2261	1041	1982	2114	547	742	1990	2241	1199	519	836	1689	1473	1174	414	
119	1801	1491	837	959	171	720	1538	969	488	2066	1909	1054	2062	466	1449	
120	1733	1005	1756	810	1237	1032	2246	2092	1050	167	2050	1218	963	626	240	
121	1444	1828	1470	84	676	942	1188	1415	575	2291	496	354	273	1632	288	
123	801	1694	25	1306	1044	1224	1802	313	2263	728	2266	1919	1672	1058	2293	
124	475	518	2002	1388	1615	386	639	754	1769	1685	1851	1490	255	1894	2258	
128	133	903	1503	1206	502	662	1269	744	1864	1450	1539	249	85	481	1308	
129	1999	159	511	375	2006	308	1281	787	1772	682	331	1552	677	596	716	
132	776	1361	69	40	1956	49	1117	2286	1315	1086	1478	1889	2054	2003	56	
137	130	838	867	60	1859	1065	1621	785	2226	923	1030	1762	1376	1823	1120	
140	2046	124	945	473	1121	66	533	623	1366	1171	2284	1458	1726	991	456	

Total: 13 cambios

Mostrar solo versiones con nombre

Deivid Zhang Figueroa
Julián Alejandro Úsuga Ortiz

30 de enero, 15:20
Julián Alejandro Úsuga Ortiz
Deivid Zhang Figueroa

30 de enero, 11:37
Julián Alejandro Úsuga Ortiz
Deivid Zhang Figueroa

30 de enero, 2:45
Deivid Zhang Figueroa

SÁBADO

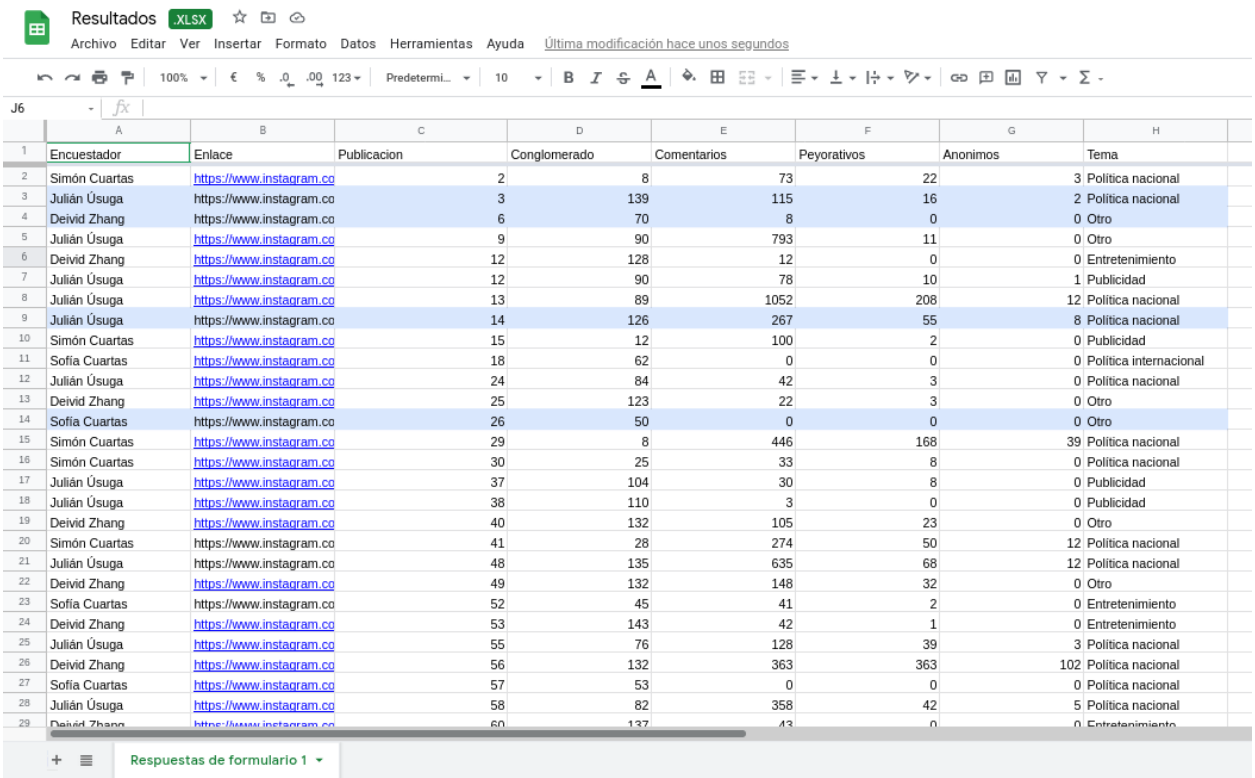
29 de enero, 23:12
Sofía Cuartas Garcia
Julián Alejandro Úsuga Ortiz

