

Hipótesis y decisiones

Preparado por Luis M. Molinero (Alce Ingeniería)

CorreoE: bioestadistica@alceingenieria.net

[Artículo en formato PDF](#)

Abril 2001

Introducción

La difusión de la metodología estadística, así como la amplia disponibilidad de ordenadores y sofisticados programas de análisis de datos, permiten "producir" un número creciente de resultados, y sin embargo conviene que en algún momento nos detengamos a pensar en las características, y por qué no las debilidades, del método que estamos empleando. En opinión de muchos autores la onnipresencia de los niveles de significación en la literatura clínica es totalmente desafortunada, dado que la práctica clínica se basará más en la magnitud de la diferencia observada que en el nivel de probabilidad, por lo que se aconseja el uso de los intervalos de confianza.

En este artículo se revisa el concepto de prueba estadística de contraste de hipótesis, así como los errores asociados, justificando mediante ejemplos la utilidad de los intervalos de confianza. Asimismo se plantea la problemática de los estudios en los que se busca confirmar una hipótesis de igualdad y no su rechazo.

Se completa con un interesante enlace a una calculadora on line para predeterminar tamaños de muestra o calcular la potencia de una prueba.

Pruebas estadísticas de contraste de hipótesis

Es habitual que en la investigación científica se utilicen hipótesis simples para explicar la realidad, debido a que son más fáciles de contrastar de manera empírica y de esa forma determinar su validez o deficiencia. En un ensayo clínico típico el objetivo es estudiar la respuesta al tratamiento y compararla con la respuesta a una terapia alternativa, ya sea ésta el tratamiento habitual o un placebo. La respuesta al tratamiento se determinará, en cada caso, en base a una medida numérica, por ejemplo descenso en la presión arterial, o mediante una clasificación según una variable cualitativa, que en la situación más simple será de tipo dicotómico: el paciente responde o no al tratamiento. La diferencia observada en la respuesta (medias para variables cuantitativas, proporciones cuando se trate de variables cualitativas) entre el grupo de tratamiento y el grupo control constituye una estimación de la efectividad del tratamiento.

Aunque en realidad la efectividad del nuevo tratamiento y del tratamiento control fuera la misma, es previsible que la diferencia observada no sea exactamente cero, aunque sí próxima a cero, siendo más improbables (más raros) los resultados a medida que se alejen de ese valor.

En el procedimiento clásico de contraste de hipótesis se denomina *hipótesis nula* a la que considera que ambos tratamientos son iguales y si, en el supuesto de que sea cierta, la probabilidad de que se observe una diferencia tan grande o mayor que la obtenida en nuestro estudio es muy baja (típicamente 0.05, aunque hay autores más exigentes que sugieren otros valores como 0.01) se rechaza dicha hipótesis y se acepta la contraria, denominada *hipótesis alternativa*, que establece que ambos tratamientos son diferentes.

A pesar de la amplia aceptación en las publicaciones científicas de esta metodología desde sus orígenes, existe un gran debate respecto a su validez metodológica y sobre todo respecto a su aplicación rutinaria, controversia que va calando, aunque de forma muy lenta, en la comunidad investigadora frente a la inercia de su aplicación como meras recetas para la toma de decisiones.

Estudios de equivalencia

En primer lugar hay que destacar que la hipótesis nula nunca puede ser demostrada o establecida, siendo posible sólo, con esta metodología, refutarla mediante un experimento. Nos encontramos entonces en una situación paradójica, ya que en algunos estudios lo que se busca es precisamente demostrar una igualdad. Por ejemplo cuando se pretende demostrar que un tratamiento menos agresivo que el tradicional es igualmente eficaz. O cuando se pretende demostrar que un tratamiento que es más eficaz tiene una tasa similar de efectos adversos.

Probabilidad de error en un contraste de hipótesis

Es evidente que en el contraste estadístico de hipótesis se pueden dar dos posibles errores. El denominado **error tipo I o error alfa**, que es el que se produce cuando se rechaza la hipótesis nula y en realidad es cierta. La probabilidad de cometer este error se fija de antemano por el investigador cuando sitúa el nivel de rechazo, habitualmente 0.05.

