

# S12-P2 (Sesión 24)

## Repaso guiado

Departamento de Física.

---

Corodinadora: C Loyola

Profesores C Femenías / F Bugini / D Basantes

Primer Semestre 2025

Universidad Andrés Bello

Departamento de Física y Astronomía



## Ruta de la sesión (60 min)

Repaso & Calentamiento (40 min)

Trabajo en equipo (reforzar)

Cierre

## Repaso & Calentamiento (40 min)

---

## ¿Recuerdas...?

1. ¿Para qué sirve `df.info()`?
2. ¿Cómo obtengo la moda de una columna numérica?
3. ¿Qué gráfico elegirías para detectar *outliers*? <sup>1</sup>

Respuestas rápidas en voz alta, sin nota.

---

<sup>1</sup>En un conjunto de datos, un **outlier** (dato atípico) es un valor que se aleja de manera notable del patrón general. Respuesta : **boxplot**

# Repaso & Calentamiento (40 min) ∈ Creamos un dataset sintético de ¡pandas gigantes! (7 min)

```
1 import pandas as pd, numpy as np
2 np.random.seed(42)
3 n = 200
4 df_syn = pd.DataFrame({
5     "height_m": np.random.normal(2.5, 0.3, n),      # altura
6     "weight_kg": np.random.normal(150, 15, n),      # peso
7     "bamboo_daily_kg": np.random.normal(25, 4, n),  # bambú
8     ↪ ingerido
9     "continent": np.random.choice(
10         ["Asia", "Europa", "América"], n, p=[.6,.25,.15])
11 })
```

- Datos inventados → sólo práctica.
- Recordamos `head()`, `describe()`, `value_counts()`.

## Repaso & Calentamiento (40 min) ∈ Trucos rápidos de inspección (5 min)

```
1 df_syn.isnull().sum()                                # ¿faltantes?  
2 df_syn.groupby("continent")["weight_kg"].mean()  
3 df_syn.quantile([.25, .75])                          # cuartiles
```

## Repaso & Calentamiento (40 min) ∈ ¿Usemos seaborn?

- Se apoya en **matplotlib** todo lo que aprendes de la *librería base* sigue siendo válido.
- **Ajustes estéticos por defecto:** paletas perceptualmente uniformes, estilos limpios y tipografías coherentes sin configuración extra.
- **API de alto nivel:** llamadas concisas (**`sns.boxplot`**, **`sns.violinplot`**, **`sns.regplot`**) que automatizan pasos de formateo y estadísticas.
- **Gráficos estadísticos listos** (distribuciones, relaciones, catplots) sin escribir cálculos preliminares.
- **Uso en investigación:** astrónomos y astrofísicos lo emplean para explorar catálogos masivos (Gaia, SDSS) y publicar figuras claras en revistas científicas.
- Conocemos **matplotlib** porque es *el* motor de renderizado, pero mostramos **seaborn** para que veas que existen atajos y estilos listos para producción.

### Mensaje clave

Aprende la base (**matplotlib**), pero no dudes en aprovechar atajos como **seaborn** cuando necesites rototipar visualizaciones estadísticas rápidamente.

```
1 # Scatter altura vs peso
2 df_syn.plot.scatter("height_m", "weight_kg",
3                     title="Pandas gigantes sintéticos")
4
5 # Heatmap de correlación exprés
6 import seaborn as sns, matplotlib.pyplot as plt
7 sns.heatmap(df_syn.corr(numeric_only=True), annot=True,
8             ↪ square=True)
9 plt.title("Matriz de correlación")
```

- Repasamos scatter y heatmap.
- Discusión rápida: ¿correlación altura-peso?, ¿ingesta de bambú?



Repaso completado. ¡Vamos a la Tarea!

Recuerde archivar su notebook para entrega.

## Trabajo en equipo (reforzar)

---

- 500 estrellas (muestra HIPPARCOS).
- Archivo csv : [https://gitarra.cl/lectures/gfiles/-/raw/main/pcfi161/S12/stars\\_brightness.csv](https://gitarra.cl/lectures/gfiles/-/raw/main/pcfi161/S12/stars_brightness.csv)
- Columnas: `magnitude_app`, `temperature_K`, `spectral_class`, `radius_solar`, y `metallicity_Fe_H`.

Trabajo en grupos de 2-3

1. Leer el CSV con **pandas**.
2. Calcular: media, mediana, desvío estándar, IQR.
3. Graficar:
  - Histograma de **magnitude\_app** (20 bins)
  - Boxplot de **temperature\_K** por **spectral\_class**
  - *Opcional*: violin plot comparativo
4. Guarden su trabajo para estudiar.

Cierre

---

- Guarden/descarguen su notebook antes de terminar la clase.
- Si tiene dudas: ¡consulte!.
- ¡Prepárense para la Solemne 2!