

Contraste de hipótesis

Dentro de la inferencia estadística, un **contraste de hipótesis** (también denominado **test de hipótesis** o **prueba de significación**) es un procedimiento para juzgar si una propiedad que se supone en una población estadística es compatible con lo observado en una muestra de dicha población. Fue iniciada por Ronald Fisher y fundamentada posteriormente por Jerzy Neyman y Karl Pearson.

Mediante esta teoría, se aborda el problema estadístico considerando una hipótesis determinada H_0 y una hipótesis alternativa H_1 , y se intenta dirimir cuál de las dos es la hipótesis verdadera, tras aplicar el problema estadístico a un cierto número de experimentos.

Está fuertemente asociada a los considerados errores de tipo I y II en estadística, que definen respectivamente, la posibilidad de tomar un suceso falso como verdadero, o uno verdadero como falso.

Existen diversos métodos para desarrollar dicho test, minimizando los errores de tipo I y II, y hallando por tanto con una determinada potencia, la hipótesis con mayor probabilidad de ser correcta. Los tipos más importantes son los test centrados, de hipótesis y alternativa simple, aleatorizados, etc. Dentro de los tests no paramétricos, el más extendido es probablemente el test de la U de Mann-Whitney.

Introducción

Si sospechamos que una moneda ha sido trucada para que se produzcan más caras que cruces al lanzarla al aire, podríamos realizar 30 lanzamientos, tomando nota del número de caras obtenidas. Si obtenemos un valor demasiado alto, por ejemplo 25 o más, consideraríamos que el resultado es poco compatible con la hipótesis de que la moneda no está trucada, y concluiríamos que las observaciones contradicen dicha hipótesis.

La aplicación de cálculos probabilísticos permite determinar a partir de qué valor debemos rechazar la hipótesis garantizando que la probabilidad de cometer un error es un valor conocido *a priori*. Las hipótesis pueden clasificarse en dos grupos, según:

1. Especifiquen un valor concreto o un intervalo para los parámetros del modelo.
2. Determinen el tipo de distribución de probabilidad que ha generado los datos.

Un ejemplo del primer grupo es la hipótesis de que la media de una variable es 10, y del segundo que la distribución de probabilidad es la distribución normal.

Aunque la metodología para realizar el contraste de hipótesis es análoga en ambos casos, distinguir ambos tipos de hipótesis es importante puesto que muchos problemas de contraste de hipótesis respecto a un parámetro son, en realidad, problemas de estimación, que tienen una respuesta complementaria dando un intervalo de confianza (o conjunto de intervalos de confianza) para dicho parámetro. Sin embargo, las hipótesis respecto a la forma de la distribución se suelen utilizar para validar un modelo estadístico para un fenómeno aleatorio que se está estudiando.

Planteamiento clásico del contraste de hipótesis

Se denomina hipótesis nula H_0 a la hipótesis que se desea contrastar. El nombre de "nula" significa "sin valor, efecto o consecuencia", lo cual sugiere que H_0 debe identificarse con la hipótesis de no cambio (a partir de la opinión actual); no diferencia, no mejora, etc. H_0 representa la hipótesis que mantendremos a no ser que los datos indiquen su falsedad, y puede entenderse, por tanto, en el sentido de "neutra". La hipótesis H_0 nunca se considera probada, aunque puede ser rechazada por los datos. Por ejemplo, la hipótesis de que dos poblaciones tienen la misma media puede ser rechazada fácilmente cuando ambas difieren mucho, analizando muestras suficientemente grandes de ambas poblaciones, pero no puede ser "*demostrada*" mediante muestreo, puesto que siempre cabe la posibilidad de que las medias difieran en una cantidad δ lo suficientemente pequeña para que no pueda ser detectada, aunque la muestra sea muy grande.

A partir de una muestra de la población en estudio, se extrae un estadístico (esto es, un valor que es función de la muestra) cuya distribución de probabilidad esté relacionada con la hipótesis en estudio y sea conocida. Se toma entonces como región de rechazo al conjunto de valores que es más improbable bajo la hipótesis, esto es, el conjunto de valores para el que rechazaremos la hipótesis nula si el valor del estadístico observado entra dentro de él.

