

# Reto 04

## Econometría I | Universidad del Norte

Carlos Yanes Guerra

### Antes de empezar

Recuerde que los retos son para desarrollarlos en clases en el tiempo sugerido por el profesor 1H:20 Min. Trate con su grupo de trabajo (máximo 2 personas) este -tenga un toque único- donde haga uso en la gran mayoría de los códigos de clase y no los que les da la IA.

### Objetivo

Hacer uso del modelo base de Machine learning (regresión)

### Datos a usar

Sea la siguiente tabla con información:

Obs	Nombre	Edad	Altura (cm)	Peso (kg)	Horas de Ejercicio
1	Ana Martinez	30	165	60	5
2	Carlos Perez	45	175	80	3
3	Laura Gomez	25	160	55	6
4	Juan Rodriguez	35	180	85	4
5	Maria Lopez	28	170	65	5
6	Pedro Sanchez	40	178	78	2
7	Lucia Fernandez	32	168	62	4
8	Jose Ramirez	50	182	90	1
9	Sofia Torres	27	162	58	5
10	Andres Suarez	38	176	82	3
11	Paula Castro	33	167	63	4
12	Miguel Vargas	29	174	75	3
13	Elena Morales	26	161	57	6
14	Diego Rios	42	179	83	2

Obs	Nombre	Edad	Altura (cm)	Peso (kg)	Horas de Ejercicio
15	Natalia Ruiz	31	166	61	5

## Preguntas del Reto

1. Establezca primero con las variables (Altura y Peso) un **diagrama de dispersión**, luego haga lo mismo pero ahora con las variables (Altura y Horas de Ejercicio). Realice un análisis comparativo de ambas gráficas.
2. Haga un análisis del peso por edad. Pista: Puede ayudarse del comando `Table` de R.
3. Cree una variable *dicotómica* o dummy para el género de la persona y establezca una comparación entre el peso por género de las personas de la base. Pista: Debe responder finalmente cuál género tiene mayor peso promedio.
4. Realice una prueba estadística para determinar si el promedio de horas de ejercicio es diferente por género. Pista: Haga uso de la **prueba T de Student**. Explique y plantee la prueba de hipótesis. Corra ahora una *regresión con las variables*. ¿Difiere de lo obtenido en la prueba T de Student? Explique
5. Ejecute una regresión adicional, pero ahora cuando la variable dependiente es el peso y la explicativa es la altura. Pista: Muestre los resultados, incluyendo valores predichos y residuales e interprete el modelo.