Actividad 8: Pruebas de Hipótesis

Daniela Jiménez Téllez

2024-08-23

Importación de librerías

```
library(BSDA)
```

```
## Cargando paquete requerido: lattice
```

```
##
## Adjuntando el paquete: 'BSDA'
```

```
## The following object is masked from 'package:datasets':
##
##     Orange
```

Problema 1: Enlatados

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

- *Peso de las latas:* 11, 11.6, 11.6, 11.7, 10.9, 11.6, 12, 11.2, 11.5, 12, 12, 11.4, 11.2, 10.8, 10.5, 11.8, 12.2, 10.9, 11.8, 11.4, 12.1

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.

> 1. Muestra tu procedimiento siguiendo los 4 pasos de las pruebas de hipótesis.

- **Paso 1: Hipótesis**

$H_0 : \mu = 11.7$

$H_1 : \mu \neq 11.7$

¿Cómo se distribuye $\bar{x}$?

- x se distribuye como una normal
- $n < 30$
- No conocemos sigma

Entonces la distribución muestral es una *t de Student*.

- **Paso 2: Regla de decisión**

En este caso se utilizará un nivel de confianza de 0.98, por lo que la significancia será de 0.02.

Se necesita encontrar qué tantas desviaciones estándar está lejos el valor frontera.

```
n = 21
alpha = 0.02
t_f = qt(alpha/2, n - 1)
cat("t_f:", abs(t_f))
```

```
## t_f: 2.527977
```

Lo que nos deja con la siguiente regla de decisión:

- Rechazo H_0 sí:
  - $|t_e| > 2.53$
  - valor p < 0.02
- **Paso 3: Análisis de resultado**
- t_e: Número de desviaciones estándar del que $\bar{x}$ se encuentra lejos de $\mu = 11.7$.
- Valor p: Probabilidad de obtener lo que se obtuvo en la muestra o un valor más extremo.

Estadísticos de prueba:

```
x = c(11, 11.6, 11.6, 11.7, 10.9, 11.6, 12, 11.2, 11.5, 12, 12, 11.4, 11.2, 10.8, 10.5, 11.8, 1
2.2, 10.9, 11.8, 11.4, 12.1)

xb = mean(x)

s = sd(x)

miu = 11.7

te = (xb - miu) / (s / sqrt(n))

cat("te:", te)
```

```
## te: -2.068884
```

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

```
## Valor p: 0.0517299
```

- **Paso 4: Conclusión**

Para concluir se necesita comparar la Regla de Decisión contra el resultado del análisis. De acuerdo con la Regla Decisión:

- Rechazo H_0 sí:

- $|t_e| > 2.53$
- valor p < 0.02

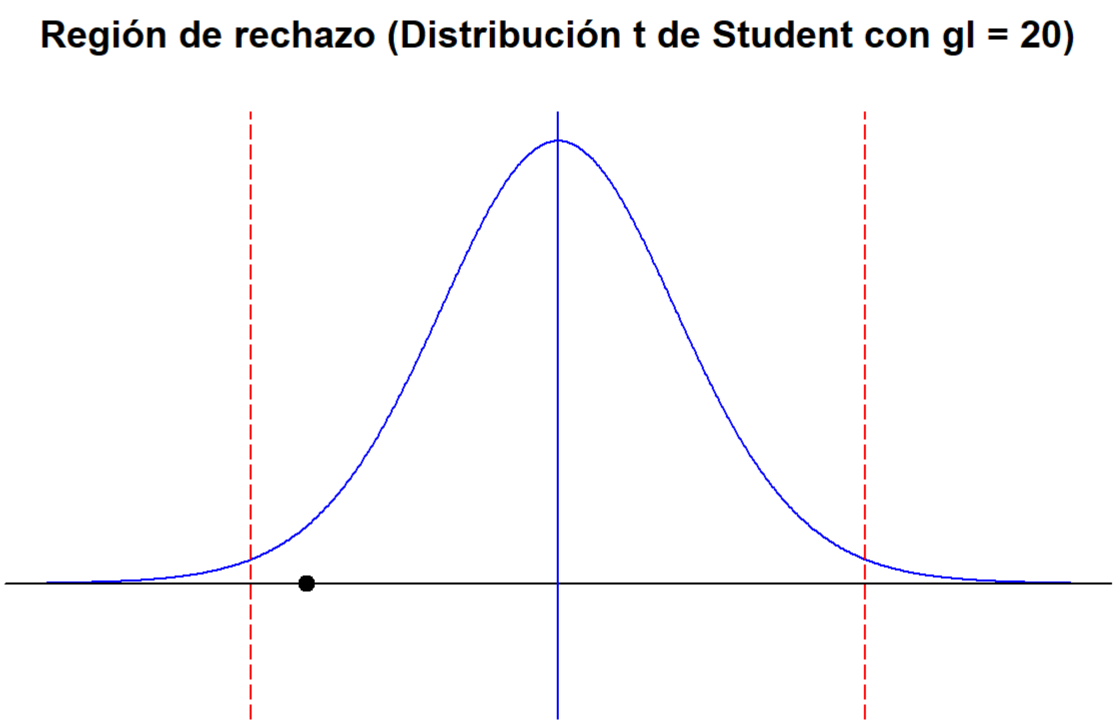
Dado que el valor t_e nos dió un total de 2.07 y este es menor a 2.53, entonces no rechazo H_0. Igualmente, el valor p nos dió un total de 0.05, el cual es mayor que 0.02, por lo tanto tampoco se rechaza H_0.

