

8. Pruebas de hipótesis

Eliezer Cavazos

2024-08-23

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 sabe que la 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: Hipotesis

$$H_0: \mu = 11.7 \quad H_1: \mu \neq 11.7$$

¿Cómo se distribuye \bar{x} ?

- X se distribuye como una Normal
- $n < 30$ (Tamaño de la muestra menor a 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

```
oLatas = 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)
iN = length(oLatas)
```

```
iGradosLibertad = iN - 1
fAlfa = 0.02
```

```
t_f = abs( qt(fAlfa/2, iGradosLibertad) )
```

```
cat("t_f =", t_f)
```

```
## t_f = 2.527977
```

Regla de decision

Rechazo H_0 si:

- $|t_e| > 2.53$
- ValorP < 0.01 (Si valor P es menor a alpha)

Paso 3: Analisis del resultado

Que se tiene que calcular? * t_e : Numero 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

```
fXb = mean(oLatas)
```

```
fMiu = 11.7
```

```
fSD = sd(oLatas)
```

```
t_e = (fXb-fMiu)/(fSD/sqrt(iN))
```

```
cat("t_e = ", t_e)
```

```
## t_e = -2.068884
```

El valor p se multiplica por 2 porque estamos calculando la probabilidad de las dos colas

*valorp = 2*pt(t_e, iN-1) # Como t_e nos salio menor a 2.53 esto tambien significara que el valor p saldra mas alto que alfa, eso significa que no se rechaza*

```
cat("Valor P = ", valorp)
```

```
## Valor P = 0.0517299
```

Mas facil: Para hacer el analisis del resultado:

```
t.test(oLatas, mu=fMiu, alternative="two.sided", conf.level=0.98)
```

```
##
```

```
## One Sample t-test
```

```
##
```

```
## data: oLatas
```

```
## 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

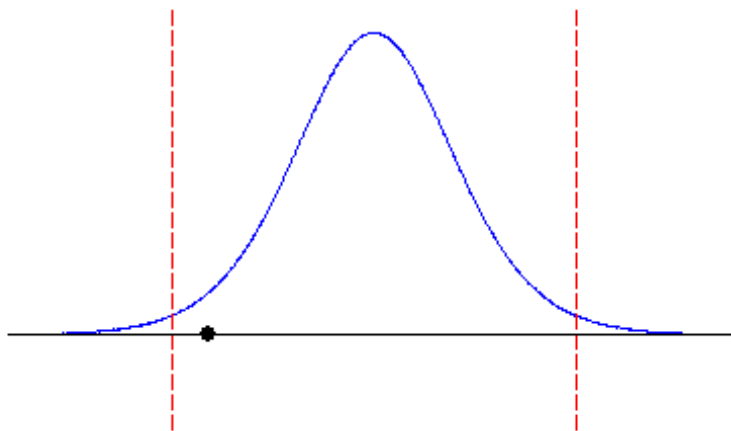
sigma = sqrt( (iN-1)/(iN-3) )
x=seq(-4*sigma,4*sigma,0.01)
y=dt(x,oLatas-1)
plot(x,y,type="l",col="blue",xlab="",ylab="",ylim=c(-0.1,0.4),frame.plot=FALSE,
xaxt="n",yaxt="n",main="Región de rechazo (distribución t de Student,
gl=20)")

#Para indicar la zona de rechazo (se ejemplifica con dos colas) y la media
(para la t de Student la  $\mu = 0$ ):

abline(v=t_f,col="red",lty=5)
abline(v=-1*t_f,col="red",lty=5)
abline(h=0)

points(fMiu,0,col="blue",pch=19)
#Para dibujar el estadístico de prueba, insertar un punto o una recta
punteada de diferente color a la zona de rechazo, para insertar un punto:
points(t_e, 0, pch=19, cex=1.1)
```

Región de rechazo (distribución t de Student, $gl=2$)



Paso 4: Conclusion

Comparar: Regla de decision vs Analisis del resultado

- $|t_e| = 2.07 < 2.53 \rightarrow$ No RH0 (no rechazo H0)
- ValorP = 0.05 > 0.02 \rightarrow No RH0

En el contexto: Las latas de durazno tienen el peso requerido

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

Paso 1: Hipotesis

$$H_0: \mu = 15 \quad H_1: \mu > 15$$

¿Como se distribuye \bar{x} ?

- X se distribuye como una Normal
- Alpha = 0.07
- $\sigma = 4$

Entonces la distribucion muestral es una Z

Paso 2: Regla de Decisión

```
oTiempo = 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)
```

```
iN = length(oTiempo)
```

```
fAlfa = 0.07
```

```
fSigma = 4
```

```
fZDs = fSigma / sqrt(iN)
```

```
Z_f = abs( qnorm(fAlfa) )
```

```
cat("Z_f =", Z_f, "\n")
## Z_f = 1.475791
cat("Z_DS =", fZDs)
## Z_DS = 0.6761234
```

Paso 3: Analisis del resultado

Se Rechaza H_0 solo si: $Z > 1.47$ ó Valor $P < 0.06$

```
fXb = mean(oTiempo)
fMiu = 15
#fSD = sd(oTiempo)

Z_e = (fXb-fMiu)/(fSigma/sqrt(iN))

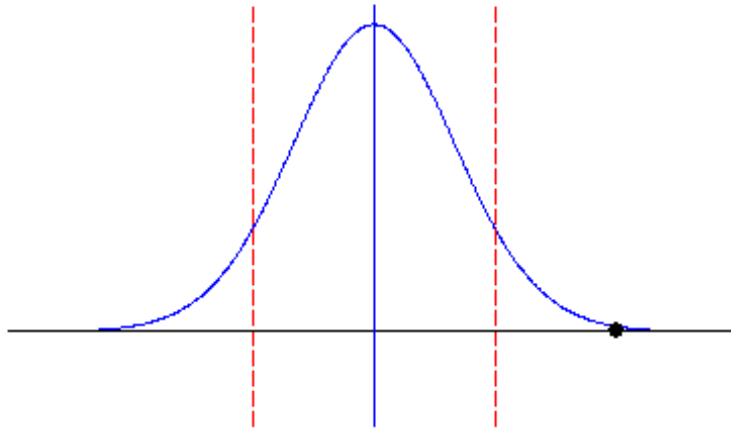
cat("Z_e = ", Z_e)
## Z_e = 2.95804
ValorP = pnorm(Z_e)

cat("Valor P = ", ValorP)
## Valor P = 0.998452

sigma =sqrt((iN-1)/(iN-3))
x=seq(-4*sigma,4*sigma,0.01)
y=dt(x,iN-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)")
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(Z_e, 0, pch=19, cex=1.1)
```

Región de rechazo (distribución Z)



Paso 4;

Conclusion

Comparar: Regla de decision vs Analisis del resultado

- $|z_e| = 2.95 > 1.47 \rightarrow$ Si RH0 (Si rechazo H0)
- ValorP = 0.99 > 0.06 \rightarrow No RH0

En el contexto: El tiempo promedio si es mayor de 15 minutos