

LA PRUEBA DE PROBABILIDAD EXACTA DE FISHER

Cuando el tamaño de la muestra es pequeño y el número de frecuencias de las celdas de la tabla de contingencia es chico (menor a 5, incluso de 0) se prefiere la prueba de Probabilidad Exacta de Fisher¹.

Se usa cuando dos puntuaciones de dos muestras caen dentro de una de dos clases mutuamente excluyentes y, partiendo del supuesto de independencia entre las variables, calcula la diferencia entre los datos observados y los datos esperados considerando los marginales.

La prueba determina si los dos grupos difieren en las proporciones en donde caen dentro de cualquiera de las clasificaciones. En otras palabras, evalúa la probabilidad de obtener una tabla tan fuerte o más fuerte que la observada, debido simplemente a lo azaroso de la muestra.

Dicho de otra forma, dados los marginales, y suponiendo que en la población las variables en cuestión no están relacionadas, ¿qué tan probable es obtener tablas con frecuencias más extremas o igual de extremas a las frecuencias observadas en la tabla actual?

Ahora bien, para poder calcular p , la probabilidad de obtener una tabla tan o más fuerte que la observada, se requiere computar la formula de la prueba que es:

$$p^2 = \frac{(r_1!r_2!c_1!c_2!)}{(n!a!b!c!d!)}$$

Para tener una idea más clara de la notación véase la tabla A, donde a , b , c y d son entradas para las celdas, r y c se refieren a los totales marginales de las filas y las columnas respectivamente, y n es el tamaño de la muestra. Asimismo, se presenta una tabla observada (tabla B) y todas las tablas posibles más fuertes (tablas C y D).

a. Notación

a	b	r_1
c	d	r_2
c_1	c_2	n

b. Valores observados

7	2	9
5	6	11
12	8	20

c. Más fuerte

8	1	9
4	7	11
12	8	20

d. La más fuerte

9	0	9
3	8	11
12	8	20

¹ Cabe aclarar que en la actualidad es posible realizar cálculos para esta prueba con tablas $[r \times c]$ y con frecuencia relativamente altas. Algunos programas estadísticos llevan a cabo esta prueba siempre y cuando $n = 100$.

² p es igual al factorial de r_1 por el factorial de r_2 por el factorial de c_1 por el factorial de c_2 entre el factorial de n por el factorial de a por el factorial de b por el factorial de c por el factorial de d . La notación matemática para un número factorial es $!$, y se refiere a la operación aritmética de multiplicar $1 \times 2 \times 3 \times n$, para $n!$. Así, por ejemplo, se tiene que $8! = 40320$. Por convención $0!=1$ y $1!=1$.

E. Valores de p

$p_b = 9!11!12!8! / 20!7!2!5!6! =$.132
$p_c = 9!11!12!8! / 20!8!1!4!7! =$.024
$p_d = 9!11!12!8! / 20!9!0!3!8! =$.001
$p_{total} =$.157

Se realiza el cálculo de p , presentado en la tabla E, y queda sólo por interpretar el resultado obtenido. Como se ha comentado anteriormente .05 se considera el umbral mínimo aceptable de nivel de significación y dado el resultado obtenido ($p = .157$) se puede concluir que la distribución que presenta la tabla B no es significativamente diferente debido simplemente a lo azaroso de la muestra, esto es *no hay evidencia* suficiente que permita descartar la H_0 .

Por último, la prueba de Probabilidad Exacta de Fisher funciona igual que la prueba de la Ji Cuadrada, sin embargo, ésta última sólo proporciona un estimado del valor de probabilidad verdadero, un estimado que no es muy preciso si los marginales son muy desiguales o si hay valores muy pequeños en una de las celdas.

EL VALOR-P (O LA PROBABILIDAD CALCULADA)

Es la probabilidad estimada de rechazar la H_0 cuando esa hipótesis es verdadera. Es una medida de la fuerza de los resultados, más allá de sólo aceptar o rechazar una hipótesis.

