5-pruebas-de-hipotesis

August 23, 2023

1 5. Pruebas de hipótesis

Francisco Mestizo Hernández A01731549

1.1 Problema 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.

1.1.1 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 (no conocemos la desviación)

$\mu_{\bar{x}} = 11.7$, $\sigma_{\bar{x}} = \frac{s}{\sqrt{n}}$

1.1.2 Paso 2

Regla de decisión

Nivel de confianza = 0.98

$\alpha = 0.02$

[ ]:
```
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)

alfa = 0.02
n = length(x)
t0 = qt(alfa/2, n-1) #Valor frontera
```

```
cat("t0 = ", t0)
```

```
t0 =  -2.527977
```

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

H_0 se rechaza si:

- |t*| > 2.53

1.1.3 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)

Cálculo de t

[ ]:
```
m = mean(x)
s = sd(x)
sm = s/sqrt(n)

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

```
t* = -2.068884
```

Calculo de valor p

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

```
Valor p = 0.0517299
```

1.1.4 Paso 4

Conclusiones

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

En el contexto del problema esto significa que mantenemos la hipótesis, ya que el valor calculado no se sale de los rangos marcados. Por lo tanto, podemos decir que el peso de las latas no ha cambiado, no es ni menor ni mayor de lo esperado.

Mas fácil:

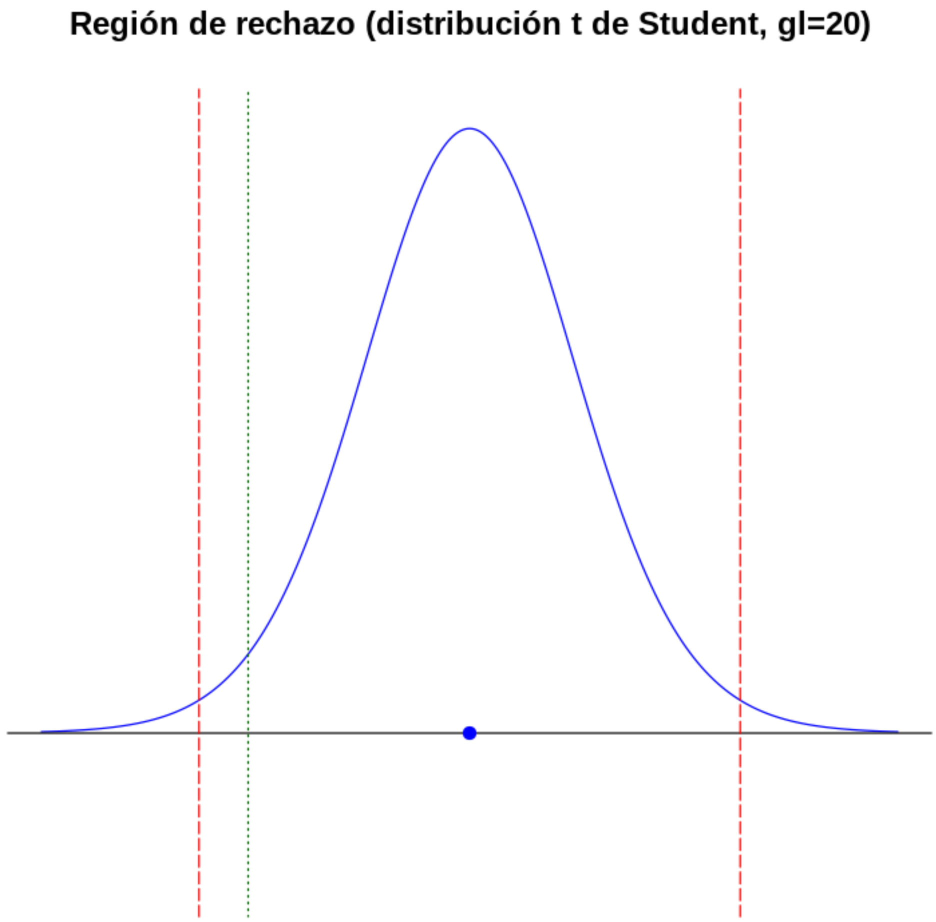
[ ]:
```
t.test(x, alternative="two.sided", mu=11.7, conf.level=0.98)
```

```
        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
```

Podemos graficar los resultados para ver que la t* (raya verde) quedó dentro del intervalo de confianza (rojo). Además, la media está marcada con un punto

[ ]:
```
x=seq(-4,4,0.01)
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(h=0)
abline(v=t0,col="red",lty=5)
abline(v=-1*t0,col="red",lty=5)
points(0,0,col="blue",pch=19)
abline(v=te,col="darkgreen",lty=3)
```



1.2 Problema 2. Fowle Marketing Research

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 =4 minutos. Usando un nivel de significación de 0.07, ¿está justificada la tarifa adicional?

- Muestra tu procedimiento siguiendo los 4 pasos de solución
- Grafica la regla de decisión y el valor del estadístico de prueba.
- Concluye en el contexto del problema

1.2.1 Paso 1

Definir las hipótesis

$H_0 : \mu <= 15$

$H_1 : \mu > 15$

Estadístico: $\bar{x}$

Distribución del estadístico: Normal (Conocemos la desviación)

$\mu_{\bar{x}} = 15$, $\sigma_{\bar{x}} = 4$

1.2.2 Paso 2

Regla de decisión

Nivel de confianza = 0.86 Nivel de significacion = 0.07

$\alpha = 0.07$

[ ]:
```
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)

n = length(x)
alfa = 0.07
z0 = qnorm(alfa/2) #Valor frontera
cat("z0 = ", t0)
```

```
z0 =  -2.527977
```

z* es el numero de desviaciones estándar al que $\bar{x}$ está lejos de μ

H_0 se rechaza si:

- |z*| > 2.53

1.2.3 Paso 3

Análisis del resultado

Tenemos que calcular:

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

Cálculo de z

[ ]:
```
m = mean(x)
s = 4
```

```
sm = s/sqrt(n)

ze = (m-15)/sm
cat("z* =", ze)
```

```
z* = 2.95804
```

Calculo de valor p

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

```
Valor p = 0.00154801
```

1.2.4 Paso 4

Conclusiones

- Como valor p (0.0015) es menor que 0.07, entonces RH_0
- Como |t*| (2.96) es mayor que 2.53, entonces RH_0

En el contexto del problema esto significa que rechazamos la hipotesis de que las llamadas duran 15 minutos o menos. Por lo tanto, el cobro extra por las llamadas está justificado porque estas duran más de 15 minutos.

Mas fácil:

Podemos graficar los resultados para ver que la z* (raya verde) quedó más a la derecha del intervalo de confianza (rojo). Además, la media está marcada con un punto.

[ ]:
```
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 t de␣
  ↪Student, gl=20)")
abline(h=0)

abline(v=-1*z0,col="red",lty=5)
points(0,0,col="blue",pch=19)
abline(v=ze,col="darkgreen",lty=3)
```

Región de rechazo (distribución t de Student, gl=20)