

# Intervalos de confianza

Catherine Rojas

2024-08-21

## Problema 2

### El misterioso Helio

#### Parte 1

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.

Parámetros conocidos: Desviación estándar ( $\sigma$ ) = 0.75 Tamaño de la muestra ( $n$ ) = 20 Promedio muestral ( $\bar{x}$ ) = 4.85 Nivel de confianza = 97%

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

El intervalo de confianza para la media se calcula utilizando la siguiente fórmula:  $\bar{x} \pm Z_{\alpha/2} \times \frac{\sigma}{\sqrt{n}}$

Donde:

$Z_{\alpha/2}$  es el valor crítico de la distribución normal estándar correspondiente al nivel de confianza.  
 $\sigma$  es la desviación estándar.  $n$  es el tamaño de la muestra.

```
# Parámetros
sigma <- 0.75
n <- 20
x_bar <- 4.85
confidence_level <- 0.97

# Valor crítico Z
Z <- qnorm((1 + confidence_level) / 2)

# Error estándar
SE <- sigma / sqrt(n)

# Intervalo de confianza
```

```

lower_bound <- x_bar - Z * SE
upper_bound <- x_bar + Z * SE

# Mostrar el resultado
cat("El intervalo de confianza del 97% para el promedio de porosidad es:
[", lower_bound, ", ", upper_bound, "]\n")

## El intervalo de confianza del 97% para el promedio de porosidad es: [
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

Desviación estándar ( $\sigma$ ) = 0.75 Tamaño de la muestra ( $n$ ) = 16 Promedio muestral ( $\bar{x}$ ) = 4.96 Nivel de confianza = 97%

```

# Nuevos parámetros
sigma <- 0.75
n_1 <- 16
x_bar_1 <- 4.56

# Error estándar
SE_1 <- sigma / sqrt(n_1)

# Intervalo de confianza
lower_bound_1 <- x_bar_1 - Z * SE_1
upper_bound_1 <- x_bar_1 + Z * SE_1

# Mostrar el resultado
cat("El intervalo de confianza del 97% para el promedio de porosidad con
la nueva muestra es: [", lower_bound_1, ", ", upper_bound_1, "]\n")

## El intervalo de confianza del 97% para el promedio de porosidad con la
nueva muestra es: [ 4.153108 , 4.966892 ]

```

## c) ¿Podemos afirmar que la porosidad del helio ha disminuido?

Si el valor promedio histórico de 5.3 no está incluido en el intervalo de confianza calculado, esto podría sugerir que la porosidad ha disminuido con un 97% de confianza. Si el valor histórico está dentro del intervalo, no hay suficiente evidencia para afirmar que la porosidad ha disminuido.

En este caso, se puede afirmar que la porosidad del helio a disminuido.

**Parte 2** 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.

### ¿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?

Desviación estándar ( $\sigma$ ) = 0.75 Ancho máximo del intervalo de confianza = 0.4  
Nivel de confianza = 95%

```
# Parámetros conocidos
sigma <- 0.75
desired_width <- 0.4
confidence_level <- 0.95

# Valor crítico Z
Z <- qnorm((1 + confidence_level) / 2)

# Cálculo del tamaño de la muestra
n_required <- (2 * Z * sigma / desired_width)^2

# Redondear hacia arriba al entero más cercano
n_required <- ceiling(n_required)

# Mostrar el resultado
cat("El tamaño de la muestra necesario es:", n_required, "\n")

## El tamaño de la muestra necesario 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%?

Desviación estándar ( $\sigma$ ) = 0.75 Margen de error deseado = 0.2 Nivel de confianza = 99%

```
# Parámetros conocidos
sigma <- 0.75
desired_margin <- 0.2
confidence_level <- 0.99

# Valor crítico Z
Z <- qnorm((1 + confidence_level) / 2)

# Cálculo del tamaño de la muestra
n_required <- (Z * sigma / desired_margin)^2

# Redondear hacia arriba al entero más cercano
n_required <- ceiling(n_required)

# Mostrar el resultado
cat("El tamaño de la muestra necesario es:", n_required, "\n")
```

