Contrastes de Hipótesis

(Basado parcialmente en *La Estadistica: Una Orquesta Hecha Instrumento* (1996), Jaime Llopis Pérez, Ed. Ariel)

Se considera primero la teoría general de contrastes y luego se particulariza en contrastes concretos. Se usa indistintamente el término *contraste de hipótesis* y el de *test de hipótesis*. La mayor parte de las veces que se utiliza la estadística en el ámbito laboral, se hace a través del contraste de hipótesis.

El contraste de hipótesis es una filosofía considerablemente distinta a lo visto anteriomente. Con la estimación puntual se pretende obtener un valor para un parámetro desconocido y con la estimación por intervalos se pretende dar un intervalo donde esté el valor poblacional buscado con mucha confianza. Sin embargo, en el contraste de hipótesis se trata de hacer alguna afirmación acerca de lo desconocido, y después ver hasta qué punto puede mantenerse dicha afirmación. Es dedir, hasta qué punto la imagen de lo que afirmamos está en sintonía con lo que vemos.

Lo que afirmamos y lo que vemos. Aquí está el punto clave del contraste: poner en relación cómo deberían ser las cosas según una hipótesis con cómo son en realidad. El contraste de hipótesis tiene importantes conexiones con la propia filosofía de la ciencia. En el corazón de su proceder está la pregunta de cómo avanzar, cómo progresar en el conocimiento científico. También gira en tomo a este procedimiento el concepto de verdad científica.

El contraste de hipótesis consiste en la puesta a prueba de una hipótesis, llamada **nula**, frente a otra hipótesis, llamada **alternativa**. Las dos hipótesis **no** tienen la misma importancia. No son simétricas en cuanto a nuestra consideración inicial.

La hipótesis nula la tomaremos como cierta y la mantendremos hasta que no se demuestre lo contrario. La mantendremos hasta que se rompa la sintonía entre el mundo visto bajo la suposición de veracidad de la hipótesis nula y el mundo visto a la luz de nuestras observaciones.

Hay sorprendentes paralelismos entre este planteamiento del contraste de hipótesis en estadística y el proceso del juicio a una persona. Cuando se juzga a una persona es como si al comienzo de la sesión el juez se dirigiera al público afirmando: *La persona que hoy juzgamos es inocente*, refiriéndose al acusado, e inmediatamente después empezara el juicio, empezara a hablar el fiscal, el abogado, los testigos, etc. Así, una persona es inocente hasta que no se demuestre lo contrario.

Si tomamos como cierto lo afirmado al inicio del juicio por el juez, ¿qué esperamos del comportamiento del acusado? Pues el comportamiento que se espera ver en una persona inocente: que no vaya robando bancos, matando a gente por la calle, violando a personas, etc. Si durante el juicio el jurado contempla, por las pruebas aportadas, que no hay coherencia entre lo que suponemos de una persona justa y lo que observamos de la persona que se juzga, entonces retiraremos la hipótesis de inocencia y aceptaremos la de culpabilidad. Pero si no se observa incoherencia entre lo observado en el acusado y lo supuesto en una persona inocente, no se rechazará la hipótesis de inocencia, se mantendrá la hipótesis de inocencia y se le dejará libre.

El planteamiento general de un contraste de hipótesis es, pues, el siguiente. Tenemos dos hipótesis, una a la que llamaremos hipótesis nula y otra, a la que llamaremos hipótesis alternativa que escribiremos,

$$\begin{array}{ll} H_0 \equiv & \text{Afirmación } A \\ H_1 \equiv & \text{Afirmación } B \end{array}$$

Tanto en una como en la otra hacemos una afirmación que puede ser en un sentido muy general. ¿Qué afirmaciones podemos incluir en las hipótesis? Pues cosas como que la media de una población es un determinado valor, que la varianza de la población es un determinado valor, que la media de dos poblaciones son iguales, que la varianza de dos poblaciones son iguales, que dos variables no están relacionadas, que la distribución de una población es una normal, etc.

Es importante que destacar que las afirmaciones implicadas en las hipótesis son **siempre** de toda la **población**. Son afirmaciones hechas acerca de una globalidad que se nos escapa por amplia o por inaccesible.

