PRÁCTICA 2: Intervalos de confianza.

DESARROLLO DE LA PRÁCTICA

El software R no dispone de un programa específico para calcular intervalos de confianza, pero si como parte de otros análisis, en particular de los test de hipótesis. Trabajaremos con el fichero "HIPER200.Rdata" que deberíamos al menos recordar su contenido.

Vamos a utilizar diversas librerías para la obtención de los intervalos de confianza clásicos (medias y varianzas del caso normal y una o dos proporciones). La librería **Desctools** nos permite la obtención de varios intervalos de confianza.

Para las variables Peso y Talla vamos a calcular sus intervalos de confianza de la media. Previamente, podemos calcular sus estadísticos descriptivos relevantes al caso

```
        peso   talla
media 67.715 161.890
        peso    talla
sd  12.81659 10.11034
```

Para obtener un intervalo de confianza de la media, basado en una distribución normal con varianza conocida o desconocida, elegimos el comando indicado para la variable que deseemos.

MeanCI(peso,sd=10.0,conf.level=0.95)
MeanCI(peso,conf.level=0.95)

Sus salidas de resultados son

```
data:  peso
   mean    lwr.ci    upr.ci
67.7150   66.3291   69.1009
```

$$\left[\bar{x} \pm z_{\frac{\alpha}{2}} \frac{\sigma}{\sqrt{n}}\right]$$

```
data:  peso
    mean     lwr.ci    upr.ci
67.71500   65.92787  69.50213
```

$$\left[\bar{x} \pm t_{n-1,\frac{\alpha}{2}} \frac{s}{\sqrt{n}}\right]$$

Si deseamos obtener un intervalo de confianza para la varianza, en el caso de ser desconocida la media, podríamos usar:

VarCI(peso, conf.level=0.95)

```
data:  peso
     var    lwr.ci    upr.ci
164.2651  136.2260  201.9992
```

$$\left[\frac{(n-1)s^2}{\chi^2_{n-1,\alpha/2}}, \frac{(n-1)s^2}{\chi^2_{n-1,1-\alpha/2}}\right]$$

Para la diferencia de medias, cuando las varianzas son desconocidas y distintas, utilizamos el commando

MeanDiffCI(peso~genero, conf.level=0.95)

sus resultados son

```
data:  peso by genero

meandiff     lwr.ci     upr.ci
 6.708350   3.266825  10.149875
```

$$\left[(\bar{x}-\bar{y}) \pm t_{f,\frac{\alpha}{2}} \sqrt{\frac{s_1^2}{n_1}+\frac{s_2^2}{n_2}} \right]$$

Cuando se tiene una proporción, si disponemos de los recuentos necesarios podemos calcular el intervalo de confianza para una proporción, según varias versiones.

BinomCI(x=18, n=106, conf.level=0.95, method="wilson")
BinomCI(x=18, n=106, conf.level=0.95, method="wald")

```
      est    lwr.ci     upr.ci
0.1698113  0.1101996  0.2525182
```

$$\left[\frac{2n\hat{p}+z_{\alpha/2}^2 \pm z_{\alpha/2}\sqrt{4n\hat{p}(1-\hat{p})+z_{\alpha/2}^2}}{2(n+z_{\alpha/2}^2)} \right]$$

```
      est     lwr.ci     upr.ci
0.1698113  0.09833423  0.2412884
```

$$\left[\hat{p} \pm z_{\frac{\alpha}{2}} \sqrt{\frac{\hat{p}(1-\hat{p})}{n}} \right]$$

Para la misma situación de una proporción pero en dos muestras independientes, si deseamos un intervalo de confianza para la diferencia de proporciones

```
BinomDiffCI(x1, n1, x2, n2, conf.level = 0.95,
            method = c("wald","waldcor","ac","exact","newcombe","newcombecor","fm","ha")
            )
```

Para nuestro caso el comando y sus resultados son

BinomDiffCI(x1=24, n1=106, x2=41, n2=96, method="wald")

```
       est     lwr.ci     upr.ci
0.20975512 0.08170177 0.33780847
```

$$\left[(\hat{p}_1-\hat{p}_2) \pm z_{\frac{\alpha}{2}} \sqrt{\frac{\hat{p}_1(1-\hat{p}_1)}{n_1}+\frac{\hat{p}_2(1-\hat{p}_2)}{n_2}} \right]$$

El intervalo de confianza para el parámetro de una distribución de Poison, puede obtenerse mediante el comando

PoissonCI(x=36, n=200, conf.level=0.98,method=c("score", "wald"))

sus resultados son

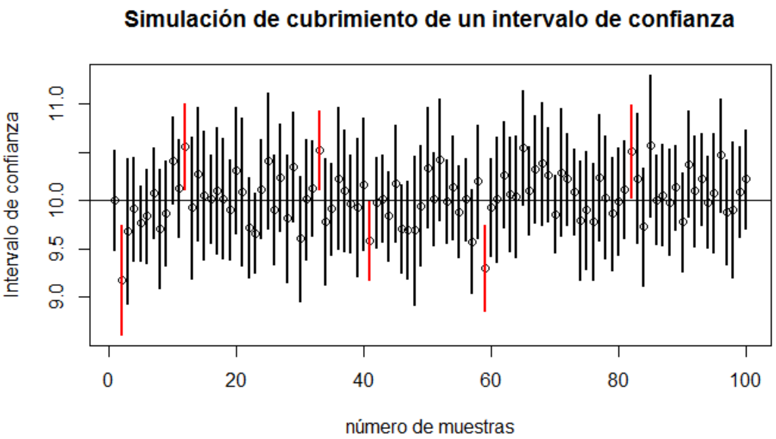
```
      est    lwr.ci    upr.ci
0.3947368 0.2183809 0.7135111
```

