Act 7

Ozner Leyva

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:

Haz la simulación de 150 muestras de tamaño 150 extraídas de una población normal con miu = 70 y sigma = 9

```r
# Semilla para reproducibilidad
set.seed(123)

# Parámetros de la población
mu <- 70
sigma <- 9

# Número de muestras y tamaño de muestra
n_muestras <- 150
n_obs <- 150

# Crear vector para almacenar los límites inferiores y superiores de los
intervalos de confianza
lim_inf <- numeric(n_muestras)
lim_sup <- numeric(n_muestras)

# Crear vector para almacenar si cada intervalo contiene o no la media
verdadera
contiene_media <- logical(n_muestras)

# Realizar la simulación
for (i in 1:n_muestras) {
  # Generar una muestra de la población
  muestra <- rnorm(n_obs, mean = mu, sd = sigma)

  # Calcular el intervalo de confianza del 95% para la media
  lim_inf[i] <- mean(muestra) - 1.96 * (sigma / sqrt(n_obs))
  lim_sup[i] <- mean(muestra) + 1.96 * (sigma / sqrt(n_obs))

  # Verificar si el intervalo de confianza contiene la media verdadera
  contiene_media[i] <- (mu >= lim_inf[i] && mu <= lim_sup[i])
```

```r
}

# Calcular el porcentaje de intervalos que contienen la media verdadera
porcentaje_contiene_media <- mean(contiene_media) * 100

# Imprimir el resultado
cat("El porcentaje de intervalos de confianza que contienen la media verdadera es:", porcentaje_contiene_media, "%.")

## El porcentaje de intervalos de confianza que contienen la media
verdadera es: 96 %.
```

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

```r
# Crear vector para almacenar los límites inferiores y superiores de los intervalos de confianza
lim_inf_97 <- numeric(n_muestras)
lim_sup_97 <- numeric(n_muestras)

# Realizar la simulación
for (i in 1:n_muestras) {
  # Generar una muestra de la población
  muestra <- rnorm(n_obs, mean = mu, sd = sigma)

  # Calcular el intervalo de confianza del 97% para la media
  lim_inf_97[i] <- mean(muestra) - 2.576 * (sigma / sqrt(n_obs))
  lim_sup_97[i] <- mean(muestra) + 2.576 * (sigma / sqrt(n_obs))
}

# Imprimir los resultados
for (i in 1:n_muestras) {
  # Generar una muestra de la población
  muestra <- rnorm(n_obs, mean = mu, sd = sigma)

  # Calcular el intervalo de confianza del 97% para la media
  lim_inf_97[i] <- mean(muestra) - 2.576 * (sigma / sqrt(n_obs))
  lim_sup_97[i] <- mean(muestra) + 2.576 * (sigma / sqrt(n_obs))
}

# Imprimir los resultados
cat("Límites inferiores de los intervalos de confianza del 97%:\n")

## Límites inferiores de los intervalos de confianza del 97%:

head(lim_inf_97)

## [1] 67.46036 68.68502 67.88041 68.05569 67.48119 67.94721

cat("\nLímites superiores de los intervalos de confianza del 97%:\n")
```

