Test de Hipótesis

Hasta ahora hemos estudiado el problema de estimar un parámetro desconocido a partir de una muestra aleatoria.

En muchas situaciones se requiere tomar una decisión entre aceptar o rechazar una proposición sobre algún parámetro.

Esta proposición recibe el nombre de ***hipótesis estadística*** y el procedimiento de toma de decisión sobre la hipótesis se conoce como ***prueba o test de hipótesis.***

Hipótesis Estadística

- Es la afirmación acerca de:
    - el valor de un solo parámetro
    - Los valores de varios parámetros
    - La forma de una distribución de probabilidad completa.
- Ejemplos de hipótesis
    - La afirmación "**μ=0.75**" siendo μ el diámetro promedio de una pieza de PVC.
    - La expresión "**p<0.10**" donde p es la proporción de productos defectuosos de una línea de producción.

Ejercicio 9.1.

Cuáles de las siguientes afirmaciones son hipótesis estadísticas válidas?

- $H : \sigma > 100$
- $H : x = 45$
- $H : s \leq 0.20$
- $H : \sigma_1 / \sigma_2 < 1$
- $H : \lambda \leq 0.01$

(siendo λ el parám.de una distrib. exponencial)

Test de Hipótesis

En cualquier problema de test de hipótesis hay **dos hipótesis contradictorias** en consideración.

Ejemplo

- "**μ = 0.75**" y "**μ ≠ 0.75**"
- "**p<0.10**" y "**p $\geq$ 0.10**"

El objetivo es decidir, con base en la información muestral, cual de las dos hipótesis es correcta.

Hipótesis Nula e Hipótesis Alternativa

- La **hipótesis nula**, denotada como $\mathbf{H_0}$, es la afirmación que se supone al principio como cierta.
- La **hipótesis alternativa**, denotada como $\mathbf{H_1}$, es la afirmación que contradice a H_0.
- $\mathbf{H_0}$ se rechaza a favor de $\mathbf{H_1}$ si la evidencia muestral hace pensar que $\mathbf{H_0}$ es falsa; si no, se continua con la creencia de que es verdadera.
- Los dos resultados posibles de un test de hipótesis son: *rechazar H_0 o no rechazar H_0.*

Ejemplo 9.1

El 10% de las tarjetas que produjo cierto fabricante resultaron defectuosas. Se sugieren modificaciones en el proceso de producción con el objetivo de reducir este valor. Identifique $\mathbf{H_0}$ y $\mathbf{H_1}$

Hipótesis Nula

En el test de hipótesis, $\mathbf{H_0}$ siempre será expresada como una afirmación de igualdad.

Si θ denota el parámetro de interés, la hipótesis nula tendrá la forma

$$H_0 : \theta = \theta_0$$

siendo θ_0 un valor dado denominado **valor nulo** del parámetro.

Procedimiento de Prueba

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

Se especifica a partir de

- Un **estadístico de prueba**, una función de los datos muestrales en la que se basará la decisión (rechazar H_0).
- Una **región de rechazo** el conjunto de valores de la estadística de prueba para los que se rechazará H_0.

La hipótesis nula será rechazada si y sólo si el valor del estadístico de prueba cae en la región de rechazo.

Ejemplo de Procedimiento de Prueba

Volviendo al ejemplo de las tarjetas que produjo cierto fabricante. Se sabe que tasa de defectuosas fue del 10%. Se desea efectuar un procedimiento de prueba para ver si las modificaciones introducidas han logrado reducir esta proporción. Es decir que se va a probar

$$H0 : p = 0.10 \text{ contra } H1 : p < 0.10$$

examinando una muestra de 200 artículos.

Se prueba $H_0 : p = 0.10$ contra $H_1 : p < 0.10$

Sea X el número de tarjetas defectuosas de la muestra. $X \sim B(200,0.10)$.

El valor esperado para X es 200*0.10=20 tarjetas

Si H_1 es cierta, se esperan muchos menos.

Un procedimiento de prueba sería por ejemplo rechazar H_0 si $x \leq 15$ y no rechazar en caso contrario.

Errores en la Prueba de Hipótesis

Un **error de tipo I** consiste en rechazar la hipótesis nula H_0 cuando es verdadera.

- En el ejemplo de las tarjetas defectuosas usando como región de rechazo ***x≤15*** aunque ***H_0 : p=0.10* sea cierta** podría ocurrir que una muestra inusual diera como resultado ***x=13***. Es decir que, ***es un error rechazar H_0.***

Un **error de tipo II** consiste en NO rechazar la hipótesis nula H_0 cuando es falsa.

