

Capítulo 3: Contrastes de hipótesis

Aunque gran parte de la investigación médica esta relacionada con la recogida de datos con propósito descriptivo, otra buena parte lo está con la recolección de información con el plan de responder a cuestiones puntuales, es decir, **contrastes de hipótesis**.

Las técnicas de contrastes de hipótesis están muy vinculadas a las de cálculo de intervalos de confianza, como mencionábamos al hablar de estos últimos. Aunque la aproximación es diferente:

- Al hacer un intervalo de confianza establecemos una región donde esperamos que esté el valor del parámetro.
- Al hacer un contraste de hipótesis establecemos posibles valores para unos parámetros y calculamos la probabilidad de que se obtengan muestras tan discrepantes o más que la obtenida, bajo la suposición de que la hipótesis es cierta. Si dicha probabilidad es muy baja (por debajo de una cantidad denominada *nivel de significación*) la hipótesis es rechazada.

Vamos a ilustrarlo retomando un ejemplo del tema de intervalos de confianza.

Supongamos que con un método tradicional, los pacientes afirman no sufrir trastornos con un tratamiento en el 50% de los casos, pero con un nuevo tratamiento probado sobre 80 personas, 60 afirman no haberlos sufrido. Previamente habíamos calculado que un intervalo de confianza al 95% sería que hay un 75% de individuos sin trastornos con un margen de error de 9.6%. Este intervalo parece indicarnos que el tratamiento es preferible ya que claramente ofrece resultados mejores (el valor 50% no está contenido en el intervalo de confianza).

En términos de contraste de hipótesis resolveríamos la cuestión con cálculos muy parecidos pero se enunciaría de otra manera. Diríamos que si fuese cierta la hipótesis de que el nuevo tratamiento provoca trastornos con frecuencia similar al 50%, la muestra obtenida sería muy improbable. Lo que nos llevaría a rechazar dicha hipótesis, con un nivel de significación del $100\% - 95\% = 5\%$. Como se aprecia, tiene una interpretación clínica más inmediata el resolver una cuestión con un intervalo de confianza que con un contraste de hipótesis. Será mejor hacerlo así donde se pueda.

3.1 ¿Qué es una hipótesis?

Las hipótesis que se contrastan hay que entenderlas como una declaración, no como una pregunta a responder. No se formulan como la cuestión “¿El tratamiento tiene efecto?”. Se formulan bien como “El tratamiento no tiene efecto” o como “el tratamiento sí tiene efecto”. A una hipótesis que traduce las ideas de “no hay efecto”, “no hay relación”, “los resultados en los grupos son similares”, se la denomina hipótesis nula. La hipótesis alternativa es la que se deduce al rechazar a la primera.

Las hipótesis se formularán normalmente como una declaración sobre una o más poblaciones, especialmente sobre sus parámetros. Ejemplos de hipótesis pueden ser:

- La edad media de los individuos con la enfermedad X es 50 años.
- El tratamiento X tiene el mismo efecto que un placebo (por tanto no tiene valor terapéutico). Se podría traducir como que el efecto medio del placebo es igual al efecto medio del tratamiento.

- Para determinar la cantidad de anestésico a aplicar a un paciente, se utiliza un modelo de regresión lineal, siendo el peso y la edad del mismo variables independientes. Alguna hipótesis podría ser que no deberíamos tener en cuenta la edad, por ejemplo. Esto se traduciría en que el coeficiente asociado en el modelo de regresión es cero.
- Al tratar de identificar factores de riesgo para el desarrollo de una enfermedad podríamos estar interesados en saber si fumar es uno de ellos. Para ello podríamos realizar una regresión logística y contrastar si la “odds ratio” es igual a uno, lo que equivale a decir que no es un factor de riesgo.

En todo contraste de hipótesis se enfrenta la denominada hipótesis nula (“no efecto”), frente a la hipótesis alternativa, que la niega. La hipótesis nula puede interpretarse como aquella que normalmente sería aceptada mientras los datos no indiquen otra cosa. Cuando los datos se muestran contrarios a la hipótesis nula a la vez que favorables a la hipótesis alternativa, se rechaza la nula y se acepta la alternativa. El mecanismo no puede ser más simple, pero hay que tener cuidado con interpretarlo bien, pues pueden esconderse trampas de no emplearse con atención.

Un contraste se declara como **estadísticamente significativo**, cuando a partir de los resultados muestrales concluimos que se rechaza la hipótesis nula.

En los capítulos siguientes trataremos contrastes de hipótesis de frecuente uso en la ciencia médica, indicando en cada caso la hipótesis nula y la alternativa.

3.2 Tipos de error, significación, nivel de significación y potencia

Al realizar un contraste de hipótesis hay dos tipos de errores posibles:

- Rechazar la hipótesis nula, cuando esta es cierta. A esto se le denomina error de tipo I.
- No rechazar la hipótesis nula, cuando esta es falsa. Lo denominamos error de tipo II, y a la probabilidad de que ocurra se la denomina β .

