3.4 Pruebas para medias y proporciones
Prueba de Hipótesis Referida al Parámetro Media Poblacional μ (Muestras Grandes)

Utilizaremos un ejemplo para introducir este tema.

De acuerdo a ciertos informes de un organismo encargado del estudio del medio ambiente, la lluvia ácida, causada por la reacción de ciertos contaminantes del aire con el agua de lluvia, disminuye la acidez del aire, afectando las tierras de cultivo y causando corrosión en los metales expuestos. La lluvia pura que cae a través del aire limpio registra un valor de pH (el pH es una medida de la acidez), de 5.7.

La sospecha de que la instalación masiva de fábricas en un parque industrial cercano aun radio urbano puede estar contaminando el ambiente, llevó a este organismo a iniciar una investigación sobre la contaminación ambiental.

Los investigadores formularon la hipótesis de que el ambiente estaba contaminado por el aumento de actividad fabril en la zona. En otras palabras, pensaban que el pH de la lluvia que cae en la zona era inferior a 5.7.

Se establecieron entonces, las siguientes hipótesis:

H_0) $\mu \geq 5.7$

H_1) $\mu < 5.7$

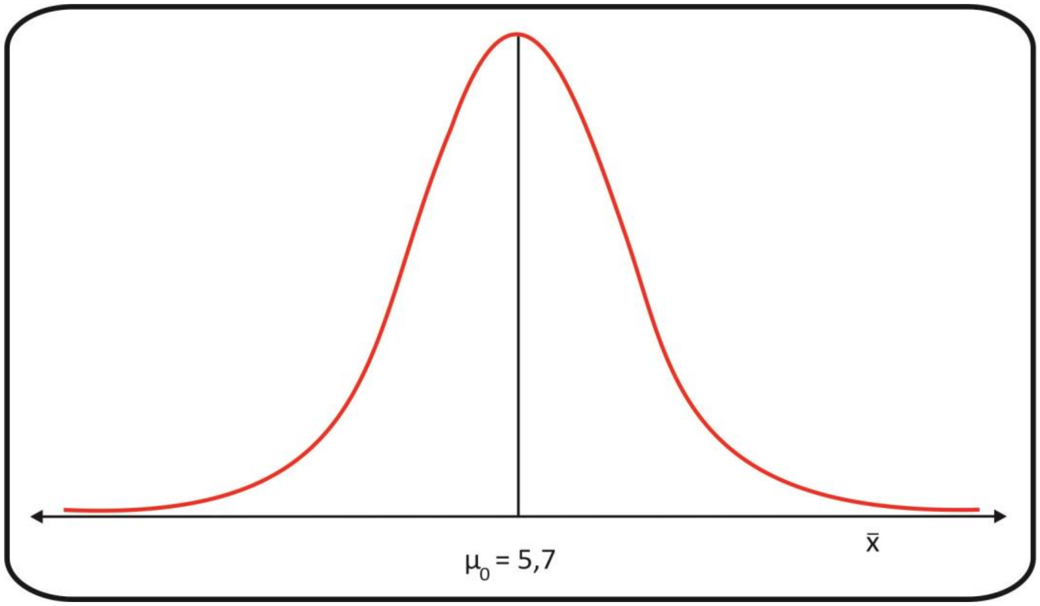
siendo μ el promedio de pH de la lluvia caída en la totalidad de precipitaciones que se dan en el lugar. (Hacemos la aclaración de que a menor pH, mayor acidez).

En el procedimiento de prueba de hipótesis, al igual que en el caso de estimación, es lógico pensar que la evidencia para rechazar o no la hipótesis nula será proporcionada por una muestra aleatoria extraída de la población de interés.

Como estamos tratando de probar una hipótesis referida al parámetro media poblacional, utilizaremos como estadístico para tomar una decisión su contrapartida en la muestra, la media muestral x .

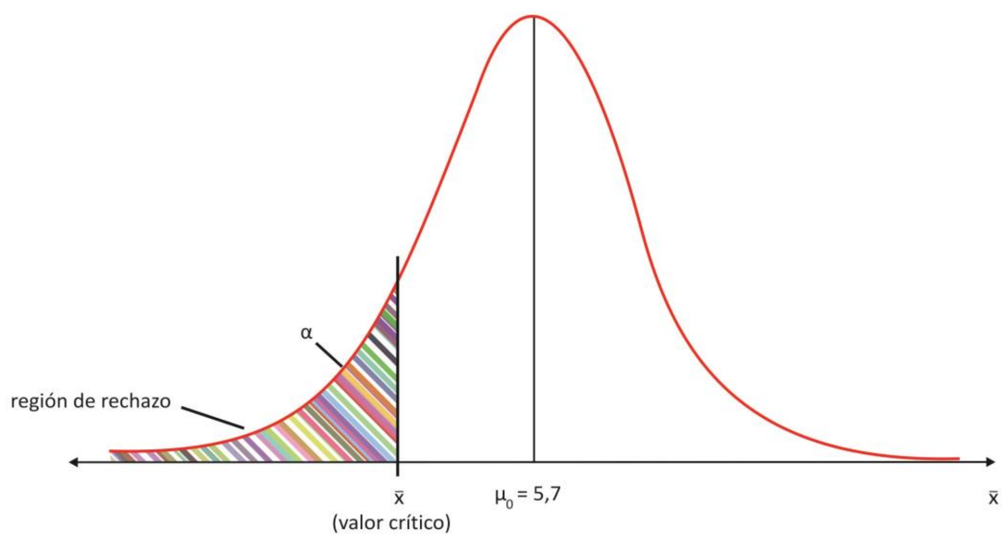
Ya hemos visto que la media muestral es un estimador insesgado, de mínima varianza y que tiende a distribuirse normalmente a medida que se incremento el tamaño de la muestra.

Gráficamente, tendremos:



En este caso en particular, rechazaremos la hipótesis nula cuando el estimador puntual x tome valores más pequeños que 5.7. Cuanto más alejado esté de este valor, con mayor razón vamos a rechazar H_0.

La región de rechazo de la hipótesis nula va a estar a la izquierda de la distribución tal como se observa en la siguiente figura:



Existirá un valor de x, que denominaremos valor crítico y simbolizaremos con x* que separará la región de rechazo de H_0 de la región de no rechazo.

La probabilidad de rechazar la hipótesis nula cuando es verdadera será el área sombreada a la izquierda de x* bajo la curva normal. Esta probabilidad es la probabilidad de cometer el error de Tipo I que hemos llamado α.

Si la zona de rechazo de una prueba de hipótesis se encuentra en uno de los extremos de la distribución, decimos que se trata de una prueba unilateral. En nuestro caso en particular, tenemos una prueba unilateral izquierda porque la región de rechazo de H_0 se encuentra a la izquierda de la distribución.

