



# Tecnológico de Monterrey

*Instituto Tecnológico y de Estudios Superiores de Monterrey*

## Actividad M1. Intervalos

**TC3006C.101 Inteligencia artificial avanzada para la ciencia de  
datos I**

### Profesores:

*Ivan Mauricio Amaya Contreras*

*Blanca Rosa Ruiz Hernandez*

*Antonio Carlos Bento*

*Frumencio Olivas Alvarez*

*Hugo Terashima Marín*

### Alumno:

*Alberto H Orozco Ramos – A00831719*

23 de Agosto de 2023

# Instrucciones

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.

```
In [ ]: # Cargamos el lenguaje de R para utilizarlo en Google Colab
        %load_ext rpy2.ipynthon
```

```
In [ ]: %%R

# Instalamos los paquetes necesarios a utilizar dentro de todo el Colab
install.packages('BSDA')

library(BSDA)
```

## 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. para el promedio de porosidad para evaluar si ha disminuido.

1. Haga una estimación por intervalo con una confianza del 97%

```
In [ ]: %%R

# Número de elementos tomados de la muestra
n <- 20
# Media Arimética de la Muestra
sample_m <- 4.85
# Desviación Estándar de la Población
pop_std_dev <- 0.75
# Nivel de Confianza
conf_lvl <- 0.97

# Calculamos la puntuación t para el nivel de confianza proporcionado y el grado de
z <- abs(qnorm((1 + conf_lvl) / 2))
# Calculamos el error estándar de la media
E <- (z * pop_std_dev) / sqrt(n)
# Calculamos el intervalo de confianza
confidence_interval <- c(sample_m - E, sample_m + E)

cat("Intervalo de confianza: ", confidence_interval)
```

Intervalo de confianza: 4.486065 5.213935

2. 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

```
In [ ]: %%R

# Número de elementos tomados de La muestra
n <- 16
# Media Arimética de La Muestra
sample_m <- 4.85
# Desviación Estándar de La Población
pop_std_dev <- 0.75
# Nivel de Confianza
conf_lvl <- 0.97

# Calculamos la puntuación t para el nivel de confianza proporcionado y el grado de
z <- abs(qnorm((1 + conf_lvl) / 2))
# Calculamos el error estándar de la media
E <- (z * pop_std_dev) / sqrt(n)
# Calculamos el intervalo de confianza
confidence_interval <- c(sample_m - E, sample_m + E)

cat("Intervalo de confianza: ", confidence_interval)
```

Intervalo de confianza: 4.443108 5.256892

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

En base a los resultados obtenidos en los 2 puntos anteriores, y comparándolos con la media de población original, podemos llegar a la conclusión de que resulta muy difícil determinar que la porosidad del helio ha disminuido, inclusive lo sería si quisiéramos determinar si ha aumentado debido a que el intervalo de confianza abarca un rango lo suficientemente grande como para no permitirnos tener la certeza de que la porosidad disminuyó. Ambos intervalos se encuentran en extremos muy opuestos, siendo que el mínimo tiende a ser 4.44 o 4.48 y el máximo a 5.21 o 5.25. Si estos intervalos fueran más pequeños, probablemente tendríamos una mejor noción de qué es lo que ha pasado con la porosidad del helio en los últimos años.

## 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. promedio verdadera dentro de 0.2 unidades alrededor de la

1. ¿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?
2. ¿Qué tamaño de muestra necesita para estimar la porosidad

media muestral con una confianza de 99%?

```
In [ ]: %%R

# Desviación Estándar de La Población
pop_std_dev <- 0.75
# Nivel de Confianza
conf_lvl <- 0.95
# Ancho del intervalo deseado
E <- 0.4

# Calculamos el puntaje z para el valor de confianza deseado
z <- abs(qnorm((1 + conf_lvl) / 2))

n <- ceiling((2 * z * pop_std_dev / E)^2)

cat("Tamaño de la muestra deseado: ", n)
```

