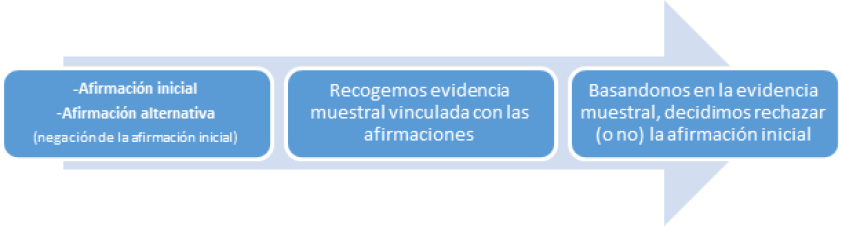
**Introducción**

Supongamos que un amigo nuestro afirma que en cada partido de fútbol que juega, mete tres o cuatro goles. Impresionados con su excelente performance, vamos a verlo jugar cinco partidos seguidos. Pero ocurre que en esos cinco partidos no mete ningún gol. ¿No sospecharíamos que tal vez nos mintió? ¿No es muy incompatible «lo observado» con su afirmación inicial de que mete tres o cuatro goles por partido? Este mini-ejemplo muestra la lógica que hay detrás de una prueba de hipótesis estadística.

Una prueba de hipótesis es un procedimiento, con el que se busca tomar una decisión sobre el valor de verdad de una hipótesis estadística. Al realizar una prueba de hipótesis decidimos si rechazar o no rechazar esa hipótesis estadística. Basamos la decisión en la evidencia muestral.

Un esquema muy simplificado que resume el proceso sería el siguiente

Por eso muchas veces se compara un proceso de prueba de hipótesis con un juicio: hay que recoger evidencias para analizar si la hipótesis de base (la inocencia del acusado en el caso del juicio) se sostiene o se rechaza.

**Hipótesis estadísticas**

Un principio general de la investigación científica es escoger siempre la hipótesis más simple capaz de explicar la realidad observada. La razón es que una hipótesis simple es más fácil de contrastar empíricamente y descubrir sus deficiencias, lo que permite aprender de los datos con mayor rapidez y seguridad.

Este principio justifica que muchas investigaciones estadísticas tengan por objeto contrastar una hipótesis simplificadora del tipo: una población es idéntica a otra de referencia (el efecto de un medicamento nuevo es igual a otro existente…); dos o más poblaciones son iguales entre sí…

Una hipótesis se contrasta comparando sus predicciones con la realidad: si coinciden, dentro del margen de error admisible, mantendremos la hipótesis; en caso contrario, la rechazaremos.

Una **hipótesis estadística** es una afirmación que se hace sobre una o más características de una población (decir que la vida media de una batería son x horas, que un tratamiento reduce el dolor, que un determinado producto hace adelgazar, que a la gente le gusta mayoritariamente el morbo…). Un **test** o **contraste de hipótesis** es algún procedimiento para aceptar o rebatir dicha hipótesis o afirmación.

La teoría de los test de hipótesis surge a partir de 1925, cuando Ronald Fisher publicó el libro *Métodos estadísticos para investigadores*. En él, Fisher definía los **contrastes de significación**, cuyo funcionamiento básico sigue el esquema del falsacionismo, como veremos.

**Test de significación (NHST)**

Una prueba de significación consta de una hipótesis  (hipótesis nula) que consiste en establecer que el valor de un parámetro (llamémosle θ) es un número concreto (llamémosle ):

La hipótesis de partida del investigador fue bautizada con este nombre por Fisher porque significaba que no había cambio alguno con el uso de un nuevo fertilizante, que su efecto era nulo. Es decir, para poder demostrar que un nuevo fertilizante es efectivo, supondremos que no lo es, que tiene efecto nulo, y trataremos de **falsar** esta afirmación. El acrónimo NHST viene de “Null hypothesis signification testing”.