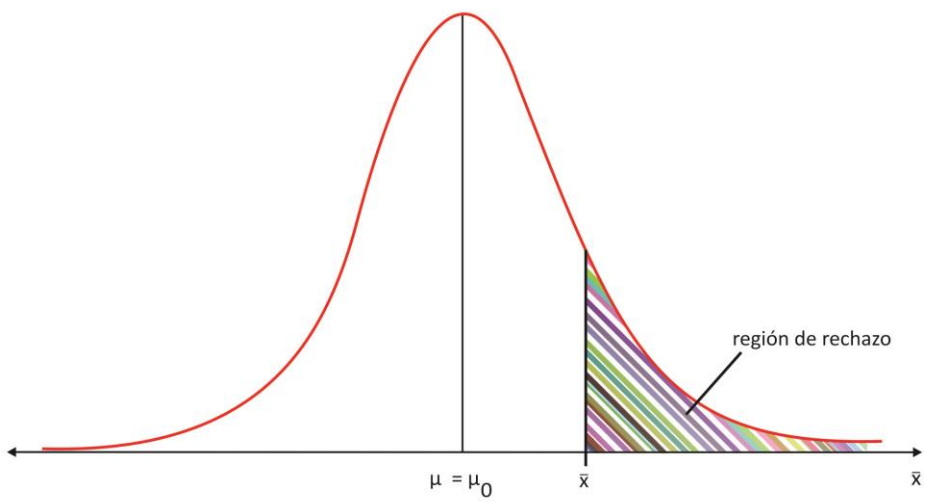
Si las hipótesis planteadas hubieran sido:

$H_0)\ \mu \leq \mu_0$

$H_1)\ \mu > \mu_0$

(siendo μ_0 el valor del parámetro en la hipótesis nula), la región de rechazo estaría en el extremo derecho de la distribución y tendríamos una prueba unilateral derecha.

Gráficamente, tendremos:



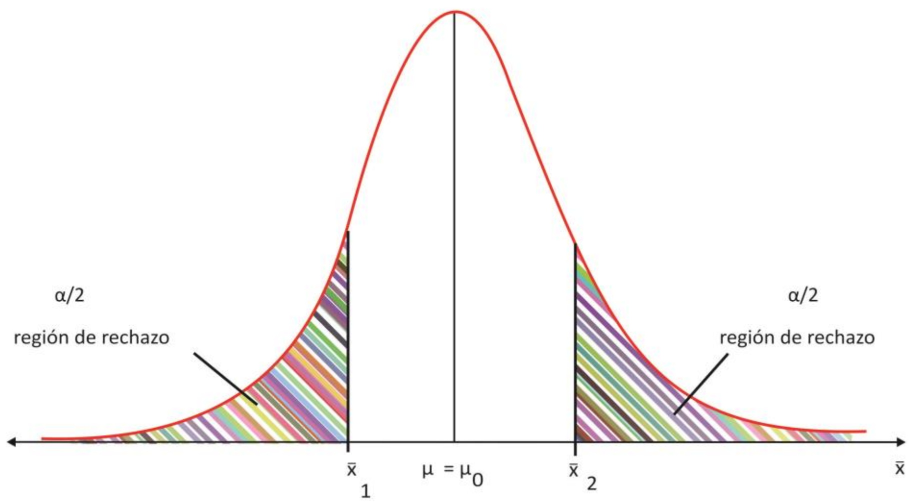
A su vez, si las hipótesis planteadas fueran:

$H_0)\ \mu = \mu_0$

$H_1)\ \mu \neq \mu_0$

tendríamos dos zonas de rechazo de la hipótesis nula en ambos extremos de la distribución, cada una con probabilidad $\alpha/2$ y estaríamos en presencia de una prueba bilateral.

Gráficamente, tendremos:



* Reglas de decisión

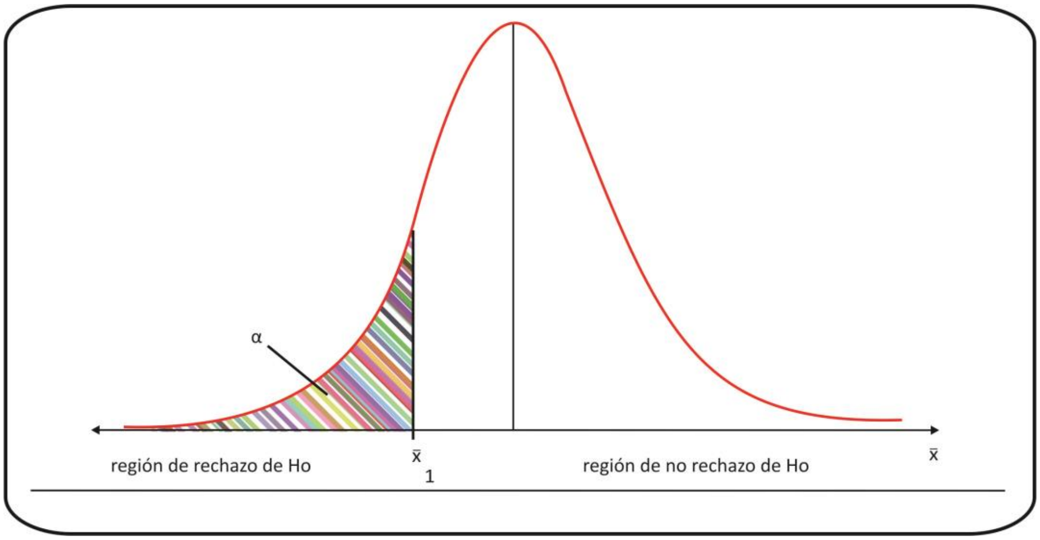
Veremos a continuación como calculamos el valor crítico que separará las zonas de rechazo y no rechazo de la hipótesis nula.

Como las zonas de rechazo y no rechazo de H_0 son áreas de la distribución normal, debemos estandarizar al valor crítico x*, que establece el límite entre las dos zonas, de acuerdo a la probabilidad de cometer el error de Tipo I que se ha fijado.

En el caso de una prueba unilateral izquierda:

H_0) $\mu \geq \mu_0$
H_1) $\mu < \mu_0$

Gráficamente:



Luego,

$$Z^* = \frac{(\bar{X}^* - \mu_0)}{\sigma/\sqrt{n}}$$

De esta expresión podemos despejar el valor del estimador x* que separará las zonas de rechazo y no rechazo de H_0:

$$\bar{X}^* = Z^* \bullet \frac{\sigma}{\sqrt{n}} + \mu_0$$

Conocido este valor crítico, podemos establecer las siguientes reglas de decisión:

Si $x \leq x^*$ se rechaza H_0
Si $x > x^*$ no se rechaza H_0

Cuando se elige una determinada prueba para la prueba de una hipótesis estadística, las reglas de decisión deben ser formuladas de manera tal que indiquen claramente cuando los datos experimentales son consistentes con la hipótesis nula y cuando no lo son.

En este caso en particular, donde se está hablando de probar hipótesis con respecto al parámetro media poblacional μ, las reglas de decisión deben señalar un rango de valores para x (media muestral) para los cuales la decisión sea rechazar H_0.

La hipótesis alternativa determina la localización de la región de rechazo de H_0 y el nivel de significación α el tamaño de dicha región.

Así, si se plantean las siguientes hipótesis nula y alternativa en una prueba unilateral derecha;

$H_0)$ $\mu \leq \mu_0$
$H_1)$ $\mu > \mu_0$

las reglas de decisión serán:

