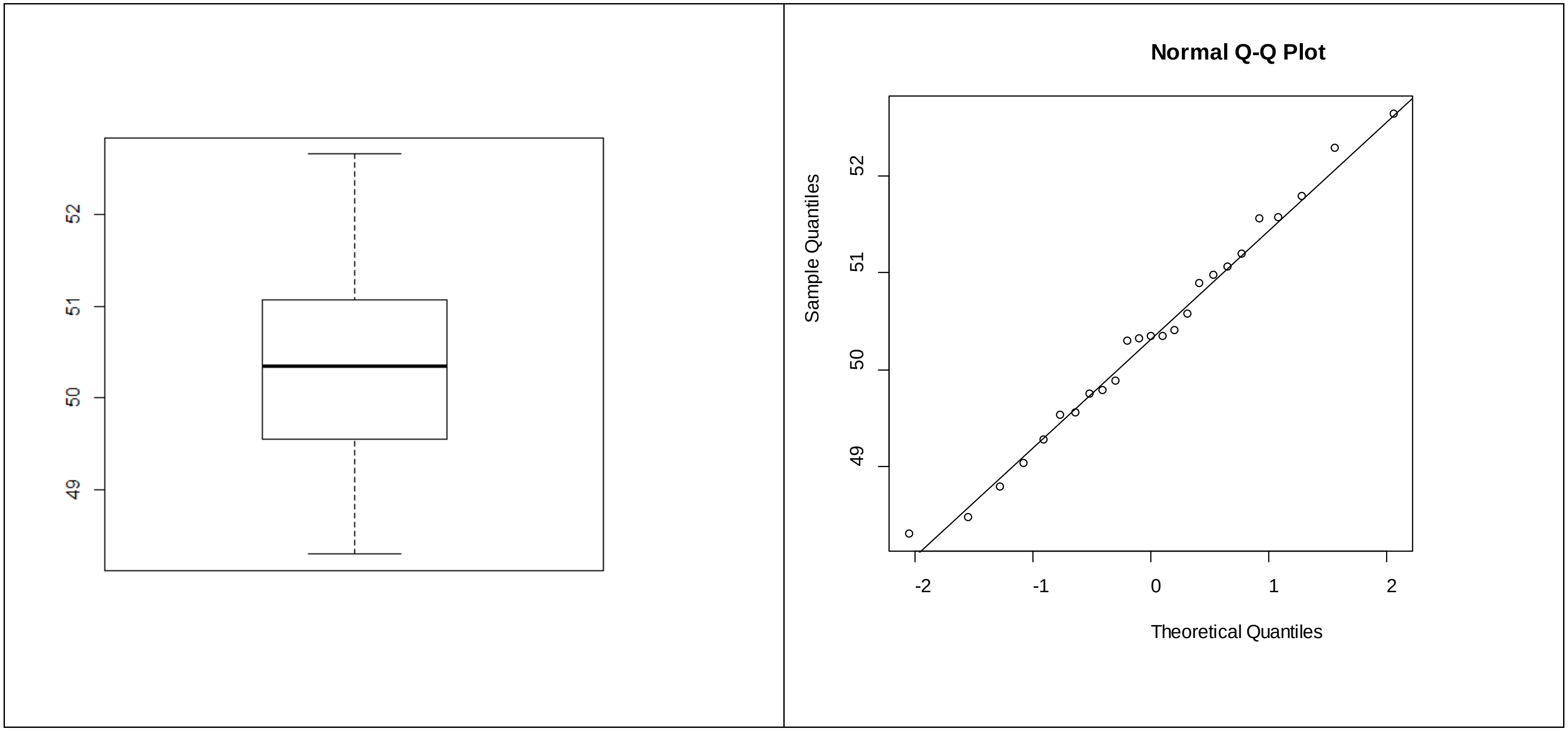
Inferencia estadística – Tests de hipótesis

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
52
51
50
49
-2
-1
0
1
2
Theoretical Quantiles

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

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

Entonces

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

¿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,

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

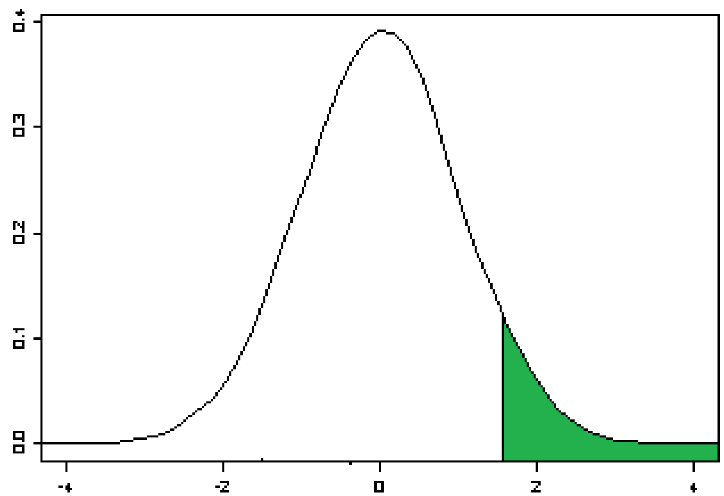
Notemos que rechazaríamos H_o para valores grandes de este cociente.

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.

<u>Definición</u>: 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 .

