Instituto Tecnológico y de Estudios Superiores de Monterrey

Actividad M1. Pruebas de hipótesis

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

24 de Agosto de 2023

Instrucciones

1. Resuelve las dos partes del problema "Enlatados" que se encuentran al final de la presentación de Pruebas de hipótesis. Muestra tu procedimiento siguiendo los 4 pasos

- Elabora un gráfico que muestre la regla de decisión y el punto donde queda el estadístico de prueba.
- Concluye en el contexto del problema.

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

In [ ]:
```
%%R

# Cargamos las librerías necesarias
install.packages('ggplot2')

library(ggplot2)
```

Parte 1: Enlatados

Los pesos de 21 latas de duraznos empacados elegidas al azar fueron:

11.0, 11.6, 10.9, 12.0, 11.5, 12.0, 11.2, 10.5, 12.2, 11.8, 12.1, 11.6, 11.7, 11.6, 11.2, 12.0, 11.4, 10.8, 11.8, 10.9, 11.4

Por estudios anteriores se saber que población del peso de las latas se distribuye normalmente. Si a los dueños no les conviene que el peso sea menor, pero tampoco mayor a 11.7, prueba la afirmación de que el verdadero peso de las latas es de 11.7 con un nivel de confianza de 0.98 haciendo uso de los datos obtenidos en la muestra.

Paso 1: Definir las hipótesis

- $H_0 : \mu = 11.7$
- $H_1 : \mu \neq 11.7$
- Estadístico: $\bar{x}$
- Distribución del estadístico: t de Student
- $\mu_{\bar{x}} = 11.7$, $\sigma_{\bar{x}} = \frac{s}{\sqrt{n}}$

Paso 2. Definir las hipótesis

- Nivel de confianza = 0.98

- $\alpha = 0.02$

In [ ]:
```
%%R

x = c(11.0, 11.6, 10.9, 12.0, 11.5, 12.0, 11.2, 10.5, 12.2, 11.8, 12.1,
11.6, 11.7, 11.6, 11.2, 12.0, 11.4, 10.8, 11.8, 10.9, 11.4)

mu_h0 = 11.7
alpha = 0.02
n = length(x)
t0 = qt(alpha/2, n - 1) # Valor frontera
cat("t0 =", t0)
```

```
t0 = -2.527977
```

t*: es el número de desviaciones estándar al que $\bar{x}$ está lejos de μ

H_0 se rechaza si:

- |t*| > 2.53
- valor p < 0.02

Paso 3. Análisis del resultado

Tenemos que calcular:

- t* (qué tan lejos está $\bar{x}$ de μ)
- Valor p (la probabilidad de que $\bar{x}$ esté en las colas de la distribución)

In [ ]:
```
%%R

m = mean(x)
s = sd(x)
sm = s / sqrt(n)

te = (m - 11.7) / sm
cat("t* =", te)
```

```
t* = -2.068884
```

Cálculo de valor p

In [ ]:
```
%%R

valorp = 2*pt(te, n - 1)
cat("Valor p = ", valorp)
```

```
Valor p =  0.0517299
```

