

# Análisis exploratorios

Santos G

## Tabla de contenidos

<b>1 Contexto del proyecto</b>	<b>1</b>
<b>2 Carga y verificación inicial de datos</b>	<b>1</b>

```
1 # Librerías
2 library(tidyverse) # Manipulación de datos: dplyr, tidyr, readr
3 library(janitor)   # Limpieza: clean_names(), tabyl()
4 library(ggplot2)   # Gráficos profesionales
5 library(skimr)     # EDA rápido y completo (skim())
6 library(GGally)    # Matriz de gráficos para variables múltiples
7 library(knitr)     # Tablas en Quarto
8 library(kableExtra) # Tablas formateadas para informes
```

## 1 Contexto del proyecto

Se realizó una exploración y control de calidad de los datos de entrada para identificar variables relevantes, evaluar supuestos básicos y priorizar rutas analíticas. El objetivo es generar una guía reproducible que permita a futuros analistas (o a un equipo de consultoría) replicar y ampliar los análisis según objetivos específicos (p. ej. comparar tratamientos, modelar abundancias o construir índices de condición).

## 2 Carga y verificación inicial de datos

```
1 #|label: data-load
2 # Carga de datos (ejemplo iris) y limpieza mínima
3 data("iris")
4 df <- as_tibble(iris) %>%
5   janitor::clean_names() # convierte a snake_case: sepal_length, etc.
6
7 # Información básica
8 n_rows <- nrow(df); n_cols <- ncol(df)
9 glimpse(df)
```

```

Rows: 150
Columns: 5
$ sepal_length <dbl> 5.1, 4.9, 4.7, 4.6, 5.0, 5.4, 4.6, 5.0, 4.4, 4.9, 5.4, 4.~
$ sepal_width  <dbl> 3.5, 3.0, 3.2, 3.1, 3.6, 3.9, 3.4, 3.4, 2.9, 3.1, 3.7, 3.~
$ petal_length <dbl> 1.4, 1.4, 1.3, 1.5, 1.4, 1.7, 1.4, 1.5, 1.4, 1.5, 1.5, 1.~
$ petal_width  <dbl> 0.2, 0.2, 0.2, 0.2, 0.2, 0.4, 0.3, 0.2, 0.2, 0.1, 0.2, 0.~
$ species      <fct> setosa, setosa, setosa, setosa, setosa, setosa, setosa, s~

```

```

1 tbl1<-skim(df)
2 tbl1 # Resumen compacto por variable

```

Tabla 1: Data summary

Name	df
Number of rows	150
Number of columns	5
Column type frequency:	
factor	1
numeric	4
Group variables	None

#### Variable type: factor

skim_variable	n_missing	complete_rate	ordered	n_unique	top_counts
species	0	1	FALSE	3	set: 50, ver: 50, vir: 50

#### Variable type: numeric

skim_variable	n_missing	complete_rate	mean	sd	p0	p25	p50	p75	p100	hist
sepal_length	0	1	5.84	0.83	4.3	5.1	5.80	6.4	7.9	
sepal_width	0	1	3.06	0.44	2.0	2.8	3.00	3.3	4.4	
petal_length	0	1	3.76	1.77	1.0	1.6	4.35	5.1	6.9	
petal_width	0	1	1.20	0.76	0.1	0.3	1.30	1.8	2.5	

El dataset contiene **N = 150 observaciones** y **5 variables**. Cuatro son cuantitativas continuas en centímetros (*Sepal.Length*, *Sepal.Width*, *Petal.Length*, *Petal.Width*), y una categórica (*Species*), que clasifica en tres grupos balanceados ( $n = 50$  por especie). No se detectaron valores faltantes ni duplicados tras la inspección inicial. Esta estructura balanceada y sin NA permite aplicar análisis univariados, comparativos y multivariados con mínimo preprocesamiento.

La tabla de estadísticos descriptivos muestra lo siguiente:

- **Sepal.Length:** media 5.84 cm, SD 0.83, rango 4.3–7.9. Variación moderada, con solapamiento esperado entre especies.

- **Sepal.Width:** media 3.06 cm, SD 0.44, rango 2.0–4.4. Es la variable más estable, aunque con ligera asimetría negativa.
- **Petal.Length:** media 3.76 cm, SD 1.77, rango 1.0–6.9. Mayor dispersión relativa, con clara separación de *setosa*.
- **Petal.Width:** media 1.20 cm, SD 0.76, rango 0.1–2.5. Alta variabilidad, con potencial de discriminación entre las tres especies.

Aspectos destacados del dataset:

- **Escala homogénea de medidas:** todas las variables en centímetros → comparaciones y análisis multivariados sin necesidad de reescalado inmediato.
- **Colinealidad esperada:** Petal.Length y Petal.Width muestran alta correlación, lo que debe considerarse en regresiones o PCA.
- **Grupos biológicos claros y balanceados:** un escenario ideal para aprendizaje, aunque poco frecuente en estudios ecológicos reales.
- **Potencial de discriminación:** las variables de pétalos concentran el mayor poder de separación, coherente con su relevancia funcional en la biología reproductiva de las plantas.