2 - Pruebas de Hipótesis

16 July 2020 17:12

Necesarias cuando se debe decidir si una afirmación relativa a un parámetro es verdadera o falsa. Es una prueba de hipótesis *relativa* a un parámetro.
Por ahora son relativas a la media

Ej:
Hipótesis nula: $\mu = 20$
Hipótesis alternativa: $\mu > 20$
Se toman la media y la desviación estándar muestrales y se calcula el valor z para usar tablas normalizadas

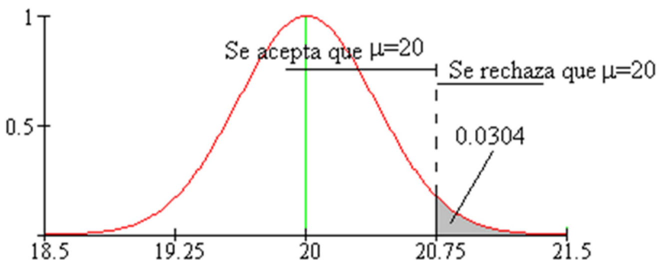
$$z = \frac{\bar{x} - \mu}{\frac{\sigma}{\sqrt{n}}} \longrightarrow t = \frac{\bar{x} - \mu}{\frac{s}{\sqrt{n}}} \longrightarrow$$

Si:

- Tamaño pequeño de muestra
- Se desconoce σ
- Proviene de una población normal

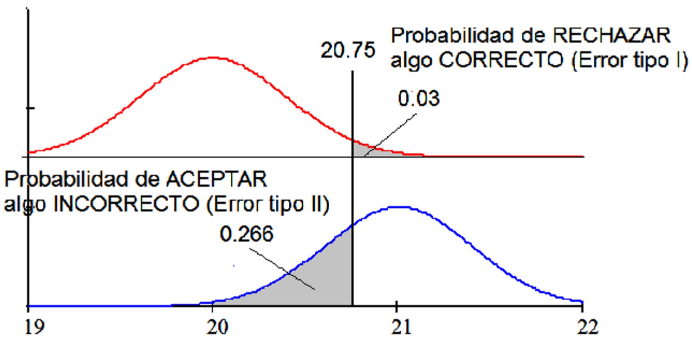
Con $\nu = n - 1$ grados de libertad

Para generar el criterio:



El área bajo la curva en la región de rechazo (α), es la ***probabilidad de rechazar erróneamente a la hipótesis nula***. *Error de Tipo I*

Si ***resulta que la media poblacional real es otro valor***, se toma el valor z otra vez.
El área bajo la curva en la región de aceptación (β), representa la ***probabilidad de equivocarse en la aceptación***. *Error de Tipo II*



	Se Acepta H	Se Rechaza H
H es Verdadera	Decisión Correcta	Error de Tipo I
H es Falsa	Error de Tipo II	Decisión Correcta

Su probabilidad se llama Nivel de Significación (α).

A menor Error de Tipo I, mayor Error de Tipo II (β)

Valor P = 1 - α

Si no se puede rechazar la hipótesis nula y se decide *reservar la decisión*, **no** hay posibilidad de este error

Hipótesis:

- Nula: Proposición que *se desea probar* como cierta o falsa
- Alternativa: Hipótesis que se aceptara cuando la *hipótesis nula* debe ser *rechazada*
- Simple: Si los dos subconjuntos de hipótesis (nula y alternativa) se componen de un solo elemento
- Compuesta: Si al menos un subconjunto contiene mas de un elemento

Prueba de significancia:

Pruebas diseñadas para poder determinar que una discrepancia puede razonablemente atribuirse o no al azar.

Diseñada para determinar si un estadístico es estadísticamente (significativamente) diferente a un valor.

2 Tipos:

- Unilateral o de 1 cola: Hipótesis alternativa donde el parámetro puede ser < o > que el propuesto en la Hipótesis Nula
- Bilateral o de 2 colas: Hipótesis alternativa donde el parámetro es ≠ que el propuesto en la Hipótesis Nula

Hipótesis alterna	Se rechaza H0 si
$\mu < \mu_o$	$z < -z_{\alpha}$
$\mu > \mu_o$	$z > z_{\alpha}$
$\mu <> \mu_o$	$z < -z_{\alpha/2}$ ó $z > z_{\alpha/2}$

