

# Prueba de hipótesis estadísticas

## 1.1 Introducción

La prueba de una hipótesis estadística tiene una fuerte relación con el concepto de estimación.

Una hipótesis estadística es una afirmación con respecto a alguna característica desconocida de una población de interés. La esencia de probar una hipótesis estadística es el de decidir si la afirmación se encuentra apoyada por la evidencia experimental que se obtiene a través de una muestra aleatoria.

## 1.2 Conceptos básicos para la prueba de hipótesis estadísticas

La idea principal es determinar si el valor de un parámetro es igual a un valor específico o si es diferente de ese valor. A la afirmación igual a un valor específico se le llama *hipótesis nula* y se escribe como:

$$H_0 : \mu = 10.$$

Debemos tomar en cuenta las consecuencias que puede originarse como resultado del verdadero estado de la naturaleza al tomar alguna de las posibles decisiones. Es decir, existe dos posibles decisiones con respecto a  $H_0$ : Rechazar o equivocarse al rechazar  $H_0$ . Sin embargo, cada una de estas decisiones tiene las siguientes dos consecuencias con respecto al estado de la naturaleza:

$$\begin{array}{cc} \text{Rechazar } H_0 \left\{ \begin{array}{l} \text{cuando } H_0 \text{ es cierta} \\ \text{cuando } H_0 \text{ es falsa} \end{array} \right. & \text{Equivocarse al rechazar } H_0 \left\{ \begin{array}{l} \text{cuando } H_0 \text{ es cierta} \\ \text{cuando } H_0 \text{ es falsa} \end{array} \right. \end{array}$$

Cuando se toma una decisión con respecto a una hipótesis nula, dos de las posibles consecuencias relativas al verdadero estado de la naturaleza conducen a errores inferenciales. El rechazo de la hipótesis  $H_0$  cuando en realidad es cierta, constituye lo que se denomina *error de tipo I*. Equivocarse al rechazar  $H_0$  cuando en realidad  $H_0$  es falsa, constituye lo que se denomina *error de tipo II*. Debemos tomar en cuenta que sólo es posible el error de tipo I cuando la decisión es el rechazar la hipótesis nula, mientras que el error de tipo II sólo es posible cuando la decisión es el no rechazar  $H_0$ . En otras palabras, si la hipótesis nula es falsa, sólo puede cometerse un error de tipo I; si la hipótesis nula realmente es falsa, sólo puede cometerse un error de tipo II.

Es necesario tener alguna cantidad que mida la posibilidad de cometer alguno de estos errores. Esta medida es una probabilidad.

**Definición 1.1** La probabilidad de rechazar  $H_0$ , dado que  $H_0$  es cierta, se define como la probabilidad (o tamaño) de error de tipo I y se denota por  $\alpha$ ,  $0 \leq \alpha \leq 1$ .

**Definición 1.2** La probabilidad de no rechazar  $H_0$ , dado que  $H_0$  es falsa, se define como la probabilidad (o tamaño) de error de tipo II y se denota por  $\beta$ ,  $0 \leq \beta \leq 1$ .

Por lo tanto, las probabilidades de error de tipo I y tipo II están dadas por las proposiciones

$$P(\text{rechazar } H_0 | H_0 \text{ es cierta}) = \alpha.$$

$$P(\text{no poder rechazar } H_0 | H_0 \text{ es falsa}) = \beta.$$

Cuando una afirmación se incorpora en la proposición de la hipótesis nula, se necesita una regla que indique que decisión tomar con respecto a  $H_0$  una vez que se encuentra disponible la evidencia muestral. Esta regla recibe el nombre de prueba de una hipótesis estadística.

**Definición 1.3** Una *prueba* de una hipótesis estadística con respecto a alguna característica desconocida de la población de interés es cualquier regla para decidir si se rechaza la hipótesis nula con base en una muestra aleatoria de la población.

