

UNIVERSIDAD TECNOLÓGICA NACIONAL
Facultad Regional San Nicolás

PROBABILIDAD
y
ESTADÍSTICA II

UNIDAD N°4

Licenciatura en Enseñanza de la Matemática
Año 2011
Mg. Lucía C. Sacco

UNIDAD N°5 Pruebas de Hipótesis

Test de hipótesis. Test para una muestra de observaciones. Test e intervalo de confianza para la media de una población normal con varianza conocida y desconocida. Test e intervalo de confianza para dos muestras normales independientes. Región crítica. P – valor. Nivel potencia de test.

Propósitos:

Brindar oportunidades para la construcción de herramientas que permitan:

- ✓ Comprender que significa realizar un test de hipótesis.
- ✓ Conocer los diferentes tipos de pruebas de hipótesis.
- ✓ Reconocer la importancia de realizar una buena prueba de hipótesis.
- ✓ Comprender en qué consiste y para qué sirve realizar el análisis de la varianza de una distribución de datos.

Bibliografía sugerida:

- Alfonso Lopes, Paulo (2000). *Probabilidad & Estadística. Conceptos, modelos y aplicaciones en Excel*. Prentice Hall.
- Mendanhall, Beaver, Beaver (2010). *Introducción a la probabilidad y estadística*. Décima tercera edición. CENGAGE Learning.
- Meyer, P. (1973). *Probabilidad y aplicaciones estadísticas*. Versión en español. Universidad Católica de Chile. Fondo Educativo Interamericano.
- Zylberberg, A. (2004). *Probabilidad y Estadística P(X)*. Nueva Librería.

1. Introducción

Como dijimos anteriormente, los dos principales procedimientos de la estadística inferencial son la **estimación de parámetros** (puntual o por intervalos) y las **pruebas de hipótesis**.

En esta unidad analizaremos, como es posible la toma de decisiones en el campo de la estadística. Ello lo haremos, estudiando las diferentes maneras de realizar "pruebas de hipótesis".

2. El concepto de prueba de hipótesis

A través de un ejemplo veremos ciertas consideraciones que nos aproximarán al concepto de prueba de hipótesis.

Supongamos que se establece la siguiente hipótesis:

"La droga A mata a las bacterias x"

A partir de esta hipótesis, el investigador formula la siguiente deducción:

Si se exponen colonias de bacterias x a la droga A, las bacterias morirán.

Si la hipótesis es verdadera, se sabrá que la deducción también lo es, pues si la droga realmente mata, las bacterias morirán.

Ahora bien, ¿si al exponer las bacterias x a la droga A, las bacterias no mueren, se puede afirmar que la hipótesis es falsa?

Si se denomina A a la hipótesis y B a la deducción se puede resumir que:

- Si la deducción no se cumple, la hipótesis es falsa: $\bar{B} \rightarrow \bar{A}$
- Si la deducción se cumple, decir que la hipótesis es verdadera: $B \rightarrow A$ no es totalmente cierto

(Las bacterias pueden haber muerto por muchas otras causas: temperatura moderada, medio de cultivo mal preparado, pH incorrecto, etc. y no sólo por la droga).

En resumen, cuando se analiza sólo una parte de la población (una muestra), siempre se puede rechazar una hipótesis pero no podemos aceptarla. (La evidencia únicamente nos llevará a no rechazarla).

Otro ejemplo muy simple puede ayudarnos a clarificar estas ideas.

Supongamos que establecemos la siguiente hipótesis y su correspondiente deducción con respecto a un aula de la Facultad que está con las puertas cerradas:

A: la habitación está vacía (hipótesis)

B: no se ve a nadie (deducción).



Si miramos por el ojo de la cerradura y vemos un alumno sentado en un banco, podemos sentirnos seguros al decir que la hipótesis A es falsa.