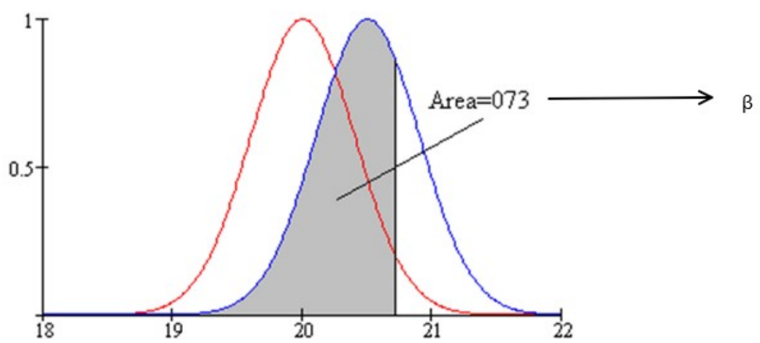
Preceptos para la resolución sistemática de problemas de Prueba de Hipótesis

1. Se formula una Hipótesis Nula simple y una Alternativa apropiada.
2. Especificar la probabilidad de Error Tipo I (si es posible) y se pueden especificar también las probabilidades de Error Tipo II, para alternativas particulares.
3. Construir un criterio para aprobar la Hipótesis Nula.
4. Calcular el valor del estadístico sobre el cual se basa la decisión.
5. Decidir. Se puede aceptar la hipótesis nula, rechazarla o abstenerse de tomar una decisión.

Curvas características de operación

Análisis para cola derecha

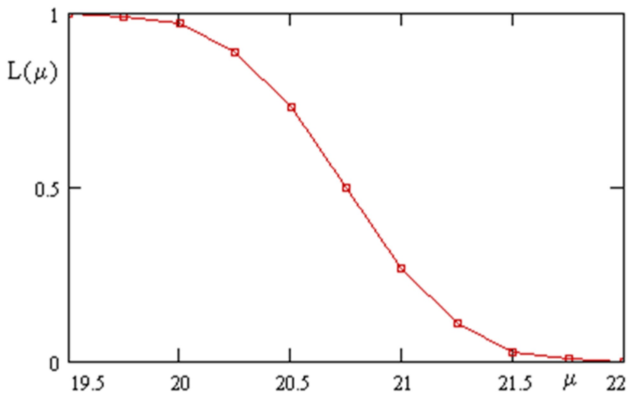
Hasta ahora no se atendieron los Errores de Tipo II
Si se toma una hipótesis alternativa donde el parámetro es **mayor** que un valor.
Se puede *analizar* que sucede con el Error de Tipo II para *distintos valores* del parámetro analizado.



Parámetro ← | → β

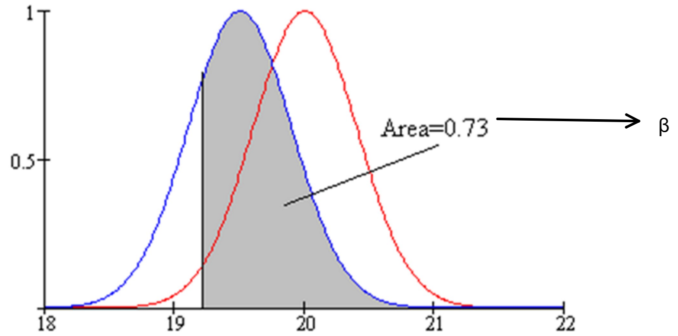
Valor de $\cancel{\mu}$	Valor de z	Probabilidad de Aceptar la Ho $\cancel{L(\mu)}$
19.50	3.125	0.999
19.75	2.5	0.99
20.00	1.875	0.97
20.25	1.25	0.89
20.50	0.625	0.73
20.75	0	0.50
21.00	-0.625	0.27
21.25	-1.25	0.11
21.50	-1.875	0.03
21.75	-2.5	0.01
22.00	-3.125	0.001

Se obtiene un grafico como el siguiente.



Análisis para cola izquierda

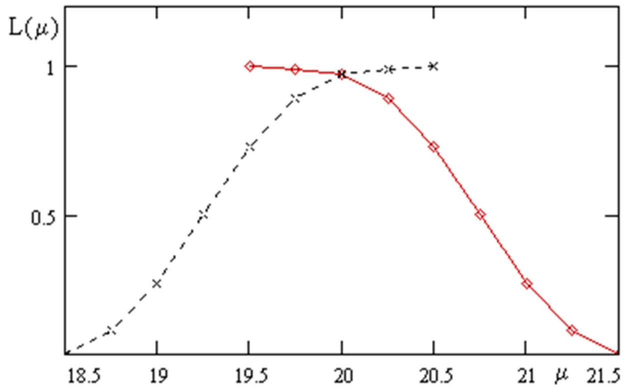
Si se toma una hipótesis alternativa donde el parámetro es **menor** que un valor.
Se puede *analizar* que sucede con el Error de Tipo II para *distintos valores* del parámetro analizado.