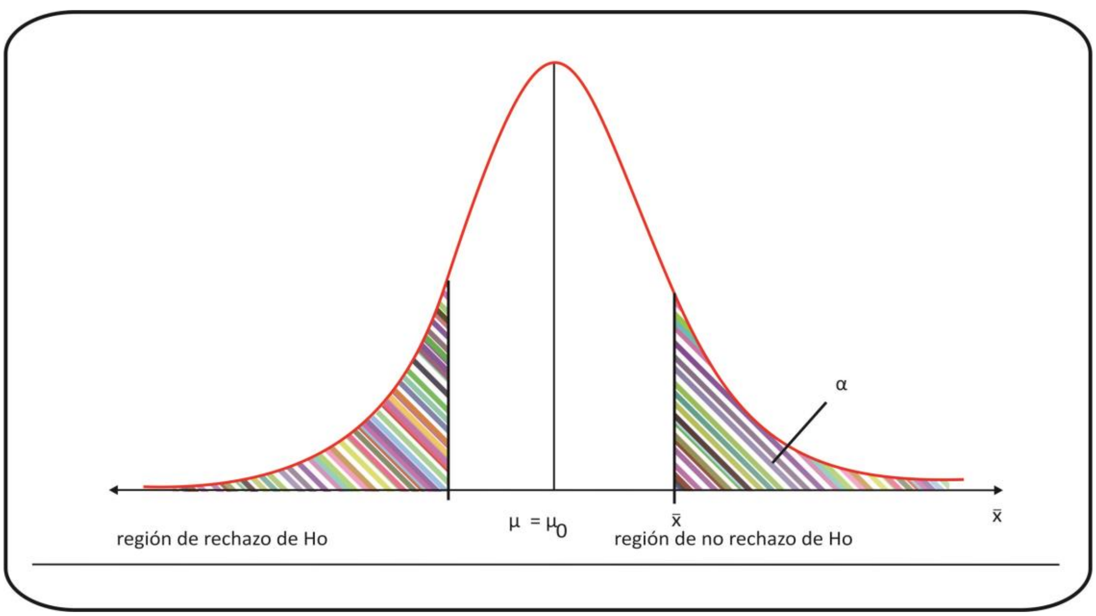
Rechazar H_0 cuando $x \geq x$ *
No rechazar H_0 cuando $x < x$ *

De acuerdo a la hipótesis alternativa y a la especificación de α, podemos representar gráficamente la región de rechazo de H_0 como:

En cambio, si la hipótesis alternativa se define como:
$H_1)$ $\mu \neq \mu 0$ (prueba bilateral)
la región de rechazo para H_0 incluye tanto al extremo izquierdo como el derecho de la distribución de x cuando se cumple la hipótesis nula.
Las reglas de decisión cuando se establece una alternativa bilateral serán:

$$\text{Si } x \leq x\ {*1} \text{ o } x \geq x\ {*2} \text{ se rechaza } H_0$$
$$\text{Si } x\ {*_1} < x < x\ {*_2} \text{ no se rechaza } H_0$$

Gráficamente, tendremos:



Para el caso de la prueba bilateral, el tamaño de cada cola de la distribución (zonas de rechazo de H_0), es igual a la mitad del nivel de significación α. Dividiendo a α en partes iguales se provee de la misma potencia a las hipótesis alternativas (izquierda y derecha) de H_0.

Siguiendo con el ejemplo de la contaminación y para poder calcular el valor crítico, el organismo a cargo de la investigación decide tomar una muestra de n = 40 precipitaciones. En cada una mide el *pH* y obtiene un *promedio* $x = 3.7$ con una desviación estándar s = 0.5.

Como no conocemos la desviación estándar poblacional se la debe estimar a partir de la desviación estándar muestral$_i$ s. Como el tamaño de la muestra es grande (n = 40), z seguirá distribuyéndose normal (0,1).

Si el investigador decide equivocarse un 5% de las veces al tomar una decisión con respecto a la hipótesis nula, $\alpha = 0.05$.

Para determinar el valor crítico que separa las zonas de aceptación y de rechazo de H_0, debemos encontrar el valor de z que acumula 0.05 de probabilidad. Si vamos a una tabla de probabilidades normales, encontraremos que ese valor de z es - 1.645.

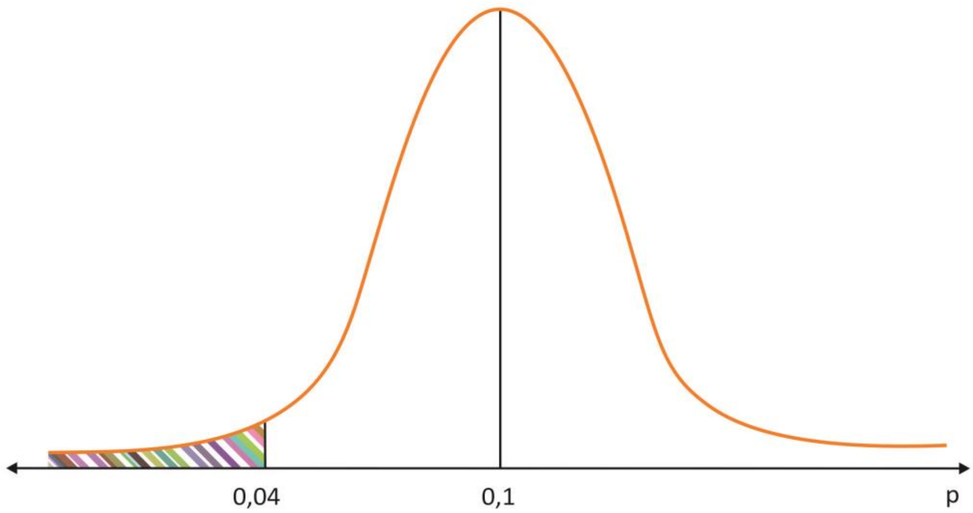
Entonces:

$$x^* = -1{,}645 \bullet \frac{0{,}5}{\sqrt{40}} + 5{,}7 = -0{,}13 + 5{,}7 = 5{,}57$$

Conocido este valor crítico, podemos establecer las siguientes reglas de decisión:

Si $x \leq 5.57$ se rechaza H_0
Si $x > 5.57$ no se rechaza H_0

Gráficamente, tenemos:



Como el valor de la *media muestral* $x = 3.7$ cae en la zona de rechazo de H_0, el organismo encargado de la investigación decidirá rechazar la hipótesis nula a favor de la hipótesis alternativa. Se concluye luego que el ambiente presenta un contaminante producido por la lluvia ácida.

Al llegar a la conclusión de que el pH promedio no es 5.7, el investigador deseará efectuar una estimación por intervalos para averiguar cuál es el verdadero pH en la población.

Si construimos tal intervalo, tendremos:

$$P(3{,}7 - 1{,}96\ \frac{0{,}5}{\sqrt{40}} \leq \mu \leq 3{,}7 + 1{,}96\ \frac{0{,}5}{\sqrt{40}}) = 0{,}95$$

$$P(3{,}55 \leq \mu \leq 3{,}85) = 0{,}95$$

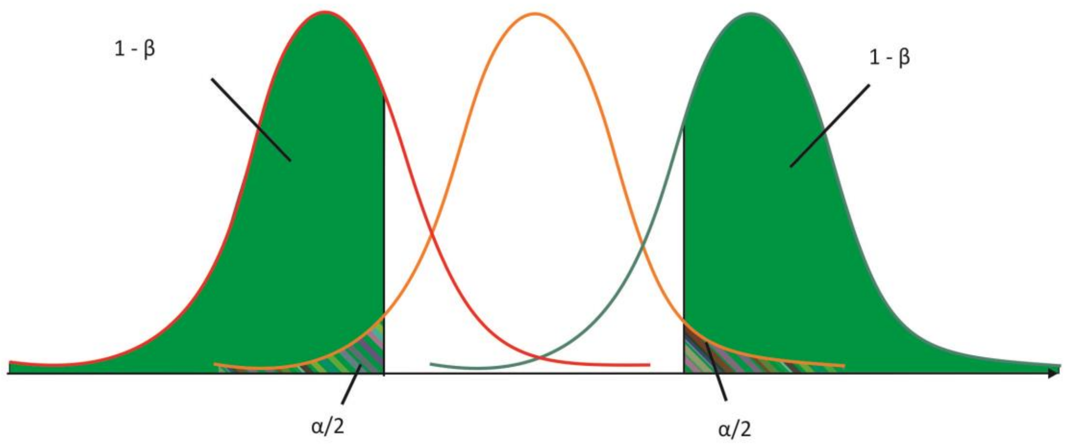
El valor del *pH* estimado es muy inferior al planteado en la hipótesis nula por lo cual puede concluirse que la contaminación por lluvia ácida es importante.