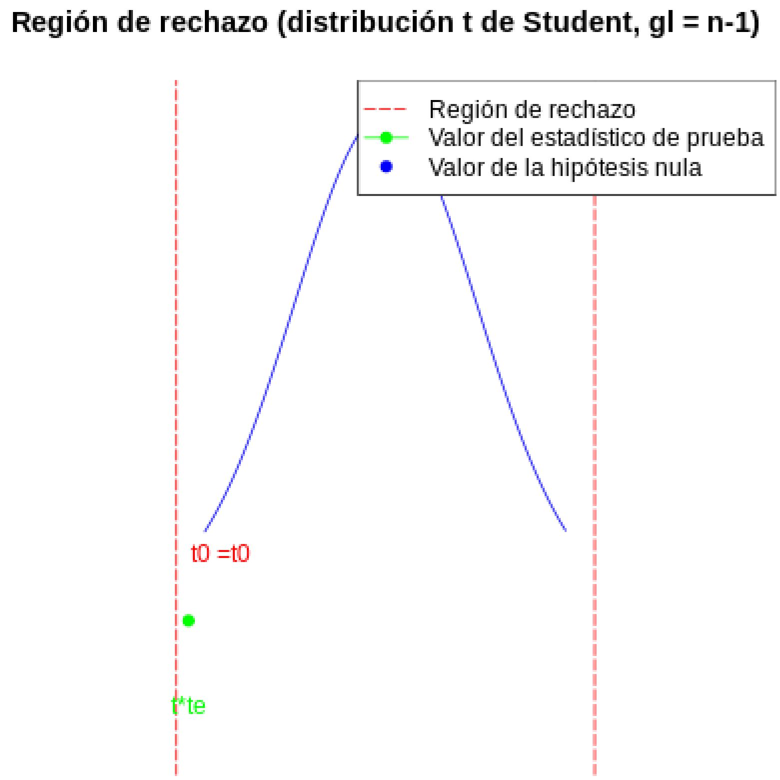
Paso 4. Conclusiones

- Como el valor p (0.05173) es mayor que 0.02, entonces RH_0
- Como |t*| (2.07) es menor que 2.53, entonces no RH_0

En el contexto del problema esto significa que, con un nivel de confianza del 98%, no nos es posible afirmar que el verdadero peso de las latas sea diferente de 11.7. Esto sugiere que no tenemos suficiente evidencia para contradecir la afirmación de que el peso promedio de las latas es de 11.7 unidades. Por lo tanto, no hay base para concluir que el peso promedio de las latas sea distinto de 11.7 unidades con el nivel de confianza y los datos proporcionados.

Paso 5. Graficamos los resultados

In [ ]:
```
%%R

# Datos y parámetros
x = seq(-4 * s, 4 * s, 0.01)
y = dt(x, n - 1)
t0 = qt(alpha, n - 1)
te = (m - mu_h0) / (s / sqrt(n))

# Crear el gráfico
plot(x, y, type = "l", col = "blue", xlab = "", ylab = "", ylim = c(-0.1, 0.4),
     frame.plot = FALSE, xaxt = "n", yaxt = "n", main = "Región de rechazo (distrib
     xlim = c(-8 * s, 8 * s))

# Indicar la zona de rechazo y la media
abline(v = t0, col = "red", lty = 5)
abline(v = -t0, col = "red", lty = 5)
points(mu_h0, 0, col = "blue", pch = 19)

# Dibujar el estadístico de prueba
points(te, 0, col = "green", pch = 19)

# Etiquetas y leyenda
text(t0, 0.05, expression(paste("t0 =", t0)), pos = 4, col = "red")
text(te, -0.05, expression(paste("t*", te)), pos = 1, col = "green")
text(mu_h0, -0.05, expression(paste("miu =", mu_h0)), pos = 1, col = "blue")

legend("topright", legend = c("Región de rechazo", "Valor del estadístico de prueba
       col = c("red", "green", "blue"), pch = c(NA, 19, 19), lty = c(5, 1, NA))
```



Parte 2. Encuestas Telefónicas

Fowle Marketing Research, Inc., basa los cargos a un cliente bajo el supuesto de que las encuestas telefónicas (para recopilación de datos) pueden completarse en un tiempo medio de 15 minutos o menos. Si el tiempo es mayor a 15 minutos entonces se cobra una tarifa adicional. Compañías que contratan estos servicios piensan que el tiempo promedio es mayor a lo que especifica Fowle Marketing Research Inc. así que realizan su propio estudio en una muestra aleatoria de llamadas telefónicas y encuentran los siguientes datos: Tiempo: 17, 11, 12, 23, 20, 23, 15, 16, 23, 22, 18, 23, 25, 14, 12, 12, 20, 18, 12, 19, 11, 11, 20, 21, 11, 18, 14, 13, 13, 19, 16, 10, 22, 18, 23

Paso 1: Definir las hipótesis

- $H_0 : \mu \leq 15$ (No hay tarifa adicional)
- $H_1 : \mu > 15$ (Hay tarifa adicional)

Donde:

- μ es el tiempo promedio de las encuestas telefónicas.

Paso 2: Escoger el nivel de significación

Dado que se nos proporciona un nivel de significación de 0.07, tenemos $\alpha = 0.07$.

Paso 3. Calcular el valor estadísitico de prueba el valor p

In [ ]:
```
%%R

# Datos
tiempo <- c(17, 11, 12, 23, 20, 23, 15, 16, 23, 22, 18, 23, 25, 14, 12, 12, 20, 18,
n <- length(tiempo)
mu_h0 <- 15
sigma <- 4
alpha <- 0.07

# Estadístico de prueba (z)
z <- (mean(tiempo) - mu_h0) / (sigma / sqrt(n))
cat("Valor del estadístico de prueba z:", z, "\n")

# Valor p (probabilidad de obtener un z mayor al observado)
valorp <- 1 - pnorm(z)
cat("Valor p:", valorp, "\n")
```

```
Valor del estadístico de prueba z: 2.95804
Valor p: 0.00154801
```

Paso 4. Tomar la decisión y concluir

Si el valor p es menor que el nivel de significación (α), rechazamos la hipótesis nula (H_0) a favor de la hipótesis alternativa (H_1). En este caso, si $valorp < \alpha$, entonces rechazamos la hipótesis nula.