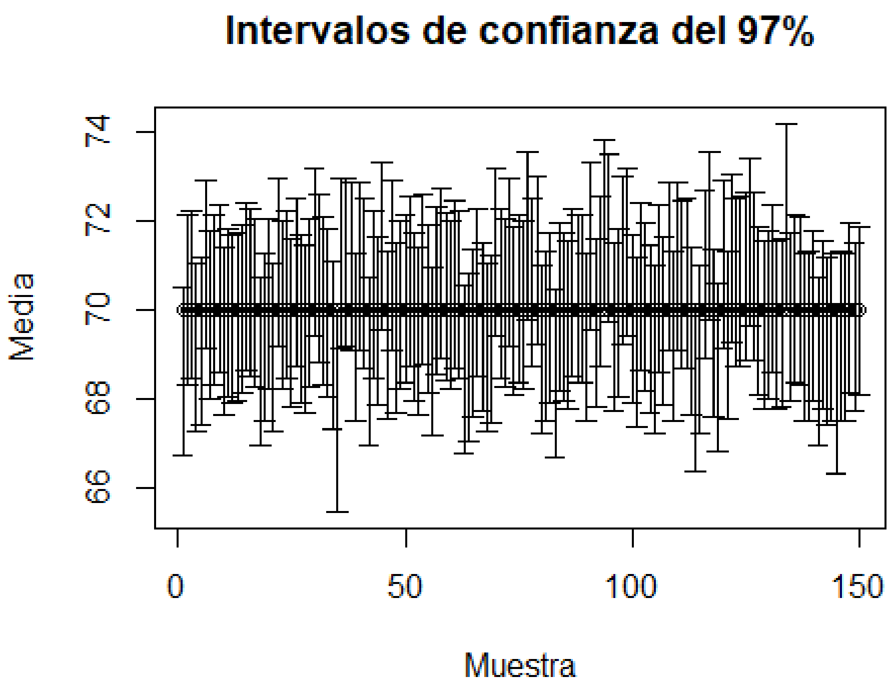
```
##
## Límites superiores de los intervalos de confianza del 97%:
```

```r
head(lim_sup_97)
```

```
## [1] 71.24629 72.47096 71.66634 71.84162 71.26712 71.73314
```

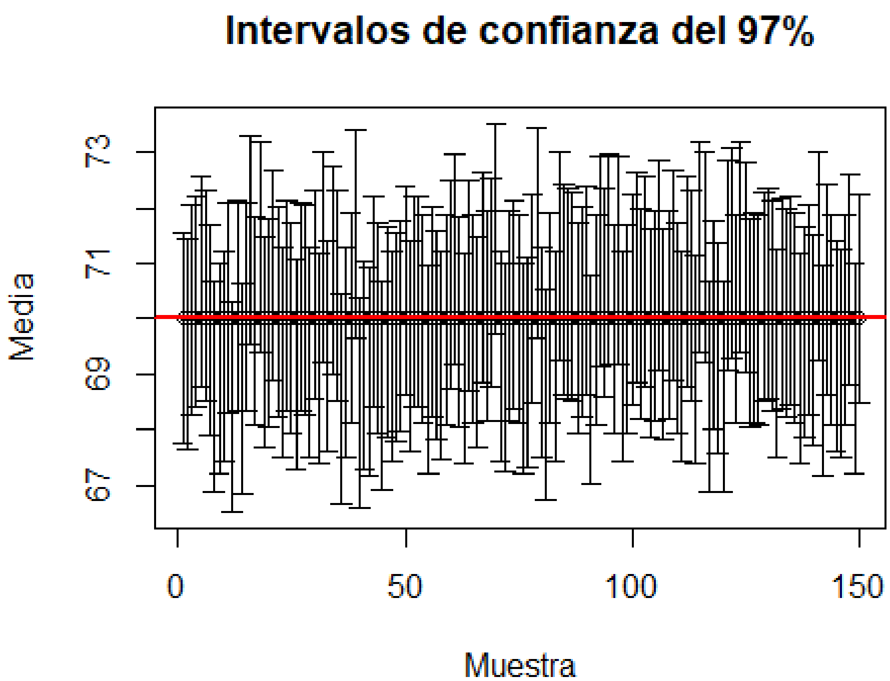
Grafica los 150 intervalos de confianza

```r
# Cargar la librería plotrix
library(plotrix)

# Crear vector para almacenar los límites inferiores y superiores de los
intervalos de confianza
lim_inf_97 <- numeric(n_muestras)
lim_sup_97 <- numeric(n_muestras)

# Realizar la simulación
for (i in 1:n_muestras) {
  # Generar una muestra de la población
  muestra <- rnorm(n_obs, mean = mu, sd = sigma)

  # Calcular el intervalo de confianza del 97% para la media
  lim_inf_97[i] <- mean(muestra) - 2.576 * (sigma / sqrt(n_obs))
  lim_sup_97[i] <- mean(muestra) + 2.576 * (sigma / sqrt(n_obs))
}

# Crear el gráfico de intervalos de confianza
plotCI(1:n_muestras, rep(mu, n_muestras), uiw = lim_sup_97 - mu, liw = mu
- lim_inf_97, xlab = "Muestra", ylab = "Media", main = "Intervalos de
confianza del 97%")
```