29 de enero, 18:26
Simón Cuartas Rendón
Julián Alejandro Úsuga Ortiz
Sofía Cuartas Garcia
Alejandro Aristizabal Londono
Deivid Zhang Figueroa

29 de enero, 8:26
Simón Cuartas Rendón

29 de enero, 8:05
Simón Cuartas Rendón

Figure 5: Progreso y evidencia de muestreo realizado por cada uno de los integrantes.



Encuestador	Enlace	Publicacion	Conglomerado	Comentarios	Peyorativos	Anonimos	Tema
Simón Cuartas	https://www.instagram.co	2	8	73	22	3	Política nacional
Julián Úsuga	https://www.instagram.co	3	139	115	16	2	Política nacional
Deivid Zhang	https://www.instagram.co	6	70	8	0	0	Otro
Julián Úsuga	https://www.instagram.co	9	90	793	11	0	Otro
Deivid Zhang	https://www.instagram.co	12	128	12	0	0	Entretenimiento
Julián Úsuga	https://www.instagram.co	12	90	78	10	1	Publicidad
Julián Úsuga	https://www.instagram.co	13	89	1052	208	12	Política nacional
Julián Úsuga	https://www.instagram.co	14	126	267	55	8	Política nacional
Simón Cuartas	https://www.instagram.co	15	12	100	2	0	Publicidad
Sofía Cuartas	https://www.instagram.co	18	62	0	0	0	Política internacional
Julián Úsuga	https://www.instagram.co	24	84	42	3	0	Política nacional
Deivid Zhang	https://www.instagram.co	25	123	22	3	0	Otro
Sofía Cuartas	https://www.instagram.co	26	50	0	0	0	Otro
Simón Cuartas	https://www.instagram.co	29	8	446	168	39	Política nacional
Simón Cuartas	https://www.instagram.co	30	25	33	8	0	Política nacional
Julián Úsuga	https://www.instagram.co	37	104	30	8	0	Publicidad
Julián Úsuga	https://www.instagram.co	38	110	3	0	0	Publicidad
Deivid Zhang	https://www.instagram.co	40	132	105	23	0	Otro
Simón Cuartas	https://www.instagram.co	41	28	274	50	12	Política nacional
Julián Úsuga	https://www.instagram.co	48	135	635	68	12	Política nacional
Deivid Zhang	https://www.instagram.co	49	132	148	32	0	Otro
Sofía Cuartas	https://www.instagram.co	52	45	41	2	0	Entretenimiento
Deivid Zhang	https://www.instagram.co	53	143	42	1	0	Entretenimiento
Julián Úsuga	https://www.instagram.co	55	76	128	39	3	Política nacional
Deivid Zhang	https://www.instagram.co	56	132	363	363	102	Política nacional
Sofía Cuartas	https://www.instagram.co	57	53	0	0	0	Política nacional
Julián Úsuga	https://www.instagram.co	58	82	358	42	5	Política nacional

Figure 6: Resultados del muestreo, extraidos a formato EXCEL usando Office Forms.