Potencia de la Prueba

Si el investigador está sólo interesado en rechazar H_0 cuando la hipótesis alternativa tiene una dirección especificada con respecto a H_0, es preferible trabajar con una región de rechazo localizada en una sola cola de la distribución, o sea, con una prueba unilateral. De esta manera, se aumenta la potencia de la prueba que no es más que la probabilidad de rechazar H_0 cuando es falsa, objetivo primordial de una prueba de hipótesis.

Esta situación se visualiza en el siguiente gráfico:

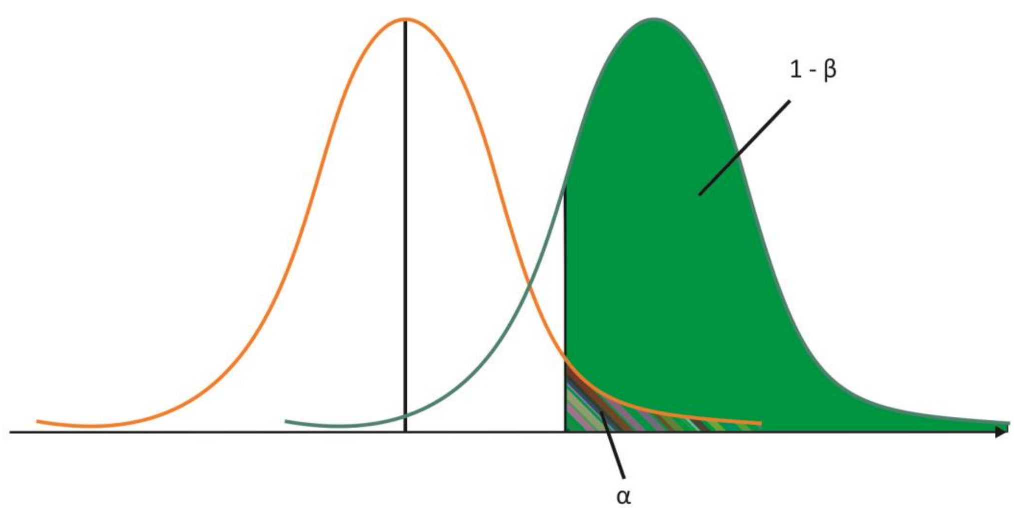
a)



La potencia para una prueba de dos colas se observa en el área sombreada de la Figura a). En cambio, la potencia de una prueba unilateral derecho, en el área sombreada de la Figura b).

Aunque la probabilidad de cometer error de Tipo I (α), es la misma en ambas pruebas, la potencia se incrementa en el caso de una prueba unilateral.

b)



Ahora bien, el investigador puede tener interés en calcular la probabilidad de cometer el error de Tipo II (β); para comprender mejor este cálculo presentamos la Figura c):

La primera de las distribuciones normales considera que la hipótesis nula es verdadera; por ello el estadístico x se distribuye alrededor de $\mu = \mu_0$. En cambio, la segunda distribución normal supone que la hipótesis alternativa es la verdadera y que x varía alrededor de μ_a, siendo μ_a un valor posible entre los infinitos valores de μ que establece la hipótesis alternativa.

Dijimos que el error de Tipo II consiste en no rechazar la hipótesis nula cuando es falsa, por lo cual estará representado por la parte sombreada correspondiente a la aceptación de H_0 pero bajo la curva normal dada por la hipótesis alternativa.

Para calcular el valor de β cuando la hipótesis alternativa establece que μ_a es un valor cualquiera, por ejemplo 5.5, debemos calcular el área bajo la curva normal con media $\mu = 5.5$ desde que la variable toma el valor crítico x * hasta ∞.

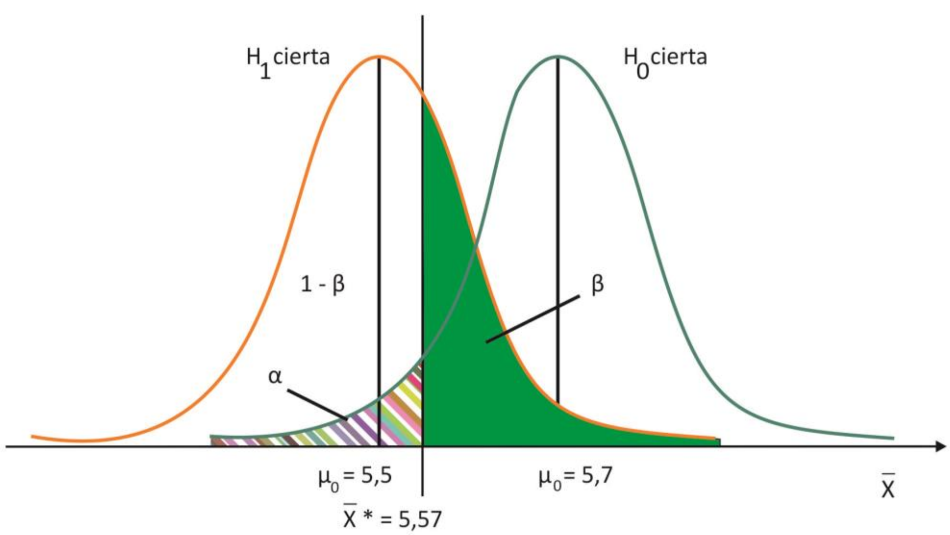
Debemos buscar la siguiente probabilidad:

$$P(5.57 < x < \infty / \mu_a = 5.5)$$

Esto se interpreta como la probabilidad de que la media muestral se encuentre entre 5.57 e ∞ dado que $\mu = 5.5$.

Gráficamente, tendremos:

c)



Esta probabilidad será:

$$1 - P(\bar{x} < 5{,}57 \,/ \mu_a = 5{,}5)$$

$$= 1 - P\left(\frac{\bar{x} - 5{,}5}{\frac{0{,}5}{\sqrt{40}}} < \frac{5{,}57 - 5{,}5}{\frac{0{,}5}{\sqrt{40}}}\right)$$

$$= 1 - P(z < 0{,}89)$$

$$= 1 - 0{,}8133 = 0{,}1867$$

$$\beta = 0{,}1867$$

Si quisiéramos calcular el valor de la potencia para esta alternativa en particular ($\mu = 5.5$), tendríamos:

Potencia $= 1 - \beta = 1 - 0.1867 = 0.8133$

Al calcular la potencia para varias alternativas y graficar convenientemente esta función, obtenemos lo que se denomina curva de potencia de una prueba.

Aclaremos este punto por medio de un ejemplo.

Supongamos que se desea saber si ha variado el valor de la canasta familiar mínima calculada en base a una familia tipo en la República Argentina en el rubro Alimentación luego de haber transcurrido un año desde la implementación del Plan de Convertibilidad.

A principios del año 1991, su cálculo era de \$ 400 con una desviación estándar de \$ 50.