Tamaño de la muestra deseado: 55

```
In [ ]: %%R

# Desviación Estándar de La Población
pop_std_dev <- 0.75
# Nivel de Confianza
conf_lvl <- 0.99
# Ancho del intervalo deseado
E <- 0.2

# Calculamos el puntaje z para el valor de confianza deseado
z <- abs(qnorm((1 + conf_lvl) / 2))

n <- ceiling((z * pop_std_dev / E)^2)

cat("Tamaño de la muestra deseado: ", n)
```

Tamaño de la muestra deseado: 94

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

- Intensidad de pulsos con y sin Marcapasos
- Periodo entre pulso con y sin Marcapasos

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.

```
In [ ]: # Montamos Google Drive dentro de Colab
from google.colab import drive
drive.mount('/content/drive')
```

Mounted at /content/drive

```
In [ ]: # Obtenemos la ruta del archivo CSV a utilizar
file_path = '/content/drive/MyDrive/Colab Notebooks/El marcapasos.csv'
```

```
In [ ]: # Pasamos la variable de la ruta de los datos a R
%Rpush file_path
```

```
In [ ]: # Extraemos los datos del archivo CSV
%%R

M <- read.csv(file_path)
```

## Intensidad de pulsos con y sin Marcapasos

```
In [ ]: %%R

# Obtenemos los valores correspondientes a intensidad de pulso
pace_maker <- M[['Marcapasos']]
pulse_intensity <- M[['Intensidad.de.pulso']]
pulse_intensity_wMP <- pulse_intensity[pace_maker == 'Con MP']
pulse_intensity_woMP <- pulse_intensity[pace_maker == 'Sin MP']

# d = sd(pulse_intensity_filtered)
# n <- length(pulse_intensity_filtered)
# m <- mean(pulse_intensity_filtered)
# alpha = 0.05
# t = abs(qt(alpha/2))

# Intervalo
# E <- (t * d) / sqrt(n)
# Li <- m - E
# LiS <- m + E

# Calculamos el intervalo de confianza utilizando t.test()
conf_interval_1 <- t.test(pulse_intensity_woMP, conf.level=0.95)$conf.int

cat("Intervalo de confianza para Intensidad de Pulso (Sin MP): \n")
cat("Límite Inferior: ", conf_interval_1[1], "\n")
cat("Límite Superior: ", conf_interval_1[2], "\n")
cat("\n")

# Calculamos el intervalo de confianza utilizando t.test()
conf_interval_2 <- t.test(pulse_intensity_wMP, conf.level=0.95)$conf.int

cat("Intervalo de confianza para Intensidad de Pulso (Sin MP): \n")
cat("Límite Inferior: ", conf_interval_2[1], "\n")
cat("Límite Superior: ", conf_interval_2[2], "\n")
cat("\n")

# Grafica los intervalos de confianza para intensidad de pulso
plot(0, ylim = c(0, 3), xlim = c(min(conf_interval_1), max(conf_interval_2)), yaxt
axis(2, at = c(1, 2), labels = c("Sin MP", "Con MP"))

arrows(conf_interval_1[1], 1, conf_interval_1[2], 1, angle = 90, code = 3, length =
```

```
arrows(conf_interval_2[1], 2, conf_interval_2[2], 2, angle = 90, code = 3, length =
points(mean(pulse_intensity_woMP), 1, pch = 19, cex = 1.1)
points(mean(pulse_intensity_wMP), 2, pch = 19, cex = 1.1)
```

Intervalo de confianza para Intensidad de Pulso (Sin MP):