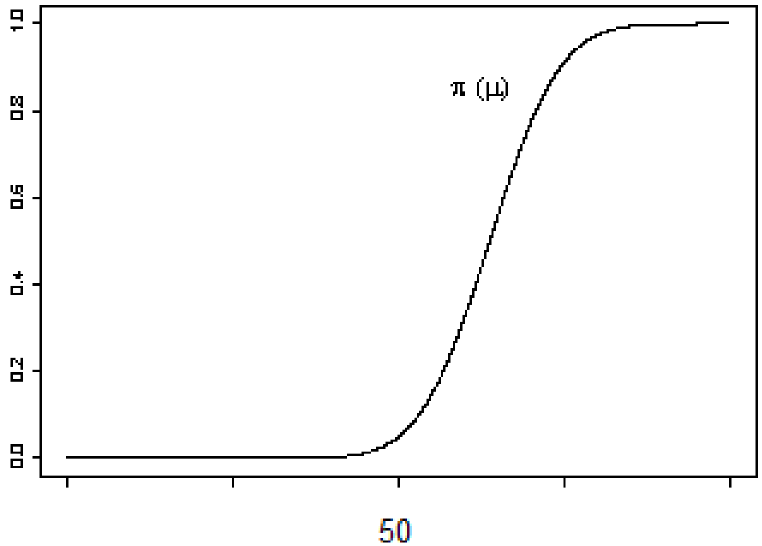
Utilizando la función de potencia es posible obtener una expresión general para los dos tipos de errores, pues

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

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:

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

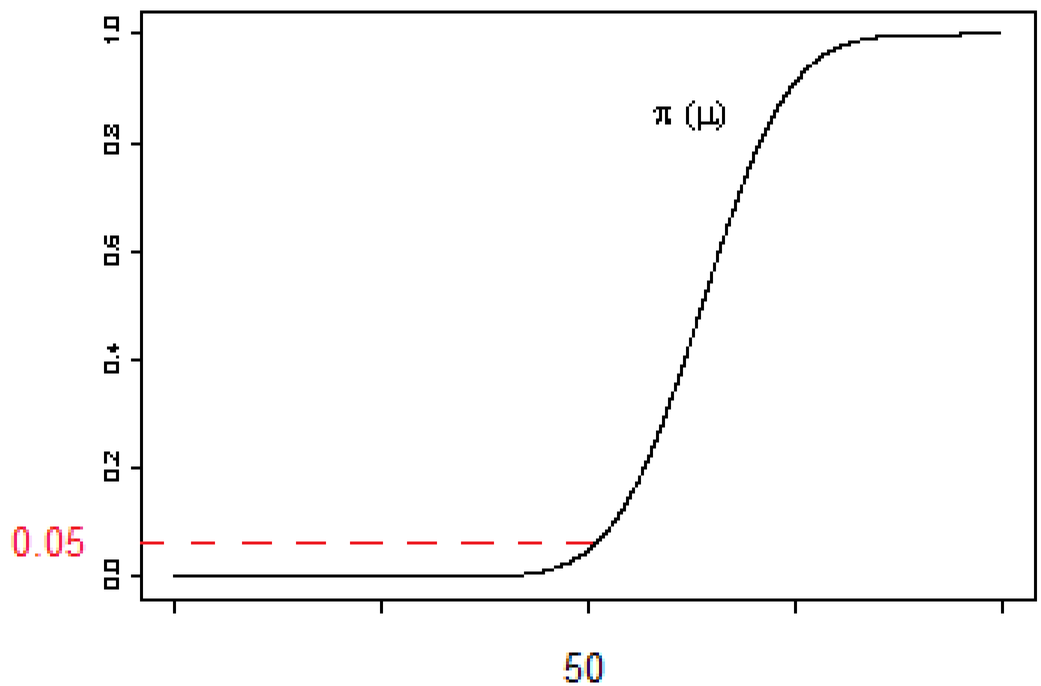


Función creciente de μ

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

H_o: $\mu = 50$ vs. H_1: $\mu > 50$

también para H_o: $\mu \leq 50$ vs. H_1: $\mu > 50$



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 $X_1, X_2, \dots, X_n$ 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,

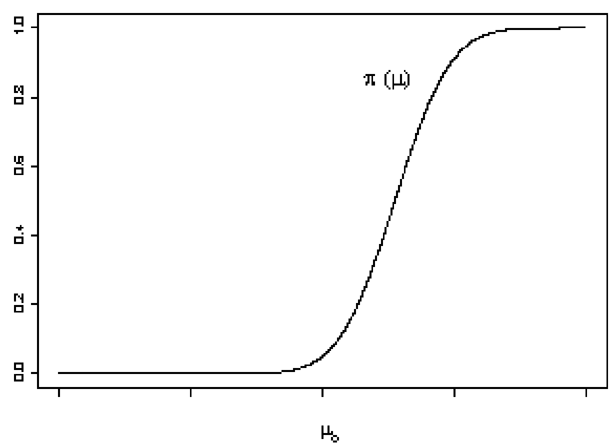
$$P_{\mu_o}\left(\left|\sqrt{n}\,\frac{\overline{X}-\mu_o}{\sigma_o}\right| \geq K\right) = \alpha \Leftrightarrow 1 - P_{\mu_o}\left(\left|\sqrt{n}\,\frac{\overline{X}-\mu_o}{\sigma_o}\right| < K\right) = 1 - P_{\mu_o}\left(-K < \sqrt{n}\,\frac{\overline{X}-\mu_o}{\sigma_o} < K\right) = \alpha$$