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

Conglomerado	Comentarios	Peyorativos	Anonimo
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:

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

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:

$$\widehat{Var}[\hat{\tau}_c] = \frac{N(N-n)}{N}\sum_{i=1}^{n}\frac{(\tau_i - \hat{\mu}_c)^2}{n-1}$$
$$\widehat{Var}[\hat{\mu}] = \frac{1}{M_0^2}\widehat{Var}[\hat{\tau}_c]$$

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:

$$n = \frac{n_0}{1 + \frac{n_0}{N}}, \qquad n_0 = \frac{Z^2_{\alpha/2} S^2_{con}}{B^2_\mu M^2}$$

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

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

Para el cual, al realizar los cálculos con R, se obtiene que $S^2_{con} = 2786.7658$. 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:

$$n = \frac{N S^2_{con}}{N D + S^2_{con}}, \qquad D = \frac{B^2_\tau}{Z^2_{\alpha/2} N^2}$$

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:

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

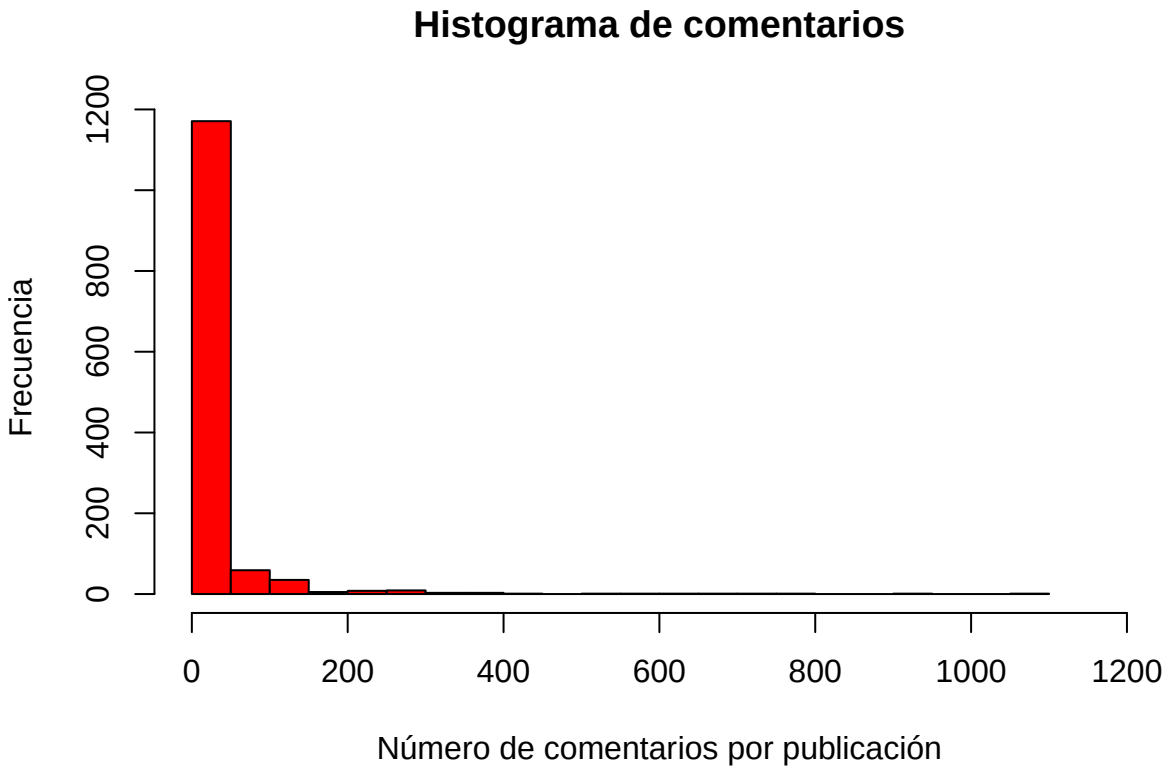
Y como se ve el resultado es el mismo que para el tamaño de la muestra de la media poblacional, es decir, que se deben muestrear un total de 87 conglomerados.