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

## 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  $\mu = 70$  y  $\sigma = 9$**

Media poblacional ( $\mu$ ) = 70 Desviación estándar ( $\sigma$ ) = 9  $\text{\text{Tamaño de cada muestra}}$  (n) = 150  $\text{\text{Número de muestras}}$  (N) = 150  $\text{\text{Nivel de confianza}}$  = 95%

```
# Parámetros
mu <- 70
sigma <- 9
n <- 150
N <- 150
confidence_level <- 0.95
Z <- qnorm((1 + confidence_level) / 2)

# Simulación
set.seed(123) # Fijar la semilla para reproducibilidad
count <- 0 # Contador de intervalos que contienen la verdadera media

for (i in 1:N) {
  # Generar una muestra aleatoria
  sample_data <- rnorm(n, mean = mu, sd = sigma)

  # Calcular el promedio muestral
  x_bar <- mean(sample_data)

  # Calcular el error estándar
  SE <- sigma / sqrt(n)

  # Calcular el intervalo de confianza
  lower_bound <- x_bar - Z * SE
  upper_bound <- x_bar + Z * SE

  # Verificar si el intervalo contiene la verdadera media
  if (lower_bound <= mu & upper_bound >= mu) {
    count <- count + 1
  }
}

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

```
percentage <- (count / N) * 100

# Mostrar el resultado
cat("El porcentaje de intervalos de confianza que contienen la verdadera
media es:", percentage, "%\n")

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

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

```
# Parámetros
mu <- 70
sigma <- 9
n <- 150
N <- 150
confidence_level <- 0.97
Z <- qnorm((1 + confidence_level) / 2)

# Lista para almacenar los intervalos de confianza
intervals <- matrix(NA, nrow = N, ncol = 2)
colnames(intervals) <- c("Lower Bound", "Upper Bound")

# Simulación
set.seed(123) # Fijar la semilla para reproducibilidad

for (i in 1:N) {
  # Generar una muestra aleatoria
  sample_data <- rnorm(n, mean = mu, sd = sigma)

  # Calcular el promedio muestral
  x_bar <- mean(sample_data)

  # Calcular el error estándar
  SE <- sigma / sqrt(n)

  # Calcular el intervalo de confianza
  lower_bound <- x_bar - Z * SE
  upper_bound <- x_bar + Z * SE

  # Guardar los intervalos en la matriz
  intervals[i, ] <- c(lower_bound, upper_bound)
}

# Mostrar los primeros 10 intervalos de confianza como ejemplo
print("Primeros 10 intervalos de confianza (97%):")

## [1] "Primeros 10 intervalos de confianza (97%):"
```

```
print(intervals[1:10, ])

##           Lower Bound Upper Bound
## [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
## [6,]      68.96493     72.15430
## [7,]      68.80405     71.99342
## [8,]      68.49602     71.68539
## [9,]      67.89586     71.08523
## [10,]     69.09620     72.28557

# Opcional: Guardar todos Los intervalos en un data frame y mostrar
intervals_df <- data.frame(intervals)
colnames(intervals_df) <- c("Lower Bound", "Upper Bound")

# Mostrar todos Los intervalos (puedes mostrar solo Los primeros si
prefieres)
head(intervals_df, 10)

##           Lower Bound Upper Bound
## 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
## 6      68.96493     72.15430
## 7      68.80405     71.99342
## 8      68.49602     71.68539
## 9      67.89586     71.08523
## 10     69.09620     72.28557
```