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

```
# Crear vector para almacenar los límites inferiores y superiores de los
intervalos de confianza
lim_inf_97 <- numeric(n_muestras)
lim_sup_97 <- numeric(n_muestras)

# Realizar la simulación
for (i in 1:n_muestras) {
  # Generar una muestra de la población
  muestra <- rnorm(n_obs, mean = mu, sd = sigma)

  # Calcular el intervalo de confianza del 97% para la media
  lim_inf_97[i] <- mean(muestra) - 2.576 * (sigma / sqrt(n_obs))
  lim_sup_97[i] <- mean(muestra) + 2.576 * (sigma / sqrt(n_obs))
}

# Crear el gráfico de intervalos de confianza
par(mar = c(5, 4, 4, 2) + 0.1) # Ajustar los márgenes del gráfico
plotCI(1:n_muestras, rep(mu, n_muestras), uiw = lim_sup_97 - mu, liw = mu
- lim_inf_97, xlab = "Muestra", ylab = "Media", main = "Intervalos de
confianza del 97%")
abline(h = mu, col = "red", lwd = 2) # Agregar la línea horizontal de la
media poblacional
```



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

```
# Crear vector para almacenar los límites inferiores y superiores de los intervalos de confianza
lim_inf_97 <- numeric(n_muestras)
lim_sup_97 <- numeric(n_muestras)

# Crear vector para almacenar si cada intervalo contiene o no la media verdadera
contiene_media <- logical(n_muestras)

# Realizar la simulación
for (i in 1:n_muestras) {
  # Generar una muestra de la población
  muestra <- rnorm(n_obs, mean = mu, sd = sigma)

  # Calcular el intervalo de confianza del 97% para la media
  lim_inf_97[i] <- mean(muestra) - 2.576 * (sigma / sqrt(n_obs))
  lim_sup_97[i] <- mean(muestra) + 2.576 * (sigma / sqrt(n_obs))

  # Verificar si el intervalo de confianza contiene la media verdadera
  contiene_media[i] <- (mu >= lim_inf_97[i] && mu <= lim_sup_97[i])
}

# Calcular el número de intervalos que contienen la media verdadera
```

```r
num_intervalos_contienen_media <- sum(contiene_media)

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

# Imprimir los resultados
cat("Número de intervalos de confianza del 97% que contienen la media
verdadera:", num_intervalos_contienen_media, "\n")

## Número de intervalos de confianza del 97% que contienen la media
verdadera: 150

cat("Porcentaje de intervalos que contienen la media verdadera:",
porcentaje_contiene_media, "%")

## Porcentaje de intervalos que contienen la media verdadera: 100 %
```

Problema 2

Primera Parte

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.

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

```r
sigma = 0.75
alfa = 0.03
xb1 = 4.85 # media de la muestra, no de x
n1 = 20

E1 = abs(qnorm(0.03/2))*sigma/sqrt(n1)
A1 = xb1 -E1
B1 = xb1 + E1
cat("La verdadera media actual está entre", A1, "y", B1)

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

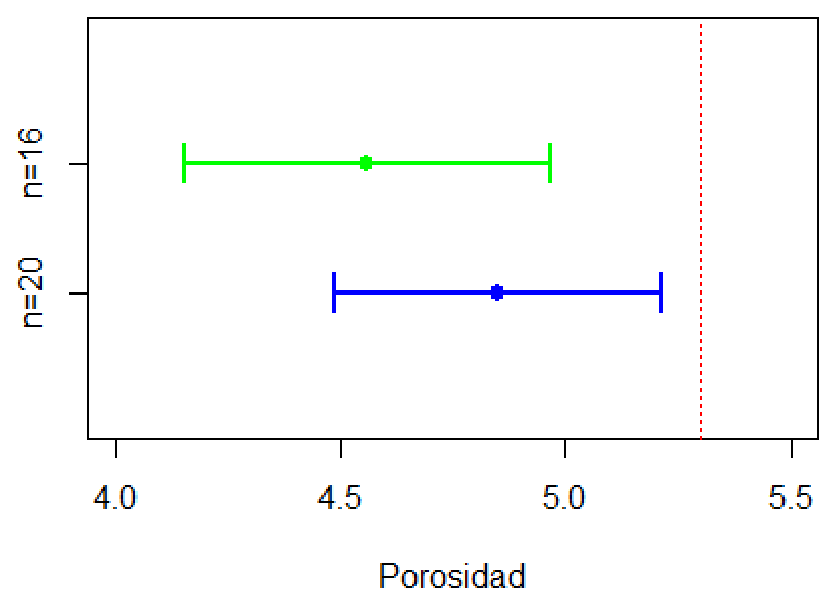
Se toma otra muestra de tamaño 16. El promedio de la muestra fue de 4.56. Calcule el intervalo de confianza al 97% de confianza