Deducción no se cumple \longrightarrow Hipótesis falsa

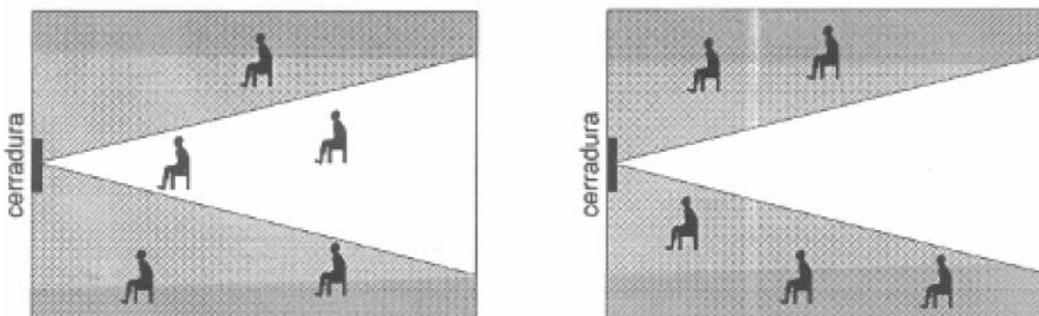
$$\bar{B} \longrightarrow \bar{A}$$



Si miramos por el ojo de la cerradura y no vemos a nadie en el aula, decir que la hipótesis es verdadera es incorrecto.

Deducción se cumple \longrightarrow Aceptar la hipótesis es una afirmación incierta

B \longrightarrow A es una afirmación incierta.



El hecho de mirar por el ojo de la cerradura equivale a observar sólo una parte del aula. Podemos asimilar esta situación con la inferencia estadística por cuanto debemos tomar una decisión en base a una información limitada (muestra) respecto a la totalidad del aula (población).

Generalmente, en una investigación científica se trabaja con muestras obtenidas a partir de una cierta población. Por ello se debe tener mucho cuidado al establecer conclusiones a partir de ellas. Si la evidencia de la muestra lleva a la decisión de rechazar la hipótesis planteada estaremos mucho más seguros de nuestra determinación que cuando no existan evidencias para su rechazo.

3. Hipótesis nula y alternativa

Una hipótesis estadística es un supuesto que se establece sobre las características de una distribución poblacional.

Pasaremos ahora a comentar el razonamiento que se sigue al efectuar una prueba de hipótesis.

Este razonamiento es muy similar al que se utiliza en una corte de justicia cuando se debe tomar la decisión de declarar o no culpable a una persona acusada de haber cometido un delito.

Los actores de un juicio pueden resumirse en un acusado, un fiscal y un juez, que es quien debe tomar la decisión de declarar o no culpable al acusado.