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

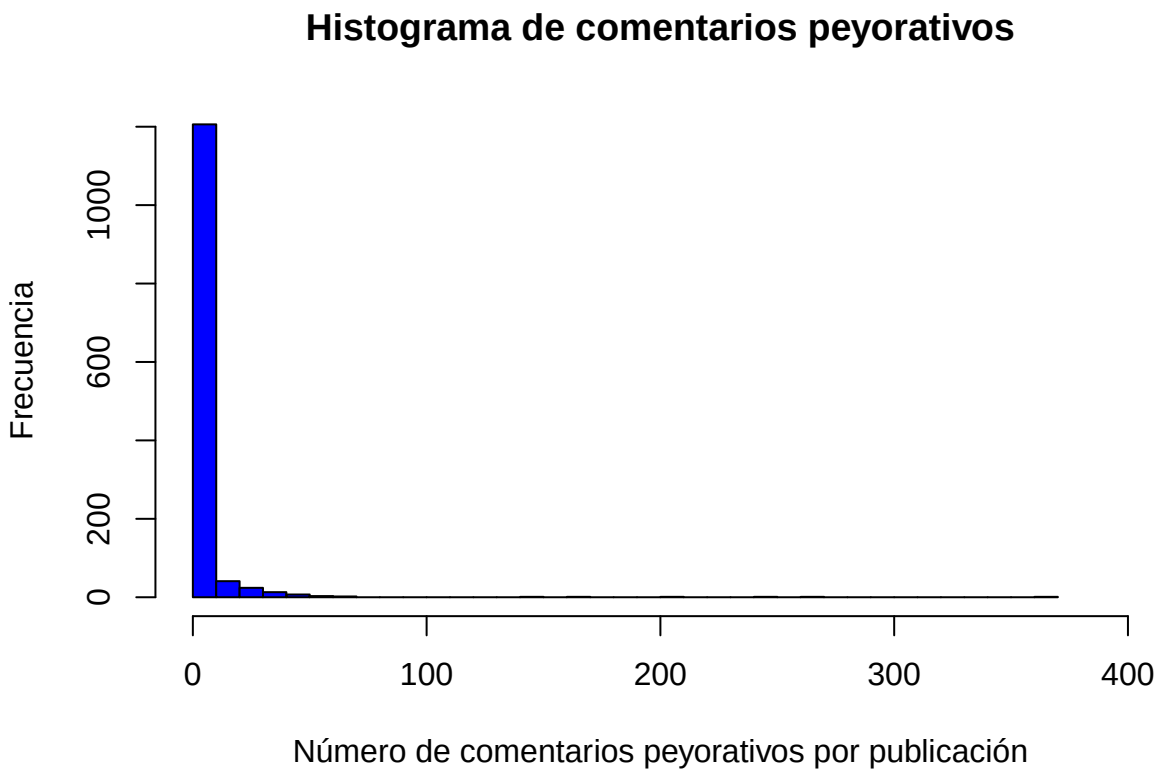


vemos que el número de comentarios que más se repite es cero con una frecuencia de 301, además como se puede ver en el gráfico de barra el número de comentarios varía mucho, y es evidente que un número de comentarios más pequeño es lo más frecuente para una publicación.