## Grafica los 150 intervalos de confianza

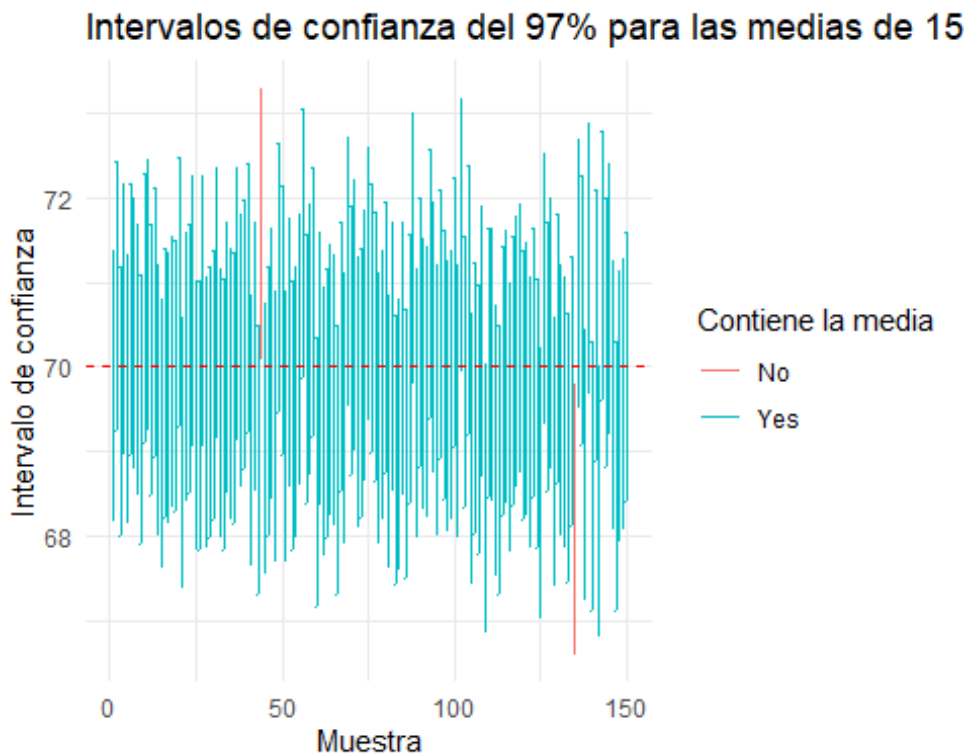
```
# Cargar La librería ggplot2
library(ggplot2)

## Warning: package 'ggplot2' was built under R version 4.3.3

# Crear un data frame con Los intervalos de confianza
intervals_df <- data.frame(
  Sample = 1:N,
  Lower_Bound = intervals[, 1],
  Upper_Bound = intervals[, 2]
)

# Añadir una columna que indique si el intervalo contiene La verdadera
media
intervals_df$Contains_Mu <- ifelse(intervals_df$Lower_Bound <= mu &
intervals_df$Upper_Bound >= mu, "Yes", "No")
```

```
# Crear La gráfica
ggplot(intervals_df, aes(x = Sample)) +
  geom_errorbar(aes(ymin = Lower_Bound, ymax = Upper_Bound, color =
Contains_Mu), width = 0.5) +
  geom_hline(yintercept = mu, linetype = "dashed", color = "red") +
  labs(title = "Intervalos de confianza del 97% para las medias de 150
muestras",
x = "Muestra",
y = "Intervalo de confianza",
color = "Contiene la media") +
  theme_minimal()
```



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

```
# Cargar la librería ggplot2
library(ggplot2)

# Crear un data frame con los intervalos de confianza
intervals_df <- data.frame(
  Sample = 1:N,
  Lower_Bound = intervals[, 1],
  Upper_Bound = intervals[, 2]
)

# Añadir una columna que indique si el intervalo contiene la verdadera
media
```