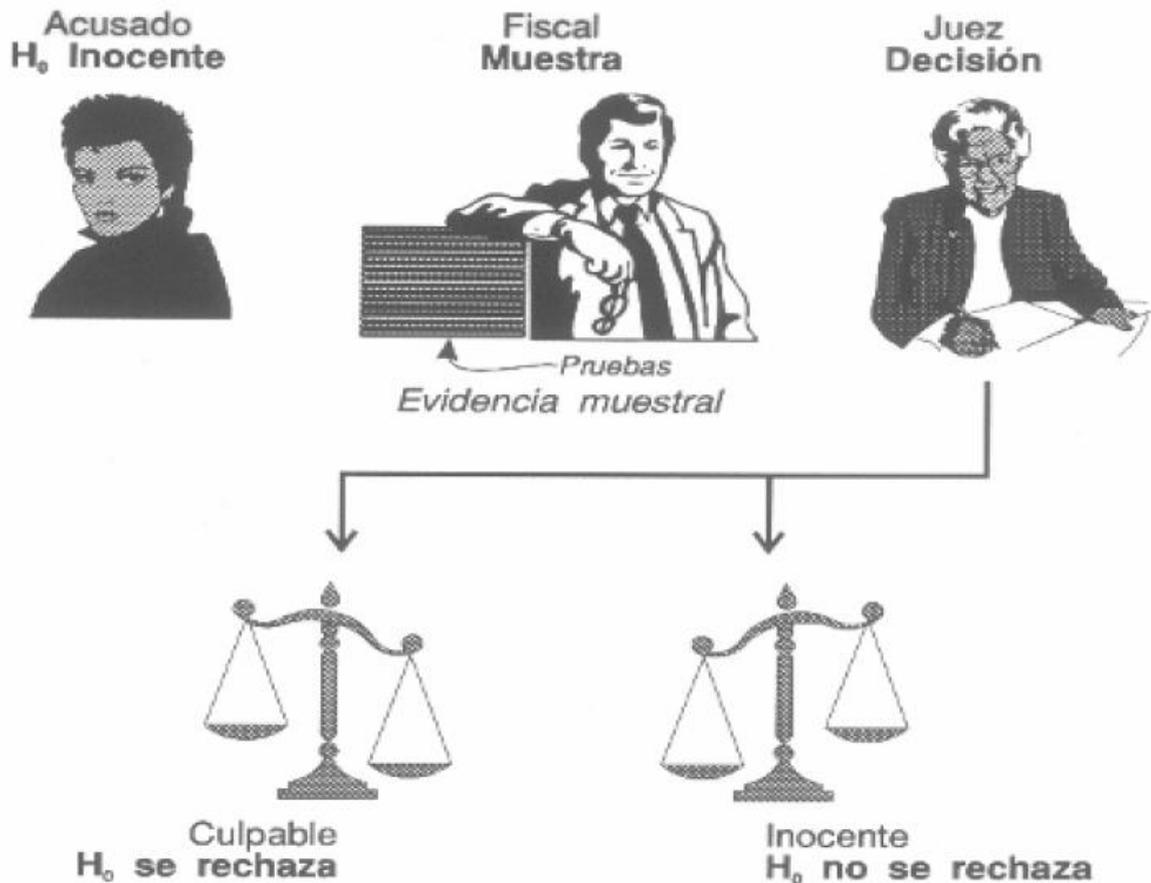
El acusado será considerado inocente hasta tanto las pruebas presentadas por el fiscal demuestren lo contrario.

Haciendo una analogía con la prueba de hipótesis, la hipótesis nula asume el papel del acusado en el juicio y el investigador el papel de fiscal.

La hipótesis nula es un supuesto acerca de uno o más parámetros de la población que debe ser rechazado o no en base a la evidencia muestral.

Simbolizaremos la hipótesis nula como H_0 .

Se denomina nula en el sentido que supone que no existe diferencia real entre el verdadero valor del parámetro de la población de la que hemos obtenido la muestra y el valor hipotetizado (supuesto de inocencia).



Ahora bien, si la evidencia presentada al juez no es contundente, éste decidirá por la inocencia del acusado (no rechazo de la hipótesis nula). En caso contrario, cuando la evidencia condene al acusado, el juez tiene a mano una alternativa, la acusación de culpabilidad (lo que equivale al rechazo de la hipótesis nula planteada).

Si la hipótesis nula es falsa, deberá existir otra hipótesis que sea verdadera. Esta hipótesis recibe el nombre de **hipótesis alternativa**.

La hipótesis alternativa será simbolizada con H_1 .

Aplicaremos los conceptos estudiados a un ejemplo real.

El Instituto Pasteur de París y el Laboratorio Merieux bajo la rectoría de la Agencia Nacional de Investigación contra el SIDA (ANRS) han elaborado una vacuna contra tal enfermedad. En estos días comenzó a probarse en seres humanos, luego de que los ensayos con chimpancés resultaran positivos. Cincuenta voluntarios, hombres y mujeres de 18 a 55 años seleccionados con minuciosidad entre más de 600 candidatos, serán los cobayos en esta experiencia histórica que acerca un paso más a la prevención de la enfermedad. Los candidatos debían ser todos cero-negativos, y con un alto riesgo de contraer la enfermedad (prácticamente se conoce que todos van a enfermarse a causa de su situación de alto riesgo).

Previamente a esta experiencia, el grupo francés ensayó la inmunidad lograda en chimpancés inoculados con la vacuna, obteniendo una respuesta favorable en un 75% de los casos. Por este motivo se decidió que la vacuna podrá ser considerada efectiva en seres humanos cuando ella prevenga de la enfermedad a más del 80% de los pacientes tratados.

