

T de student para varianza desconocida

Fernanda Sarmiento

Introducción

La T de estudiante se basa en que los datos de la muestra provienen de una distribución normal, los datos son continuos y la muestra fue tomada de forma aleatoria de la población.

Nota: en caso de conocer el sigma de la población, se debe aplicar la z.

Para cualquier test realizado se emplean intervalos de confianza, estos suelen estar comúnmente entre el 95% y el 99%, siendo estos dos extremos los más empleados acordes al estudio y sus necesidades.

Como primer paso, debemos realizar una prueba de normalidad para establecer si es viable o no realizar el estudio. En ese caso, emplearemos la función `shapiro.test`.

```
x <- c(18, 11, 17, 10, 20, 25, 13, 16, 25, 20, 19, 20)
shapiro.test(x)
```

```
Shapiro-Wilk normality test
```

```
data: x
W = 0.9413, p-value = 0.5152
```

Al presentar un p-valor de 0.5152, se puede aceptar la hipótesis nula que nos indica que los datos presentan una distribución de tipo normal.

El siguiente paso es crear el intervalo de confianza con ayuda de la prueba `t.test`.

```
t.test(x)
```

```
One Sample t-test
```

```
data: x
t = 12.864, df = 11, p-value = 5.673e-08
alternative hypothesis: true mean is not equal to 0
95 percent confidence interval:
 14.78220 20.88447
sample estimates:
mean of x
 17.83333
```

El análisis nos arroja que el intervalo de confianza es de 14.78220 a 20.99447, este intervalo representa la media de la población; por otra parte, nos indica que la media de la muestra es de 17.83.

Esta función también nos brinda información adicional como el valor calculado de la $t = 12.86$, los grados de libertad que son 11, y la probabilidad resultante $p\text{-valor} = 5.673e-08$, lo cual nos da que es menor a 0.05 y esto indica que la hipótesis nula se rechaza la hipótesis nula que indica que la media es igual a 0.

A continuación, se va a realizar un análisis en el cual se trabajará con un intervalo de confianza del 99% para el rendimiento promedio.

```
t.test(x, conf.level = 0.99)
```

One Sample t-test

```
data: x
t = 12.864, df = 11, p-value = 5.673e-08
alternative hypothesis: true mean is not equal to 0
99 percent confidence interval:
 13.52788 22.13879
sample estimates:
mean of x
 17.83333
```

El análisis de confianza del 99% nos indica que la media poblacional del estudio se encuentra entre 13.52788 y 22.13879.

Para determinar otros parámetros se van a realizar algunas suposiciones para ir mejorando la comprensión y el funcionamiento de esta función.

μ diferente a 23, con un intervalo de confianza del 99%.

```
t.test(x, mu = 23, alternative = "two.sided", conf.level = 0.99)
```

One Sample t-test

```
data: x
t = -3.7271, df = 11, p-value = 0.003341
alternative hypothesis: true mean is not equal to 23
99 percent confidence interval:
 13.52788 22.13879
sample estimates:
mean of x
 17.83333
```

En este caso, la t es de -3.7271 , la desviación estándar es de 11 y el $p\text{-valor}$ de 0.003341 , lo que nos indica que la hipótesis que nos dice que media es diferente a 23 no se rechaza.

Media mayor a 23 % al 99 % de confianza.

```
t.test(x, mu = 23, alternative = "greater", conf.level = 0.99)
```

One Sample t-test

```
data: x
t = -3.7271, df = 11, p-value = 0.9983
alternative hypothesis: true mean is greater than 23
99 percent confidence interval:
 14.06537      Inf
sample estimates:
mean of x
 17.83333
```

En este caso, la t es de -3.7271, la desviación estándar es de 11 y el p-valor es de 0.9983, lo que nos indica que la hipótesis de que la media es mayor a 23 se rechaza.

Media menor a 6, al 95% de confianza.

```
t.test(x, mu = 6, alternative = "less", conf.level = 0.95)
```

One Sample t-test

```
data: x
t = 8.5362, df = 11, p-value = 1
alternative hypothesis: true mean is less than 6
95 percent confidence interval:
 -Inf 20.3229
sample estimates:
mean of x
 17.83333
```

En este caso, la t es de 8.5362, la desviación estándar es de 11 y el p-valor es de 1, lo que se rechaza la hipótesis que nos indica que la media es menor a 6.

Como último paso a realizar, se recomienda sacar algunas gráficas, entre ellas el histograma, para evidenciar la simetría de los datos.

```
par(mfrow = c (1,2))
mean(x)
```

```
[1] 17.83333
```

```
boxplot(x, col = "gray85", main = "Standar\nBoxplot")
points(mean(x), pch = 20, cex = 1.5)
text(17.5, "Promedio", col = "red", font = 8)
text(19, "Mediana", font=8)
text(14, "Cuantil 1", font = 8)
```

```

text(20.4, "Cuantil 2", font = 8)
quantile(x)

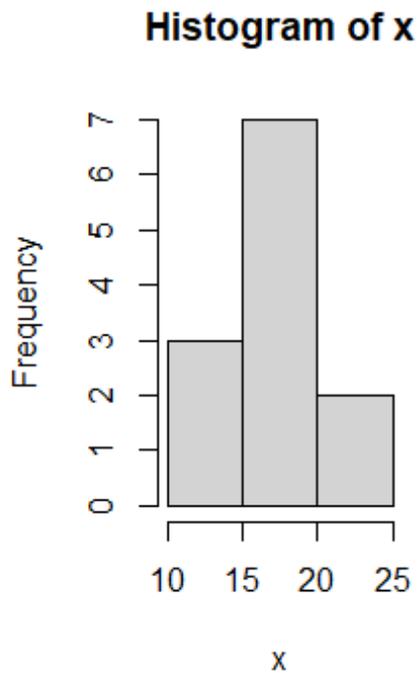
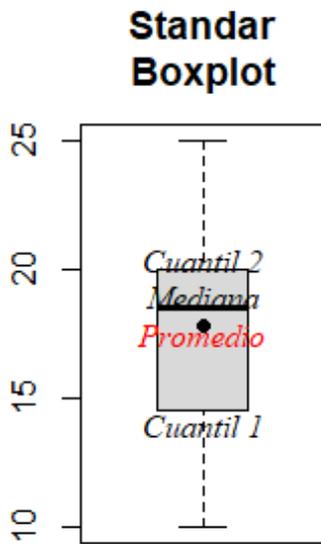
  0%   25%   50%   75%  100%
10.00 15.25 18.50 20.00 25.00

median(x)

[1] 18.5

hist(x)

```



En el boxplot obtenido se puede evidenciar que no se presentan datos atípicos, además la media está un poco por encima de la mitad del gráfico. Por otra parte, en el histograma podemos ver que presenta un leve sesgo negativo, se podría considerar leptocúrtica.