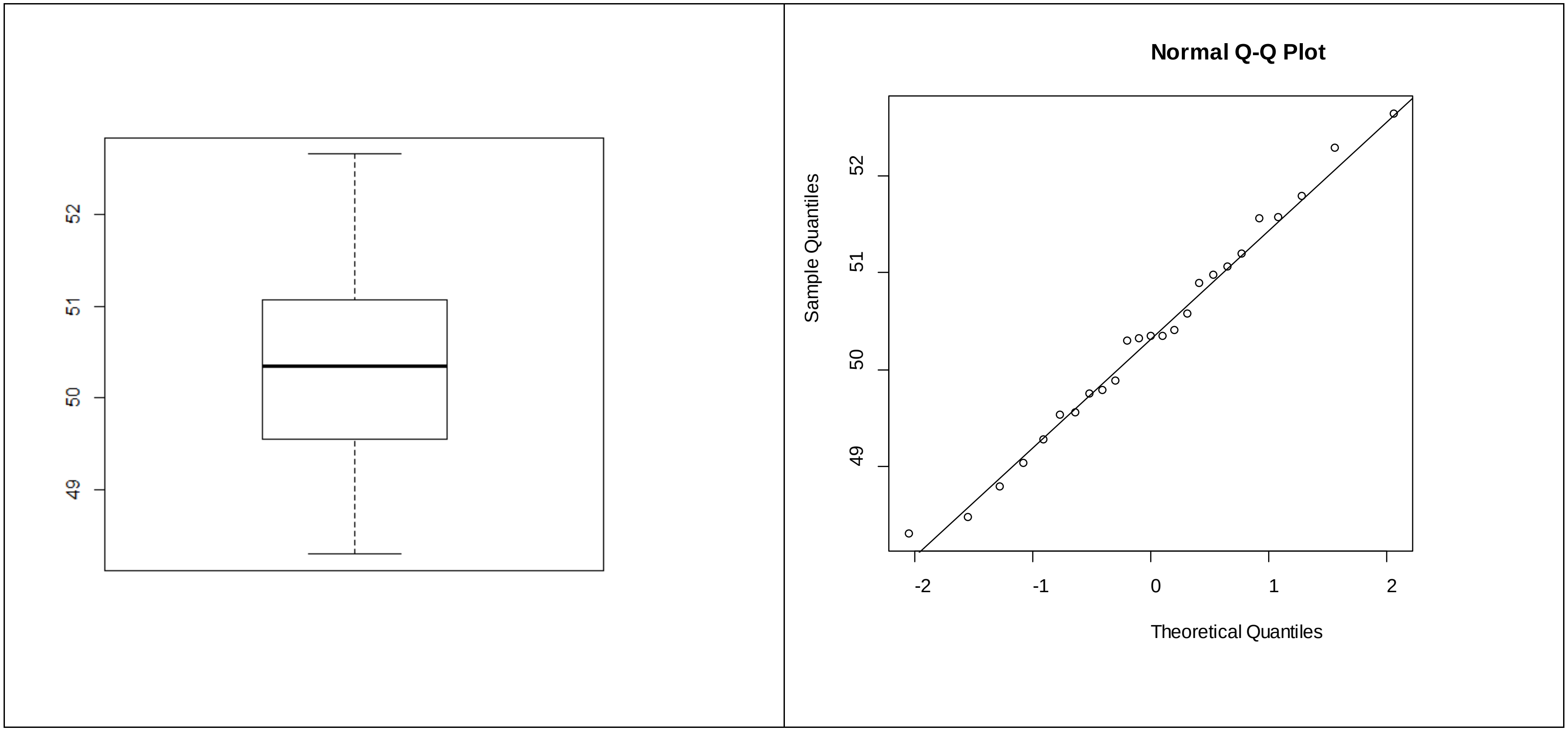
Inferencia estadística – Tests de hipótesis

Hasta ahora hemos visto como obtener, a partir de una muestra, un estimador puntual o un intervalo de confianza para un parámetro θ, ya sea la media, la varianza o la proporción poblacionales. Frecuentemente el objetivo del estudio es decidir, en base a la información que provee la muestra, entre dos hipótesis relativas a un parámetro.