Como dijimos que la hipótesis nula asume el rol del acusado y que éste es inocente (no efectiva) hasta que se demuestre su culpabilidad (efectiva), la hipótesis nula deberá expresar:

$$H_0) P \leq 0,80$$

Donde P es la proporción poblacional, es decir, el parámetro del cual hacemos un juicio provisorio.

Los 50 voluntarios de la muestra serán inoculados con la vacuna y se les efectuará un seguimiento de un año luego de lo cual se obtendrían las primeras conclusiones.

Podemos comparar al equipo de investigación con el fiscal, pues cuando realizan su experimento están buscando pruebas respecto a la falsedad de la hipótesis nula especificada. En otras palabras, están tratando de demostrar que el acusado es realmente culpable. La muestra de personas a las que se les dio la vacuna será la evidencia que el equipo de investigación llevará ante el tribunal encargado de dictar sentencia (ANSR).

También dijimos que en el procedimiento de prueba de una hipótesis nula debe existir lo que se denomina una hipótesis alternativa.

Si en nuestro ejemplo la hipótesis nula especificaba que $P \leq 0,80$, evidentemente, H_1 deberá especificar que $P > 0,80$.

$$H_0) P \leq 0,80 \text{ y } H_1) P > 0,80$$

Hasta aquí hemos resuelto el primer paso a seguir en el planteamiento de hipótesis, o sea, el establecimiento de la hipótesis nula y de la hipótesis alternativa.

4. Reglas de decisión

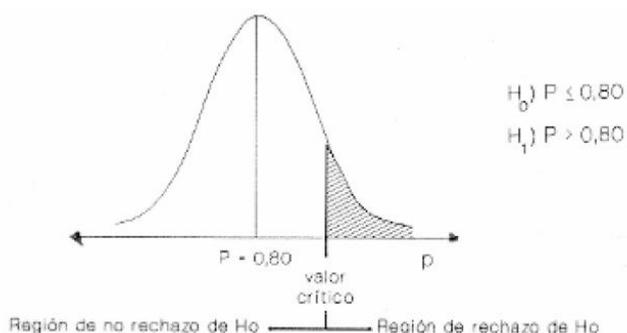
Ahora bien, si en la muestra considerada, el 15% de las personas no contrae la enfermedad, es muy probable que la vacuna sea declarada como no efectiva. Por el contrario, si el 95% de las personas no contraen la enfermedad, con seguridad ella será considerada efectiva. Pero, ¿qué decisión se tomará si, por ejemplo, el porcentaje de personas que no se enferman oscila entre el 70% y el 85%?

Evidentemente, debería establecerse un criterio que especificara, de manera objetiva, a partir de qué valor de p (proporción muestral), se tomará la decisión de rechazar o no la hipótesis nula. Este criterio, conocido como **criterio de test**, se calcula en base a la información aportada por la muestra.

Como consecuencia del criterio de test adoptado se podrá establecer un valor crítico que determinará una **región de rechazo** y una **región de no rechazo** de la hipótesis nula.

En el ejemplo de la vacuna, el estadístico para efectuar el test será la proporción muestral p de personas que quedaron inmunizadas por su acción y las reglas de decisión determinarán en cuál de las dos regiones se posiciona el valor muestral.

Al llegar a este punto debemos recordar que se está decidiendo algo con respecto al parámetro poblacional P a partir de la información proporcionada por una muestra. En realidad, estamos mirando por el ojo de la cerradura y, en consecuencia, el investigador podrá cometer dos tipos de errores.



Por una parte, podría rechazar la hipótesis nula diciendo que la vacuna es efectiva cuando en realidad no lo es. Es fácil imaginarse la peligrosidad de esta aseveración en cuanto a la salud de los pacientes que creen estar inmunizados y no toman las precauciones necesarias para no contraer la enfermedad.

Otro error que puede cometer consiste en no rechazar la hipótesis cuando en realidad es falsa. Estaría diciendo que la vacuna no es efectiva cuando en realidad lo es. La ciencia médica perdería la posibilidad de erradicar el SIDA cuando tenía todo a su alcance para hacerlo.

