Análisis paramétrico:

Pearson, T Student

Filosofía

INFORMÁTICA

Análisis Paramétrico:

Introducción

Análisis Paramétrico

- **¿Qué es?**

Conjunto de pruebas estadísticas que asumen que los datos provienen de distribuciones conocidas (normal o aproximadamente normal).

- **Requisitos comunes:**
  - Escala de intervalo o razón
  - Distribución normal
  - Varianzas homogéneas (según la prueba)

Correlación de Pearson

- **Objetivo:**

Evaluar la fuerza y dirección de la relación lineal entre dos variables cuantitativas.

- **Coeficiente (r):**

$$r = \frac{\sum (x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum (x_i - \bar{x})^2 \sum (y_i - \bar{y})^2}}$$

Valores:

- r = 1: correlación positiva perfecta
- r = -1: correlación negativa perfecta
- r = 0: sin correlación lineal

Ejemplo 1:

Pearson

Datos:

Horas de estudio: [1, 2, 3, 4, 5]
Nota en examen: [2, 4, 5, 4, 5]

Cálculos:

$$\bar{x} = 3, \quad \bar{y} = 4$$

$$r = \frac{(1-3)(2-4) + (2-3)(4-4) + (3-3)(5-4) + (4-3)(4-4) + (5-3)(5-4)}{\sqrt{[(1-3)^2 + ...][(2-4)^2 + ...]}} = 0.8$$

Interpretación: Correlación positiva alta

Prueba T de Student

(2 muestras independientes)

¿Para qué sirve?

Comparar las medias de dos grupos independientes para ver si la diferencia es significativa.

Fórmula:

$$t = \frac{\bar{X}_1 - \bar{X}_2}{\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}}$$

🔍 Elementos de la fórmula explicados:

- $\bar{X}_1$: Media muestral del grupo 1
- $\bar{X}_2$: Media muestral del grupo 2
- s_1^2: Varianza muestral del grupo 1
- s_2^2: Varianza muestral del grupo 2
- n_1: Tamaño de la muestra del grupo 1
- n_2: Tamaño de la muestra del grupo 2
- t: Valor estadístico que indica cuántas desviaciones estándar hay entre las medias muestrales

□ Interpretación del resultado:

- Si $|t| > t_{crítico}$ (de la tabla t de Student con gl grados de libertad), se **rechaza** H_0 (hay diferencia significativa).
- Si $|t| \leq t_{crítico}$, **no se rechaza** H_0 (no hay evidencia suficiente de diferencia).