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

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

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

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



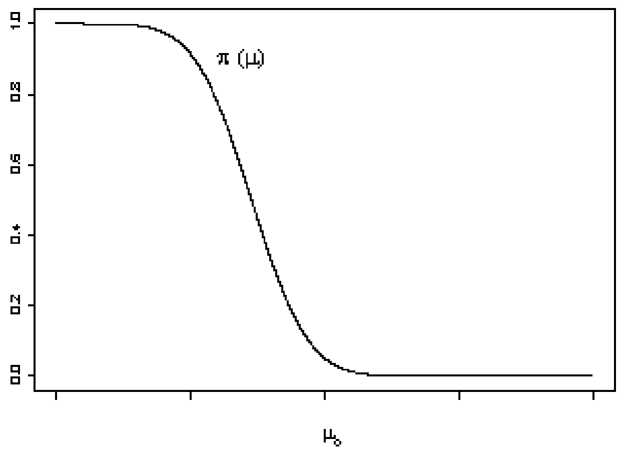
Observemos que esta función es creciente y $\pi(\mu_o) = \alpha$, entonces, si $\mu < \mu_o$, $\pi(\mu) < \alpha$. Por esta razón el test es de nivel α para las hipótesis

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

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

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

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



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

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

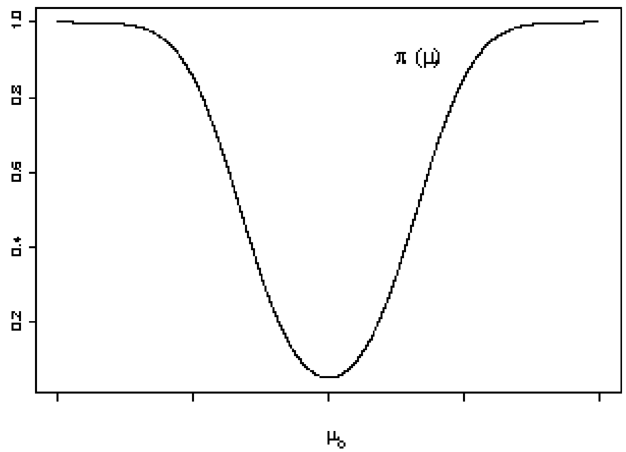
en el sentido de que la probabilidad de error tipo I es a lo sumo α.

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

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

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

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



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$_o$: $\mu \le \mu_o$ vs H$_1$: $\mu > \mu_o$

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

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

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

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

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.

Recordemos el resultado que vimos en el contexto de Intervalos de Confianza

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

a) $\bar{X} \sim N\left(\mu, \frac{\sigma^2}{n}\right) \quad \Leftrightarrow \quad \sqrt{n}\frac{\bar{X}-\mu}{\sigma} \sim N(0,1)$

b) $\frac{(n-1)S^2}{\sigma^2} \sim \chi^2_{n-1}$ con $S^2 = \frac{\sum_{i=1}^{n}(X_i - \bar{X})^2}{n-1}$

c) $\bar{X}$ y S^2 son independientes

d) $\sqrt{n}\frac{\bar{X}-\mu}{S} \sim t_{n-1}$

Las posibles hipótesis son:

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: $\mu = \mu_o$, tenemos que $T \sim t_{n-1}$

<u>Región de rechazo</u>: 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.

Veamos un ejemplo.

Tests para la varianza cuando la media es desconocida: Las hipótesis a testear son

a) H_o: $\sigma^2 = \sigma_o^2$ (ó $\sigma^2 \leq \sigma_o^2$) vs H_1: $\sigma^2 > \sigma_o^2$

b) H_o: $\sigma^2 = \sigma_o^2$ (ó $\sigma^2 \geq \sigma_o^2$) vs H_1: $\sigma^2 < \sigma_o^2$

c) H_o: $\sigma^2 = \sigma_o^2$ vs H_1: $\sigma^2 \neq \sigma_o^2$

Estadístico del test: $U = \dfrac{(n-1)S^2}{\sigma_o^2}$.

Bajo H_o: $\sigma^2 = \sigma_o^2$, tenemos que $U \sim \chi^2_{n-1}$

Región de rechazo: Como siempre la forma de la zona de rechazo depende de la hipótesis alternativa. En este caso, estará dada por

a) $U \geq \chi^2_{n-1,\alpha}$

b) $U \leq \chi^2_{n-1,1-\alpha}$

c) $U \geq \chi_{n-1,\alpha/2}$ ó $U \leq \chi_{n-1,1\text{-}\alpha/2}$

El tamaño de la zona de rechazo depende del nivel. Por ejemplo, consideremos el caso b). Como la alternativa es $\sigma^2 < \sigma_o^2$, la forma de la región es $U \leq K$, pero como la probabilidad de rechazar H_o siendo cierta (P(Error tipo I)) debe ser α,

$$P_{\sigma_o^2}\left(\frac{(n-1)S^2}{\sigma_o^2} \leq K \right) = \alpha \Leftrightarrow K = \chi^2_{n-1,\,\alpha}$$

Ejemplo: Se toman 25 determinaciones de la temperatura en cierto sector de un reactor, obteniéndose

$$\overline{x} = 249^{\circ}C \qquad \text{y} \qquad s = 2.8^{\circ}C$$

Interesa saber, a nivel 0.05

a) si existe evidencia para decidir que la temperatura media en ese sector del reactor es menor que $250^{\circ}C$. Calcular el p-valor.
b) si existe evidencia para decidir que la varianza de la temperatura en ese sector del reactor es mayor que $\left(2^{\circ}C\right)^{2}$.

a) Las hipótesis a testear son

$$H_o: \mu = 250 \text{ (ó } \mu \geq 250) \quad \text{vs} \quad H_1: \mu < 250$$

El estadístico del test será $T = \sqrt{n}\,\dfrac{\overline{X} - 250}{S}$

y la región de rechazo estará dada por los valores de *T* tales que

$$T=\sqrt{n}\,\frac{\overline{X}-250}{S}\leq -t_{n-1,\,0.05}$$

En nuestro caso, n = 25 y por lo tanto $-t_{24,\,0.05}=-1.71$. Además el valor observado de T es T_{obs}= -1.7857 y por lo tanto se rechaza H_o, es decir que a nivel 0.05 encontramos evidencia de que la temperatura media del reactor es menor que $250^{\circ}C$.