Pasaremos a definir estos dos tipos de errores que son: error de Tipo I y error de Tipo II.

El error de Tipo I es aquel que se comete al rechazar una hipótesis nula cuando ella es verdadera.

La probabilidad de cometer un error de Tipo I se simboliza con α , conocida como p o nivel de significación de un test.

El error de Tipo II es aquel que se comete al no rechazar una hipótesis nula cuando en realidad es falsa.

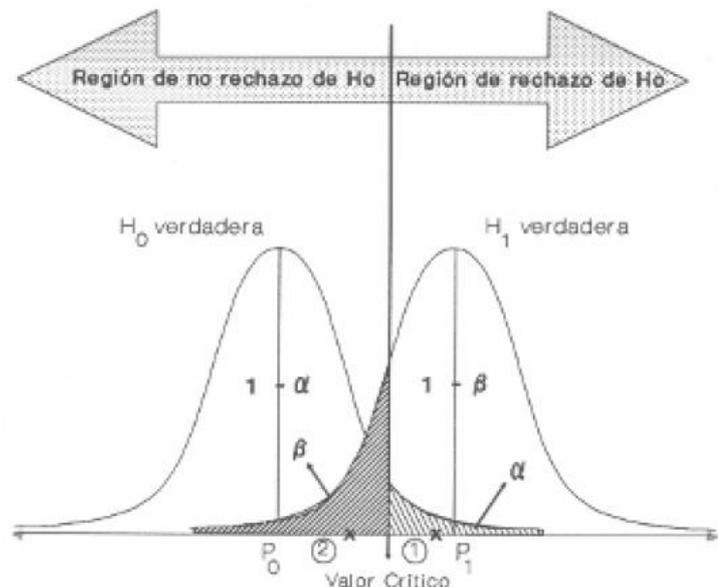
La probabilidad de cometer un error de Tipo II se simboliza con β .

El siguiente cuadro resume todas las acciones que se pueden llevar a cabo en un procedimiento de prueba de hipótesis.

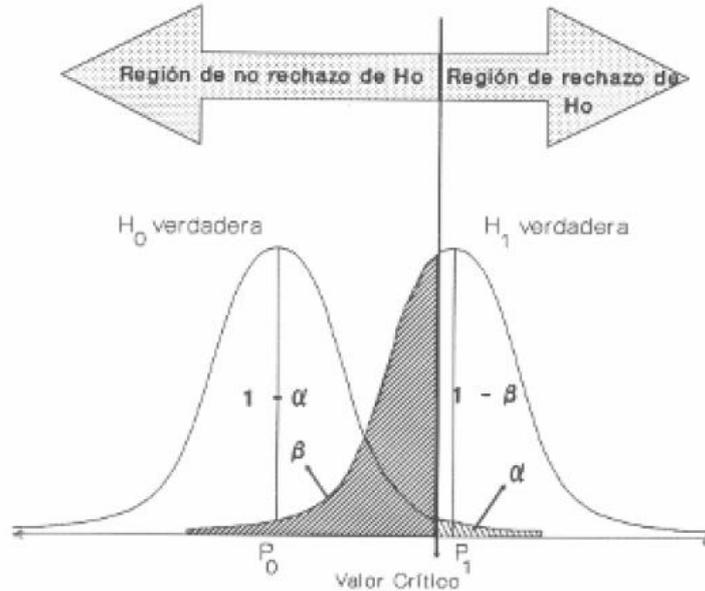
Decisión	Hipótesis nula	
	Verdadera	Falsa
Rechazar	Error Tipo I $P(E_I) = \alpha$	Decisión correcta $(1 - \beta)$
No rechazar	Decisión correcta $(1 - \alpha)$	Error Tipo II $P(E_{II}) = \beta$

También podemos observarlo gráficamente:

a)



b)



En los gráficos a) y b), la curva de la izquierda representa la distribución de p cuando H_0 es verdadera, es decir, la muestra proviene de una población con media P_0 . La curva de la derecha representa la distribución del estadístico cuando se cumple una hipótesis alternativa H_1 (una alternativa podría ser $P = P_1$ con $P_1 > P_0$). Las regiones de rechazo y no rechazo de H_0 siempre se definen con referencia a la distribución que supone que H_0 es verdadera.