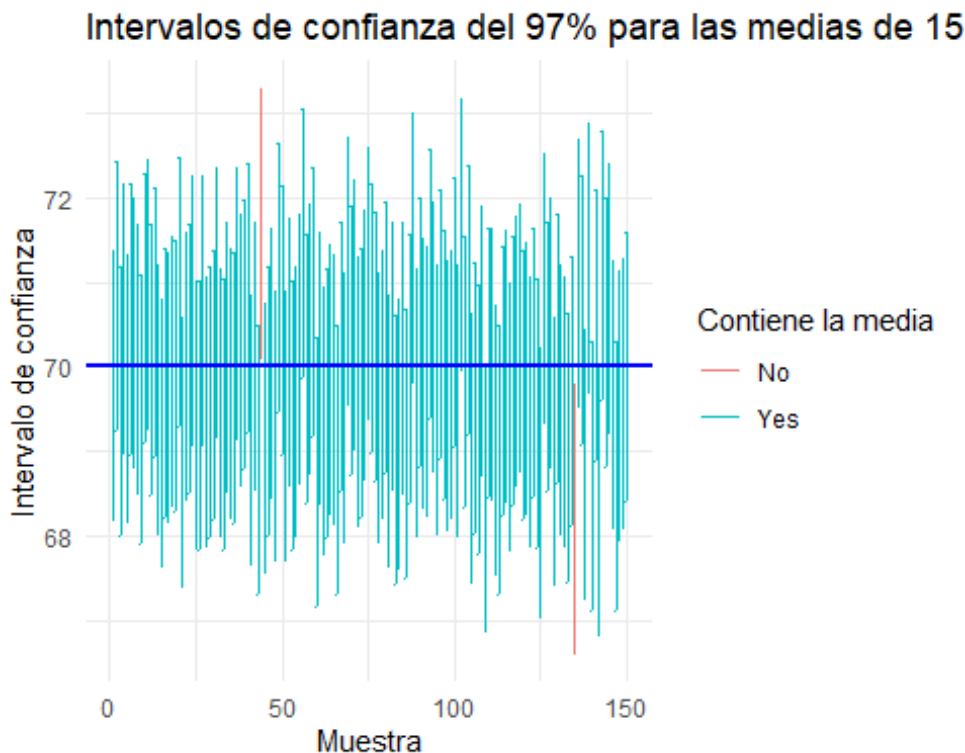
```

intervals_df$Contains_Mu <- ifelse(intervals_df$Lower_Bound <= mu &
intervals_df$Upper_Bound >= mu, "Yes", "No")

# Crear la gráfica
ggplot(intervals_df, aes(x = Sample)) +
  geom_errorbar(aes(ymin = Lower_Bound, ymax = Upper_Bound, color =
Contains_Mu), width = 0.5) +
  geom_hline(yintercept = mu, linetype = "solid", color = "blue", size =
1) + # Línea horizontal para la media poblacional
  labs(title = "Intervalos de confianza del 97% para las medias de 150
muestras",
    x = "Muestra",
    y = "Intervalo de confianza",
    color = "Contiene la media") +
  theme_minimal()

## Warning: Using `size` aesthetic for lines was deprecated in ggplot2
3.4.0.
## i Please use `linewidth` instead.
## This warning is displayed once every 8 hours.
## Call `lifecycle::last_lifecycle_warnings()` to see where this warning
was
## generated.

```



### Interpretación de gráficas

Las gráficas muestran que los intervalos de confianza al 97% son efectivos para contener la media verdadera en la mayoría de los casos (alrededor del 97% de las



muestras). Los pocos intervalos que no contienen la media verdadera son un recordatorio de que el nivel de confianza no garantiza que todos los intervalos contengan la media, sino que lo harán en el porcentaje especificado (en este caso, 97%).

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

```
# Contar cuántos intervalos contienen la verdadera media
count_contains_mu <- sum(intervals_df$Contains_Mu == "Yes")

# Calcular el porcentaje
percentage_contains_mu <- (count_contains_mu / N) * 100

# Mostrar el resultado
cat("Número de intervalos que contienen la verdadera media:",
    count_contains_mu, "\n")

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

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

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

## Problema 3

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

```
library(readr)

## Warning: package 'readr' was built under R version 4.3.3

data <- read_csv("El marcapasos.csv")

## Rows: 102 Columns: 3
## — Column specification
## Delimiter: ","
## chr (1): Marcapasos
## dbl (2): Periodo entre pulsos, Intensidad de pulso
##
## i Use `spec()` to retrieve the full column specification for this
data.
## i Specify the column types or set `show_col_types = FALSE` to quiet
this message.

