Tema 7 : Pruebas de Hipótesis

Contenido

- Concepto de Hipótesis.
- Hipótesis con alternativa única y múltiple.
- Región de aceptación y región de rechazo de la hipótesis planteada.
- Errores tipo I y II. P-valor.
- Pruebas que involucran una sola población

¿Qué es una Hipótesis?

Hipótesis

- La vida media de las computadoras es 6 años.
- El consumo moderado de alcohol en las comidas reduce el riesgo de infarto de miocardio.
- El 20% de las mujeres maneja mal.

La hipótesis es una suposición o conjetura respecto de lo que se observa en la realidad.

Las hipótesis son las preguntas que se hace el investigador sobre el tema por indagar.

Hipótesis Estadística

Una *hipótesis estadística* es una afirmación, generalmente sobre los parámetros de una población.

Por ejemplo, la afirmación el “20% de las mujeres maneja mal” se refiere a la población de todas las mujeres que conducen un automóvil particular o algún otro tipo de vehículo.

¿Puede alguien saber exactamente cual es ese porcentaje? Nadie puede saberlo; por lo tanto, esa afirmación no corresponde necesariamente a un hecho real. Se trata de una hipótesis, la cual deber ser comprobarda.

Hipótesis Estadística

La estadística propone la formulación de 2 hipótesis. Las denominadas Hipótesis Nula e Hipótesis Alternativa. Aunque luego el análisis se realiza sobre la hipótesis nula, la cual se aceptará o se rechazará. En caso de rechazarla se considera cierta la hipótesis alternativa.

Hipótesis Nula **H_0**: establece una igualdad, se plantea lo conocido o lo que se sabe hasta ese momento.

Hipótesis alternativa **H_1**: o **H_a**: describe generalmente la gama de posibilidades que puedan ser verdades cuando H_0 es falsa.

Ejemplo de Hipótesis

La proporción poblacional que prefiere el producto de “Morrell” supera el 50 %.

Hipótesis de interés o de investigación

$$H_0: p = 0.50$$

$$Ha: p > 0.50$$

Hipótesis estadística

Como verán, en la hipótesis nula se plantea una igualdad y representa lo que “ya es sabido” generalmente por estudios previos y en la hipótesis alternativa se plantea lo que se quiere demostrar. Existe como un protocolo establecido que para demostrar algo “nuevo” hay que tener evidencia con los datos que hipótesis nula no es cierta.

Prueba de Hipótesis

La ***prueba de la hipótesis*** estadística es un proceso formalizado que utiliza la información en una muestra para decidir si aceptar o rechazar la *hipótesis nula* H_0.

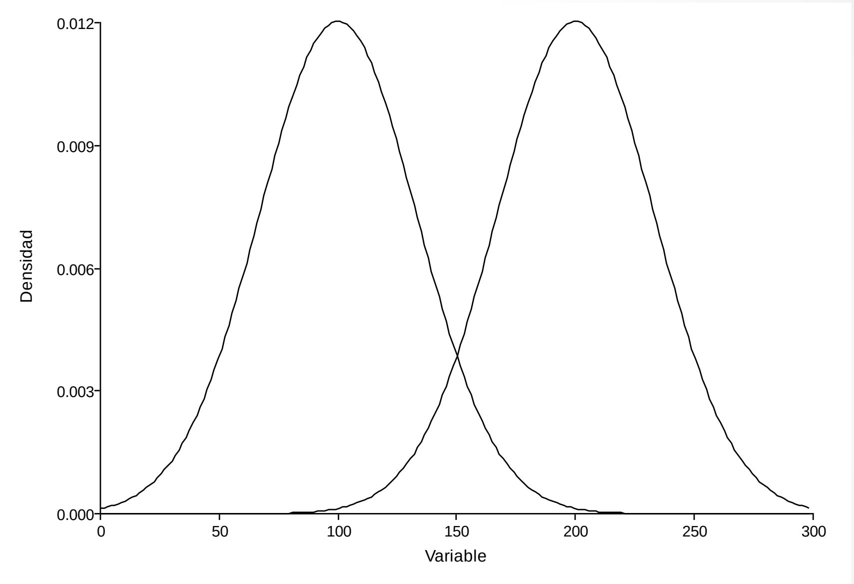
Se dedica a conocer la probabilidad de ocurrencia del resultado obtenido en el estudio del problema, basándose en una distribución (normal o muestral) de la estadística utilizada para medir un resultado.

