ReaR
ISNN 1989 4090
Revista electrónica de AnestesiaR
Mayo 2021

ORIGINAL

Horror vacui. Corrección de continuidad de Yates.

Molina Arias M.

Hospital Infantil Universitario La Paz, Madrid.

Resumen

Con frecuencia utilizamos aproximaciones con distribuciones de probabilidad continuas para resolver contrastes de hipótesis con variables que siguen una distribución discreta. En estos casos, debemos aplicar una corrección de continuidad, siendo la más conocida la corrección de continuidad de Yates.

Introducción

Con frecuencia utilizamos aproximaciones con distribuciones de probabilidad continuas para resolver contrastes de hipótesis con variables que siguen una distribución discreta. En estos casos, debemos aplicar una corrección de continuidad, siendo la más conocida la corrección de continuidad de Yates.

A lo largo de la historia del arte nos encontramos de forma repetida con el *horror vacui* que, para aquellos que no fueseis tan afortunados de estudiar latín en vuestros años jóvenes, no es otra cosa que el miedo al vacío.

Hay numerosos ejemplos de obras de arte en los que puede verse un empeño obsesivo en rellenar todo el espacio con algún elemento, no dejando nada vacío. Pensad en la decoración islámica o en las obras del periodo rococó o, sobre todo, en la recargada decoración de la época victoriana.

¿Y por qué os cuento todo esto? Pues porque me lo ha recordado el tema de hoy, que tiene que ver con las distribuciones de probabilidad discretas y continuas y de cómo las primeras permiten este vacío y las segundas no, y de cómo se complica la cosa cuando usamos unas para aproximar otras. Esto parece un trabalenguas, pero que nadie desespere, vamos a ver si lo aclaramos.

Una prueba muy popular, pero aproximada

Probablemente la prueba de contraste de hipótesis más frecuentemente utilizada sea la prueba de la ji-cuadrado de independencia, que utilizaremos para comparar las proporciones de dos variables cualitativas y tratar de determinar si ambas variables están asociadas o son independientes.

Como todos sabemos, construimos una tabla de contingencia con los valores observados, calculamos los valores esperados bajo el supuesto de la hipótesis nula de que las dos variables son independientes y, por último,

calculamos la probabilidad (bajo la hipótesis nula) de observar por azar una tabla tan alejada o más de la teórica que la que hemos observado en nuestro experimento.

El problema surge con el uso indiscriminado de la prueba que, a veces, nos lleva a olvidar que el estadístico que utilizamos para el contraste, la ji-cuadrado, sigue una distribución aproximada que solo es útil cuando el número de observaciones es relativamente grande, pero que pierde efectividad cuando la información de que disponemos es escasa, lo que ocurre con cierta frecuencia.

Por eso, una vez construida la tabla de contingencia, comprobamos que no haya celdas con frecuencias menores de 5. Si esto ocurre, tenemos dos formas de solucionar el problema.

Mejor una prueba exacta

La primera forma de solucionarlo es utilizar una prueba exacta, como la prueba exacta de Fisher.

Las pruebas exactas calculan la probabilidad de forma directa, generando para ello todos los escenarios posibles en los que se produce la condición que queremos estudiar. Esto se hace construyendo todas las tablas de contingencia más extremas que la observada y que cumplen con la dirección de la asociación de la tabla observada.

Una vez calculada esta probabilidad exacta, se comparará con el nivel de significación estadística y se procederá a resolver el contraste de hipótesis.

El problema de estos métodos es que son mucho más laboriosos, lo que ha dificultado su mayor utilización hasta disponer de la potencia de cálculo necesaria. Esto explica la predilección por el uso de las pruebas aproximadas como la de la ji-cuadrado.

La corrección de continuidad de Yates

Dijimos que había dos formas de solucionar el problema de los datos escasos. Pues la segunda forma es aplicar la corrección de continuidad de Yates, que supone restar 0,5 a la diferencia entre valores observados y esperados al calcular el valor del estadístico ji-cuadrado.

Todo el mundo conoce la corrección de Yates, tan popular como la prueba de la ji-cuadrado, no cabe duda. Pero que levanten la mano aquellos que sepan qué es exactamente una corrección de continuidad como, por ejemplo, la de Yates.

Para entenderlo bien, primero tenemos que saber con qué tipo de distribución de probabilidad estamos tratando.

Distribuciones continuas y discretas

Las variables cuantitativas pueden ser continuas y discretas. Una variable es continua cuando, entre dos valores de la variable, existen infinitos (al menos, en teoría) valores posibles. Por ejemplo, pensemos en el peso de un recién nacido. Puede pesar 3 kg y puede pesar, digamos, 4 kg. Pero entre los 3 y los 4 kg hay infinitos valores posibles de peso (aunque en la práctica este infinito se limita al número que nos permita la precisión de nuestra báscula).

Ahora pensemos en el número de hijos. Uno puede tener 2, tener 3, o un número diferente, pero lo que no puede es tener un número de hijos entre dos y tres, por ejemplo 2,5 (ya sé que a veces vemos este tipo de cosas, pero es un recurso que facilita el análisis de la variable pero carece de sentido desde el punto de vista de la vida cotidiana).

Lo mismo ocurre con las distribuciones de probabilidad. Entre los valores 3 y 4 de una distribución de probabilidad discreta hay un vacío completo. Sin embargo, las distribuciones continuas son como un dormitorio victoriano y padecen de *horror vacui*: entre 3 y 4 hay todo un intervalo de valores posibles.

Esto, en sí, no supone ningún problema. El problema surge cuando tenemos un contraste que precisaría para su resolución el usar una distribución discreta y realizamos una aproximación utilizando una distribución continua. Veamos algún ejemplo.

