

Actividad7_intervalos_confianza

Rodolfo Jesús Cruz Rebollar

2024-08-22

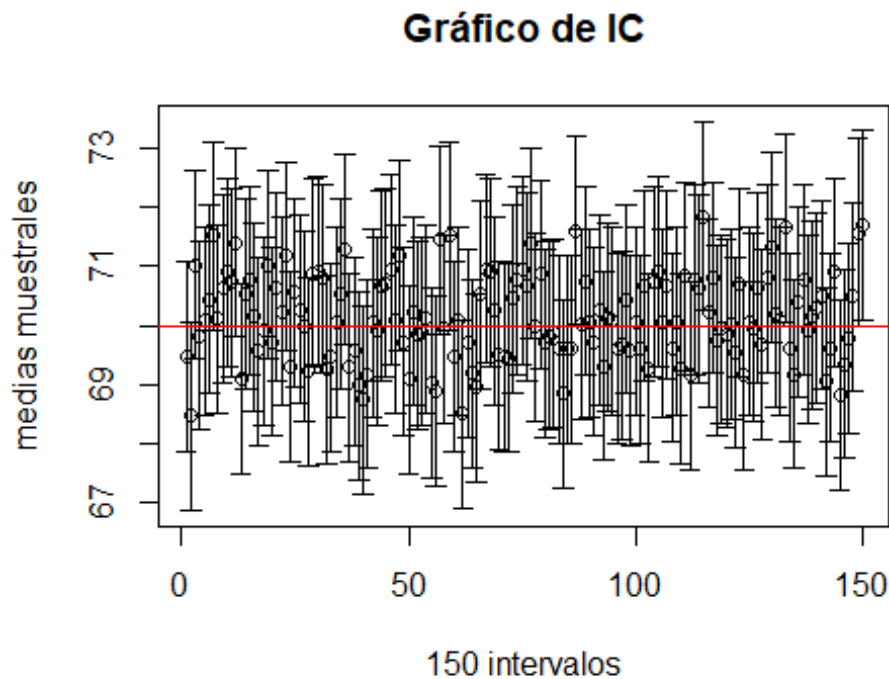
Problema 1

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:

```
# Librería para graficar Los intervalos de confianza correspondientes a  
cada una de las  
# muestras, además de la media poblacional  
  
library(plotrix)  
  
# Definir el tamaño de las muestras a extraer  
  
n = 150  
  
# Definir la media de la población normal  
  
miu = 70  
  
# Definir la desviación estándar de la población normal  
  
sigma = 9  
  
# Definir el nivel de significancia para los intervalos de confianza  
  
alfa = 0.03  
  
xb = rnorm(n, miu, sigma/sqrt(n)) # simulación de una muestra de tamaño  
n=150  
  
E = abs(qnorm(alfa/2))*sigma/sqrt(n) # Margen de error  
  
m = 150 # número de muestras de tamaño n=150  
  
# Graficar las muestras con sus intervalos de confianza y trazar una  
línea recta roja  
# horizontal sobre el valor de la media poblacional
```

```
plotCI(1:m, xb, E, main="Gráfico de IC", xlab="150 intervalos", ylab="medias muestrales")
```

```
abline(h=miu, col="red") # Media poblacional
```



```
# Calcular cotas inferiores de Los 150 intervalos de confianza
```

```
cotas_inferiores = xb - E
```

```
# Buscar cuántas cotas inferiores son superiores a La media miu = 70
```

```
cinf_sup_miu= length(cotas_inferiores[cotas_inferiores > miu])
```

```
cinf_sup_miu
```

```
## [1] 4
```

```
# Calcular cotas superiores de Los 150 intervalos de confianza
```

```
cotas_superiores = xb + E
```

```
# Buscar cuántas cotas superiores son inferiores a La media miu = 70
```

```
csup_inf_miu = length(cotas_superiores[cotas_superiores < miu])
```

```
csup_inf_miu
```

```
## [1] 0
```

En el gráfico anterior de los 150 intervalos de confianza, además del cálculo de la cota inferior y superior para cada uno de ellos, es posible observar que de los 150 intervalos de confianza graficados en total, 146 de ellos sí contienen a la verdadera media $\mu = 70$, mientras que solamente 4 de los 150 intervalos no contienen a dicha media, por lo tanto, el porcentaje de intervalos de confianza que sí contienen a la media verdadera es del 98%.

