

Parcial computacional sobre Test de Hipótesis: Afinidad de Género

Métodos Estadísticos en Física Experimental

2023

1 Enunciado

El mundo de la física se divide en áreas, donde en algunas hay mayor proporción de mujeres trabajando que en otras ¿se imaginan cuáles?. Esta tendencia se suma a que existe un altísimo grado de homofilia en los trabajos publicados (investigadoras buscan colaborar más con otras investigadoras mientras que investigadores hombres hacen lo propio).

Uno de los factores que se postula que contribuyen a este fenómeno es que las becarias que entran al ámbito académico eligen hacerlo en grupos con mayor proporción de mujeres, donde se sienten más escuchadas, seguras y bienvenidas. Podemos testear si esto se replica acá. Para esto, se cuenta con el número de candidatas (m) y candidatos (h) en convocatorias a becas doctorales en física del CONICET de los últimos cuatro años, y se cuenta además con el número de directoras (M) y directores (H) propuestos para dichas becas cada año.

Se conoce además, para cada año, cual fue la probabilidad de que la directora propuesta sea mujer cuando la candidata a beca es mujer $p(M|m)$ y que el director propuesto sea hombre cuando el candidato a beca es hombre y $p(H|h)$.

2 Consignas

Se quiere aplicar un test de hipótesis para estudiar afinidad de género. Es decir, si es más probable que las candidatas mujeres se presenten con directoras mujeres y los candidatos varones con directores varones. 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 reportadas cuando la hipótesis de que no existe afinidad de género 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 $N=10000$ veces cada candidate con una directore (sin reposición), para cada una de las N asignaciones calcule las probabilidades condicionales $p(M|m)$ y $p(H|h)$ y grafique sus distribuciones.

Ayuda sobre como construir cada una de las N asignaciones para el año 2019: Imagine que tiene dos bolsitas, en una de ellas hay 21 papelitos que dicen *candidata mujer* y 49 papelitos que dicen *candidato hombre*, mientras que en la otra bolsita,

	2019	2020	2021	2022
m	21	11	11	15
h	49	41	50	58
M	16	8	9	13
H	54	44	52	60
p(M m)	0.3809	0.5455	0.2727	0.2667
p(H h)	0.8367	0.9512	0.8800	0.8448

Table 1: Número de postulantes a becas doctorales del CONICET en física mujeres (m) y hombres (h) y número de directoras mujeres (M) y directores hombres (H) propuestos para los años 2019, 2020, 2021 y 2022. Se reportan además las probabilidades condicionales observadas $p(M|m)$ y $p(H|h)$ para cada año.

hay 16 papelitos que dicen *directora mujer* y 54 papelitos que dicen *director hombre*. Luego, sacas un papelito de cada bolsa para conformar una postulación a beca, que dirá si es de un candidato o candidata a beca y además si se propone un director o directora. Repetis esto hasta agostar los papelitos de las bolsas. Como resultado tendrás una realización de la Tabla 1 bajo la hipótesis de que no hay correlación entre candidate y directorx. Luego calculas $p(M|m)$ y $p(H|h)$.

2. **Definiendo el test:** Sobre la base de las dos distribuciones anteriores establezca las probabilidades condicionales críticas $p_C(M|m)$ y $p_C(H|h)$ que determinan las regiones de rechazo con una significancia $\alpha=0.05$.
3. **Aplicando el test:** Ahora, utilizando las probabilidades condicionales reportadas en la Tabla 1 y las distribuciones obtenidas diga, año a año, y por género, en que casos rechaza la hipótesis nula de que no hay afinidad de género.
4. **Potenciando el test:** Ahora calcule el p-valor correspondiente a cada probabilidad condicional observada utilizando las distribuciones simuladas y luego combine todos los p-valores a partir del estadístico $\chi_{2n} = -2 \ln(\prod_i^n p_i)$. 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.

Consideración importante El armado de la Tabla 1 se hizo clasificando el género de forma binaria tanto para aspirantes como directorxs. La asignación se realizó utilizando el nombre de pila mediante un algoritmo para asignación de género. Este es un punto a mejorar en futuros análisis donde se cuente con mayor información que haga posible una asignación no binaria.

Referencia Esta propuesta está basada en un trabajo que hicieron integrantes del DF-FCEN, Laura Estrada (lestrada@df.uba.ar) y Tomas Cicchini (tomas.cicchini@gmail.com)

como parte de las actividades de relevamiento que realiza la Subcomisión de género de la Asociación de Física Argentina (generoafa.wixsite.com/subcomision).

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.