```r
sigma = 0.75
alfa = 0.03
xb2 = 4.56 # media de la muestra, no de x
```

```
n2 = 16

E2 = abs(qnorm(0.03/2))*sigma/sqrt(n2)
A2 = xb2 - E2
B2 = xb2 + E2
cat("La verdadera media actual está entre", A2, "y", B2)

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

¿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")
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")
```



La gráfica superior muestra una línea horizontal de color morado en 5.3, que representa la media original de porosidad. Debajo de esta línea se observan dos intervalos de confianza: uno en azul para una muestra de 20 elementos (n=20) y otro en rosa para una muestra de 16 elementos (n=16).

Es notable que en ambos intervalos de confianza se sitúan completamente por debajo del valor 5.3 y no lo incluyen en sus rangos. Esto nos permite concluir, con un alto grado de certeza estadística, que la porosidad del helio ha experimentado una disminución significativa respecto a su media original.

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.

```
sigma = 0.75
W = 0.4
alpha = 0.05


z_alpha_2 = qnorm(1 - alpha / 2)


n_required = (2 * z_alpha_2 * sigma / W)^2
n_required = ceiling(n_required)  # Redondear al número entero más
cercano

cat("El tamaño de muestra necesario para que el ancho del intervalo no
sobrepase de 0.4 con un 95% de confianza es:", n_required)

## El tamaño de muestra necesario para que el ancho del intervalo no
sobrepase de 0.4 con un 95% de confianza es: 55
```

¿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%?

```
sigma = 0.75
E = 0.2
alpha = 0.01

z_alpha_2 = qnorm(1 - alpha / 2)


n_required = (z_alpha_2 * sigma / E)^2
n_required = ceiling(n_required)  # Redondear al número entero más
cercano

cat("El tamaño de muestra necesario para estimar la porosidad promedio
dentro de 0.2 unidades con una confianza de 99% es:", n_required)
```

```
## El tamaño de muestra necesario para estimar la porosidad promedio dentro de 0.2 unidades con una confianza de 99% es: 94
```

Problema 3

A) Con el archivo de data de El Marcapasos Download El Marcapasoshaz los intervalos de confianza para la media de las siguientes variables:

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

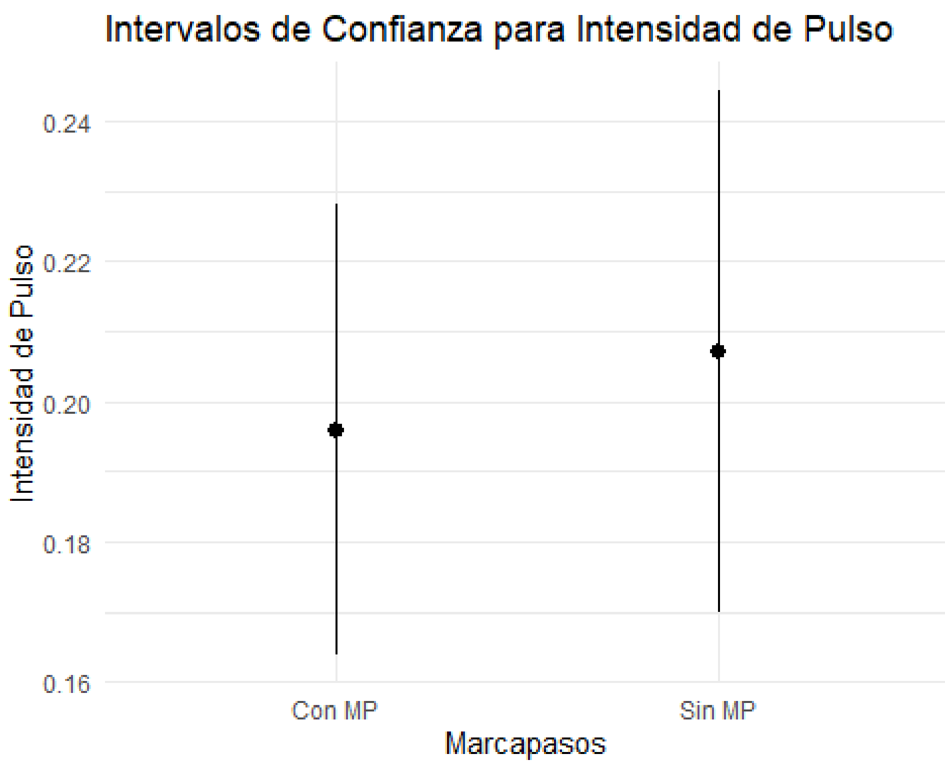
```
data <- read.csv("C:/Users/ozner/Desktop/Computer Science/R/El marcapasos.csv")