Problema 2

Resuelve las dos partes del problema “El misterioso Helio”.

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.

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
```

```
x_barra1 = 4.85
```

```
n1 = 20
```

```
E1 = abs(qnorm(0.03/2)) * sigma / sqrt(n1)
```

```
A1 = x_barra1 - E1
```

```
B1 = x_barra1 + E1
```

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

```
## La verdadera media actual 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 ¿Podemos afirmar que la porosidad del helio ha disminuido?

```
sigma = 0.75

alfa = 0.03

x_barra2 = 4.56

n2 = 16

E2 = abs(qnorm(0.03/2)) * sigma / sqrt(n2)

A2 = x_barra2 - E2

B2 = x_barra2 + E2

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

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

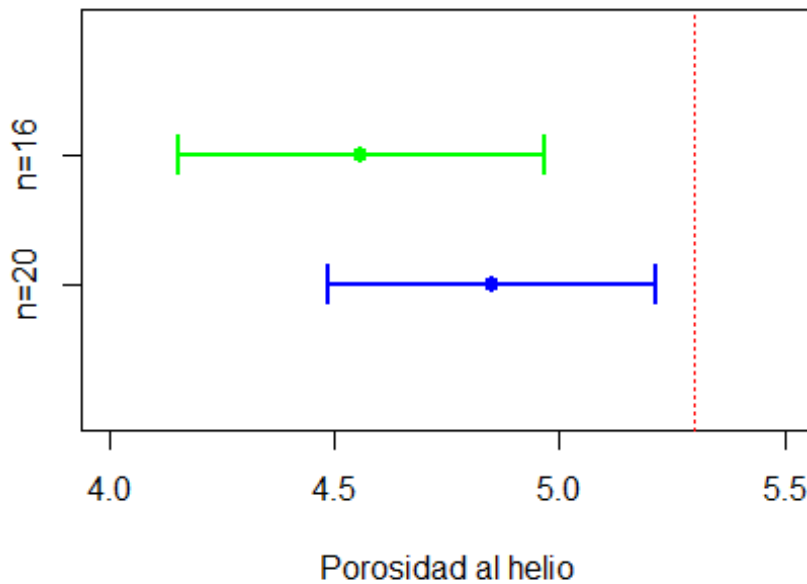
Gráfico de intervalos de confianza para n=20 y n=16 con 97% de confianza

```
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(x_barra1, 1, pch=19, cex=1.1, col="blue")
points(x_barra2, 2, pch=19, cex=1.1, col="green")

abline(v = 5.3, lty=3, col="red")
```



En el gráfico anterior de los intervalos de confianza, se puede apreciar que el intervalo de confianza correspondiente a la muestra de tamaño 20 con 97% de confianza (en color azul), tiene el punto central (valor del estadístico \bar{x} de la muestra) desplazado más hacia la derecha en comparación con el valor del estadístico del intervalo de confianza de la muestra de tamaño 16 con 97% de confianza (en color verde), mismo que se encuentra más a la izquierda que el estadístico del intervalo azul, motivo por el cual, al disminuir ligeramente el tamaño de la muestra de 20 a 16, se observa que de igual forma el valor del estadístico \bar{x} también disminuye, por lo tanto, se afirma que la porosidad del helio sí 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 ($\sigma = 0.75$).

¿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?