Muestra cuanta evidencia se tiene en contra de la H_0 y, normalmente se puede interpretar con el nivel de significación $[\alpha]$ que convencionalmente es de $p < 0.05$, esto es 5% -menos de 1 probabilidad en 20 de estar equivocado, o, para un nivel más riguroso como $p < 0.001$, es decir 0.1% -menos de una probabilidad en mil de estar equivocado- de tal suerte que si el valor-p es menor a alguno de los niveles de umbral establecidos se puede rechazar la H_0 .

El nivel de rechazo puede ser más o menos fuerte en función del valor-p obtenido del análisis estadístico. Al respecto algunos estudiosos sugieren una interpretación como la que se muestra en la tabla F y otros prefieren hablar sólo de significación estadística³ -tabla G.

³ Entendida como la probabilidad de ocurrencia que va más allá que la simple suerte o coincidencia

F. Interpretación del valor-p

<i>P-value</i>	<i>Interpretation</i>
$P < 0.01$	Very strong evidence against H_0
$0.01 \leq P < 0.05$	Moderate evidence against H_0
$0.05 \leq P < 0.10$	Suggestive evidence against H_0
$0.10 \leq P$	Little or no real evidence against H_0

G. La significación estadística del valor-p

<i>P value</i>	<i>Wording</i>
>0.05	Not significant
0.01 to 0.05	Significant
0.001 to 0.01	Very significant
<0.001	Extremely significant

Resalta el hecho que no se incluyen todos los niveles de significación en ambas tablas⁴: la tabla G. incluye el nivel más riguroso ($p < 0.001$) y la tabla F se queda en el nivel anterior ($p < 0.01$). Esta última incluye el nivel (0.10), que normalmente es considerado fuera del umbral mínimo para establecer algún tipo de significación estadística (0.05). No obstante, si se toma el valor 0.01 como evidencia muy fuerte en contra de la H_0 , es evidente que esta conclusión se puede extender para el valor p de 0.001.

⁴ Recuérdese que los niveles de significación (empezando por el nivel umbral mínimo y así hasta el nivel más riguroso) son: [0.05], [0.025], [0.01], [0.005] y [0.001].

NUMEROS FACTORIALES

1! = 1
2! = 2
3! = 6
4! = 24
5! = 120
6! = 720
7! = 5040
8! = 40320
9! = 362880
10! = 3628800
11! = 39916800
12! = 479001600
13! = 6227020800
14! = 87178291200
15! = 1307674368000
16! = 20922789888000
17! = 355687428096000
18! = 6402373705728000
19! = 121645100408832000
20! = 2432902008176640000
21! = 51090942171709440000
22! = 112400072777607680000
23! = 25852016738884976640000
24! = 620448401733239439360000
25! = 15511210043330985984000000
26! = 403291461126605635584000000
27! = 10888869450418352160768000000
28! = 304888344611713860501504000000
29! = 8841761993739701954543616000000
30! = 265252859812191058636308480000000
31! = 8222838654177922817725562880000000
32! = 263130836933693530167218012160000000
33! = 8683317618811886495518194401280000000
34! = 295232799039604140847618609643520000000
35! = 10333147966386144929666651337523200000000
36! = 371993326789901217467999448150835200000000
37! = 13763753091226345046315979581580902400000000
38! = 523022617466601111760007224100074291200000000
39! = 20397882081197443358640281739902897356800000000
40! = 815915283247897734345611269596115894272000000000
41! = 33452526613163807108170062053440751665152000000000
42! = 1405006117752879898543142606244511569936384000000000
43! = 60415263063373835637355132068513997507264512000000000
44! = 2658271574788448768043625811014615890319638528000000000
45! = 119622220865480194561963161495657715064383733760000000000
46! = 5502622159812088949850305428800254892961651752960000000000
47! = 258623241511168180642964355153611979969197632389120000000000
48! = 12413915592536072670862289047373375038521486354677760000000000
49! = 608281864034267560872252163321295376887552831379210240000000000
50! = 30414093201713378043612608166064768844377641568960512000000000000