Parámetro ← | → β

Valor de $\cancel{\mu}$	Valor de z	Probabilidad de Aceptar la Ho $\cancel{L(\mu)}$
18.50	-1.875	0.03
18.75	-1.25	0.11
19.00	-0.625	0.27
19.25	0	0.5
19.50	0.625	0.73
19.75	1.25	0.89
20.00	1.875	0.97
20.25	2.5	0.99
20.50	3.125	0.999

Si se *combina* el grafico que generan estos datos con los del análisis para cola derecha, se obtiene el siguiente gráfico



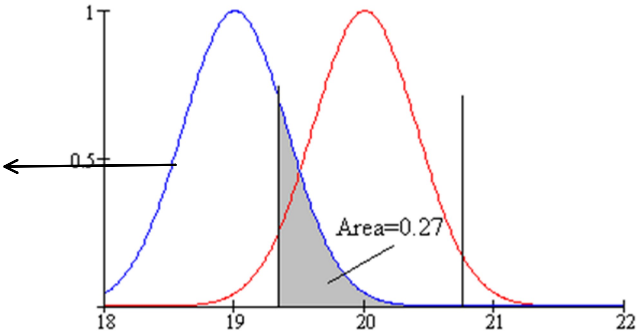
Esta comparación muestra que ambos análisis son la imagen espejo uno del otro.

Análisis para dos colas

Si se toma una hipótesis alternativa donde el parámetro es **distinto** que un valor.
Se puede *analizar* que sucede con el Error de Tipo II para *distintos valores* del parámetro analizado.

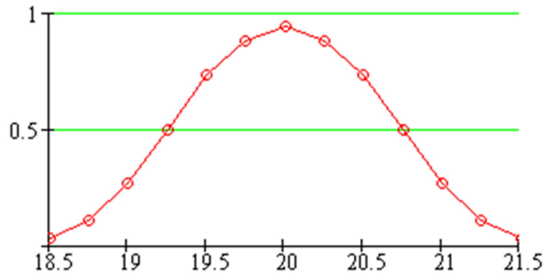
En este ejemplo se tiene como criterio $\bar{x} = 19.25$ y $\bar{x} = 20.75$ con μ = 20

Esta curva se va moviendo a medida que cambiamos el valor del parámetro



Valor de μ	Valor de z_1	Valor de z_2	Probabilidad de Aceptar la Ho L(μ)
18.50	1.875	5.625	0.03
18.75	1.25	5	0.106
19.00	0.625	4.375	0.27
19.25	0	3.75	0.5
19.50	-0.625	3.125	0.733
19.75	-1.25	2.5	0.88
20.00	-1.875	1.875	0.939
20.25	-2.5	1.25	0.88
20.50	-3.125	0.625	0.733
20.75	-3.75	0	0.5
21.00	-4.375	-0.625	0.27
21.25	-5	-1.25	0.106
21.50	-5.625	-1.875	0.03

Lo que se ve gráficamente como:

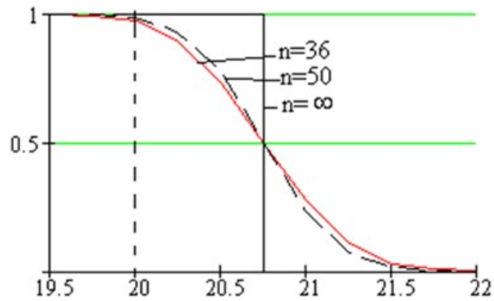


Dado que z1 y z2 obtienen gráficos espejo (como vimos en el análisis anterior), el gráfico producto de un análisis para dos colas equivale a la combinación ambos análisis.