- En el ejemplo de las tarjetas defectuosas usando como región de rechazo ***x≤15*** aunque ***H_1 : p<0.10* sea cierta** podría ocurrir que una muestra inusual diera como resultado ***x=20***. Es decir que, ***es un error NO rechazar H_0.***

Ejemplo 9.2

Para determinar si las soldaduras de una tubería cumplen con las especificaciones se selecciona una muestra de soldaduras. Suponga que las especificaciones establecen que la resistencia promedio de las soldaduras debe pasar de 100 lb/pulg2.

El equipo de inspección decide probar

$H_0 : \mu = 100$ contra $H_1 : \mu > 100$

¿ Por qué podría ser preferible utilizar esta H_1 en lugar de $\mu < 100$?

Ejercicio 9.2

Represente con μ el nivel de radiactividad promedio real. El valor 5 es considerado la línea divisoria entre el agua segura e insegura. Qué recomendaría probar

a) $H_0 : \mu = 5$ contra $H_1 : \mu > 5$

b) $H_0 : \mu = 5$ contra $H_1 : \mu < 5$

Debe plantearse de manera que el error de Tipo I sea más grave que el de tipo II

Ejercicio 9.2

Antes de comprar una gran cantidad de discos se busca tener una prueba concluyente de que la desviación estándar verdadera del diámetro es menor que 0.1 mm. ¿Qué hipótesis debe probar y por qué?

Ejercicio 9.3

Suponga que un fabricante de cigarrillos afirma que el contenido promedio de nicotina μ de los cigarrillos marca B es (a lo sumo) 1.5 mg. Sería desaconsejable rechazar la afirmación del fabricante sin una prueba contradictoria firme. ¿Qué hipótesis debe probar?

Nivel de significación del test

La elección de un valor particular para la región de rechazo fija las probabilidades de que ocurran los errores de tipo I y II. Estas probabilidades se denotan generalmente por α y β.

$$\alpha = P(\text{error de tipo I})$$

$$\beta = P(\text{error de tipo II})$$

A α = *P(error de tipo I)* se lo conoce como ***nivel de significación del test***.

Ejemplo 9.3

Cierto tipo de automóvil ha mostrado ser resistente al 25% de las pruebas de colisión a 10 mph. Buscando incrementar este porcentaje se le han realizado modificaciones.

Sea *p* la proporción de colisiones que puede soportar con las modificaciones de diseño.

¿Qué hipótesis debe probar?

$H_0 : p = 0.25$ ***(niguna mejora)*** **contra** $H_1 : p > 0.25$

La prueba se basará en un experimento que requiere 20 colisiones independientes con prototipos del nuevo diseño.

- **Estadístico de Prueba**

  X= "nro. de colisiones sin daño en las 20 pruebas"

- **Región de rechazo**:

  R = {8,9,10,…, 19, 20}

  es decir se rechaza H_0 si $x \geq 8$

Cuando H_0 es cierta, X="nro. de colisiones sin daño en las 20 pruebas" tiene distribución B(20, 0.25)

Entonces

$$\alpha = P(\text{error de tipo I}) = P(\text{se rechaza } H_0 \text{ cuando es verdadera})$$

$$= P(X \geq 8) = 1 - P(X < 8) = 1 - 0.898 = 0.102$$

Es decir que, cuando en realidad H_0 es verdadera, el 10% de los experimentos que consisten en 20 colisiones darían lugar a un error de tipo I .

Si se busca calcular β = *P(error de tipo II)* se encontrará un valor para cada *p* diferente que pase de 0.25

Por ejemplo, si p=0.3 (es decir X~B(20,0.3))

$$\beta(0.3) = P(\text{error de tipo II cuando } p=0.3) = P(X \leq 7 \text{ cuando } X \sim B(20,0.3)) = 0.772$$

Es decir que, cuando *p es* en realidad 0.3 y no 0.25, más del 70% de los experimentos darían como resultado el rechazo incorrecto de H_1

Si calculamos β para diferentes valores de *p* obtendremos lo siguiente

p	0.3	0.4	0.5	0.6	0.7	0.8
$\beta(p)$	0.772	0.416	0.132	0.021	0.001	0.000

Es decir que β disminuye a medida que *p* se aparta del valor nulo 0.25 .