Si no se rechaza la hipótesis nula, cuando el valor de probabilidad es inferior al nivel fijado, también se corre el riesgo de cometer un error que se denomina **error tipo II o error beta β** . Ahora las cosas no son tan sencillas, la probabilidad de cometer un error tipo II no es un valor único como la que corresponde al error tipo I. La probabilidad de un error tipo I se calcula suponiendo que la hipótesis nula (no existen diferencias) es correcta, mientras que la probabilidad de un error tipo II se tiene que calcular cuando ésta es falsa, es decir cuando existen diferencias entre los tratamientos. Pero la magnitud **D** de esa diferencia puede tomar en principio cualquier valor y la probabilidad de error dependerá de esa magnitud. Hay que fijar pues la diferencia para la que se desea acotar ese error. Habitualmente se utilizará un valor a partir del cual se puede considerar como diferencia relevante en términos del proceso en estudio.

Inconvenientes de las pruebas de contraste de hipótesis

El principal problema de las pruebas estadísticas de contraste de hipótesis radica en que la decisión de rechazar o no la hipótesis de igualdad descansa en el tamaño de muestra, ya que una diferencia muy pequeña puede ser estadísticamente significativa si la muestra es suficientemente grande y por contra, una diferencia de magnitud relevante puede no llegar a ser estadísticamente significativa si la muestra es pequeña. Aquí es donde entra en juego el concepto de significación física (léase fisiológica, clínica, o el término que corresponda en cada campo de aplicación) frente al de significación estadística, que pone de relieve que lo importante no es demostrar que existen diferencias –algo con lo que ya contábamos pues estamos estudiando dos pautas terapéuticas que sabemos que son diferentes–, sino que éstas son de cierta importancia.

Otro problema a subrayar es la evidente arbitrariedad a la hora de fijar el punto de corte y el planteamiento maximalista del todo o nada.

Veamos un ejemplo. En la siguiente tabla se resumen los datos obtenidos en un ensayo clínico:

	Media	Desv. típica	Tamaño muestra
Tratamiento	128	9	20
Control	123	8	20

Como vemos la diferencia entre las medias en ambos grupos es de 5. Supongamos que una diferencia de 4 o más se considera de importancia clínica. Existe pues una diferencia clínicamente importante entre los dos grupos.

Si se realiza un contraste de hipótesis basado en la *t de Student*, se obtiene un valor de probabilidad para una diferencia igual o mayor que la observada de 0.07, que es inferior al nivel de rechazo universalmente aceptado de 0.05. Por lo que aplicando la "doctrina" estadística no se rechaza la hipótesis de igualdad de los

tratamientos.

Sin embargo se trata de una diferencia de importancia clínica y es cierto que también una probabilidad de 7 entre 100 es bastante baja. Es más que probable que nuestra muestra de 20 pacientes por grupo sea insuficiente, pero en nuestro ejemplo es muy difícil conseguir más pacientes.

Intervalos de confianza

A pesar de la inercia intelectual que nos conduce la aceptación universal del uso de pruebas de hipótesis como un asunto de todo o nada, hay otras alternativas que no se quedan en la mera receta: el uso de los intervalos de confianza. Esta alternativa ha sido defendida vehementemente por diferentes autores y secundada por diferentes editores de revistas médicas. El intervalo de confianza nos da el margen de valores en los que es previsible esperar que se encuentre la verdadera diferencia entre los tratamientos, para una probabilidad dada (habitualmente el 95 %). En realidad se sustenta sobre la misma teoría, pero el enfoque es mucho más expresivo, proporcionando información y no sólo documentando una mera decisión, como es el caso del contraste de hipótesis.

En nuestro ejemplo, el intervalo de confianza va aproximadamente desde -0.5 a 10.5 , lo que revela que es insuficiente para descartar el 0 (igualdad de tratamientos) con toda certidumbre, pero se encuentra muy escorado hacia el lado positivo llegando a valores de gran importancia clínica, por lo que de acuerdo con otros datos (por ejemplo de coste, disponibilidad, posología, efectos secundarios, etc) proporcionan una mayor información a la hora de determinar el interés del nuevo tratamiento.