Como existen opiniones divergentes sobre si ha aumentado o disminuido, se decide tomar una muestra al azar de 100 familias para probar las siguientes hipótesis:

$H_0)$ $\mu = 400$
$H_1)$ $\mu \neq 400$

La decisión a adoptar puede seguir cualquiera de estos caminos:

Decisión	$\mu = 400$	$\mu < 400$	$\mu > 400$
No cambió	Correcto	β	β
Cambió	α	Correcto	Correcto

La decisión de rechazar H_0 puede ser expresada de la siguiente manera:
Rechazar H_0 si:

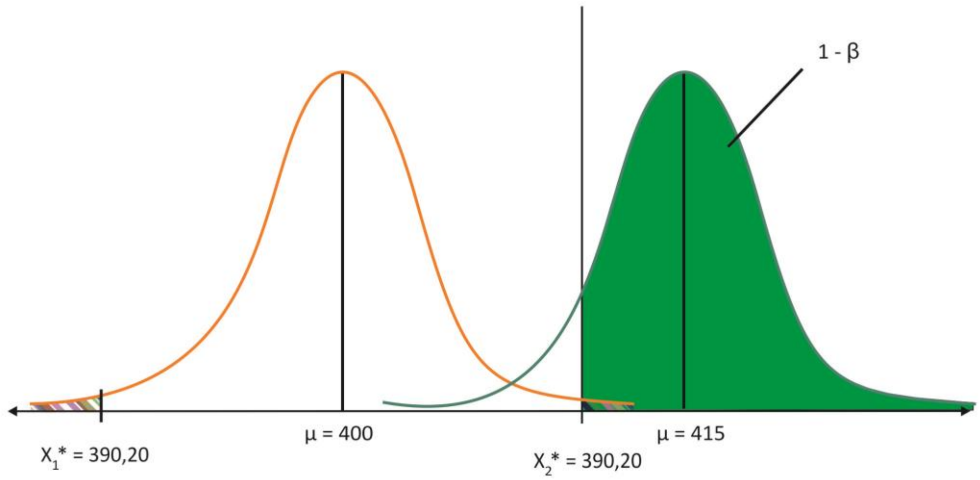
$$
\begin{aligned}
\bar{x} < \bar{x}_1 * &= \mu_0 - Z_{1-\alpha/2} \bullet \frac{\sigma}{\sqrt{n}} \\
&= 400 - 1{,}96 \bullet \frac{50}{\sqrt{100}} \\
&= 400 - 9{,}80 \\
&= 390{,}20 \ \$
\end{aligned}
$$

ó

$$
\begin{aligned}
\bar{x} < \bar{x}_2 * &= \mu_0 - Z_{1-\alpha/2} \bullet \frac{\sigma}{\sqrt{n}} \\
&= 400 - 1{,}96 \bullet \frac{50}{\sqrt{100}} \\
&= 400 - 9{,}80 \\
&= 409{,}80 \ \$
\end{aligned}
$$

En cambio, no se rechazará H_0 si $390.20 < x < 409.80$

Gráficamente, tendremos:



Dijimos que la potencia de una prueba es la probabilidad de rechazar H_0 cuando es falsa. En otras palabras, rechazar H_0 cuando en realidad la hipótesis alternativa es verdadera.

Si, en nuestro ejemplo, el verdadero valor del parámetro es $\mu = 415$, para calcular la potencia debemos determinar la probabilidad expresada por el área sombreada.

El área sombreada coincide con la zona de rechazo de Ho en la cola derecha de la distribución pero dibujada en la distribución normal cuya media es $\mu = 415$, uno de los posibles valores de¡ parámetro en la hipótesis alternativa.

Esta probabilidad puede escribirse como:

$$P(\ x > 409.8|\mu = 415)$$

Leemos esta probabilidad como: probabilidad de que la media muestral sea mayor a 409.8 cuando el valor de μ es 415.

$$\begin{aligned} P(\bar{x} > 409{,}8/\mu = 415) &= P\left(\frac{\bar{X} - 415}{5} > \frac{409{,}8 - 415}{5}\right) \\ &= P(z > -1{,}04) \\ &= 1 - F(-1{,}04) \\ &= 1 - [1 - F(1{,}04)] \\ &= F(1{,}04) \\ &= 0{,}8508 \end{aligned}$$

La potencia de la prueba cuando $\mu = 415$ es 0.8508.

Efectuaremos el mismo cálculo cuando $\mu = 410$.

La potencia de la prueba cuando $\mu = 410$ es igual a 0.5160.

Ahora bien, ¿Qué pasa si nos movemos en la zona de rechazo de H_0 expresada en la cola izquierda de la distribución?

Supongamos, por ejemplo, que el verdadero valor del parámetro es $\mu = 380$.

Para calcular la potencia en este punto, calcularemos la siguiente probabilidad:

$$
\begin{aligned}
P(\bar{x} < 390{,}2/\mu = 380) &= P\left(\frac{\bar{X} - 380}{5} < \frac{309{,}2 - 380}{5}\right) \\
&= P(z < 2{,}04) \\
&= F(2{,}04) \\
&= 0{,}9793
\end{aligned}
$$

Cuando $\mu = 380$, la potencia es igual a 0.9793.

Calculemos, por último, la potencia cuando $\mu = 385$.

La potencia será:

$$
\begin{aligned}
P(\bar{x} < 390{,}2/\mu = 385) &= P\left(\frac{\bar{X} - 385}{5} < \frac{309{,}2 - 385}{5}\right) \\
&= P(z < 1{,}04) \\
&= F(1{,}04) \\
&= 0{,}8508
\end{aligned}
$$

La potencia de la prueba cuando $\mu = 385$ es igual a 0.8508.

Vamos a calcular ahora la potencia cuando $\mu = 400$, exactamente el valor que toma el parámetro en la hipótesis nula.

En este caso, tendremos:

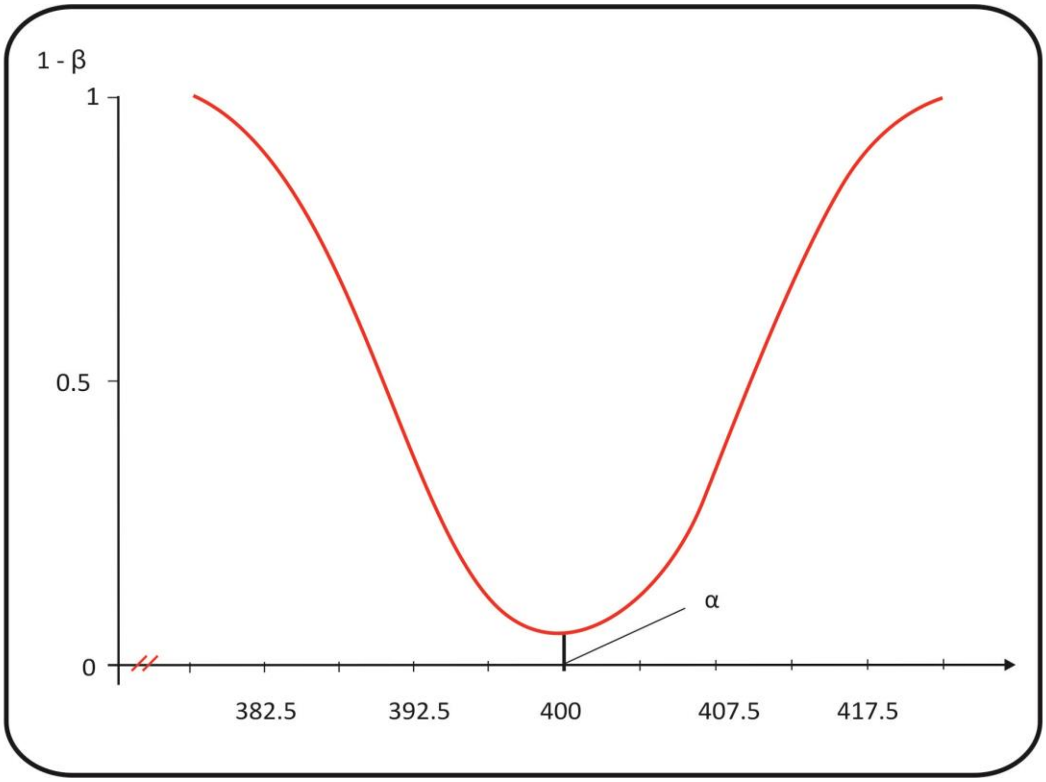
$$
\begin{aligned}
P(\bar{x} < 390{,}2/\mu = 400) = P(\,\bar{X} > 409{,}8/\mu = 400\,) \\
&= P\left(\frac{\bar{X} - 400}{5} < \frac{309{,}2 - 400}{5}\right) \\
&= P\left(\frac{\bar{X} - 400}{5} > \frac{409{,}8 - 400}{5}\right) \\
&= P(z < -1{,}96) + P(z > 1{,}96) \\
&= F(-1{,}96) + [1 - F(1{,}96)] \\
&= [1 - F(1{,}96)] + [1 - F(1{,}96)] \\
&= 2\,[1 - F(1{,}96)] \\
&= 2 \times (1 - 0{,}975) \\
&= 2 \times 0{,}025 \\
&= 0{,}05
\end{aligned}
$$