$$\left[(\hat{\lambda} \pm z_{\frac{\alpha}{2}} \sqrt{\frac{\hat{\lambda}}{n}} \right]$$

```
      est    lwr.ci    upr.ci
0.3947368 0.1576340 0.6318396
```

$$\left[\hat{\lambda} + \frac{z^2_{\alpha/2}}{2n} \pm \frac{z_{\alpha/2}}{2} \sqrt{(\frac{z^2_{\alpha/2}}{n} + 4\hat{\lambda})\frac{1}{n}} \right]$$

Para la finalización de la práctica, mediante varias líneas de código en R, podemos hacer varias simulaciones para realizar una interpretación frecuentista de un intervalo de confianza, en el caso de una media de una distribución normal con varianza conocida.



Aunque el libre crecimiento (más bien exponencial) del desarrollo de librerías por los autores es una cuestión deseable, no lo es tanto hace que no se disponga de una información precisa de los procedimientos que ya han sido desarrollados por otro autor. Algunas veces sentimos que perdemos el tiempo. A modo de ejemplo tenemos que la librería TeachingDemos, tiene un comando que realiza el proceso anterior.

library(TeachingDemos)
ci.examp(mean.sim = 10, sd = 1.4, n = 25, reps = 100, conf.level = 0.95)

EJERCICIOS:

1.- Para el fichero de datos “HIPER200.Rdata”, se solicita (supuesta la normalidad de los datos):
a) El intervalo de confianza al 96% para la media de la variable edad.
b) El intervalo de confianza al 97% para la diferencia de medias de la variable Tensión Arterial Sistólica inicial (TAsist0) en los grupos formados por las personas que no toman café y para las que si toman café.
c) El intervalo de confianza al 90% para la diferencia de medias de las variables Tensión Arterial Diastólica inicial (TAdias0) y Tensión Arterial Diastólica final (TAdias1).

2.- Para una proporción, realizar 1000 simulaciones de una m.a.s. de tamaño 30 de una distribución Bernoulli de parámetro 0.78, e indicar la probabilidad de cubrimiento obtenida en la simulación.

3.- Para una distribución normal de media 5.8 y de varianza 2.2, se pide realizar 1000 simulaciones de una m.a.s. de tamaño 40 de dicha distribución Normal, e indicar la probabilidad de cubrimiento obtenida por los intervalos de confianza para la media en la simulación, supuesta desconocida la varianza.

4.- Para una distribución normal de media 5.8 y de varianza 2.2, se pide realizar 2000 simulaciones de una m.a.s. de tamaño 50 de dicha distribución Normal, e indicar la probabilidad de cubrimiento obtenida por los intervalos de confianza para la varianza en la simulación, supuesta desconocida la media.

5.- Se propone el siguiente ejercicio:

- La siguiente función (copiar y ejecutar en R) permite calcular el intervalo de confianza de Wald para la proporción p de distribución binomial X~Binomial(n,p):

```
waldInterval <- function(x, n, conf.level = 0.95){
  p <- x/n
  sd <- sqrt(p*((1-p)/n))
  z <- qnorm(c( (1 - conf.level)/2, 1 - (1-conf.level)/2))
  ci <- p + z*sd
  return(ci)
}
```

- Para hallar el intervalo de confianza de p cuando X=15 y n=20:

waldInterval (x=15, n=20, conf.level=0.95)

- Suponer que n=100 y p=0.2. Generar Nsim=10,000 muestras simuladas de una variable X con distribución Binomial(n,p) (utilizar el comando **rbinom**, cuidado con los argumentos **n** de número de observaciones y **size** que es el número de ensayos de la binomial).

x<- ………

- Hallar para cada una de las Nsim muestras el intervalo de Wald al 95% de confianza y calcular la proporción de ellos que contienen el valor p=0.2, es decir, la cobertura.

```
incluye_p <- vector()
for (j in 1:Nsim) {
  # calcula el interval de confianza
   intervalo <- waldInterval(x = x[j], n = 100)

   # si p=0.2 está dentro del intervalo retorna 1, si no un 0
   incluye_p[j] <- (intervalo [1] < 0.2 & (0.2 < intervalo [2])
 }
cobertura<-sum(incluye_p)/Nsim
```

¿Es igual la cobertura al nivel de confianza 1-α? Explicar razonadamente a qué se debe la diferencia que pueda existir entre la cobertura y el nivel de confianza.

- Repetir el cálculo anterior de la cobertura para diferentes valores de p (aparte del 0.2). Por ejemplo, desde p=0.05 hasta 0.95, con saltos de 0.1 (NOTA: p<-seq(0.05, 0.95, 0.1)). Comentar cómo es ahora la cobertura del intervalo y a qué se debe las discrepancias que se puedan observar.