EVIDENCIAS EN PEDIATRÍA
Toma de decisiones clínicas basadas en las mejores pruebas científicas
www.evidenciasenpediatria.es

Fundamentos de medicina basada en la evidencia

Inferencia estadística: contraste de hipótesis

Ochoa Sangrador C[1], Molina Arias M[2], Ortega Páez E[3]
[1]Servicio de Pediatría. Hospital Virgen de la Concha. Zamora. España.
[2]Servicio de Gastroenterología. Hospital Universitario La Paz. Madrid. España.
[3]Unidad de Gestión clínica de Maracena. Distrito Granada-Metropolitano. Granada. España.

Correspondencia: Carlos Ochoa Sangrador, cochoas2@gmail.com

Palabras clave en español: estadística; inferencia estadística; contraste de hipótesis.
Palabras clave en inglés: statistics; statistical inference; contrast hypothesis.

Fecha de recepción: 11 de febrero de 2020 • **Fecha de aceptación:** 19 de febrero de 2020
Fecha de publicación del artículo: 26 de febrero de 2020

Evid Pediatr. 2020;16:11.

CÓMO CITAR ESTE ARTÍCULO

Ochoa Sangrador C, Molina Arias M, Ortega Páez E. Inferencia estadística: contraste de hipótesis. Evid Pediatr. 2020;16:11.

Este artículo está disponible en: http://www.evidenciasenpediatria.es/EnlaceArticulo?ref=2020;16:11.
 • ISSN: 1885-7388

Inferencia estadística: contraste de hipótesis

Ochoa Sangrador C[1], Molina Arias M[2], Ortega Páez E[3]
[1]Servicio de Pediatría. Hospital Virgen de la Concha. Zamora. España.
[2]Servicio de Gastroenterología. Hospital Universitario La Paz. Madrid. España.
[3]Unidad de Gestión clínica de Maracena. Distrito Granada-Metropolitano. Granada. España.

Correspondencia: Carlos Ochoa Sangrador, cochoas2@gmail.com

En artículos previos de esta serie hemos planteado los fundamentos de la inferencia estadística. Diferenciamos en ella dos estrategias: la estimación por intervalos y el contraste de hipótesis. En este artículo expondremos los fundamentos del contraste de hipótesis, tal y como ha sido entendido clásicamente, asumiendo la utilidad de su planteamiento categórico, aunque finalizaremos exponiendo las limitaciones y potenciales errores que entrañan dicho abordaje.

CONTRASTE DE HIPÓTESIS

El contraste de hipótesis nos permite comparar dos o más alternativas, cuantificando la probabilidad de que las diferencias entre ellas sean esperables por azar. Para el cálculo de esta probabilidad nos basaremos en las propiedades de las distribuciones de probabilidad conocidas. Si la probabilidad de encontrar por azar la diferencia observada es muy baja, podemos considerar la opción de que una de las alternativas comparadas sea superior a las demás.

Recordemos un ejemplo presentado anteriormente: en un ensayo clínico se compararon dos tratamientos, A y B, en dos grupos de 100 pacientes, para prevenir recaídas de una enfermedad. En el contraste de hipótesis se plantean dos alternativas:

- Hipótesis nula: "no hay diferencias de eficacia entre A y B", o lo que es lo mismo, la diferencia de proporciones no es distinta de 0.

- Hipótesis alternativa; tenemos dos opciones: "sí hay diferencias entre A y B" (contraste bilateral sin definir la dirección de las diferencias) o "A es más eficaz que B" (contraste unilateral); dicho de otra manera, que la diferencia de proporciones es distinta (bilateral)/mayor (unilateral) que 0, según la opción elegida. La elección de un contraste bilateral o unilateral es elección del investigador y va a depender de nuestro conocimiento previo del problema. Como el contraste bilateral es más conservador, esto es, necesita mayores diferencias para que alcancen el umbral de significación estadística, es la opción más elegida, aunque ambas sean correctas.

Recordemos que en el grupo A recayeron un 20%, mientras que en el grupo B un 40%. En el artículo previo de estimación por intervalos calculamos para el mismo ejemplo que la diferencia de proporciones era del 20%, con un intervalo de confianza del 95% de 7,6 a 32,4. Como ese intervalo no incluye el valor nulo, que para una diferencia es "0", parece que el tratamiento A es más eficaz que el B. Sin embargo, para resolver el contraste de hipótesis debemos cuantificar la probabilidad exacta de que la diferencia encontrada sea mayor que "0", asumiendo la validez de las asunciones requeridas por la prueba de contraste de hipótesis elegida.

