PRUEBAS DE HIPÓTESIS PARA MEDIAS DE UNA MUESTRA CON EXCEL, WINSTATS Y GEOGEBRA

En vez de estimar el valor de un parámetro, a veces se debe decidir si una afirmación relativa a un parámetro es verdadera o falsa. Es decir, *probar una hipótesis* relativa a un parámetro. Se realiza una prueba de hipótesis cuando se desea probar una afirmación realizada acerca de un parámetro o parámetros de una población.

Una *hipótesis* es un enunciado acerca del valor de un parámetro (media, proporción, etc.).

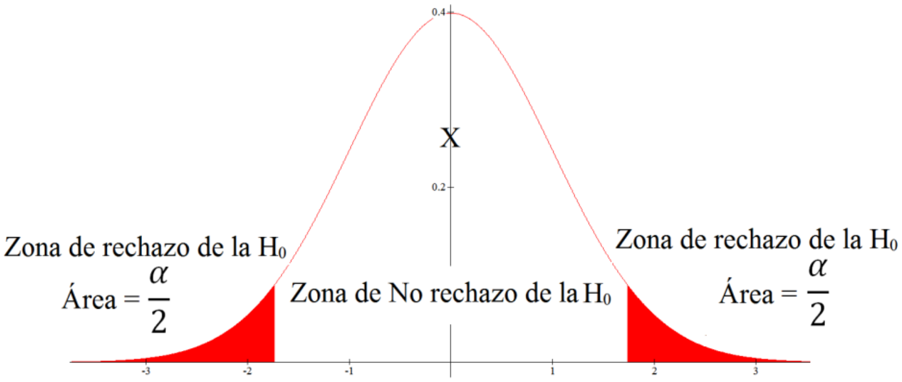
Prueba de Hipótesis es un procedimiento basado en evidencia muestral (estadístico) y en la teoría de probabilidad (distribución muestral del estadístico) para determinar si una hipótesis es razonable y no debe rechazarse, o si es irrazonable y debe ser rechazada.

La hipótesis de que el parámetro de la población es igual a un valor determinado se conoce como *hipótesis nula*. Una hipótesis nula es siempre una de status quo o de no diferencia. Se simboliza con el símbolo H_0 .Y cuando se desarrolla la prueba se asume que la hipótesis nula es verdadera y este supuesto será rechazado solo si se encuentran suficientes evidencias en base a la información muestral.

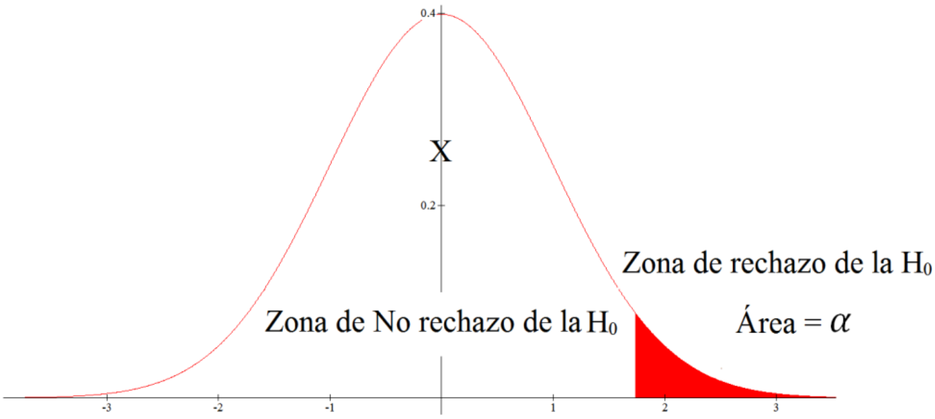
Siempre que se especifica una hipótesis nula, también se debe especificar una *hipótesis alternativa*, o una que debe ser verdadera si se encuentra que la hipótesis nula es falsa. La hipótesis alternativa se simboliza H_1. La hipótesis alternativa representa la conclusión a la que se llegaría si hubiera suficiente evidencia de la información de la muestra para decidir que es improbable que la hipótesis nula sea verdadera, y por tanto rechazarla. Es siempre opuesta a la Hipótesis Nula.

