PRUEBA DE HIPÓTESIS DE DOS COLAS

Introducción

Este tipo de pruebas son aquellas en las que la hipótesis nula que queremos probar es igual a un valor, mientras que la alternativa puede ser tanto mayor como menor que la nula:

$$H_0: \mu = \theta$$

$$H_0: \mu \neq \theta$$

Se habla de prueba de hipótesis con dos colas, porque la hipótesis alternativa es la que puede ir "para ambos lados", es decir, la hipótesis alternativa nos muestra la opción de que este valor puede ser tanto mayor como menor que el de la hipótesis nula. Por ende las condiciones para nuestros valores de significancia también se verán un poco alterados, ya veremos dicha razón.

Ejemplo

Una empresa que fabrica pelotas y accesorios para golf, trabaja conjuntamente con un campo de golf muy exclusivo en el cual se le exige que las pelotas que fabrica tengan una distancia media de recorrida igual a 295 yardas, ya que es el rendimiento que le exigen los clientes a dicho club de golf. Tanto al dueño de la empresa de fabricación de accesorios de golf como al del campo de golf les conviene que esto se mantenga, pero ha habido algunos problemas con la nueva tecnología adquirida, y se cree que pueden estar produciendo algunas pelotas con distinto rendimiento, por lo que se quiere probar si esto es así o no.

En este caso la hipótesis nula y la alternativa serán:

$$H_0: \mu = 295$$

$$H_0: \mu \neq 295$$

Suponga también, que por estudios hechos anteriormente se ha calculado que la desviación estándar de la población es conocida y es igual a σ=12, el tamaño de la muestra es de n=50, por lo tanto podemos calcular el error estándar como:

$$\sigma_x = \frac{\sigma}{\sqrt{n}} = \frac{12}{\sqrt{50}} = 1.7$$

Además dado que la muestra es lo suficientemente grande, por el teorema central del límite podemos asumir que la distribución de la distancia que recorren las pelotas de golf se aproxima a la normal. Suponga además que quiere hacer las pruebas a un 5% nivel de significancia.

Dado los 50 datos de la muestra también se ha calculado que el valor de $\bar{x} = 297.6$

En este caso como en el caso de la prueba con una cola la prueba se calcula de la misma forma que antes, es decir, dado que asumimos la normalidad de la distribución, y tenemos los datos podemos calcular nuestro estadístico de prueba como antes:

$$z = \frac{\bar{x} - H_0}{\sigma_{\bar{x}}} = \frac{297.6 - 295}{1.7} = 1.53$$

Ya tenemos el valor z de prueba, ahora debemos encontrar los valores de decisión para poder inferir si se puede o no rechazar la hipótesis nula.

Valor crítico

Lo primero que haremos será calcular el método del valor crítico, para esto debemos tomar el nivel de significancia de 5%, es decir α=0.05, pero nótese en este caso que estamos trabajando con una hipótesis alternativa que tiene dos colas, es decir, que nos interesaría tanto saber si el valor puede estar tanto a la izquierda (ser menor) o a la derecha (mayor) de la nula. Por lo tanto, el nivel de significancia que nos importa también se divide entre ambas colas, es decir, el valor que nos interesa es α/2 = 0.05/2 = 0.025, por lo tanto es dicho valor el que buscamos en la tabla de distribución normal. Dicho valor es $\bar{z} = -1{,}96$, pero como es simétrico y hacia ambos lados, también tenemos el mismo valor critico en $\bar{z} = 1{,}96$.

Tal como lo vimos antes, lo que nosotros queremos, es poder ver de qué manera comparar dicho valor con el valor estadístico de prueba que calculamos anteriormente. Como antes el nivel de significancia entrega el nivel máximo que estamos dispuestos a aceptar del error tipo I, por lo tanto dado este valor z, nosotros queremos que este dentro del rango, es decir que sea o mayor que el valor $\bar{z}$ de la izquierda, o menor que el valor $\bar{z}$ de la derecha, por lo tanto la prueba de rechazo seria; Rechazar la hipótesis nula si $z \leq \bar{z}_{0.025}\ o\ z \geq \bar{z}_{0.975}$. O alternativamente no rechazar la hipótesis nula si $\bar{z}_{0.025} \leq z \leq \bar{z}_{0.975}$.