PRUEBAS DE HIPÓTESIS

- Se basan en formular al principio de la investigación de forma clara la llamada **hipótesis nula o H_0** que someteremos a comprobación o contraste para evaluar su grado de certeza. Tenemos también la necesidad de formular la **hipótesis alternativa o H_1.** Podemos decir que una prueba de hipótesis es un procedimiento de elección entre H_0 y H_1.

Razonamiento para realizar una Prueba de Hipótesis

Suponga dos distribuciones de probabilidad y no se sabe cuál es la que corresponde a la población bajo estudio.



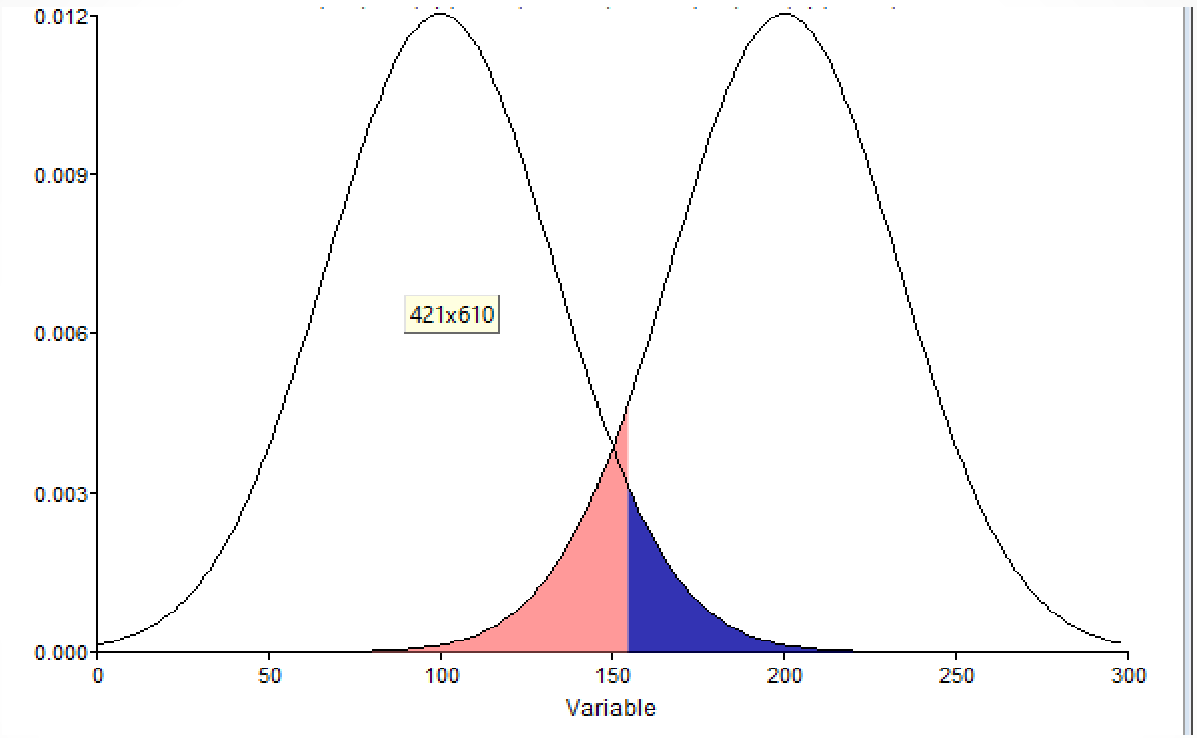
La diferencia entre ambas distribuciones es el parámetro μ