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

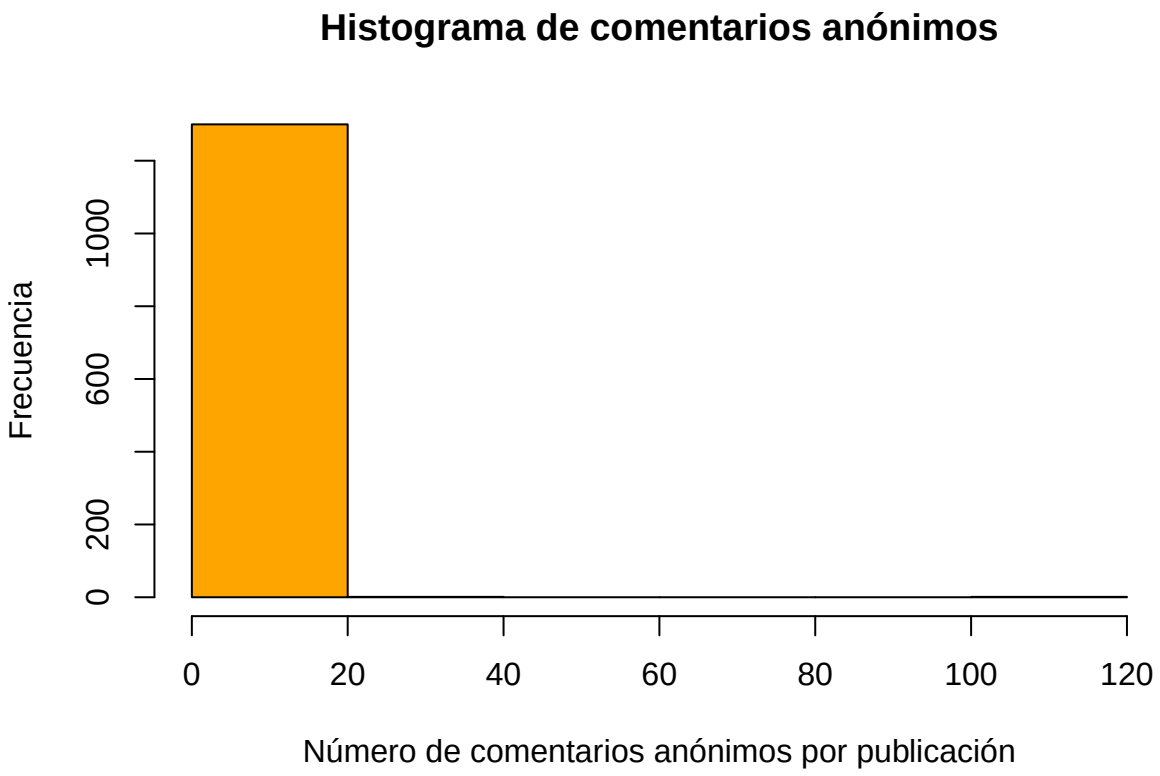
Número de comentarios peyorativos por publicación	
Minimo	0
Primer Cuantil (Q1)	0
Mediana	0
Tercer cuantil (Q3)	1
Máximo	363
Moda	0



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.