La muestra son las pruebas del juicio que haremos a la hipótesis nula. Son el proceso propiamente del juicio. Son el relato de los testigos. Las afirmaciones contenidas en las hipótesis nulas son siempre *conservadoras*. Nos muestran siempre un mundo indiferenciado, donde las cosas son iguales, donde las variables no están relacionadas, etc. En definitiva, una hipótesis nula muestra un mundo *primitivo*, donde todo está por demostrar y se parte del supuesto de la veracidad de la misma. Son las diferencias entre las cosas, las relaciones entre las cosas, lo que hay que demostrar, no lo contrario.

De entrada, un fármaco, por ejemplo, antes de comercializarse, hemos de pensar que no produce ningún efecto. En todo caso deberemos demostrar que hay un cambio en las personas que lo toman. Deberemos demostrar que la hipótesis de que no produce ningún efecto es absurdo mantenerla.

Importante: un estadístico de test es un estadístico usado en un contraste de hipótesis, es decir, un estadístico del que podemos saber su distribución bajo la hipótesis nula y bajo la alternativa. O sea, un estadístico del que sabemos cómo se comportaría si fuesen ciertas las hipótesis. Se puede destacar que el *comportamiento* de un estadístico es equivalente a *su distribución*.

Ejemplo: Supongamos que sabemos que la distribución de la población es una normal con varianza conocida y queremos realizar el contraste de hipótesis:

$$\begin{array}{ll} H_0 \equiv & \mu = 5 \\ H_1 \equiv & \mu = 8 \end{array}$$

Si la varianza es conocida lo único desconocido es, entonces, μ.

Suponiendo, por ejemplo, que la varianza es igual a 1, si utilizamos la media muestral podemos saber su distribución bajo la hipótesis nula y bajo la alternativa:

Bajo H_0

$$\overline{X}_n \sim N\left(5, \frac{1}{\sqrt{n}}\right)$$

Cuando se dice *bajo* H_0, se refiere a las condiciones expuestas por H_0.

Del mismo modo,

Bajo H_1

$$\overline{X}_n \sim N\left(8, \frac{1}{\sqrt{n}}\right)$$

Estas distribuciones están perfectamente definidas bajo las dos hipótesis.

Hay dos errores posibles en un contraste de hipótesis:

1. Error de tipo 1: Rechazar H_0 siendo cierta H_0.

2. Error de tipo 2: Aceptar H_0 siendo falsa H_0 (rechazar H_1 siendo cierta H_1).

Evidentemente hay también dos aciertos posibles:

Aceptar H_0 siendo cierta H_0 y aceptar H_1 siendo cierta H_1. El último de estos aciertos tiene un nombre: se le denomina **potencia del test**.

Ante un contraste de hipótesis debe crearse un mecanismo de decisión. Para llegar a este mecanismo es fundamental tomar como cierta H_0 y ver entonces la distribución del estadístico de test bajo esta suposición y bajo la suposición de la H_1.

Al visualizar la distribución del estadístico bajo una hipótesis y bajo la otra es cuando se establecen criterios de coherencia para establecer un mecanismo de decisión. Debemos tomar una decisión a partir de los datos de una muestra y de un planteamiento teórico que hacemos, tras aceptar unas suposiciones acerca del modelo de la población. Es fundamental darse cuenta de que, si no fuera por este planteamiento teórico que construimos paralelamente a los datos, no podríamos elaborar un criterio coherente de elección entre las hipótesis.

Por ejemplo, desde el punto de vista muestral, la media de dos poblaciones casi siempre es distinta. ¿Hasta qué punto esta diferencia muestral se puede considerar una diferencia debida al azar y hasta qué punto puede atribuirse a que estamos ante poblaciones realmente distintas?

Únicamente con los n *escasos* datos de una muestra no podemos establecer un procedimiento de decisión que sea coherente. Es necesario considerar un fundamento teórico.

Nota: el error de tipo I se *fija*. El error de tipo 2 queda determinado por nuestra forma de realizar el contraste, es decir, por la elección que hagamos de lo que se llama **zona crítica**.

Suponiendo que es cierta la H_0 se trata de crear una zona crítica. Una zona crítica es un subconjunto de los números reales que tiene baja probabilidad, bajo H_0, cuando se calcula el estadístico de test.

El criterio de decisión será: aceptaremos H_0, si el estadístico visto como *número* cae fuera de la zona crítica y, por el contrario, la rechazaremos si cae en la zona crítica.

