7. Intervalos de confianza

Juan Bernal

2024-08-21

Problema 1

Muestra que el nivel de confianza indica el porcentaje de intervalos de confianza extraídos de una misma población que contienen a la verdadera media a través de la simulación de intervalos:

A. Haz la simulación de 150 muestras de tamaño 150 extraídas de una población normal con $\mu = 70$ y $\sigma = 9$

```
n = 150 # Número de datos
media = 70 # Media poblacional
de = 9 # Desviación estándar poblacional
alfa1 = 0.03 # Nivel de confianza
ErrorEst = de/sqrt(n) # Media muestral
X_ = rnorm(n,media,ErrorEst) #
E = abs(qnorm(alfa1/2))*ErrorEst
```

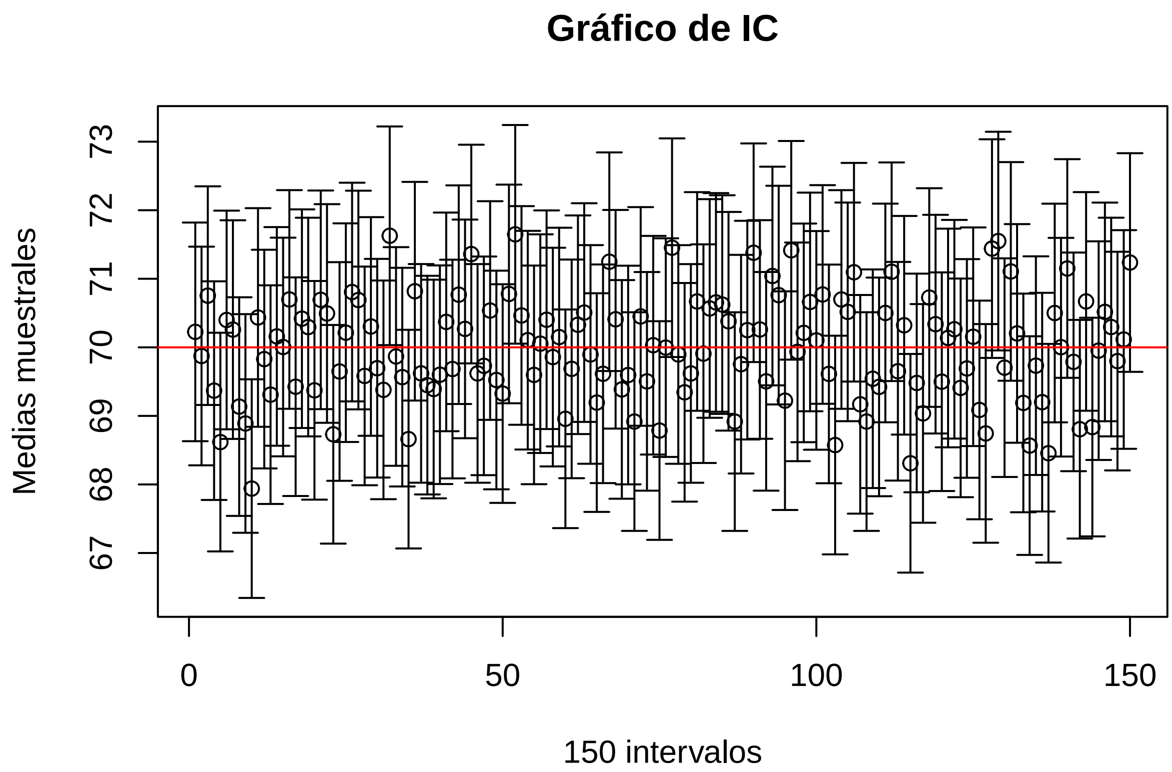
B. Calcula el intervalo con un nivel de confianza del 97% para cada una de esas medias. Obtendrás 150 intervalos de confianza.

```
ici = X_ - E # Cota inferior
ics = X_ + E # Cota superior
```

C. Grafica los 150 intervalos de confianza

D. Grafica la media poblacional ($\mu = 70$) como una linea horizontal

```
library(plotrix)
plotCI(1:n,X_,E,main="Gráfico de IC", , xlab="150 intervalos", ylab="Medias muestrales")
abline(h = media, col='red')
```



E. Cuenta cuántos intervalos de confianza contienen a la verdadera media, ¿qué porcentaje representan?

Hay 145 intervalos que contienen la verdadera media, mientras que hay 5 intervalos que o son superiores o inferiores a la media, que ni el intervalo la contiene. Entonces el 97% de los intervalos de confianza contienen a la verdadera media.