Las reglas de decisión de un test especifican que H_0 será rechazada cuando el estadístico muestral utilizado como criterio para efectuar el correspondiente test (p), cae en la región de rechazo de la misma (cae en el punto (1) del gráfico). La probabilidad de que un estadístico "caiga" en esta región es igual a α cuando H_0 es verdadera.

La probabilidad del error de Tipo II, es decir la probabilidad de no rechazar H_0 cuando es falsa, está localizada bajo la curva donde se cumple H_1 pero a la izquierda del valor crítico, es decir, en la frontera de la zona de aceptación de H_0 .

Comparando ambos gráficos, vemos que en el b), la probabilidad del error de Tipo I es menor que en la de a) pero, el área correspondiente al error de Tipo II, se ha incrementado.

Entonces, cuando tratamos de reducir la probabilidad de cometer el error Tipo I (α), aumenta la probabilidad de cometer el error de Tipo II (β).

La única manera que tiene el investigador de disminuir ambos tipos de errores a la vez consiste en aumentar el tamaño de la muestra (n).

5. Potencia de un test

La elección de un criterio de test que defina reglas de decisión adecuadas a la hipótesis que queremos probar, depende de la **potencia** que tenga el test para detectar cuando rechazar o no una hipótesis nula.

En otras palabras, se debe buscar un test que sepa discriminar correctamente cuándo las diferencias entre los valores muestrales realmente observados por el investigador y el valor verdadero que toma el parámetro a ser probado se deban solamente al azar

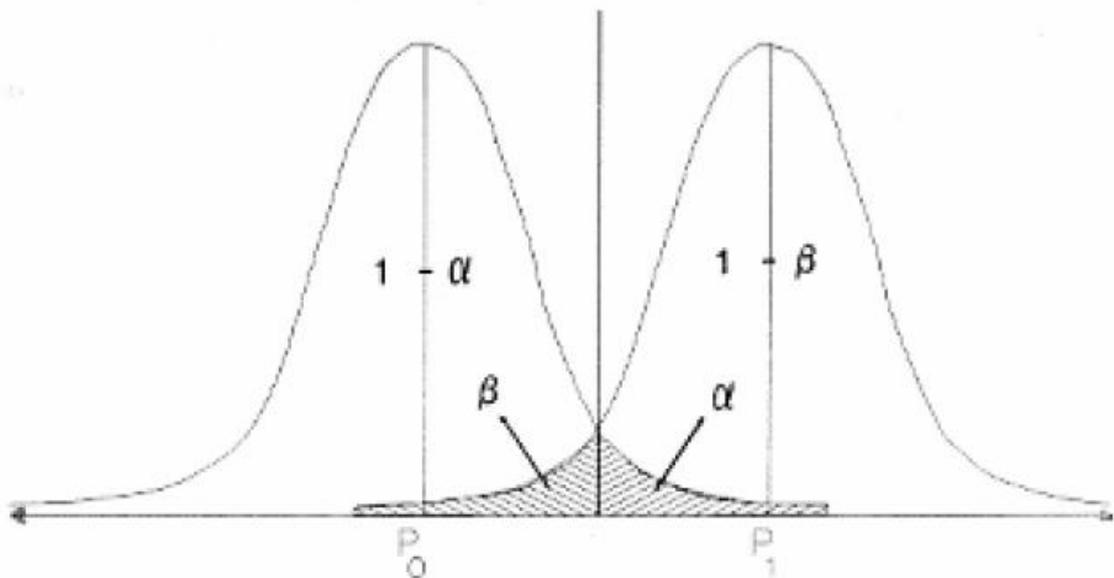
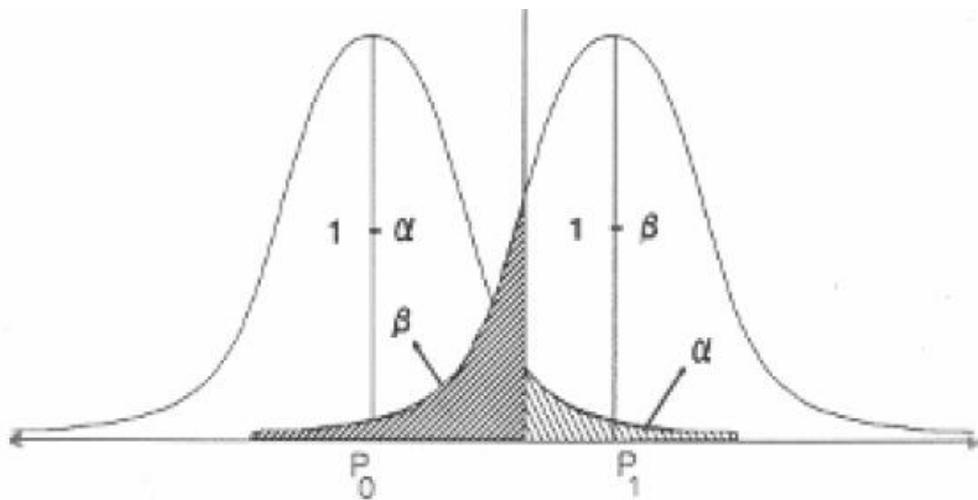
(no se rechaza H_0) y cuándo la divergencia es tan grande que la evidencia muestral no pueda sostener al valor que toma el parámetro en la hipótesis nula (rechazo de H_0).