Por ejemplo, se decide rechazar  $H_0 : \mu = 10$ . Para un tamaño  $n$  dado de la muestra, supóngase que se decide rechazar  $H_0$  si se observa un valor de la media muestral  $\bar{X}$  que sea más grande que 12. Entonces,  $\bar{X}$  es la estadística de prueba, el valor  $\bar{X} = 12$  es el *valor crítico*, y el conjunto de valores mayores que 12 constituyen la región crítica de la prueba.

El área de la región crítica es igual al tamaño del error de tipo I. Es decir,

$$P(\bar{X} > 12 | \mu = 10) = \alpha.$$

La probabilidad de  $\alpha$  es sólo una referencia con respecto a la región  $\bar{X} > 12$ . Pero la decisión de rechazar  $H_0$  se tomará con base en una sola muestra de tamaño  $n$ , a partir de la cual se calculará el estimador de  $\bar{x}$ . De esta forma, si  $\bar{x} > 12$ , esto no significa que la probabilidad de que  $H_0$  sea correcta es  $\alpha$ ; más bien, esto implica una interpretación de frecuencia para  $\alpha$  cuando se toman muchas muestras. En otras palabras, si el valor de  $\mu$  es realmente 10, y si se tomasen en forma repetida muestras de tamaño  $n$  de la población, debe esperarse que en un  $100\alpha\%$  de las veces, se encuentre un valor de la estadística de prueba  $\bar{X}$  mayor que 12 y de esta forma debe rechazarse la hipótesis nula. Sólo en este sentido puede decirse que la confiabilidad al rechazar  $H_0$ , cuando el estimador  $\bar{X} > 12$  es igual al complemento del error  $\alpha$  de tipo I, o  $1 - \alpha$ .

También es necesario establecer una *hipótesis alternativa* que refleje el valor posible o intervalo de valores del parámetro de interés si la hipótesis nula ( $H_0$ ) es falsa. Esto es,  $H_1$  representa alguna forma de negación de  $H_0$ .

Cabe mencionar que el rechazo de la hipótesis nula implicaría que ha sido capaz encontrar evidencia suficiente para garantizar la hipótesis alternativa. Por otro lado, si no se presenta evidencia sustancial, se aceptará la hipótesis nula. Esta decisión no implica necesariamente que "el acusado sea inocente", más bien hace énfasis en la falta de evidencia sustancial necesaria para condenar al acusado. Por lo tanto, en cierto sentido un veredicto de culpable (rechazo de  $H_0$ ) debe ser considerado como una decisión más fuerte que un veredicto de inocente (no rechazar  $H_0$ ). A lo que, el error de tipo I se considera un error mucho más grave que el error de tipo II.

Como resultado se tiene que muchas veces se selecciona con anticipación el tamaño del error de tipo I que puede tolerarse y se intenta construir un procedimiento de prueba que minimice el tamaño del error de tipo II. Es decir, no es posible fijar tanto a  $\alpha$  como a  $\beta$  y diseñar alguna regla de decisión para probar  $H_0$  contra  $H_1$ , dada una muestra aleatoria de tamaño  $n$ . Es por esta razón que se dice "equivocación al

rechazar  $H_0$  mas que aceptar  $H_0$  cuando la evidencia muestral no apoya el rechazo de la hipótesis nula.

También cabe mencionar que el error de tipo II aumenta conforme el error de tipo I disminuya. Y la probabilidad  $\alpha$  del error de tipo I también se conoce como el *nivel de significancia estadístico*. En este contexto la palabra "significancia" sólo implica que la evidencia muestral es tal que garantiza el rechazo de  $H_0$  a un nivel dado  $\alpha$ .