Se plantean las siguientes <u>hipótesis estadísticas,</u>
$H_0: \mu = 100 \quad H_a: \mu > 100$

Se <u>obtienen datos</u> para decidir sobre cual hipótesis se debe aceptar. Dicha decisión será a partir del cálculo de un estadístico y de su probabilidad (P-Valor)

<u>Regla de Decisión</u>
Deteminar un valor de probabilidad α valor que constituye un lìmite y decimos de decisión.
Si **P-valor < α** entonces **Rechazo H_0**

Razonamiento para realizar una Prueba de Hipótesis



La distribución de la izquierda sería lo que plantea H0 y la de la derecha lo que plantea Ha (en realidad Ha plantea cualquier distribución cuya media sea mayor que 100).

Entonces si el valor promedio obtenido tiene poca probabilidad bajo H0, la rechazamos y decimos que tiene mas probabilidad de pertenecer a otra distribución.

Errores Estadísticos

La estadística no es una ciencia exacta, es una ciencia basada en probabilidades, por lo tanto siempre habrá un margen de error. Este margen lo decide el investigador y en base a eso propone un criterio para la decisión estadística

Cuando se toma la decisión de aceptar o rechazar la hipótesis nula, se pueden cometer errores ya que se desconoce la verdadera situación poblacional.

Atendiendo a los resultados del estadístico se decide:

		No rechazar	Rechazar
H_0	Verdadera	Decisión correcta	Error tipo I (α)
	Falsa	Error tipo II (β)	Decisión correcta

Cuando los datos nos conducen a RECHAZAR la H_0 y H_0 es VERDADERA se comete el error **α.**

Cuando los datos nos conducen la ACEPTAR a H_0 y ella era FALSA se cometе el error **β.**

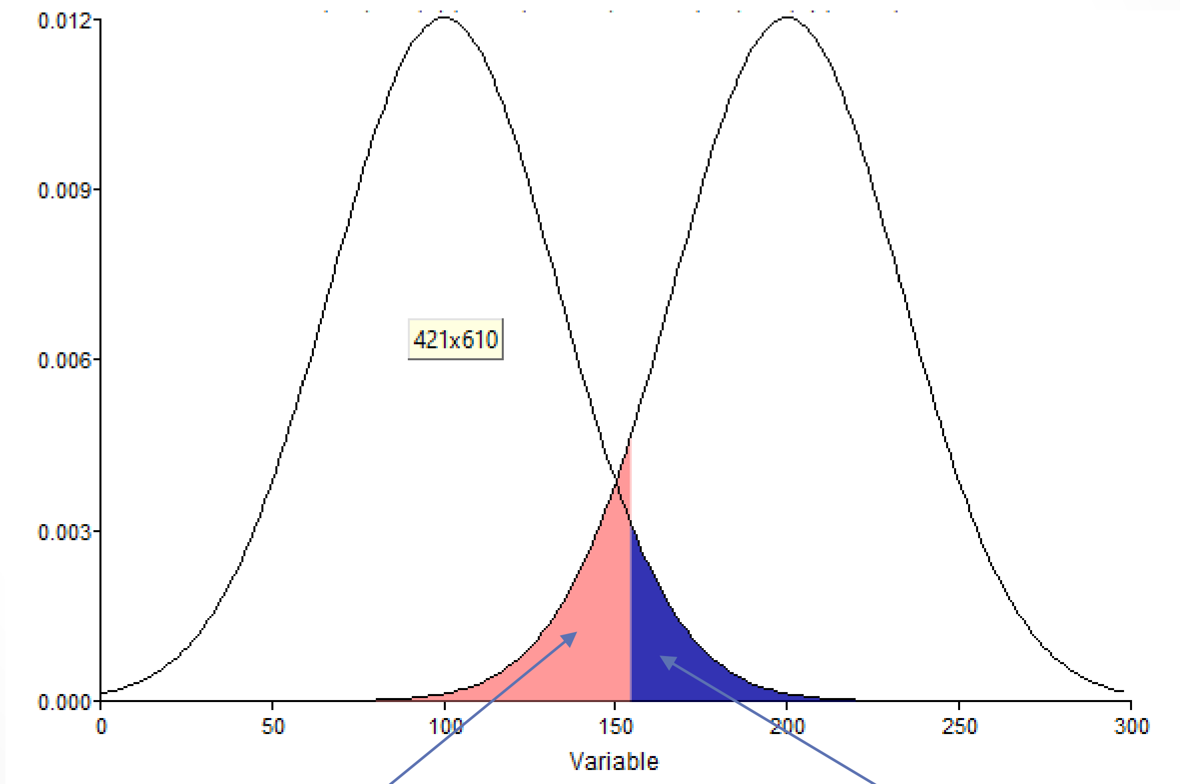
La hipótesis H_0 propone igualdad y la H_a puede ser del tipo > ó < ó ≠

