Act7 Intervalos de confianza

Luis Maximiliano López Ramírez

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 miu = 70 y sigma = 9

```r
# Parámetros de la población
miu <- 70
sigma <- 9
n_muestras <- 150
tamaño_muestra <- 150
alfa <- 0.05
Z_alfa_2 <- abs(qnorm(alfa / 2))

# Simulación
set.seed(123)  # Para reproducibilidad
dentro <- 0

for (i in 1:n_muestras) {
  muestra <- rnorm(tamaño_muestra, mean = miu, sd = sigma)
  media_muestra <- mean(muestra)
  error_est <- Z_alfa_2 * sigma / sqrt(tamaño_muestra)

  # Construir intervalo de confianza
  A <- media_muestra - error_est
  B <- media_muestra + error_est

  # Verificar si la verdadera media está dentro del intervalo
  if (miu >= A & miu <= B) {
    dentro <- dentro + 1
  }
}

# Proporción de intervalos que contienen la verdadera media
nivel_confianza_observado <- dentro / n_muestras

cat("El nivel de confianza observado es:", nivel_confianza_observado)

## El nivel de confianza observado es: 0.96
```

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

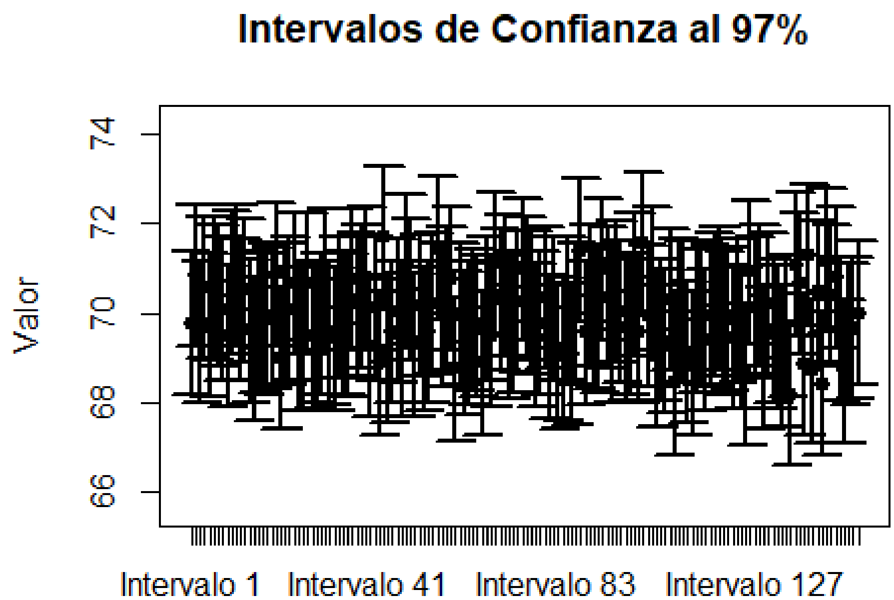
```
# Parámetros de la población
miu <- 70
sigma <- 9
n_muestras <- 150
tamaño_muestra <- 150
alfa <- 0.03  # Nivel de significancia para un nivel de confianza del 97%
Z_alfa_2 <- abs(qnorm(alfa / 2))

# Simulación
set.seed(123)  # Para reproducibilidad
intervalos <- matrix(NA, nrow = n_muestras, ncol = 2)  # Para almacenar
los intervalos

for (i in 1:n_muestras) {
  muestra <- rnorm(tamaño_muestra, mean = miu, sd = sigma)
  media_muestra <- mean(muestra)
  error_est <- Z_alfa_2 * sigma / sqrt(tamaño_muestra)

  # Calcular el intervalo de confianza
  A <- media_muestra - error_est
  B <- media_muestra + error_est

  # Guardar el intervalo en la matriz
  intervalos[i, ] <- c(A, B)
}

# Mostrar los primeros 5 intervalos como ejemplo
print(intervalos[1:5, ])

##          [,1]     [,2]
## [1,] 68.18605 71.37542
## [2,] 69.24453 72.43390
## [3,] 67.99116 71.18053
## [4,] 68.98344 72.17280
## [5,] 68.14462 71.33398

# El número total de intervalos calculados
cat("Se han calculado", n_muestras, "intervalos de confianza con un nivel
del 97%.")

## Se han calculado 150 intervalos de confianza con un nivel del 97%.
```