Problema 2

Resuelve las dos partes del problema "El misterioso Helio".

Primera parte. Suponga que la porosidad al helio (en porcentaje) de muestras de carbón, tomadas de cualquier veta en particular, está normalmente distribuida con una desviación estándar verdadera de 0.75. Se sabe que 10 años atrás la porosidad media de helio en la veta era de 5.3 y se tiene interés en saber si actualmente ha disminuido. Se toma una muestra al azar de 20 especímenes y su promedio resulta de 4.85.

X: Porosidad del helio

$X \sim N(\mu =?, \sigma = 0.75)$

A. Haga una estimación por intervalo con una confianza del 97% para el promedio de porosidad para evaluar si ha disminuido.

```
de = 0.75 # Desviación estándar
alfa1 = 0.03 # Nivel de confianza del 97%
x_med1 = 4.85 # Media muestral de 20 especímenes
```

```
n = 20 # Número de especímenes

E1 = abs(qnorm(alfa1/2))*de/sqrt(n) # Error del 97% confianza
A1 = x_med1 - E1 # Cota inferior
B1 = x_med1 + E1 # Cota superior

cat('La verdadera media actual está entre [',A1,', ',B1,']')
```

```
## La verdadera media actual está entre [ 4.486065 ,  5.213935 ]
```

B. Se toma otra muestra de tamaño 16. El promedio de la muestra fue de 4.56. Calcule el intervalo de confianza al 97% de confianza

```
n = 16 # Número de especímenes
x_med2 = 4.56 # Media muestral de 16 especímenes

E2 = abs(qnorm(alfa1/2))*de/sqrt(n) # Error del 97% confianza
A2 = x_med2 - E2 # Cota inferior
B2 = x_med2 + E2 # Cota superior

cat('La verdadera media actual está entre [',A2,', ',B2,']')
```

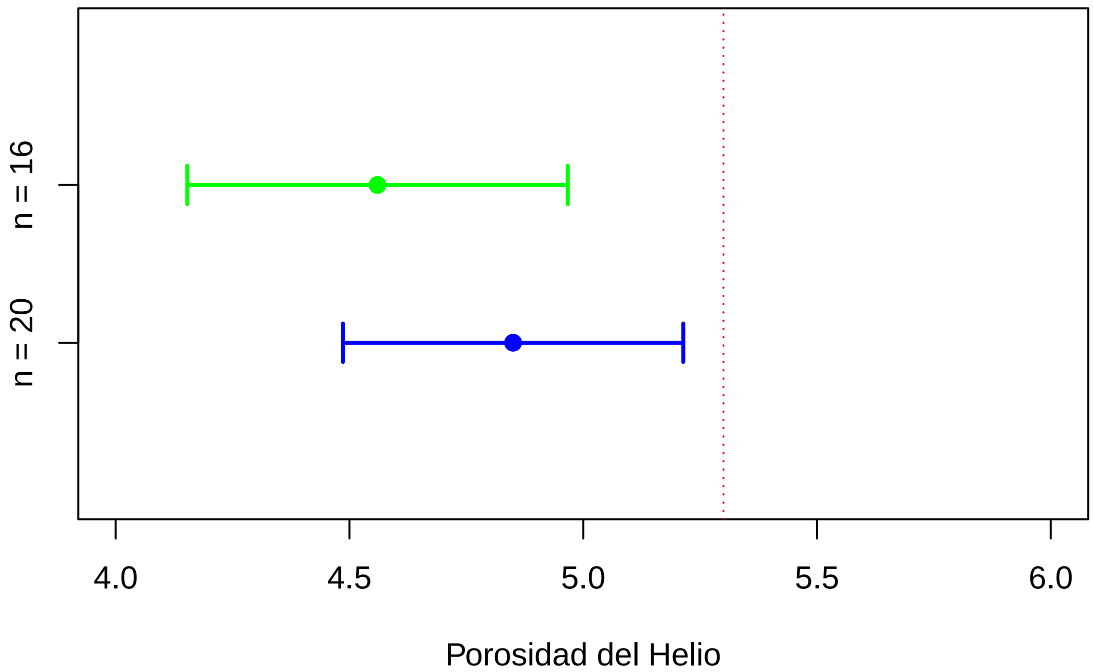
```
## La verdadera media actual está entre [ 4.153108 ,  4.966892 ]
```