¿Cuánto vale el p-valor? El p-valor lo calculamos como la probabilidad de observar un valor tan extremo del estadístico como el observado o más extremo aún, bajo H_o.

En este ejemplo, rechazamos cuando observamos valores pequeños del estadístico, por lo tanto calculamos

$$p-valor=P_{\mu=250}\left(T\leq T_{OBS}\right)=P_{\mu=250}\left(T\leq -1.7857\right)$$

siendo $T=\sqrt{25}\,\dfrac{\overline{X}-250}{S}\sim t_{24}$

Con R o cualquier otro paquete que lo permita, podemos calcular esta probabilidad, si solo disponemos de las tablas, en general nos tendremos que conformar con acotarla.
En R usariamos la función pt:

```
pt(-1.7857,24)
[1] 0.04339571
```

es decir que el p-valor es 0.04339571. Esto quiere decir que rechazaremos H_o cuando realicemos un test de t como el que hemos hecho cuando el nivel de significación sea $\alpha>0.04339571$, mientras que no rechazaremos H_o cuando el nivel de significación del test de t sea $\alpha<0.04339571$. Así , rechazamos para $\alpha=0.05$, pero no rechazamos H_o si $\alpha=0.01$.

b) Las hipótesis a testear son

$$H_o: \sigma^2 = 4 \text{ (ó } \sigma^2 \leq 4\text{)} \quad \text{vs} \quad H_1: \sigma^2 > 4$$

El estadístico del test será $U = \dfrac{(n-1)S^2}{\sigma^2}$, y la región de rechazo estará dada por los valores de *U* tales que

$$U = \frac{(n-1)S^2}{4} \geq \chi^2_{n-1,\,0.05}$$

En nuestro caso, *n* = 25 y por lo tanto $\chi^2_{24,\,0.05} = 36.42$. Como el valor observado de *U* es U_{obs}=47.04, se rechaza H_o. Es decir, a nivel 0.05 hay evidencia de que la varianza de la temperatura del reactor es mayor que $(2^\circ C)^2$.

Tests de hipótesis de nivel aproximado (o asintótico) α para la media de una distribución cualquiera:

Sea $X_1, X_2, \ldots, X_n$ una m.a. de una distribución con media μ y varianza $\sigma^2 < \infty$. Aplicando el Teorema Central del Límite, sabemos que

$$\frac{\overline{X}-\mu}{\sigma/\sqrt{n}} \xrightarrow{d} Z \sim N(0,1)$$

Además, utilizando la propiedad enunciada al construir intervalos de confianza de nivel asintótico (1- α) para la media de una distribución cualquiera,

$$\left.\begin{array}{l} \sqrt{n}\,\dfrac{\overline{X}-\mu}{\sigma} \xrightarrow{d} N(0,1) \\ \dfrac{\sigma}{S} \xrightarrow{p} 1 \end{array}\right\} \quad \Rightarrow \quad \sqrt{n}\,\frac{\overline{X}-\mu}{S} \xrightarrow{d} N(0,1)$$

Por lo tanto, si *n* es suficientemente grande,

$$\sqrt{n}\,\frac{\overline{X}-\mu}{S} \overset{(a)}{\sim} N(0,1)$$

Supongamos que se desea testear a nivel aproximado α alguna de las hipótesis siguientes:

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: $\mu \neq \mu_o$

y que n es suficientemente grande. Utilizando como estadístico $T = \sqrt{n}\,\frac{\overline{X}-\mu_o}{s}$, las siguientes regiones de rechazo proveen tests del nivel requerido para cada una de las hipótesis:

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

Test de hipótesis de nivel aproximado (o asintótico) α para una proporción (parámetro p de la distribución binomial):

Sea $X_1, X_2, \ldots, X_n$ una m.a. de una distribución Bi(1,p). Entonces, $X = \sum_{i=1}^{n} X_i$ ~ Bi(n,p).

Aplicando el **Teorema Central del Límite**, si n es suficientemente grande,

$$\frac{\overline{X} - p}{\sqrt{\frac{p(1-p)}{n}}} \xrightarrow{d} Z \sim N(0,1)$$

siendo $\overline{X}$ la proporción muestral o frecuencia relativa de éxitos.

Un test de nivel aproximado α para las hipótesis:

a) H_o: $p = p_o$ vs H_1: $p > p_o$
b) H_o: $p = p_o$ vs H_1: $p < p_o$
c) H_o: $p = p_o$ vs H_1: $p \neq p_o$

se basa en el estadístico $\dfrac{\overline{X} - p_o}{\sqrt{\dfrac{p_o(1-p_o)}{n}}}$, el cual, si H_o es cierta, tiene distribución aproximada N(0,1). Las regiones de rechazo estarán dadas por

a) $\dfrac{\overline{X} - p_o}{\sqrt{\dfrac{p_o(1-p_o)}{n}}} \geq z_\alpha$

b) $\dfrac{\overline{X} - p_o}{\sqrt{\dfrac{p_o(1-p_o)}{n}}} \leq -z_\alpha$

c) $\left|\dfrac{\overline{X} - p_o}{\sqrt{\dfrac{p_o(1-p_o)}{n}}}\right| \geq z_{\alpha/2}$

Relación entre tests de hipótesis bilaterales e intervalos de confianza

Introduciremos esta relación a través de un ejemplo. Sea $X_1, X_2, ..., X_n$ una m.a. de una distribución $N(\mu, \sigma^2)$. Sabemos que, cuando la varianza es desconocida, el intervalo de confianza para μ de nivel $1 - \alpha$ está dado por

$$\left[\overline{X} - t_{n-1, \alpha/2} \frac{s}{\sqrt{n}}, \overline{X} + t_{n-1, \alpha/2} \frac{s}{\sqrt{n}} \right].$$

Supongamos ahora que deseamos testear a nivel α las siguientes hipótesis:

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

Dado que el intervalo construido contiene con alta probabilidad al valor verdadero de μ, si μ_o no pertenece al intervalo, esto nos llevaría a sospechar que la hipótesis nula es falsa.