### 1.3 Tipos de regiones críticas y la función de potencia

Se desea estudiar los tipos de regiones críticas que pueden surgir. Considérese la hipótesis nula simple

$$H_0 : \theta = \theta_0$$

con respecto al parámetro de interés  $\theta$ , cuando se muestrea una distribución cuya función de densidad de probabilidad es  $f(x|\theta)$  en donde  $\theta_0$  es el valor propuesto de  $\theta$ . Si la hipótesis alternativa es de la forma

$$H_1 : \theta > \theta_0 \quad \text{o} \quad H_1 : \theta < \theta_0,$$

se dice que  $H_1$  es una hipótesis *alternativa unilateral*. La región crítica también recibe el nombre de región de *rechazo unilateral*. Vale la pena notar que la hipótesis alternativa unilateral debe formularse sólo si el valor de uno de los parámetros que se encuentre en el lado opuesto, no tiene sentido para el investigador. De otro modo, debe establecerse una *hipótesis alternativa bilateral*. Esto es, si la hipótesis alternativa no proporciona una dirección con respecto al valor propuesto de  $\theta_0$ . Es decir,

$$H_1 : \theta \neq \theta_0.$$

El tamaño del error de tipo II se obtiene como una función de los valores alternativos de  $\theta$  ( $\theta >$ ) bajo  $H_1$ . Debe notarse que  $\beta(\theta)$  se conoce como la *función característica de operación*, y cuando se grafica  $\beta(\theta)$  para diversos valores de  $\theta$  de  $H_1$ , se obtiene una curva característica de operación (CO).

Dado que  $\beta(\theta)$  es la probabilidad de que un valor de la estadística de prueba no se encuentre en la región crítica cuando  $H_0$  es falsa, entonces  $1 - \beta(\theta)$  representa la probabilidad de que un valor de la estadística de prueba se encuentre dentro de la región crítica cuando  $H_0$  es falsa. Esta probabilidad se conoce como la *función potencia de la prueba*. En otras palabras, las funciones potencia y característica de operación son complementarias.

**Definición 1.4** La función  $P(\theta) = 1 - \beta(\theta)$  recibe el nombre de función potencia y representa la probabilidad de rechazar la hipótesis nula cuando esta es falsa; es decir, cuando el valor del parámetro de  $H_1$  es cierto.

En esencia, la potencia de una prueba es la probabilidad de detectar que  $H_0$  es, en forma verdadera, falsa; de aquí el uso de la palabra potencia.

### 1.4 Las mejores pruebas

Sea  $X_1, X_2, \dots, X_n$  una muestra aleatoria de tamaño  $n$  de una población cuya función de densidad de probabilidad es  $f(x|\theta)$ , y considérese la hipótesis

$$H_0 : \theta = \theta_0$$

contra

$$H_1 : \theta = \theta_1.$$

en donde se especifican  $\theta_0$  y  $\theta_1$ . Supóngase que  $\alpha$  es el tamaño máximo del error de tipo I que se puede tolerar. Entonces la mejor prueba para  $H_0$  contra  $H_1$  es aquella que tiene el tamaño más pequeño del error de tipo II (y de esta forma la mayor potencia) de entre todas las pruebas que tengan un tamaño del error de tipo I no mayor que  $\alpha$ . Se puede determinar las regiones críticas para estas pruebas mediante el uso del siguiente teorema, el cual se conoce como el **lema de Neyman Pearson**:

**Teorema 1.1** Si existe una región crítica  $C$  de tamaño  $\alpha$  y una constante positiva  $k$  tal que

$$\frac{L_0(x_1, x_2, \dots, x_n; \theta_0)}{L_1(x_1, x_2, \dots, x_n; \theta_1)} \leq k \quad \text{interior } C,$$

$$\frac{L_0(x_1, x_2, \dots, x_n; \theta_0)}{L_1(x_1, x_2, \dots, x_n; \theta_1)} \geq k \quad \text{exterior } C,$$

entonces  $C$  es la mejor región crítica de tamaño  $\alpha$  para probar  $H_0 : \theta = \theta_0$  contra  $H_1 : \theta = \theta_1$ , en donde  $L_0$  y  $L_1$  son las funciones de verosimilitud relativa a  $H_0$  y  $H_1$  respectivamente. ■

## 1.5 Principios generales para probar una $H_0$ simple contra una $H_1$ uni o bilateral

No existen pruebas uniformes más potentes para hipótesis alterativas bilaterales a pesar de que, en forma usual, existen para hipótesis alternativas unilaterales.