```
sigma = 0.75
```

```
alfa = 0.05
```

```
n = (qnorm(alfa / 2) * sigma / 0.2) ^ 2
```

```
cat("El tamaño de la muestra debe ser: ", n)
```

```
## El tamaño de la muestra debe ser: 54.02051
```

¿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
```

```
alfa = 0.01
```

```
n = (qnorm(alfa / 2) * sigma / 0.2) ^ 2
```

```
cat("El tamaño de la muestra debe ser: ", n)
```

```
## El tamaño de la muestra debe ser: 93.30323
```

Problema 3

Nota: se utilizará un nivel de significancia α de 0.05 (5%).

Además, para calcular el margen de error E que será necesario sumar o restar al valor de la media para calcular los intervalos de confianza, se encuentra dado por la siguiente expresión:

$$E = \frac{Z_{\alpha/2} \sigma_x}{\sqrt{n}}$$

Por lo tanto, los intervalos de confianza estarán dados por la expresión:

$$\bar{x}_0 - E \leq \bar{x}_0 \leq \bar{x}_0 + E$$

```
# Importar Los datos del archivo csv
```

```
marcapasos = read.csv("El marcapasos.csv")
```

```
# Verificar La correcta importación de Los datos
```

```
head(marcapasos)
```

```
##   Periodo.entre.pulsos Intensidad.de.pulso Marcapasos
## 1                1.2             0.131      Sin MP
## 2                0.9             0.303      Sin MP
## 3                0.9             0.297      Sin MP
## 4                0.8             0.416      Sin MP
## 5                0.7             0.585      Sin MP
## 6                1.2             0.126      Sin MP
```

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

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

Intervalo de confianza para la intensidad media de pulsos con marcapasos

```
# Calcular La intensidad media o promedio de pulsos con marcapasos

iprom_MP = mean(marcapasos[marcapasos$Marcapasos == "Con MP",
]$Intensidad.de.pulso)

# Calcular La desviación estándar de pulsos con marcapasos

isd_MP = sd(marcapasos[marcapasos$Marcapasos == "Con MP",
]$Intensidad.de.pulso)

# Definir el valor de alfa = 0.05 para el intervalo de confianza

alfa = 0.05

# Definir la cantidad de registros del dataframe

n = nrow(marcapasos[marcapasos$Marcapasos == "Con MP", ])

# Calcular el margen de error E

E1 = abs(qnorm(alfa / 2)) * isd_MP / sqrt(n)

# Mostrar intervalo de confianza para la media de pulsos con marcapasos

cat("La verdadera intensidad media de pulsos con marcapasos está entre",
    iprom_MP - E1, "y", iprom_MP + E1)

## La verdadera intensidad media de pulsos con marcapasos está entre
0.1645811 y 0.2273013
```

Intervalo de confianza para la intensidad media de pulsos sin marcapasos

```
# Calcular La intensidad media o promedio de pulsos sin marcapasos

iprom_sin_MP = mean(marcapasos[marcapasos$Marcapasos == "Sin MP",
]$Intensidad.de.pulso)

# Calcular La desviación estándar de pulsos sin marcapasos

isd_sin_MP = sd(marcapasos[marcapasos$Marcapasos == "Sin MP",
]$Intensidad.de.pulso)
```

```

# Definir el valor de alfa = 0.05 para el intervalo de confianza
alfa = 0.05

# Definir la cantidad de registros del dataframe
n = nrow(marcapasos[marcapasos$Marcapasos == "Sin MP", ])

# Calcular el margen de error E
E2 = abs(qnorm(alfa / 2)) * isd_sin_MP / sqrt(n)

# Mostrar intervalo de confianza para la media de pulsos sin marcapasos
cat("La verdadera intensidad media de pulsos sin marcapasos está entre",
    iprom_sin_MP - E2, "y", iprom_sin_MP + E2)

## La verdadera intensidad media de pulsos sin marcapasos está entre
0.1708292 y 0.2433669