# Función para calcular intervalos de confianza
calcular_ic <- function(x, alpha = 0.05) {
  n <- length(x)
  media <- mean(x)
  sd <- sd(x)
  error_estandar <- sd / sqrt(n)
  t_critico <- qt(1 - alpha / 2, df = n - 1)

  ic_inf <- media - t_critico * error_estandar
  ic_sup <- media + t_critico * error_estandar

  return(c(ic_inf, ic_sup))
}

# Intervalos de confianza para Intensidad de pulso
intensidad_sin_mp <- data$Intensidad[data$Marcapasos == "Sin MP"]
intensidad_con_mp <- data$Intensidad[data$Marcapasos == "Con MP"]

ic_intensidad_sin_mp <- calcular_ic(intensidad_sin_mp)
ic_intensidad_con_mp <- calcular_ic(intensidad_con_mp)

# Intervalos de confianza para Periodo entre pulso
periodo_sin_mp <- data$Periodo[data$Marcapasos == "Sin MP"]
periodo_con_mp <- data$Periodo[data$Marcapasos == "Con MP"]

ic_periodo_sin_mp <- calcular_ic(periodo_sin_mp)
ic_periodo_con_mp <- calcular_ic(periodo_con_mp)

# Mostrar resultados
cat("A continuación, estos son los intervalos de confianza para la media de las variables especificadas:", "\n", "\n")

## A continuación, estos son los intervalos de confianza para la media de las variables especificadas:
##
```

```
cat("Intensidad de pulsos sin Marcapasos:", ic_intensidad_sin_mp, "\n")

## Intensidad de pulsos sin Marcapasos: 0.16993 0.2442661

cat("Intensidad de pulsos con Marcapasos:", ic_intensidad_con_mp, "\n")

## Intensidad de pulsos con Marcapasos: 0.1638035 0.2280788

cat("Periodo entre pulso sin Marcapasos:", ic_periodo_sin_mp, "\n")

## Periodo entre pulso sin Marcapasos: 1.002887 1.220643

cat("Periodo entre pulso con Marcapasos:", ic_periodo_con_mp, "\n")