Nota: una zona crítica se construye cogiendo la distribución del estadístico del test bajo H_0 y buscando una zona que tenga, bajo esta distribución, una área pequeña y fijada. Esta área fijada se denomina *nivel de significación*, se simboliza generalmente por α, y acostumbra a cogerse un valor pequeño, un valor de 0,1, 0,05 o 0,01.

A la hora de hacer un contraste de hipótesis son muchas las posibles zonas críticas que podemos elegir, de hecho, cualquier zona que tenga una probabilidad igual al nivel de significación fijado, bajo la H_0, es una zona crítica. Esta probabilidad pequeña es el error fijado por nosotros. El nivel de significación es el error de tipo 1, en realidad es una probabilidad calculada tomando como si fuera cierta H_0.

Por lo tanto, al rechazar H_0, si el valor del estadístico cae en la zona crítica, estamos fijando la probabilidad de aceptar H_1, siendo cierta H_0. Esta probabilidad hay que hacerla lo más pequeña posible. Rechazar una hipótesis nula que, en realidad, es cierta, es muy peligroso. Es como condenar a muerte a un inocente. Cuando se condena en un juicio hay que estar muy seguro de que no se está cometiendo un error. Aquí sucede lo mismo, por eso el nivel de significación se coge muy pequeño.

Dado un estadístico de test, son muchas las posibles zonas críticas. De modo que el problema es encontrar una zona crítica que haga del mecanismo de decisión un procedimiento encaminado a minimizar el error de tipo 2, porque el error de tipo 1 lo habremos fijado ya nosotros al elegir la probabilidad que deseamos que tenga la zona crítica.

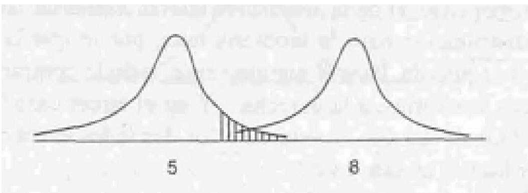
Son infinitas las zonas críticas posibles, pero no todas son igual de interesantes. Cuando tengamos una zona crítica, el valor del estadístico caerá dentro de la zona crítica o fuera. Aceptaremos H_0, si el estadístico visto como número cae fuera de la zona crítica. La rechazaremos, por el contrario, si cae en la zona crítica.

Es muy importante observar lo que representa la distribución de un estadístico bajo una hipótesis. Es el modelo de la variabilidad del valor del estadístico calculado a todas la muestras de tamaño n posibles, si fuera cierta la hipótesis. Tenemos, pues, una imagen teórica de cómo deberían ser las cosas si fuera cierta la hipótesis.

Esto es lo mismo que en el ejemplo del juicio: Cuando a alguien lo declaramos inocente, nos esperamos un tipo de comportamiento adecuado a una persona inocente. Lo que en el juicio son las *pruebas*, aquí es el valor del estadístico calculado a partir de la muestra.

La zona crítica más adecuada será coger la zona donde tenga más importancia la hipótesis alternativa; o sea, una zona donde la función de densidad del estadístico cuando la alternativa sea cierta tenga un área grande. Si el valor del estadístico nos cae en zona crítica rechazaremos la hipótesis nula y aceptaremos la alternativa. Se trata, entonces, de coger una zona crítica con mucha influencia de la alternativa, una zona donde sea muy probable tener aquel valor del estadístico bajo la hipótesis alternativa.

En nuestro ejemplo, como la distribución de la media muestral, bajo la hipótesis alternativa, está a la derecha de la distribución de la media, bajo la hipótesis nula (bajo la alternativa la media poblacional es 8 y bajo la nula es 5), entonces, construiremos la zona crítica hacia la derecha.



Imaginemos que se construye a la izquierda, y se supone, entonces, que tu media muestral, tu número, fuera, por ejemplo 1.5, ¿rechazarías la hipótesis o la aceptarías? Supón que la zona crítica, en este caso, fuera desde el 3 hacia la izquierda, hasta menos infinito. ¿Rechazarías la hipótesis o la aceptarías? Obviamente la rechazarías. Pero si rechazas la nula, en este caso, con más motivo, deberías rechazar la alternativa.