Es decir, podríamos construir un test de nivel α , rechazando Ho si μ_o no pertenece al intervalo de confianza, dado que

$$P(EI) = P_{\mu_o}\left(\mu_o \notin \left[\overline{X} - t_{n-1,\alpha/2}\frac{s}{\sqrt{n}}, \overline{X} + t_{n-1,\alpha/2}\frac{s}{\sqrt{n}} \right] \right)$$

$$= 1 - P_{\mu_o}\left(\mu_o \in \left[\overline{X} - t_{n-1,\alpha/2}\frac{s}{\sqrt{n}}, \overline{X} + t_{n-1,\alpha/2}\frac{s}{\sqrt{n}} \right] \right) = 1 - (1-\alpha) = \alpha$$

Proposición: Sea $IC(X_1, X_2, ..., X_n)$ un intervalo de confianza de nivel 1 - α para un parámetro θ, obtenido a partir de una m.a. $X_1, X_2, ..., X_n$. Consideremos el problema de testear las hipótesis

$$H_o: \theta = \theta_o \qquad \text{vs.} \qquad H_1: \theta \neq \theta_o$$

El test que rechaza H_o cuando $\theta_o \notin IC(X_1, X_2, ..., X_n)$, tiene nivel α.

Ejemplo: Sea $X_1, X_2, \ldots, X_n$ una m.a. de una distribución $N(\mu, \sigma^2)$. Recordemos que hemos obtenido el siguiente intervalo de confianza de nivel exacto 1 - α para σ^2

$$IC = \left(\frac{(n-1)S^2}{\chi^2_{n-1,\alpha/2}}, \frac{(n-1)S^2}{\chi^2_{n-1,1-\alpha/2}} \right)$$

Si deseamos testear las hipótesis

$$H_o: \sigma^2 = \sigma^2_o \qquad \text{vs.} \qquad H_1: \sigma^2 \neq \sigma^2_o$$

El test que rechaza H_o si $\sigma^2_o \notin IC$ tiene nivel α.

Por lo tanto, revisitando el ejemplo que analizamos en las clases de IC de 49 observaciones independientes de una distribución normal y tales que $\overline{x} = 160$ y $s = 35$ y el IC para σ^2 de nivel 0.95 resultaba

$$\left(\frac{48 \cdot 35^2}{69.02}, \frac{48 \cdot 35^2}{30.75} \right) = (851.93, 1912.20)$$

(ya que $\chi^2_{n-1,\alpha/2} = \chi^2_{48,0.025} = 69.02$ y $\chi_{n-1,1-\alpha/2} = \chi_{48,0.975} = 30.75$), si nos interesara testear

$$H_o: \sigma^2 = 800 \text{ vs. } H_1: \sigma^2 \neq 800$$

con un nivel de significación igual a 0.05, rechazaríamos que H_o en favor de la hipótesis H_1: $\sigma^2 \neq 800$.

1 Test para comparar medias de dos muestras normales independientes

Sean $X_1, \ldots, X_{n_1} \sim N\left(\mu_1, \sigma_1^2\right)$, e $Y_1, \ldots, Y_{n_2} \sim N\left(\mu_2, \sigma_2^2\right)$ dos muestras aleatorias independientes entre sí. Se quiere testear

$$H_0 : \mu_1 - \mu_2 = \delta \text{ versus alguna de las alternativas siguientes}$$
$$H_1 : \mu_1 - \mu_2 \neq \delta \qquad H_1 : \mu_1 - \mu_2 > \delta \qquad H_1 : \mu_1 - \mu_2 < \delta$$

1.1 Caso varianzas conocidas

Estadístico del test

$$\frac{\overline{X} - \overline{Y} - \delta}{\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}} \underset{\substack{\uparrow \\ \text{bajo } H_0}}{\sim} N(0,1)$$

1.2 Caso varianzas desconocidas pero iguales

Estadístico del test

$$\frac{\overline{X} - \overline{Y} - \delta}{\sqrt{s_p^2 \left(\frac{1}{n_1} + \frac{1}{n_2} \right)}} \underset{\text{bajo } H_0}{\underset{\uparrow}{\sim}} t_{n_1+n_2-2}$$

donde

$$\begin{aligned} s_p^2 &= \frac{s_1^2 (n_1 - 1) + s_2^2 (n_2 - 1)}{n_1 + n_2 - 2} \\ &= \frac{1}{n_1 + n_2 - 2} \left(\sum_{i=1}^{n_1} (X_i - \overline{X})^2 + \sum_{i=1}^{n_2} (Y_i - \overline{Y})^2 \right) \end{aligned}$$