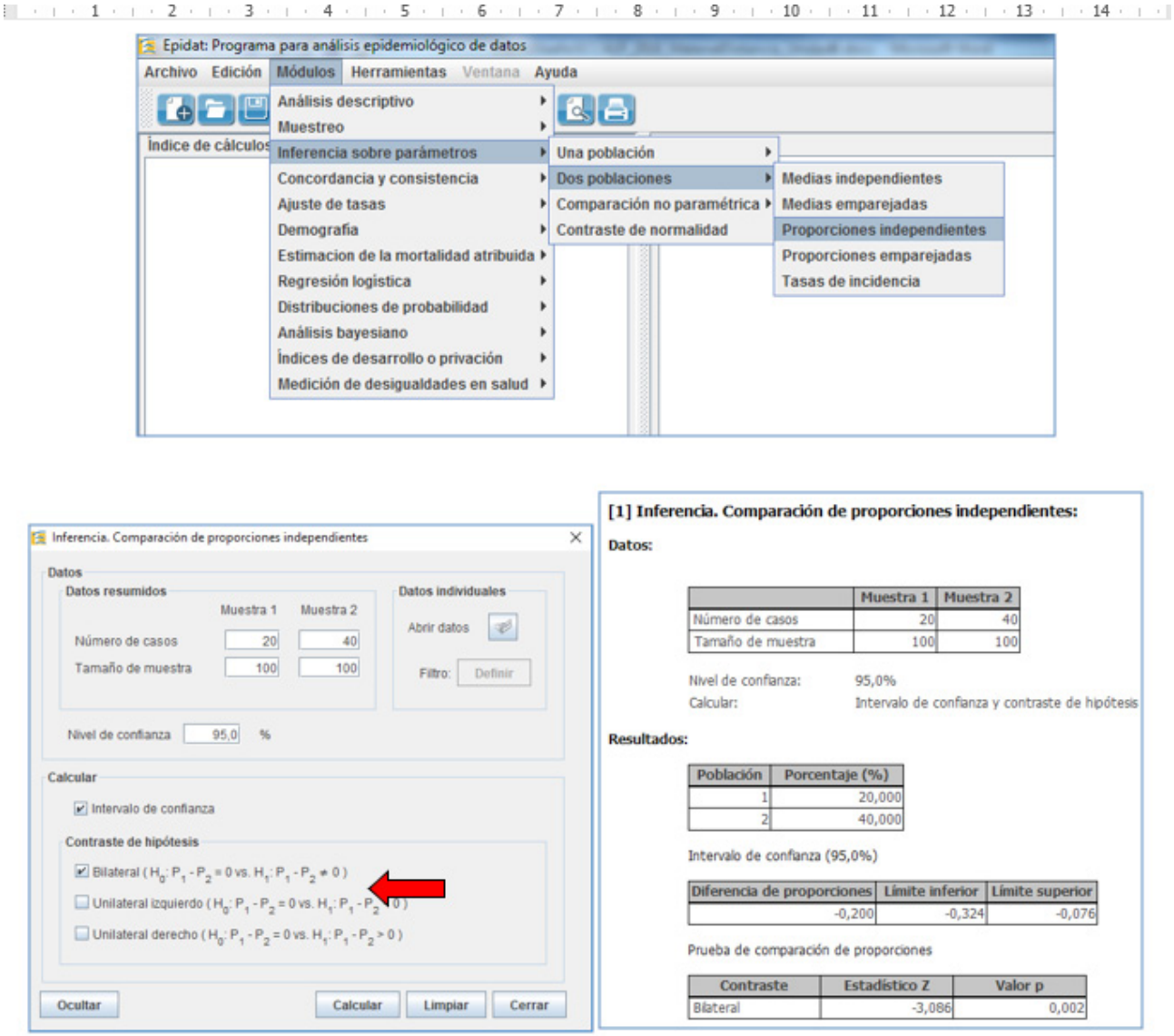
Contamos con varias pruebas con las que calcular esta probabilidad. Una de las pruebas es la aproximación a la distribución normal de la diferencia de proporciones, cuyo error estándar es:

$$EE\ diferencia\ proporciones = \sqrt{\frac{\hat{p}_1\,(1-\hat{p}_1)}{n_1} + \frac{\hat{p}_2\,(1-\hat{p}_2)}{n_2}}$$

Podríamos realizar los cálculos por nosotros mismos, con el error estándar y nuestros conocimientos de la distribución normal, algo que no recomendamos. La mejor alternativa es usar alguna calculadora epidemiológica. En la figura 1 se presenta el cálculo, realizado con el programa gratuito Epidat 4.2 (disponible gratuitamente en https://www.sergas.es/Saude-publica/EPIDAT-4-2?idioma=es) para el contraste bilateral, esto es, aquel en el que la hipótesis alternativa defiende que la eficacia de A y B son distintas, o lo que es lo mismo, la diferencia entre A y B es distinta de 0 (mayor o menor). Los cálculos serían distintos si hubiéramos elegido la hipótesis alternativa unilateral, en la que solo consideramos que A sea más eficaz que B.

