PRÁCTICA 11-12: Estimación puntual e intervalos de confianza.

DESARROLLO DE LA PRÁCTICA

Para obtener los estimadores de máxima verosimilitud (ML) de los parámetros de determinadas distribuciones, podemos utilizar la librería "**MASS**" y si queremos obtener un estimador ML en condiciones generales debemos recurrir a la librería "**bbmle**". Respecto a los intervalos de confianza, el software R no dispone de un programa específico (librería) pero si como parte de otros análisis, en particular de los test de hipótesis. También es posible rebuscar entre otras librerías que nos permitan realizar parte de los intervalos de confianza exclusivamente. Trabajáremos con el fichero "HIPER200.Rdata" que deberíamos al menos recordar su contenido.

ESTIMACIÓN PUNTUAL: Estimadores de máxima verosimilitud

En esta ocasión optamos por generar valores aleatorios de determinadas distribuciones y solicitar los estimadores de máxima verosimilitud

Caso1: Distribución Normal
Generamos 100 valores aleatorios de una distribución normal $N(5,2)$ y después le indicamos que nos obtenga los estimadores de ML de μ y σ . Se consigue con los comandos

```
x<-rnorm(100,5,2)
library(MASS)
fitdistr(x,"normal")
```

sus resultados son

```
     mean          sd
  4.8832049   1.9897776
 (0.1989778) (0.1406985)
```

$$\hat{\mu} = \bar{x} = \frac{1}{n}\sum_{i=1}^{n} x_i \qquad \hat{\sigma} = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(x_i - \bar{x})^2}$$

Caso2: Distribución Exponencial
Generamos 100 valores aleatorios de una distribución exponencial $\text{Exp}(0.5)$ y después le indicamos que nos obtenga el estimador de ML de λ . Se consigue con los comandos

```
x<-rexp(100,0.5)
fitdistr(x,"exponential")
```

su resultado es

```
    rate
  0.48001178
 (0.04800118)
```

$$\hat{\lambda} = \frac{1}{\bar{x}}$$

Caso3: Distribución Gamma

Generamos 100 valores aleatorios de una distribución normal $Ga(4, 0.5)$ y después le indicamos que nos obtenga los estimadores de ML de p y λ . Se obtiene con los comandos

x<-rgamma(100,4,0.5)
fitdistr(x,"gamma")

sus resultados son

```
    shape       rate
  4.1271790   0.5577373
 (0.5616698) (0.0807149)

Su resolución es mediante
un algoritmo
```

Caso4: Distribución Binomial

Generamos 100 valores aleatorios de una distribución normal $Bi(12, 0.3)$ y después le indicamos que nos obtenga el estimador de ML de p . Se logra con los comandos

x<-rbinom(100,12,0.3)
library(bbmle)
mtmp <- function(size,prob)
{-sum(dbinom(x,size,prob,log=TRUE))}
(m0 <- mle2(mtmp,start=list(prob=0.8),data=list(size=12)))

su resultado es

```
     prob
0.2925003
(0.013132)
```

$$\hat{p} = \frac{\bar{x}}{n}$$

INTERVALOS DE CONFIANZA

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.

Intervalos de Confianza para una muestra

En el caso de disponer de una variable normal $N(\mu, \sigma)$. Para la variable Peso vamos a calcular sus intervalos de confianza bajo distintos supuestos de sus parámetros. Para obtener un intervalo de confianza, basado en una distribución normal con varianza conocida o desconocida, elegimos los siguientes comandos:

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

MeanCI(peso,conf.level=0.95)

Sus resultados son, respectivamente:

```
data:  peso
   mean  lwr.ci  upr.ci
67.0200 65.6341 68.4059
```

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

```
data:  peso
   mean  lwr.ci  upr.ci
67.0200 65.3217 68.7183
```

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

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
148.3413 123.0203 182.4175
```

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

Si la media fuese conocida, no hemos encontrado un programa que realice dicho intervalo, pero en varias líneas se puede construir rápidamente

```
data:  peso
     var   lwr.ci   upr.ci
 148.64 123.3231 182.6852
```

$$\left[\frac{\sum_{i=1}^{n}(x_i-\mu)^2}{\chi^2_{n,\alpha/2}}, \frac{\sum_{i=1}^{n}(x_i-\mu)^2}{\chi^2_{n,1-\alpha/2}}\right]$$

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

BinomCI(x=40, n=100, conf.level=0.95, method="wald")

```
est    lwr.ci    upr.ci
0.4 0.3039818 0.4960182
```

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

Intervalos de Confianza para dos muestras independientes

Para la realización de estos intervalos de confianza, necesitamos dos muestras independientes, que se puede conseguir de varias formas desde un data frame o una sola m.a.s. Por ejemplo, crear dos grupos de datos de acuerdo con los valores de una variable continua

```
ed<-ifelse(edad<=40,1,2)
ed<-factor(ed, labels=c("<=40",">40"))
```

Otra posibilidad es recodificar una variable cualitativa

```
a_f<-HIPER200$act_fisi
levels(a_f)<-c("escasa", "moderada","moderada")
```

para que solo tenga dos modalidades.
Una tercera opción, consiste en seleccionar casos de acuerdo a una variable diferente, por ejemplo