### 1.5.1 Principios generales para el caso 1

Considérese la prueba de la hipótesis nula

$$H_0 : \theta = \theta_0$$

contra la alternativa

$$H_1 : \theta \neq \theta_0,$$

donde  $\theta_0$  es el valor propuesto de algún parámetro  $\theta$  bajo  $H_0$ . Dada una muestra aleatoria de tamaño  $n$  de la distribución de interés, el procedimiento general para probar  $H_0$  es escoger el mejor estimador de  $\theta$ ,  $T$  y rechazar  $H_0$  cuando el estimado  $t$  obtenido de la muestra, es en forma suficiente, diferente del valor propuesto de  $\theta_0$ . Esto es, si el estimado  $t$  es lo suficientemente distinto del valor propuesto  $\theta_0$ , entonces se ha observado un evento raro (y la hipótesis nula es correcta), o se ha observado un valor de la estadística que sugiere un valor  $\theta$  diferente del propuesto  $\theta_0$ . Cuando el estimado  $t$  es en forma suficiente distinto de  $\theta_0$ , se asumirá la última posibilidad y se dejará el tamaño del error de tipo II igual a la probabilidad del anterior. En particular, para un tamaño preseleccionado  $\alpha$ , del error de tipo I se obtiene una región crítica bilateral en los extremos de la distribución de muestreo de  $T$ , de esta manera tal que el área, en cualquier lado, más allá del valor crítico es igual a  $\alpha/2$ . Entonces se rechaza  $H_0$  en favor de  $H_1$  cuando el estimado  $t$  se encuentra dentro de la región crítica. Cuando el estimado  $t$  no se encuentra dentro de la región crítica, no puede rechazarse la hipótesis nula. De esta forma, cualquier diferencia con respecto al valor de  $\theta_0$  se considera causada por fluctuaciones en el muestreo del estimador  $T$ .

Este enfoque es muy similar a la construcción de un intervalo de confianza bilateral para  $\theta$ . Para cualquier valor propuesto de  $\theta_0$ , que se encuentre dentro de un intervalo de confianza del  $100(1 - \alpha)\%$  para  $\theta$ ,  $H_0$  no será rechazada. y si se encuentra fuera se rechazara la hipótesis nula.

### 1.5.2 Principios generales para el caso 2

Considérese la prueba de la hipótesis nula

$$H_0 : \theta = \theta_0$$

## 1.6. PRUEBA DE HIPÓTESIS CON RESPECTO A LAS MEDIAS CUANDO SE MUESTREAN DISTRIBUCIONES NORMALES

contra la alternativa

$$H_1 : \theta > \theta_0.$$

Para un tamaño  $\alpha$ , del error tipo I, la región crítica se encuentra localizada en el extremo superior de la distribución de muestreo  $T$  y  $H_0$  se rechaza si el estimado  $t$  no es menor que el valor crítico.

### 1.5.3 Principios generales para el caso 3

Para probar la hipótesis nula

$$H_0 : \theta = \theta_0$$

contra

$$H_1 : \theta < \theta_0,$$

el procedimiento general es rechazar a  $H_0$  cada vez que el estimado  $t$  sea, e forma suficiente menor que el valor propuesto  $\theta_0$ .

Con respecto a la prueba de hipótesis estadísticas, y debido a que se coloca un gran énfasis en el tamaño del error de tipo I generalmente se formula la hipótesis nula tal que esta se rechace si la evidencia experimental apoya esta decisión. Es decir, lo que realmente se desea es concluir que la hipótesis alternativa es la correcta.

En la práctica escogemos el tamaño del error de tipo I antes de la determinación de la muestra aleatoria. Si se obtiene como resultado que la hipótesis nula no puede rechazarse con el valor escogido de  $\alpha$  debe evitarse aumentar el tamaño del error de tipo I con la idea de rechazar la hipótesis nula.