Ya se ha comentado que en determinados estudios la hipótesis que de hecho se pretende verificar es la de igualdad. A estos estudios se les suele denominar estudios de "**equivalencia**" o también "**negativos**". Puesto que lo que se espera es no rechazar la hipótesis de igualdad, ahora nuestro interés se centra en el error de tipo II o β que podemos cometer. Estamos ante la otra cara de la moneda: si la muestra es suficientemente pequeña, diferencias de importancia clínica pueden no ser estadísticamente significativas. En estos casos es incluso más importante indicar el intervalo de confianza para la diferencia ya que nos permitirá comprobar si incluye valores de trascendencia clínica, lo que justificaría la necesidad de una muestra mayor para poder aceptar que no existen diferencias entre tratamientos. El intervalo de confianza nos proporciona no sólo información en cuanto a las diferencias sino también en cuanto a la sensibilidad de nuestro estudio.

Potencia de un contraste

Es frecuente manejar un concepto relacionado con el error tipo II o β , el de potencia de una prueba estadística. Para una diferencia D dada, la probabilidad de no cometer un error de tipo II, es decir la probabilidad de detectar una diferencia D en el estudio es

$$1 - \beta$$

Como alternativa al intervalo de confianza, en un ensayo de tipo negativo puede ser interesante indicar cuál era la potencia de la prueba para detectar una diferencia clínicamente relevante, siendo importante que ésta no sea pequeña.

Es evidente que el ejemplo anterior no puede ser aceptado como prueba de que no existen diferencias entre los tratamientos, por el mero hecho de que no se encontrase diferencias estadísticamente significativas. Si con un estudio mucho mayor los resultados hubieran sido

	Media	Desv. típica	Tamaño muestra
Tratamiento	128	9	200
Control	127.5	8	200

El intervalo de confianza es ahora $(-1.2, 2.2)$, que engloba el cero y que no incluye valores clínicamente relevantes, lo que nos aporta más evidencia para aceptar una eficacia similar de ambos tratamientos.

Si se calcula la potencia de la prueba en el primer ejemplo con dos muestras de tamaño 20, desviaciones típicas 9 y 8, para detectar una diferencia de 4, veríamos que es 0.297, a todas luces insuficiente.

Por el contrario, la potencia de la prueba para los mismos valores anteriores salvo el tamaño que sería ahora de 200 es de 0.9967.

Y finalmente, para complicar las cosas supongamos que el resultado observado es el de la siguiente tabla

	Media	Desv. típica	Tamaño muestra
Tratamiento	128	9	200
Control	126	8	200

Se puede comprobar que un contraste mediante la t de Student resulta estadísticamente significativo con $p = 0.019$. Sin embargo la magnitud de la diferencia 2 puede no ser de importancia clínica, dado que comentábamos que se consideraban relevantes diferencias iguales o superiores a 4. Una vez más el intervalo de confianza puede ser más ilustrativo que el simple valor de p, en este caso es

0.3 , 3.7

no llegando a incluir el valor considerado relevante, aunque se aproxima.

La importancia, en un sentido o en otro, de estos resultados solo se podrá considerar a la luz del proceso fisiológico concreto que se analiza y no con meros números como aquí se ha expuesto, lo que recalca el absurdo de la utilización cómoda y simplista de las "recetas estadísticas", siendo esta metodología una herramienta más para ayudar en la investigación y no estando exenta ella misma de controversia.

Enlaces de interés

- Como complemento a este artículo se indica una dirección en la que se dispone de calculadoras on line para determinar la potencia de diferentes pruebas o estimar el tamaño de muestra necesario para un contratos de hipótesis
[Power calculator](#)
-

Bibliografía seleccionada

- **Cultura estadística e investigación científica en el campo de la salud: una mirada crítica**
Luis Carlos Silva Ayçaguer
Ed. Díaz de Santos
Madrid 1997
-

 [Índice de artículos](#)

[Principio de la página](#) 