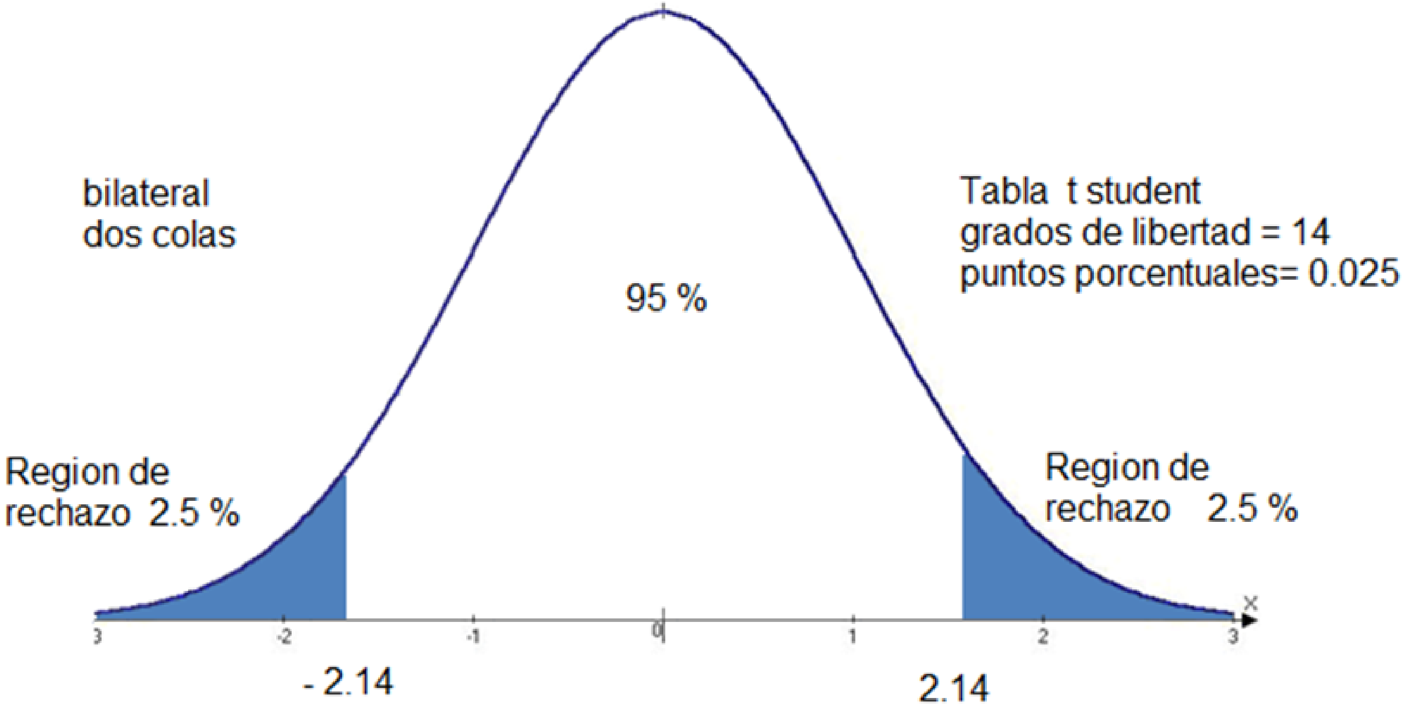
Curva de Probabilidad

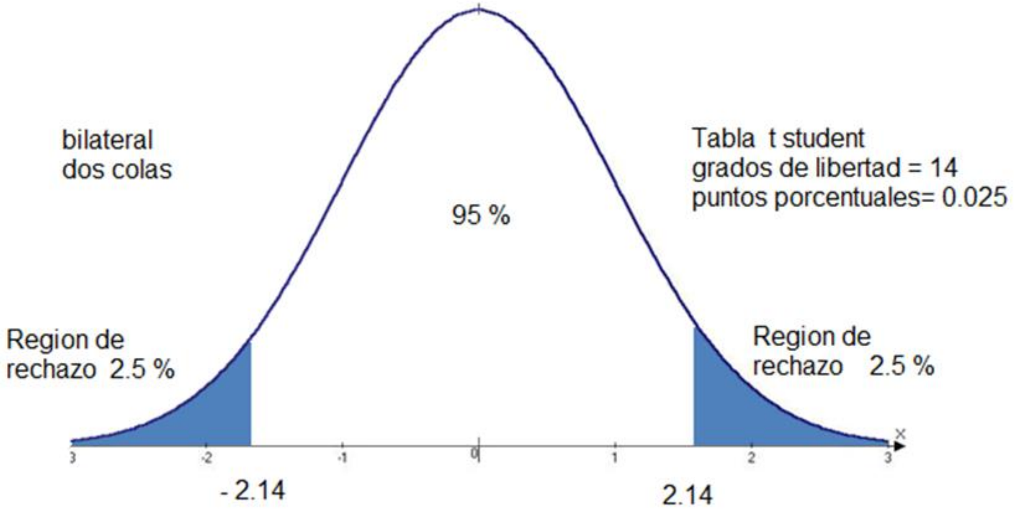
T de student

🎯 **Tipo de prueba representada:**

- **Prueba bilateral (dos colas)**: estamos analizando si **hay diferencia significativa** (sin importar la dirección).
- Se usa cuando la hipótesis alternativa es:

$$H_1 : \mu_1 \neq \mu_2$$





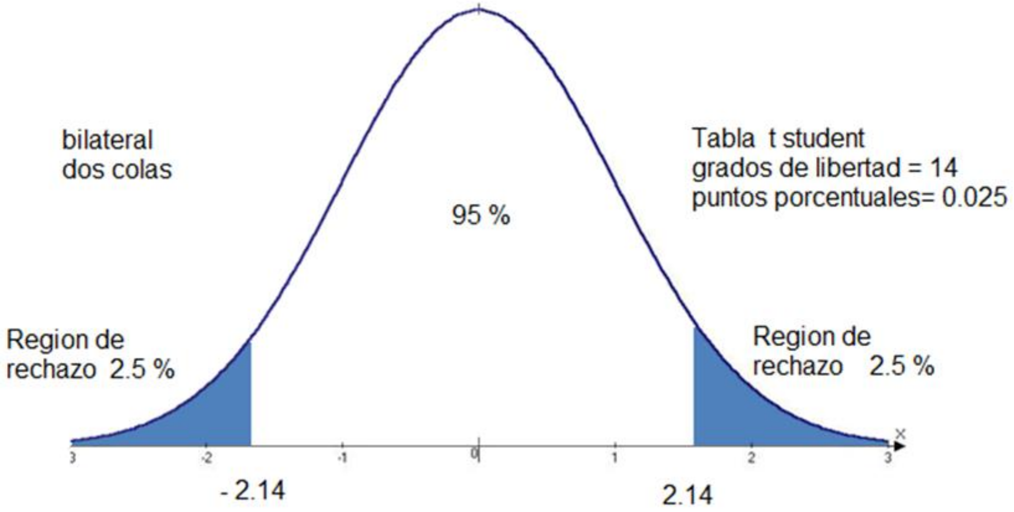
📊 Componentes de la gráfica:

✅ Región central (blanca):

- Área = 95%
- Corresponde al intervalo de **no rechazo de la hipótesis nula (H_0)**.
- Si el valor calculado de t cae dentro de este intervalo (entre -2.14 y 2.14), **no hay evidencia suficiente para rechazar H_0.**

🚫 Regiones sombreadas (azul, 2 colas):

- **Área total = 5% (2.5% en cada cola)**
- Representan la **región de rechazo** para un nivel de significancia $\alpha = 0.05$.
- Si el valor de t cae en cualquiera de estas regiones, **rechazamos H_0** porque la diferencia observada es estadísticamente significativa.



Valor crítico t = ±2.14

- Este valor se obtiene de la **tabla t de Student** para:
  - **Grados de libertad (gl) = 14**
  - **Nivel de significancia** $\alpha = 0.05$ (distribuido como 0.025 a cada cola)
- **Interpretación:**
  - Si t < -2.14 o t > 2.14 → **Rechazamos H_0**
  - Si -2.14 ≤ t ≤ 2.14 → **No se rechaza H_0**

📌 Resumen para tomar decisiones

Valor t calculado	Conclusión
$t < -2.14$ o $t > 2.14$	Rechazar H_0 (diferencia significativa)
$-2.14 \leq t \leq 2.14$	No rechazar H_0 (no hay evidencia de diferencia)

✅ ¿Qué son los **grados de libertad?**

> En estadística, los **grados de libertad (gl)** indican **cuántos valores de un conjunto de datos pueden variar libremente**, sin violar una restricción (como una media fijada).

Cuando impones una condición (por ejemplo, que la media sea un valor específico), **no todos los valores pueden cambiar libremente.** Esa restricción reduce la libertad de elección.

Supón que tienes 3 valores x_1, x_2, x_3 cuya media debe ser **exactamente 10.**

Es decir:

$$\frac{x_1 + x_2 + x_3}{3} = 10 \Rightarrow x_1 + x_2 + x_3 = 30$$

Paso a paso:

- Puedes **elegir libremente dos valores**, por ejemplo:
  - $x_1 = 9$
  - $x_2 = 11$
- Entonces el tercer valor ya **no puede ser cualquiera:**

$$x_3 = 30 - (x_1 + x_2) = 30 - 20 = 10$$

Solo **dos valores** pudieron elegirse libremente. El tercero **queda determinado por la restricción** (la media fija).

Gados de libertad

🎯 Vamos paso a paso:

✅ 1. ¿Qué estás analizando con la prueba de hipótesis?