La discusión anterior constituye el método clásico para probar hipótesis estadísticas. El cual, se dirigieron algunas críticas directas hacia este enfoque debido a que la decisión final de rechazar o no una  $H_0$  dada, es demasiado cortante y seca y no proporciona una medida real de que la decisión sea correcta en términos de la probabilidad. Para esto lo que se ha sugerido es el cálculo del llamado **valor p**. El valor p es la probabilidad, dado que  $H_0$  es cierta, de que la estadística de prueba tome un valor mayor o igual que el calculado con base a la muestra aleatoria. Un valor relativamente pequeña puede sugerir que si  $H_0$  es realmente cierta, el valor de la estadística de prueba sea poco probable. Puede entonces optarse por rechazar  $H_0$  debido a que esta decisión tendrá una alta probabilidad de ser correcta.

Se recomienda el cálculo del valor p acoplado con el enfoque clásico de escoger un tamaño del error de tipo I antes de la determinación de la muestra aleatoria. Entonces, la decisión de rechazar o no a  $H_0$  puede basarse en una región crítica de tamaño  $\alpha$ , con el valor p proporcionando una medida real en términos de la probabilidad de que la decisión sea correcta. De acuerdo con lo anterior, se sugiere lo siguiente regla:

Si el valor p es menor o igual a  $\alpha$ , se rechaza  $H_0$ ; de otra forma no puede rechazarse la hipótesis nula.

## 1.6 Prueba de hipótesis con respecto a las medias cuando se muestrean distribuciones normales

### 1.6.1 Pruebas para una muestra

Sea  $X_1, X_2, \dots, X_n$  una muestra aleatoria de una distribución normal con media  $\mu$  desconocida. En este caso el interés recae en probar uno de los siguientes conjuntos de hipótesis con respecto a  $\mu$ .

$$\begin{array}{lll} H_0 : \mu = \mu_0 & H_0 : \mu = \mu_0 & H_0 : \mu = \mu_0 \\ H_1 : \mu \neq \mu_0 & H_1 : \mu > \mu_0 & H_1 : \mu < \mu_0 \end{array}$$

Primero, supóngase que el valor de la varianza poblacional  $\sigma^2$  es conocido. Entonces la estadística de prueba es la media muestral  $\bar{X}$ , misma que, bajo la hipótesis nula, tiene una distribución normal con media  $\mu_0$  y desviación estándar  $\sigma/\sqrt{n}$ . La región crítica de tamaño  $\alpha$  para la hipótesis bilateral es de la forma

$$\text{Rechazar } H_0 \text{ si } \begin{cases} \bar{X} \geq \bar{x}_{1-\alpha/2} \\ \text{o} \\ \bar{X} \leq \bar{x}_{\alpha/2}. \end{cases}$$

donde  $\bar{x}_{1-\alpha/2}$  y  $\bar{x}_{\alpha/2}$  son los valores cuantiles críticos de  $\bar{X}$  de manera tal que

$$P(\bar{X} \geq \bar{x}_{1-\alpha/2}) = \alpha/2 \quad \text{y} \quad P(\bar{X} \leq \bar{x}_{\alpha/2}) = \alpha/2.$$

Dado que bajo  $H_0$ ,  $\bar{X} \sim N(\mu_0, \sigma/\sqrt{n})$ , entonces en forma equivalente

$$P\left(Z \geq \frac{\bar{x}_{1-\alpha/2} - \mu_0}{\sigma/\sqrt{n}}\right) = \alpha/2 \quad \text{y} \quad P\left(Z \leq \frac{\bar{x}_{\alpha/2} - \mu_0}{\sigma/\sqrt{n}}\right) = \alpha/2.$$

o

$$z_{1-\alpha/2} = \frac{\bar{x}_{1-\alpha/2} - \mu_0}{\sigma/\sqrt{n}} \quad \text{y} \quad z_{\alpha/2} = \frac{\bar{x}_{\alpha/2} - \mu_0}{\sigma/\sqrt{n}}$$

en donde  $z_{1-\alpha/2}$  y  $z_{\alpha/2}$  son los correspondientes valores cuantiles de  $Z$ . Por lo tanto, se sigue que  $H_0$  debe rechazarse cuando un valor  $\bar{x}$  de la media muestral  $\bar{X}$  es tal que

$$\bar{x} \geq \frac{\sigma z_{1-\alpha/2}}{\sqrt{n}} + \mu_0 \quad \text{o} \quad \bar{x} \leq \frac{\sigma z_{\alpha/2}}{\sqrt{n}} + \mu_0.$$

