Intervalos de confianza

Jacobo Hirsch Rodriguez

2024-08-23

##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.

Para ello haz el gráfico de 150 intervalos de confianza obtenidos de la misma población. Guiáte de los siguientes pasos:: Haz la simulación de 150 medias de muestras de tamaño 150 extraídas de una población normal con miu = 70 y sigma = 9

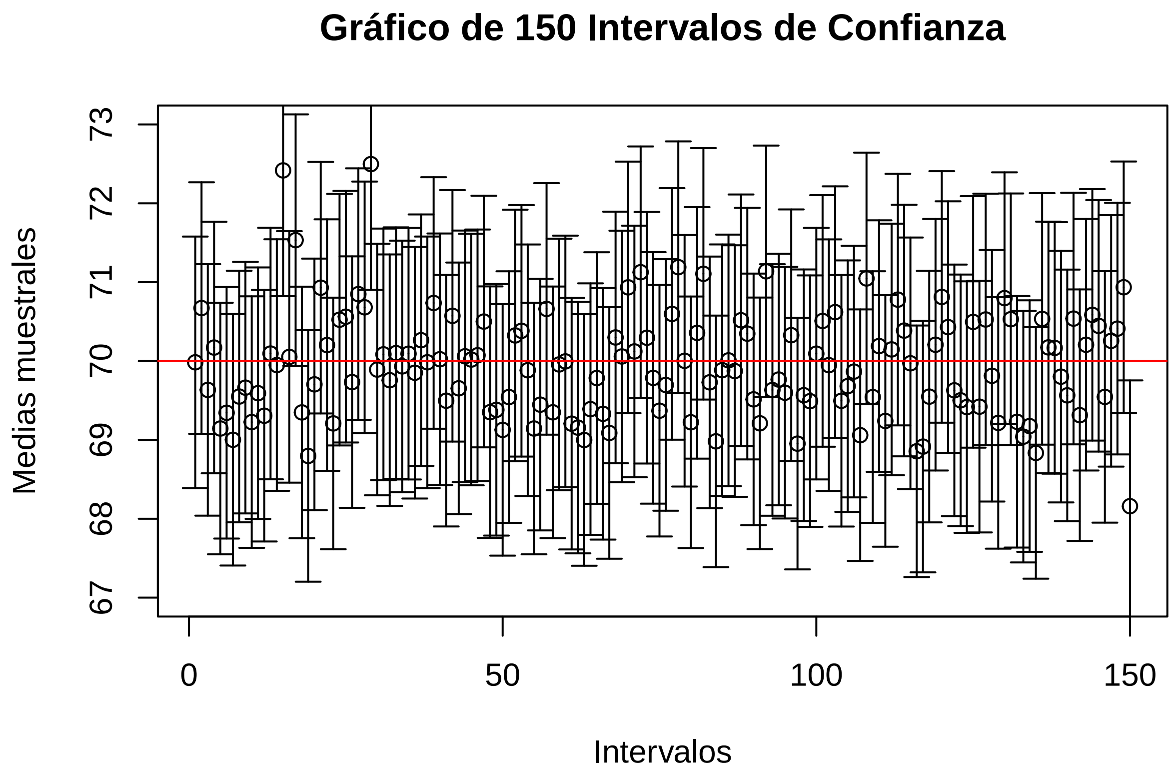
```
library(plotrix)

# Parámetros de la simulación
m <- 150  # número de muestras
n <- 150  # tamaño de la muestra
miu <- 70  # media poblacional
sigma <- 9  # desviación estándar poblacional
alfa <- 0.03  # nivel de significancia para un nivel de confianza del 97%

# Simulación de m medias muestrales
xb <- rnorm(m, miu, sigma/sqrt(n))  # medias muestrales simuladas
```

Calcula el margen de error con un nivel de confianza del 97% para esas medias.

```
# Cálculo del margen de error
E <- abs(qnorm(alfa/2)) * sigma / sqrt(n)  # Margen de error
```

Grafica las medias con el margen de error calculado en el inciso B. El comando plotCI calcula automáticamente los intervalos de confianza si le proporcionas las medias y el margen de error y Grafica la media poblacional (= 70) como una linea horizontal

```
# Gráfico de los intervalos de confianza
plotCI(1:m, xb, E, main="Gráfico de 150 Intervalos de Confianza",
       xlab="Intervalos", ylab="Medias muestrales", ylim=c(67, 73))
abline(h=miu, col="red")
```