Ejemplo:

Las aguas minerales pueden clasificarse de acuerdo a su contenido mineral en cálcicas, magnésicas, hiposódicas, fluoradas o carbónicas. Para que pueda ser calificada como magnésica, con lo cual podría ayudar en los tratamientos de la osteoporosis o en la re-mineralización ósea, debe tener un **contenido medio de magnesio mayor a de 50 mg/l.**

Con el propósito de evaluar la posibilidad de calificar una determinada marca de agua mineral como magnésica, se desea estimar su contenido medio de magnesio y se tomaron en forma independiente 25 muestras de dicha agua. El investigador analizó en forma exploratoria los datos, concluyendo que el supuesto de normalidad es razonable.



52
51
50
49
Normal Q-Q Plot
Sample Quantiles
Theoretical Quantiles
-2
-1
0
1
2

Situación 1:

Asumamos que obtuvo un promedio muestral de 50.35 mg/l y **por el momento que se sabe que el desvío estándar es 1 mg/l.**

¿Proveen estos datos evidencia de que el contenido medio de magnesio es superior a 50 mg/l o los resultados observados se deben simplemente al azar?

De acuerdo con la descripción, podríamos asumir que nuestros datos corresponden a una muestra aleatoria, es decir que las 25 mediciones son independientes e idénticamente distribuidas con una distribución que podría asumirse normal, por lo tanto corresponderían a

$$X_1,\ldots,X_{25} \quad \text{independientes, donde} \quad X_i \sim N(\mu,1)$$

Entonces

$$\bar{X} \sim N\left(\mu, \frac{1}{25}\right) \qquad \Leftrightarrow \qquad \frac{X-\mu}{\sqrt{1/25}} \sim N(0,1)$$

¿Qué es lo que estamos tratando de decidir? Nuestras hipótesis se refieren a μ, el contenido medio de magnesio, y se podrían enunciar así:

- $\mu \leq 50$ mg/l. En este caso no se califica como magnésica
- $\mu > 50$ mg/l. En este caso se la califica como magnésica

En primera instancia consideremos una versión simplificada:

i) $\mu = 50$ mg/l. En este caso no se califica como magnésica
ii) $\mu > 50$ mg/l. En este caso se la califica como magnésica

Si el verdadero contenido medio de magnesio en mg/l fuera 50, ¿cuál sería la probabilidad de que una v.a. normal con media 50 y varianza 1/25 tomase un valor igual o mayor aún que el observado, 50.35?

$$P\left(\overline{X} \geq 50.35\right) = P\left(\frac{\overline{X} - 50}{\sqrt{1/25}} \geq \frac{50.35 - 50}{\sqrt{1/25}}\right) = 1 - \Phi(1.75) = 0.04005916 \cong 0.04$$

Esta probabilidad se denomina **p-valor**.

¿Qué nos dice? Hubiese sido muy poco probable que se observase un valor promedio tan extremo como el observado o más aún si la media verdadera fuera 50.

Consideremos la versión simplificada:

iii) μ = 50 mg/l. En este caso no se califica como magnésica
iv) μ > 50 mg/l. En este caso se la califica como magnésica

A la primera hipótesis se la denomina **hipótesis nula** y se designa H_o. Esta hipótesis implica que no habrá cambio o que no hay efecto, es la hipótesis del status quo, o sea del no cambio respecto a la situación inicial. La segunda hipótesis se denomina **hipótesis alternativa** y se designa H_1 . Se la suele llamar la hipótesis del investigador.

Expresadas en términos del parámetro de interés las hipótesis del ejemplo serán

$$H_o: \mu = 50 \qquad \text{vs.} \qquad H_1: \mu > 50$$

Un test es una regla de decisión basada en un **estadístico** o función de la muestra, en este caso $\overline{X}$, y en una **zona de rechazo**, es decir un conjunto de valores para los cuáles se rechaza la hipótesis nula H_o.

¿Cómo se elige la zona de rechazo?

Observemos que al tomar una decisión en base a una muestra, podemos cometer dos tipos de error.

	No se rechaza H_o	**Se rechaza H_o**
H_o es cierta	OK	Error tipo I
H_o no es cierta	Error tipo II	OK

Debido a la variabilidad muestral, es imposible construir tests en los cuáles estemos absolutamente seguros de tomar la decisión correcta. Lo que podemos hacer es tratar de mantener bajas las probabilidades de error.

Llamaremos **nivel de significación del test**, y lo designaremos α, a la *probabilidad de error tipo I* (en realidad a la máxima probabilidad de error tipo I) y designaremos β a la *probabilidad de error tipo II.*

Como el estadístico se construye bajo la condición de que H_o es verdadera, lo que podemos controlar es la probabilidad de error tipo I. Elegiremos la zona de rechazo del test de manera de que la probabilidad de error tipo I sea un valor α predeterminado.