La potencia, calculada en el valor del parámetro dado por la hipótesis nula es exactamente igual a $\alpha(1 - \beta = \alpha)$.

Si hubiéramos calculado varios valores de la potencia para diferentes valores de μ, podríamos construir la siguiente tabla:

Hipótesis	P	Potencia
H_1	375,0	1,0000
H_1	382,5	0,9380
H_1	387,5	0,7050
H_1	392,5	0,3230
H_1	397,5	0,0790
H_0	400,0	0,0500
H_1	402,5	0,0790
H_1	407,5	0,3230
H_1	412,5	0,7050
H_1	417,5	0,9380
H_1	425,0	1,0000

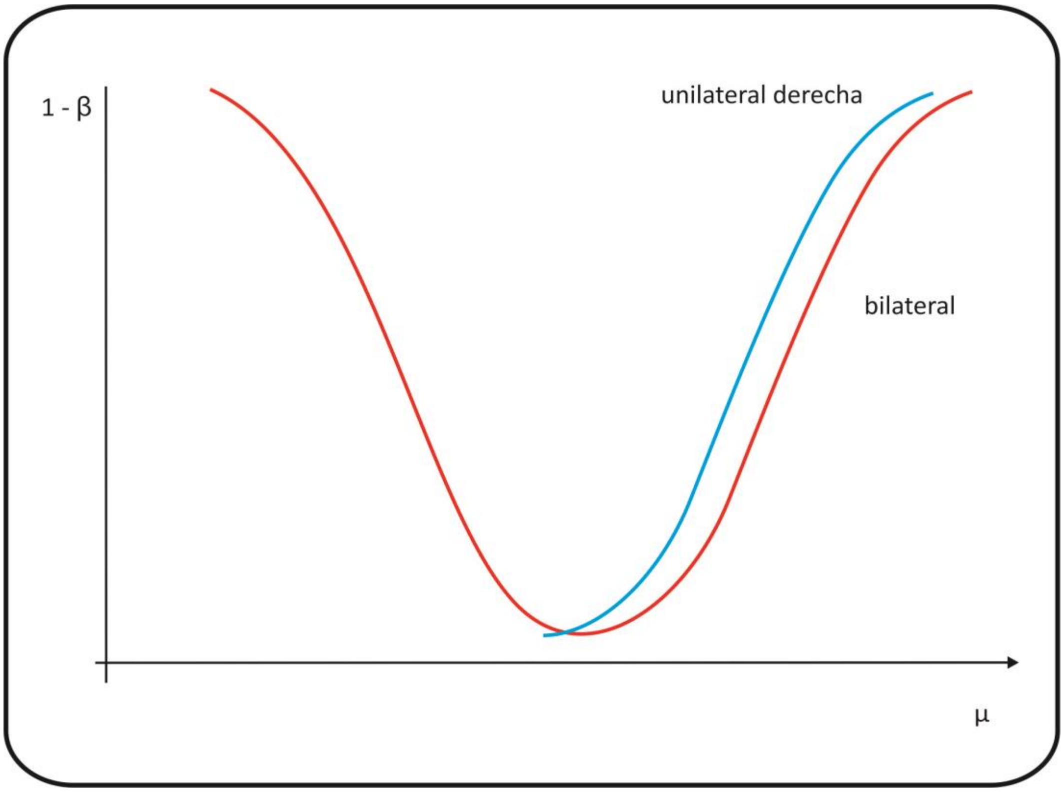
Graficando esta función de potencia, obtenemos la curva de potencia correspondiente a la prueba planteado en el ejemplo.



Si en el ejemplo de la canasta familiar, la opinión más fuerte hubiera sido que ella habría aumentado, hubiéramos planteado las siguientes hipótesis:

H_0) $\mu \leq 400$
H_1) $\mu > 400$

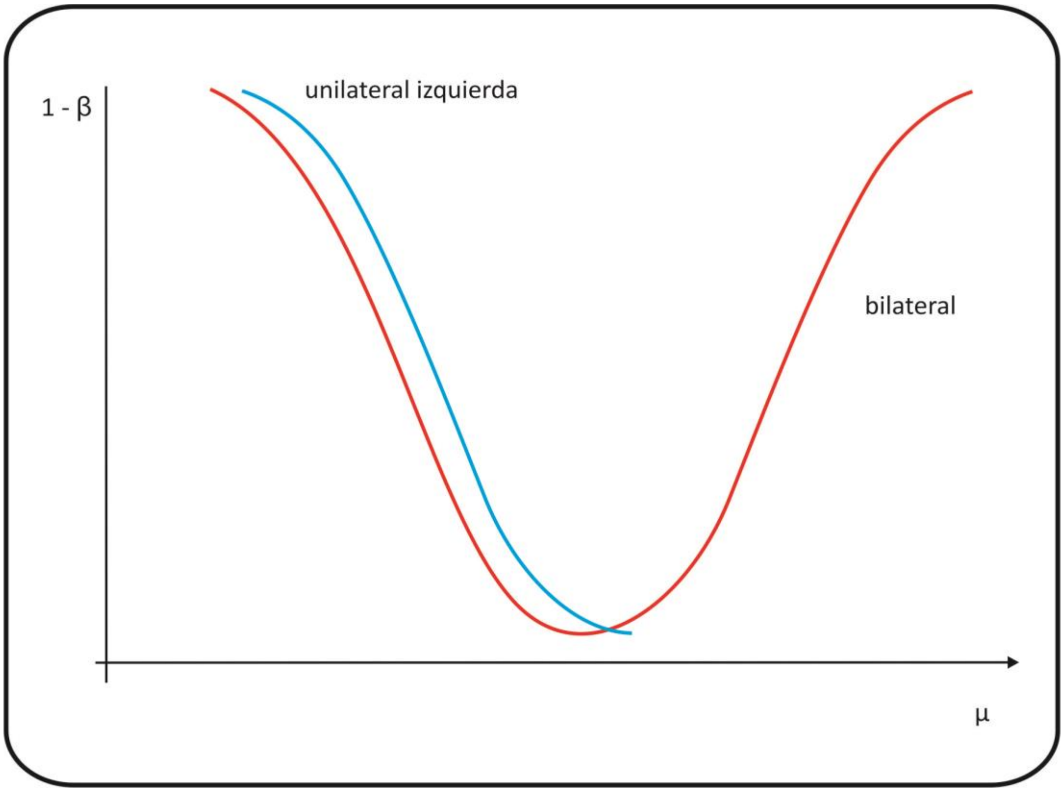
Al emplear el procedimiento anterior para calcular la potencia de la prueba y comparar esta curva con la de potencia correspondiente a la de la prueba bilateral, tendremos:



En cambio, si en el equipo encargado de la investigación, hubiera prevalecido la opinión de que la canasta promedio ha disminuido, las hipótesis postuladas, serían:

H_0) $\mu \geq 400$
H_1) $\mu < 400$

Cuando calculamos la correspondiente curva de potencia en el caso en que consideramos una prueba unilateral izquierda y la comparamos con la de la prueba bilateral, tenemos:



Siempre las pruebas unilaterales son uniformemente más potentes que los bilaterales en el caso de considerar que el tamaño de la muestra y el nivel de significación α se mantienen constantes. Por ello, cuando se tiene algún indicio de la dirección que toma el parámetro en la hipótesis alternativa es preferible trabajar con pruebas unilaterales.

