Actividad 8. Hipotesis

Adrian Pineda Sanchez

2024-08-23

Enlatados

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: Hipotesis

$$H_0: \mu = 11.7$$

$$H_1: \mu \neq 11.7$$

Como se distribuye $\bar{X}$?

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

Paso 2: Regla de desicion

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

Necesito encontrar a cuantas desviaciones estandar estara esa regla de desicion, el valor frontera

```
# Datos de los pesos de las latas
pesos <- 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)

# Parámetros de la hipótesis
```

```r
mu_hipotetico <- 11.7
nivel_confianza <- 0.98
alpha = 1 - nivel_confianza

# Tamaño de la muestra
n <- length(pesos)

# Grados de libertad
gl <- n - 1

# Valor crítico t para el nivel de confianza (bilateral)
t_frontera <- qt(alpha/2, df = gl)
cat("t frontera =",t_frontera)

## t frontera = -2.527977
```

Regla Desicion

Rechazo H_0 si:

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

Paso 3: Analisis del resultado

- $|t_e|$: " 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

```r
# Datos de los pesos de las latas
pesos <- 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)

# Parámetros de la hipótesis
mu_hipotetico <- 11.7
nivel_confianza <- 0.98
alpha = 1 - nivel_confianza

# Cálculo de la media y desviación estándar muestral
media_muestral <- mean(pesos)
cat("La media muestral es =",media_muestral,"\n","\n")

## La media muestral es = 11.48571
##

desv_est_muestral <- sd(pesos)
cat("La desviacion es =",desv_est_muestral,"\n","\n")

## La desviacion es = 0.4746427
##
```

```r
# Tamaño de la muestra
n <- length(pesos)

# Estadístico de prueba t
t_e <- (media_muestral - mu_hipotetico) / (desv_est_muestral / sqrt(n))
cat("El Estadístico de prueba t =",t_e ,"\n","\n")

## El Estadístico de prueba t = -2.068884
##

# Grados de libertad
gl <- n - 1

# Valor crítico t para el nivel de confianza (bilateral)
t_f <- qt(alpha/2, df = gl)
cat("t frontera =",t_f,"\n","\n")

## t frontera = -2.527977
##

# P-valor
p_valor <- 2 * pt(t_e, df = gl)
cat("El valor p=",p_valor)

## El valor p= 0.0517299
```

Mas facil:

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

##
##  One Sample t-test
##
## data:  pesos
## 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: Conclusion

Comparar: Regla de decision vs Analisis del resultado

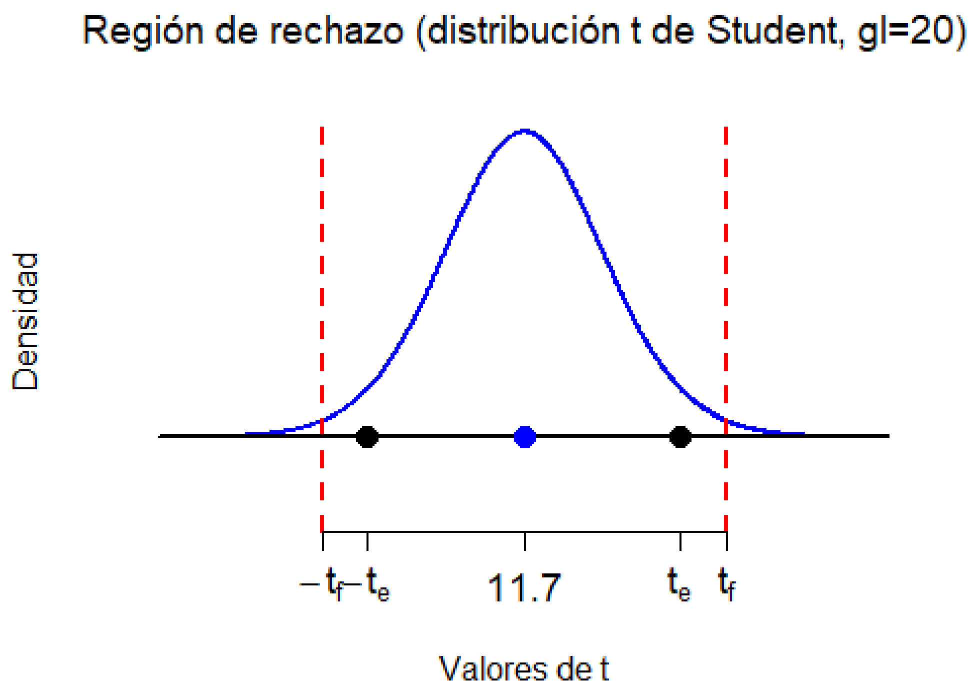
- $|t_e| = 2.06 < 2.53$ -> no Rechazo H_0
- valor p = 0.05 > 0.02 -> no Rechazo H_0

