Intervalo de confianza 2

J. Abellán

22/10/2017

Intervalo de confianza II.

¿Cómo se construye el *intervalo de confianza* para la media poblacional μ cuando no conocemos la varianza poblacional σ^2 ?

Parece obvio que habrá que utilizar la varianza muestral s^2. Ahora bien, en este caso se utiliza la función de distribución *t-student* y no la *normal* porque el teorema dice:

$$T \equiv \frac{\overline{X} - \mu}{\frac{S}{\sqrt{n}}} \sim t_{n-1}$$

Ahora el *intervalo de confianza* varía de muestra a muestra puesto que depende de la desviación muestral S. También dependen los valores $z_{\alpha/2,n-1}$ del número de medidas n. Por ejemplo, para $n = 5$, tenemos:

```
# Número de medidas
n <- 5

# Valores de z
z <- c( 1, 2, 3, 4 )

# Probabilidades según t-student
alfa <- 2 * pt( - z, n - 1 )

cbind( z, alfa, 1 - alfa )
```

```
##      z       alfa
## [1,] 1 0.37390097 0.6260990
## [2,] 2 0.11611652 0.8838835
## [3,] 3 0.03994197 0.9600580
## [4,] 4 0.01613009 0.9838699
```

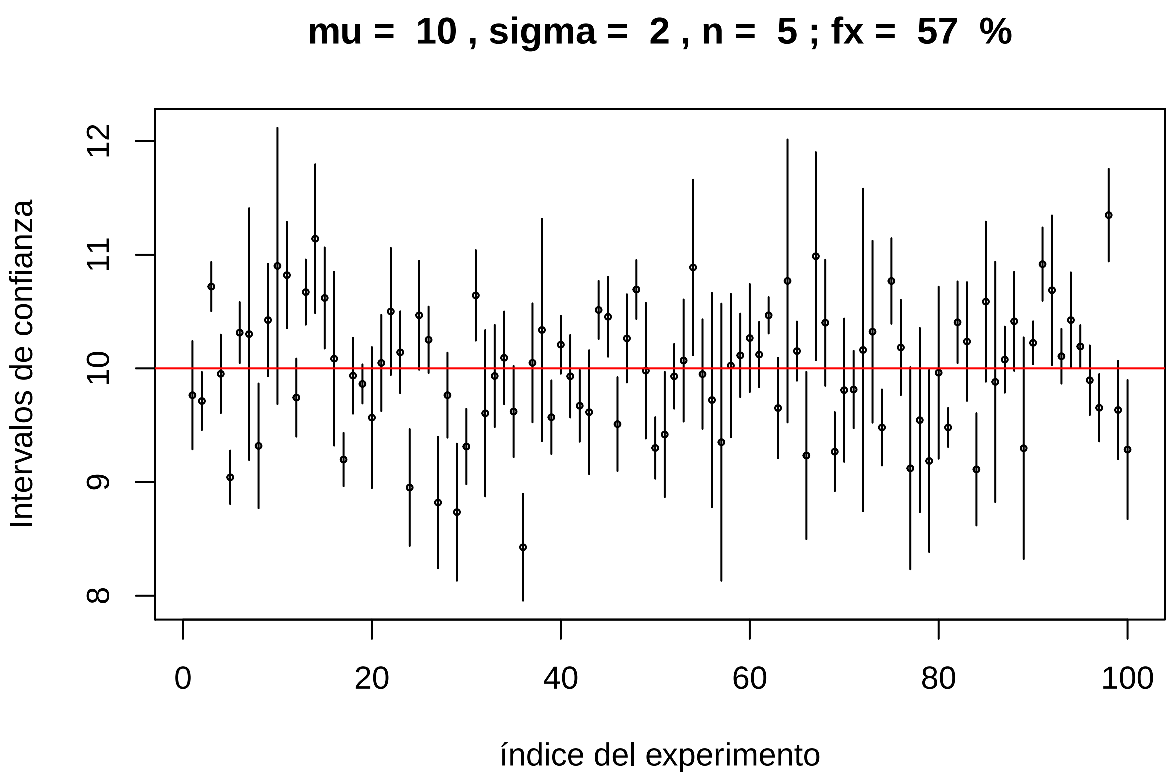
Como se ve, para $n = 5$ y $z = 1$ tenemos que $1 - \alpha = 0.626$ y no 0.68 como antes.

Definimos la función *intervalot*, que usa la *t-student* para generar los números aleatorios.

```
intervalot <- function( mu, sigma, n ) {

  # N grande para una buena estadística
  N <- 100

  #La nueva variable aleatoria: media muestral
  Xm <- S <- rep( 0, N )

  #Hacemos el experimento. Variable aleatoria Xm, media muestral
  mu0 <- rep( mu, n )

  for (i in 1 : N ) {

    # generamos n datos según una t-student con n-1 grados de libertad
```

```r
  # que sumamos a la línea base
  muestra <- mu0 + rt( n, n - 1 )

  # Desviación estándar de la muestra
  S[ i ] <- sd( muestra )

  # Media de la muestra
  Xm[ i ] <- mean( muestra )

}

# Los intervalos de confianza (son distintos entre sí)
IC <- S / sqrt( n )

#Variable aleatoria límite inferior del intervalo
Li <- Xm - IC

#Variable aleatoria límite superior del intervalo
Ls <- Xm + IC

# para el dibujo
y1 <- min( Li ) ; y2 <- max( Ls )

#Dibujamos las medias muestrales obtenidas
plot( Xm,

      cex = .4,

      ylim = c( y1, y2 ),

      xlab = " índice del experimento ",

      ylab = " Intervalos de confianza "

    )

#y los intervalos de confianza IC
i <- 1 : N

segments( i, Li, i, Ls )

#Dibujamos el valor verdadero (conocido por el ordenador)
abline( h = mu, col = 2 )

# Número de intervalos por encima de mu
nIa <- length( Li[ Li > mu ] )

# Número de intervalos por debajo de mu
nId <- length( Ls[ Ls < mu ] )

# Número de intervalos que contienen a mu
nI <- N - ( nIa + nId )
```

```
  # Fracción de intervalos (probabilidad) que contienen a mu
  fx <- nI / N

  title( paste( " mu = ", mu,

                ", sigma = ",sigma,

                ", n = ",n,

                "; fx = ", round( fx * 100 ), " % " )
      )

  return( fx )

  }

# Valores de los parámetros
mu <- 10

sigma <- 2

# Número de medidas por muestra
n <- 5

nx <- intervalot( mu, sigma, n )
```



El número de intervalos que contienen la media poblacional μ es 0.57, muy próximo al valor 0.63 teórico.

También se puede ver que no hay correlación entre los valores de las medias muestrales y la magnitud de los intervalos. Es decir, las variables aleatorias X_m y S son independientes.