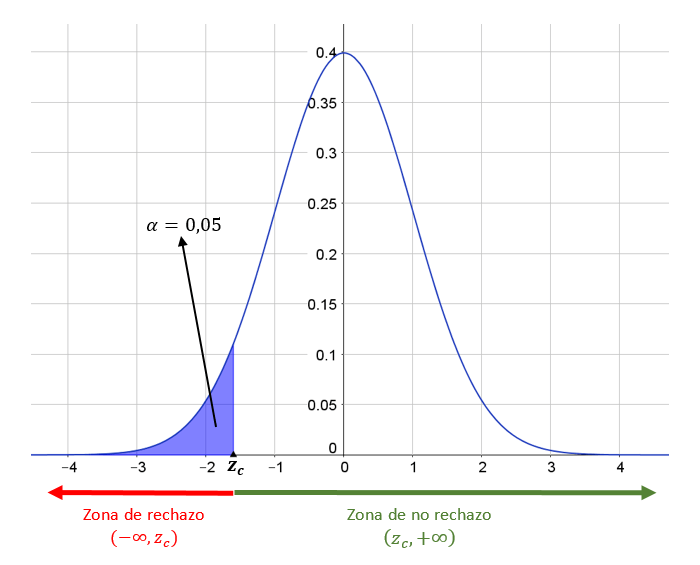
Entonces, una vez que se delimita la hipótesis nula que se desea poner a prueba, la forma general de proceder es elegir una muestra de esa población, y ver si los resultados de esa muestra son coherentes con la afirmación que se está realizando. Básicamente, que la evidencia que nos proporcione la muestra sea lo suficientemente fuerte para poder tomar una decisión.

Para comprobar la coherencia de los resultados, o la fuerza de la evidencia, se tratará de ver la diferencia entre lo observado en la muestra y lo que dice la hipótesis nula. Para ello se elige el llamado **estadístico**  (estadístico pivote) del test y se calcula su valor sobre los datos de una muestra observada , lo que se denota como . Dado que la distribución en el muestreo del estadístico  ha de ser conocida, se determina la probabilidad de que el estadístico tome un valor igual o más extremo que el valor observado  bajo la suposición de que la hipótesis nula es cierta.

Se escribe simbólicamente

, y este número se denomina  valor.

Entonces, si el   *-valor* es excesivamente pequeño (en general, por debajo de 0.05), el resultado del test se dice que es significativo, porque permite rechazar la hipótesis nula . En otro caso, el resultado del test no es significativo y no puede rechazarse esta hipótesis.



De acuerdo con esta lógica se entiende que, con un nivel de significación del 5%, en promedio cinco de cada 100 veces que la hipótesis nula sea cierta se rechazará por azar. Es decir, 5 de cada 100 veces que rechazamos la hipótesis nos estaremos equivocando, pues se asume que es cierta y la prueba no demuestra ni su falsedad ni su certeza.

La hipótesis nula sólo se rechaza si la probabilidad de observar una muestra como la dada es demasiado baja. Esto es, la hipótesis nula se rechaza si la muestra es demasiado rara cuando la hipótesis nula es cierta. El razonamiento estadístico se basa en la siguiente disyunción lógica: “o bien ha ocurrido un suceso excepcional (muy improbable) o bien la hipótesis nula no es correcta.”

El *P***-valor** o probabilidad de significación funciona como una suerte de la evidencia en contra de la hipótesis nula: cuanto menor es, más evidencia en contra. Un valor demasiado pequeño indica que la muestra observada se separa de lo esperado mucho más de lo que sería achacable al azar, a las circunstancias del muestreo aleatorio. Por lo tanto, el investigador se encuentra ante una hipótesis nula inverosímil, descartable.