Si se modifica la región de rechazo, los valores de α y β se modifican

Si se considera R_9={9,10,…, 20}

α = *P(se rechaza H_0 cuando p=0.25)* = *P(X* $\geq$ *9) = 1 – P(X < 9) = 0.041*

Antes era 0.102

β*(0.3) = P(NO se rechaza H_0 cuando p=0.30) = P(X* $\leq$ *8 cuando X~B(20,0.3)) = 0.887*

Antes 0.772

β*(0.5) =P(X* $\leq$ *8 cuando X~B(20,0.5)) = 0.252*

Antes 0.132

Tamaño de la región de rechazo

- Suponga que permanecen fijos un experimento y un tamaño de muestra y se elige un estadístico de prueba.
- La disminución del tamaño de la región de rechazo para obtener un valor **más pequeño** de α dará como resultado un valor **más grande** de β para algún parámetro particular consistente con H_1.
- Es decir que no puede elegirse una misma región que permita reducir los valores de α y β a la vez.
- Dado que generalmente un error de tipo I es más grave que uno de tipo II, se acostumbra fijar el valor más grande de α que pueda ser tolerado y se determina la región de rechazo que tiene ese valor de α y no otro más pequeño.
- Esto hace a β tan chica como sea posible en función del valor de α.

Nivel de significación

- El valor de α se conoce como el **nivel de significación** del test.
- Los niveles tradicionales son 0.10, 0.05 y 0.01.
- En general, cuanto más grave sea el error de tipo I menor debe ser el grado de significación.
- El procedimiento de prueba correspondiente se llama **prueba de nivel** α.
- Una prueba con nivel de significación α es una para la cual el error de tipo I se controla al nivel especificado.

Ejemplo 9.4

Consideremos nuevamente la situación en que μ fue el contenido promedio de nicotina de los cigarrillos marca B. El objetivo es probar

$H_0 : \mu = 1.5$ contra $H_1 : \mu > 1.5$

con base en una muestra aleatoria $X_1, \ldots, X_{32}$ de contenidos de nicotina.

Suponga que se sabe que la distrib. es $N(\mu,\sigma^2)$ con σ=0.20.

El estadístico de prueba será

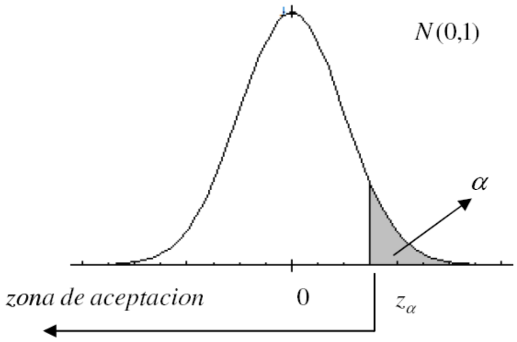
$$Z = \frac{\overline{X} - 1.5}{\sigma / \sqrt{n}} = \frac{\overline{X} - 1.5}{0.20 / \sqrt{32}} = \frac{\overline{X} - 1.5}{0.0354}$$

Si se fija la región de rechazo en $x \geq 1.6$ la regla de decisión sería

$$\begin{cases} \text{Rechazar } H_0 \text{ si } Z \geq 2.825 \\ \text{Aceptar } H_0 \text{ si } Z < 2.825 \end{cases} \Longleftarrow \quad \frac{1.6 - 1.5}{0.0354} = 2.825$$

y la probabilidad de cometer un error de Tipo I sería $\alpha = P(Z \geq 2.825) = 1 - 0.9976 = 0.0024$

Note que la media muestral se aparta de la media real en la misma proporción en que el valor z de Z se aparta de cero. Por lo tanto la forma de la región de rechazo será $z \geq c$.



$$Z = \frac{\overline{X} - 1.5}{0.0354}$$

Si se busca establecer las reglas para un test de nivel α=0.05 sólo resta hallar c tal que

α = *P(se rechaza H_0 cuando es verdadera)*=P($Z \geq c$ cuando Z~N(0,1)) = 0.05

Por lo tanto, c = 1.645 y la región de rechazo será

$$Z = \frac{\overline{X} - 1.5}{0.0354} \geq 1.645 \Rightarrow \overline{X} \geq 1.5 + 0.0354 * 1.645 = 1.5582 \qquad \therefore \quad \overline{X} \geq 1.5582$$

