Pruebas de hipótesis

Erika Martínez Meneses

2024-08-23

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.

Muestra tu procedimiento siguiendo los 4 pasos de las pruebas de hipótesis

Paso 1. Establecer hipótesis

$$H_0: \mu = 11.7$$

El peso de las latas es $= 11.7$

$$H_1: \mu =\neq 11.7$$

El peso de las latas es $\neq 11.7$

¿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. Definir la regla de decisión

- Nivel de confianza es de 0.98
- Nivel de significancia es de 0.02

Necesito encontrar a cuántas desviaciones estándar está lejos el valor frontera

```
n = 21
alfa = 0.02
```

```r
t_f = qt(alfa/2,n-1) # valor frontera a partir del cuál voy a rechazar h0
cat("t_f=", t_f)

## t_f= -2.527977
```

Si el estadístico de prueba es mayor a $|tf|$ o si el valor p es < alpha rechazo H_0

Regla de decisión

Rechazo H_0 si:

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

Paso 3. Analizar el resultado

- t_e: Número de desviaciones al que $\bar{x}$ se encuentra lejos de $\mu = 11.7$
- Valor p: Probabilidad de obtener lo que obtuve en la muestra o un valor más extremo

Estadístico de prueba

```r
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, 12.2, 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
```

Más fácil: Para hacer el análisis del resultado

```r
t.test(X, mu = 11.7, alternative = "two.sided", conf.level = 0.98)

##
##  One Sample t-test
##
## data:  X
## t = -2.0689, df = 20, p-value = 0.05173
## alternative hypothesis: true mean is not equal to 11.7
## 98 percent confidence interval:
##  11.22388 11.74755
## sample estimates:
## mean of x
##  11.48571
```

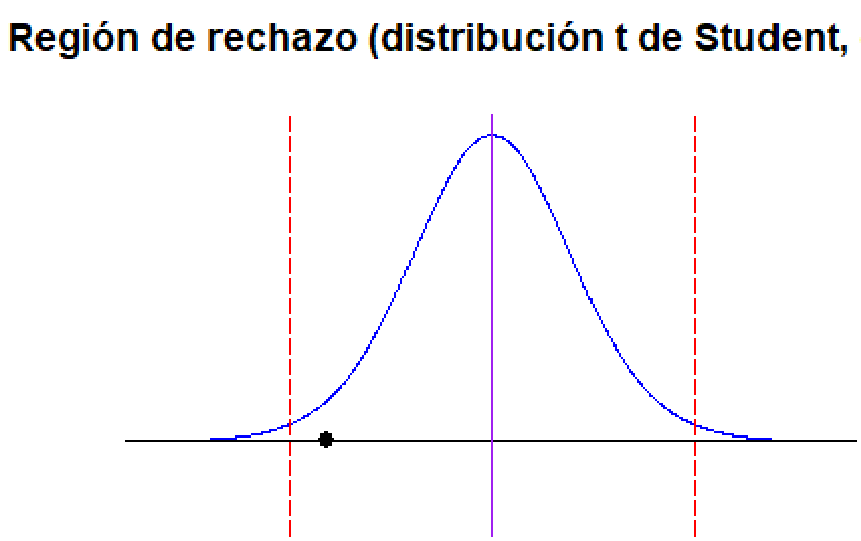
Elabora un gráfico que muestre la regla de decisión y el punto donde queda el estadístico de prueba.

```
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=
FALSE,xaxt="n",yaxt="n",main="Región de rechazo (distribución t de Studen
t, gl=20)")

abline(v=t_f,col="red",lty=5) # zona de rechazo
abline(v=-1*t_f,col="red",lty=5) # zona de rechazo
abline(h=0)
abline(v = 0,col="purple",pch=19) # Media

points(te, 0, pch=19, cex=1.1) # estadístico de prueba
```



Paso 4. Conclusión (estadística y en el contexto)

Comparar: Regla de decisión vs Análisis del resultado (Para conclusión estadística)

- $|t_e| = 2.07 < 2.53$ -> No Rechazo H_0
- valor p = 0.05 < 0.02 -> No Rechazo H_0

En el contexto: Las latas 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?

Muestra tu procedimiento siguiendo los 4 pasos de las pruebas de hipótesis

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

Paso 1. Establecer hipótesis

$$H_0: \mu = 15$$

Las encuestas telefónicas se completan en un tiempo medio $\leq$ 15min

$$H_1: \mu > 15$$

Las encuestas telefónicas se completan en un tiempo medio > 15min

¿Cómo se distribuye $\bar{X}$?