La calculadora nos informa del valor Z (distribución normal estandarizada) correspondiente a una diferen-

Figura 1. Contraste de hipótesis para una diferencia de proporciones mediante aproximación a la normal realizado con Epidat 4.2. Se presenta el menú desplegado en el que se accede a la ventana correspondiente (en calcular, debe señalarse el tipo de contraste; en este caso se ha optado por contraste bilateral)

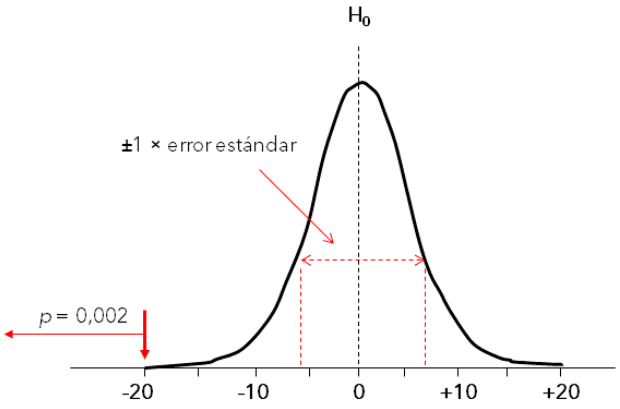


cia de 0,20 (20%) en una distribución normal de media 0 y desviación estándar equivalente a nuestro error estándar (figura 2). El valor Z es 3,086 (valor más alejado que 1,96 de 0; aparece con signo negativo porque el grupo con menor recaídas lo hemos puesto primero, pero para el cálculo es irrelevante; si cambiamos el orden el valor Z sería positivo, pero a ambos les corresponde el mismo valor *p*), al que le corresponde una probabilidad (valor *p*) de 0,002 (0,2%). Como esta probabilidad es menor de 0,05 (5%), consideramos que la hipótesis nula es menos verosímil que la hipótesis alternativa; expresado con la terminología clásica, concluiremos que se rechaza la hipótesis nula (no hay diferencias) y se acepta la alternativa (el tratamiento A es mejor que el B).

Rechazando la hipótesis nula y aceptando la alternativa asumimos un error de 0,002 (0,2%). A este error lo denominamos error tipo I, o error de falso positivo (porque asumimos que hay diferencias en la población, de la que procede nuestra muestra, cuando no las hay en la población), y a su probabilidad la llamamos alfa. Es importante advertir de que, aunque el error sea muy pequeño, siempre existe cierto riesgo de error.

Veamos otro ejemplo. Supongamos que el estudio realizado anteriormente, en vez de contar con 100 sujetos en cada grupo, solo contara con 30 sujetos y que la proporción de recaídas fuera la misma: 20% en el grupo A (6/30) y 40% en el grupo B (12/30). En la figura 3 presentamos el nuevo cálculo.