En toda prueba de hipótesis se presentan 3 casos de *zonas críticas* o llamadas también *zonas de rechazo de la hipótesis nula*, estos casos son los siguientes:

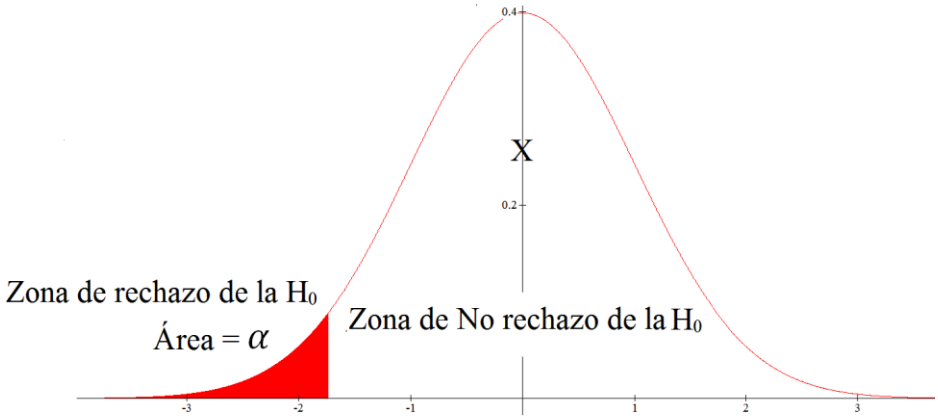
1) Prueba Bilateral o a dos colas: $H_0: \mu = X; H_1 \neq X$



2) Prueba Unilateral con cola hacia la derecha: $H_0: \mu \leq X; H_1 > X$



3) Prueba Unilateral con cola hacia la izquierda: $H_0: \mu \geq X; H_1 < X$



En toda prueba de hipótesis se pueden cometer 2 *tipos de errores*:

1) Error tipo I: se comete error tipo I, cuando se rechaza la H_0, siendo esta realmente verdadera. A la probabilidad de cometer error tipo I, se le conoce como nivel de significación y se le denota como α

2) Error tipo II: se comete error tipo II, cuando no se rechaza la H_0, siendo esta realmente falsa. A la probabilidad de cometer error tipo II, se le denota como β

El complemento de la probabilidad de cometer error tipo II, se le llama potencia de la prueba y se denota como $1 - \beta$

Como resumen se da la siguiente tabla:

	Se Acepta H_0	Se Rechaza H_0
H_0 es Verdadera	Decisión Correcta	Error de Tipo I
H_0 es Falsa	Error de Tipo II	Decisión Correcta

Se utiliza una prueba de una muestra para probar una afirmación con respecto a una media de una población única.

Si se conoce la desviación estándar de la población (σ), la distribución de muestreo adecuada es la distribución normal. Si la población que se muestra es normal, la distribución de muestreo será normal en el caso de todos los tamaños de la muestra, y el valor estadístico de prueba a utilizar es:

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

Si la población no es normal, o si se desconoce su forma, se emplea la ecuación anterior solamente para tamaños de muestra iguales o mayores 30, es decir, para $n \geq 30$

Si no se conoce la desviación estándar de la población (σ), el valor estadístico de prueba es:

$$t_{prueba} = \frac{\bar{x} - \mu}{\frac{S}{\sqrt{n}}}$$

Nota: Se considera práctico utilizar la distribución t solamente cuando se requiera que el tamaño de la muestra sea menor de 30, ya que para muestras más grandes los valores t y z son aproximadamente iguales, y es posible emplear la distribución normal en lugar de la distribución t.

