Relación entre tests de hipótesis bilaterales e intervalos de confianza

Introduciremos esta relación a través de un ejemplo. Sea $X_1, X_2, \ldots, X_n$ 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:

$$H_o: \mu = \mu_o \qquad \text{vs} \qquad H_1: \mu \neq \mu_o$$

Dado que el intervalo construido contiene con alta probabilidad al valor verdadero de μ, si μ_o no pertenece al intervalo, ésto 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: $\theta = \theta_o$ vs. 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

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

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

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

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

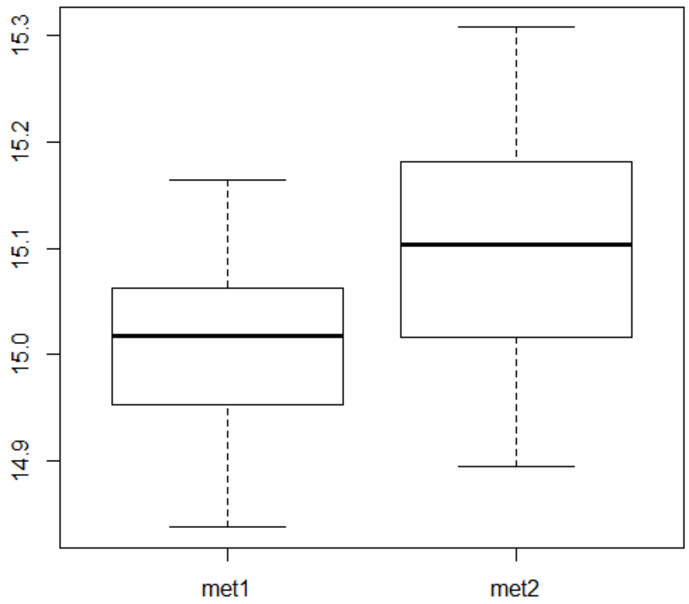
$$
\frac{s_1^2}{s_2^2} \underset{\substack{\uparrow \\ \text{bajo } H_0}}{\sim} F_{n_1-1, n_2-1}
$$