Visualización de los Errores Estadísticos

Se muestra a continuación los tipos de error que pueden cometerse al tomar una decisión. La hipótesis H_a puede ser del tipo > ó < ó ≠

H_0:μ = 100

H_a:μ >100



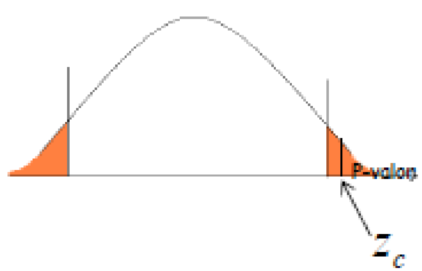
Error de Tipo II o β.

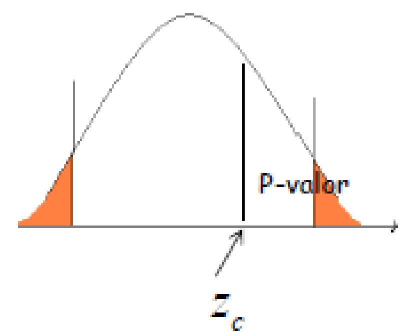
Error de Tipo I o α.

P- valor

El P- valor es la probabilidad, en la distribución teórica, de obtener un valor mas extremo que el encontrado.

Se debe entender que un valor muy pequeño de P es una evidencia que la hipótesis nula no es cierta y la debemos rechazar





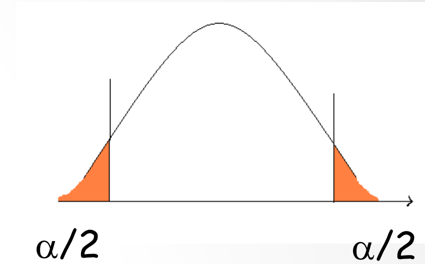
Hipótesis Alternativas

La H_a puede ser del tipo $\neq$, $<$ ó $>$

- **Alternativa múltiple :**

$H_0: \theta = \theta_0$
$H_a: \theta \neq \theta_0$

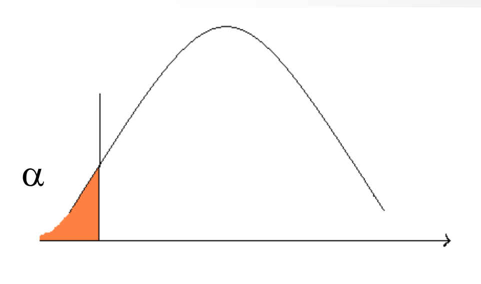
Pruebas de dos colas



- **Alternativa única :**

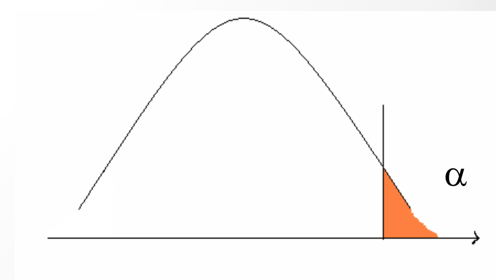
$H_0: \theta = \theta_0$
$H_a: \theta < \theta_0$