Volviendo al ejemplo, sabemos que, si H_o fuera cierta,

$$\frac{\overline{X}-50}{\sqrt{1/25}} \sim N(0,1)$$

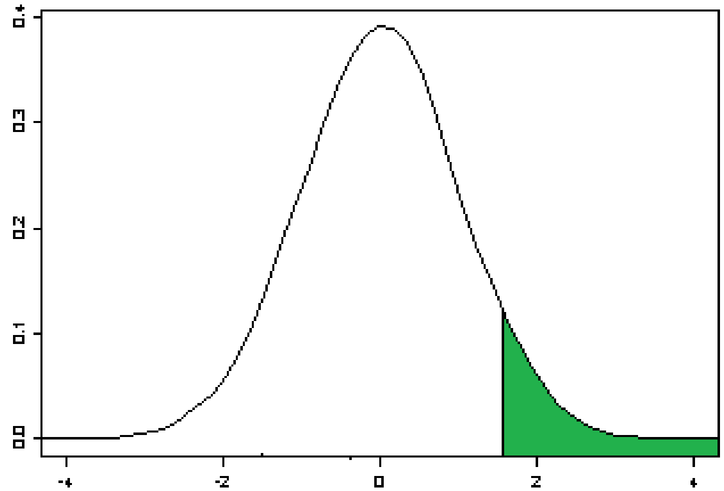
Notemos que rechazaríamos H_o para valores grandes de este cociente.

Si queremos que el test tenga nivel de significación α = 0.05, usaríamos la regla:

Rechazamos H_o si

$$\frac{\overline{X}-50}{\sqrt{1/25}} \geq 1.64$$

De hecho: P(rechazar H_o cuando μ = 50)= $P_{\mu=50}\left(\frac{\overline{X}-50}{\sqrt{1/25}} \geq 1.64\right) = 0.05$



Esta es la zona de rechazo del test de nivel 0.05. Si observamos un promedio igual a 50.35, el valor del estadístico es 1.75 y por lo tanto se rechaza H_o, mientras que si hubiésemos observado un promedio muestral igual a 50.25, el valor del estadístico habría sido 1.25 y no se rechazaría H_o.

Si quisiéramos que el test tuviese un nivel de significación α = 0.01, usaríamos la regla

Rechazamos H_o si

$$\frac{\overline{X} - 50}{1/5} \geq 2.32$$

Esta es la zona de rechazo del test de nivel 0.01.

Como hemos visto, al seleccionar la región de rechazo controlamos la probabilidad de error tipo I, pero ¿qué ocurre con el error tipo II?

Volvamos al test de nivel de significación α = 0.05, con

Zona de rechazo

$$\frac{\overline{X}-50}{1/5} \geq 1.64$$

Imaginemos la siguiente

Situación 2

Supongamos que en nuestro ejemplo observamos un contenido promedio de magnesio en la muestra de tamaño 25 igual a 50.25 mg/l. En este caso, el estadístico del test valdría 1.25, es decir

$$\frac{50.25-50}{1/5} = 1.25$$

y por lo tanto, no rechazamos H_o , luego si estuviéramos cometiendo un error, éste sería de tipo II.

¿Cuánto valdría esta probabilidad de cometer un error de tipo II?

Esta pregunta no la podemos responder si no fijamos un valor determinado de la hipótesis alternativa.

Por ejemplo:

Queremos calcular la probabilidad de no rechazar la hipótesis nula si el verdadero contenido medio de magnesio fuera 50.50 mg/l.

$$P_{\mu=50.50}\left(\frac{\overline{X}-50}{1/5}<1.64\right)=P_{\mu=50.50}\left(\overline{X}<1.64\cdot\frac{1}{5}+50\right)=P_{\mu=50.50}\left(\frac{\overline{X}-50.50}{1/5}<\frac{1.64\cdot\frac{1}{5}+50-50.50}{1/5}\right)$$

$$P_{\mu=50.50}\left(\frac{\overline{X}-50.50}{1/5}<-0.86\right)=P_{\mu=50.50}\left(\frac{\overline{X}-50.50}{1/5}<1.64+\frac{50-50.50}{1/5}\right)=\Phi(-0.86)=0.195$$

Es decir, que la probabilidad de error tipo II para el valor de μ = 50.50 es aproximadamente 0.20.