```
HIPER200_25 <- subset(HIPER200, subset=edad > 25, select=c(TAsist,genero))
```

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

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

sus resultados son

```
data:  peso by genero

meandiff    lwr.ci    upr.ci
 7.000000  3.738674 10.261326
```

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

Para la misma situación de una proporción, pero en dos muestras independientes, si deseamos un intervalo de confianza para la diferencia de proporciones. Para nuestro caso el comando y sus resultados son

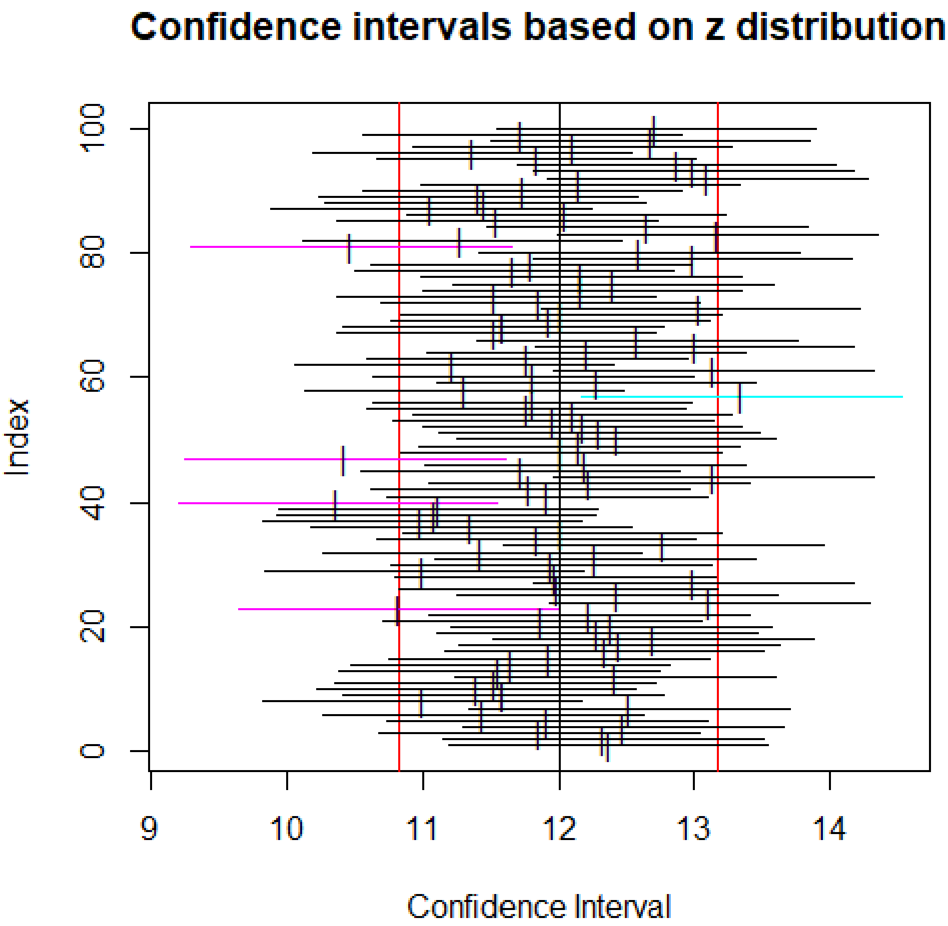
```
BinomDiffCI(x1=40, n1=100, x2=89, n2=100, method="wald")
```

```
      est     lwr.ci     upr.ci
0.4900000  0.3760689  0.6039311
```

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

Para la finalización de la práctica, podemos llevar a cabo varias simulaciones para realizar una interpretación frecuentista de un intervalo de confianza, para el caso de una media de una distribución normal con varianza conocida.

library(TeachingDemos)
ci.examp(mean.sim = 12, sd = 3, n = 25, reps = 100, conf.level = 0.95)



EJERCICIOS:

1.- En una gran ciudad, para una muestra de 500 familias con televisión se comprobó que 280 habían contratado cierto canal privado de televisión. Determinar un intervalo de confianza del 95% para estimar la proporción real de familias suscritas a dicho canal

2.- Se tiene que, en una muestra de 200 votantes de un distrito, 132 votaron al candidato A mientras que, de una muestra de 150 votantes de otro distrito, 90 votaron por el mismo candidato. Determinar un intervalo de confianza del 95% para la diferencia de proporciones de votos a favor del candidato A en los dos distritos.

3.- Para una distribución normal de media 5.8 y de varianza 2.2, se pide realizar 200 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.

4. Un informe reciente revela, a partir de una muestra de tamaño 900, que el 65% de las familias de esta región tienen dificultades para llegar a final de mes. Construye un intervalo para la verdadera proporción, al 95% de confianza.

5.- 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 TA sistólica final.
b) El intervalo de confianza al 97% para la media de la variable TA sistólica final siempre que tenga más de 50 años.
c) El intervalo de confianza al 98% para la media de la variable TA sistólica final siempre tena más de 50 años y sea mujer.
d) El intervalo de confianza al 97% para la diferencia de medias de la variable TA sistólica final en los grupos formados por las personas que no toman café y para las que si toman café.