8. Pruebas de hipótesis

Juan Bernal

2024-08-23

Resuelve el problema “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: Hipótesis

$H_0 : \mu = 11.7$ Es el verdadero peso de las latas

$H_1 : \mu \neq 11.7$ No es el verdadero peso de las latas

- X : Pesos de las latas, además $X \sim N(\mu_x, \sigma_x)$
- $n < 30$
- $\sigma = ?$

Entonces, la distribución muestral es una T de Student. Donde el estadístico de prueba está dado como $T = \frac{\bar{X} - \mu}{\frac{s}{\sqrt{n}}}$

Paso 2: Regla de decisión

- Nivel de confianza del 0.98
- Nivel de significancia del 0.02
- Encontrar a cuántas desviaciones estándar está lejos el valor frontera

```
n = 21 # Grados de libertad
alfa = 0.02 # Nivel de significancia
t_f = qt(alfa/2, n-1) # Valor frontera
cat('La frontera de rechazo será ', abs(t_f), ' desviaciones estándar alrededor de la media')
```

```
## La frontera de rechazo será  2.527977  desviaciones estándar alrededor de la media
```

La regla de decisión estará dada por:

*Si $|t_e| > 2.53$, entonces rechazamos H_0

*Si $p - value < 0.02$, entonces rechazamos H_0

Paso 3: Análisis del resultado

t_e: Número de desviaciones al que $\bar{X}$ se encuentra lejos de $\mu = 11.7$.

$P - value$: Probabilidad de obtener lo que obtuve en la muestra o un valor más extremo.

```
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) # Pesos de las latas
x_bar = mean(x) # Media de la muestra
s = sd(x) # Desviación estándar de la muestra
mu = 11.7


te = (x_bar - mu)/(s/sqrt(n)) # Estadístico de prueba
cat('El estadístico de prueba t es igual a ', abs(te))
```

```
## El estadístico de prueba t es igual a  2.068884
```

```
p_value = 2*pt(te, n-1) # Valor P de dos colas
cat('El p-value es ', p_value)
```

```
## El p-value es  0.0517299
```

Una manera más fácil de hacerlo

```
t.test(x, mu = mu, 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
```

Paso 4: Conclusión

Regla de decisión vs Análisis del resultado

*Dado que $|t_e| = 2.07 < 2.53$, entonces no hay suficiente evidencia para rechazar H_0

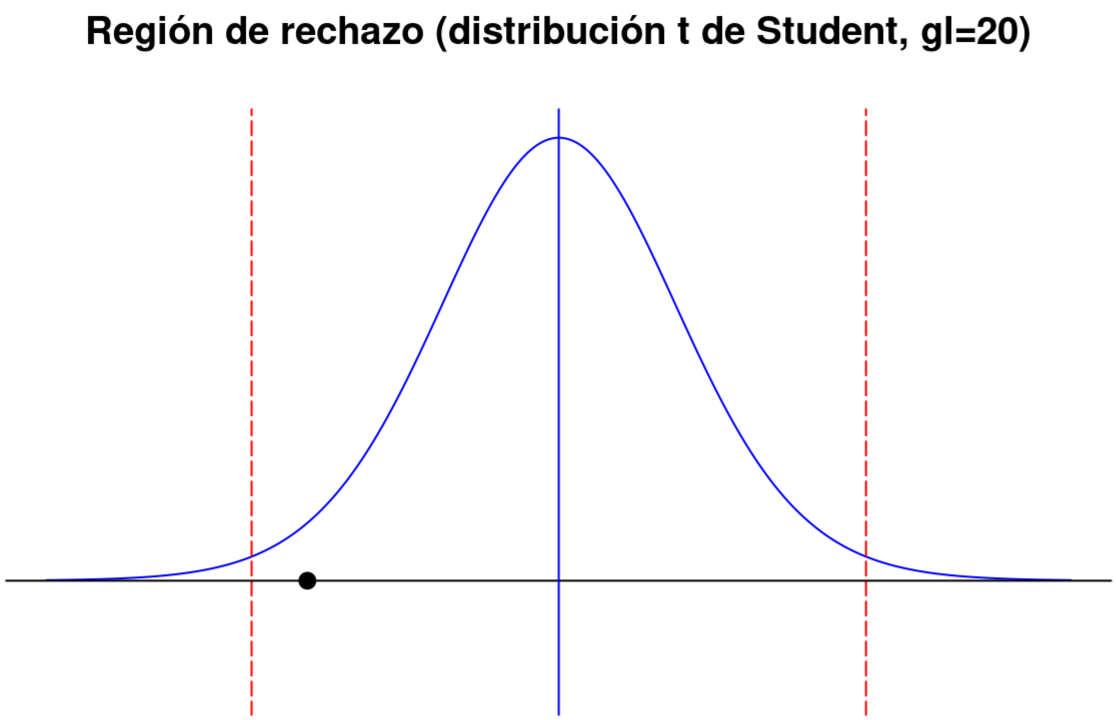
*Dado que $p - value = 0.0517 > 0.02$, entonces no hay suficiente evidencia para rechazar H_0