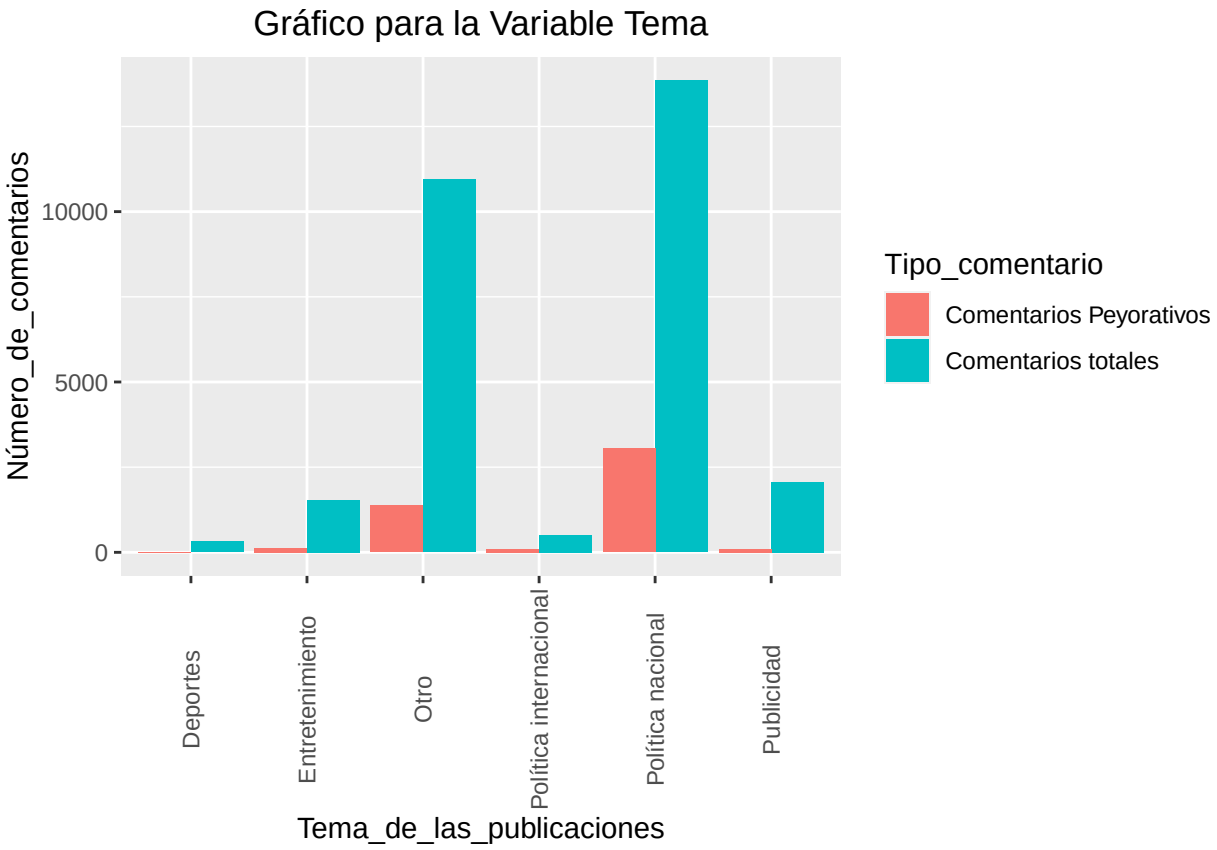
Variable *anonimos*.

Número de comentarios anonimos por publicación	
Minimo	0
Primer Cuantil (Q1)	0
Mediana	0
Tercer cuantil (Q3)	0
Máximo	102
Moda	0



Es evidente que el número de comentarios para esta variable se reduce en gran medida, y por lo que nos muestra los estadísticos de resumen parece ser que no suele haber comentarios realizados por cuentas anónimas.

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.

9B

Se halla los 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 95% 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
```

El error estándar estimado de $\hat{\tau}_c = 776.964$.

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

El error estándar estimado de $\hat{\mu} = 0.3392856$.

```
## Intervalo de confianza al 95% para tau:  6702.053 9790.653
```

```
## Intervalo de confianza al 95% para mu:  2.918815 4.267549
```

Luego con un nivel de confianza de 95%, el número total τ de comentarios peyorativos que hay en las 2290 publicaciones se encuentra entre 6703 comentarios y 9791 comentarios, y la cantidad media μ de comentarios peyorativos por publicación se encuentra entre 3 comentarios y 5 comentarios.

Por último, se halla el B usando 88 conglomerados para comprobar si este mejoró.

Primeramente, reemplazando los respectivos valores de la fórmula:

$$n = \frac{n_0}{1 + \frac{n_0}{N}} = 88 = \frac{n_0}{1 + \frac{n_0}{153}} \Longrightarrow n_0 = \frac{13464}{65}$$

Luego, despejando la fórmula siguiente:

$$n_0 = \frac{z^2 S_{con}^2}{B^2 M^2}$$

.

se tiene que,

$$B = \sqrt{\frac{z^2 S_{con}^2}{n_0 M^2}}$$

Reemplazando los valores obtenidos se tiene que,

$$B = \sqrt{\frac{1.96^2 * 5341.702}{\frac{13464}{65} * 2290^2}}$$

Luego, B = 0.0043464, y como podemos evidenciar esta disminuyó, lo que quiere decir que nuestro I.C es más preciso.