Las anteriores ecuaciones se aplican para poblaciones infinitas, pero cuando la población es finita y el tamaño de la muestra n constituye más del 5% del tamaño de la población N, es decir:

$$\frac{n}{N} \cdot 100\% > 5\%$$

En este caso se debe usar el factor finito de corrección para modificar las desviaciones estándar, por lo tanto se aplican las siguientes ecuaciones para (σ) conocida y desconocida, respectivamente.

$$Z_{prueba} = \frac{\bar{x} - \mu}{\frac{\sigma}{\sqrt{n}} \cdot \sqrt{\frac{N-n}{N-1}}} \qquad t_{prueba} = \frac{\bar{x} - \mu}{\frac{S}{\sqrt{n}} \cdot \sqrt{\frac{N-n}{N-1}}}$$

Ejemplos ilustrativos:

1) La duración media de una muestra de 300 focos producidos por una compañía resulta ser de 1620 horas. Se conoce que desviación típica de la población es 150 horas. Comprobar la hipótesis $\mu = 1600$ contra la hipótesis alternativa $\mu \neq 1600$ horas con un nivel de significación de 0,05 si la muestra fue tomada de 5000 focos.

Solución:

Los datos son:
$n = 300$
$\bar{x} = 1620$
$\sigma = 150$
$\alpha = 0{,}05$
$N = 5000$

Las hipótesis son:
$H_0: \mu = 1600$
$H_1: \mu \neq 1600$

Al observar $H_1: \mu \neq 1600$ se trata de una prueba a dos colas, por lo que se tiene que calcular:
$$\frac{\alpha}{2} = \frac{0{,}05}{2} = 0{,}025$$

Como se conoce la desviación estándar de la población σ se debe utilizar la distribución normal. Con lectura en la tabla para un área de 0,025 le corresponde un valor $Z_{tabla} = \pm 1{,}96$. Se toma en cuenta el valor positivo y el negativo porque se trata de una prueba de hipótesis a dos colas.

Como se tiene como dato el tamaño de la población se tiene que verificar si cumple con la condición para utilizar el factor finito de corrección.

$$\frac{n}{N} \cdot 100\% > 5\%$$

$$\frac{300}{5000} \cdot 100\% = 6\%$$

Entonces para calcular el valor de Z_{prueba} se emplea la siguiente ecuación:

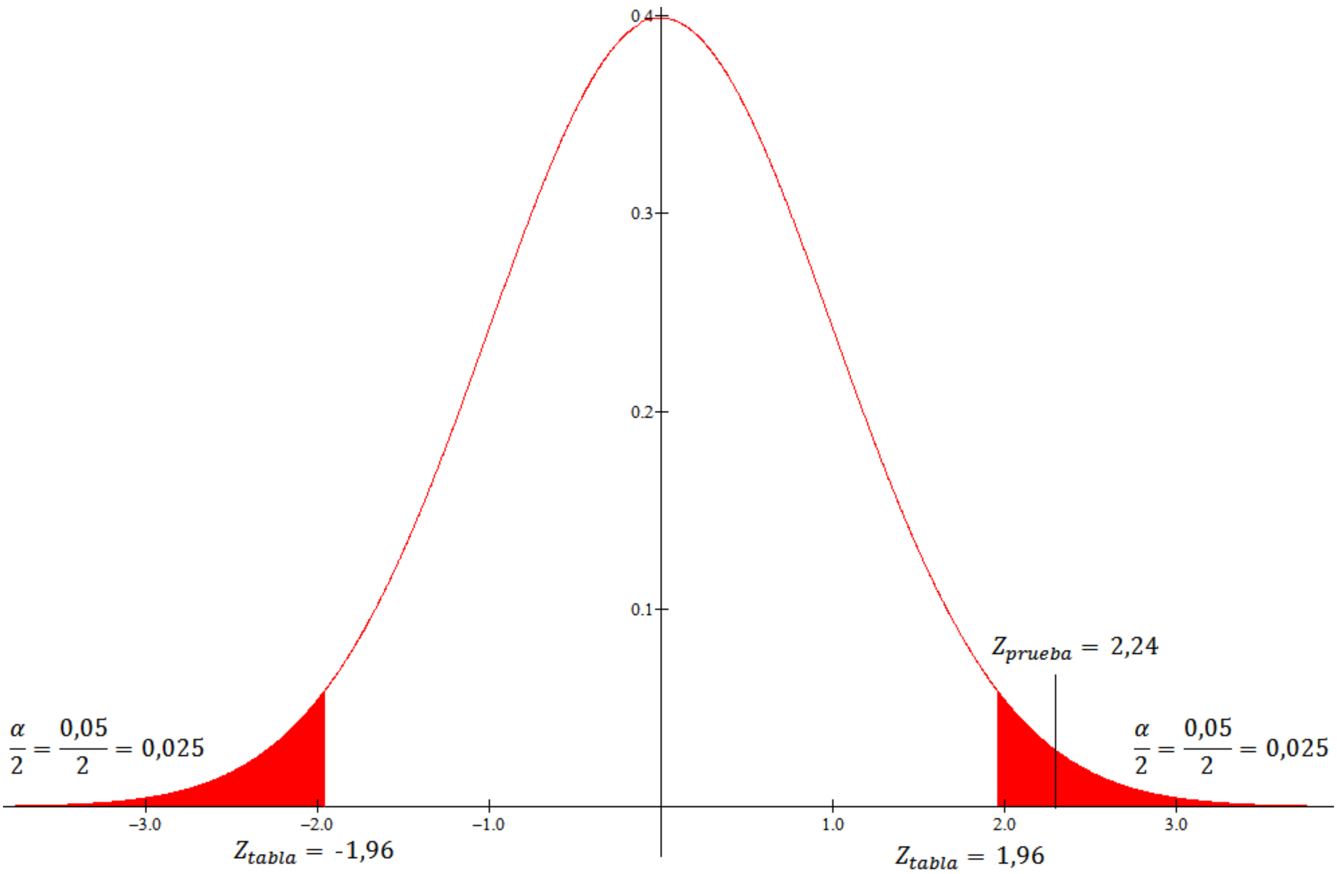
$$Z_{prueba} = \frac{\bar{x} - \mu}{\frac{\sigma}{\sqrt{n}} \cdot \sqrt{\frac{N-n}{N-1}}}$$

$$Z_{prueba} = \frac{1620 - 1600}{\frac{150}{\sqrt{300}} \cdot \sqrt{\frac{5000-300}{5000-1}}} = 2{,}24$$

Los cálculos en Excel se muestran en la siguiente imagen:

	A	B	C	D	E	F
1	N	5000				
2	n	300				
3	$\bar{x}$	1620				
4	σ	150				
5	μ	1600				
6	$H_0: \mu = 1600$					
7	$H_1: \mu \neq 1600$					
8	α	0,05				
9	$\frac{\alpha}{2}$	0,025	=B8/2			
10						
11	$\frac{n}{N} \cdot 100\% > 5\%$	6,00	=(B2/B1)*100			
12						
13	Z_{tabla}	-1,96	=INV.NORM.ESTAND(B9)			
14		1,96	=B13*-1			
15	$Z_{prueba} = \frac{\bar{x} - \mu}{\frac{\sigma}{\sqrt{n}} \cdot \sqrt{\frac{N-n}{N-1}}}$	2,24	=(B3-B5)/(B4/RCUAD(B2))*RCUAD((B1-B2)/(B1-1))			
16						
17						

El gráfico elaborado con Winstats y Paint se muestra en la siguiente imagen:



Decisión: Dado que $Z_{prueba} 2{,}24 > Z_{tabla} \pm 1{,}96$ se rechaza la H_0, y por lo tanto se acepta H_1

2) La duración media de lámparas producidas por una compañía han sido en el pasado de 1120 horas. Una muestra de 8 lámparas de la producción actual dio una duración media de 1070 horas con una desviación típica de 125 horas. Comprobar la hipótesis $\mu = 1120$ horas contra la hipótesis alternativa $\mu < 1200$ horas mediante un error tipo I de 0,05.