- ¿Quieres comparar medias entre dos grupos?
- ¿Quieres ver correlación entre dos variables?
- ¿Vas a aplicar un test t de Student, ANOVA, chi-cuadrado, etc.?

🔁 El tipo de prueba determina cómo se calculan los grados de libertad.

Pearson

📌 **Caso A: Correlación (por ejemplo: Pearson)**

Supongamos que tienes dos variables (independiente y dependiente), y aplicaste tu cuestionario a n personas.

Grados de libertad:

$$gl = n - 2$$

> Se restan 2 porque se estima la media de ambas variables.

T de Sudent

FR

📌 **Caso B: Prueba t de Student para dos muestras independientes**

Supón que tienes dos grupos (por ejemplo: control y experimental), con n_1 y n_2 personas respectivamente.

Grados de libertad:

$$gl = n_1 + n_2 - 2$$

Xi cuadrado

📌 **Caso C: Prueba de chi-cuadrado para independencia (ej. cruces entre variables categóricas)**

Supón que quieres saber si la variable independiente está asociada con alguna variable categórica (por ejemplo, respuestas agrupadas).

Grados de libertad:

$$gl = (r - 1)(c - 1)$$

> Donde:

- r = número de filas (niveles de una variable)
- c = número de columnas (niveles de la otra variable)

Observación

⚠️ **OJO: No confundir "número de preguntas" con "número de observaciones"**

- Si hiciste un cuestionario con 11 ítems pero encuestaste a 50 personas, los grados de libertad dependen de **esas 50 personas (n)**, no del número de ítems directamente.
- El número de ítems sí puede afectar el **análisis factorial**, **validez** o **consistencia interna**, pero no los grados de libertad en hipótesis estadísticas básicas como correlaciones, t-test o chi-cuadrado.

Ejemplo

📝 Ejercicio 2 (T Student)

Un nuevo medicamento se prueba en dos grupos:

- Grupo A: 14, 15, 16, 14, 15 (n=5)
- Grupo B: 10, 11, 9, 12, 10 (n=5)

¿El medicamento tiene efecto significativo? (α = 0.05, bilateral)

✅ Solución al Ejercicio 2

$$\bar{X}_A = 14.8, \quad \bar{X}_B = 10.4, \quad s_A^2 = 0.7, \quad s_B^2 = 1.3$$

$$t = \frac{14.8 - 10.4}{\sqrt{\frac{0.7}{5} + \frac{1.3}{5}}} = \frac{4.4}{\sqrt{0.14 + 0.26}} = \frac{4.4}{0.77} \approx 5.71$$

*Valor crítico bilateral t ≈ 2.306**

Conclusión: Como 5.71 > 2.306, rechazamos H_0. Hay evidencia de efecto.

¿Cómo se obtiene el valor t* de la tabla t de Student?

Necesitas 3 elementos clave:

1. **Nivel de significancia (α):** Es la probabilidad máxima que aceptas de cometer un error tipo I (rechazar H_0 siendo verdadera).
    - Comúnmente: α = **0.05** (5%)
2. **Tipo de prueba:**
    - **Bilateral (dos colas)** si quieres saber si hay diferencia (≠).
    - **Unilateral (una cola)** si quieres saber si una media es mayor o menor (>, <).
3. **Grados de libertad (gl):**
    - Para dos muestras:

$$gl = n_1 + n_2 - 2$$

🔍 Consulta en la tabla t (bilateral, α = 0.05, gl = 8):

$$t^* = \boxed{2.306}$$

¿De dónde sale este número?

Lo sacas de una **tabla t de Student** como esta:

gl	0.10 (bilat)	0.05 (bilat)	0.025 (bilat)
8	1.860	2.306	2.896
9	1.833	2.262	2.821
10	1.812	2.228	2.764

Columna usada: $\alpha = 0.05$ bilateral → 2.5% en cada cola

Fila usada: $gl = 8$

📌 Comparación final

- Valor t calculado: 5.71
- Valor t crítico: 2.306
- Como $5.71 > 2.306$, está en la **región de rechazo**, por lo tanto:

$$\boxed{\text{Rechazamos H}_0\text{. Hay evidencia significativa de efecto.}}$$

FIN