Como observamos que las pruebas pasaron la regla de decisión y terminamos sin rechazar la hipótesis nula, entonces concluimos que las latas de durazno pesan 11.7

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="Regi
ón 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)
abline(v = 0, col = "blue", pch = 19)
points(te, 0, pch=19, cex=1.1)
```



Concluye en el contexto del problema.

Como observamos que las pruebas pasaron la regla de decisión y terminamos sin rechazar la hipótesis nula, entonces concluimos que las latas de durazno pesan 11.7.

Resuelve el problema: “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

Paso 1: Hipótesis

$H_0: \mu = 15$ Las encuestas telefónicas pueden completarse en 15 minutos.

$H_1: \mu > 15$ Las encuestas telefónicas se completan en más de 15 minutos.

- X : Minutos que dura la encuesta telefónica, además $X \sim N(\mu_x, \sigma_x)$
- $n = 35 > 30$
- $\sigma = 4$ minutos

Entonces, la distribución muestral es Normal, pues hay más de 30 datos y contamos con la desviación estándar. Donde el estadístico de prueba está dado como $Z = \frac{\bar{X}-\mu}{\frac{\sigma}{\sqrt{n}}}$

Paso 2: Regla de decisión

- Nivel de confianza del 0.93
- Nivel de significancia del 0.07
- Encontrar a cuántas desviaciones estándar está lejos el valor frontera

```
x1 = 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) # Tiempo

n = length(x) # Número de observaciones
alfa = 0.07 # Nivel de significancia
med = mean(x1)
miu = 15
de = 4 # Desviación estándar de la población
z_f = qnorm(alfa/2) # Valor frontera
cat('La frontera de rechazo será ', abs(z_f), ' desviaciones estándar alrededor de la media')
```

```
## La frontera de rechazo será  1.811911  desviaciones estándar alrededor de la media
```

La regla de decisión estará dada por:

*Si $|z_e| > 1.81$, entonces rechazamos H_0

*Si $p - value < 0.07$, entonces rechazamos H_0

Paso 3: Análisis del resultado

z_e: Número de desviaciones al que $\bar{X}$ se encuentra lejos de $\mu = 15$.

$P - value$: Probabilidad de obtener lo que obtuve en la muestra o un valor más extremo.