Análisis para muestras de mayor tamaño

Si a los ejemplos anteriores se los realiza a partir de una muestra mayor se obtendrían resultados como los siguientes

Valor de $\ddot{\mu}$	Valor de $\ddot{z}$	Probabilidad de Aceptar la Ho $L(\mu)$
19.50	3.683	1
19.75	2.946	0.998
20.00	2.21	0.986
20.25	1.473	0.93
20.50	0.737	0.769
20.75	0	0.50
21.00	-0.737	0.231
21.25	-1.473	0.07
21.50	-2.21	0.014
21.75	-2.946	0.0016
22.00	-3.683	0

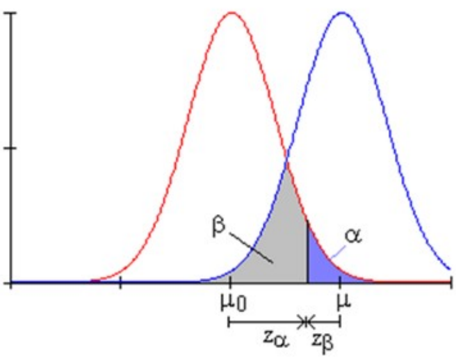


Por lo que a *mayor tamaño* de muestra, *mas próximo* se está al *parámetro* sujeto de la prueba

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

Algoritmo para el trazado de las curvas de operación

Se pretende poder graficar el error tipo II en su forma mas *general* para *distintos niveles de significación*



En el esquema podemos observar:

$$z_\alpha = \frac{\bar{x} - \mu_0}{\frac{\sigma}{\sqrt{n}}} \qquad z_\beta = \frac{\bar{x} - \mu}{\frac{\sigma}{\sqrt{n}}}$$

Al restarlos entre sí obtenemos:

$$z_\alpha - z_\beta = \frac{\mu - \mu_0}{\sigma}\sqrt{n}$$

Llamando **d** a $\frac{|\mu-\mu_0|}{\sigma}$, podemos obtener la siguiente función:

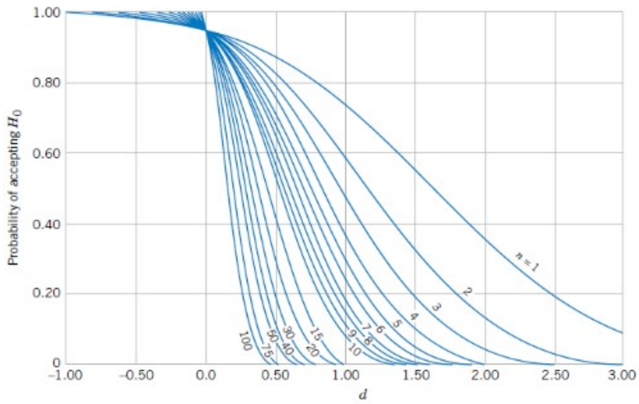
$$z_{\beta_{(d)}} = z_\alpha - d\sqrt{n}$$

Finalmente el error de tipo II es:

$$\boxed{\beta(d,n,\alpha) = 0.5 + \int_0^{qnorm(1-\alpha,0,1)-d\cdot\sqrt{n}} dnorm(x,0,1)\,dx}$$

Se han hecho gráficos para calcular el Error de Tipo II para distintos valores de **d**, **tamaños muestrales** (n) y valores de **nivel de significancia**, como parámetros, para muestras de una cola y de dos colas.

Para 1 cola y α =0.05:



Hipótesis relativa a dos medias

Se consideran *dos poblaciones* con medias μ_1 y μ_2 , y varianzas $\sigma^2{}_1$ y $\sigma^2{}_2$.
Las muestras tendrán tamaño n_1 y n_2

Hipótesis Nula:
- $\mu_1 - \mu_2 = \delta$

Hipótesis Alternativa:
- $\mu_1 - \mu_2 \neq \delta$
- $\mu_1 - \mu_2 < \delta$
- $\mu_1 - \mu_2 > \delta$

Hipótesis Alterna	Se rechaza la Hipótesis Nula si:
$\mu_1-\mu_2<\delta$	$z < -z_\alpha$
$\mu_1-\mu_2>\delta$	$z > z_\alpha$
$\mu_1-\mu_2<>\delta$	$z < -z_{\alpha/2}$ ó $z > z_{\alpha/2}$

Tomando δ como una *constante* (generalmente 0)

Por lo que la prueba dependerá de las diferencias entre las medias muestrales. $\overline{x_1} - \overline{x_2}$
Si ambas provienen de **poblaciones normales** se define el estadístico:

$$z = \frac{\left[\bar{x}_1 - \bar{x}_2\right] - \delta}{\sigma_{\bar{x}_1-\bar{x}_2}}$$

$\sigma_{\overline{x_1}-\overline{x_2}}$ es la desviación estándar de la *distribucion muestral de la diferencia* entre las medias muestrales.