C. ¿Podemos afirmar que la porosidad del helio ha disminuido?

```
plot(0, ylim = c(0,2+1), xlim = c(4,6), yaxt = "n", ylab = "", xlab = "Porosidad del Helio")
axis(2, at = c(1,2), labels = c("n = 20", "n = 16"))

arrows(A1, 1, B1, 1, angle=90, code=3, length = 0.1, lwd = 2, col = 'blue')
arrows(A2, 2, B2, 2, angle=90, code=3, length = 0.1, lwd = 2, col = 'green')

points(x_med1, 1, pch = 19, cex = 1.1, col = 'blue')
points(x_med2, 2, pch = 19, cex = 1.1, col = 'green')
abline(v = 5.3, lty = 3, col = 'red')
```



Sabemos con un 97% de confianza que la porosidad del helio ha disminuido, pues en los gráficos podemos notar que la media verdadera anterior es mayor que las medias muestrales de tamaño 20 y 16.

Segunda parte. Suponga que la porosidad al helio (en porcentaje) de muestras de carbón, tomadas de cualquier veta en particular, está normalmente distribuida con una desviación estándar verdadera de 0.75.

A. ¿Qué tan grande tiene que ser el tamaño de la muestra si se desea que el ancho del intervalo con un 95% de confianza no sobrepase de 0.4?

Para calcular el tamaño que debe tener la muestra para que el intervalo de 95% de confianza no sobrepase 0.4, debemos despejar la fórmula que nos da el error de confianzal, es decir:

$E = z_{\frac{\alpha}{2}} \frac{\sigma_x}{\sqrt{n}}$

Si despejamos n que es el número de muestras, obtendremos:

$n = (z_{\frac{\alpha}{2}} \frac{\sigma_x}{E})^2$

Y como contamos con $\sigma = 0.75$, $\alpha = 0.05$ y $E = 0.2$ debido a que 0.4 es el intervalo completo, finalmente obtendremos la expresión:

$n = (z_{\frac{0.05}{2}} \frac{0.75}{0.2})^2$

```
n = (abs(qnorm(0.05/2))*0.75/0.2)^2
n
```

```
## [1] 54.02051
```

El tamaño de la muestra debe de ser mínimo de 55, para que no se sobrepase el ancho de 0.4 en un intervalo con un 95% de confianza.

B. ¿Qué tamaño de muestra necesita para estimar la porosidad promedio verdadera dentro de 0.2 unidades alrededor de la media muestral con una confianza de 99%?

Usaremos la misma ecuación despejada para el número de muestras mínimo, pero ahora con una confianza del 99%, es decir, $\alpha = 0.01$:

$n = (z_{\frac{0.01}{2}} \frac{0.75}{0.2})^2$

```
n = (abs(qnorm(0.01/2))*0.75/0.2)^2
n
```

```
## [1] 93.30323
```

El tamaño de la muestra debe de ser mínimo de 94, para poder estimar la porosidad promedio verdadera dentro de 0.2 unidades alrededor de la media muestral con un 99% de confianza.

Problema 3.