Notemos que la probabilidad de no rechazar la hipótesis nula si el verdadero contenido medio de magnesio fuera μ_1 quedaría:

$$P_{\mu=\mu_1}\left(\frac{\overline{X}-50}{1/5}<1.64\right)=P_{\mu=\mu_1}\left(\overline{X}<1.64\cdot\frac{1}{5}+50\right)=P_{\mu=\mu_1}\left(\frac{\overline{X}-\mu_1}{1/5}<\frac{1.64\cdot\frac{1}{5}+50-\mu_1}{1/5}\right)$$

$$P_{\mu=\mu_1}\left(\frac{\overline{X}-\mu_1}{1/5}<1.64+\frac{50-\mu_1}{1/5}\right)=\Phi(1.64+\frac{50-\mu_1}{1/5})$$

Es decir, que la probabilidad de error tipo II como función de μ_1 es decreciente.

Definición: La **función de potencia** de un test, $\pi(\mu)$, es la probabilidad de rechazar la hipótesis nula cuando el valor verdadero del parámetro es μ .

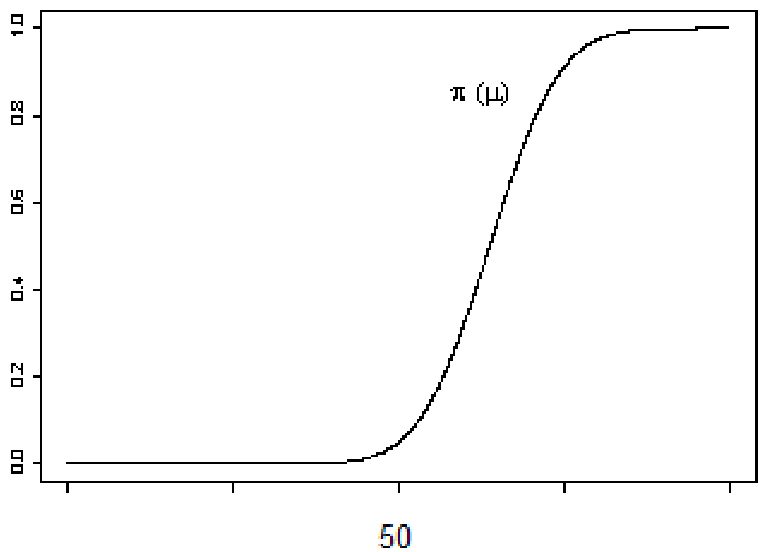
Utilizando la función de potencia es posible obtener una expresión general para los dos tipos de errores, pues

$$\pi(\mu) = \begin{cases} \alpha(\mu) & \text{si } \mu \in \mathrm{H}_o \\ 1-\beta(\mu) & \text{si } \mu \in \mathrm{H}_1 \end{cases}$$

donde $\alpha(\mu)$ y $\beta(\mu)$ denota las probabilidades de error tipo I y tipo II respectivamente cuando el verdadero valor del parámetro es μ.

Notemos que en nuestro ejemplo resulta:

$$\pi(\mu) = P_{\mu}\left(\frac{\overline{X}-50}{1/5} \geq 1.64\right) = P_{\mu}\left(\frac{\overline{X}-\mu}{1/5} \geq 1.64 + \frac{50-\mu}{1/5}\right) = 1 - \Phi(1.64 + \frac{50-\mu}{1/5})$$