C. Grafica los 150 intervalos de confianza

```
# Parámetros de la población
miu <- 70
sigma <- 9
n_muestras <- 150
```

```r
tamaño_muestra <- 150
alfa <- 0.03  # Nivel de significancia para un nivel de confianza del 97%
Z_alfa_2 <- abs(qnorm(alfa / 2))

# Simulación
set.seed(123)  # Para reproducibilidad
intervalos <- matrix(NA, nrow = n_muestras, ncol = 2)
medias <- numeric(n_muestras)

for (i in 1:n_muestras) {
  muestra <- rnorm(tamaño_muestra, mean = miu, sd = sigma)
  media_muestra <- mean(muestra)
  error_est <- Z_alfa_2 * sigma / sqrt(tamaño_muestra)

  # Calcular el intervalo de confianza
  A <- media_muestra - error_est
  B <- media_muestra + error_est

  # Guardar el intervalo y la media
  intervalos[i, ] <- c(A, B)
  medias[i] <- media_muestra
}

# Establecer los límites del gráfico
min_y <- min(intervalos) - 1  # Un poco más pequeño que el valor mínimo
del intervalo
max_y <- max(intervalos) + 1  # Un poco más grande que el valor máximo
del intervalo

# Crear el gráfico con ejes intercambiados
plot(0, xlim=c(0, n_muestras + 1), ylim=c(min_y, max_y), xaxt="n",
ylab="Valor", xlab="", main="Intervalos de Confianza al 97%", horiz =
TRUE)

## Warning in plot.window(...): "horiz" is not a graphical parameter

## Warning in plot.xy(xy, type, ...): "horiz" is not a graphical
parameter

## Warning in axis(side = side, at = at, labels = labels, ...): "horiz"
is not a
## graphical parameter

## Warning in axis(side = side, at = at, labels = labels, ...): "horiz"
is not a
## graphical parameter

## Warning in box(...): "horiz" is not a graphical parameter

## Warning in title(...): "horiz" is not a graphical parameter
```

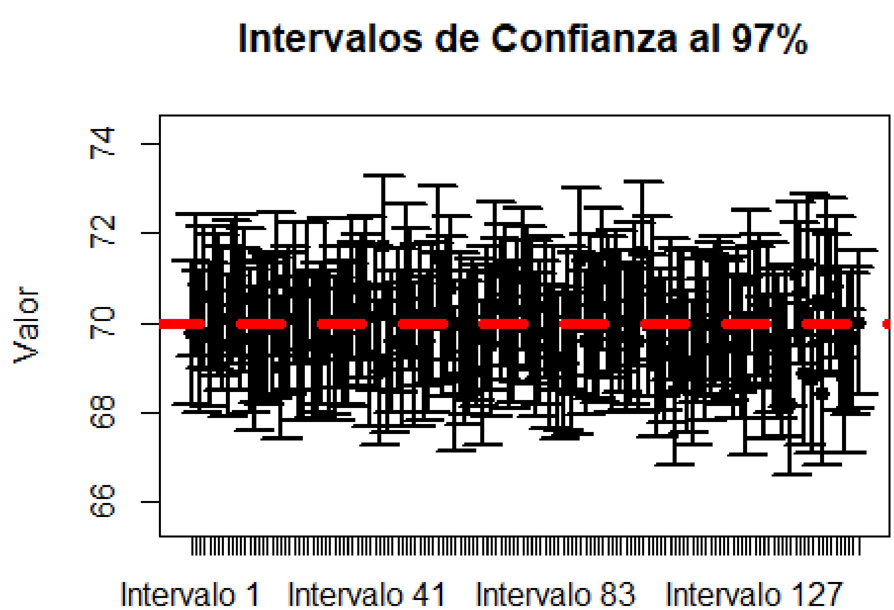
```r
axis(1, at=1:n_muestras, labels=paste("Intervalo", 1:n_muestras))

# Graficar cada intervalo y la media correspondiente
for (i in 1:n_muestras) {
  arrows(i, intervalos[i, 1], i, intervalos[i, 2], angle=90, code=3,
length = 0.1, lwd = 2)
  points(i, medias[i], pch=19, cex=1.1)
}
```