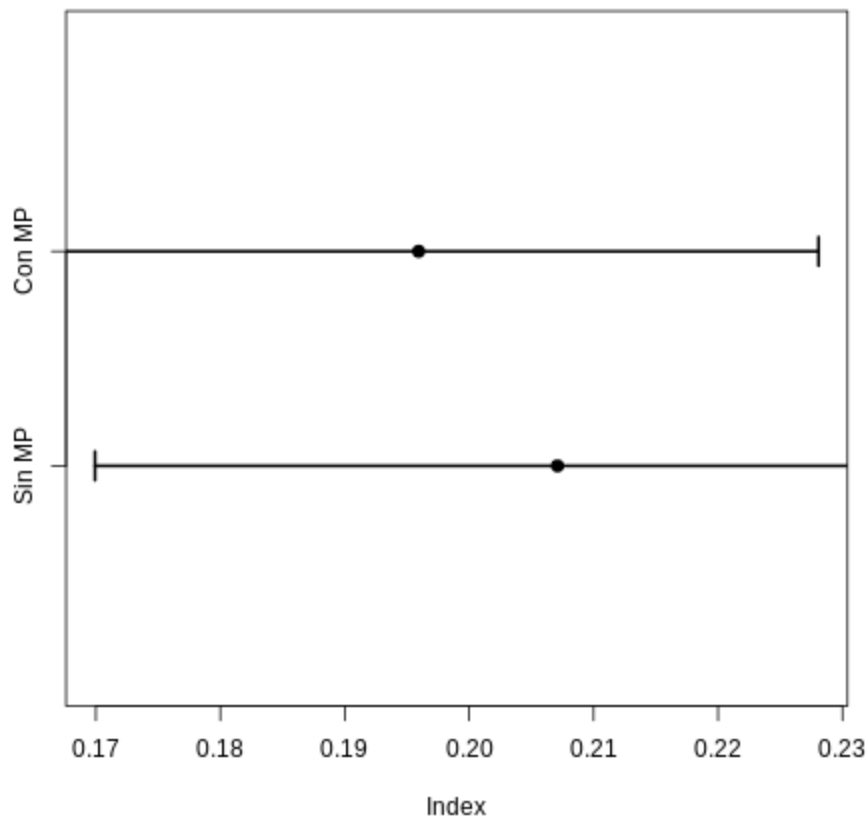
Límite Inferior: 0.16993

Límite Superior: 0.2442661

Intervalo de confianza para Intensidad de Pulso (Sin MP):

Límite Inferior: 0.1638035

Límite Superior: 0.2280788



## Periodo entre pulso con y sin Marcapasos

```
In [ ]: %%R

# Obtenemos Los valores correspondientes a periodo entre pulsos
period_btw_pulses <- M[['Periodo.entre.pulsos']]
period_btw_pulses_wMP <- period_btw_pulses[pace_maker == 'Con MP']
period_btw_pulses_woMP <- period_btw_pulses[pace_maker == 'Sin MP']

# d = sd(pulse_intensity_filtered)
# n <- length(pulse_intensity_filtered)
# m <- mean(pulse_intensity_filtered)
# alpha = 0.05
```

```

# t = abs(qt(alpha/2))

# Intervalo
# E <- (t * d) / sqrt(n)
# li <- m - E
# liS <- m + E

# Calculamos el intervalo de confianza utilizando t.test()
conf_interval_3 <- t.test(period_btw_pulses_woMP, conf.level=0.95)$conf.int

cat("Intervalo de confianza para Periodos entre Pulsos (Sin MP): \n")
cat("Límite Inferior: ", conf_interval_3[1], "\n")
cat("Límite Superior: ", conf_interval_3[2], "\n")
cat("\n")

# Calculamos el intervalo de confianza utilizando t.test()
conf_interval_4 <- t.test(period_btw_pulses_wMP, conf.level=0.95)$conf.int

cat("Intervalo de confianza para Periodos entre Pulsos (Con MP): \n")
cat("Límite Inferior: ", conf_interval_4[1], "\n")
cat("Límite Superior: ", conf_interval_4[2], "\n")
cat("\n")

# Grafica los intervalos de confianza para periodo entre pulsos
plot(0, ylim = c(0, 3), xlim = c(min(conf_interval_3), max(conf_interval_4)), yaxt
axis(2, at = c(1, 2), labels = c("Sin MP", "Con MP"))

arrows(conf_interval_3[1], 1, conf_interval_3[2], 1, angle = 90, code = 3, length =
arrows(conf_interval_4[1], 2, conf_interval_4[2], 2, angle = 90, code = 3, length =

points(mean(period_btw_pulses_woMP), 1, pch = 19, cex = 1.1)
points(mean(period_btw_pulses_wMP), 2, pch = 19, cex = 1.1)