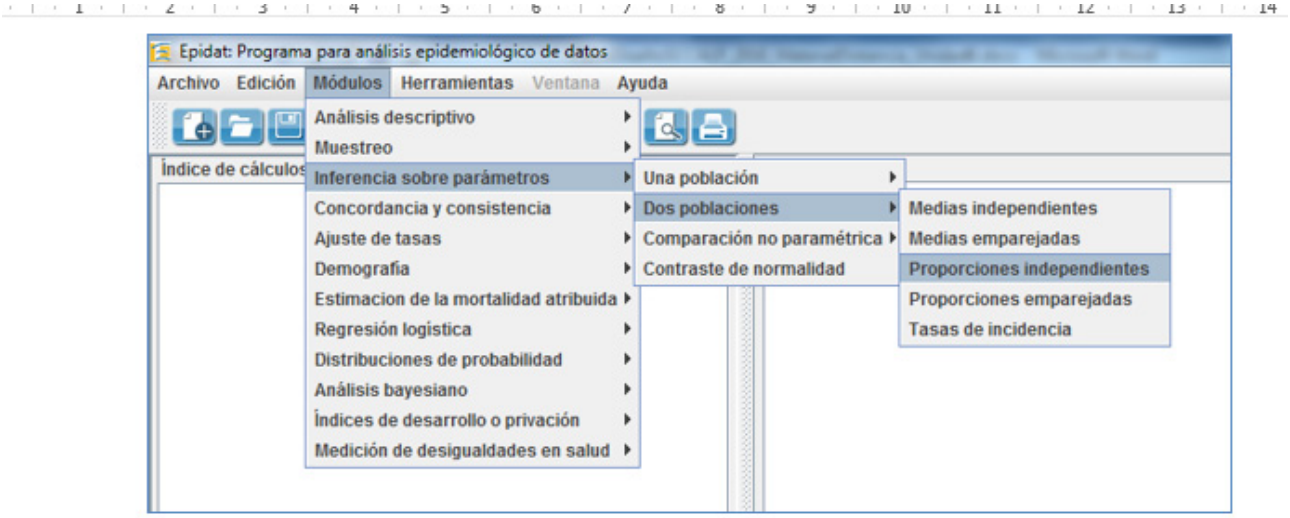
Figura 2. Distribución normal de las diferencias de proporciones de media 0 (hipótesis nula [H_0]) y desviación típica equivalente a su error estándar (para muestras de tamaño 100)

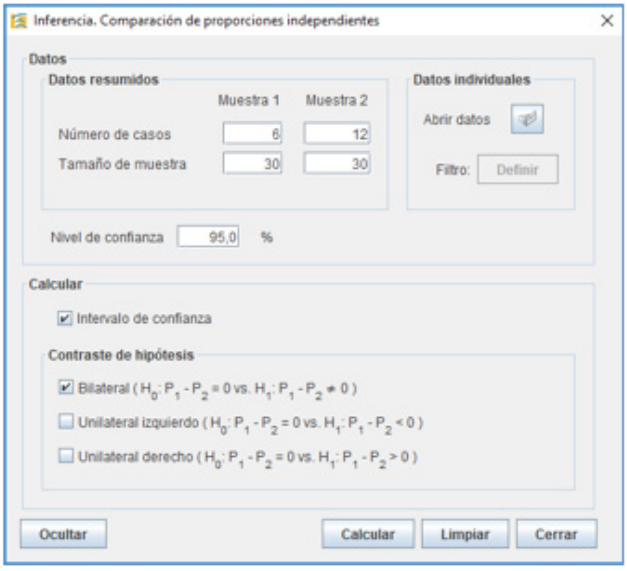


Vemos cómo, aunque la diferencia porcentual entre tratamientos es la misma, al disminuir el tamaño muestral, el error estándar aumenta y la probabilidad asociada a la diferencia encontrada cambia. La calculadora nos da un valor Z de 1,69 (menos alejado que 1,96 de 0), lo que para un contraste bilateral (figura 4A) le corresponde una probabilidad (valor *p*) de 0,091 (9,1%). Con este resultado no podemos considerar la hipótesis nula menos verosímil que la alternativa, porque la probabilidad de encontrar la diferencia observada no es suficientemente baja; expresándolo con la terminología clásica, no podemos rechazar la hipótesis nula ni aceptar la alternativa, ya que el error tipo I (o de falso positivo) en el que incurriríamos sería mayor de 0,05 (5%).

¿Qué ha pasado? Que el nuevo estudio ha perdido potencia, aumentando el riesgo de error tipo II (riesgo beta) o de falso negativo (probabilidad de no encontrar

Figura 3. Contraste de hipótesis para una diferencia de proporciones mediante aproximación a la normal Epidat 4.2





[3] Inferencia. Comparación de proporciones independientes:

Datos:

	Muestra 1	Muestra 2
Número de casos	6	12
Tamaño de muestra	30	30

Nivel de confianza: 95,0%
Calcular: Intervalo de confianza y contraste de hipótesis

Resultados:

Población	Porcentaje (%)
1	20,000
2	40,000

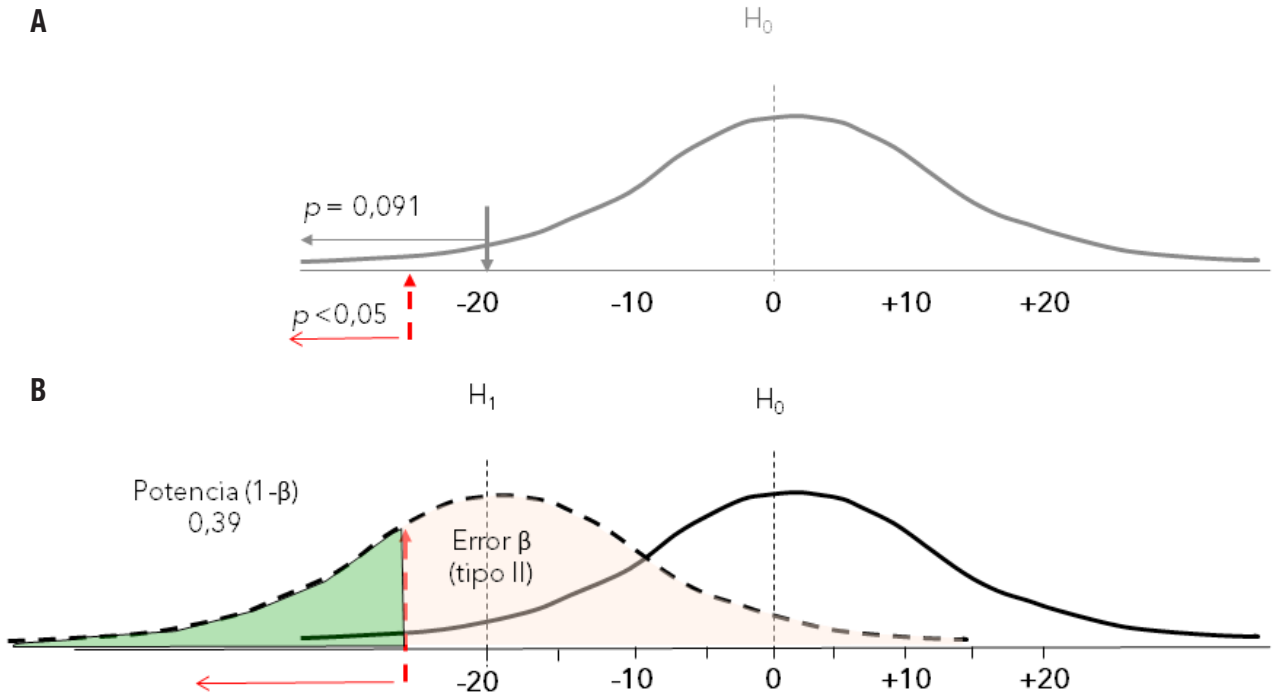
Intervalo de confianza (95,0%)

Diferencia de proporciones	Límite inferior	Límite superior
-0,200	-0,426	0,026

Prueba de comparación de proporciones

Contraste	Estadístico Z	Valor p
Bilateral	-1,690	0,091

Figura 4. A: distribución normal de las diferencias de proporciones de media 0 (hipótesis nula [H_0]) y desviación típica equivalente a su error estándar (para muestras de tamaño 30). B: distribuciones normales de medias 0 (H_0) y -20 (hipótesis alternativa [H_1])



diferencias en la muestra cuando sí las hay en la población). El tratamiento A podría ser más eficaz que el B, pero nosotros no hemos sido capaces de observarlo con suficiente confianza. Al aumentar el error estándar la distribución normal es tan amplia que, aunque la diferencia sea grande el valor nulo "0" es muy probable que quede dentro del intervalo de confianza (figura 4B).

Cuando las diferencias observadas no nos permiten descartar la hipótesis nula (en terminología clásica: no hay diferencias estadísticamente significativas) los resultados solo son valorables si el estudio tiene un error tipo II, cuantificado en el riesgo "beta", menor de 0,20 (20%). Al complementario del riesgo beta lo llamamos Potencia (1-beta). Por ello, estos resultados solo son aceptables si la Potencia es mayor del 80%.