De manera equivalente, se rechazará  $H_0$  cuando

$$z \geq z_{1-\alpha/2} \quad \text{o} \quad z \leq z_{\alpha/2},$$

donde  $z = (\bar{x} - \mu_0)/(\sigma/\sqrt{n})$  es el valor de la correspondiente normal estándar al valor  $\bar{x}$  de  $\bar{X}$ .

Para la hipótesis alternativa unilateral  $H_1 : \mu > \mu_0$  la región crítica de tamaño  $\alpha$  es el extremo derecho de la distribución de muestreo  $\bar{X}$ ; esta es de la forma

$$\text{Rechazar } H_0 \text{ si } \bar{X} \geq \bar{x}_{1-\alpha}.$$

en donde  $\bar{x}_{1-\alpha}$  es el valor cuantil de  $\bar{X}$ , tal que  $P(\bar{X} \geq \bar{x}_{1-\alpha}) = \alpha$ . En forma similar, para la hipótesis alternativa  $H_1 : \mu < \mu_0$ , la región crítica de la forma

$$\text{Rechazar } H_0 \text{ si } \bar{X} \leq \bar{x}_{\alpha}.$$

en donde el valor  $\bar{x}$  es tal que  $P(\bar{X} \leq \bar{x}_{\alpha}) = \alpha$ .

Antes de resolver un ejemplo, se desarrollará una expresión general del error de tipo II para uno de los casos. Considérese la hipótesis nula  $H_0 : \mu = \mu_0$  contra la alternativa  $H_1 : \mu > \mu_0$ . Supóngase que en realidad  $\mu = \mu_1 > \mu_0$ . De acuerdo a

$$\text{Rechazar } H_0 \text{ si } \bar{X} \geq \bar{x}_{1-\alpha},$$

no puede rechazarse  $H_0$  si un valor de  $\bar{X}$  es menor que  $(\sigma z_{1-\alpha}/\sqrt{n}) + \mu_0$ . Dado que la probabilidad de error de tipo II es igual a la probabilidad de no rechazar un  $H_0$  falsa, es necesario determinar

$$\beta = P\left(\bar{X} < \frac{\sigma z_{1-\alpha}}{\sqrt{n}} + \mu_0 \mid \mu = \mu_1 > \mu_0\right),$$

## 1.6. PRUEBA DE HIPÓTESIS CON RESPECTO A LAS MEDIAS CUANDO SE MUESTREAN DISTRIBUCIONES NORMALES

la que es términos de la normal estándar es

$$\beta = P \left( Z < \frac{\frac{\sigma z_{1-\alpha}}{\sqrt{n}} + \mu_0 - \mu_1}{\sqrt{n}} + \mu_0 \middle| \mu = \mu_1 \right).$$

Al sustituir cualquier valor  $\mu_1$  de la hipótesis alternativa, se puede calcular el correspondiente valor de la probabilidad de error de tipo II y de esta forma, la potencia.

Hipótesis nula	Valor de la estadística de prueba bajo $H_0$
$H_0 : \mu = \mu_0$	$z = \frac{\bar{x} - \mu_0}{\sigma / \sqrt{n}}$
Hipótesis alternativa	Criterios de rechazo
$H_1 : \mu \neq \mu_0$	Rechazar $H_0$ cuando $z \leq z_{\alpha/2}$ o cuando $z \geq z_{1-\alpha/2}$
$H_1 : \mu > \mu_0$	Rechazar $H_0$ cuando $z \geq z_{1-\alpha}$
$H_1 : \mu < \mu_0$	Rechazar $H_0$ cuando $z \leq z_{\alpha}$

### 1.6.2 Pruebas para dos muestras