Con el archivo de datos de El Marcapasos haz los intervalos de confianza para la media de las siguientes variables:

```
data = read.csv('El marcapasos.csv') # Lectura de la base de datos
head(data) # Primeras observaciones del dataset
```

```
##   Periodo.entre.pulsos Intensidad.de.pulso Marcapasos
## 1                  1.2               0.131     Sin MP
## 2                  0.9               0.303     Sin MP
## 3                  0.9               0.297     Sin MP
## 4                  0.8               0.416     Sin MP
## 5                  0.7               0.585     Sin MP
## 6                  1.2               0.126     Sin MP
```

A. Intensidad de pulsos con y sin Marcapasos (2 intervalos de confianza)

```
miu = mean(data$Intensidad.de.pulso) # Media de la intensidad de pulsos
de = sd(data$Intensidad.de.pulso) # Desviación estándar de la intensidad de pulsos
alfa2 = 0.05 # Nivel de confianza

Ints = data$Intensidad.de.pulso[data$Marcapasos == 'Sin MP'] # Intensidad de pulsos sin marcapasos
Int = data$Intensidad.de.pulso[data$Marcapasos == 'Con MP'] # Intensidad de Pulsos con marcapasos

med_ints = mean(Ints) # Media de la intensidad de pulsos sin marcapasos
med_int = mean(Int) # Media de la intensidad de pulsos con marcapasos

E_ints = abs(qnorm(alfa2/2))*de/sqrt(length(Ints)) # Error del 95% de confianza para Sin MP
E_int = abs(qnorm(alfa2/2))*de/sqrt(length(Int)) # Error del 95% de confianza para Con MP

A_ints =  med_ints - E_ints # Cota inferior para intensidad de pulsos sin MP
B_ints =  med_ints + E_ints # Cota superior para intensidad de pulsos sin MP

A_int = med_int - E_int # Cota inferior para intensidad de pulsos con MP
B_int = med_int + E_int # Cota superior para intensidad de pulsos con MP

cat('La verdadera media actual para intensidad de pulsos sin marcapasos está entre [',A_ints,', ',B_ints
```

```
## La verdadera media actual para intensidad de pulsos sin marcapasos está entre [ 0.1733278 ,  0.24086
```

```
cat('La verdadera media actual para intensidad de pulsos sin marcapasos está entre [',A_int,', ',B_int,
```

```
## La verdadera media actual para intensidad de pulsos sin marcapasos está entre [ 0.1621709 ,  0.22971
```

```r
cat('La verdadera media es ',mean(data$Intensidad.de.pulso))
```

```
## La verdadera media es  0.2015196
```

B. Periodo entre pulso con y sin Marcapasos (2 intervalos de confianza)

```r
miu2 = mean(data$Periodo.entre.pulsos) # Media del periodo entre pulsos
de2 = sd(data$Periodo.entre.pulsos) # Desviación estándar del periodo entre pulsos
alfa2 = 0.05 # Nivel de confianza

Pers = data$Periodo.entre.pulsos[data$Marcapasos == 'Sin MP'] # Periodo entre pulsos sin marcapasos
Per = data$Periodo.entre.pulsos[data$Marcapasos == 'Con MP'] # Periodo entre Pulsos con marcapasos

med_pers = mean(Pers) # Media del periodo entre pulsos sin marcapasos
med_per = mean(Per) # Media del periodo entre pulsos con marcapasos

E_pers = abs(qnorm(alfa2/2))*de2/sqrt(length(Pers)) # Error del 95% de confianza para Sin MP
E_per = abs(qnorm(alfa2/2))*de2/sqrt(length(Per)) # Error del 95% de confianza para Con MP

A_pers =  med_pers - E_pers # Cota inferior para periodo entre pulsos sin MP
B_pers =  med_pers + E_pers # Cota superior para periodo entre pulsos sin MP

A_per = med_per - E_per # Cota inferior para periodo entre pulsos con MP
B_per = med_per + E_per # Cota superior para periodo entre pulsos con MP

cat('La verdadera media actual para periodo de pulsos sin marcapasos está entre [',A_pers,', ',B_pers,']
```

```
## La verdadera media actual para periodo de pulsos sin marcapasos está entre [ 1.028899 ,  1.194631 ]
```

```r
cat('La verdadera media actual para periodo de pulsos sin marcapasos está entre [',A_per,', ',B_per,'] 
```

```
## La verdadera media actual para periodo de pulsos sin marcapasos está entre [ 0.8083103 ,  0.9740427 ]
```

```r
cat('La verdadera media es ',mean(data$Periodo.entre.pulsos))
```

```
## La verdadera media es  1.001471
```