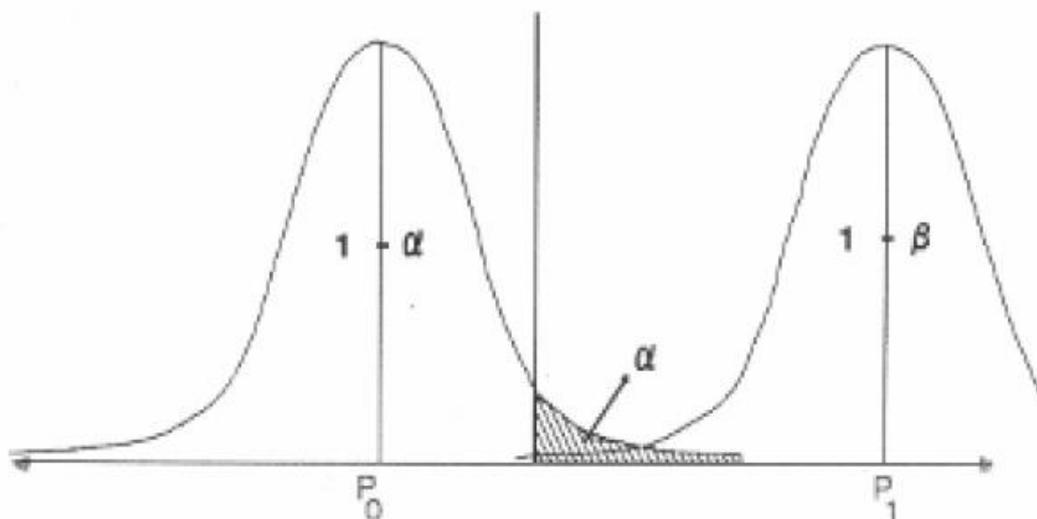
Esto se resume diciendo que se debe buscar el test más potente. Por ello, pasamos a definir, a continuación, lo que se entiende por potencia de un test estadístico.

La potencia de un test se define como la función que establece la probabilidad de rechazar la hipótesis nula cuando es falsa.

De acuerdo a la definición anterior, la potencia de un test se calcula tomando $1 - \beta$, o sea, uno menos la probabilidad de cometer un error de Tipo II.

Cada uno de los posibles valores que puede tomar el parámetro en la hipótesis alternativa tiene su propia potencia. Así, cuanto más próximo se encuentre un valor del parámetro definido por la hipótesis alternativa al valor que toma en la hipótesis nula, mayor será la superposición de las correspondientes distribuciones y menor será la potencia del test con respecto a esta alternativa.