La probabilidad de que se obtenga un valor del estadístico que entre en la región de rechazo aún siendo cierta la hipótesis puede calcularse. De esta manera, se puede escoger dicha región de tal forma que la probabilidad de cometer este error sea suficientemente pequeña.

Siguiendo con el anterior ejemplo de la moneda trucada, la muestra de la población es el conjunto de los treinta lanzamientos a realizar, el estadístico escogido es el número total de caras obtenidas, y la región de rechazo está constituida por los números totales de caras iguales o superiores a 25. La probabilidad de cometer el error de admitir que la moneda está trucada a pesar de que no lo está es igual a la probabilidad binomial de tener 25 "éxitos" o más en una serie de 30 ensayos de Bernoulli con probabilidad de "éxito" 0,5 en cada uno, entonces: 0,0002, pues existe la posibilidad, aunque poco probable, que la muestra nos dé más de 25 caras sin haber sido la moneda trucada.

Procedimientos de prueba

Un procedimiento de prueba es una regla con base en datos muestrales, para determinar si se rechaza H_0 .

Ejemplo

Una prueba de $H_0: p = .10$ contra $H_1: p < .10$, podría estar basada en el examen de una muestra aleatoria de $n = 200$ objetos. Representamos con X el número de objetos defectuosos de la muestra, una variable aleatoria binomial; x representa el valor observado de X . Si H_0 es verdadera, $E(X) = np = 200(.10) = 20$, mientras, podemos esperar menos de 20 objetos defectuosos si H_1 es verdadera. Un valor de x ligeramente debajo de 20 no contradice de manera contundente a H_0 así que es razonable rechazar H_0 solo si x es considerablemente menor que 20. Un procedimiento de prueba es rechazar H_0 si $x \leq 15$ y no rechazar H_0 de otra forma. En este caso, la región de rechazo está formada por $x = 0, 1, 2, \dots, y 15$. H_0 no será rechazada si $x = 16, 17, \dots, 199$ o 200.

Un procedimiento de prueba se especifica por lo siguiente:

1. Un estadístico de prueba: una función de los datos muestrales en los cuales se basa la decisión de rechazar H_0 o no rechazar H_0 .
2. Una región de rechazo, el conjunto de todos los valores del estadístico de prueba para los cuales H_0 será rechazada.

Entonces, la hipótesis nula será rechazada si y solo si el valor observado o calculado del estadístico de prueba se ubica en la región de rechazo

En el mejor de los casos podrían desarrollarse procedimientos de prueba para los cuales ningún tipo de error es posible. Pero esto puede alcanzarse solo si una decisión se basa en un examen de toda la población, lo que casi nunca es práctico. La dificultad al usar un procedimiento basado en datos muestrales es que debido a la variabilidad en el muestreo puede resultar una muestra no representativa.

Un buen procedimiento es aquel para el cual la probabilidad de cometer cualquier tipo de error es pequeña. La elección de un valor particular de corte de la región de rechazo fija las probabilidades de errores tipo I y II. Estas probabilidades de error son representadas por α y β , respectivamente.

Enfoque actual de los contrastes de hipótesis

El enfoque actual considera siempre una hipótesis alternativa a la hipótesis nula. De manera explícita o implícita, la hipótesis nula, a la que se denota habitualmente por H_0 , se enfrenta a otra hipótesis que denominaremos hipótesis alternativa y que se denota H_1 . En los casos en los que no se especifica H_1 de manera explícita, podemos considerar que ha quedado definida implícitamente como “ H_0 es falsa”.

Si por ejemplo deseamos comprobar la hipótesis de que dos distribuciones tienen la misma media, estamos implícitamente considerando como hipótesis alternativa “ambas poblaciones tienen distinta media”. Podemos, sin embargo considerar casos en los que H_1 no es la simple negación de H_0 . Supongamos por ejemplo que sospechamos que en un juego de azar con un dado, este está trucado para obtener 6. Nuestra hipótesis nula podría ser “el dado no está trucado” que intentaremos contrastar, a partir de una muestra de lanzamientos realizados, contra la hipótesis alternativa “el dado ha sido trucado a favor del 6”. Cabría realizar otras hipótesis, pero, a los efectos del estudio que se pretende realizar, no se consideran relevantes.