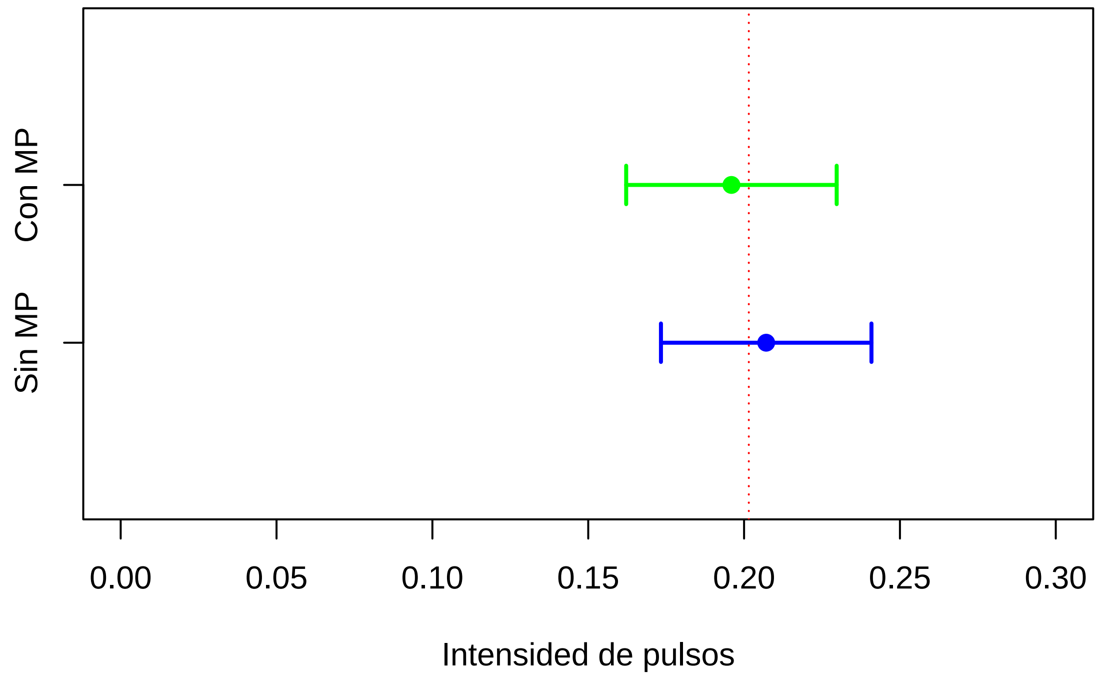
Grafica los intervalos de confianza obtenidos en "El marcapasos":

A. Grafica en un mismo eje coordenado la intensidad de pulso con y sin marcapasos

```r
plot(0, ylim = c(0,2+1), xlim = c(0,0.3), yaxt = "n", ylab = "", xlab = "Intensided de pulsos")
axis(2, at = c(1,2), labels = c("Sin MP", "Con MP"))

arrows(A_ints, 1, B_ints, 1, angle=90, code=3, length = 0.1, lwd = 2, col = 'blue') # Sin marcapasos
arrows(A_int, 2, B_int, 2, angle=90, code=3, length = 0.1, lwd = 2, col = 'green') # Con marcapasos

points(med_ints, 1, pch = 19, cex = 1.1, col = 'blue')
points(med_int, 2, pch = 19, cex = 1.1, col = 'green')
abline(v = miu, lty = 3, col = 'red')
```