```

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

IC para el periodo entre pulso con marcapasos

```

# Calcular el periodo entre pulso promedio con marcapasos
period_prom_MP = mean(marcapasos[marcapasos$Marcapasos == "Con MP",
]$Periodo.entre.pulsos)

# Calcular la desviación estándar de periodo entre pulso con marcapasos
period_sd_MP = sd(marcapasos[marcapasos$Marcapasos == "Con MP",
]$Periodo.entre.pulsos)

# Definir el valor de alfa = 0.05 para el intervalo de confianza
alfa = 0.05

# Definir la cantidad de registros del dataframe
n = nrow(marcapasos[marcapasos$Marcapasos == "Con MP", ])

# Calcular el margen de error E
E3 = abs(qnorm(alfa / 2)) * period_sd_MP / sqrt(n)

# Mostrar intervalo de confianza para la media de periodo entre pulso con
marcapasos

```



```
cat("La verdadera media de periodo entre pulso con marcapasos está
entre",
    period_prom_MP - E3, "y", period_prom_MP + E3)

## La verdadera media de periodo entre pulso con marcapasos está entre
0.8644566 y 0.9178964
```

IC para el periodo entre pulso sin marcapasos

```
# Calcular el periodo entre pulso promedio sin marcapasos

period_prom_SinMP = mean(marcapasos[marcapasos$Marcapasos == "Sin MP",
]$Periodo.entre.pulsos)

# Calcular la desviación estándar de periodo entre pulso sin marcapasos

period_sd_SinMP = sd(marcapasos[marcapasos$Marcapasos == "Sin MP",
]$Periodo.entre.pulsos)

# Definir el valor de alfa = 0.05 para el intervalo de confianza

alfa = 0.05

# Definir la cantidad de registros del dataframe

n = nrow(marcapasos[marcapasos$Marcapasos == "Sin MP", ])

# Calcular el margen de error E

E4 = abs(qnorm(alfa / 2)) * period_sd_SinMP / sqrt(n)

# Mostrar intervalo de confianza para la media de periodo entre pulso sin
marcapasos

cat("La verdadera media de periodo entre pulso sin marcapasos está
entre",
    period_prom_SinMP - E4, "y", period_prom_SinMP + E4)

## La verdadera media de periodo entre pulso sin marcapasos está entre
1.005521 y 1.218009
```

Grafica los intervalos de confianza obtenidos en “El marcapasos”

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

```
# Crear un eje cartesiano cuyos límites inferior y superior sean más
pequeños y
# grandes que el mínimo y el máximo de los intervalos y además cuyo
límite superior
# hace alusión a la cantidad de intervalos de confianza a graficar
```

```

plot(0, ylim = c(0, 2 + 1), xlim = c(0.1, 0.3), yaxt = "n", ylab = "",
     xlab="Intensidad de pulso")

axis(2, at = c(1, 2), labels = c("Con MP", "Sin MP"))

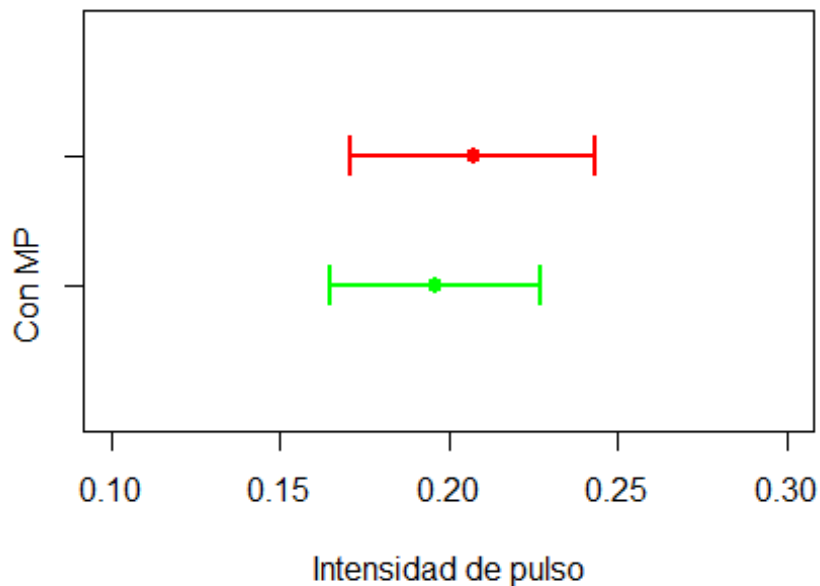
# Graficar intervalo de confianza para intensidad promedio con marcapasos
arrows(iprom_MP - E1, 1, iprom_MP + E1, 1, angle=90, code=3, length =
0.1,
      lwd = 2, col = "green")

# Graficar intervalo de confianza para intensidad promedio sin marcapasos
arrows(iprom_sin_MP - E2, 2, iprom_sin_MP + E2, 2, angle=90, code=3,
      length = 0.1, lwd = 2, col = "red")

# Punto para la media del primer intervalo (intensidad con marcapasos)
points(iprom_MP, 1, pch=19, cex=1.1, col="green")

# Punto para la media del segundo intervalo (intensidad sin marcapasos)
points(iprom_sin_MP, 2, pch=19, cex=1.1, col="red")