Un test de hipótesis se entiende, en el enfoque moderno, como una función de la muestra, corrientemente basada en un estadístico. Supongamos que se tiene una muestra $X = (X_1, X_2, \dots, X_n)^t$ de una población en estudio y que se han formulado hipótesis sobre un parámetro θ relacionado con la distribución estadística de la población. Supongamos que se dispone de un estadístico $T(X)$ cuya distribución con respecto a θ , $F_\theta(t)$ se conoce. Supongamos, también, que las hipótesis nula y alternativa tienen la formulación siguiente:

$$\begin{cases} H_0 : \theta \in \Theta_0 \\ H_1 : \theta \in \Theta_1 \end{cases}$$

Un **contraste, prueba o test** para dichas hipótesis sería una función de la muestra de la siguiente forma:

$$\phi(X) = \begin{cases} 1 & \text{si } T(X) \in \Omega \\ 0 & \text{si } T(X) \notin \Omega \end{cases}$$

Donde $\phi(X) = 1$ significa que debemos rechazar la hipótesis nula, H_0 (aceptar H_1) y $\phi(X) = 0$, que debemos aceptar H_0 (o que no hay evidencia estadística contra H_0). A Ω se la denomina región de rechazo. En esencia, para construir el test deseado, basta con escoger el estadístico del contraste $T(X)$ y la región de rechazo Ω .

Se escoge Ω de tal manera que la probabilidad de que $T(X)$ caiga en su interior sea baja cuando se da H_0 .

Errores en el contraste

Una vez realizado el contraste de hipótesis, se habrá optado por una de las dos hipótesis, H_0 o H_1 , y la decisión escogida coincidirá o no con la que en realidad es cierta. Se pueden dar los cuatro casos que se exponen en el siguiente cuadro:

	H_0 es cierta	H_1 es cierta
Se escogió H_0	No hay error	Error de tipo II
Se escogió H_1	Error de tipo I	No hay error

Si la probabilidad de cometer un error de tipo I está unívocamente determinada, su valor se suele denotar por la letra griega α , y en las mismas condiciones, se denota por β la probabilidad de cometer el error de tipo II, esto es:

$$\begin{aligned} P(\text{escoger } H_1 | H_0 \text{ es cierta}) &= \alpha \\ P(\text{escoger } H_0 | H_1 \text{ es cierta}) &= \beta \end{aligned}$$

En este caso, se denomina **Potencia del contraste** al valor $1 - \beta$, esto es, a la probabilidad de escoger H_1 cuando ésta es cierta

$$P(\text{escoger } H_1 | H_1 \text{ es cierta}) = 1 - \beta.$$

Cuando es necesario diseñar un contraste de hipótesis, sería deseable hacerlo de tal manera que las probabilidades de ambos tipos de error fueran tan pequeñas como fuera posible. Sin embargo, con una muestra de tamaño prefijado, disminuir la probabilidad del error de tipo I, α , conduce a incrementar la probabilidad del error de tipo II, β .

Usualmente, se diseñan los contrastes de tal manera que la probabilidad α sea el 5% (0,05), aunque a veces se usan el 10% (0,1) o 1% (0,01) para adoptar condiciones más relajadas o más estrictas. El recurso para aumentar la potencia del contraste, esto es, disminuir β , probabilidad de error de tipo II, es aumentar el tamaño muestral, lo que en la práctica conlleva un incremento de los costes del estudio que se quiere realizar.

Contraste más potente

El concepto de potencia nos permite valorar cual entre dos contrastes con la misma probabilidad de error de tipo I, α , es preferible. Si se trata de contrastar dos hipótesis sencillas sobre un parámetro desconocido, θ , del tipo:

$$\begin{cases} H_0 : \theta = \theta_0 \\ H_1 : \theta = \theta_1 \end{cases}$$

Se trata de escoger entre todos los contrastes posibles con α prefijado aquel que tiene mayor potencia, esto es, menor probabilidad β de incurrir en el error de tipo II.

En este caso el Lema de Neyman-Pearson garantiza la existencia de un contraste de máxima potencia y determina cómo construirlo.

