

Parcial computacional sobre Test de Hipótesis: La Grieta

Métodos Estadísticos en Física Experimental

2023

1 Enunciado

Se vienen las elecciones y una vez más mucho de lo que se dice sucede a través de las redes sociales, en particular en Twitter. Con intenciones de testear y cuantificar la famosa Grieta, se cuenta con Tweets de las elecciones 2019. En ellos fue posible clasificar los mensajes de dos grupos de medios según fueron a favor de una u otra de las fórmulas presidenciales que llegaron al ballottage.

Tenemos noticias pertenecientes a dos grupos. El grupo A corresponde a noticias de los diarios La Nación, Clarín e Infobae. El grupo B corresponde a noticias de Página 12 y el Destape. Usando un estadístico (el sesgo de sentimiento -sb-) que se calcula en torno a dos conjuntos de palabras claves, "ff" : palabras relativas a Alberto y Cristina, "mp" : palabras relativas a Macri y Picheto, se clasificaron las noticias de A y de B según hablan mejor de 'ff' que de 'mp'. Se cuenta entonces con la Tabla 1 donde se indica el número de Tweets que pueden considerarse a favor de cada fórmula presidencial discriminado por el medio de comunicación que la publicó.

2 Consignas

Se quiere aplicar un test para estudiar la hipótesis nula de que no existe tal cosa como una Grieta. Para ello, se propone aplicar dos test de hipótesis diferentes, a saber. Un test que a partir de simulaciones permita reconstruir la distribución de las probabilidades condicionales $p(\text{ff}|B)$ y $p(\text{mp}|A)$ reportadas cuando la hipótesis de que no existe polarización es verdadera y un Test de Contingencia (Test exacto de Fisher).

1. **Distribuciones:** A partir de los datos consignados en la Tabla 1, realice una simulación que le permita asignar $N=1000$ veces cada Tweet a favor de una de las fórmulas presidenciales con un medio de comunicación (sin reposición). Para cada una de las N asignaciones calcule las probabilidades condicionales $p(\text{ff}|B)$ y $p(\text{mp}|A)$ y a partir de ellas grafique sus distribuciones.

Ayuda sobre como construir cada una de las N asignaciones: Imagine que tiene dos bolsitas, en una de ellas hay $(139+337)$ papelitos que dicen *a favor de ff* y $(191+116)$ papelitos que dicen *a favor de mp*, mientras que en la otra bolsita, hay $(139+191)$ papelitos que dicen *lo publicó A* y $(337 + 116)$ papelitos que dicen

Table 1: Tweets publicados por los medios A y B que pueden considerarse en favor de las fórmulas presidenciales ff o mp.

	A	B
ff	139	337
mp	191	116

lo publicó B. Luego, saca un papelito de cada bolsa, para conformar un Tweet completo, que dirá si es a favor de ff o de mp y además si lo publicó A o B. Repite esto hasta agostar los papelitos de las bolsas. Como resultado tendrá una realización de la Tabla 1 bajo la hipótesis de que no hay correlación entre el Tweet y quién lo publicó. Luego calcula $p(\text{ff}|B)$ y $p(\text{mp}|A)$.

2. **Definiendo el test:** Sobre la base de las dos distribuciones anteriores esablezca las probabilidades condicionales críticas $p_C(\text{ff}|B)$ y $p_C(\text{mp}|A)$ que determinan las regiones de rechazo con una significancia $\alpha=0.05$.
3. **Aplicando el test:** Ahora, calculando $p(\text{ff}|B)$ y $p(\text{mp}|A)$ a partir de los datos reportados en la Tabla 1 diga en que casos rechaza la hipótesis nula de que no hay polarización.
4. **Potenciando el test:** Calcule el p-valor correspondiente a $p(\text{ff}|B)$ y $p(\text{mp}|A)$ y luego combine ambos p-valores a partir del estadístico $\chi_4 = -2 \ln(\text{p-valor}_1 \text{ p-valor}_2)$. Calcule el p-valor conjunto y diga si acepta o rechaza la hipótesis nula con una significancia $\alpha=0.05$.
5. **Test de Contingencia:** Ahora aplique el test de Fisher. Para ello construya la tabla de contingencia y calcule la probabilidad de todas las tablas posibles sujetas a las condiciones de contorno impuestas por los datos de la Tabla 1. Calcule el p-valor correspondiente para la tabla de contingencia que devuelve las probabilidades condicionales reportadas en la Tabla 1. Compare con el resultado obtenido utilizando el test 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>).

3 Generalidades

1. Indicar nombre, apellido, libreta o DNI en el informe.
2. Dar una descripción clara y precisa de la metodología utilizada.
3. Incluir todos los gráficos como figuras con sus correspondientes leyendas.
4. Justifique las hipótesis en la que se sustenten sus resultados y discuta los resultados obtenidos.
5. La fecha límite para la entrega es el martes 11 de julio a las 11 hs.

6. Enviar por mail a dariorodriguesfm@gmail.com. Nombrar al archivo de la siguiente manera: TH-SuApellido.pdf y adjuntar en el mismo mail los códigos desarrollados. Utilizar como asunto del mail: "TH - Afinidad de género".
7. Por consultas sobre interpretación de los enunciados escribir al grupo: mefe-1c-2023@googlegroups.com, así todos tienen acceso a todas las respuestas.