Pruebas de una cola



O bien

$H_0: \theta = \theta_0$
$H_a: \theta > \theta_0$



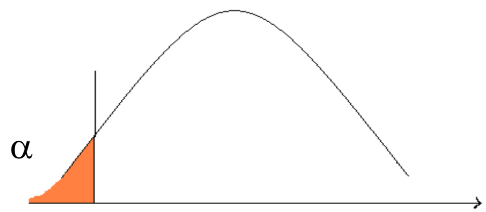
Procedimiento para probar hipótesis

- Formulación de Hipótesis.
- Especificación del Error de tipo I (α) expresado en términos de probabilidad, también se denomina Nivel de Significación.
- Selección del estadístico de prueba y en consecuencia de la Distribución de Probabilidades que se utilizará.
- Establecimiento del criterio de decisión.
- Toma de datos y realización de los Cálculos.
- Toma de decisión

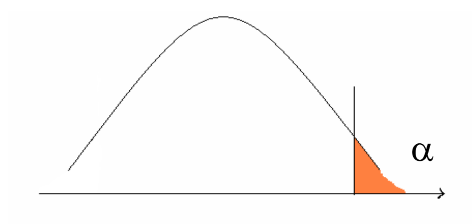
Pruebas que involucran una sola población

Estas pruebas se refieren a las hipótesis donde μ es igual a un número determinado simbolizado como μ_0, se plantean 3 situaciones posibles tendiendo en cuenta la la hipótesis alternativa.

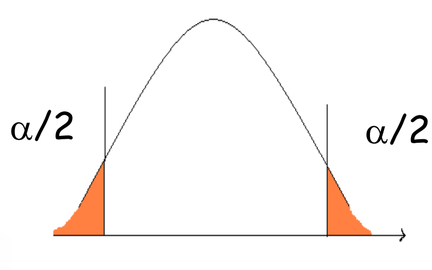
- $H_o: \mu = \mu_0$
  $H_a: \mu < \mu_0$



- $H_o: \mu = \mu_0$
  $H_a: \mu > \mu_0$



- $H_o: \mu = \mu_0$
  $H_a: \mu \neq \mu_0$



Estadístico

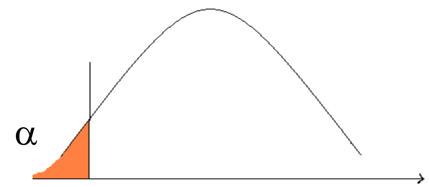
$$z = \frac{\bar{x} - \mu}{\sigma / \sqrt{n}}$$

En los tres casos el Estadístico de prueba será Z si n>30 o T si n<30. Y tendremos que obtener el P-Valor de Z o T y compararlo con α si estamos en una de las 2 primeras situaciones o $\alpha/2$ en la ultima.

Pruebas que involucran una sola población

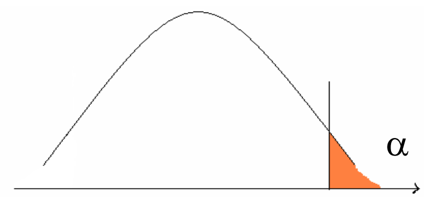
Estas pruebas se refieren a las hipótesis donde π (una proporción) es igual a un número determinado simbolizado como π_0, se plantean 3 situaciones posibles tendiendo en cuenta la la hipótesis alternativa.