Cuenta cuántos intervalos de confianza contienen a la verdadera media, ¿qué porcentaje representan? (Si quieres tener los 150 intervalos de confianza, calcula xb -E para las cotas inferiores y xb + E para las cotas superiores de los intervalos. Esto puede ayudarte a contar de manera automática cuantos intervalos contienen a la verdadera media)

```
# Cálculo de los límites de los intervalos de confianza
lower_bound <- xb - E
upper_bound <- xb + E

# Contar cuántos intervalos contienen la verdadera media
contain_true_mean <- sum(lower_bound <= miu & upper_bound >= miu)

# Calcular el porcentaje de intervalos que contienen la media verdadera
percentage <- contain_true_mean / m * 100

# Mostrar el resultado
contain_true_mean
```

```
## [1] 147
```

```
percentage
```

```
## [1] 98
```

#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.

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

```
# Datos
sigma <- 0.75
n1 <- 20
xbar1 <- 4.85
alpha1 <- 0.03

# Valor crítico Z para el 97% de confianza
z_critical <- qnorm(1 - alpha1 / 2)

# Margen de error
E1 <- z_critical * (sigma / sqrt(n1))

# Intervalo de confianza
lower_bound1 <- xbar1 - E1
upper_bound1 <- xbar1 + E1

lower_bound1
```

```
## [1] 4.486065
```

```
upper_bound1
```

```
## [1] 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

```
# Datos para la segunda muestra
n2 <- 16
xbar2 <- 4.56

# Margen de error para la segunda muestra
E2 <- z_critical * (sigma / sqrt(n2))

# Intervalo de confianza
lower_bound2 <- xbar2 - E2
upper_bound2 <- xbar2 + E2

lower_bound2
```

```
## [1] 4.153108
```

```
upper_bound2
```

```
## [1] 4.966892
```

¿Podemos afirmar que la porosidad del helio ha disminuido?

si ha disminuido por que ninguno de los dos intervalos contiene la media historica, un fuerte indicio de que cambió.

##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.

¿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?

```
# Datos para el cálculo del tamaño de muestra
alpha3 <- 0.05
z_critical3 <- qnorm(1 - alpha3 / 2)
E3 <- 0.2  # Queremos un ancho total de 0.4, por lo que E = 0.2

# Cálculo del tamaño de muestra
n_required1 <- (z_critical3 * sigma / E3) ^ 2
n_required1 <- ceiling(n_required1)  # Redondeo hacia arriba al entero más cercano

n_required1
```

```
## [1] 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%?

```
# Datos para el cálculo del tamaño de muestra con 99% de confianza
alpha4 <- 0.01
z_critical4 <- qnorm(1 - alpha4 / 2)
E4 <- 0.2  # Queremos un margen de error de 0.2

# Cálculo del tamaño de muestra
n_required2 <- (z_critical4 * sigma / E4) ^ 2
n_required2 <- ceiling(n_required2)  # Redondeo hacia arriba al entero más cercano

n_required2
```

```
## [1] 94
```

Problema 3. Con el archivo de datos de El Marcapasos Download El Marcapasos haz los intervalos de confianza para la media de las siguientes variables:

```
# Cargar el dataset
marcapasos <- read.csv("./marcapasos.csv") #leer la base de datos
```

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

```
# Función para calcular el intervalo de confianza al 95%
calc_confidence_interval <- function(data, variable) {
  mean_val <- mean(data[[variable]])
  sd_val <- sd(data[[variable]])
```

```r
  n <- length(data[[variable]])
  error_margin <- qt(0.975, df = n - 1) * sd_val / sqrt(n)

  lower_bound <- mean_val - error_margin
  upper_bound <- mean_val + error_margin

  return(c(mean = mean_val, lower = lower_bound, upper = upper_bound))
}

# Intervalos de confianza para 'Intensidad de pulso'
intensidad_sin_mp <- calc_confidence_interval(marcapasos[marcapasos$Marcapasos == "Sin MP", ], "Intensi
intensidad_con_mp <- calc_confidence_interval(marcapasos[marcapasos$Marcapasos == "Con MP", ], "Intensi

# Intervalos de confianza para 'Periodo entre pulsos'
periodo_sin_mp <- calc_confidence_interval(marcapasos[marcapasos$Marcapasos == "Sin MP", ], "Periodo.en
periodo_con_mp <- calc_confidence_interval(marcapasos[marcapasos$Marcapasos == "Con MP", ], "Periodo.en
```

```r
cat("intensidad sin mp: ",intensidad_sin_mp , "\n")
```

```
## intensidad sin mp:  0.207098 0.16993 0.2442661
```

```r
cat("intensidad con mp:",intensidad_con_mp ,  "\n")
```

```
## intensidad con mp: 0.1959412 0.1638035 0.2280788
```

```r
cat("periodo sin mp:", periodo_sin_mp,  "\n")
```

```
## periodo sin mp: 1.111765 1.002887 1.220643
```