Ejemplo1 Hostigamiento

En un estudio acerca de las situaciones en las cuales las personas amenazan con suicidarse saltando desde un edificio, un puente o una torre, se advirtió que el abucheo o el hostigamiento por parte de la multitud como espectadora ocurría sólo en algunos casos. Varias teorías proponen que un estudio psicológico de disminución de la identidad y la autoconciencia, conocido como *deindividuación*, puede contribuir al fenómeno de hostigamiento. Se conocen algunos factores que inducen reacciones en las multitudes, incluidos, la temperatura, el ruido y la fatiga. En un esfuerzo por evaluar varias hipótesis concernientes al hostigamiento por parte de las multitudes, un investigador revisó 21 artículos publicados y el mes del año; esto último se refería más bien al índice de temperatura. La hipótesis es que habrá un incremento en el hostigamiento por parte de los espectadores cuando hiciera calor.

Los datos extraídos de los artículos se muestran en la tabla de contingencia [2x 2] a continuación.

1. Incidencia de hostigamiento en episodios de intento de suicidio

Mes	Multitud		
	Hostigamiento	No hostigamiento	Combinación
Junio-Septiembre	8	4	12
Octubre- Mayo	2	7	9
Total	10	11	21

Pasos a seguir

1. Determinar la hipótesis nula y la hipótesis alterna.

H_0 : El hostigamiento por parte de las multitudes no varía durante los meses calurosos.

H_1 : El hostigamiento por parte de las multitudes aumenta durante los meses calurosos

2. Determinar la prueba estadística a utilizar

Prueba de probabilidad exacta de Fisher.

3. Determinar el nivel de significación α .

$$\alpha = 0.01$$

4. Región de rechazo. Puesto que H_1 predice la dirección de la diferencia entre los grupos, la región de rechazo es unidireccional. H_0 será rechazada si los valores de la celda observados difieren en la dirección predicha y si son de tal magnitud que la que la probabilidad asociada con su ocurrencia cuando H_0 es verdadera, es igual o menor que α .

5. Decisión. La información que arrojan los artículos muestra que hubo 10 multitudes que hostigaron a los suicidas y 11 multitudes que no lo hicieron. La revisión de la tabla muestra que **existen 2 tablas adicionales** que producirían un resultado más extremo (unidireccional). Entonces, la probabilidad de observar un conjunto de frecuencias de celdas tan extremas o más extremas que la actualmente observada se determina al utilizar **la ecuación de la prueba exacta de Fisher**.

Importante

La decisión sobre la direccionalidad debe tomarse antes de recolectar los datos.

Izquierda: Se usa esta direccionalidad cuando la alternativa hacia la independencia es que hay una asociación negativa entre las variables. En otras palabras, las observaciones tienden a caer en la parte baja de la izquierda y en la parte alta de la derecha de la tabla de contingencia.

Derecha: Se usa esta direccionalidad cuando la alternativa hacia la independencia es que hay una asociación positiva entre las variables. En otras palabras, las observaciones tienden a caer en la parte alta de la izquierda y en la parte baja de la derecha de la tabla de contingencia.

Bidireccional: Se usa cuando no hay una alternativa previa a la recolección de los datos.

Dato obtenidos

A. Incidencia de hostigamiento en episodios de intento de suicidio

Mes	Multitud		
	Hostigamiento	No hostigamiento	Combinación
Junio-Septiembre	8	4	12
Octubre- Mayo	2	7	9
Total	10	11	21

Tablas más extremas

B. Primera tabla más extrema

MES DEL AÑO	HOSTIGAMIENTO	NO HOSTIGAMIENTO	Total
JUNIO-SEPTIEMBRE	9	3	12
OCTUBRE-MAYO	1	8	9
Total	10	11	21

C. Segunda tabla más extrema

MES DEL AÑO	HOSTIGAMIENTO	NO HOSTIGAMIENTO	Total
JUNIO-SEPTIEMBRE	10	2	12
OCTUBRE-MAYO	0	9	9
Total	10	11	21

Valores de P de cada tabla de contingencia

$$P_A = \frac{12! \cdot 9! \cdot 10! \cdot 11!}{21! \cdot 8! \cdot 4! \cdot 2! \cdot 7!} = 0.0505$$

$$P_B = \frac{12! \cdot 9! \cdot 10! \cdot 11!}{21! \cdot 9! \cdot 3! \cdot 1! \cdot 8!} = 0.0056$$

$$P_C = \frac{12! \cdot 9! \cdot 10! \cdot 11!}{21! \cdot 10! \cdot 2! \cdot 0! \cdot 9!} = 0.0002$$

$$P_A (0.0505) + P_B (0.0056) + P_C (0.0002) = 0.0563$$

$$P = 0.0563$$

Conclusión:

