

Polarización en Tweets y Elecciones Legislativas

Parcial Computacional - MEFE 2025

Entrega

Subir al campus:

- Un informe en formato PDF.
- El código con el que generó los gráficos y resultados.

En ambos archivos, indicar el nombre, apellido y número de libreta o DNI. La página del curso es: <https://campus.exactas.uba.ar/course/view.php?id=362>. Si aún no habían entrado, la clave de automatriculación es: `completar`

Sobre el informe:

- **Largo:** aproximadamente 1 hoja de texto (500 palabras o 3000 caracteres) + figuras.
- Dar una descripción clara y precisa de la metodología utilizada.
- Mostrar todos los gráficos que considere necesarios y los resultados esperados.
- ¿Cuál es la conclusión?

Contexto

Se acercan las elecciones legislativas en Argentina, y las redes sociales juegan un rol clave en la difusión de información política. Para estudiar sesgos en la cobertura mediática en Twitter, analizamos publicaciones relacionadas con dos coaliciones:

- **Coalición U**
- **Coalición L**

Se clasifican los tweets en dos grandes grupos según el tipo de usuario que los publicó:

- **Grupo A:** usuarios que siguen mayoritariamente a medios públicos o periodistas oficialistas.
- **Grupo B:** usuarios que siguen mayoritariamente a medios opositores o periodistas independientes.

Los tweets se clasifican automáticamente como más favorables a U o a L utilizando palabras clave:

- "U_support": palabras favorables al partido U.
- "L_support": palabras favorables al partido L.

Se construye la siguiente tabla cruzada que resume los tweets analizados:

	Grupo A	Grupo B
U_support	141	307
L_support	373	221

Cuadro 1: Tweets clasificados por tipo de usuario y tendencia política.

Consignas

Se quiere investigar si existe polarización entre los tipos de usuarios y su apoyo en Twitter a una de las dos coaliciones. Para ello se propone realizar dos tests de hipótesis: uno que, a partir de simulaciones, permita reconstruir la distribución de probabilidades condicionales $p(U_support | B)$ y $p(L_support | A)$ bajo la hipótesis de que no existe polarización, y un Test de Contingencia (Test exacto de Fisher, <https://doi.org/10.2307/2340521>, existe la clase `fisher_exact` en `scipy.stats`). Realizá los siguientes pasos:

1. **Simulación por permutación.** Bajo la hipótesis nula de que el tipo de usuario es independiente de a quién apoya el tweet, partiendo de la Tabla 1, generá una simulación que asigne $N = 1000$ veces cada Tweet con un medio de comunicación (sin reposición). Para cada permutación calculá:
 - $p(U_support | B)$: probabilidad condicional que un tweet del Grupo B sea a favor de U.
 - $p(L_support | A)$: probabilidad condicional que un tweet del Grupo A sea a favor de L.

Graficá las distribuciones obtenidas para ambos estimadores.

Ayuda: imaginá que hay dos bolsas, en una de ellas hay $180 + 220$ papeles a favor de U y $110 + 90$ papeles a favor de L, mientras que en la otra bolsa hay $180 + 110$ papeles que dicen que lo publicó A y $220 + 90$ papeles que dicen que lo publicó B. Luego, sacá un papel de cada bolsa para generar un Tweet completo que dice a favor de quién está y además quién lo publicó. Repetir esto hasta agotar los papelitos. Como resultado, se consigue una realización de la Tabla 1 bajo la hipótesis de que no hay correlación entre el Tweet y quién lo publicó. Luego calcular $p(U_support | B)$ y $p(L_support | A)$.

2. **Región crítica.** Determiná las regiones críticas para ambas distribuciones al nivel $\alpha = 0,05$ (percentiles 2,5 y 97,5).
3. **Test de hipótesis.** Calculá los valores reales a partir de la tabla y verificá si caen en las zonas críticas para rechazar o no la hipótesis nula de que no hay polarización.
4. **P-valores y p-valor combinado.** Calculá los p-valores empíricos para $p(U_support | B)$ y $p(L_support | A)$ y combinalos mediante el estadístico propuesto por Fisher para combinar p-valores:

$$\chi_4^2 = -2 \ln(p_1 p_2),$$

y evaluá el p-valor conjunto para decidir si con una significancia $\alpha = 0,05$ se acepta o rechaza la hipótesis nula.

5. **Test exacto de Fisher.** Realizá también el test exacto de Fisher para la tabla y compará su p-valor con el obtenido en el paso anterior.

Referencia: Esta propuesta está basada en un trabajo de Cicchini, del Pozo, Tagliazucchi y Balenzuela (<https://doi.org/10.1140/epjds/s13688-022-00360-8>).