La distribución de la suma (o resta) de dos distribuciones de variables aleatorias tiene:
- $\mu_{\mu_1-\mu_2} = \mu_1 - \mu_2$
- $\sigma^2{}_{\sigma_1{}^2-\sigma_2{}^2} = \sigma_1{}^2 + \sigma_2{}^2$

Si se sabe que:

$$\sigma^2_{\bar{x}_1} = \frac{\sigma_1^2}{n_1} \qquad \sigma^2_{\bar{x}_2} = \frac{\sigma_2^2}{n_2}$$

Entonces:

$$\sigma^2_{\bar{x}_1-\bar{x}_2} = \frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}$$

$$\sigma^2_{\bar{x}_1-\bar{x}_2} = \frac{\sigma_1{}^2}{n_1} + \frac{\sigma_2{}^2}{n_2}$$

Luego:

$$\boxed{z = \frac{\left[\bar{x}_1-\bar{x}_2\right]-\delta}{\sqrt{\frac{\sigma_1{}^2}{n_1}+\frac{\sigma_2{}^2}{n_2}}}}$$

Si las muestras son grandes pero se desconoce σ_1 y σ_2:

$$z = \frac{\left(\bar{x}_1-\bar{x}_2\right)-\delta}{\sqrt{\frac{(s_1)^2}{n1}+\frac{(s_2)^2}{n2}}}$$

Si se quiere conocer el Error de Tipo II, se debe tomar:

$$d = \frac{|\delta - \delta'|}{\sqrt{(\sigma_1)^2+(\sigma_2)^2}}$$

En el caso de que n_1 y n_2 sean distintos:

$$n = \frac{(\sigma_1)^2+(\sigma_2)^2}{\frac{(\sigma_1)^2}{n_1}+\frac{(\sigma_2)^2}{n_2}}$$

En el caso de que las muestras sean pequeñas, se *desconozcan* las **varianzas poblacionales** y *ambas* **poblaciones** sean **normales** con $\sigma_1 = \sigma_2 = \sigma$
Se aplica la prueba t-bimuestral:

$$t = \frac{\left[\bar{x}_1-\bar{x}_2\right]-\delta}{\hat{\sigma}_{\bar{x}_1-\bar{x}_2}} \quad \text{para} \quad \nu = n_1+n_2-1$$

Donde:

$$\hat{\sigma}^2_{\bar{x}_1-\bar{x}_2} = \sigma^2\cdot\left(\frac{1}{n_1}+\frac{1}{n_2}\right)$$

Siendo σ^2:

$$\sigma^2 = \frac{\left[\sum_{i=1}^{n_1}\left(\overline{x1_i}-\overline{x1}\right)^2+\sum_{i=1}^{n_2}\left(\overline{x2_i}-\overline{x2}\right)^2\right]}{n_1+n_2-2} = \frac{(n_1-1)\cdot(s_1)^2+(n_2-1)\cdot(s_2)^2}{n_1+n_2-2}$$

Sustituyendo en **t** se obtiene:

$$t = \frac{\left[\left(\overline{x_1}-\overline{x_2}\right)-\delta\right]}{\sqrt{(n_1-1)\cdot(s_1)^2+(n_2-1)\cdot(s_2)^2}}\cdot\sqrt{\frac{n_1\cdot n_2\cdot(n_1+n_2-2)}{n_1+n_2}}$$

Con $\nu = n_1 - n_2 - 2$

Pero es importante que las muestras sean INDEPENDIENTES.

Si se quiere trabajar con muestras de "Antes y Después" se utiliza la *diferencia de datos apareados* (con su signo)

Hipótesis Nula:

- $\mu = \delta$

Hipótesis Alternativa:

- $\mu \neq \delta$
- $\mu > \delta$
- $\mu < \delta$

En la <u>Prueba t para Muestras Apareadas:</u>

$$t = \frac{\bar{x} - \mu}{\frac{s}{\sqrt{n}}}$$

Donde $\bar{x}$ y s son:

$$\bar{x} = \sum \frac{x_{1,i} - x_{2,i}}{n}$$

$$s = \sqrt{\left(\sum \frac{(x_{1,i} - x_{2,i})^2}{n}\right) - \frac{n \times \bar{x}^2}{n}}$$