En este caso el valor z que es de 1.53, este valor se encuentra dentro del intervalo donde no podemos rechazar la hipótesis nula, por lo tanto, podemos inferir que estadísticamente no tenemos evidencia para pensar que las pelotas de golf no tienen un rendimiento promedio de 295 yardas al lanzarse.

Valor-p

Alternativamente podemos hacer el mismo ejercicio pero tomando la decisión mediante el cálculo del p-valor, en este caso el p-valor que depende del z, es decir debemos encontrar cual es la probabilidad de estar a la derecha de dicho valor, pero a su vez también cual es la probabilidad de estar a la izquierda, claro que esta vez del valor -1.53, ya que sabemos que al estar ante una prueba con hipótesis alternativa con dos colas, debemos calcular el valor simétrico sobre la distribución. En otras palabras dados estos valores de z, nuestro p-valor será la suma de las probabilidades de estar a la izquierda de -1.53 y a la derecha de 1.53, es decir:

$$p - valor = P(z \leq -1.53) + P(z \geq 1.53)$$

$$p - valor = 0.063 + 0.063 = 0.126$$

Como en el caso anterior, el criterio de rechazo de la hipótesis nula será: rechazar la hipótesis nula si p-valor $\leq \alpha$, dicho de otra manera, si el p-valor es mayor al nivel de significancia no podemos rechazar la hipótesis nula. Como en este caso el nivel de significancia es de 0.05, y el p-valor 0.126, no podemos rechazar la hipótesis nula.

Resumen

Hipótesis nula	*Valor de la estadística de prueba bajo H_0*
$H_0: \mu = \mu_0$	$z = \dfrac{\bar{x} - \mu_0}{\sigma/\sqrt{n}}$
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 $z \leq z_{\alpha}$

Stata

Vamos a seguir ocupando la base de datos CASEN. Fijémonos en varios ejemplos

ttest ytotaj=500000 if edad<30
ttest ytotaj=500000 if ecivil==1
ttest ytotaj=200000 if edad<20

Para cada uno de estos resultados, ¿qué podemos inferir de la hipótesis alternativa con dos colas?

```
. ttest ytotaj=500000 if edad<30


One-sample t test
------------------------------------------------------------------------------
Variable |     Obs        Mean    Std. Err.   Std. Dev.   [95% Conf. Interval]
---------+--------------------------------------------------------------------
  ytotaj |    1269    353411.3    13823.47      492434    326291.9    380530.7
------------------------------------------------------------------------------
    mean = mean(ytotaj)                                           t = -10.6043
Ho: mean = 500000                                degrees of freedom =     1268

  Ha: mean < 500000            Ha: mean != 500000              Ha: mean > 500000
 Pr(T < t) = 0.0000         Pr(|T| > |t|) = 0.0000          Pr(T > t) = 1.0000


. ttest ytotaj=500000 if ecivil==1

One-sample t test
```

```
------------------------------------------------------------------------------
Variable |     Obs        Mean    Std. Err.   Std. Dev.   [95% Conf. Interval]
---------+--------------------------------------------------------------------
  ytotaj |    1874      879319    31586.56     1367374    817370.4    941267.5
------------------------------------------------------------------------------
    mean = mean(ytotaj)                                           t =  12.0089
Ho: mean = 500000                                degrees of freedom =     1873

  Ha: mean < 500000           Ha: mean != 500000            Ha: mean > 500000
 Pr(T < t) = 1.0000         Pr(|T| > |t|) = 0.0000         Pr(T > t) = 0.0000

.
. ttest ytotaj=200000 if edad<20

One-sample t test
------------------------------------------------------------------------------
Variable |     Obs        Mean    Std. Err.   Std. Dev.   [95% Conf. Interval]
---------+--------------------------------------------------------------------
  ytotaj |     341    63721.97    5866.154    108325.5    52183.45     75260.5
------------------------------------------------------------------------------
    mean = mean(ytotaj)                                           t = -23.2312
Ho: mean = 200000                                degrees of freedom =      340

  Ha: mean < 200000           Ha: mean != 200000            Ha: mean > 200000
 Pr(T < t) = 0.0000         Pr(|T| > |t|) = 0.0000         Pr(T > t) = 1.0000
```