D. Grafica la media poblacional (miu = 70) como una linea horizontal

```r
# Establecer los límites del gráfico
min_y <- min(intervalos) - 1  # Un poco más pequeño que el valor mínimo
del intervalo
max_y <- max(intervalos) + 1  # Un poco más grande que el valor máximo
del intervalo

# Crear el gráfico con ejes intercambiados
plot(0, xlim=c(0, n_muestras + 1), ylim=c(min_y, max_y), xaxt="n",
ylab="Valor", xlab="", main="Intervalos de Confianza al 97%", horiz =
TRUE)

## Warning in plot.window(...): "horiz" is not a graphical parameter

## Warning in plot.xy(xy, type, ...): "horiz" is not a graphical
parameter
```

```
## Warning in axis(side = side, at = at, labels = labels, ...): "horiz"
is not a
## graphical parameter

## Warning in axis(side = side, at = at, labels = labels, ...): "horiz"
is not a
## graphical parameter

## Warning in box(...): "horiz" is not a graphical parameter

## Warning in title(...): "horiz" is not a graphical parameter

axis(1, at=1:n_muestras, labels=paste("Intervalo", 1:n_muestras))

# Graficar cada intervalo y la media correspondiente
for (i in 1:n_muestras) {
  arrows(i, intervalos[i, 1], i, intervalos[i, 2], angle=90, code=3,
length = 0.1, lwd = 2)
  points(i, medias[i], pch=19, cex=1.1)
}

# Dibujar una línea horizontal en la verdadera media de la población
abline(h = miu, col = "red", lwd = 5, lty = 2)
```



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

```r
# Parámetros de la población
miu <- 70
sigma <- 9
n_muestras <- 150
tamaño_muestra <- 150
alfa <- 0.03  # Nivel de significancia para un nivel de confianza del 97%
Z_alfa_2 <- abs(qnorm(alfa / 2))

# Simulación
set.seed(123)  # Para reproducibilidad
intervalos <- matrix(NA, nrow = n_muestras, ncol = 2)
medias <- numeric(n_muestras)
contiene_miu <- 0

for (i in 1:n_muestras) {
  muestra <- rnorm(tamaño_muestra, mean = miu, sd = sigma)
  media_muestra <- mean(muestra)
  error_est <- Z_alfa_2 * sigma / sqrt(tamaño_muestra)

  # Calcular el intervalo de confianza
  A <- media_muestra - error_est
  B <- media_muestra + error_est

  # Guardar el intervalo y la media
  intervalos[i, ] <- c(A, B)
  medias[i] <- media_muestra

  # Verificar si el intervalo contiene la verdadera media
  if (miu >= A & miu <= B) {
    contiene_miu <- contiene_miu + 1
  }
}

# Calcular el porcentaje de intervalos que contienen la verdadera media
porcentaje_contienen_miu <- (contiene_miu / n_muestras) * 100

cat("Número de intervalos que contienen la verdadera media:",
contiene_miu, "\n")

## Número de intervalos que contienen la verdadera media: 148

cat("Porcentaje de intervalos que contienen la verdadera media:",
porcentaje_contienen_miu, "%\n")

## Porcentaje de intervalos que contienen la verdadera media: 98.66667 %
```

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 al 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.

```
sigma = 0.75
alfa = 0.03
xb1 = 4.85
n1 = 20

E1 = abs(qnorm(alfa/2))*sigma/sqrt(n1)
A1 = xb1 - E1
B1 = xb1 + E1

cat("La verdadera media está entre", A1, "y", B1)

## La verdadera media está entre 4.486065 y 5.213935
```