En el contexto Las latas de durazno tienen el peso requerido

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

Se grafica una secuencia de números para x que abarque 4 desviaciones estándar alrededor de la media (se ejemplifica con la t de student) con su respectivo valor de y:

```
# Parámetros
n <- 21
t_f <- 2.528  # Valor crítico t para gl=20 y 98% de confianza
t_e <- 1.96   # Ejemplo de valor del estadístico t
mu_hipotetico <- 11.7  # Media hipotética

# Calcula sigma
sigma <- sqrt((n-1)/(n-3))

# Genera los valores de x y y
x <- seq(-4 * sigma, 4 * sigma, 0.01)
y <- dt(x, df = n-1)

# Configura el gráfico
plot(x, y, type = "l", col = "blue", lwd = 2,
     xlab = "Valores de t", ylab = "Densidad",
     ylim = c(-0.1, 0.4), frame.plot = FALSE,
     xaxt = "n", yaxt = "n",
     main = expression(paste("Región de rechazo (distribución ", t, " de
Student, gl=20)")))

# Añade líneas de la región de rechazo
abline(v = t_f, col = "red", lty = 2, lwd = 2)
abline(v = -t_f, col = "red", lty = 2, lwd = 2)
abline(h = 0, col = "black", lwd = 2)

# Añade puntos para los valores del estadístico t y la media hipotética
points(0, 0, col = "blue", pch = 19, cex = 1.5)
points(t_e, 0, col = "black", pch = 19, cex = 1.5)
points(-t_e, 0, col = "black", pch = 19, cex = 1.5)

# Añade etiquetas de los ejes
axis(1, at = c(-t_f,-t_e, 0, t_e, t_f),
     labels = c(expression(-t[f]), expression(-t[e]),"11.7",
expression(t[e]), expression(t[f])),
     las = 1, cex.axis = 1.2)
```



Concluye en el contexto del problema.

La evidencia estadística no es suficiente para rechazar la hipótesis de que el peso promedio de las latas de durazno es 11.7. Esto sugiere que, dentro del nivel de confianza del 98%, no se puede concluir que el peso promedio difiera de 11.7. Por lo tanto, las latas de durazno cumplen con el peso requerido según la especificación dada.

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: Hipotesis

$$H_0: \mu = 15$$

$$H_1: \mu > 15$$

Como se distribuye $\bar{X}$?

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

Paso 2: Regla de Decisión

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

Dado que conocemos la desviación estándar $\sigma = 4$ minutos, podemos utilizar la distribución normal para calcular el valor crítico (Z crítico) asociado a un nivel de significancia de 0.07 en una prueba unilateral. Buscaremos el valor de Z para $1 - \alpha = 0.93$ en la tabla de la distribución normal.

```r
# Datos proporcionados
tiempos <- c(17, 11, 12, 23, 20, 23, 15, 16, 23, 22, 18, 23, 25, 14, 12, 19,
11, 11, 20, 21, 11, 18, 14, 13, 13, 19, 16, 10, 22, 18, 23)

# Parámetros conocidos
mu_0 <- 15  # media bajo la hipótesis nula
sigma <- 4   # desviación estándar conocida
n <- length(tiempos)  # tamaño de la muestra
alpha <- 0.07  # nivel de significancia

# Calcular el valor crítico (Z crítico) para un nivel de significancia de
0.07 en una prueba unilateral
z_critico <- qnorm(1 - alpha)
cat("El Z critico a partir es:", z_critico)

## El Z critico a partir es: 1.475791
```

Regla Desicion

Rechazo H_0 si:

- $ Z-Estadistico > 1.475791$
- valor p < 0.07

Paso 3: Analisis del resultado

- $Z-Estadistico $: ” Numero de desviaciones al que $\bar{X}$ se encuentra lejos de $\mu = 15$

- valor p: probabilidad de obtener lo que obtuve en la muestra o un valor mas extremo

```r
# Datos proporcionados
tiempos <- 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)

# Parámetros conocidos
mu_0 <- 15  # media bajo la hipótesis nula
sigma <- 4   # desviación estándar conocida
n <- length(tiempos)  # tamaño de la muestra
alpha <- 0.07  # nivel de significancia

# Paso 1: Calcular la media muestral
media_muestral <- mean(tiempos)

# Paso 2: Calcular el estadístico z
z <- (media_muestral - mu_0) / (sigma / sqrt(n))

# Paso 3: Calcular el valor crítico (z crítico) para un nivel de
significancia de 0.07 en una prueba unilateral
z_critico <- qnorm(1 - alpha)

# Paso 4: Calcular el valor p asociado
valor_p <- 1 - pnorm(z)

# Resultados
cat("Media muestral:", media_muestral, "\n")

## Media muestral: 17

cat("Estadístico Z:", z, "\n")

## Estadístico Z: 2.95804

cat("Z crítico:", z_critico, "\n")

## Z crítico: 1.475791

cat("Valor p:", valor_p, "\n")

## Valor p: 0.00154801
```