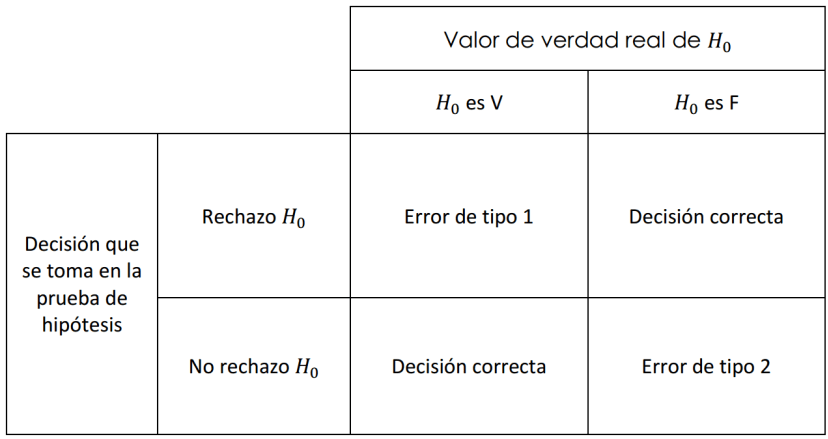
**La teoría de Neyman-Pearson**

Buscando fortalecer las bases lógicas de los test de significación de Fisher, Egon Pearson (1895-1980) (hijo de Karl Pearson) y Jerzy Neyman (1894-1981) idearon varias mejoras. El eje principal de su investigación era el siguiente interrogante: ¿qué hacer si se obtiene un resultado significativo en un test estadístico? Se rechaza la hipótesis nula, pero los test de significación no arrojaban ninguna pista sobre qué hipótesis elegir a cambio.

La teoría de Neyman-Pearson utilizó el NHST de Fisher y el -valor como parte de un **proceso formal de decisión**. Así, plantearon una elección real entre dos hipótesis rivales. El contraste de hipótesis quedó convertido en un método para discernir entre dos hipótesis: la hipótesis nula y la hipótesis **alternativa**

Todo contraste de hipótesis conduce pues, a aceptar o rechazar la hipótesis nula planteada (aceptando, en este último caso, la hipótesis alternativa). Ahora bien, pueden darse las siguientes situaciones.

* Se acepta la hipótesis nula siendo verdadera. Esta es una decisión correcta.
* Se rechaza la hipótesis nula siendo falsa. Esta es otra situación correcta.
* Se rechaza la hipótesis nula siendo verdadera. Estamos cometiendo un error, que se llama *error de tipo uno*. La probabilidad de cometer este error viene dada por el nivel de significación , fijado de antemano.
* Se acepta la hipótesis nula siendo falsa. También cometemos un error, que se llama *error de tipo II*. La probabilidad de cometer este error se representa por β, y la probabilidad  se llama *potencia del contraste*, que cuantifica la probabilidad de rechazar la hipótesis nula cuando es falsa.



Posibles opciones en un test de hipótesis.

Si tenemos una hipótesis nula: : un tratamiento nuevo no es efectivo, frente a : el tratamiento sí es efectivo siempre es posible construir más de un test de hipótesis para contrastar la hipótesis nula frente a la alternativa.

Por otro lado, si tiramos una moneda al aire. Si sale cara, aceptamos . Si sale cruz, rechazamos .

La probabilidad de cometer un error de tipo I es 0.5, igual que la probabilidad de cometer un error de tipo II.

Si en vez de tirar una moneda tiramos un dado y decidimos mediante la regla: “aceptamos la hipótesis nula si sale un 1, la rechazamos si sale cualquier otro número”, la probabilidad de error de tipo I es 5/6 y la de error de tipo II es 1/6.

Obviamente ambos test son bastante absurdos, pero nos sirve para ver que siempre existen test con sus correspondientes errores.

**Contrastes paramétricos y no paramétricos**

Los contrastes pueden ser de tipo **paramétrico** o **no paramétrico**, según se refieran o no a parámetros de una población (a la media, a la varianza, a una proporción…).

Una **hipótesis paramétrica** es una afirmación sobre una o más características (parámetros) de una población. Si dicha hipótesis especifica un único valor para el parámetro la llamaremos **hipótesis simple**.

**Contrastes de hipótesis paramétricas**

* **Tipos de contrastes: bilaterales y unilaterales**

Un contraste es **bilateral** cuando tiene la forma

: (Por  nos referimos a un parámetro teórico y por  a un valor constante, un número) frente a

Un contraste **unilateral** es de la forma: : frente a

Contraste unilaterales

**Pasos a seguir al realizar un contraste de hipótesis**

1.- Especificar las hipótesis nula y alternativa.

2.- Elegir un **estadístico**  para el contraste (para medir la discrepancia entre lo observado y lo teórico). Este estadístico tendrá una función de densidad determinada que nos servirá para calcular el - valor.

3.- Tomar una muestra  y evaluar el estadístico del contraste .

4.- Calcular el - valor , que viene a ser la probabilidad de obtener esos datos, si  es cierta. Si ese valor es muy pequeño, significa que esos datos son muy improbables bajo la hipótesis nula, con lo cual tenderemos a pensar que no es cierta.