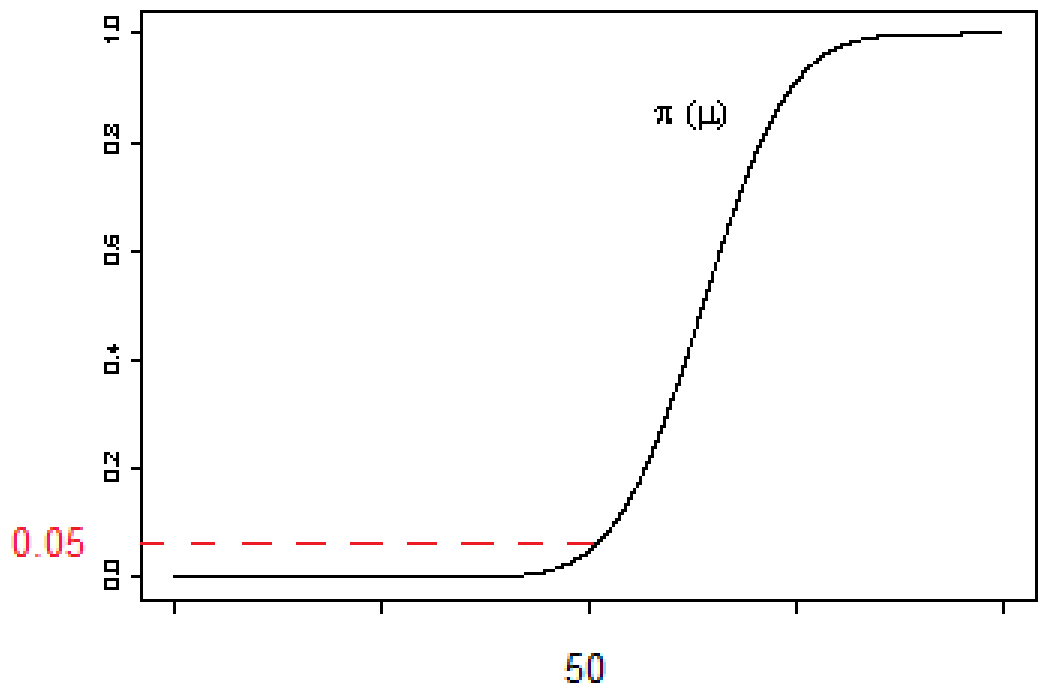


Función creciente de μ

El comportamiento de $\pi(\mu)$ justifica que usemos el test diseñado para

H_o: μ = 50 vs. H_1: μ > 50

también para H_o: μ≤ 50 vs. H_1: μ > 50


1.0
0.8
0.6
0.4
0.2
0.0
0.05
π (μ)
50

Tipos de hipótesis a testear:

Hipótesis unilaterales:

$H_o: \mu = \mu_o$ (ó $\mu \leq \mu_o$) vs. $H_1: \mu > \mu_o$

$H_o: \mu = \mu_o$ (ó $\mu \geq \mu_o$) vs. $H_1: \mu < \mu_o$

Hipótesis bilaterales:

$H_o: \mu = \mu_o$ vs. $H_1: \mu \neq \mu_o$

La forma de la región de rechazo dependerá de la hipótesis alternativa a testear.

Así, en el ejemplo presentado anteriormente, la zona de rechazo consiste en un intervalo de valores en la cola derecha de la distribución porque la hipótesis alternativa es de la forma $\mu > \mu_o$.

Tests de hipótesis de nivel α para los parámetros de la distribución normal:

Sea $X_1, X_2, \ldots, X_n$ una m.a. de una distribución $N(\mu, \sigma^2)$.

Tests para la media cuando la varianza es conocida: Supongamos que $\sigma^2 = \sigma_o^2$ es conocida y consideremos las siguientes hipótesis

a) H_o: $\mu = \mu_o$ (ó $\mu \leq \mu_o$) vs. H_1: $\mu > \mu_o$
b) H_o: $\mu = \mu_o$ (ó $\mu \geq \mu_o$) vs. H_1: $\mu < \mu_o$
c) H_o: $\mu = \mu_o$ vs. H_1: $\mu \neq \mu_o$

Estadístico del test: $Z = \sqrt{n}\dfrac{\overline{X} - \mu_o}{\sigma_o}$. Bajo H_o: $\mu = \mu_o$, $Z \sim N(0,1)$

<u>Región de rechazo</u>: Como dijimos, la zona de rechazo depende de la hipótesis alternativa. Estará dada, en cada caso, por

a) $Z \geq z_\alpha$ b) $Z \leq -z_\alpha$ c) $|Z| \geq z_{\alpha/2}$

Veamos el caso c)

Observemos que, así como la forma de la región de rechazo depende de la alternativa, su tamaño depende del nivel. Por ejemplo, consideremos el caso c). Como la alternativa es $\mu \neq \mu_o$, la forma de la región es $|T| \geq K$, pero como la probabilidad de rechazar H_o siendo cierta (P(Error tipo I)) debe ser α,