Se trata de crear la zona crítica donde pese más la alternativa, no sólo se está aceptando o rechazando una hipótesis. En la recámara, preparada para entrar en juego, se tiene otra hipótesis, la alternativa, por lo que en el criterio de decisión debe estar contemplada también. Se debe coger siempre una zona crítica que caiga en una zona donde la alternativa tenga mucho peso. Haciéndolo así se minimiza el error de tipo 2, que es mantener la hipótesis nula, siendo cierta la hipótesis alternativa.

Si la hipótesis alternativa, en lugar de ser de un valor concreto de la media poblacional, fuera un poco más compleja, la elección de la zona crítica se complica un poco. Supongamos una hipótesis nula,

$$
\begin{aligned}
H_0 &\equiv \mu = 5 \\
H_1 &\equiv \mu \neq 5
\end{aligned}
$$

contra la alternativa *bilateral*.

Ahora tenemos no una distribución bajo la alternativa, sino infinitas. En este caso los dos lados están ocupados por la alternativa y así construiríamos una zona crítica dividida en dos mitades: una hacia la cola de la izquierda y otra hacia la cola de la derecha. Tanto si obtenemos valores grandes como pequeños del estadístico de test, tendremos sospechas de que la hipótesis nula no funciona y, en cambio, la alternativa puede ser coherente.

Ejemplo

Supongamos que tenemos una población normal, con varianza igual a 1, y que queremos hacer el contraste de hipótesis siguiente:

$$
\begin{aligned}
H_0 &\equiv \mu = 5 \\
H_1 &\equiv \mu = 8
\end{aligned}
$$

Supongamos que el tamaño de la muestra es 16, la media muestral es 5.7, y que el α que cogemos, el nivel de significación, es 0.05. Hay que decidir si aceptmamos la hipótesis nula o si la rechazaríamos.

La distribución bajo la hipótesis nula de la media muestral sería en este caso, $N(5, 1/4)$ por lo tanto, como la zona crítica la hemos de construir hacia la derecha, debemos buscar en esta distribución normal un valor a que cumpla la siguiente condición:

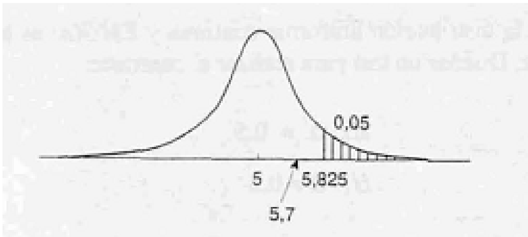
$$
\begin{aligned}
P\left(N\left(5, \frac{1}{4}\right) > a\right) &= 0{,}05 \\
P\left(\frac{N\left(5, \frac{1}{4}\right) - 5}{\sqrt{\frac{1}{4}}} > \frac{a-5}{\sqrt{\frac{1}{4}}}\right) &= 0{,}05 \\
P\left(Z > \frac{a-5}{\sqrt{\frac{1}{4}}}\right) &= 0{,}05
\end{aligned}
$$

Calculándolo en tabla

$$\frac{a-5}{\frac{1}{2}} = 1{,}65$$

por lo que a será igual a 5.825

Visto esto, la hipótesis nula se aceptaría en nuestro caso porque el estadístico (como número) cae en la zona de aceptación, es decir, no cae en la zona crítica. El valor de la media muestral para nuestra muestra hemos dicho que es 5.7 y hemos visto que la zona crítica empezaba en 5.825 y se va hacia la derecha. Haciendo un dibujo:



Es, claramente, zona de aceptación. Si no cae en la zona crítica se decide mantener la hipótesis nula.

Hay una forma más moderna de hacer un contaste de hipótesis. No es que sea contradictoria a ésta. Es, simplemente, una forma de perfilar con más precisión lo que hemos hecho hasta ahora. Fijaos que hasta ahora una hipótesis nula se aceptaba cayera donde cayera el valor del estadístico, siempre que lo hiciera en zona no crítica. Lo mismo ocurre en el caso contrario. Pero aparece una situación muy simplista de que una hipótesis sólo pertenece a una de dos categorías: aceptación o rechazo.

Se define, entonces un concepto más informativo, que se denomina *p-valor*.

Dado un contraste de hipótesis, el **p-valor** es la probabilidad de la mínima zona crítica según la cual rechazaríamos la hipótesis nula. Como es la probabilidad de una zona crítica, es un valor entre 0 y 1.