# Filtrar datos para "Con MP" y "Sin MP"
data_con_mp <- subset(data, Marcapasos == 'Con MP')
data_sin_mp <- subset(data, Marcapasos == 'Sin MP')
```

```
# Definir nivel de confianza
confidence_level <- 0.95
alpha <- 1 - confidence_level
```

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

```
# Calcular intervalos de confianza ara Intesidad de pulso con y sin MP

# Intensidad de pulsos con y sin Marcapasos
intensidad_con_mp <- data_con_mp$`Intensidad de pulso`

intensidad_sin_mp <- data_sin_mp$`Intensidad de pulso`

# Calcular La media de pulsos con y sin Marcapasos
mean_intensidad_con_mp <- mean(intensidad_con_mp)
mean_intensidad_sin_mp <- mean(intensidad_sin_mp)

# Calcular Los intervalos de confianza de pulsos con y sin Marcapasos
ci_intensidad_con_mp <- mean_intensidad_con_mp + c(-1, 1) * qnorm(1 -
alpha/2) * sd(intensidad_con_mp) / sqrt(length(intensidad_con_mp))

ci_intensidad_sin_mp <- mean_intensidad_sin_mp + c(-1, 1) * qnorm(1 -
alpha/2) * sd(intensidad_sin_mp) / sqrt(length(intensidad_sin_mp))

# Mostrar Los resultados
cat("Intervalos de confianza al 95%:\n")

## Intervalos de confianza al 95%:

cat("Intensidad de pulso (Con MP): [", round(ci_intensidad_con_mp[1], 3),
", ", round(ci_intensidad_con_mp[2], 3), "]\n")

## Intensidad de pulso (Con MP): [ 0.165 , 0.227 ]

cat("Intensidad de pulso (Sin MP): [", round(ci_intensidad_sin_mp[1], 3),
", ", round(ci_intensidad_sin_mp[2], 3), "]\n")

## Intensidad de pulso (Sin MP): [ 0.171 , 0.243 ]
```

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

```
# Calcular intervalos de confianza para Periodo entre pulsos con y sin MP

# Periodo entre pulso con y sin Marcapasos
periodo_con_mp <- data_con_mp$`Periodo entre pulsos`
periodo_sin_mp <- data_sin_mp$`Periodo entre pulsos`

# Calcular La media del periodo entre pulso con y sin Marcapasos
mean_periodo_con_mp <- mean(periodo_con_mp)
mean_periodo_sin_mp <- mean(periodo_sin_mp)
```

```

# Calcular Los intervalos de confianza de Periodo entre pulso con y sin
# Marcapasos
ci_periodo_con_mp <- mean_periodo_con_mp + c(-1, 1) * qnorm(1 - alpha/2)
* sd(periodo_con_mp) / sqrt(length(periodo_con_mp))
ci_periodo_sin_mp <- mean_periodo_sin_mp + c(-1, 1) * qnorm(1 - alpha/2)
* sd(periodo_sin_mp) / sqrt(length(periodo_sin_mp))

# Mostrar Los resultados
cat("Periodo entre pulsos (Con MP): [", round(ci_periodo_con_mp[1], 3),
", ", round(ci_periodo_con_mp[2], 3), "]\n")

## Periodo entre pulsos (Con MP): [ 0.864 ,  0.918 ]

cat("Periodo entre pulsos (Sin MP): [", round(ci_periodo_sin_mp[1], 3),
", ", round(ci_periodo_sin_mp[2], 3), "]\n")