Contraste uniformemente más potente

En el caso de que las hipótesis sean **compuestas**, esto es, que no se limiten a especificar un único posible valor del parámetro, sino que sean del tipo:

$$\begin{cases} H_0 : \theta \in \Theta_0 \\ H_1 : \theta \in \Theta_1 \end{cases}$$

donde Θ_0 y Θ_1 son conjuntos de varios posibles valores, las probabilidades α y β ya no están unívocamente determinadas, sino que tomarán diferentes valores según los distintos valores posibles de θ . En este caso se dice que un contraste $\phi(X)$ tiene tamaño α si

$$\alpha = \max_{\theta \in \Theta_0} P_{\theta}(\phi(X) = 0)$$

esto es, si la máxima probabilidad de cometer un error de tipo I cuando la hipótesis nula es cierta es α . En estas circunstancias, se puede considerar β como una función de θ , puesto que para cada posible valor de θ en la hipótesis alternativa se tendría una probabilidad distinta de cometer un error de tipo II. Se define entonces

$$\beta(\theta) = P_{\theta}(\phi(X) = 1) \quad \forall \theta \in \Theta_1$$

y, la **función de potencia del contraste** es entonces

$$\text{Pot}(\theta) = 1 - \beta(\theta) \quad \forall \theta \in \Theta_1$$

esto es, la probabilidad de discriminar que la hipótesis alternativa es cierta para cada valor posible de θ dentro de los valores posibles de esta misma hipótesis.

Se dice que un contraste es **uniformemente más potente de tamaño α** cuando, para todo valor $\theta \in \Theta_1$ $\text{Pot}(\theta)$ es mayor o igual que el de cualquier otro contraste del mismo tamaño. En resumen, se trata de un contraste que garantiza la máxima potencia para todos los valores de θ en la hipótesis alternativa.

Es claro que el caso del contraste uniformemente más potente para hipótesis compuestas exige el cumplimiento de condiciones más exigentes que en el caso del contraste más potente para hipótesis simples. Por ello, no existe un equivalente al Lema de Neyman-Pearson para el caso general.

Sin embargo, sí existen muchas condiciones en las que, cumpliéndose determinadas propiedades de las distribuciones de probabilidad implicadas y para ciertos tipos de hipótesis, se puede extender el Lema para obtener el

contraste uniformemente más potente del tamaño que se desee.

Aplicaciones de los contrastes de hipótesis

Los contrastes de hipótesis, como la inferencia estadística en general, son herramientas de amplio uso en la ciencia en general. En particular, la moderna Filosofía de la ciencia desarrolla el concepto de falsabilidad de las teorías científicas basándose en los conceptos de la inferencia estadística en general y de los contrastes de hipótesis. En este contexto, cuando se desea optar entre dos posibles teorías científicas para un mismo fenómeno (dos hipótesis) se debe realizar un contraste estadístico a partir de los datos disponibles sobre el fenómeno que permitan optar por una u otra.

Las técnicas de contraste de hipótesis son también de amplia aplicación en muchos otros casos, como ensayos clínicos de nuevos medicamentos, control de calidad, encuestas, etcétera .

Enlaces externos

- Inferencia estadística, apuntes del Departamento de Matemáticas de la Universidad de La Coruña ^[1]
- HESTADIS - Cálculo del contraste de hipótesis para la media con varianza poblacional conocida (gratuito) ^[2]
- Carlos Reynoso - Atolladeros del pensamiento aleatorio: Batallas en torno de la prueba estadística. ^[3]

Referencias

[1] http://www.udc.es/dep/mate/estadistica2/sec1_3.html

[2] http://www.vaxasoftware.com/soft_edu/hestadis.html

[3] <http://carlosreynoso.com.ar/atolladeros-del-pensamiento-aleatorio-batallas-en-torno-de-la-prueba-estadistica>

Fuentes y contribuyentes del artículo

Contraste de hipótesis *Fuente:* <http://es.wikipedia.org/w/index.php?oldid=61887553> *Contribuyentes:* Acratta, Alakasam, Califasuseso, Cgb, Davius, Elpolaco08, Fenicio, Folkvanger, Hu12, Ialad, Isha, Jagarsoft, Jcaraballo, Jmvkrecords, Jorge c2010, Joseaperez, Juan Mayordomo, LauraFarina, Lloux, LuchoX, Matdroses, Mxcatania, Niqueco, Pabloallo, Pólux, Raimundo Pastor, Sageo, Varyatanil, 62 ediciones anónimas

Licencia

Creative Commons Attribution-Share Alike 3.0 Unported
[//creativecommons.org/licenses/by-sa/3.0/](https://creativecommons.org/licenses/by-sa/3.0/)