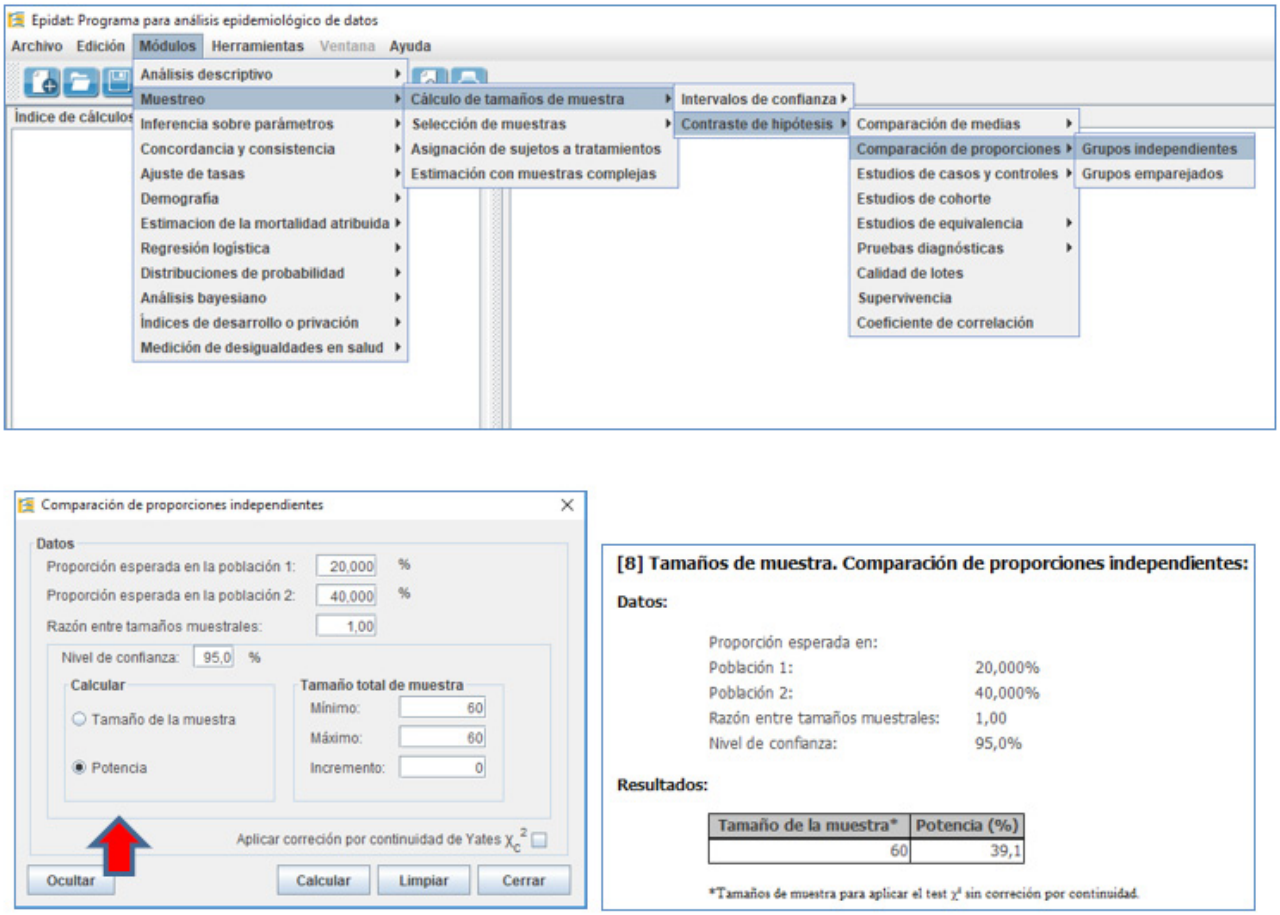
Para calcular el error tipo II (riesgo beta) o la potencia (1-beta), recomendamos usar una calculadora epidemiológica. En la figura 4b se muestra el planteamiento en el que se sustenta el cálculo de la potencia. Si realmente existieran diferencias (H_1 cierta), existiría una distribución de diferencias de proporciones alternativa (H_1) a la hipótesis nula (H_0). En ese caso nuestro estudio podría haber encontrado cualquier valor comprendido en la distribución alternativa, pero solo los que quedan más alejados de la hipótesis nula (H_0), darían una probabilidad menor de 0,05 en ella; podemos ver que ese valor es un valor más extremo que el que nosotros hemos encontrado (-0,20). En la figura 5 se muestra el cálculo para nuestro estudio.

La calculadora ha estimado que, con una muestra de 30 sujetos por grupo (60 en total), la potencia para estimar una diferencia del 20% es 39,1%. Como vemos, no alcanza la potencia mínima requerida del 80%. Por ello nuestro resultado no sería valorable. Si estamos convencidos de que el tratamiento A es mejor que el B (así lo sugiere la diferencia encontrada), lo más razonable es plantear un estudio con mayor tamaño muestral.

Es importante destacar que en el cálculo de la potencia del estudio debemos introducir diferencias (proporciones esperadas en cada población) que consideremos clínicamente importantes, que no tienen por qué coincidir con las observadas en nuestro estudio. En nuestro supuesto hemos usado los datos del estudio, ya que un 20% es aceptable como diferencia clínicamente importante. Si el estudio hubiera encontrado diferencias muy pequeñas (por ejemplo 2%), para el cálculo de la potencia deberíamos haber usado diferencias que consideremos clínicamente importantes, por ejemplo, un 10, un 15 o 20%. La potencia calculada sería interpretada como que, aunque el estudio no ha encontrado suficientes diferencias, tenía potencia suficiente para haber encontrado diferencias mayores de 10, 15 o 20%. La elección de la diferencia requiere conocimientos del problema en estudio y no responde a criterios estadísticos.