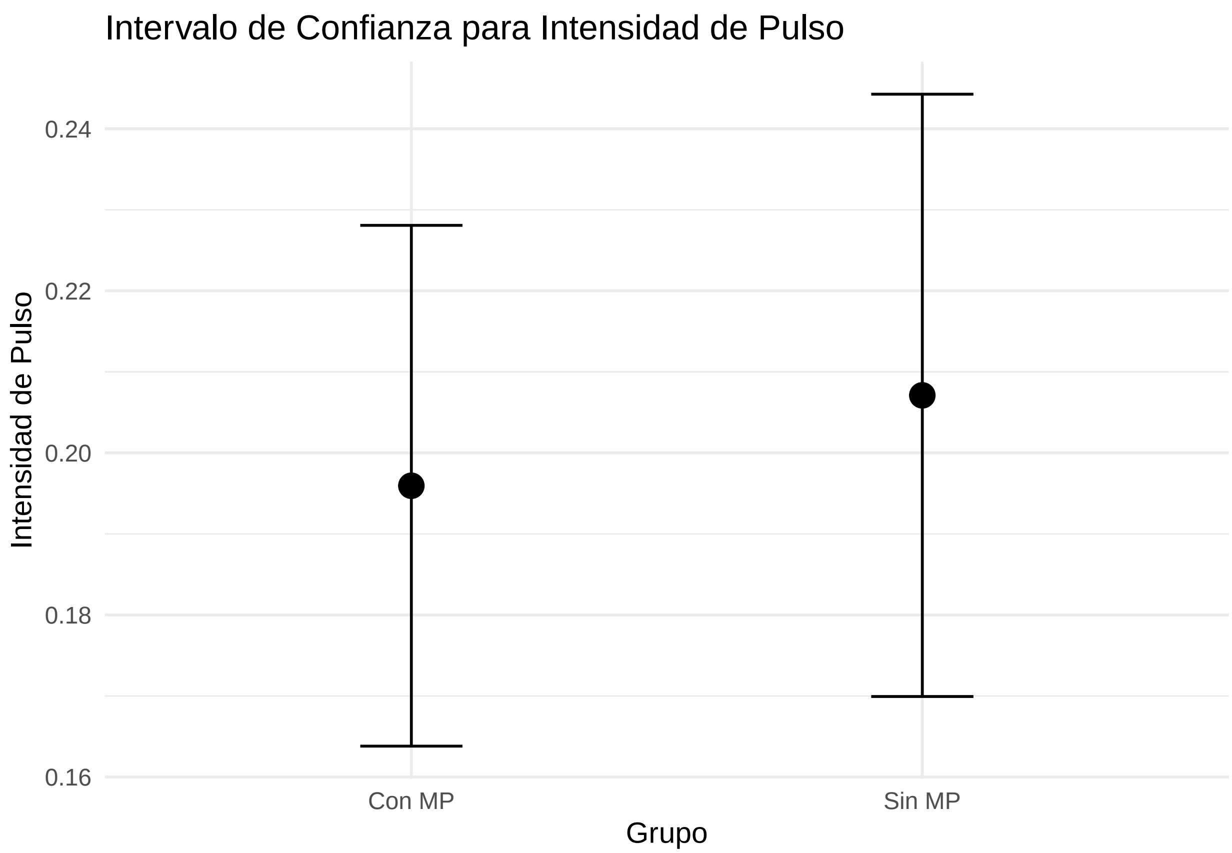
```r
cat("periodo con mp:", periodo_con_mp,  "\n")
```

```
## periodo con mp: 0.8911765 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

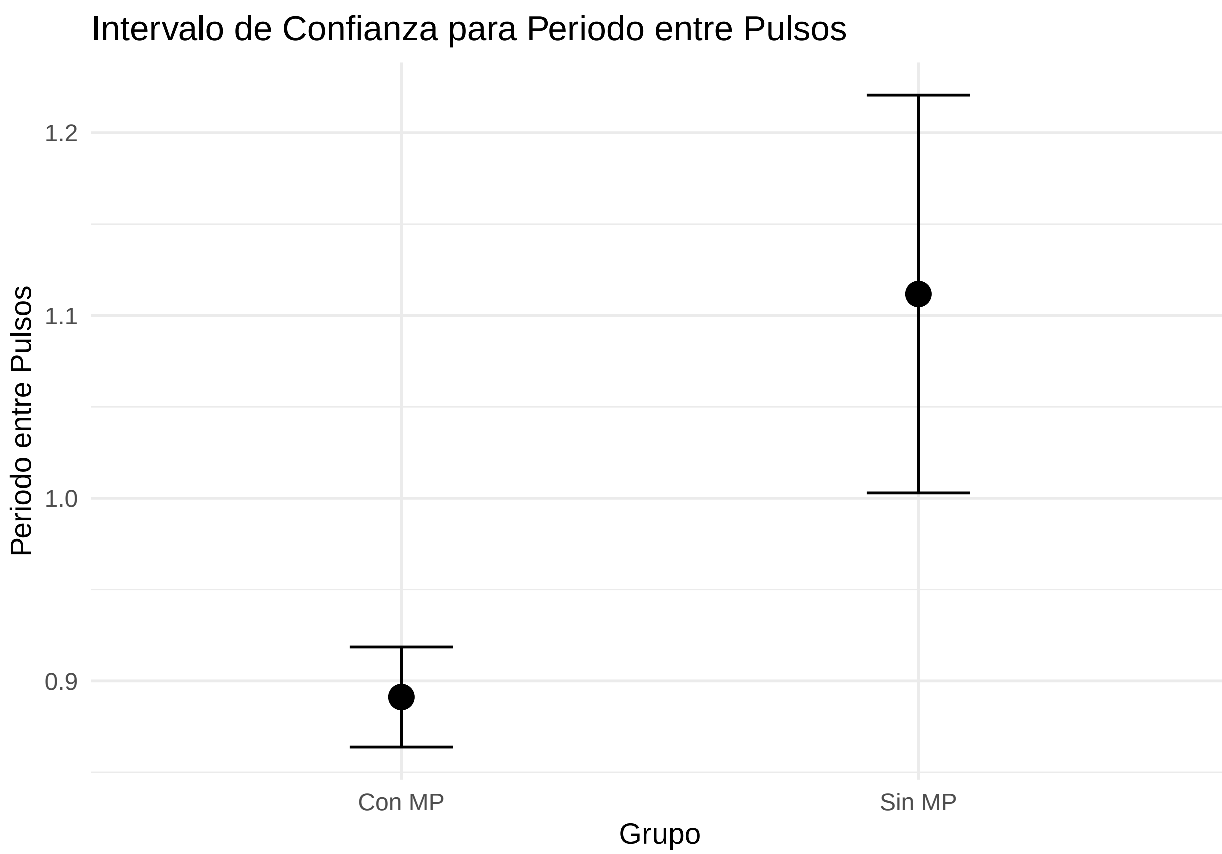
```r
library(ggplot2)

