

# Resumen de Análisis Estadístico Avanzado

Técnicas estadísticas para mejorar la toma de decisiones.  
importancia y algunas definiciones

Fortalece la capacidad para interpretar y aplicar análisis de datos complejos.

 por conectiva oficial





# Objetivo

Ir mas allá de los datos observados

Ejemplo: ¿La edad promedio de los hombres solteros es diferente a la de los casados?

¿Cómo?

Mediante test estadísticos que nos dicen si una diferencia es real o puede deberse al azar.

# Hipótesis, $H_0$ y $H_1$

## ¿Qué es una hipótesis?

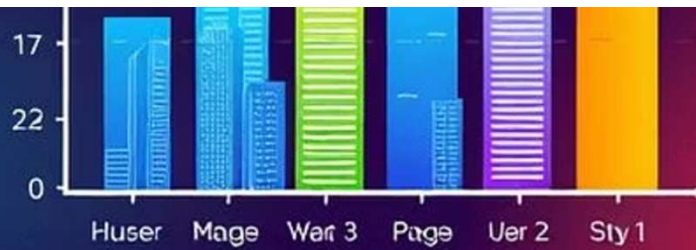
Es una propuesta que se quiere probar con los datos.

- $H_0$  (hipótesis nula): No hay efecto / No hay diferencia / No hay relación.
- $H_1$  (hipótesis alternativa): Sí hay efecto / Hay diferencia / Hay relación.

## ¿Cómo funciona?

- Se parte asumiendo que  $H_0$  es cierta.
- El test estadístico calcula la probabilidad de observar tus datos si  $H_0$  fuera cierta.
- Si esa probabilidad es muy baja ( $p < 0.05$ ), se rechaza  $H_0$ .





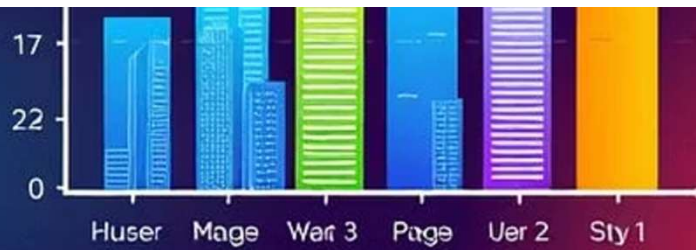
# p-value

## ¿Qué es?

Es la probabilidad de resultados observados bajo la hipótesis nula

## Interpretación

p-valor	Conclusión
$< 0.05$	Hay evidencia contra $H_0 \rightarrow$ Rechazar
$\geq 0.05$	No hay evidencia suficiente $\rightarrow$ No rechazar



# Tabla de contingencia

## ¿Qué es?

Una tabla que resume frecuencias entre dos variables categóricas

Sexo	Soltero	Casado
Hombre	50	30
Mujer	40	35



# Chi-cuadrado ( $\chi^2$ )

¿Qué hace?

Evalúa si hay relación entre dos variables categóricas..

¿Ejemplo?

¿El estado civil está relacionado con el sexo?

Sexo	Soltero	Casado
Hombre	50	30
Mujer	40	35

- Se aplica  $\chi^2$
- Si  $p < 0.05 \rightarrow$  hay relación entre sexo y estado civil



# Anova

## ¿Qué hace?

Compara la media de una variable numérica entre más de dos grupos para detectar diferencias.

## ¿Ejemplo?

¿La edad promedio varía según el estado civil?

Estado civil	Media Años
Soltero	25
Casado	35
Divorciado	40

- Anova dice si algún grupo es diferente  $p < 0.05$
- No dice que grupo. Para ello se utiliza Tukey





# Test de Tukey

## ¿Qué son?

- Después de ANOVA, el test de Tukey compara los grupos de a dos para ver dónde está la diferencia.
- Se usa tras rechazo de hipótesis nula.

## ¿Por qué usarlo?

- ANOVA solo dice “hay diferencia”
- Tukey dice “entre qué grupos”



# Regresión lineal

## ¿Qué hace?

- Estima cómo una o varias variables explican o predicen otra variable numérica.
- Modela variable dependiente mediante variables independientes.
- R-Cuadrado ( $R^2$ ) Valor cercano a 1 indica mejor ajuste.
- $y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \dots + \varepsilon$

## Ejemplo

Predecir ingresos según edad y educación

$$\text{Ingreso} = \beta_0 + \beta_1 \cdot \text{Edad} + \beta_2 \cdot \text{Educación} + \text{error}$$



# Variables dummies

## ¿Qué son?

- Variables categóricas convertidas a números binarios (0 y 1) para usar en modelos estadísticos (como regresión).
- Incorporan datos cualitativos en análisis cuantitativos.

## Ejemplo

Soltero	Casado	Divorciado
1	0	0
0	1	0
0	0	1

- Anova dice si algún grupo es diferente  $p < 0.05$
- No dice que grupo. Para ello se utiliza Tukey

# Colinealidad o multicolinealidad

## ¿Qué es?

- Cuando dos o más variables independientes están muy relacionadas entre sí.
- Una variable predice muy bien a otra
- Esto genera problemas al estimar sus efectos individuales en una regresión

## Ejemplo

$$\text{ingreso} = \beta_0 + \beta_1 \cdot \text{edad} + \beta_2 \cdot \text{casados} + \beta_3 \cdot \text{divorciados}$$

S./ Convertir en variable dummie y eliminar la primera columna

Estado civil	Casado	Divorciado
Soltero (referencia)	0	0
Casado	1	0
Divorciado	0	1

## Interpretación

Los coeficientes se interpretan en comparación con el grupo "Soltero":

- Ingreso estimado para un soltero de edad 0
- Por cada año más de edad, el ingreso sube \$10.23
- $\beta_2$  (casado) = Un casado gana \$301 más que un soltero
- $\beta_3$  (divorciado) = Un divorciado gana \$98 menos que un soltero



# Modelo de ecuaciones estructuradas (SEM)

## ¿Qué hace?

- Analiza relaciones complejas y causales entre múltiples variables.
- Incluye análisis factorial confirmatorio y trayectoria.