Otra cuestión que hay que advertir es que si en los cálculos de la potencia queremos usar riesgos alfa o beta alternativos (por ejemplo, riesgo alfa 0,01 o riesgo

Figura 5. Cálculo de la potencia de un contraste de hipótesis para una diferencia de proporciones con Epidat 4.2.



beta 0,10), los umbrales de cálculo de probabilidad cambiarán. Si disminuimos el riesgo alfa aumentará el beta y viceversa; solo aumentando el tamaño muestral disminuirán los dos.

En la tabla 1 se resumen todas las situaciones posibles del contraste de hipótesis. Debemos tener en cuenta que sea cual sea la decisión de nuestro contraste, siempre existe un cierto riesgo de error, ya que la población es inaccesible. Recordemos que si el riesgo alfa es menor de 0,05 solo tendremos en cuenta la primera fila de la tabla. Cuando el riesgo alfa sea mayor, nos plantearemos el cálculo de la segunda fila, estimando el riesgo beta.

PRUEBAS DE CONTRASTE DE HIPÓTESIS

En el apartado anterior hemos empleado una prueba de contraste de hipótesis (aproximación a la normal de la diferencia de proporciones), pero existen muchas otras pruebas, entre las que tendremos que elegir la más apropiada para cada contraste.

En la elección del test estadístico tendremos que considerar los siguientes factores:

- Cuántas variables están implicadas: 1, 2 o más.
- Cuáles son las variables dependientes e independientes.
- Qué escalas de medida siguen las variables implicadas (nominal, ordinal, continua normal, continua no normal).
- Cuántos grupos de estudio hay: 1, 2 o más.
- Si los grupos de estudio son independientes o están relacionados (o apareados; por ejemplo, mediciones repetidas en los mismos sujetos).
- Si queremos un contraste uni- o bilateral.
- Qué umbrales de errores tipo I y II elegimos (0,05 y 0,20 respectivamente o inferiores).

En la tabla 2 se presenta un esquema simplificado para la elección de la prueba de contraste más apropiada. En próximos artículos iremos desarrollando las principales pruebas de contraste de hipótesis.

LIMITACIONES DEL CONTRASTE DE HIPÓTESIS

En los últimos años va creciendo una opinión crítica con el planteamiento categórico del contraste de hipótesis. Se critica fundamentalmente que la interpretación de los resultados de un estudio y, en consecuencia, la

Tabla 1. Alternativas del contraste de hipótesis.

	Realidad (¡¡desconocida!!)	
Decisión	**H_o Cierta**	**H_o Falsa**
H_0 rechazada H_1 aceptada	Error tipo I (α) Falsos (+)	Decisión correcta
H_0 no rechazada	Decisión correcta	Error tipo II (β) Falsos (-)

Error alfa = probabilidad de equivocarnos si rechazamos la hipótesis nula (H_0) cuando esta es cierta
Error beta = probabilidad de equivocarnos si no rechazamos la hipótesis nula, a pesar de que sea falsa (H_1 cierta)
Potencia del test (1-beta) = probabilidad de rechazar la hipótesis nula cuando es falsa (encontrar diferencias cuando estas existen)

asunción de jerarquías de superioridad en la comparación de alternativas, se sustente exclusivamente en un umbral de significación estadística, establecido arbitrariamente en el nivel de probabilidad 0,05 (5%). Un error muy común, que observamos con frecuencia en textos y exposiciones científicas, es interpretar una *p* no significativa como una prueba de ausencia de efecto o asociación. También es frecuente interpretar una *p* significativa como una prueba de la existencia de un efecto o relación. Ni la ausencia de significación estadística (*p* mayor de 0,05) permite probar la hipótesis nula, ni la presencia de significación (*p* menor de 0,05) permite probar la hipótesis alternativa. Cualquier decisión sobre superioridad o inferioridad está sujeta a incertidumbre, que no se resuelve en función de que la *p* sea superior o inferior a 0,05. La interpretación de los resultados requiere tener en cuenta otros factores, como la magnitud del efecto o asociación, la adecuación de las hipótesis contrastadas, los posibles errores cometidos en el diseño o ejecución del estudio y la validez de las asunciones inherentes a la prueba estadística empleada.