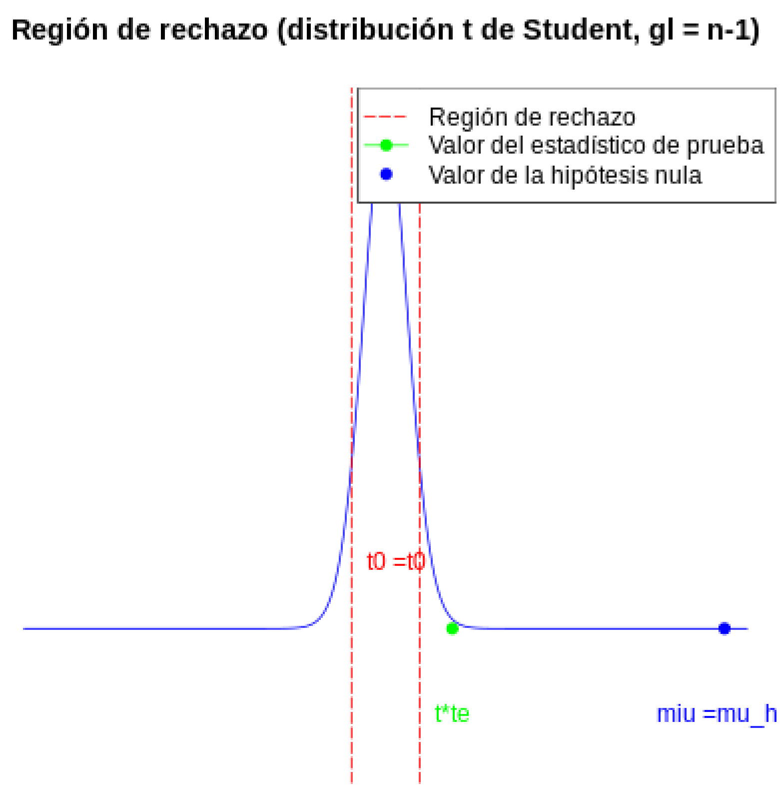
In [ ]:
```
%%R

if (valorp < alpha) {
  cat("Rechazamos H0. Hay evidencia para afirmar que el tiempo promedio es mayor a
} else {
  cat("No rechazamos H0. No hay suficiente evidencia para afirmar que el tiempo pro
}
```

```
Rechazamos H0. Hay evidencia para afirmar que el tiempo promedio es mayor a 15 minut
os.
```

Paso 5. Graficamos

In [ ]:
```
%%R

# Datos y parámetros
x = seq(-4 * sigma, 4 * sigma, 0.01)
y = dt(x, n - 1)
t0 = qt(alpha, n - 1)
```

```
te = (mean(tiempo) - mu_h0) / (sigma / sqrt(n))

# Crear el gráfico
plot(x, y, type = "l", col = "blue", xlab = "", ylab = "", ylim = c(-0.1, 0.4),
     frame.plot = FALSE, xaxt = "n", yaxt = "n", main = "Región de rechazo (distrib

# Indicar la zona de rechazo y la media
abline(v = t0, col = "red", lty = 5)
abline(v = -t0, col = "red", lty = 5)
points(mu_h0, 0, col = "blue", pch = 19)

# Dibujar el estadístico de prueba
points(te, 0, col = "green", pch = 19)

# Etiquetas y leyenda
text(t0, 0.05, expression(paste("t0 =", t0)), pos = 4, col = "red")
text(te, -0.05, expression(paste("t*", te)), pos = 1, col = "green")
text(mu_h0, -0.05, expression(paste("miu =", mu_h0)), pos = 1, col = "blue")

legend("topright", legend = c("Región de rechazo", "Valor del estadístico de prueba
       col = c("red", "green", "blue"), pch = c(NA, 19, 19), lty = c(5, 1, NA))
```



Dentro delo contexto del problema, dado el nivel de significación de 0.07, y considerando la muestra de tiempos de encuestas telefónicas, rechazamos la hipótesis nula. Por lo tanto, hay evidencia estadística para afirmar que el tiempo promedio de las encuestas telefónicas es mayor a 15 minutos, justificando así la tarifa adicional.