Solución:

Los datos son:
$\mu = 1120$
$n = 8$
$\bar{x} = 1070$
$S = 125$
$\alpha = 0{,}05$

Las hipótesis son:
$H_0: \mu = 1120$
$H_1: \mu < 1120$

Como se conoce la desviación estándar de la muestra S se debe utilizar la distribución t de Student. Con lectura en la tabla para un área de 0,05 y con $n - 1 = 8 - 7$ grados de libertad le corresponde un valor $t_{tabla} = -1{,}8946$. Se toma en cuenta el valor negativo porque se trata de una prueba de hipótesis a cola izquierda como se puede observar en la H_1.

Entonces para calcular el valor de t_{prueba} se emplea la siguiente ecuación:

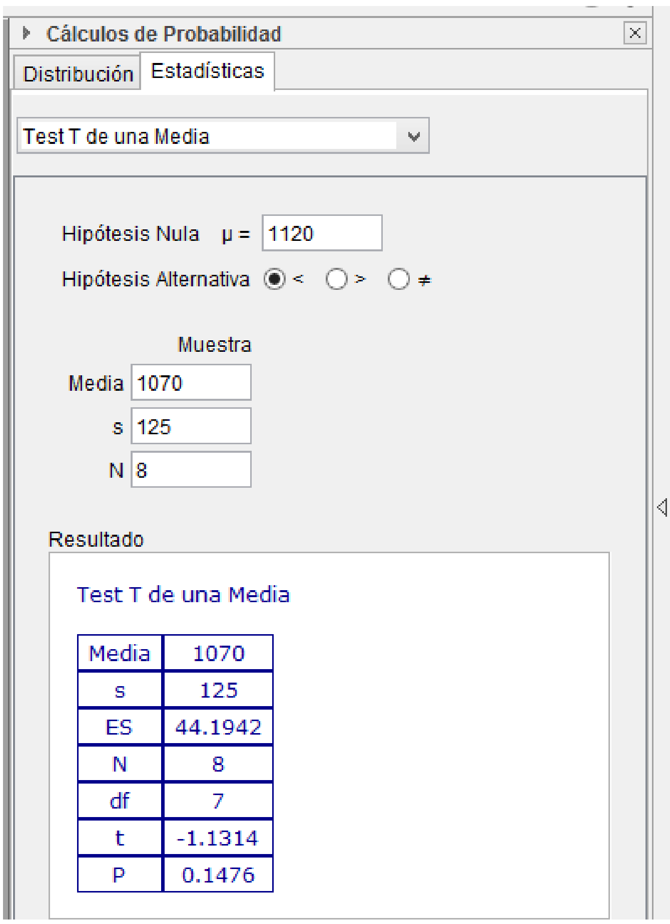
$$t_{prueba} = \frac{\bar{x} - \mu}{\frac{S}{\sqrt{n}}}$$