```
library(BSDA)
```

```
## Loading required package: lattice
```

```
##
## Attaching package: 'BSDA'
```

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

```
z.test(x1, mu=miu, sigma.x = de, alternative="two.sided", conf.level=0.93)
```

```
##
##  One-sample z-Test
##
## data:  x1
## 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
```

El estadístico de prueba z es igual a 2.958 y el $p - value$ es aproximadamente 0.0031.

Paso 4: Conclusión

Regla de decisión vs Análisis del resultado

*Dado que $|z_e| = 2.958 > 1.81$, entonces hay suficiente evidencia para rechazar H_0

*Dado que $p - value = 0.0031 < 0.07$, entonces hay suficiente evidencia para rechazar H_0

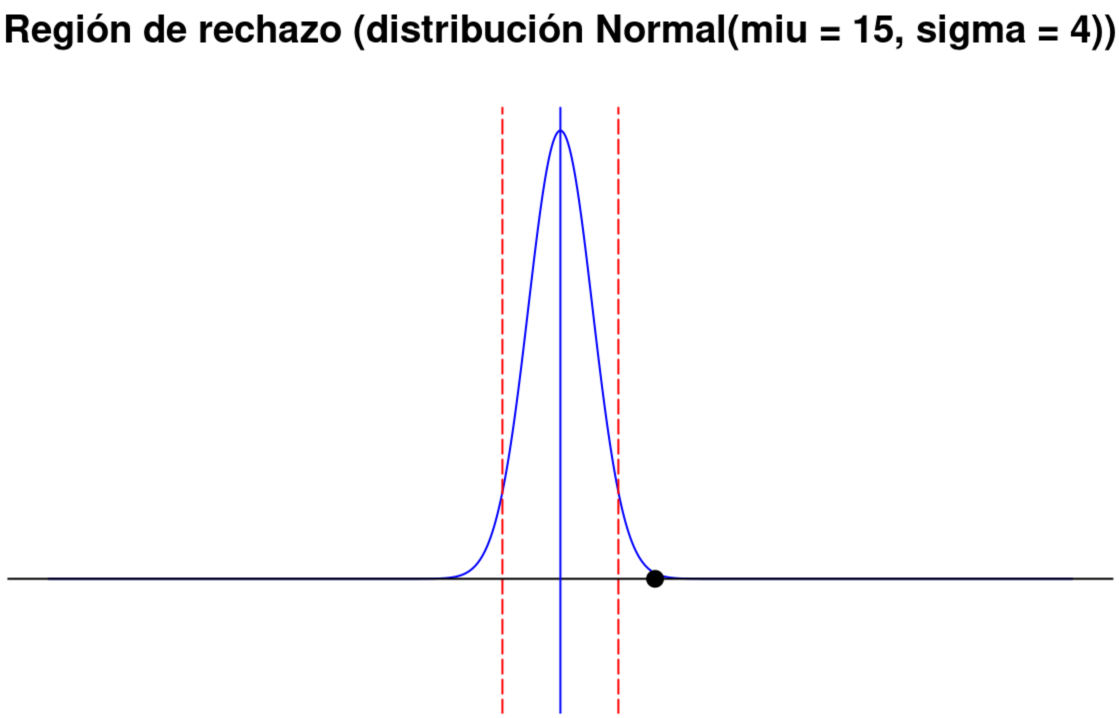
Como observamos que las pruebas no pasaron la regla de decisión y terminamos por rechazar la hipótesis nula, entonces concluimos que las encuestas telefónicas (para recopilación de datos) pueden completarse en un tiempo medio de más de 15 minutos, al contrario de lo que la compañía afirmaba.

Elabora un gráfico que muestre la regla de decisión y el

punto donde queda el estadístico de prueba.

```
x=seq(-4*de,4*de,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="Regi
ón de rechazo (distribución Normal(miu = 15, sigma = 4))")
abline(v=z_f,col="red",lty=5)
abline(v=-1*z_f,col="red",lty=5)
abline(h=0)
abline(v = 0, col = "blue", pch = 19)
points(2.958, 0, pch=19, cex=1.1)
```



Concluye en el contexto del problema.

Como observamos que las pruebas no pasaron la regla de decisión y terminamos por rechazar la hipótesis nula, entonces concluimos que las encuestas telefónicas (para recopilación de datos) pueden completarse en un tiempo medio de más de 15 minutos, al contrario de lo que la compañía afirmaba.