- X se distribuye como una Normal
- n > 30 (son 35)
- σ = 4
- Nivel de significación = 0.07 (α)

Entonces la distribución muestral es una t de Student

Paso 2. Definir la regla de decisión

- Nivel de confianza es de 0.93
- Nivel de significancia es de 0.07
- Estadístico de prueba $Z = \frac{\bar{X}-\mu_0}{\frac{\sigma}{\sqrt{n}}}$
- Valor crítico Z_a para una prueba unilateral con $\alpha = 0.07$.

```
n = 35
alfa = 0.07
sigma = 4
```

Regla de decisión

Rechazo H_0 si:

- Si $Z > Z_a$ rechazamos la hipótesis nula.
- valor p < 0.07

Paso 3. Analizar el resultado

Estadístico de prueba

```
X = 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(X)
miu = 15
z = (xb - miu) / (sigma / sqrt(n)) # Estadístico z
z

## [1] 2.95804
```

Valor crítico Z

```
z_alfa <- qnorm(1 - alfa)
z_alfa

## [1] 1.475791

p_value = 1 - pnorm(z)
cat("Valor p =", p_value)

## Valor p = 0.00154801
```

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

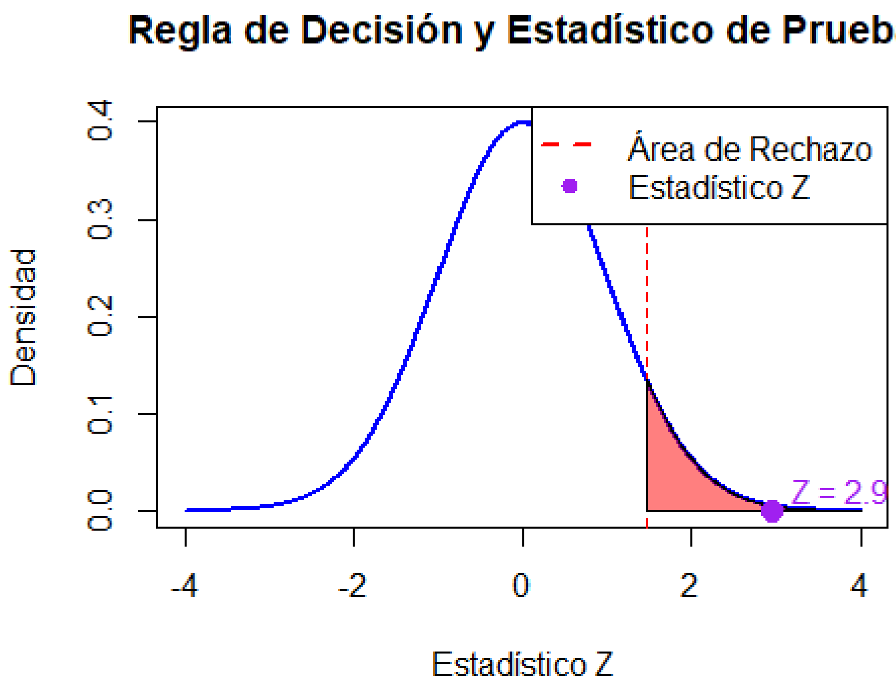
```
# Gráfico
x <- seq(-4, 4, length=1000)
y <- dnorm(x)
plot(x, y, type="l", lwd=2, col="blue", main="Regla de Decisión y Estadís
tico de Prueba", xlab="Estadístico Z", ylab="Densidad")

# Área de rechazo
abline(v = z_alfa, col="red", lty=2)
polygon(c(z_alfa, x[x > z_alfa], max(x)), c(0, y[x > z_alfa], 0), col=rgb
(1,0,0,0.5))

# Estadístico de prueba
points(z, 0, col="purple", pch=19, cex=1.5)
text(z, 0.02, labels = paste("Z =", round(z, 2)), pos=4, col="purple")

# Leyenda
```

```
legend("topright", legend=c("Área de Rechazo", "Estadístico Z"), col=c("r
ed", "purple"), lwd=2, lty=c(2, NA), pch=c(NA, 19))
```



Paso 4. Conclusión

Comparar: Regla de decisión vs Análisis del resultado (Para conclusión estadística)

- $Z = 2.95804 > Z_a = 1.475791$ -> Rechazo H_0
- valor p = 0.00154801 < 0.07 -> Rechazo H_0

En el contexto: Existe evidencia suficiente para afirmar que el tiempo promedio de las encuestas telefónicas es mayor a 15 minutos.

¿Está justificada la tarifa adicional? La tarifa adicional no está justificada porque el tiempo promedio es mayor a 15 minutos, lo cual incumple con la especificación original de la empresa.