> 2. Elabora un gráfico que muestre la regla de decisión y el punto donde queda el estadístico de prueba.

```
sigma = sqrt((n - 1) / (n - 3))
x = seq(-4 * sigma, 4 * sigma, 0.01)
y = dt(x, n - 1)

plot(x, y, type = "l", col = "blue", xlab = "", ylab = "", ylim = c(-0.1, 0.4), frame.plot = FAL
SE, xaxt = "n", yaxt = "n", main = "Región de rechazo (Distribución t de Student con gl = 20)")


abline(v = t_f, col = "red", lty = 5)
abline(v = -1 * t_f, col = "red", lty = 5)
abline(h = 0)
abline(v = 0, col = "blue", pch = 19)
points(te, 0, pch = 19, cex = 1.1)
```



> 3. Concluye en el contexto del problema.

Como se dijo antes, dado que el valor t_e nos dió un total de 2.07 y este es menor a 2.53, entonces no se rechaza H_0. Igualmente, el valor p nos dió un total de 0.05, el cual es mayor que 0.02, por lo tanto tampoco se rechaza H_0. Esto nos permite concluir que las tazas de durazno tienen el peso requerido.

Problema 2: La decisión de Fowle Marketing Research, Inc.

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

Por experiencias anteriores, se sabe que σ = 4 minutos. Usando un nivel de significación de 0.07, ¿está justificada la tarifa adicional?

> 1. Muestra tu procedimiento siguiendo los 4 pasos de las pruebas de hipótesis.

- **Paso 1: Hipótesis**

$H_0 : \mu \leq 15$

$H_1 : \mu > 15$

¿Cómo se distribuye $\bar{x}$?

- x se distribuye como una normal
- $n > 30$
- Conocemos que $\sigma = 4$

Entonces se asume que la distribución muestral es una *normal*.

- **Paso 2: Regla de decisión**

En este caso se utilizará un nivel de confianza de 0.93, por lo que la significancia será de 0.07.

Se necesita encontrar qué tantas desviaciones estándar está lejos el valor frontera.

```
n = 35
alpha = 0.07
z_f = qnorm(1 - alpha)
cat("z_f:", z_f)
```

```
## z_f: 1.475791
```

Lo que nos deja con la siguiente regla de decisión:

- Rechazo H_0 sí:

  - $|z_e| > 1.48$
  - valor p < 0.07

- **Paso 3: Análisis de resultado**
- z_e: Número de desviaciones estándar del que $\bar{x}$ se encuentra lejos de $\mu = 15$.
- Valor p: Probabilidad de obtener lo que se obtuvo en la muestra o un valor más extremo.