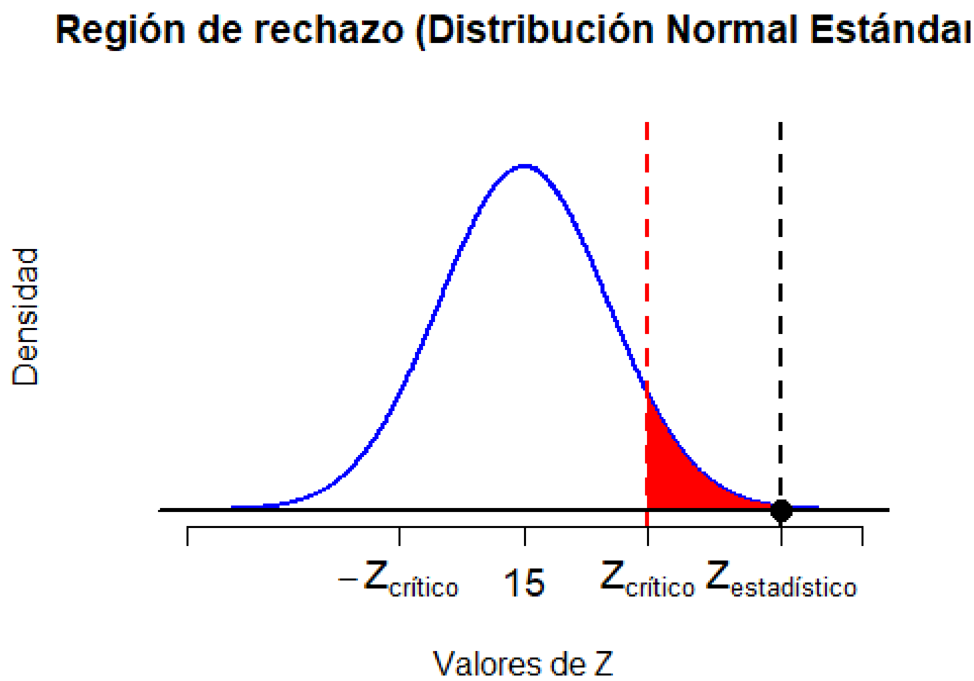
Paso 4: Conclusion

Comparar: Regla de decision vs Analisis del resultado

- $ Z-Estadistico = 3.05329 > 1.475791$ -> Rechazamos H_0
- valor p = 0.001131735 < 0.07 -> Rechazamos H_0

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

```
# Parámetros
n <- 31  # Tamaño de la muestra
z_critico <- 1.475791  # Valor crítico Z para 93% de confianza (α = 0.07)
z_estadistico <- 3.05329  # Valor del estadístico Z calculado
mu_hipotetico <- 15  # Media hipotética

# Genera los valores de x y la distribución normal estándar
x <- seq(-4, 4, 0.01)
y <- dnorm(x)

# Configura el gráfico
plot(x, y, type = "l", col = "blue", lwd = 2,
     xlab = "Valores de Z", ylab = "Densidad",
     ylim = c(0, 0.45), frame.plot = FALSE,
     xaxt = "n", yaxt = "n",
     main = "Región de rechazo (Distribución Normal Estándar)")

# Añade la región de rechazo
polygon(c(z_critico, seq(z_critico, 4, 0.01), 4),
        c(0, dnorm(seq(z_critico, 4, 0.01)), 0),
        col = "red", border = NA)

# Añade líneas de la región de rechazo y estadístico
abline(v = z_critico, col = "red", lty = 2, lwd = 2)
abline(v = z_estadistico, col = "black", lty = 2, lwd = 2)
abline(h = 0, col = "black", lwd = 2)

# Añade puntos para los valores del estadístico Z y la media hipotética
points(z_estadistico, 0, col = "black", pch = 19, cex = 1.5)

# Añade etiquetas de los ejes
axis(1, at = c(-4, -z_critico, 0, z_critico, z_estadistico, 4),
     labels = c("", expression(-Z[crítico]), "15", expression(Z[crítico]),
expression(Z[estadístico]), ""),
     las = 1, cex.axis = 1.2)
```



Concluye en el contexto del problema.

En el contexto del problema, esto significa que hay suficiente evidencia para concluir que el tiempo promedio de las encuestas telefónicas realizadas por Fowle Marketing Research, Inc. es significativamente mayor a los 15 minutos estipulados. Por lo tanto, la tarifa adicional que la empresa cobra por tiempos de encuesta que exceden los 15 minutos está justificada, ya que el tiempo promedio observado excede claramente ese límite.