En esta definición hay que ir con cuidado con una cosa: aquí el concepto de zona crítica se usa en un sentido más general al que antes hemos usado. Aquí zona crítica es cualquier subconjunto de la recta real que nos sirva como criterio de decisión para rechazar la hipótesis nula si el valor del estadístico nos cae ahí. No debe tener necesariamente una probabilidad pequeña. Es una zona crítica con el nivel de significación que queramos. De esta forma, fijaos que la zona crítica más pequeña mediante la cual rechazaríamos la hipótesis nula es siempre una zona crítica construida a partir del punto donde cae el valor del estadístico calculado a la muestra. Una zona crítica más pequeña provocaría el no rechazo de la hipótesis y con una zona mayor rechazaríamos, pero no sería la más pequeña.

La zona crítica que se construye para calcular el p-valores siempre la que dicta la alternativa, de la misma forma que ocurre en el método clásico. Por lo tanto, p-valores grandes mostrarán coherencia de la hipótesis nula y p-valores pequeños incoherencia de la hipótesis nula.

Si el valor de α, según el procedimiento clásico, es de 0.05, aceptaríamos la hipótesis nula si el p-valor fuera superior a 0.05 y rechazaríamos si, por el contrario, el p-valor fuera inferior a 0,05.

No son. por lo tanto. dos métodos distintos, son únicamente informaciones distintas. El método del p-valor proporciona una *nota* a la hipótesis, una nota del 0 al 1. Una nota que aprueba a partir de 0,05 para arriba. El p-valor introduce un nivel de precisión importante. Proporciona una nota a la hipótesis nula.

El contraste de hipótesis es un diálogo entre el número que tengo como consecuencia de operar con un estadístico en una muestra, y todos los números posibles de este estadístico bajo una hipótesis nula, que es la distribución de este mismo estadístico bajo la hipótesis nula. Un contraste de hipótesis es un diálogo entre mi número y todos los números posibles que se podrían obtener bajo una hipótesis. Dependiendo de cómo se desarrolle este diálogo, mantendré o no mantendré la hipótesis.

Fijaos que el p-valor es un índice que mide este diálogo. Si el p-valor es grande, es que es coherente mi número con todos los números posibles. Si el p-valor es bajo, es que hay mucha distancia entre mi número y todos los números posibles bajo la hipótesis. El contraste de hipótesis es un poner en relación lo que veo con lo que debería ver si fuera cierta una hipótesis. El p-valor es un índice que mide la sintonía entre estos dos mundos. Ésta es la esencia del contraste de hipótesis.

Vamos a ver ejemplos concretos de contrastes de hipótesis, son todos ellos contrastes de hipótesis clásicos. Se diseñaron y se utilizan desde hace muchos años. Pueden encontrarse en cualquier software estadístico.

Contraste sobre la media de una población normal

Supongamos que la población (variable aleatoria) de donde tomamos la muestra sigue una distribución normal.

$$\begin{aligned} H_0 &\equiv \mu = \mu_0 \\ H_1 &\equiv \mu \neq \mu_0 \end{aligned}$$

donde μ_0 es un número cualquiera, pero conocido. Un posible estadístico de test es el siguiente:

$$\frac{\sqrt{n}\left(\overline{X}_n - \mu_0\right)}{S}.$$

Este estadístico tiene una distribución t de Student con $(n-1)$ grados de libertad como valor del parámetro, si la hipótesis nula es **cierta**. Aquí está la clave, en saber la distribución del estadístico en las condiciones poblacionales contempladas en la hipótesis nula.

De hecho, es imaginar cómo es el comportamiento del estadístico en el supuesto de la veracidad de la hipótesis para comprobar así si está en sintonía con el número que veremos, con las observaciones de la muestra, con el valor del estadístico.

Este test únicamente lo podemos usar si hemos comprobado que suponer una distribución normal para la población original no es una suposición incoherente.

Se construye una zona crítica o bien se calcula un p-valor, teniendo en cuenta cuál será la zona de influencia de la hipótesis alternativa. Aquí se muestra la hipótesis alternativa bilateral y en este caso la zona de influencia son las dos colas. Si la alternativa fuera lateral, derecha o izquierda, entonces la zona de influencia sería la que correspondiera.