En general, se ha dado mucha más importancia al nivel de significación de un test (α o el famoso p de las publicaciones científicas) y poco énfasis a la potencia de tales test.

Por ejemplo, en una investigación médica se desea estudiar el efecto de una determinada droga como terapia para una enfermedad. La nueva droga tiene la particularidad de producir efectos colaterales que provocan severas molestias al paciente cuando ella es administrada.

Supongamos que el investigador se plantea las siguientes hipótesis:

H_0) la sobrevivida promedio del paciente es superior a 2 años

H_1) la sobrevivida promedio del paciente es inferior a 2 años.

Debido a la agresividad de la terapia en cuestión, al investigador le va a interesar más disminuir el error de Tipo II que el de Tipo I. Es preferible para el paciente que no se utilice la droga a que se la utilice, cuando en realidad no produce una sobrevivida aceptable y disminuye su calidad de vida.

Lamentablemente, no existe una pauta estándar para determinar el nivel de significación y la potencia adecuada que un test debe tener. La única acción que protege al investigador para asegurarse que su decisión es lo más correcta posible está relacionada con el **buen diseño de su experimento**.

Supongamos, por ejemplo, que el investigador pudiera seleccionar una muestra de 30 pacientes con todas las características genéticas iguales (gemelos entre ellos), todos exactamente en el mismo estadio de la enfermedad y que pudiera aislarlos en una burbuja mientras administra la nueva droga en cuestión. En este caso, el investigador estaría seguro de que la sobrevivida del paciente está ligada únicamente al efecto de la droga y su inferencia tendría un mínimo error.

Todos estaremos de acuerdo que no es posible efectuar este tipo de experimentos en la vida cotidiana de investigación. Siempre habrá pacientes en mejores condiciones físicas que otros, será imposible determinar exactamente el estadio de la enfermedad y algunos vivirán en mejores condiciones ambientales que otros, mientras dura el tratamiento.

Con este análisis queremos alertar al investigador que, cuando se pone a prueba una hipótesis, no todo pasa por la consideración del famoso " p " del tanto por ciento sino más bien por la forma en que se ha diseñado el experimento para llegar a este valor.

Además, el uso frecuente de los niveles de significación del 0,05 y 0,01 es un establecimiento convencional con muy poca base lógica y científica.

Cuando se lee una publicación científica que aporta resultados de una investigación, es importante interiorizarse sobre las características de inclusión y exclusión de pacientes que se siguieron para definir la población objetivo, de la manera en que fueron seleccionados los pacientes, del cuidado que se tuvo durante el tiempo que duró el trabajo de experimentación en sí, de la precisión en las mediciones, etc., más que preocuparnos por el valor de p .

Cuando el experimento fue realizado de la mejor manera posible y con la mayor responsabilidad de parte del equipo de trabajo, podemos estar seguros de que, para el nivel de significación alcanzado existe una buena potencia del test, entendiéndose por potencia su habilidad para determinar el verdadero estado de la naturaleza.

Resumiendo:

Como síntesis hemos visto que para llevar a cabo cualquier prueba de hipótesis se necesita especificar 3 elementos:

- La hipótesis nula. (Dado que la hipótesis alternativa queda automáticamente determinada como su complemento).
- Un estadístico muestral que actúe como evidencia para la toma de decisión.
- La determinación de la región de rechazo de la hipótesis nula.

Las dos hipótesis en competencia son la **hipótesis alternativa (H_1)**, generalmente la hipótesis que el investigador desea apoyar, y la **hipótesis nula (H_0)**, una contradicción de la hipótesis alternativa.

Es más fácil presentar apoyo para la hipótesis alternativa al demostrar que la hipótesis nula es falsa. En consecuencia, el investigador estadístico siempre empieza por suponer que la hipótesis nula es verdadera. El investigador utiliza entonces los datos muestrales para decidir si la evidencia está a favor de H_1 más que de H_0 y saca una de dos **conclusiones:**

- **Rechaza H_0 y concluye que H_1 es verdadera**
- **Acepta H_0 como verdadera**

Observación:

Se adjunta a este material, ejemplos de investigaciones en el que se lleva a cabo esta metodología.