4. Intervalos de confianza

Datos del alumno

Luis Ángel Guzmán Iribe - A01741757

1. Resuelve las dos partes del problema "El misterioso Helio". Las dos partes del problema se encuentran al final de la presentación Intervalos de confianza con N. Concluye en el contexto del problema.

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.

```r
sigma <- 0.75
miu <- 5.3
n_muestra <- 20
media_muestra_1 <- 4.85
confianza <- 0.97
```

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

```r
error_estandar_1 <- sigma / sqrt(n_muestra)
z <- qnorm((1-confianza) / 2)

intervalo <- c(media_muestra_1 + z * error_estandar_1, media_muestra_1 - z * error_estandar_1)
intervalo
```

```
## [1] 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

```r
n_muestra_2 <- 16
media_muestra_2 <- 4.56

error_estandar_2 <- sigma / sqrt(n_muestra_2)
z <- qnorm((1-confianza) / 2)

intervalo_2 <- c(media_muestra_2 + z * error_estandar_2, media_muestra_2 - z * error_estandar_2)
intervalo_2
```

```
## [1] 4.153108 4.966892
```

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

Podemos afirmar con una confianza del 97%, que la porosidad media de la veta de helio ha disminuido, dado que ambos intervalos no incluyen la porosidad anterior (5.3), estos intervalos son (4.486065, 5.213935) para la primera muestra, y (4.153108, 4.966892) para la segunda, siendo que ninguno de estos intervalos incluyen la media anterior, podemos afirmar con seguridad que ha disminuido la porosidad del helio.

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?

Si despejamos n en la formula para el intervalo de confianza:

$$\mathrm{E} = \bar{x} \pm z_{\frac{\alpha}{2}} \left(\frac{\sigma}{\sqrt{n}} \right)$$

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

$$n = \left(\frac{\sigma \cdot z_{\frac{\alpha}{2}}}{E} \right)^2$$

Podemos calcular n

```
sigma <- 0.75
confianza <- 0.95
z <- qnorm((1-confianza) / 2)
E = 0.2

n = ((z*sigma)/(E))^2
n
```

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

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%?

```
sigma <- 0.75
confianza <- 0.99
z <- qnorm((1-confianza) / 2)
E = 0.2

n = ((z*sigma)/(E))^2
n
```

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

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

```
library(readr)
data <- read.csv("El marcapasos.csv")
```

```
intensidad_sin <- data$Intensidad.de.pulso[data$Marcapasos == "Sin MP"]
intensidad_con <- data$Intensidad.de.pulso[data$Marcapasos == "Con MP"]

int_intensidad_con <- t.test(intensidad_con, conf.level=0.95)
print(int_intensidad_con)
```

2.1 Intensidad de pulsos con y sin Marcapasos

```
##
##  One Sample t-test
```

```
##
## data:  intensidad_con
## t = 12.246, df = 50, p-value < 2.2e-16
## alternative hypothesis: true mean is not equal to 0
## 95 percent confidence interval:
##  0.1638035 0.2280788
## sample estimates:
## mean of x
## 0.1959412
```

```
int_intensidad_sin <- t.test(intensidad_sin, conf.level=0.95)
print(int_intensidad_sin)
```

```
##
##  One Sample t-test
##
## data:  intensidad_sin
## t = 11.192, df = 50, p-value = 3.182e-15
## alternative hypothesis: true mean is not equal to 0
## 95 percent confidence interval:
##  0.1699300 0.2442661
## sample estimates:
## mean of x
##  0.207098
```

```

periodo_sin <- data$Periodo.entre.pulsos[data$Marcapasos == "Sin MP"]
periodo_con <- data$Periodo.entre.pulsos[data$Marcapasos == "Con MP"]

int_periodo_con <- t.test(periodo_con, conf.level=0.95)
print(int_periodo_con)
```

2.2 Periodo entre pulso con y sin Marcapasos