Más adelante veremos la importancia de la determinación de la potencia para calcular tamaños de muestras y número de repeticiones de los tratamientos que intervienen en un diseño experimental.

Resumiendo, podemos establecer las siguientes indicaciones a tener en cuenta en la elección de una buena prueba de hipótesis:

a) El error β y en consecuencia, la potencia de una prueba, queda automáticamente establecido al fijarse el tamaño de la muestra y el nivel de significación α.

b) Para un tamaño de muestra dado, la reducción de la probabilidad de cometer un error de Tipo I produce un aumento del error de Tipo II.

c) Si se desean fijar α y β dándoles valores preestablecidos, debemos determinar el tamaño de muestra adecuado que satisfaga estas especificaciones. (El cálculo del tamaño adecuado de muestra para lograr una determinada potencia, se verá en el Módulo siguiente).

Intentaremos ahora aplicar los conceptos tratados.

Prueba de Hipótesis Referida al Parámetro Media Poblacional μ (Muestras Chicas n ≤ 3 0)

Desarrollaremos el tema por medio de un ejemplo.

El Departamento de Control de Calidad de una fábrica de televisores asegura que, de acuerdo al plan implementado, los televisores no deben fallar hasta los tres años de uso.

El Departamento de Comercialización, con el objetivo de corroborar esta afirmación, lleva a cabo una encuesta dirigida a compradores del producto con más de 3 años de antigüedad. Una de las preguntas estaba referida a cuándo requirió por primera vez asistencia técnica por falla del aparato.

Las respuestas obtenidas (tiempo desde que compró el aparato hasta que registró la primera asistencia técnica en años), en una muestra de 20 compradores, fueron las siguientes:

2,5	3,1	2,8	3,5	4,8	3,9	2,6	4,0	3,8	3,2
3,8	2,3	3,1	4,1	3,8	3,0	3,4	4,1	2,6	3,0

En base a esta evidencia proporcionada por la muestra, el Departamento de Comercialización desea saber si es lícito seguir publicitando que los aparatos no registran fallas hasta después de 3 años de ser utilizados.

Las hipótesis planteadas fueron:

H_0) $\mu \leq 3$ años

H_1) $\mu > 3$ años

El estadístico adecuado para efectuar la correspondiente prueba de hipótesis fue la media de la muestra, que en este caso es:

$\bar{x} = 3.37$ años

La varianza del tiempo transcurrido hasta la 1° asistencia técnica fue $s^2 = 0.4327$ y la correspondiente desviación estándar muestral de 0.66 años.

Como ocurrirá seguramente en la práctica, es imposible calcular la varianza del tiempo transcurrido entre la compra del televisor y su primer pedido de asistencia en la población total de televisores vendidos por la empresa. Por este motivo, la única medida de variabilidad es la correspondiente a la muestra de 20 televisores.

Siguiendo el mismo procedimiento que en el caso de muestras grandes, debemos calcular también un valor crítico que determine el límite de las regiones de no rechazo y de rechazo de la hipótesis nula planteada.

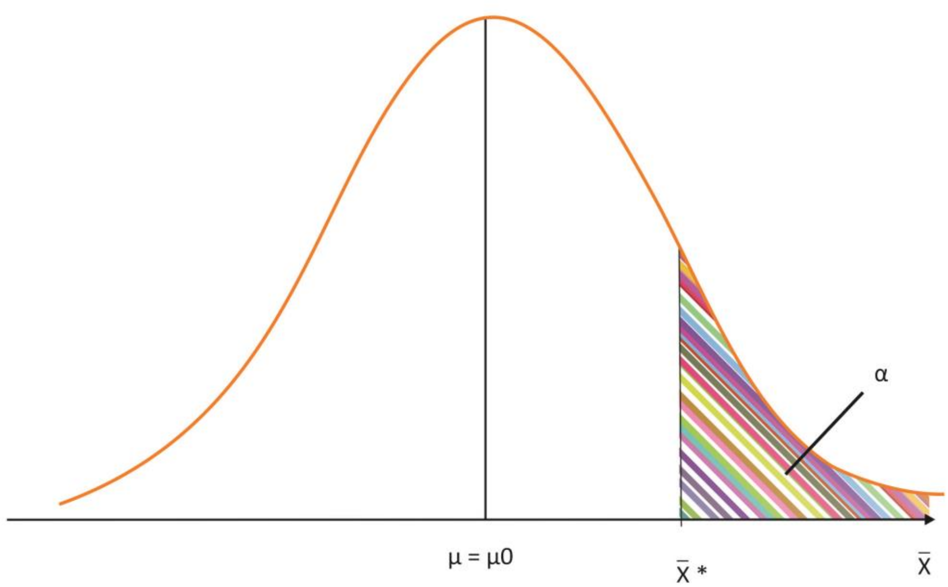
El valor estandarizado que podemos calcular es el siguiente:

$$t = \frac{(\bar{X} - \mu_0)}{S/\sqrt{n}}$$

Este valor difiere del anterior en que, en lugar de aparecer la desviación estándar de la población σ, nos encontramos con su estimador muestral insesgado s.

En nuestro ejemplo, la hipótesis alternativa establecida determina una prueba unilateral derecho por lo cual la zona de rechazo de la hipótesis nula estará en la cola derecha de la distribución t de Student.

Gráficamente, tendremos:



El paso siguiente consiste en determinar un nivel de significación α que podría ser α = 0.05.

Estableciendo el criterio de prueba como:

$$t = \frac{(\bar{X}^* - \mu_0)}{S/\sqrt{n}}$$

y considerando que $t \sim t_{n-1}$, podemos despejar el valor x * que es:

$$\bar{X}^* = \mu_0 + t \bullet \frac{S}{\sqrt{n}}$$

La regla de decisión será, entonces:

Si el promedio de tiempo transcurrido desde la compra del televisor hasta que se produce su primer asistencia técnica, calculado en la muestra de 20 televisores es mayor a x *, se rechazará la hipótesis nula; en caso contrario se concluirá que no existe una evidencia muestral contundente para rechazarla.

A los fines de determinar el valor crítico en nuestro ejemplo, sabemos que

$$t \sim t_{19}$$

o sea, t se distribuye como una variable t de Student con 19 grados de libertad.

En la tabla correspondiente, el valor de t que acumula el 0.95 de probabilidad es:

$$t_{19;\,0{,}95} = 1.729$$

Calculamos, entonces, el valor crítico que permitirá tomar una decisión con respecto a la hipótesis nula planteada.

$$\bar{X}^* = 3 + 1{,}729 \bullet \frac{0{,}66}{\sqrt{20}} = 3 + 0{,}26 = 3{,}26$$

La media en la muestra de 20 televisores fue x = 3.37.

Como es superior a 3.26, rechazamos la hipótesis nula. El valor x cae en la zona de rechazo de H_0 por lo cual llegamos a la conclusión de que el tiempo promedio hasta que se solicita asistencia técnica por primera vez, es superior a 3 años. Por ello, el Departamento de Comercialización podrá seguir publicitando la duración de los productos de la fábrica.