## Periodo entre pulso con Marcapasos: 0.8637941 0.9185589
```

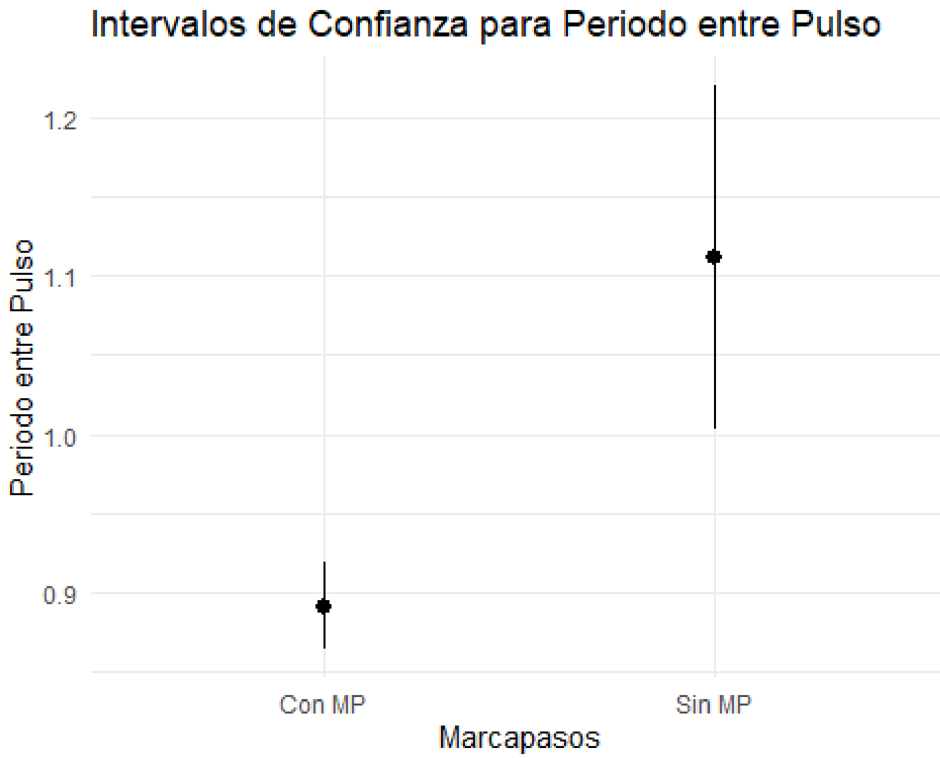
B) Grafica los intervalos de confianza obtenidos en "El marcapasos":

- Grafica en un mismo eje coordenado la intensidad de pulso con y sin marcapasos
- Grafica en un mismo eje coordenado el periodo entre pulso con y sin marcapasos

```
if (!require(ggplot2)) install.packages("ggplot2")

## Loading required package: ggplot2

library(ggplot2)

# Crear un dataframe para la gráfica de Intensidad de pulso
df_intensidad <- data.frame(
  Marcapasos = c("Sin MP", "Con MP"),
  Media = c(mean(intensidad_sin_mp), mean(intensidad_con_mp)),
  IC_inf = c(ic_intensidad_sin_mp[1], ic_intensidad_con_mp[1]),
  IC_sup = c(ic_intensidad_sin_mp[2], ic_intensidad_con_mp[2])
)

# Graficar Intensidad de pulso
ggplot(df_intensidad, aes(x = Marcapasos, y = Media, ymin = IC_inf, ymax
= IC_sup)) +
  geom_pointrange() +
  labs(title = "Intervalos de Confianza para Intensidad de Pulso",
       y = "Intensidad de Pulso") +
  theme_minimal()
```




```r
# Crear un dataframe para la gráfica de Periodo entre pulso
df_periodo <- data.frame(
  Marcapasos = c("Sin MP", "Con MP"),
  Media = c(mean(periodo_sin_mp), mean(periodo_con_mp)),
  IC_inf = c(ic_periodo_sin_mp[1], ic_periodo_con_mp[1]),
  IC_sup = c(ic_periodo_sin_mp[2], ic_periodo_con_mp[2])
)

# Graficar Periodo entre pulso
ggplot(df_periodo, aes(x = Marcapasos, y = Media, ymin = IC_inf, ymax =
IC_sup)) +
  geom_pointrange() +
  labs(title = "Intervalos de Confianza para Periodo entre Pulso",
       y = "Periodo entre Pulso") +
  theme_minimal()
```



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

El análisis de las gráficas revela información valiosa sobre el impacto del marcapasos en dos variables: la intensidad de los pulsos y el periodo entre ellos.

Intensidad de los pulsos: Los intervalos de confianza para la intensidad de los pulsos, tanto con como sin marcapasos, se encuentran en rangos muy similares. Esta nos sugiere que el marcapasos no ejerce una influencia significativa en la intensidad de los pulsos cardíacos.

Periodo entre pulsos: En contraste, la gráfica del periodo entre pulsos muestra una clara separación entre los intervalos de confianza de las condiciones con y sin marcapasos. Esta diferencia nos indica que el marcapasos tiene un efecto significativo en esta variable. Siendo especificos, se observa que el periodo entre pulsos es más corto cuando el marcapasos está activo.