```
##
##  One Sample t-test
##
## data:  periodo_con
## t = 65.37, df = 50, p-value < 2.2e-16
## alternative hypothesis: true mean is not equal to 0
## 95 percent confidence interval:
##  0.8637941 0.9185589
## sample estimates:
## mean of x
## 0.8911765
```

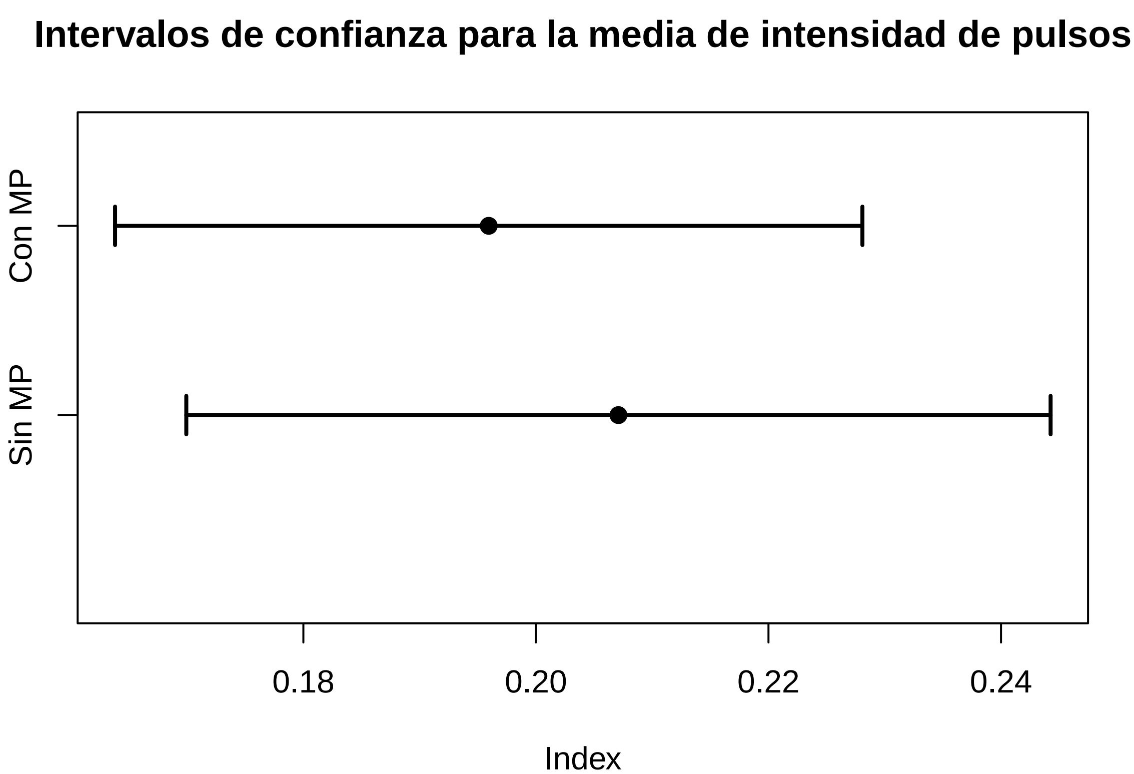
```
int_periodo_sin <- t.test(periodo_sin, conf.level=0.95)
print(int_periodo_sin)
```

```
##
##  One Sample t-test
##
## data:  periodo_sin
## t = 20.51, df = 50, p-value < 2.2e-16
## alternative hypothesis: true mean is not equal to 0
## 95 percent confidence interval:
```

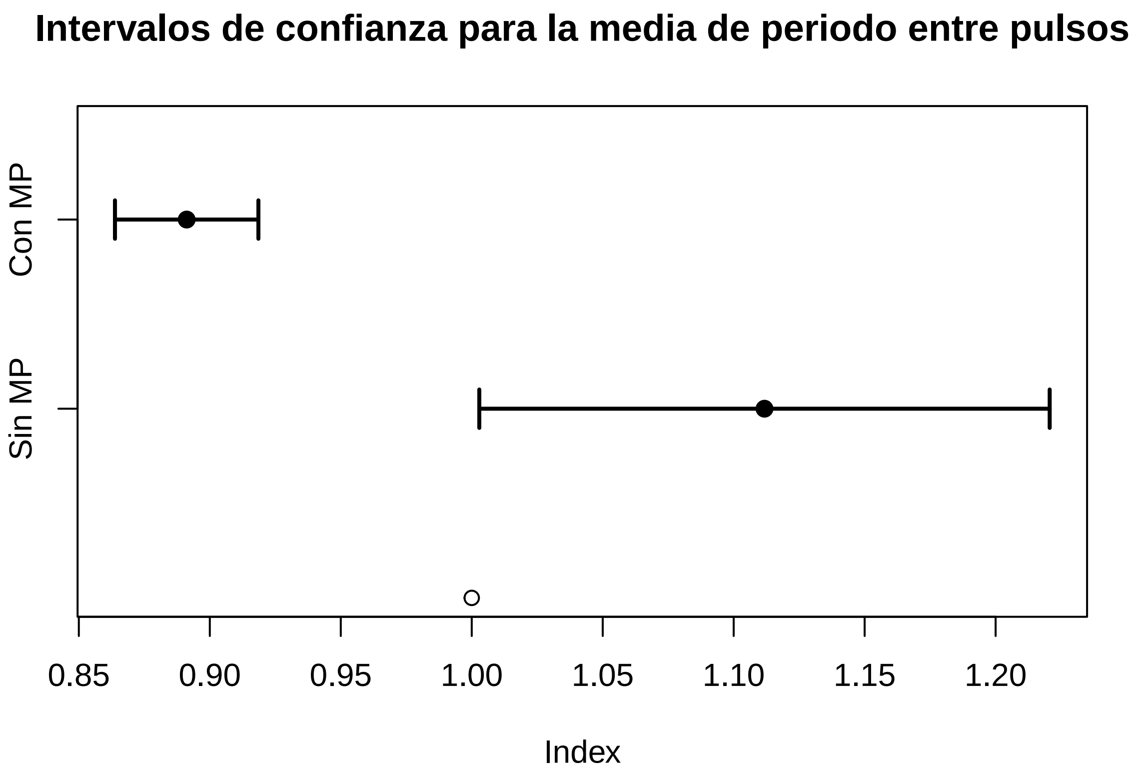
```
##  1.002887 1.220643
## sample estimates:
## mean of x
##  1.111765
```

3. Grafica los intervalos. En un gráfico la intensidad de pulso con y sin marcapasos y en otro gráfico el periodo entre pulso con y sin marcapasos. Interpreta el resultado.