1.3 Caso varianzas desconocidas y no necesariamente iguales

En este caso tenemos el test de Welch que tiene nivel aproximado. Estadístico del test

$$\frac{\overline{X} - \overline{Y} - \delta}{\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}} \quad \underset{\underset{\text{bajo } H_0}{\uparrow}}{\overset{\text{aprox}}{\sim}} \quad t_K$$

donde

$$s_1^2 = \frac{1}{n_1 - 1} \sum_{i=1}^{n_1} (X_i - \overline{X})^2 \tag{1}$$

$$s_2^2 = \frac{1}{n_2 - 1} \sum_{i=1}^{n_2} (Y_i - \overline{Y})^2 \tag{2}$$

$$K = \left[\frac{\left(\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2} \right)^2}{\frac{\left(\frac{s_1^2}{n_1} \right)^2}{n_1 - 1} + \frac{\left(\frac{s_2^2}{n_2} \right)^2}{n_2 - 1}} \right]$$

2 Test para comparar varianzas de dos muestras normales independientes

Sean $X_1, \ldots, X_{n_1} \sim N\left(\mu_1, \sigma_1^2\right)$, e $Y_1, \ldots, Y_{n_2} \sim N\left(\mu_2, \sigma_2^2\right)$ dos muestras aleatorias independientes entre sí. Se quiere testear

$$\begin{aligned} H_0 &: \sigma_1^2 = \sigma_2^2 \\ H_1 &: \sigma_1^2 \neq \sigma_2^2 \end{aligned}$$

Estadístico del test

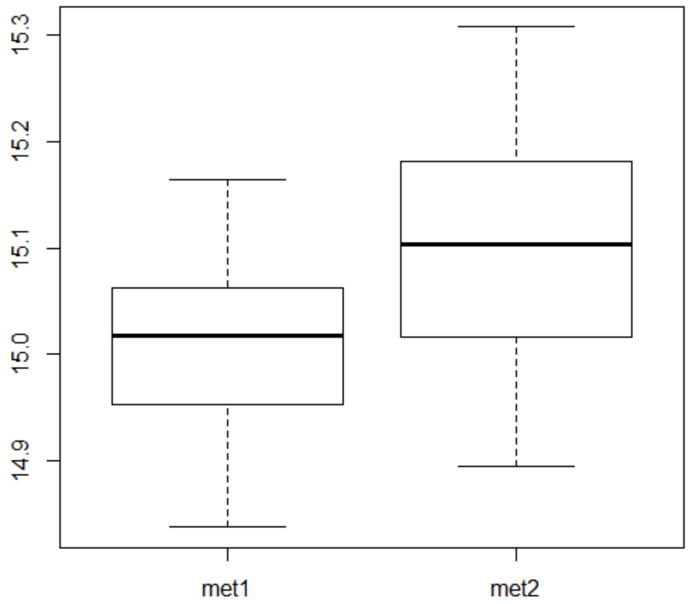
$$\frac{s_1^2}{s_2^2} \underset{\substack{\uparrow \\ \text{bajo } H_0}}{\sim} F_{n_1-1, n_2-1}$$

donde s_1 y s_2 están definidos en (1) y (2) respectivamente.

Veamos un ejemplo: Tenemos dos métodos para medir el contenido de hierro (%) en un mineral. Se realizan 17 mediciones con el primer método y 20 con el segundo. Queremos comparar los dos métodos, es decir si en media miden lo mismo o no.

```
met1
14.95310 15.14682 15.04426 15.01462 15.01715 15.05925 15.06192 15.03945 14.83784 14.90540
14.90021 14.98898 14.95323 15.08173 14.98018 15.16452 15.13923

met2
15.10411 15.00088 14.89470 14.92575 15.03312 14.97401 15.15059 15.10195 15.14152 15.13349
15.08503 15.30722 15.20021 15.27217 15.16294 15.07443 15.22297 14.94998 15.20076 15.07910

summary(met1)
   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
  14.84   14.95   15.02   15.02   15.06   15.16

summary(met2)
   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
  14.89   15.03   15.10   15.10   15.17   15.31

c(sd(met1),sd(met2))

0.09058205 0.11327112
```

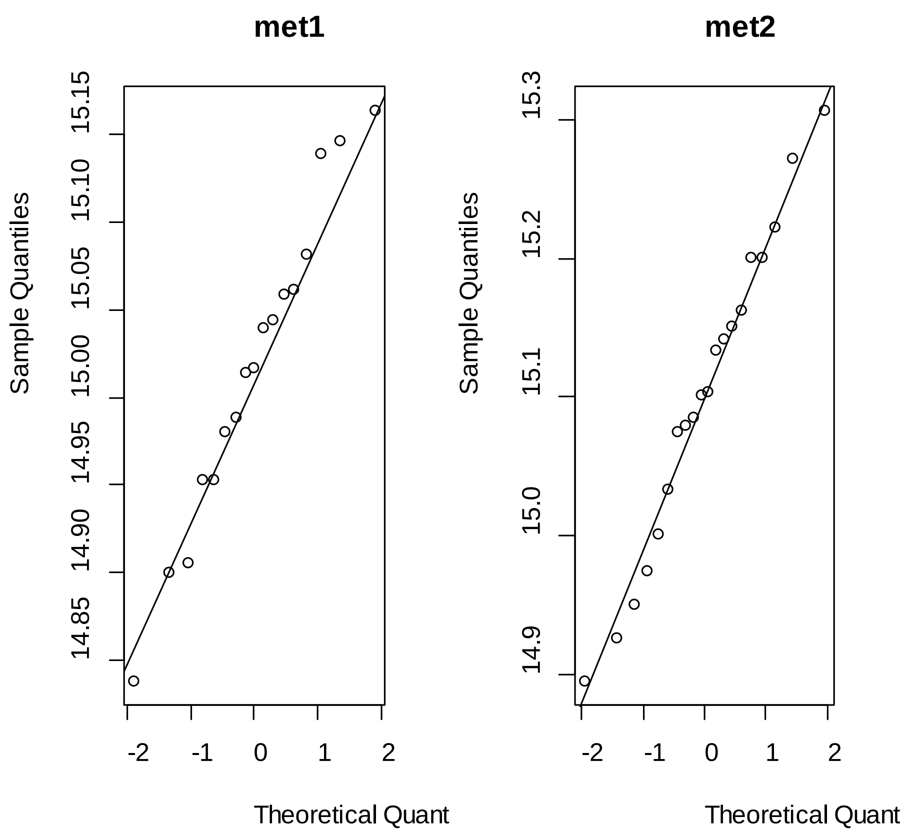
```
boxplot(met1,met2, names=c("met1","met2"))
```



El boxplot indicaría que los hay diferencias entre los dos métodos, ¿esta diferencia es significativa?