Una especie de nivel crítico es el valor 0.1, de forma que

Si el - valor es más pequeño que 0.1, tenderemos a no creer en, y, en cambio, si es más grande, tenderemos a creer en .

En muchas ocasiones (muchos libros, artículos de investigación), se trabaja fijando un **nivel de significación**  (error de tipo I) y realizando la siguiente comparación:

* Si  se rechaza
* Si  se acepta

Los valores con los que se suele trabajar son  o . El más habitual es

Para rechazar la hipótesis nula se requiere que la probabilidad del fenómeno que se produce por casualidad sea inferior al cinco por ciento (este es el origen de la anécdota del estadístico que presenció la decapitación de 25 vacas, advirtió que una sobrevivía y desechó el fenómeno por no significativo).

No rechazar una hipótesis no prueba que sea totalmente cierta. Podemos cometer un error de tipo II.

A continuación, indicamos los estadísticos que se utilizan para los principales contrastes de tipo paramétrico, y la distribución que siguen cuando la hipótesis nula es cierta.

**Para la media de una variable normal**

Supongamos que estamos trabajando con datos de una variable aleatoria XX **que sigue una distribución normal**. Supongamos que establecemos la hipótesis de que la media (teórica, que es desconocida) es igual a un cierto valor numérico:

.

 es un número concreto. Por ejemplo, se ha realizado un tratamiento a un conjunto de pacientes con dolor lumbar. A cada uno de ellos se le ha pedido que diga, en una escala de 0 a 10, cuánto dolor le ha reducido el tratamiento (0 si nada, 10 si se ha quedado sin dolor alguno).

Suponemos que la variable *X* =*reducción del dolor con el tratamiento* sigue una distribución normal. Para saber si el tratamiento es efectivo, la hipótesis nula que plantearíamos es:

De lo que se trata es de *falsar* esta hipótesis, es decir, encontrar una diferencia estadísticamente significativa entre lo que se obtenga en la muestra y lo que dice la hipótesis nula. Esta última, al considerar que la media es cero, implica que el tratamiento no tiene efectividad, puesto que, en media, no reduce el dolor.

Para realizar el contraste de hipótesis, inicialmente consideramos dos posibilidades: que conozcamos la desviación típica de la variable, o que no la conozcamos. Esta última opción es la más habitual, y la más lógica, puesto que, si no tenemos información sobre la media, es raro tenerla de la desviación típica. En el ejemplo que hemos puesto, si no se conoce la reducción media de dolor con un tratamiento (hablamos habitualmente de un tratamiento en fase experimental) pues será raro conocer su variabilidad.

En todo caso, en la mayoría de los textos realizan la distinción entre los casos de desviación típica conocida o no. Consideraremos aquí también los dos casos, igual que se hizo en el capítulo anterior de intervalos de confianza.

**Si se conoce la desviación típica**

Si tenemos una muestra  de *n* datos de una variable aleatoria normal, de parámetros  y  la media muestral verifica

Por lo tanto, si tipificamos la variable (restamos la media y dividimos por la desviación típica), obtenemos la variable , lo que quiere decir que esta variable sigue una distribución normal estándar

Este estadistico o variable aleatoria  se llama estadístico “**pivote**” o estadístico del contraste. Es con el que se realiza el contraste y se toma una decisión, según el valor que tome con los datos de la muestra.

El valor se calcula en función de la distribución que sigue el estadístico “**pivote**” del contraste, y de que el contraste sea bilateral o unilateral.

**La prueba**

En la práctica, es bastante extraño conocer la desviación típica (puesto que precisamente estamos en una situación de incertidumbre de la variable. Lo habitual es no conocer ni la media ni la desviación típica). En este caso, lo que se hace es estimar la desviación típica a partir de la muestra, utilizando para ello la cuasi-desviación típica muestral . En este caso, el estadístico que se utiliza cambia, y también la distribución del mismo. Tenemos que usar el estimador

Este estimador sigue una distribución *t* de Student, con *n−1* grados de libertad. A este contraste de hipótesis se le llama **prueba t**.

HACER EJEMPLOS EN JUPYTER

CAMPAÑA DE MARKETING ES EFECTIVA O NO

PRESENTACIÓN EN JUPYTER