3.2.1 Nivel de significación

Se denomina nivel de significación de un contraste, a la probabilidad de cometer un error de tipo I. La probabilidad de que este evento ocurra, es una cantidad que se fija de antemano (antes incluso de extraer las muestras) en un número pequeño denominado nivel de significación, y se representa con la letra α . Típicamente se elige un valor pequeño, 5% o 1%. Todo experimento, en su definición y antes de elegir las muestras, debe llevar descrito cuál es el criterio con el que rechazaremos una hipótesis. Esto se traduce en fijar el nivel de significación del contraste.

Al leer un artículo, en la sección “material y métodos”, es frecuente encontrárselo referenciado en frases del estilo: “Se declararán significativos los contrastes cuando la significación sea inferior al 5%”. Esta frase debe entenderse como que se rechazará la hipótesis nula del contraste si, al examinar la muestra, se observa que discrepa tanto de la hipótesis nula, que si esta fuese realmente cierta, la probabilidad de obtener una muestra como la obtenida (o aún peor), es inferior al 5%. Dicho con otros términos,

“Declaramos que no hay efecto, hasta que los datos nos indique otra cosa más allá de una duda razonable”.

3.2.2 Significación

Se denomina significación de un contraste, y se suele representar con la letra p , al valor calculado tras observar la muestra, y que debería tomar el nivel de significación para estar en una situación donde dudemos entre rechazar o no una hipótesis.

Visto así, podemos considerar la significación como un indicador de la discrepancia entre una hipótesis nula y los datos muestrales. Cuanto más cercano sea a cero, más evidencia tenemos en contra de la hipótesis nula; Hasta el punto que cuando la significación es inferior al nivel de significación fijado de antemano (por ejemplo el 5%), directamente se rechaza la hipótesis nula (se ha superado el umbral de los que se había fijado de antemano como una *“duda razonable a favor de que no hay efecto”*)

3.2.3 Potencia

La probabilidad de que no ocurra el error de tipo II, cuando esta puede calcularse, se la denomina potencia del contraste. Es decir, la potencia del contraste es una medida de la habilidad de un contraste para detectar un efecto que está presente.

En los casos en que es posible fijar la potencia del contraste suele tomarse valores como 85% o 90%. Aparece normalmente en contrastes donde la hipótesis alternativa afirma que no solo un tratamiento X es mejor que un placebo, sino que además es mejor en, al menos, cierta cantidad que se considera clínicamente significativa (no confundir con estadísticamente significativa).

Normalmente los programas estadísticos solamente hacen referencia a la significación de un contraste, y no a la potencia. Ello se debe a que para determinar la potencia es necesario que la hipótesis alternativa no sea simplemente una negación de la hipótesis nula, sino que debe especificar explícitamente una diferencia cuantitativa con ella.

Ejemplo: Se espera que un tratamiento X produzca un efecto beneficioso en la reducción del colesterol de al menos 10 unidades. Para ello se eligen dos grupos de pacientes de características similares y se prueba con un grupo un placebo y con el otro el tratamiento en cuestión. Mientras los datos no indiquen lo contrario pensaremos que el tratamiento no tiene efecto (no consigue reducir de media las 10 unidades). Esta es la hipótesis nula. La hipótesis alternativa es que el tratamiento sí tiene efecto (bajará el colesterol esas 10 unidades, o más). Un nivel de significación del 5% se interpreta como que a menos que los resultados del experimento discrepen suficientemente de que el tratamiento no tiene efecto, consideraremos que el tratamiento no sirve. Si el tratamiento muestra resultados muy buenos con respecto al placebo decidiremos que sí sirve, y rechazaremos la hipótesis nula. Por supuesto, excepcionalmente puede ocurrir que un tratamiento que no tiene ninguna utilidad muestre resultados buenos sobre una muestra, y acabemos concluyendo por error que es bueno. La probabilidad de que esto ocurra evidentemente esperamos que sea baja y es lo que hemos denominado nivel de significación. Supongamos que tras estudiar los resultados del experimento con algún contraste obtenemos que la significación es $p=0.001=0.1\%$. Esto quiere decir que los resultados de la muestra son incoherentes con la hipótesis nula (son más improbables

que el valor límite que fue fijado en 5%). Concluimos que hay evidencia estadísticamente significativa en contra de que el tratamiento no reduce el colesterol en al menos 10 unidades. ¿Cómo se interpreta la potencia en este ejemplo? Supongamos que el tratamiento realmente es efectivo (reduce al menos 10 unidades el colesterol). ¿Cuál sería la probabilidad de que una prueba estadística lo detecte (rechace la hipótesis nula)? Esta probabilidad sería la potencia.