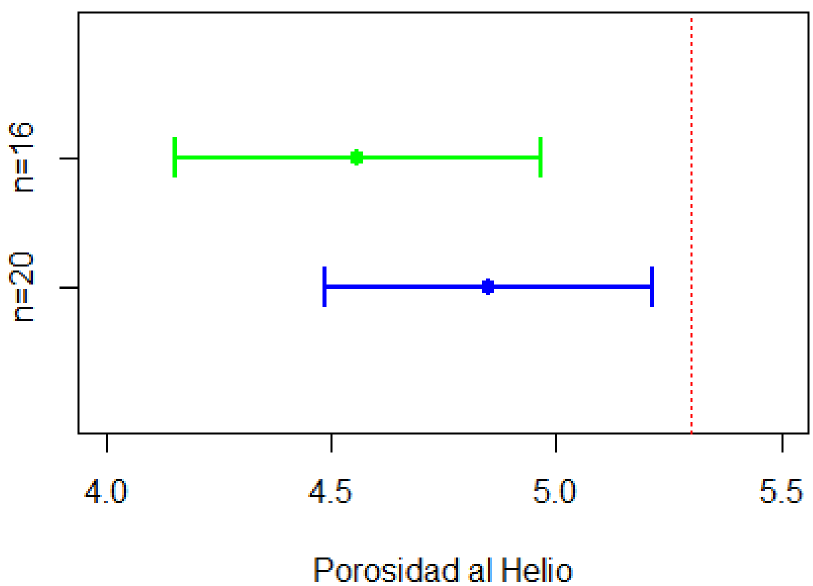
B. Se toma otra muestra de tamaño 16. El promedio de la muestra fue de 4.56. Calcule el intervalo de confianza al 97% de confianza

```
sigma = 0.75
alfa = 0.03
xb2 = 4.56
n2 = 16

E2 = abs(qnorm(alfa/2))*sigma/sqrt(n2)
A2 = xb2 - E2
B2 = xb2 + E2

cat("La verdadera media está entre", A2, "y", B2)

## La verdadera media está entre 4.153108 y 4.966892
```

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

```
plot(0, ylim=c(0,2+1), xlim=c(4,5.5), yaxt="n", ylab="", xlab="Porosidad
al 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(xb1, 1, pch=19, cex=1.1, col="blue")
points(xb2, 2, pch=19, cex=1.1, col="green")
abline(v=5.3, lty=3, col="red")
```



Como se puede observar, la porosidad del Helio ha disminuido del 5.3 que se tenía anteriormente ya que en los muestras que se tomaron, la media se encontraba en los intervalos de 4.486065 y 5.213935 para n = 20 y de 4.153108 y 4.966892 para n = 16 con un 97% de confianza, debido a que el 5.3 no está dentro de ninguno de los intervalos se puede afirmar con un 97% de confiaza que la media ha disminuido.

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?

```
sigma <- 0.75
alfa <- 0.05
E <- 0.2
Z_alfa_2 <- abs(qnorm(alfa / 2))

n_required <- (Z_alfa_2 * sigma / E)^2
n_required <- ceiling(n_required)
```

```r
cat("El tamaño de la muestra necesario es:", n_required)

## El tamaño de la muestra necesario es: 55
```

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%?

```r
sigma <- 0.75
alfa <- 0.01
E <- 0.2
Z_alfa_2 <- abs(qnorm(alfa / 2))

n_required <- (Z_alfa_2 * sigma / E)^2
n_required <- ceiling(n_required)

cat("El tamaño de la muestra necesario es:", n_required)

## El tamaño de la muestra necesario es: 94
```

Problema 3

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

```r
# Especificar el nuevo directorio
nuevo_directorio <- "C:/Users/luism/Escritorio/Documetos_2/Actividades
Concentración"

# Cambiar al nuevo directorio
setwd(nuevo_directorio)

# Carga los datos desde el archivo datosRes.csv
datos <- read.csv("El marcapasos.csv")
```

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

```r
# Definir función para calcular intervalos de confianza
intervalo_confianza <- function(x, confianza = 0.95) {
  n <- length(x)
  media <- mean(x)
  error_estandar <- sd(x) / sqrt(n)
  error_margin <- qt(confianza/2 + 0.5, df=n-1) * error_estandar
  return(c(media - error_margin, media + error_margin))
}

# Filtrar datos por Marcapasos
datos_sin_mp <- subset(datos, Marcapasos == "Sin MP")
datos_con_mp <- subset(datos, Marcapasos == "Con MP")

# Calcular intervalos de confianza para Intensidad de pulso
ic_intensidad_sin_mp <-
```

```r
intervalo_confianza(datos_sin_mp$Intensidad.de.pulso)
ic_intensidad_con_mp <-
intervalo_confianza(datos_con_mp$Intensidad.de.pulso)

# Mostrar resultados
cat("Intensidad de pulso sin MP (IC 95%):", ic_intensidad_sin_mp, "\n")

## Intensidad de pulso sin MP (IC 95%): 0.16993 0.2442661

cat("Intensidad de pulso con MP (IC 95%):", ic_intensidad_con_mp, "\n")

## Intensidad de pulso con MP (IC 95%): 0.1638035 0.2280788
```