La conclusión a la que llegó el investigador⁵ es la siguiente:

Puesto que la probabilidad obtenida (0.0563) es menor que el nivel de significación seleccionado [$\alpha = 0.10$] debemos rechazar H_0 a favor de H_1 . Concluimos que el hostigamiento por parte de las multitudes que atestiguan una amenaza de suicidio **es afectado** por la temperatura (medida según el mes del año).

Observación:

El nivel de significación utilizado por el investigador está por encima del nivel considerado como umbral para poder hablar de significación estadística.

Preguntas para resolver:

1. Si el investigador hubiera partido de este último nivel (0.05) que es el nivel inmediato inferior al utilizado originalmente, ¿hubiese cambiado su interpretación?

Efectivamente hubiera cambiado su interpretación debido a que la probabilidad obtenida fue de 0.0563, esto es, es mayor al siguiente nivel de significación [0.05]. En este caso no hubiera tenido evidencia suficiente para rechazar la H_0

2. ¿Por qué hay sólo dos tablas más fuertes con respecto de los datos recabados por el investigador?

Porque sólo dos tablas adicionales permiten observar datos tan extremos o más que los observados (sin alterar los marginales totales y, por lo tanto, n)

⁵ Mann, L., "The baiting crowd episode of threatened suicide", en *Journal of Personality and Social Psychology*, núm. 41, 1981, pp. 703-709.

3. *¿Por qué H_1 predice la dirección de la diferencia entre los grupos?*

La H_1 establece que existe un incremento en el hostigamiento por parte de las multitudes durante los meses calurosos. Esto significa que el investigador está esperando que en la columna de hostigamiento en los meses de junio-septiembre se concentre el mayor número de multitudes y, a la inversa, en los meses de octubre-mayo el no hostigamiento aumente de forma sustancial.

4. *¿Es posible sostener que los valores de las celdas observados difieren en la dirección predicha? ¿Cuál es esa dirección?*

En el ejemplo en cuestión, los valores observados de las celdas difirieron en la dirección predicha (pasaron de 8 a 9 y de 9 a 10 en la parte alta de la izquierda y de 7 a 8 y de 8 a 9 en la parte baja de la derecha) y además fueron de tal magnitud que la probabilidad asociada con su ocurrencia (o la ocurrencia de las tablas más extremas) fue menor al nivel de significación previamente seleccionado [$\alpha = 0.10$]. La direccionalidad es hacia la derecha.

Ejemplo 2⁶

Se cree que fumar está relacionado con padecer cáncer de vejiga. En una muestra aleatoria de 14 personas se obtienen los siguientes datos:

		FUMA		
		Sí	No	
Cáncer de vejiga	No	1	4	5
	Sí	7	2	9
		8	6	14

1. Escriba las hipótesis (para una prueba unidireccional)
2. Calcule el valor de p para los valores observados
3. Calcule el valor de p para cada una de las tablas más extremas posibles con los mismos marginales totales y el mismo gran total.
4. Ahora sume las probabilidades de la tabla con los valores observados y aquellas que son aún más favorables que ella para determinar si hay significación estadística a un nivel alfa de 0.05
5. Plantee sus conclusiones.

Ahora suponga que el planteamiento de las hipótesis es bidireccional.

1. Vuelva a escribir sus hipótesis
2. Vuelva a calcular el valor de p . En este caso la probabilidad se calcula de la misma manera que antes, pero en lugar de sumar se multiplica por dos. Aquí también solo se consideran los valores de p más favorables que el obtenido para la tabla con los valores observados.
3. Plantee sus conclusiones.

⁶ Tomado de: Álvarez C. R., (2007) *Estadística aplicada a las ciencias sociales*. España: Díaz de Santos.