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
```

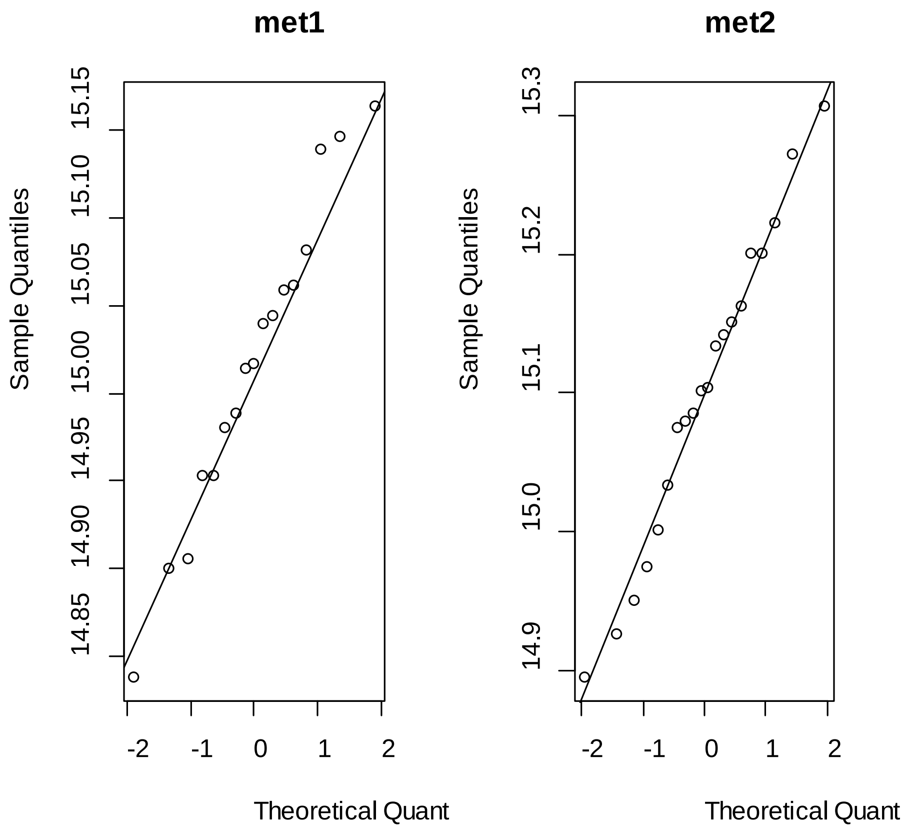
```
boxplot(met1,met2, names=c("met1","met2"))
```



El boxplot indicaría que los dos métodos no tienden a medir lo mismo, ¿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.

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

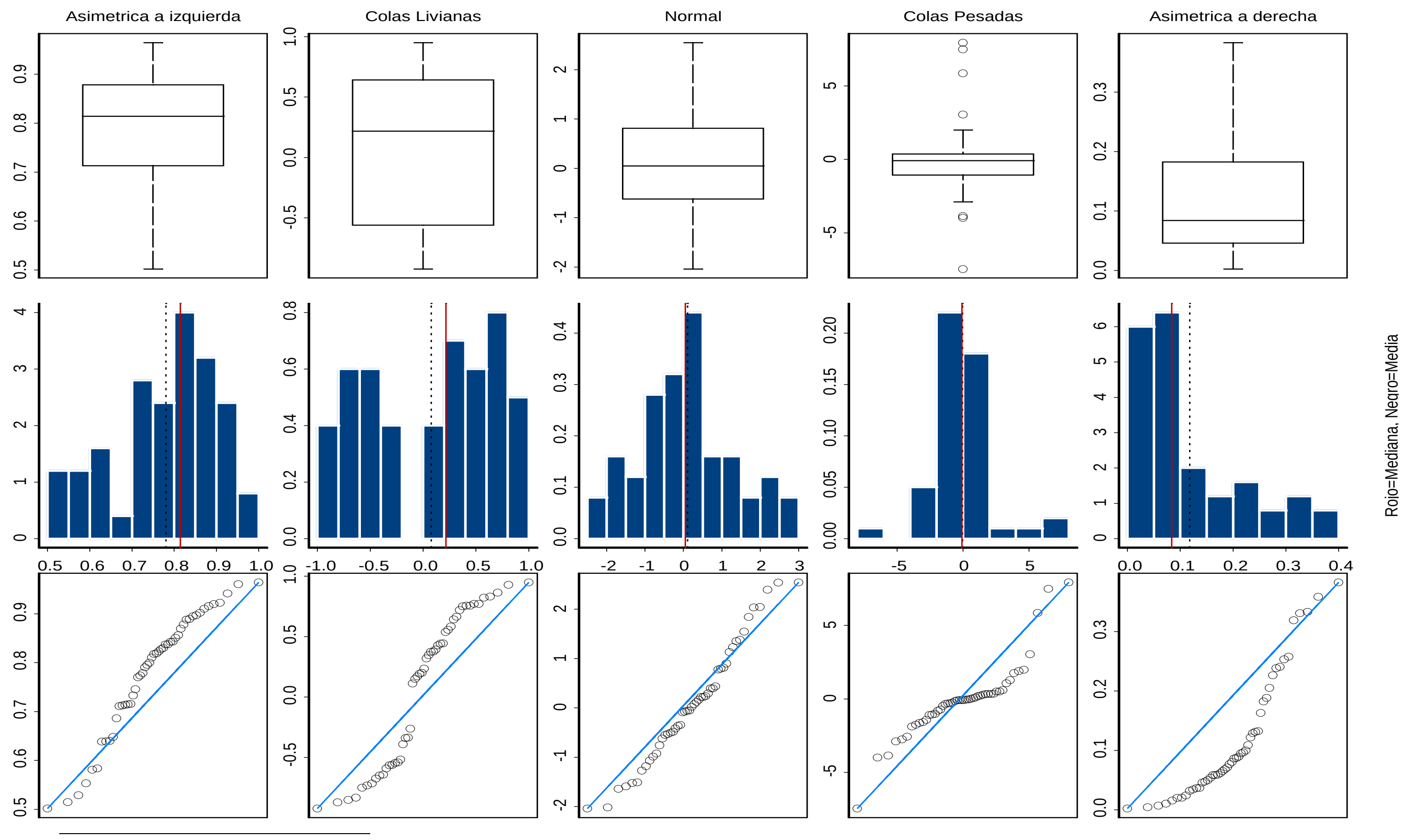
```
t.test(met1,met2,var.equal=TRUE)

       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 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(μ,σ^2), con μ real ,σ>0. Consideremos una muestra aleatoria $X_1, X_2, \ldots, X_n$, donde $X_i \sim F$. Luego, las hipótesis a testear son:

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

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.