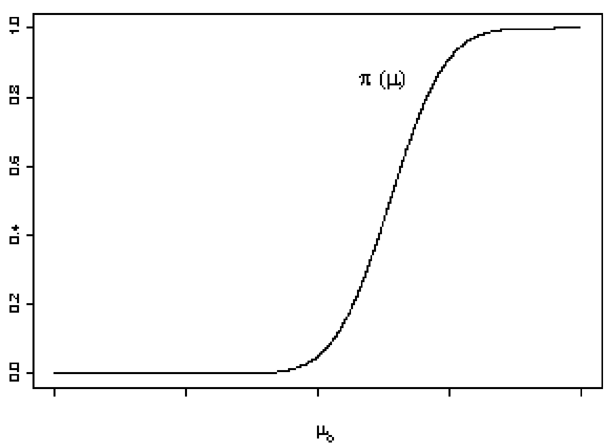
$$P_{\mu_o}\left(\left|\sqrt{n}\,\frac{\overline{X}-\mu_o}{\sigma_o}\right| \geq K\right) = \alpha \Leftrightarrow 1 - P_{\mu_o}\left(\left|\sqrt{n}\,\frac{\overline{X}-\mu_o}{\sigma_o}\right| < K\right) = 1 - P_{\mu_o}\left(-K < \sqrt{n}\,\frac{\overline{X}-\mu_o}{\sigma_o} < K\right) = \alpha$$

$$\Leftrightarrow 1 - \Phi(K) + \Phi(-K) = \alpha \Leftrightarrow 2(1-\Phi(K)) = \alpha \Leftrightarrow \Phi(K) = 1 - \frac{\alpha}{2} \Leftrightarrow K = z_{\alpha/2}$$

Función de potencia: La notación P_μ , como ya hemos visto, indicará la probabilidad cuando el valor verdadero del parámetro es μ .

a) $$\pi(\mu) = P_\mu\left(\frac{\overline{X} - \mu_o}{\sigma/\sqrt{n}} \geq z_\alpha \right) = P_\mu\left(\frac{\overline{X} - \mu + \mu - \mu_o}{\sigma/\sqrt{n}} \geq z_\alpha \right)$$

$$= P_\mu\left(\frac{\overline{X} - \mu}{\sigma_o/\sqrt{n}} \geq z_\alpha + \frac{\mu_o - \mu}{\sigma_o/\sqrt{n}} \right) = 1 - \Phi\left(z_\alpha + \frac{\mu_o - \mu}{\sigma_o/\sqrt{n}} \right)$$



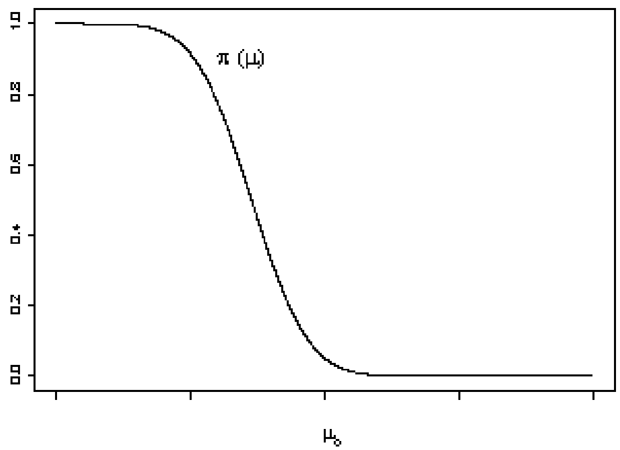
Observemos que esta función es creciente y $\pi(\mu_o)=\alpha$, entonces, si $\mu<\mu_o$, $\pi(\mu)<\alpha$. Por esta razón el test es de nivel α para las hipótesis

$$H_o:\ \mu \le \mu_o \quad \text{vs} \quad H_1:\ \mu > \mu_o$$

en el sentido de que la probabilidad de error tipo I es a lo sumo α.

b) $$\pi(\mu)=P_\mu\left(\frac{\overline{X}-\mu_o}{\sigma_o/\sqrt{n}}\le -z_\alpha\right)=P_\mu\left(\frac{\overline{X}-\mu+\mu-\mu_o}{\sigma_o/\sqrt{n}}\le -z_\alpha\right)$$

$$=P_\mu\left(\frac{\overline{X}-\mu}{\sigma_o/\sqrt{n}}\le -z_\alpha+\frac{\mu_o-\mu}{\sigma_o/\sqrt{n}}\right)=\Phi\left(-z_\alpha+\frac{\mu_o-\mu}{\sigma_o/\sqrt{n}}\right)$$