Cuando se rechaza la hipótesis nula, se puede estar interesado en estimar el verdadero valor del parámetro poblacional, en este caso μ = tiempo promedio hasta que se solicita asistencia técnica por primera vez.

Para ello se efectúa una estimación por intervalos que, en este caso, será:

$$P(3{,}37 - 2{,}093 \bullet \frac{0{,}66}{4{,}47} \leq \mu \leq 3{,}37 + 2{,}093 \bullet \frac{0{,}66}{4{,}47}) = 0{,}95$$

El valor de t, será ahora el que acumule el 0.975 de probabilidad.

Efectuando convenientemente los cálculos, tenemos:

$P(3.37 - 0.31 \leq \mu \leq 3{,}37 + 0{,}31) = 0.95$
$P(3.06 \leq \mu \leq 3.68) = 0.95$

De acuerdo a la estimación proporcionada por esta muestra, el tiempo promedio hasta que se requiere asistencia técnica es un valor comprendido en este intervalo y lo afirmamos con un 95% de confianza.

Una situación muy común en cualquier investigación radica en la necesidad de probar hipótesis referidas al parámetro proporción poblacional.

Por este motivo, comenzaremos a discutir los pasos a seguir en esta situación en particular.

Prueba de Hipótesis Referida al Parámetro Proporcional Poblacional p (Muestras Grandes)

Supongamos, por ejemplo, que se conoce por investigaciones ya realizadas, que el 20% de la población mayor de 15 años fuma.

Una Fundación que tiene como objetivo la prevención del cáncer, decide llevar adelante una campaña de lucha contra el tabaco por considerar que éste constituye uno de los principales factores de riesgo para desarrollar cáncer de pulmón.

Después de efectuar una fuerte campaña radial y televisiva durante 6 meses, se decide estudiar si la población adulta de la región ha disminuido el hábito de fumar.

Para ello selecciona una muestra aleatoria de 1000 personas adultas a las que somete a una determinada encuesta. Una de las preguntas del cuestionario utilizado estaba referida a si la persona fuma o no.

Una vez resumida la información proporcionada por el trabajo de campo, se observó que el 12% de las personas encuestadas fumaba habitualmente.

La Fundación decide poner a prueba la hipótesis estadística de que la campaña publicitaria había disminuido la cantidad de fumadores.

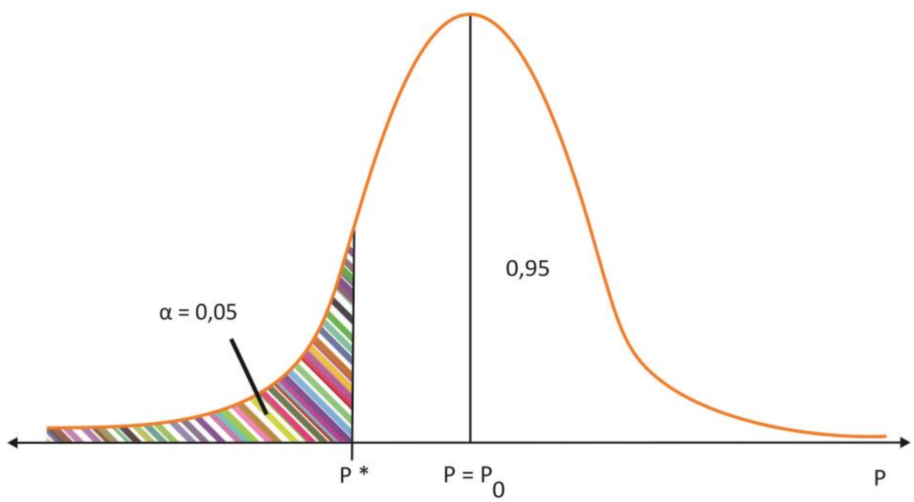
Las hipótesis postuladas fueron:
H_0) $P \geq 0.20$
H_1) $P < 0.20$

Ya hemos dicho, al hablar de la estimación del parámetro poblacional p, que cuando n (el tamaño de la muestra) es grande, la variable aleatoria proporción muestral p se distribuye normalmente con esperanza igual a p y desviación estándar igual pq/ n .

Debido a ello, podemos utilizar p (proporción muestral) como criterio de prueba para probar hipótesis con respecto al parámetro proporción poblacional.

Estandarizando, tendremos:

Gráficamente, podemos establecer la correspondiente región de rechazo de H_0 en la cola izquierda de la distribución normal, de la siguiente manera:



Las reglas de decisión adoptadas serán:
Rechazar H_0 si

$$p < p^* = P_0 + z_\alpha \bullet \sigma_p$$

Si el investigador fija un $\alpha = 0.05$, $z\alpha$ será igual a -1.645.
p_0 no es más que el valor que toma el parámetro en la hipótesis nula.

Ahora bien, y vemos que tanto p como k son valores poblacionales que desconocemos. Para solucionar este problema, conviene reemplazar al valor de p por el valor po, o sea, el valor que toma el parámetro bajo la hipótesis nula.

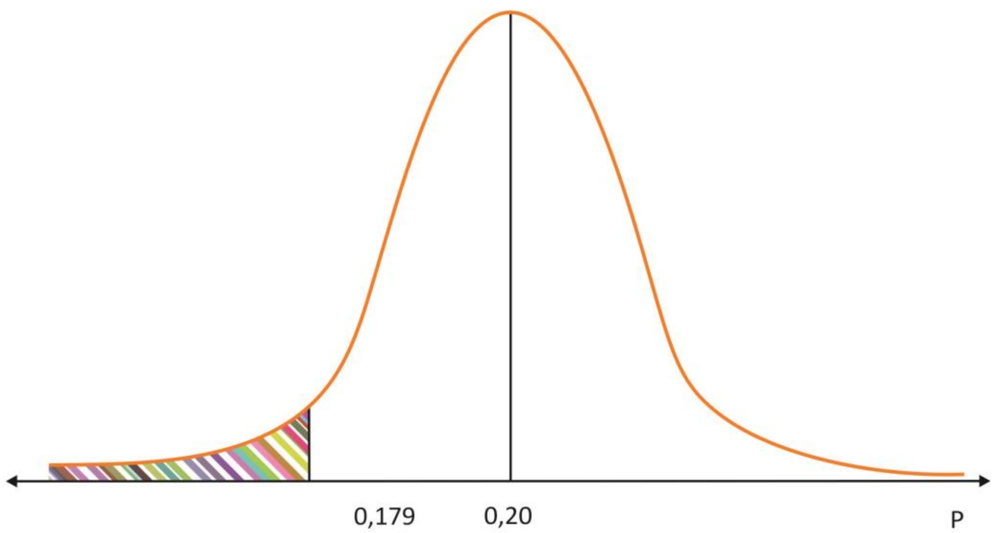
En nuestro ejemplo:

$$\sigma_p = \frac{\sqrt{P_0 Q_0}}{\sqrt{n}} = \frac{\sqrt{0{,}20 \bullet 0{,}80}}{\sqrt{1000}} = 0{,}0126$$

Entonces, rechazaremos H_0 cuando

$$p < 0.20 - 1.645 \,.\, 0.0126 = 0.179$$

Gráficamente, tenemos:



El valor de $p = 0.12$, cae en la región de rechazo de H_0 por lo cual la Fundación podrá concluir que la cantidad de fumadores después de la campaña ha disminuido significativamente.

Nos queda ahora por considerar la situación experimental en la cual queremos establecer hipótesis con respecto al parámetro varianza poblacional.