- $H_o: \pi = \pi_0$
  $H_a: \pi < \pi_0$

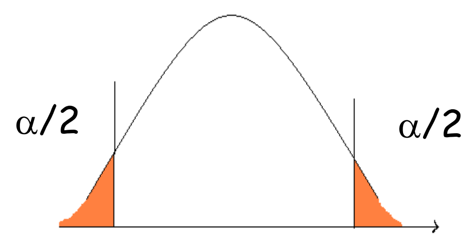


- $H_o: \pi = \pi_0$
  $H_a: \pi > \pi_0$

Unilateral



- $H_o: \pi = \pi_0$ Bilateral
  $H_o: \pi \neq \pi_0$



Estadístico:

$$z = \frac{p - \pi_0}{\sigma_p}$$

recordando

$$\sigma_p = \sqrt{\frac{\pi_0(1-\pi_0)}{n}}$$

La decisión es con la distribución Z o T, calculando el P Valor y comparándolo con α.

Ejemplo

Nos dicen que la vida media de las computadoras es 5 años pero sospechamos que es menor.

Consideramos la afirmación: “la vida media de las computadoras es 5 años” como la **hipótesis nula** (**H0**)y a nuestra sospecha: “la vida media de las computadora es menor a 5 años” como **hipótesis alternativa** (**Ha**). Llamando **μ** a la media de la vida de todas las computadoras, entonces las hipótesis se escriben:

H0: μ = 5 y **Ha: μ < 5**

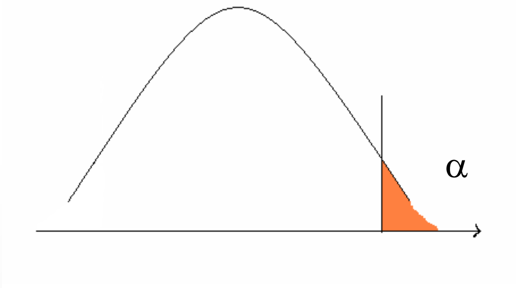
Una vez que se establecen la hipótesis nula y la alternativa, el paso siguiente consiste en hallar la evidencia para tomar la decisión. La calidad de los datos es fundamental; la información debe ser precisa y no tener sesgo. Una mayor precisión se obtiene con un mayor tamaño de muestra: para evitar el sesgo los datos deben provenir de un muestreo aleatorio simple.

Supongamos que elegimos **n=36** computadoras al azar obteniendo una media muestral $\overline{\boldsymbol{x}}$ = 4,33 años y un desvío estándar **s=1,12 años.** Queremos decidir si la diferencia entre 4,33 y el valor especificado en la hipótesis nula (μ=5) es atribuible al azar (debido a haber tomado una muestra) o tenemos suficiente evidencia para rechazar la hipótesis nula a favor de la alternativa planteada.

Debemos construir el **estadístico del test. Si la hipótesis nula es verdadera** y como n> 30 sabemos que la media muestral $\overline{\boldsymbol{x}}$ tiene una distribución aproximadamente Normal con media **μ=5** y desvío $\sigma/\sqrt{n}=1{,}12/6$

Entonces $Z=(\overline{\boldsymbol{x}}-\mu)/(s/\sqrt{n})=(4.33-5)/1.12/6)=3.589$

P-valor(3.589)=0.0001 ➔ Rechazo H0



Ejemplo

Una mezcladora-envasadora de fertilizantes produce una mezcla con 10,0 kg. promedio de nitratos por cada 100,0 kg. de producto. Se desea saber si la cantidad de nitratos de la mezcla no ha variado.
Para averiguarlo se toma una muestra de 10 envases, la que arroja un promedio de 11,3% de contenido de nitratos y un S_x= 1,25.

$$H_0 : \bar{x} = \mu$$

$$H_a : \bar{x} \neq \mu$$

$$t_c = \frac{\bar{x} - \mu}{Sx / \sqrt{n}} = \frac{11{,}3 - 10{,}0}{1{,}25 / \sqrt{10}} = 3{,}29$$

Ahora buscamos el P-Valor en la distribución T con 9 grados de libertad, y comparándolo con α, P-valor de (3.29)=0.005, entonces rechazo H_0.