Observemos que esta función es decreciente y $\pi(\mu_o) = \alpha$, entonces, si $\mu > \mu_o$, $\pi(\mu) < \alpha$. Por esta razón el test es de nivel α para las hipótesis

$$H_o: \mu \geq \mu_o \quad \text{vs} \quad H_1: \mu < \mu_o$$

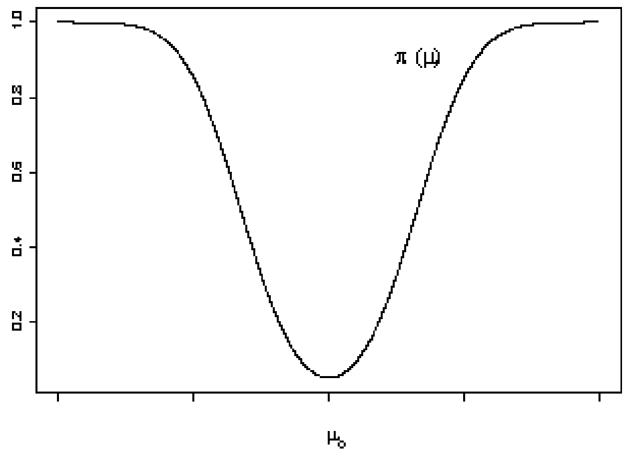
en el sentido de que la probabilidad de error tipo I es a lo sumo α.

c) $$\pi(\mu) = P_\mu\left(\left|\frac{\overline{X}-\mu_o}{\sigma_o/\sqrt{n}}\right| \ge z_{\alpha/2}\right) = 1 - P_\mu\left(\left|\frac{\overline{X}-\mu_o}{\sigma_o/\sqrt{n}}\right| < z_{\alpha/2}\right)$$

$$= 1 - P_\mu\left(-z_{\alpha/2} < \frac{\overline{X}-\mu+\mu-\mu_o}{\sigma_o/\sqrt{n}} < z_{\alpha/2}\right)$$

$$= 1 - P_\mu\left(-z_{\alpha/2} + \frac{\mu_o-\mu}{\sigma_o/\sqrt{n}} < \frac{\overline{X}-\mu}{\sigma_o/\sqrt{n}} < z_{\alpha/2} + \frac{\mu_o-\mu}{\sigma_o/\sqrt{n}}\right)$$

$$= 1 - \Phi\left(z_{\alpha/2} + \frac{\mu_o-\mu}{\sigma_o/\sqrt{n}}\right) + \Phi\left(-z_{\alpha/2} + \frac{\mu_o-\mu}{\sigma_o/\sqrt{n}}\right)$$



Observemos que esta función decrece hasta μ_o donde $\pi(\mu_o) = \alpha$ y crece a partir de allí.

Tamaño de muestra requerido para obtener una probabilidad de error tipo II dada para un valor de $\mu = \mu_1$ en la alternativa:

Recordemos que el error de tipo II se define como "no rechazar la hipótesis nula H_o cuando es falsa". Buscamos el valor de *n* para que la probabilidad de error tipo II sea menor que β cuando $\mu = \mu_1$ es un valor fijo en H_1. Consideremos el caso de nuestro ejemplo.

Vimos que cuando n=25 la probabilidad de no rechazar la hipótesis nula si el verdadero contenido medio de magnesio fuera 50.50 mg/l.

$$P_{\mu=50.50}\left(\frac{\overline{X}-50}{\sqrt{1/25}}<1.64\right)=P_{\mu=50.50}\left(\frac{\overline{X}-50.50}{\sqrt{1/25}}<1.64+\frac{50-50.50}{\sqrt{1/25}}\right)=\Phi(-0.86)=0.195$$

Si mantenemos el nivel en 0.05, la varianza es 1 y fijamos el valor de la alternativa en μ =50.50 mg/l, ¿podríamos lograr de alguna forma que esta probabilidad fuera menor? Notemos que también depende del tamaño muestral.

¿Cuál debería ser el tamaño muestral si quisiéramos que esta probabilidad fuera a lo sumo 0.10? Hagamos la cuenta.

Consideremos los distintos casos genéricos.

a) H$_o$: $\mu \leq \mu_o$ vs H$_1$: $\mu > \mu_o$

$$P_{\mu_i}\left(\frac{\overline{X}-\mu_o}{\sigma_o/\sqrt{n}} < z_\alpha\right) \leq \beta \Leftrightarrow 1-\pi(\mu_1) \leq \beta \Leftrightarrow \pi(\mu_1) \geq 1-\beta$$

$$P_{\mu_i}\left(\frac{\overline{X}-\mu_o}{\sigma_o/\sqrt{n}} < z_\alpha\right) \leq \beta \Leftrightarrow \Phi\left(z_\alpha + \frac{\mu_o-\mu_1}{\sigma_o/\sqrt{n}}\right) \leq \beta \Leftrightarrow z_\alpha + \frac{\mu_o-\mu_1}{\sigma_o/\sqrt{n}} \leq z_{1-\beta}$$