```



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

```
# Crear un eje cartesiano cuyos límites inferior y superior sean más
# pequeños y
# grandes que el mínimo y el máximo de los intervalos y además cuyo
# límite superior
# hace alusión a la cantidad de intervalos de confianza a graficar

plot(0, ylim = c(0, 2 + 1), xlim = c(0.8, 1.3), yaxt = "n", ylab = "",
     xlab="Periodo entre pulso")

axis(2, at = c(1, 2), labels = c("Con MP", "Sin MP"))

# Graficar intervalo de confianza para periodo entre pulso promedio con
# marcapasos

arrows(period_prom_MP - E3, 1, period_prom_MP + E3, 1, angle=90, code=3,
       length = 0.1,
       lwd = 2, col = "green")

# Graficar intervalo de confianza para periodo entre pulso promedio sin
# marcapasos

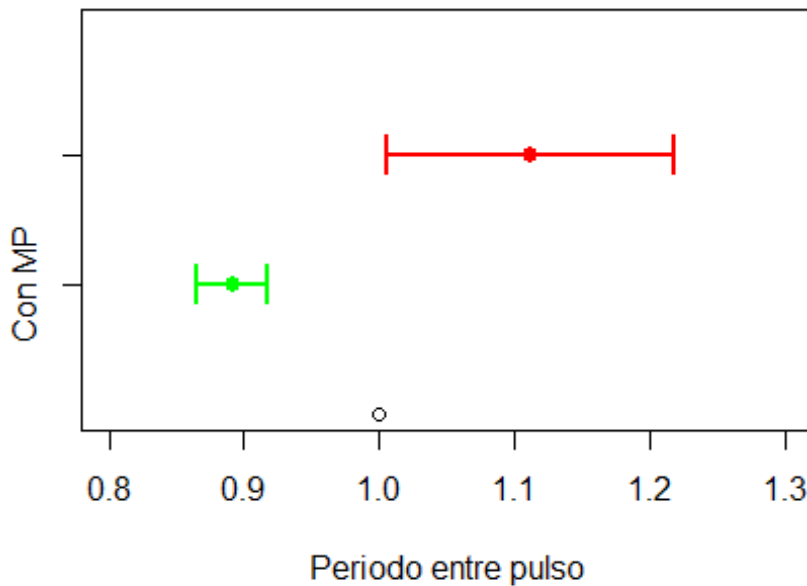
arrows(period_prom_SinMP - E4, 2, period_prom_SinMP + E4, 2, angle=90,
       code=3,
       length = 0.1, lwd = 2, col = "red")

# Punto para la media del primer intervalo (periodo entre pulso con
# marcapasos)

points(period_prom_MP, 1, pch=19, cex=1.1, col="green")

# Punto para la media del segundo intervalo (periodo entre pulso sin
# marcapasos)

points(period_prom_SinMP, 2, pch=19, cex=1.1, col="red")
```