Imaginemos que trabajamos con una distribución discreta, por ejemplo, una binomial definida por n y p: B(n,p). Resulta muy habitual que, para simplificar los cálculos de probabilidad, cuando el tamaño muestral es grande y la probabilidad del evento está alrededor de 0,5, aproximemos la solución mediante una normal. De ahí viene lo de que cuando np y n(1-p) son mayores de 5 podemos aproximar la binomial con una normal de media np y varianza igual a la raíz cuadrada de np(1-p).

Esto nos facilita los cálculos, pero estamos pasando de usar una distribución discreta a usar una continua, lo que tiene sus consecuencias, como veremos.

En una discreta, obtener la probabilidad de x>3 es sencillo. En una continua, la cosa se complica, ya que pasamos del vacío entre dos puntos de la discreta al intervalo lleno de valores posibles de la distribución continua.

Volvamos al cálculo de la probabilidad de que la variable valga más de 3: P(x>3). De 3 para abajo no hay problema, ya sea continua o discreta. De 4 para arriba, tampoco hay problema. Pero entre 3 y 4 antes había un vacío que ahora se ha llenado. ¿Cómo lo solucionamos? Pues dándole la mitad del intervalo a cada sección de la distribución por arriba y por debajo del valor. De esta forma, P(x>3) se calcularía en la aproximación normal como P(x≥3,5), incluyendo la mitad del intervalo por encima de 3, que no está incluido en el cálculo de probabilidad. Y, a la chita callando, acabamos de aplicar la corrección de continuidad de Yates.

Si queremos calcular la P(x≥3), el cálculo incluiría el 3, así que tendríamos que irnos a la mitad anterior del intervalo vacío y la calcularíamos como P(x≥2,5). Siguiendo el mismo razonamiento, la P(x≤3) = p(x≤3,5), incluyendo la mitad del intervalo por encima de 3. ¿Y la probabilidad de que x sea igual a 3? Habrá que tomar las dos partes del intervalo: P(2,5≤x≤3,5).

Dos errores para evitar

Ya hemos visto, pues, que aplicaremos la corrección de continuidad cuando queramos pasar de una distribución discreta a una continua. Cuando trabajamos con variables que siguen una distribución continua, no hay que aplicar ninguna corrección. Por ejemplo, si en una distribución normal queremos calcular P(x=3), que a nadie se le ocurra calcular la probabilidad del intervalo de 2,5 a 3,5. En este supuesto es erróneo aplicar la corrección de continuidad. La P(x=3) en una distribución normal es igual a cero. Si lo pensamos, la probabilidad es el área bajo la curva y, debajo de un punto, no hay área.

Tampoco hace falta cuando pasamos de una distribución discreta a otra también discreta. Un ejemplo puede ser cuando aproximamos una binomial con una distribución de Poisson (cuando np<5). Solo hay que aplicar la corrección de

continuidad al pasar de discreta a continua.

Volviendo a la ji-cuadrado

Ahora que ya sabemos qué es una corrección de continuidad, vamos a ver la razón por la que hay que aplicarla cuando las frecuencias de las celdas de la tabla de la ji-cuadrado son bajos.

La probabilidad exacta con muestras pequeñas se calcula utilizando distribuciones discretas de probabilidad, tales como la hipergeométrica, la binominal negativa y otras que podemos elegir en función del muestreo de los datos. Cuando la muestra es pequeña y aproximamos con la prueba de la ji-cuadrado, estamos haciendo una aproximación con una distribución de probabilidad conocida, la distribución de la ji-cuadrado que, los que estéis aun despiertos ya habréis adivinado, es una distribución de probabilidad continua.

Pasamos de discreta a continua, luego tenemos que aplicar la corrección de continuidad. Así intentamos compensar los desajustes que se producen cuando la distribución de probabilidad de las frecuencias observadas, que es discreta, es aproximada por otra de carácter continuo.

Nos vamos…

Y ya vamos a ir terminando por hoy.

Antes de irnos, solo quiero decir que no todas las expresiones artísticas pecan de este *horror vacui*. A veces, algunos artistas hacen lo contrario y usan el vacío para transmitir su mensaje. Esto es muy frecuente en fotografía, con el uso del denominado espacio negativo.

Hemos hablado todo el tiempo de la corrección del amigo Yates, que es la más conocida. Pero no penséis que es la única. Hay más, como la de Cochran o la de Mantel. Pero esa es otra historia…

Bibliografía

– Toledo E, Sánchez-Villegas A, Martínez-González MA. Probabilidad. Distribuciones de probabilidad. En: Martínez-González MA, Sánchez-Villegas A, Toledo EA, Faulin J, eds. Bioestadística amigable, 3ª ed. El sevier España SL, Barcelona;2014: 65-100. (PDF)

– Toledo E, Núñez-Córdoba JM, Martínez-González MA. Datos categóricos y porcentajes: comparación de proporciones. En: Martínez-González MA, Sánchez-Villegas A, Toledo EA, Faulin J, eds. Bioestadística amigable, 3ª ed. El sevier España SL, Barcelona;2014: 147-73. (PDF)

– Montero JM, Fernández-Avilés G. Contraste de independencia con corrección de continuidad asimétrica. Una aplicación al turismo cultural. Anales de ASEPUMA. 2010;18:801. (PDF)

Correspondencia al autor

Manuel Molina Arias
mma1961@gmail.com
Servicio de Gastroenterología.
Hospital Infantil Universitario La Paz.
Madrid. España.

Aceptado para el blog en octubre de 2020