## Periodo entre pulsos (Sin MP): [ 1.006 ,  1.218 ]

```

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

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

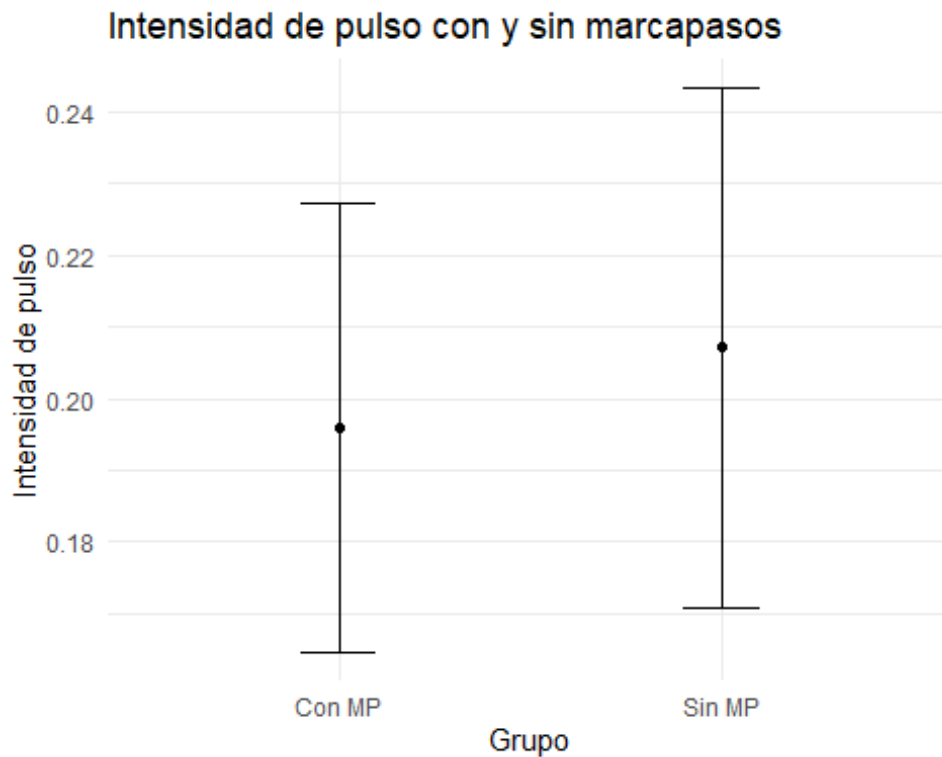
```

# Cargar la librería ggplot2
library(ggplot2)

# Datos de Los intervalos de confianza para la gráfica de Intensidad de
# pulso
intensidad_data <- data.frame(
  Grupo = c("Con MP", "Sin MP"),
  Media = c(mean_intensidad_con_mp, mean_intensidad_sin_mp),
  Lower = c(ci_intensidad_con_mp[1], ci_intensidad_sin_mp[1]),
  Upper = c(ci_intensidad_con_mp[2], ci_intensidad_sin_mp[2])
)