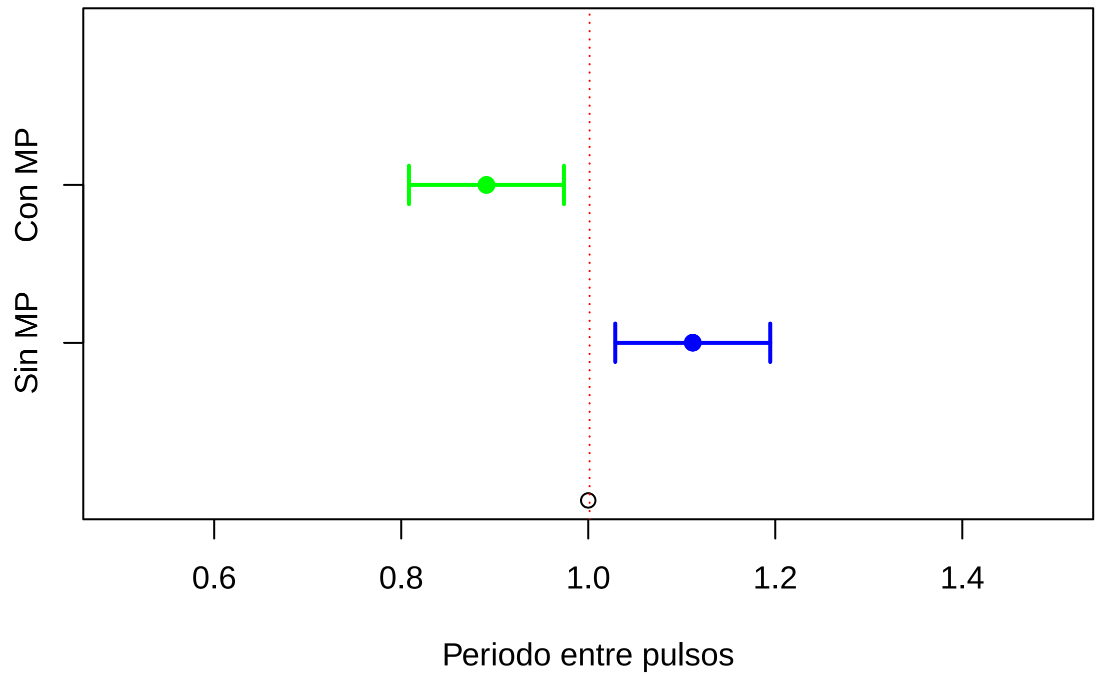


B. Grafica en un mismo eje coordenado el periodo entre pulso con y sin marcapasos

```
plot(0, ylim = c(0,2+1), xlim = c(0.5,1.5), yaxt = "n", ylab = "", xlab = "Periodo entre pulsos")
axis(2, at = c(1,2), labels = c("Sin MP", "Con MP"))

arrows(A_pers, 1, B_pers, 1, angle=90, code=3, length = 0.1, lwd = 2, col = 'blue') # Sin marcapasos
arrows(A_per, 2, B_per, 2, angle=90, code=3, length = 0.1, lwd = 2, col = 'green') # Con marcapasos

points(med_pers, 1, pch = 19, cex = 1.1, col = 'blue')
points(med_per, 2, pch = 19, cex = 1.1, col = 'green')
abline(v = miu2, lty = 3, col = 'red')
```



Compara los intervalos obtenidos e interpreta los gráficos. Concluye sobre ambas variables en la presencia y ausencia de marcapasos

Para el caso de "Intensidad de pulsos" tanto la condición "Con MP" como "Sin MP" tienen intervalos de confianza del 95% que contienen la media poblacional. Esto indica que no hay una diferencia significativa entre las medias de estas dos condiciones y la media poblacional en términos de la intensidad de pulsos. Es decir que la intensidad de los pulsos con o sin marcapasos se encuentran dentro de la media poblacional.

Y para el caso del "Periodo entre pulsos" ninguno de los intervalos de confianza del 95% para las condiciones "Con MP" y "Sin MP" contiene la media poblacional. Esto implica que hay una diferencia significativa entre las medias de estas condiciones y la media poblacional respecto al período entre pulsos. La condición "Con MP" se acerca más a la media poblacional que "Sin MP", pero dado que ninguno de los intervalos incluye la media poblacional, se sugiere que ambas condiciones difieren significativamente de lo que se espera a nivel poblacional en cuanto al período entre pulsos.

Los resultados obtenidos para cada variable sugieren que la presencia del marcapasos se vuelve un catalizador para personas con problemas del corazón, pues los mantiene en la media de intensidad de pulsos y los acerca a la media de periodo entre pulsos.