Test de Hipótesis para X~N(μ,σ^2)

Hipótesis nula: $H_0 : \mu = \mu_0$

Valor del estadístico de prueba : $z = \dfrac{\overline{x} - \mu_0}{\sigma / \sqrt{n}}$

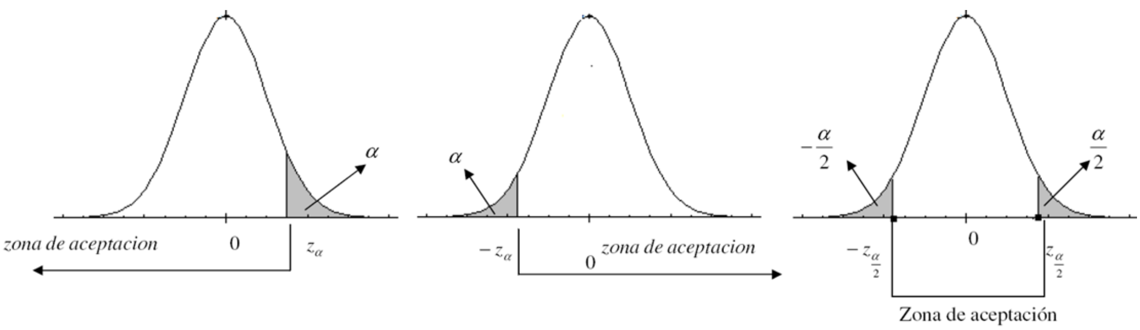
Hipótesis Alternativa	Región de rechazo para test de nivel α
$H_1 : \mu > \mu_0$	$z \geq z_\alpha$ (test de extremo superior)
$H_1 : \mu < \mu_0$	$z \leq -z_\alpha$ (test de extremo inferior)
$H_1 : \mu \neq \mu_0$	$z \leq -z_{\alpha/2}$ o $z \geq z_{\alpha/2}$ (test de dos extremos)

Distribución de probabilidad de Z cuando H_0 es verdadera

Test de extremo superior

Test de extremo inferior

Test de dos extremos



En cada figura el área sombreada corresponde a α = P(error de Tipo I)

Test de Hipótesis para muestras grandes

El estadístico $Z = \dfrac{\overline{X} - \mu}{\sigma / \sqrt{n}}$ puede reemplazarse por $Z = \dfrac{\overline{X} - \mu}{S / \sqrt{n}}$

que tiene **aproximadamente** una distribución normal estándar lo que permitirá establecer la región de rechazo (aproximada).

Para estimar una **proporción** de muestra grande puede utilizarse el estadístico que tiene **aproximadamente** una distribución normal estándar $Z = \dfrac{\hat{P} - p}{\sqrt{\dfrac{p(1-p)}{n}}}$

Ejemplo 9.5

Recientemente , de una muestra de 4115 adultos se encontró que 1276 individuos eran obesos. En una encuesta de 1998, el 20% de los adultos se consideró obeso. ¿Puede decirse que los datos recientes indican que la proporción actual de adultos obesos es más de 1.5 veces el porcentaje de la encuesta de 1998?

Sólución

Para responder se realizará un test de hipótesis con un nivel de significación de 0.10

Sea p = proporción de adultos obesos

Decir que p es 1.5 veces el porcentaje de 1998 es decir que p=30%; por lo tanto

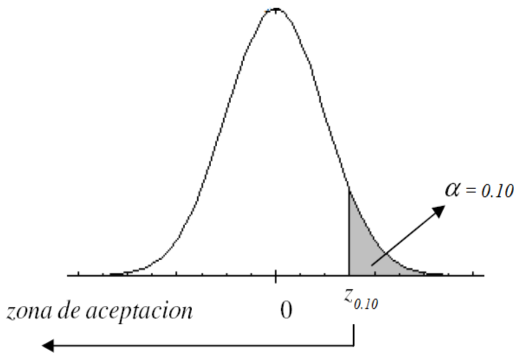
H_0 : p= 0.30

Interesa saber si p supera el 30% ; **H_1 : p > 0.30**

Dado que $np_0=4115(0.3)>10$ y $n(1-p_0)=4115(0.7)>10$ se puede usar la prueba z para muestras grandes. El valor del estadístico de prueba es