# Gráfica para La Intensidad de pulso
ggplot(intensidad_data, aes(x = Grupo, y = Media)) +
  geom_point() +
  geom_errorbar(aes(ymin = Lower, ymax = Upper), width = 0.2) +
  labs(title = "Intensidad de pulso con y sin marcapasos",
       x = "Grupo",
       y = "Intensidad de pulso") +
  theme_minimal()

```



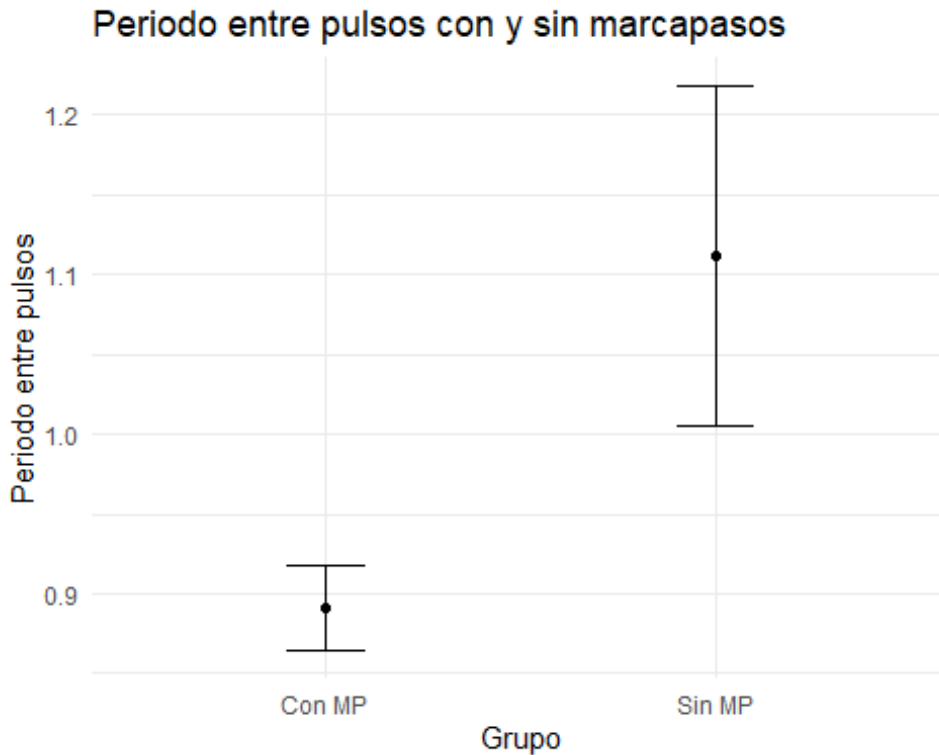
**Grafica en un mismo eje coordenado el periodo entre pulso con y sin marcapasos**

*# Datos de Los intervalos de confianza para la gráfica de Periodo entre pulsos*

```
periodo_data <- data.frame(
  Grupo = c("Con MP", "Sin MP"),
  Media = c(mean_periodo_con_mp, mean_periodo_sin_mp),
  Lower = c(ci_periodo_con_mp[1], ci_periodo_sin_mp[1]),
  Upper = c(ci_periodo_con_mp[2], ci_periodo_sin_mp[2])
)
```

*# Gráfica para el Periodo entre pulsos*

```
ggplot(periodo_data, aes(x = Grupo, y = Media)) +
  geom_point() +
  geom_errorbar(aes(ymin = Lower, ymax = Upper), width = 0.2) +
  labs(title = "Periodo entre pulsos con y sin marcapasos",
       x = "Grupo",
       y = "Periodo entre pulsos") +
  theme_minimal()
```



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

**Intensidad de Pulso: - Con Marcapasos:** El intervalo de confianza al 95% para la intensidad de pulso está entre 0.165 y 0.227. - **Sin Marcapasos:** El intervalo de confianza al 95% para la intensidad de pulso está entre 0.171 y 0.243.

Los intervalos de confianza de la intensidad de pulso con y sin marcapasos se solapan parcialmente, lo que sugiere que no hay una diferencia significativa en la intensidad de pulso entre los dos grupos. La intensidad de pulso parece estar en un rango similar, lo que indica que la presencia del marcapasos no tiene un impacto significativo en la intensidad de los pulsos.

**Periodo entre Pulsos: - Con Marcapasos:** El intervalo de confianza al 95% para el periodo entre pulsos está entre 0.864 y 0.918. - **Sin Marcapasos:** El intervalo de confianza al 95% para el periodo entre pulsos está entre 1.006 y 1.218.

Los intervalos de confianza del periodo entre pulsos con y sin marcapasos no se solapan, lo que sugiere que hay una diferencia significativa en el periodo entre pulsos entre los dos grupos. El periodo entre pulsos es menor cuando se utiliza un marcapasos. Esto sugiere que el marcapasos tiene un impacto significativo en el periodo entre pulsos, logrando su propósito principal de regular el ritmo cardíaco, mientras que la intensidad de los pulsos permanece relativamente constante.