```

Intervalo de confianza para Periodos entre Pulsos (Sin MP):

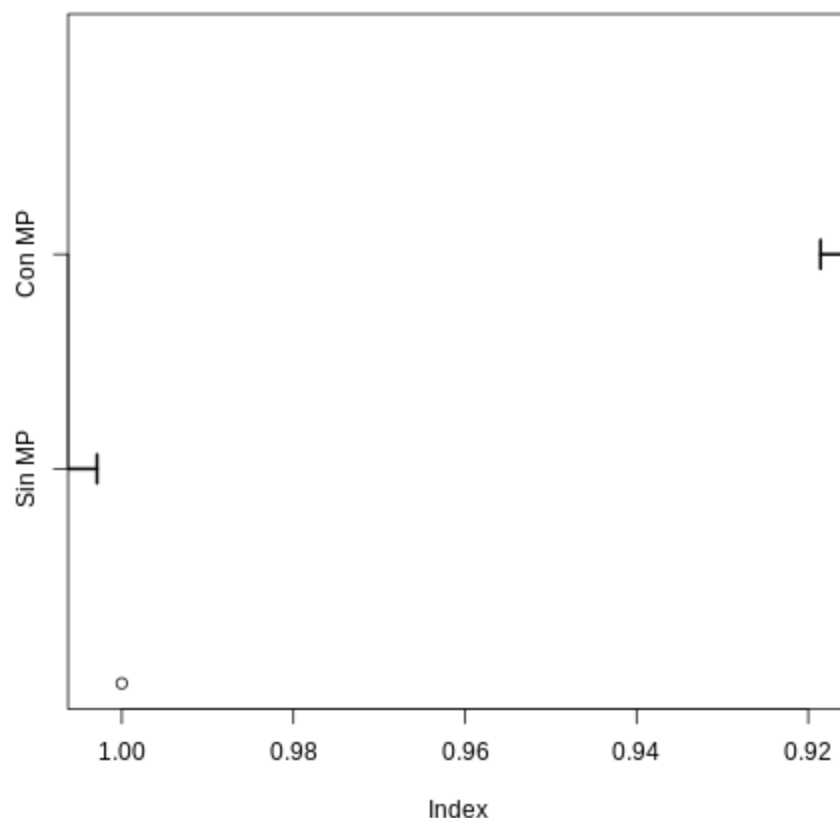
Límite Inferior: 1.002887

Límite Superior: 1.220643

Intervalo de confianza para Periodos entre Pulsos (Con MP):

Límite Inferior: 0.8637941

Límite Superior: 0.9185589



## Interpretación de los resultados

Los gráficos desplegados muestran los intervalos de confianza de las variables "Intensidad de Pulso" y "Periodos entre Pulsos" mediante un par de muestras de estos mismos que se basa en la tercera columna "Marcapasos". Como se puede apreciar, los puntos representan las medias de la muestra con respecto a los intervalos.

Si los intervalos de confianza de ambas muestras no se superponen una sobre la otra, esto quiere decir que puede existir una diferencia significativa entre las medias de estas muestras, tal es el caso del segundo gráfico que proyecta los intervalos de confianza de "Periodos entre Pulsos" con sus respectivas muestras. Ambas líneas nunca se juntan, de hecho se encuentran muy separadas la una de la otra, en cambio, la primer gráfica demuestra que no existe una diferencia significativa entre sus medias debido que sí se superpone una sobre la otra.