# Crear un dataframe para graficar
intensidad_df <- data.frame(
  Grupo = c("Sin MP", "Con MP"),
  Media = c(intensidad_sin_mp["mean"], intensidad_con_mp["mean"]),
  Inferior = c(intensidad_sin_mp["lower"], intensidad_con_mp["lower"]),
  Superior = c(intensidad_sin_mp["upper"], intensidad_con_mp["upper"])
)

# Graficar
ggplot(intensidad_df, aes(x = Grupo, y = Media)) +
  geom_point(size = 4) +
  geom_errorbar(aes(ymin = Inferior, ymax = Superior), width = 0.2) +
```

```r
  labs(title = "Intervalo de Confianza para Intensidad de Pulso",
       y = "Intensidad de Pulso") +
  theme_minimal()
```



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

```r
# Crear un dataframe para graficar
periodo_df <- data.frame(
  Grupo = c("Sin MP", "Con MP"),
  Media = c(periodo_sin_mp["mean"], periodo_con_mp["mean"]),
  Inferior = c(periodo_sin_mp["lower"], periodo_con_mp["lower"]),
  Superior = c(periodo_sin_mp["upper"], periodo_con_mp["upper"])
)

# Graficar
ggplot(periodo_df, aes(x = Grupo, y = Media)) +
  geom_point(size = 4) +
  geom_errorbar(aes(ymin = Inferior, ymax = Superior), width = 0.2) +
  labs(title = "Intervalo de Confianza para Periodo entre Pulsos",
       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

Básicamente, el marcapasos parece ser clave para hacer que los latidos ocurran más seguido, pero no tiene mucha influencia, o casi ninguna, en qué tan fuertes son esos latidos. Esto nos dice que lo que realmente hace el marcapasos, por lo menos con estos datos, es controlar qué tan rápido late el corazón, y no tanto la potencia de cada latido.