Es preciso recordar que la significación estadística no informa de la dimensión o importancia de los resultados, tan solo de la probabilidad de dichos resultados en el modelo planteado por la hipótesis nula. Si el tamaño del efecto encontrado en un estudio resulta insignificante desde el punto de vista clínico, no importa su nivel de significación, ya que su aplicabilidad será cuestionable. De hecho, cualquier diferencia, por pequeña que sea, puede alcanzar significación estadística, si el tamaño muestral del estudio es suficientemente grande. En este sentido, resulta más informativa la presentación de resultados como intervalos de confianza.

Debemos ser precisos a la hora de presentar los resultados científicos, diferenciando claramente lo que es clínicamente importante de lo que es estadísticamente significativo. Para evitar confusión, parece recomendable limitar el uso del vocablo "significativo" a la indicación del nivel de significación estadístico de un contraste de hipótesis, aportando la significación exacta, sin simplificar la información en una interpretación categó-

Tabla 2. Esquema de elección del test de contraste de hipótesis más apropiado

		Variable dependiente	
Variable independiente	**Nominal**	**Ordinal (continuas no normales)**	**Continua (razón o intervalos)**
Nominal dicotómica	**Muestras independientes:** • Test Z comparación de proporciones • **Test χ^2** • Test exacto de Fisher	Test U de Mann Whitney (Wilcoxon suma rangos)	**Test *t* de Student,** muestras independientes
(2 muestras)	**Muestras relacionadas:** Test de McNemar Test Z y método binomial	Test de Wilcoxon rangos con signo	Test *t* de Student, muestras apareadas
Nominal Politómica (> 2 muestras)	Test χ^2 Método binomial	T. Kruskal-Wallis M. apareadas: P. Friedman	**ANOVA**
Continua	Test *t* de Student	Coeficiente Correlación de Spearman También ordinal\ordinal	**Coeficiente Correlación de Pearson** Regresión Lineal
Análisis de supervivencia (tiempo hasta evento): método de Kaplan Meier y Log-Rank. Técnicas multivariantes: variable dependiente nominal: regresión logística; variable dependiente continua: regresión lineal múltiple; supervivencia: regresión de Cox			

rica de "significativo" (menor de 0,05) o "no significativo" (mayor de 0,05). Recomendamos presentar los resultados con sus intervalos de confianza que nos informan de la magnitud y verosimilitud de los resultados. Asimismo, a la hora de interpretar un intervalo de confianza no debemos limitarnos a comprobar si en su interior se incluye o no el valor nulo (0 para una diferencia; 1 para un riesgo), sino que debemos hacer una interpretación de su valor central y de todos los valores incluidos entre los límites del intervalo.

A pesar de las limitaciones mencionadas, en ausencia de un modelo alternativo ampliamente aceptado que sustituya al contraste de hipótesis clásico, no podemos desterrarlo de nuestros análisis. Dado que estamos obligados a tomar decisiones, eligiendo entre alternativas en presencia de incertidumbre, siempre será mejor cuantificar la incertidumbre que ignorarla. El contraste de hipótesis puede seguir siendo útil si somos capaces de hacer una interpretación válida y prudente de este.

BIBLIOGRAFÍA

- Altman DG. Practical statistics for medical research. Londres: Chapman & Hall; 1991.
- Greenland S, Senn SJ, Rothman KJ, Carlin JB, Poole C, Goodman SN, Altman DG. Statistical tests, P values, confidence intervals, and power: a guide to misinterpretations. Eur J Epidemiol. 2016;31:337-50.
- Milton JS. Estadística para biología y ciencias de la Salud. México: McGraw-Hill; 2001.
- Norman GR, Streiner DL. Bioestadística. México: Mosby/Doyma Libros; 1996.
- Ochoa Sangrador C, Molina Arias M, Ortega Páez E. Inferencia estadística: probabilidad, variables aleatorias y distribuciones de probabilidad. Evid Pediatr. 2019;15:27.
- Ochoa Sangrador C, Molina Arias M. Estadística. Tipos de variables. Escalas de medida. Evid Pediatr. 2018;14:29.
- Ochoa Sangrador C, Ortega Páez E, Molina Arias M. Inferencia estadística: estimación por intervalos. Evid Pediatr. 2019;15:40.
- Ochoa Sangrador C. Evaluación de la importancia de los resultados de estudios clínicos. Importancia clínica frente a significación estadística. Evid Pediatr. 2010;6:40.
- Rosner B. Fundamentals of Biostatistics, 7th Edition. Boston: Brooks/Cole, Cengage Learning 2011.