Estadísticos de prueba:

```
tiempos = c(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)

xb = mean(tiempos)

miu = 15

sigma = 4

ze = (xb - miu) / (sigma / sqrt(n))

cat("ze:", ze)
```

```
## ze: 2.95804
```

```
valorp = 1 - pnorm(ze)
cat("Valor p:", valorp)
```

```
## Valor p: 0.00154801
```

- **Paso 4: Conclusión**

Para concluir se necesita comparar la Regla de Decisión contra el resultado del análisis. De acuerdo con la Regla Decisión:

- Rechazo H_0 sí:
  - $|z_e| > 1.48$
  - valor p < 0.07

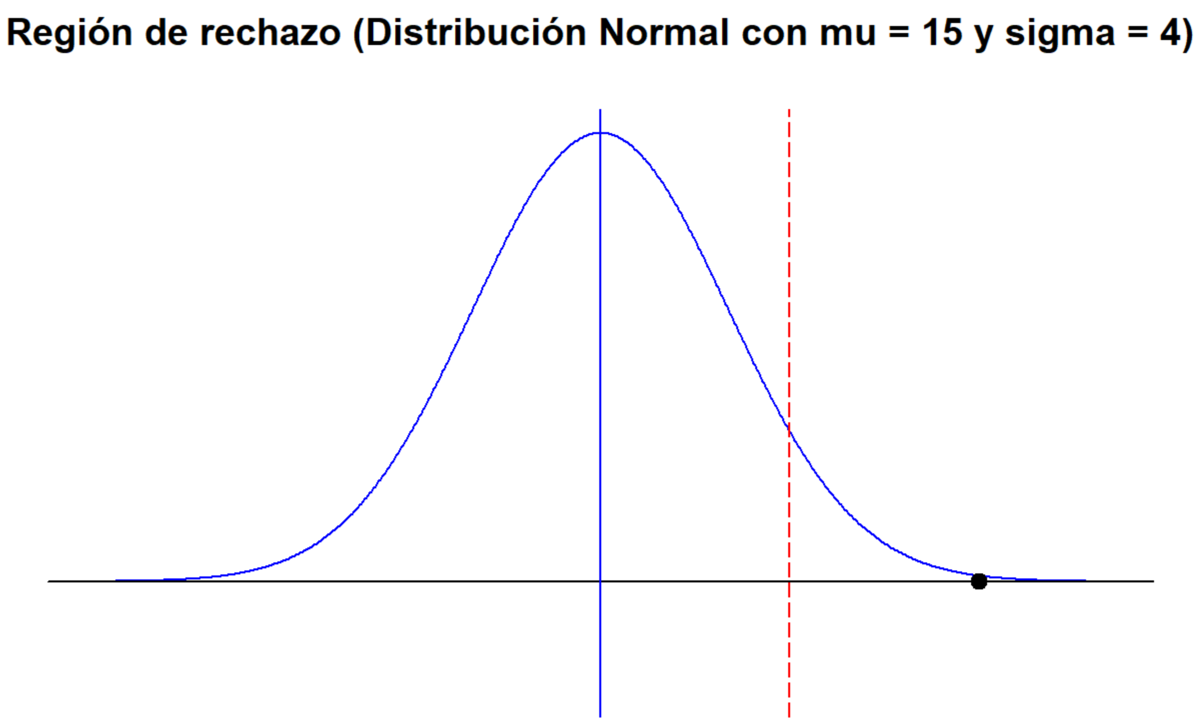
Dado que el valor z_e nos dió un total de 2.96 y este es mayor a 1.48, entonces se rechaza H_0. Igualmente, el valor p nos dió un total de 0.0015, el cual es menor que 0.07, por lo tanto también se rechaza H_0.

> 2. Elabora un gráfico que muestre la regla de decisión y el punto donde queda el estadístico de prueba.

```
x = seq(-4, 4, 0.01)
y = dnorm(x)

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 (Distribución Normal con mu = 15 y sigma = 4)")

abline(v = z_f, col = "red", lty = 5)
abline(h = 0)
abline(v = 0, col = "blue", pch = 19)
points(ze, 0, pch = 19, cex = 1.1)
```



3. Concluye en el contexto del problema.

```
resultado <- z.test(x = tiempos, mu = 15, sigma.x = 4, conf.level = 0.93)
resultado
```

```
##
##  One-sample z-Test
##
## data:  tiempos
## z = 2.958, p-value = 0.003096
## alternative hypothesis: true mean is not equal to 15
## 93 percent confidence interval:
##  15.77492 18.22508
## sample estimates:
## mean of x
##        17
```

Como se menciono anteriormente, tanto z_e y el valor p cumplen las condiciones para rechazar H_0, lo que nos dice que se acepta H_1. Por lo tanto, ya que el tiempo promedio de las llamadas es mayor a 15 minutos, está justificada la tarifa adicional.