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

En términos generales, en los gráficos de los intervalos de confianza realizados previamente, es posible apreciar que en cuanto a los intervalos correspondientes a la variable de intensidad de pulso, el intervalo en color rojo (correspondiente a datos sin marcapasos), tiene un ancho ligeramente inferior al ancho del intervalo rojo de la variable periodo entre pulso, además de que también el intervalo color verde (correspondiente a datos con marcapasos) de la variable intensidad de pulso tiene una anchura considerablemente superior a la anchura del intervalo color verde de la variable periodo entre pulso, además, de forma general, se evidencia que los valores que toma la variable intensidad de pulso son considerablemente más bajos que aquellos que toma la variable periodo entre pulso, dado que los valores de la primer variable (intensidad de pulso) se encuentran entre 0.005 y 0.585, mientras que los valores del periodo entre pulso se encuentran entre 0.690 y 2.8. Lo anterior indica que en cuanto a la intensidad de pulso, es posible ver que ambos intervalos (con y sin marcapasos) se traslapan entre sí parcialmente, lo cual indica que existe un 95% de confianza de que la media de la intensidad del pulso con marcapasos así como la media de intensidad de pulso sin marcapasos, se encuentren ambas comprendidas entre una cierta región común entre ambos intervalos, por lo cual existe cierta probabilidad de que incluso ambas medias sean iguales o en su defecto, muy cercanas entre ellas, sin embargo, también existe la probabilidad de que no sea el caso, además, el gráfico de los intervalos para la variable intensidad de pulso también indica que en 95 de cada 100 estudios practicados a pacientes que usan marcapasos, la intensidad

promedio de pulso estará situada entre 0.16 y 0.22, mientras que en 95 de 100 estudios realizados a pacientes que no usan marcapasos, la media de la intensidad de su pulso estará situada entre 0.17 y 0.24. Por otro lado, en relación a la variable de periodo entre pulso, es posible interpretar que en 95 de cada 100 estudios practicados a pacientes con marcapasos, el periodo promedio entre sus pulsos estará ubicado entre 0.86 y 0.91, mientras que en esos mismos 95 de cada 100 estudios realizados pero a pacientes que no utilizan marcapasos, el promedio del periodo entre sus pulsos estará situado entre 1 y 1.21. Finalmente, es posible concluir que en cuanto a la intensidad del pulso, cuando los pacientes no tienen marcapasos, se aprecia que la intensidad de su pulso es significativamente mayor, lo cual ocasiona a su vez que la intensidad promedio de pulso en este grupo de pacientes también sea considerablemente mayor, mientras que por el contrario, en pacientes que sí tienen marcapasos, dicha intensidad de pulso medianamente menor a la de los pacientes sin marcapasos, por lo que en este segundo grupo de pacientes (con marcapasos), el promedio de la intensidad de pulso también disminuye. Adicionalmente, en cuanto a la variable referente al periodo entre pulsos, se aprecia que en aquellos pacientes sin marcapasos, el promedio del periodo entre pulsos es mucho mayor que en el caso de aquellos otros pacientes que sí usan marcapasos, además de que también el periodo entre pulsos es considerablemente mayor en personas sin marcapasos, mientras que por el contrario, en personas que sí los utilizan, el periodo entre pulsos es significativamente menor además de que presenta mucha menor variación que el de los personas sin marcapasos que presenta una variación mucho mayor, por lo que en base a todo lo anterior, se puede concluir que el hecho de usar marcapasos, representa mayormente un apoyo para aquellas personas que tienen padecimientos cardíacos, ya que las capacidades del corazón se ven significativamente reducidas, lo cual se refleja por medio de una disminución de los valores de las variables analizadas en el análisis de los gráficos de los intervalos de confianza realizados para ambas variables.