$$t_{prueba} = \frac{1070 - 1120}{\frac{125}{\sqrt{8}}} = -1{,}131$$

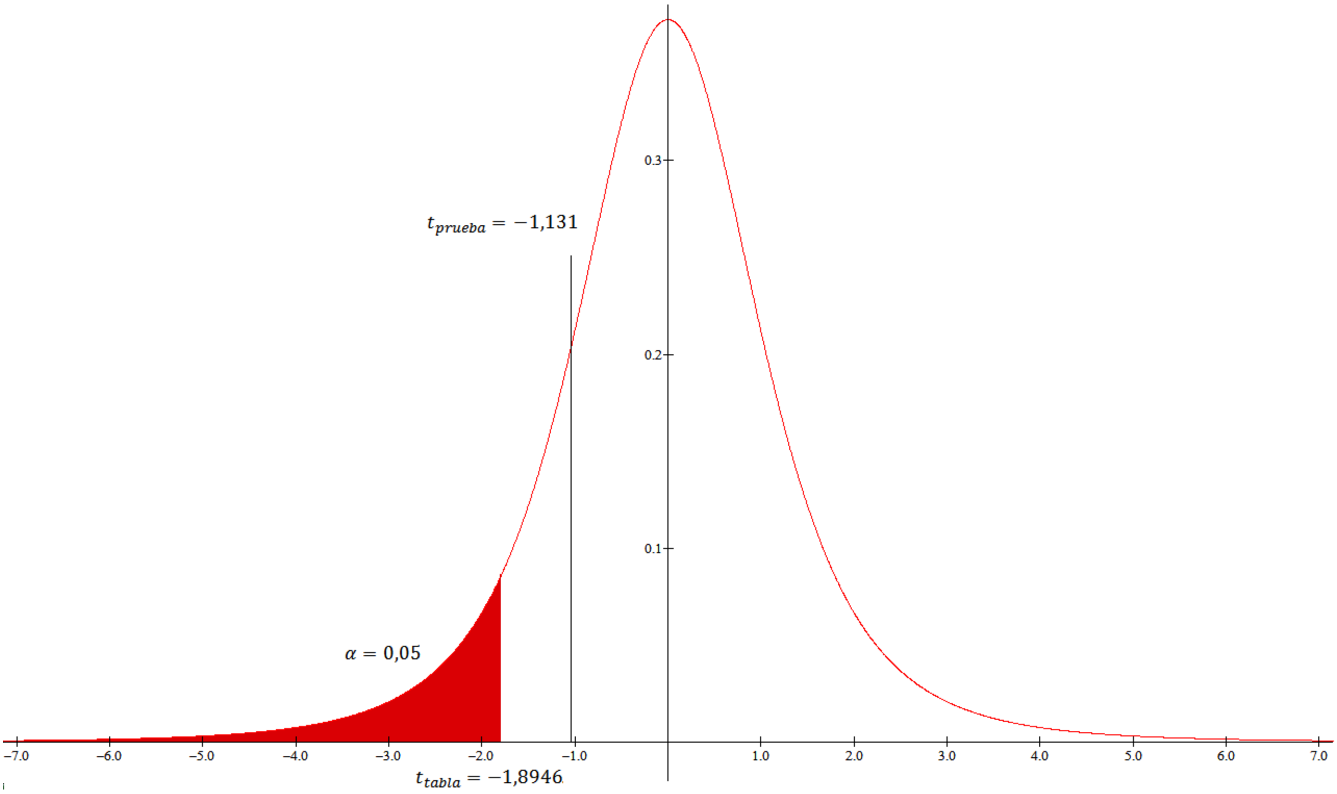
Los cálculos en Excel se muestran en la siguiente imagen:

	A	B	C	D
1	n	8		
2	$\bar{x}$	1070		
3	S	125		
4	μ	1120		
5	$H_0: \mu = 1120$			
6	$H_1: \mu < 1120$			
7	α	0,05		
8	n-1	7	=B1-1	
9	t_{tabla}	-1,8946	=INV.T(B7;B8)	
10	$t_{prueba} = \frac{\bar{x} - \mu}{\frac{S}{\sqrt{n}}}$	-1,131	=(B2-B4)/(B3/RCUAD(B1))	
11				
12				

Los cálculos en GeoGebra se muestran en la siguiente imagen:



El gráfico elaborado con Winstats y Paint se muestra en la siguiente imagen:



Decisión: Dado que $t_{prueba} - 1{,}131 > t_{tabla} - 1{,}8946$ se Acepta la H_0