```r
ints_intensidad <- data.frame(
  Marcapasos = c("Sin MP", "Con MP"),
  media = c(int_intensidad_sin$estimate, int_intensidad_con$estimate),
  inferior = c(int_intensidad_sin$conf.int[1], int_intensidad_con$conf.int[1]),
  superior = c(int_intensidad_sin$conf.int[2], int_intensidad_con$conf.int[2])
)

plot(0, ylim = c(0, 2.5), xlim = c(min(ints_intensidad$inferior), max(ints_intensidad$superior)),
     yaxt = "n", ylab = "")

axis(2, at <- c(1,2), labels <- ints_intensidad$Marcapasos )

arrows(ints_intensidad$inferior[1], 1, ints_intensidad$superior[1], 1,
       angle = 90, code = 3, length = 0.1, lwd = 2)
arrows(ints_intensidad$inferior[2], 2, ints_intensidad$superior[2], 2,
       angle = 90, code = 3, length = 0.1, lwd = 2)

points(ints_intensidad$media[1], 1, pch = 19, cex = 1.1)
points(ints_intensidad$media[2], 2, pch = 19, cex = 1.1)

title(main = "Intervalos de confianza para la media de intensidad de pulsos")
```



```r
ints_periodo <- data.frame(
  Marcapasos = c("Sin MP", "Con MP"),
  media = c(int_periodo_sin$estimate, int_periodo_con$estimate),
  inferior = c(int_periodo_sin$conf.int[1], int_periodo_con$conf.int[1]),
  superior = c(int_periodo_sin$conf.int[2], int_periodo_con$conf.int[2])
)

plot(0, ylim = c(0, 2.5), xlim = c(min(ints_periodo$inferior), max(ints_periodo$superior)),
     yaxt = "n", ylab = "")

axis(2, at <- c(1,2), labels <- ints_periodo$Marcapasos )

arrows(ints_periodo$inferior[1], 1, ints_periodo$superior[1], 1,
       angle = 90, code = 3, length = 0.1, lwd = 2)
arrows(ints_periodo$inferior[2], 2, ints_periodo$superior[2], 2,
       angle = 90, code = 3, length = 0.1, lwd = 2)

points(ints_periodo$media[1], 1, pch = 19, cex = 1.1)
points(ints_periodo$media[2], 2, pch = 19, cex = 1.1)

title(main = "Intervalos de confianza para la media de periodo entre pulsos")
```



Podemos apreciar que en la gráfica para los intervalos de confianza de la intensidad de los pulsos, estos se solapan en una considerable porsión, lo que nos lleva a creer que no existe una diferencia significativa entre ambos grupos, lo que podría interpretarse como que el marcapasos no es un elemento influyente para este parametro.

Por otro lado, podemos apreciar que en el periodo entre pusos el intervalo sin marcapasos no es solo considerablemente más largo, sino que no se solapa en ningún momento con el intervalo con marcapasos, que además es mucho más pequeño. Esto significa que la muestra que utiliza el marcapasos es mucho más precisa, y requiere de un rango menor de datos para generar confianza, lo que sugiere que el marcapasos es importante al momento de realizar mediciones para periodos entre pulsos.