Ejemplo

En una investigación se estudian los efectos que tiene el cadmio en el crecimiento de las raíces de *Allium cepa L*. Dicho estudio quiere saber si hay diferencias entre un grupo sin exposición al cadmio contra uno que sí tiene exposición a este metal pesado.

Ofrecen las siguientes muestras:

N° de muestra	Control X_1	Cadmio X_2
1	1.4	0.4
2	1.6	0.6
3	2.4	0.8
4	2.6	0.9
5	2.9	0.5
6	3.1	1
7	3.3	1.3
8	3.6	1.5
9	3.8	1.5
10	4.1	1.7
11	4.5	1.9
12	4.7	2
13	4.7	2.1
14	5.2	2.2
15	5.4	2.4
16	5.6	2.5
17	5.7	2.6
18	6.1	2.9
19	6.2	3
20	6.4	3.1
21	6.4	3.3
22	6.7	3.3
23	6.7	3.5
24	7.3	3.7
25	7.3	4.1

Tamaño de la muestra:

$$n_1 = 25$$

$$n_2 = 25$$

Promedio de la muestra:

$$\bar{x}_1 = \frac{x_1 + x_2 \dots x_i}{n_1} = 4.708$$

$$\bar{x}_2 = \frac{x_1 + x_2 \dots x_i}{n_2} = 2.112$$

Varianza de la muestra:

$$s_1^2 = \left[\frac{\sum(x_1 - \overline{X_1})^2}{n_1 - 1}\right] = 3.101$$

$$s_2^2 = \left[\frac{\sum(x_1 - \overline{X_1})^2}{n_2 - 1}\right] = 1.164$$

De acuerdo con la **prueba de F** las varianzas son diferentes, por lo que:

a. **Establecer hipótesis.**

$H_0: \mu_1 = \mu_2$ El promedio del grupo 1 (sin cadmio) es igual al promedio del grupo 2 (con Cadmio).
$H_1: \mu_1 \neq \mu_2$ El promedio del grupo 1 (sin cadmio) es diferente al promedio del grupo 2 (con Cadmio).

b. **Dibujar región de rechazo y no rechazo con un $\alpha = 0.05$**

La prueba indica que es de **dos colas** por lo que se ha de considerar $\frac{\alpha}{2} = 0.025$

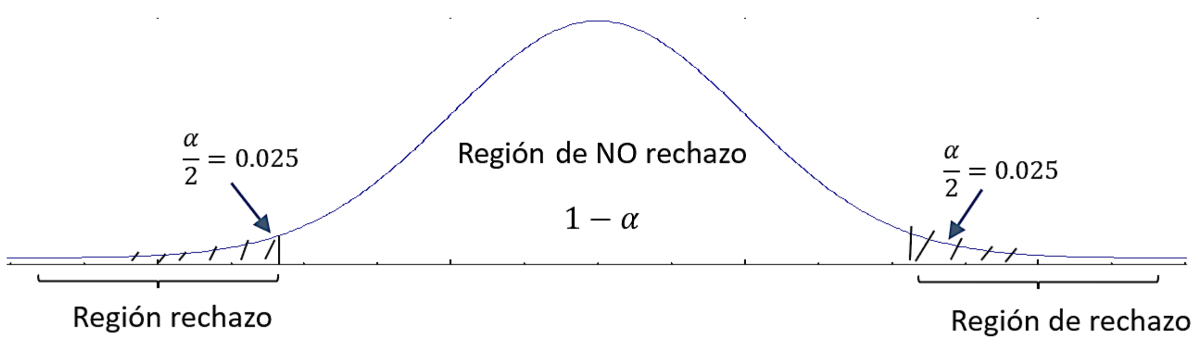

Figura 1. Región de rechazo y no rechazo para una prueba bilateral con $\frac{\alpha}{2} = 0.025$

c. Se calcula el estadístico de prueba

$$t_{cal} = \frac{\bar{X}_1 - \bar{X}_2}{\sqrt{\left(\frac{S_1^2}{n_1} + \frac{S_2^2}{n_2}\right)}} = \frac{4.708 - 2.112}{\sqrt{\left(\frac{3.101}{25} + \frac{1.164}{25}\right)}} = 6.285$$

Obtener valor teórico

$$t_{teo} = t_{(gl, \frac{\alpha}{2})} = t_{(40, 0.025)} = 2.423$$

Donde

$$gl = \frac{\left(\frac{S_1^2}{n_1} + \frac{S_2^2}{n_2}\right)^2}{\frac{\left(\frac{S_1^2}{n_1}\right)^2}{n_1 - 1} + \frac{\left(\frac{S_2^2}{n_2}\right)^2}{n_2 - 1}} = \frac{\left(\frac{3.101}{25} + \frac{1.164}{25}\right)^2}{\frac{\left(\frac{3.101}{25}\right)^2}{24} + \frac{\left(\frac{1.164}{25}\right)^2}{24}} = 39.79 \sim 40$$

Nota: para buscar el valor teórico, se requieren tablas de distribución, éstas son cálculos que representan valores estandarizados. Por lo general los grados de libertad se encuentran en la parte izquierda de la tabla y se representan con una v, mientras que la significancia se encuentra en la parte superior representados por α o bien $\frac{\alpha}{2}$, según sea el caso. Una vez se conozcan las tablas que se van a utilizar, se busca la intersección entre los grados de libertad (el valor que se calculó con la fórmula) con el valor de significancia α.

	0.1	0.05	0.025	0.01	0.005	α
v	0.20	0.10	0.05	0.02	0.01	$\frac{\alpha}{2}$
1	3.078	6.314	12.706	31.821	63.657	
2	1886	2.920	4.303	6.965	9.9925	
…						
40	1.303	1.684	2.021	2.423	2.704	
…						
∞	1.282	1.645	1.960	2.326	2.576	

d. Aplicar regla de decisión para rechazar o no H_0.

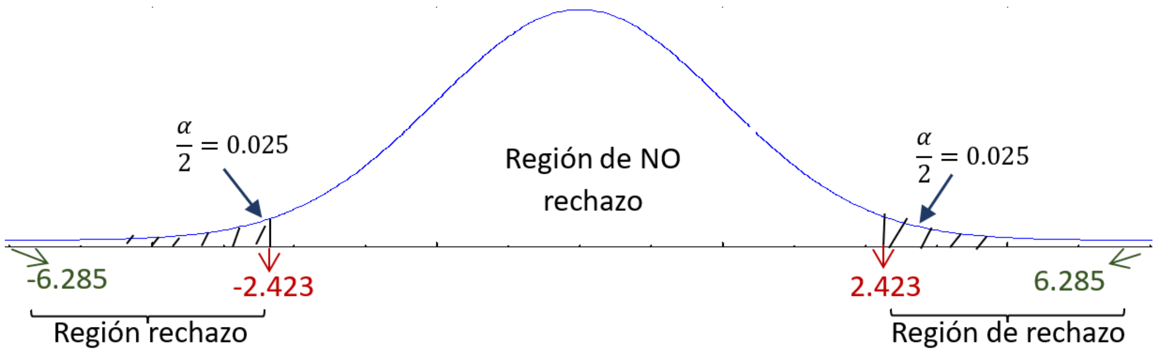

Fig. 2 El valor de t calculada cae en las regiones de rechazo.

Entonces, cómo $t_{cal} = 6.285 > t_{teo} = 2.423$ la hipótesis nula se rechaza.

e. Concluir el problema con texto

Con un valor de confianza del 95% se puede afirmar que el contenido promedio de cadmio es diferente entre los dos grupos.