Observemos que en este caso la alternativa es H$_1$: $\mu > \mu_o$, por lo tanto, $\mu_o - \mu_1 < 0$ y se obtiene

$$n \geq \left(\frac{\left(z_\alpha - z_{1-\beta}\right)\sigma_o}{\mu_1-\mu_o}\right)^2 = \left(\frac{\left(z_\alpha + z_\beta\right)\sigma_o}{\mu_1-\mu_o}\right)^2$$

b) Consideremos H$_o$: $\mu \geq \mu_o$ vs H$_1$: $\mu < \mu_o$

$$P_{\mu_i}\left(\frac{\overline{X}-\mu_o}{\sigma_o/\sqrt{n}} > -z_\alpha\right) \leq \beta \Leftrightarrow 1-\pi(\mu_1) \leq \beta \Leftrightarrow \pi(\mu_1) \geq 1-\beta$$

$$\Leftrightarrow \Phi\left(-z_\alpha + \frac{\mu_o-\mu_1}{\sigma_o/\sqrt{n}}\right) \geq 1-\beta \Leftrightarrow -z_\alpha + \frac{\mu_o-\mu_1}{\sigma_o/\sqrt{n}} \geq z_\beta$$

Observemos que en este caso la alternativa es H$_1$: $\mu < \mu_o$, por lo tanto, $\mu_o - \mu_1 > 0$ y se obtiene

$$n \geq \left(\frac{(z_\alpha + z_\beta)\sigma_o}{\mu_0 - \mu_1}\right)^2$$

c) Para el caso bilateral, el cálculo del tamaño de muestra se hace en forma aproximada, despreciando la más pequeña de las dos probabilidades.

Tests para la media cuando la varianza es desconocida: Supongamos ahora que la varianza es desconocida y consideremos las mismas hipótesis sobre μ.

a) H_o: $\mu = \mu_o$ (ó $\mu \leq \mu_o$) vs H_1: $\mu > \mu_o$
b) H_o: $\mu = \mu_o$ (ó $\mu \geq \mu_o$) vs H_1: $\mu < \mu_o$
c) H_o: $\mu = \mu_o$ vs H_1: $\mu \neq \mu_o$

Estadístico del test: $T = \sqrt{n}\,\dfrac{\overline{X} - \mu_o}{S}$. Bajo H_o: $\mu = \mu_o$, $T \sim t_{n-1}$

Región de rechazo: Como siempre la forma de la zona de rechazo depende de la hipótesis alternativa. Estará dada, en cada caso, por

a) $T \geq t_{n-1,\alpha}$

b) $T \leq -t_{n-1,\alpha}$

c) $|T| \geq t_{n-1,\alpha/2}$

El tamaño de la zona de rechazo depende del nivel. Por ejemplo, consideremos el caso a). Como la alternativa es $\mu > \mu_o$, la forma de la región es $T \geq K$, pero como la probabilidad de rechazar H_o siendo cierta (P(Error tipo I)) debe ser α,

$$P_{\mu_o}\left(\sqrt{n}\,\frac{\overline{X}-\mu_o}{S} \geq K \right) = \alpha \Leftrightarrow 1 - P_{\mu_o}\left(\sqrt{n}\,\frac{\overline{X}-\mu_o}{S} \leq K \right) = \alpha$$

$$\Leftrightarrow 1 - F_T(K) = \alpha \Leftrightarrow F_T(K) = 1-\alpha \Leftrightarrow K = t_{n-1,\,\alpha}$$

donde F_T designa la función de distribución de una v.a. *t* con *n*-1 grados de libertad.

Función de potencia y cálculo del tamaño de muestra para obtener una probabilidad de error tipo II dada:

La función de potencia de este test es complicada porque la distribución del estadístico cuando $\mu \neq \mu_o$ es una distribución *t* no central. Aunque hay tablas y gráficos que permiten obtener probabilidades para una distribución de este tipo, no los estudiaremos. Por la misma razón, no calcularemos tamaño de muestra para obtener una probabilidad de error tipo II dada para una alternativa fija.

Respecto al p-valor, cuando se utilizan tablas sólo es posible obtener una cota, ya que las tablas proveen solamente algunos valores críticos de la distribución *t*, a menos que se utilice un paquete como el R.