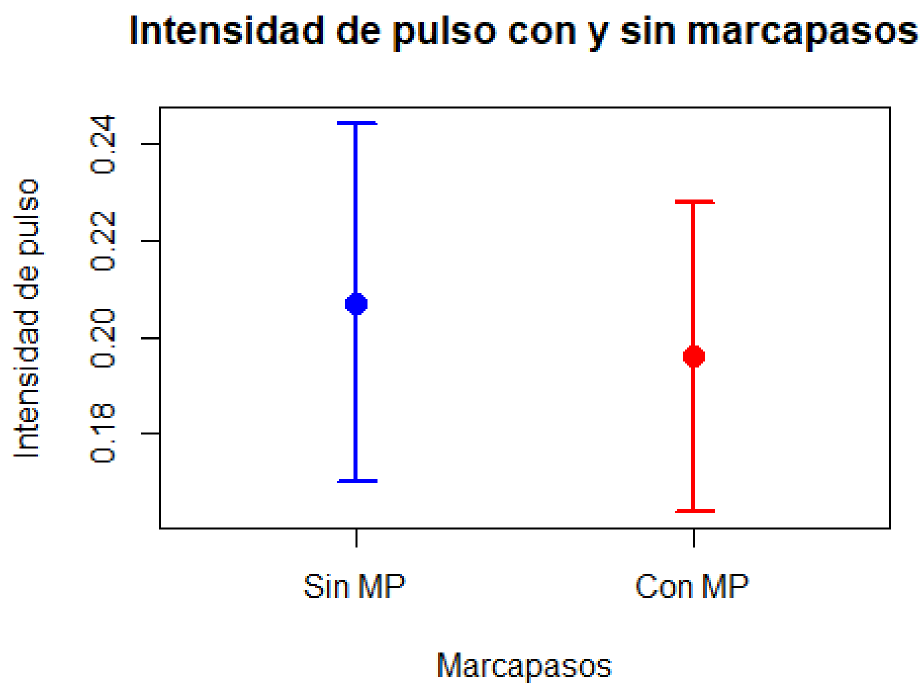
Periodo entre pulso con y sin Marcapasos (2 intervalos de confianza)

```r
# Calcular intervalos de confianza para Periodo entre pulsos
ic_periodo_sin_mp <-
intervalo_confianza(datos_sin_mp$Periodo.entre.pulsos)
ic_periodo_con_mp <-
intervalo_confianza(datos_con_mp$Periodo.entre.pulsos)

cat("Periodo entre pulsos sin MP (IC 95%):", ic_periodo_sin_mp, "\n")

## Periodo entre pulsos sin MP (IC 95%): 1.002887 1.220643

cat("Periodo entre pulsos con MP (IC 95%):", ic_periodo_con_mp, "\n")

## Periodo entre pulsos con MP (IC 95%): 0.8637941 0.9185589
```

Grafica los intervalos de confianza obtenidos en “El marcapasos”:

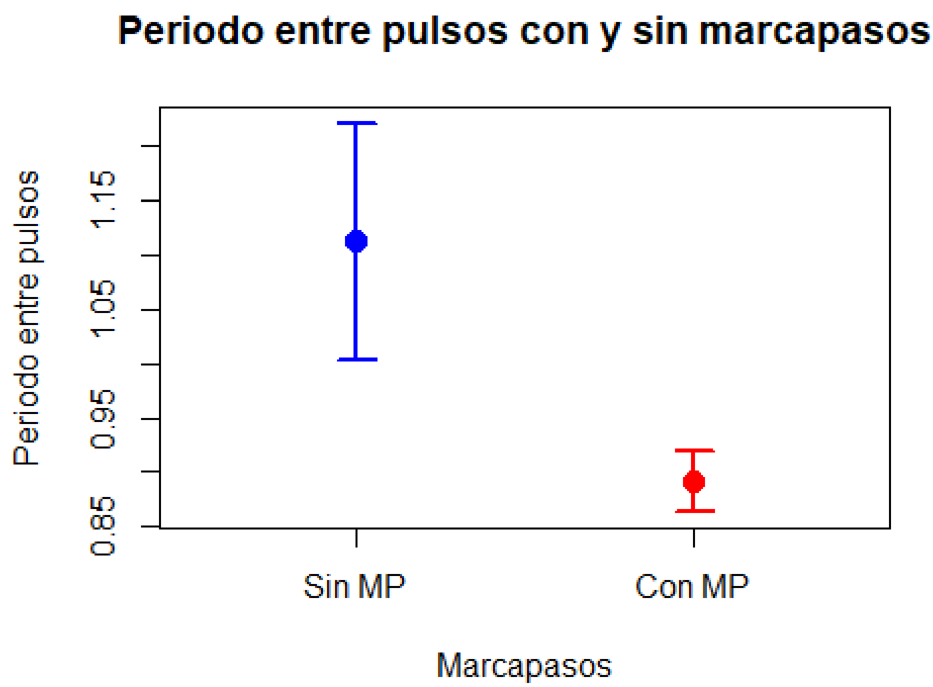
Grafica en un mismo eje coordenado la intensidad de pulso con y sin marcapasos

```r
# Graficar Intervalos de Confianza para la Intensidad de Pulso
plot(1, type="n", xlim=c(0.5, 2.5), ylim=range(c(ic_intensidad_sin_mp,
ic_intensidad_con_mp), na.rm=TRUE),
     xaxt="n", ylab="Intensidad de pulso", xlab="Marcapasos",
main="Intensidad de pulso con y sin marcapasos")

# Agregar intervalos de confianza
arrows(1, ic_intensidad_sin_mp[1], 1, ic_intensidad_sin_mp[2], angle=90,
code=3, length=0.1, col="blue", lwd=2)
points(1, mean(datos_sin_mp$Intensidad.de.pulso), pch=19, col="blue",
cex=1.5)
arrows(2, ic_intensidad_con_mp[1], 2, ic_intensidad_con_mp[2], angle=90,
code=3, length=0.1, col="red", lwd=2)
points(2, mean(datos_con_mp$Intensidad.de.pulso), pch=19, col="red",
cex=1.5)

# Agregar etiquetas de los ejes x
axis(1, at=c(1, 2), labels=c("Sin MP", "Con MP"))
```



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

```
# Graficar Intervalos de Confianza para el Periodo entre Pulsos
plot(1, type="n", xlim=c(0.5, 2.5), ylim=range(c(ic_periodo_sin_mp,
ic_periodo_con_mp), na.rm=TRUE),
     xaxt="n", ylab="Periodo entre pulsos", xlab="Marcapasos",
main="Periodo entre pulsos con y sin marcapasos")

# Agregar intervalos de confianza
arrows(1, ic_periodo_sin_mp[1], 1, ic_periodo_sin_mp[2], angle=90,
code=3, length=0.1, col="blue", lwd=2)
points(1, mean(datos_sin_mp$Periodo.entre.pulsos), pch=19, col="blue",
cex=1.5)
arrows(2, ic_periodo_con_mp[1], 2, ic_periodo_con_mp[2], angle=90,
code=3, length=0.1, col="red", lwd=2)
points(2, mean(datos_con_mp$Periodo.entre.pulsos), pch=19, col="red",
cex=1.5)

# Agregar etiquetas de los ejes x
axis(1, at=c(1, 2), labels=c("Sin MP", "Con MP"))
```



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

Intensidad de pulso: La presencia del marcapasos no parece afectar la intensidad del pulso, si bien es cierto que la media es menor con marcapasos, gran parte de los dos intervalos quedan empalmadaos por lo que no se podría afirmar su real diferencia.

Periodo entre pulsos: El marcapasos parece estabilizar el periodo entre pulsos, reduciendo la variabilidad y además de reducir drásticamente la media del periodo entre pulsos.