Act8. Pruebas de Hipotesis

Andrés Villarreal González

2024-08-23

Pruebas de Hipotesis

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.

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

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
t_f = qt(alfa/2, n-1)
cat("t_f =", t_f)

## t_f = -2.527977
```

Regla de decisión

Rechazo H_0 si: * $|t_e| > 2.53$ * valor p < 0.02

Paso 3: Análisis del 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 mas extremo

Estadistico 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, 12.2, 10.9, 11.8, 11.4, 12.1)

xb = mean(X)
s = sd(X)
miu = 11.7
n = 21

te = (xb-miu)/(s/sqrt(n))
cat("te =", te)

## te = -2.068884
```

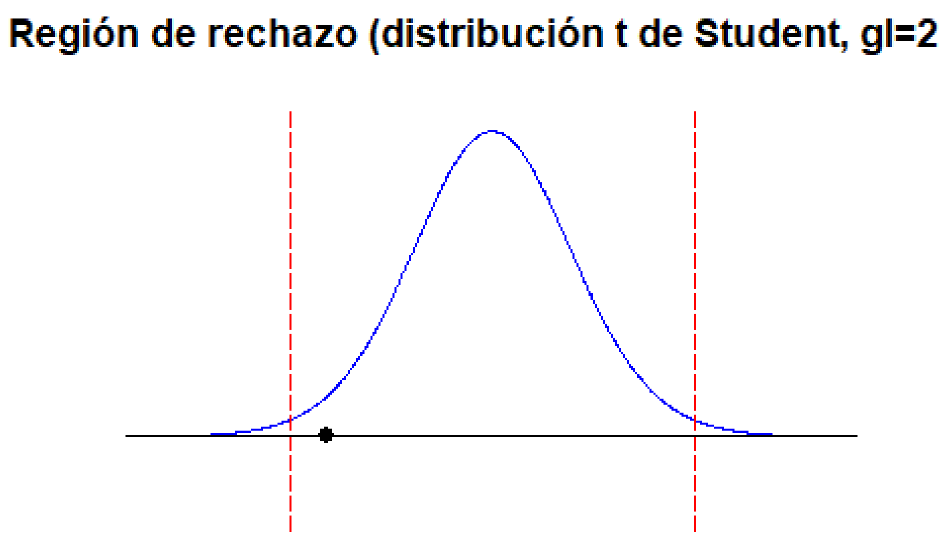
Valor p

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

## Valor p = 0.0517299
```

Mas facil Para hacer el analisis de resultado

```
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

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 Student, gl=20)")
abline(v=t_f,col="red",lty=5)
abline(v=-1*t_f,col="red",lty=5)
abline(h=0)
points(miu,0,col="blue",pch=19)
points(te, 0, pch=19, cex=1.1)
```



Paso 4: Conclusión

Comparar: Regla de decisión vs Análisis del resultado * $|t_e| = 2.07 < 2.53$ -> No RH0 * 0.051 > 0.02 -> No RH0

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?

Paso 1: Hipotesis

- $H_0 : \mu \leq 15$
- $H_1 : \mu > 15$

¿Cómo se distribuye $\bar{X}$

- X se distribuye como una Normal
- n > 30
- No conocemos sigma

Entonces: la distribución muestral es una z

Paso 2: Regla de decision

Nivel de confianza es de 0.93 Nivel de significancia es de 0.07

Necesito encontrar a cuántas desviaciones estándar está lejos el valor frontera

```
alfa <- 0.07
z_f <- qnorm(1 - alfa)
cat("z_f =", z_f)

## z_f = 1.475791
```

Regla de decisión

Rechazo H_0 si: * $|z_f| > 1.475791$ * valor p < 0.07

Paso 3: Análisis del resultado

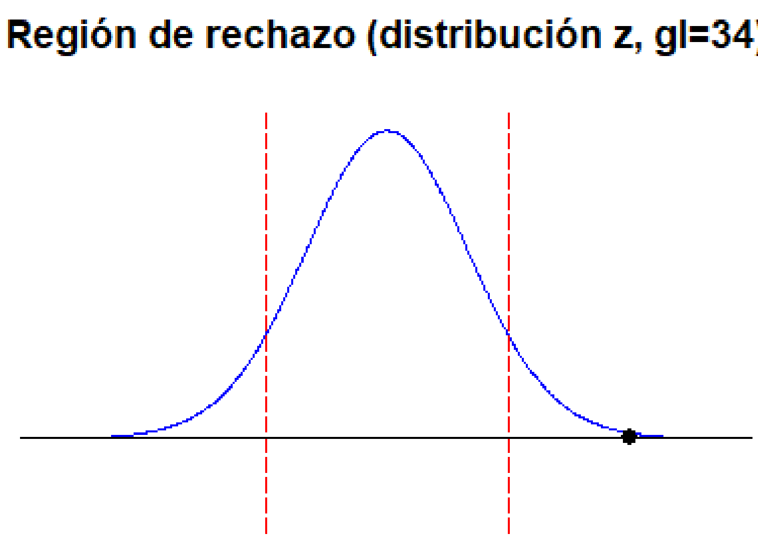
- t_e: Número de desviaciones al que $\bar{x}$ se encuentra lejos de $\mu = 11.7$
- Valor p: Probabilidad de obtener lo que obtuve en la muestra o un valor mas extremo

Estadistico 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)

mu_0 <- 15
sigma <- 4
n <- length(X)

xb <- mean(X)

z_estadistico <- (xb - mu_0) / (sigma / sqrt(n))

cat("z_estadístico =", z_estadistico)

## z_estadístico = 2.95804
```

Valor p

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

## Valor p = 0.00154801

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 z, gl=34)")
abline(v=z_f,col="red",lty=5)
abline(v=-1*z_f,col="red",lty=5)
abline(h=0)
points(mu_0,0,col="blue",pch=19)
points(z_estadistico, 0, pch=19, cex=1.1)
```



Paso 4: Conclusión

Comparar: Regla de decisión vs Análisis del resultado * $|z_{estadistico}| = 2.958 > 1.475$ -> RH0 * 0.0015 < 0.07 -> RH0

Esto indicaría que hay evidencia significativa para afirmar que el tiempo promedio de las encuestas telefónicas es mayor a 15 minutos, lo que justificaría la tarifa adicional.