Miremos si el supuesto de normalidad parece razonable:

```
qqnorm(met1,main="met1")
qqline(met1)

qqnorm(met2,main="met2")
qqline(met2)
```



Comencemos por comparar las varianzas. Esto lo podemos realizar mediante el test de F y aplicaremos la función var.test:

```
library(stats)
var.test(met1,met2, alternative="two.sided")
```

```
F test to compare two variances

data:  met1 and met2
F = 0.6395, num df = 16, denom df = 19, p-value = 0.3707
alternative hypothesis: true ratio of variances is not equal to 1
95 percent confidence interval:
 0.246852 1.725411
sample estimates:
ratio of variances
         0.6395079
```

No se rechaza el supuesto de igualdad de varianzas, el p-valor es 0.3707.

Observación: aquí el cálculo se hizo con el s^2 más pequeño en el numerador, es decir que el cociente sabemos que va a ser menor que 1, entonces el p-valor se calcula como

```
2*pf(0.6395,16,19)
[1] 0.3706808
```

Ahora apliquemos un test t. Aplicamos el test que supone igualdad de varianzas:

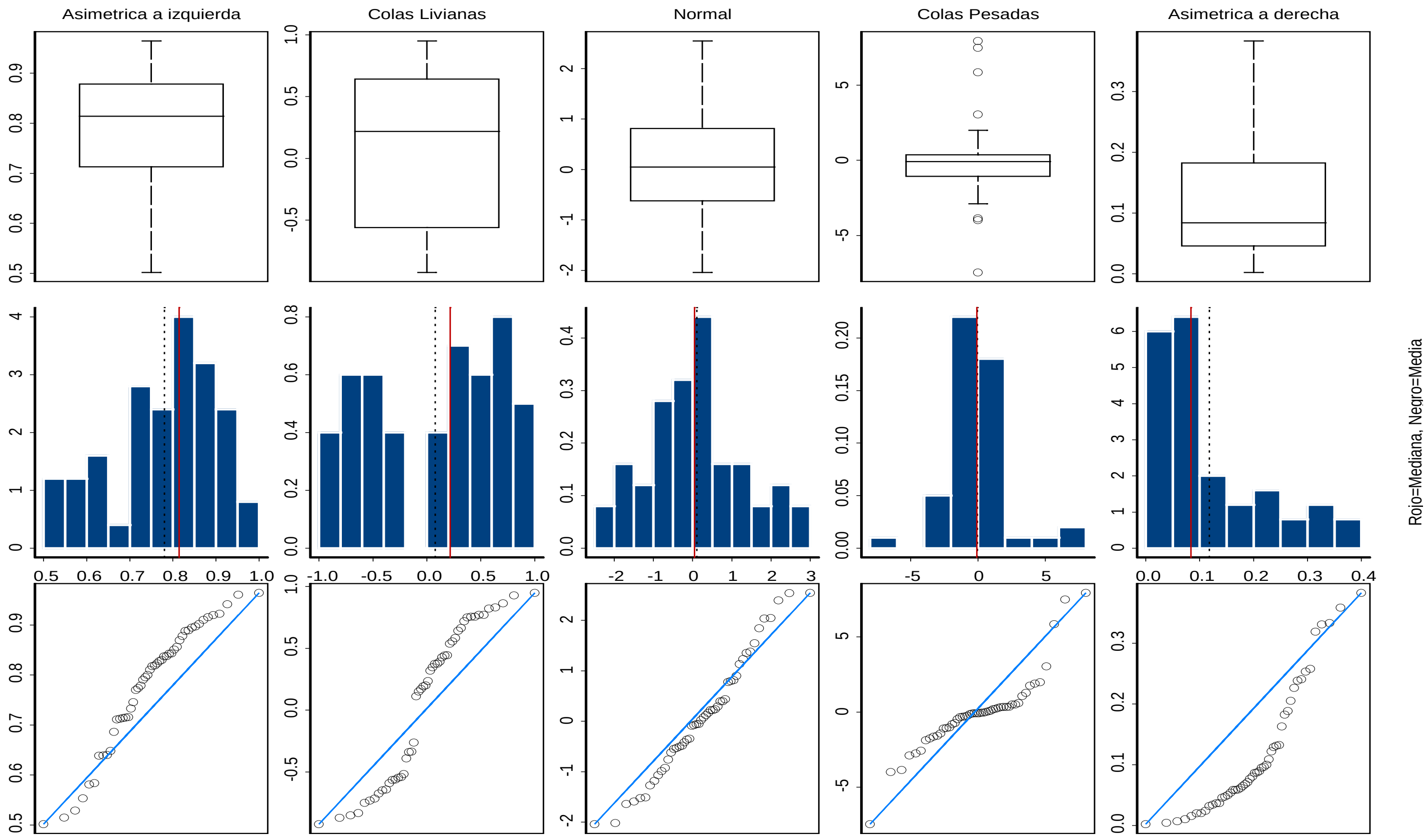
```
t.test(met1,met2,var.equal=TRUE)
       Two Sample t-test
data:  met1 and met2
t = -2.4543, df = 35, p-value = 0.01923
alternative hypothesis: true difference in means is not equal to 0
95 percent confidence interval:
 -0.15313771 -0.01448578
sample estimates:
mean of x mean of y
 15.01693    15.10075
```