$$z = (\hat{p} - 0.3) / \sqrt{(0.3)(0.7)/n} \qquad \hat{p} = 1276/4115 = 0.31$$

$$z = (0.31 - 0.3) / \sqrt{(0.3)(0.7)/4115} = 0.010/0.0071 = 1.40$$



$z_{0.10} = 1.28 \qquad z = 1.40$

La forma de **H_1 : p > 0.30** indica que se trata de una prueba de extremo superior; es decir, que

se rechazará H_0 si **$z > z_{0.10}$ = 1.28**

$z > z_{0.10}$ $\therefore$ se ubica en la zona de rechazo

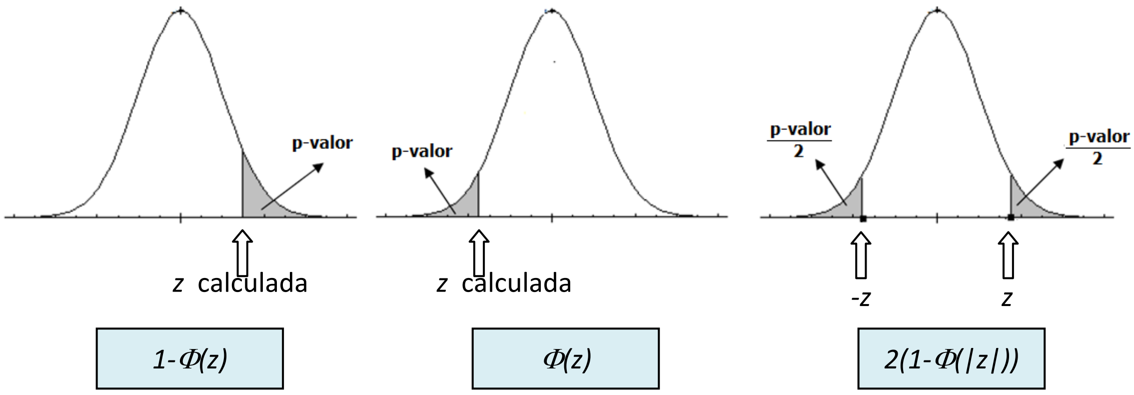
P-valor

Hasta ahora se dieron los resultados de una prueba de hipótesis estableciendo si la hipótesis nula fue o no rechazada con un valor especificado de a o nivel de significación.

Sería de utilidad saber si el valor calculado del estadístico está apenas en la región de rechazo o bien ubicado dentro de ella.

Esto es lo que puede conocerse a través del **p-valor** .

El ***valor p*** o ***p-valor*** es la probabilidad de que el estadístico de prueba tome un valor que sea al menos tan extremo como el valor observado del estadístico de prueba cuando la hipótesis nula es verdadera.



Ejemplo 9.6

El espesor de cierto tipo de placas es de 245 μm. Se toma una muestra de 50 placas y se determina el espesor de cada una. Se obtienen un espesor medio muestral de 246.18 μm y una desviación estándar muestral de 3.60 μm. Se desea saber, a partir de los datos de la muestra, si el espesor promedio real es un poco distinto del valor objetivo.

Parámetro de interés : μ = espesor promedio real

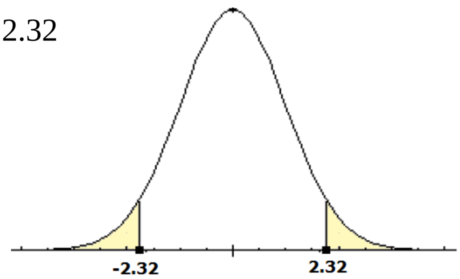
$H_o : \mu = 245$; $H_1 : \mu \neq 245$

Fórmula del estadístico de prueba $z = \dfrac{\bar{x} - 245}{s/\sqrt{n}}$

Valor del estadístico $z = \dfrac{\bar{x} - 245}{s/\sqrt{n}} = \dfrac{246.18 - 245}{3.60/\sqrt{50}} = 2.32$

$$p - valor = 2(1 - \Phi(2.32)) = 0.0204$$

p-valor = 0.0204
es el tamaño del área sombreada



Conclusión:

Usando un nivel de significación de α = 0.01 H_0 no sería rechazada ya que **0.0204 > 0.01**.

A este nivel de significación, la evidencia **es insuficiente** para concluir que el espesor promedio real difiere del valor objetivo.

P-valor

- Notar que:
  - Si a < *p–valor* entonces se acepta H_o con nivel de significancia a.
  - Si a > *p–valor* entonces se rechaza H_o con nivel de significancia a.