Aplicamos el test basado en el estadístico de Welch:

```
t.test(met1,met2)
        Welch Two Sample t-test
data:  met1 and met2
t = -2.4997, df = 34.891, p-value = 0.01728
alternative hypothesis: true difference in means is not equal to 0
95 percent confidence interval:
 -0.15188612 -0.01573738
sample estimates:
mean of x mean of y
 15.01693  15.10075
```

Volvamos al tema de la normalidad:

Lo que hicimos fue mirar el qqplot, sin embargo esto no nos da una medida de la chance de que estemos tomando una decisión equivocada al concluir que lo datos son normales.



Test de Shapiro-Wilk

Sea $\mathcal{N}$ la familia de de todas las distribuciones $N(\mu,\sigma^2)$, con μ real ,$\sigma>0$. Consideremos una muestra aleatoria $X_1, X_2,\ldots, X_n$, donde $X_i \sim F$. Luego, las hipótesis a testear son:

$$Ho: \ F \in \mathcal{N} \quad \text{vs.} \quad H1: \ F \notin \mathcal{N}$$

Con el estadístico de test de Shapiro-Wilk y su correspondiente p-valor podemos chequear la hipótesis de normalidad y podemos rechazar el supuesto de normalidad si el p-valor que nos brinda es muy pequeño.

En general, convenimos tomar como cota un p-valor superior a 0.20.

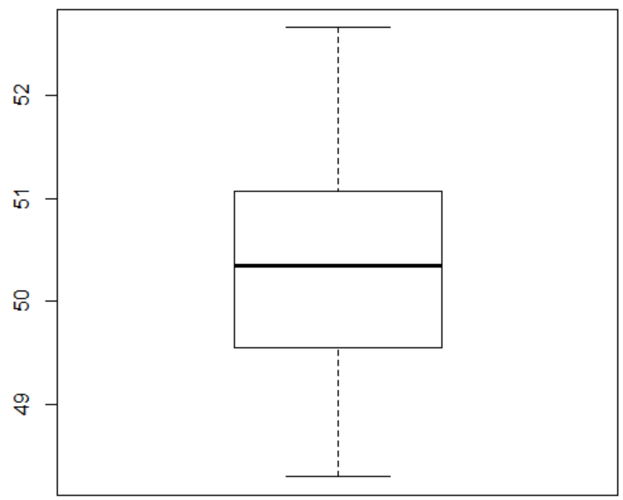
Esencialmente, lo que hace este test es medir cuán cerca de una recta está la curva que describen los puntos graficados en el QQ-plot.

```
shapiro.test(met1)

        Shapiro-Wilk normality test

data:  met1
W = 0.9771, p-value = 0.9262

shapiro.test(met2)

        Shapiro-Wilk normality test

data:  met2
W = 0.9808, p-value = 0.9439
```

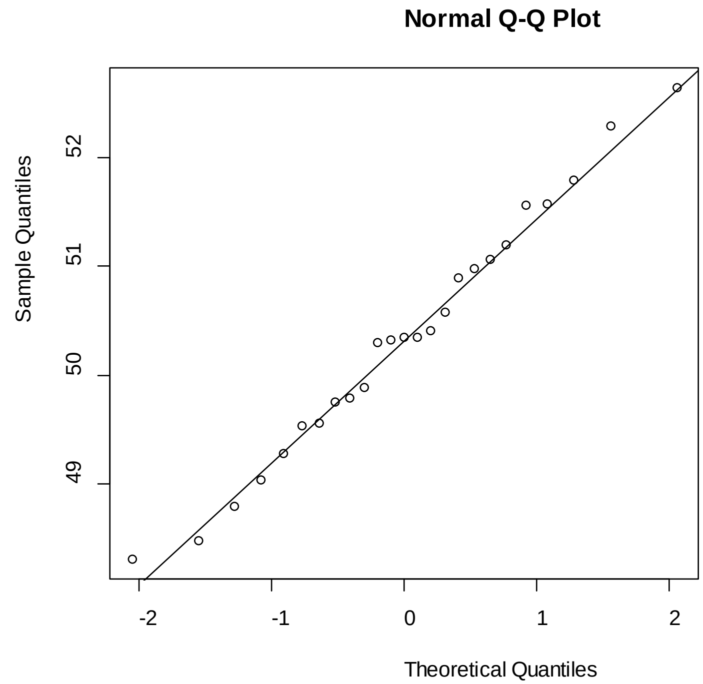
En nuestro ejemplo el p-valor del test de Shapiro-Wilk es 0.9262 en la muestra medida con el 1er. método y 0.9439 en la muestra tomada con el 2do.método, con lo cual no rechazamos el supuesto de normalidad en ninguna de ellas.

Posibles datos

```
y=scan()
49.55501  48.48144  48.79543  49.88104  49.28334  51.06887  50.57707  50.97711  50.89593  50.40595  52.65257
51.57059  52.29821  51.19372  50.29875  51.80068  49.04034  51.57614  50.34602  48.30235  49.78565  50.32727
50.35069 49.75506 49.53066

mean(y)
[1] 50.35

var(x)
[1] 1.276493
> boxplot(y)
> qqnorm(y)
> qqline(y)
```



52
51
50
49



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