

# TP4 Extra: Análisis de componentes principales

## Grupo 2

Tomás Álvarez Escalante (60127)

Alejo Francisco Caeiro (60692)

Lucas Agustín Ferreiro (61595)

Román Gómez Kiss (61003)

# ¿Por qué estandarizamos?

Si alguna de las variables toma valores mayores a los demás, tendrá mayor varianza, pero no necesariamente mayor variabilidad. Si las escalas de medida de las variables son muy distintas, la maximización de la varianza depende fuertemente de estas escalas de medida y las variables con valores más grandes tendrán más peso en el análisis.

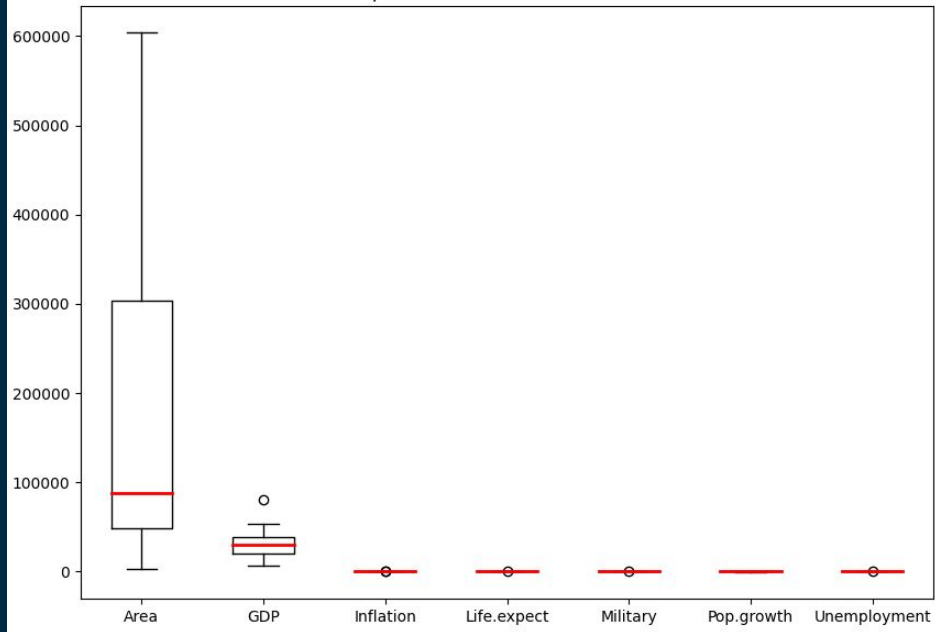
Para evitar este problema, **se deben estandarizar las variables cuando calculamos las componentes principales**. De esta manera, las magnitudes de los valores numéricos de las variables originales serán comparables.

$$\tilde{X}_i = \frac{X_i - \bar{X}_i}{s_i}$$

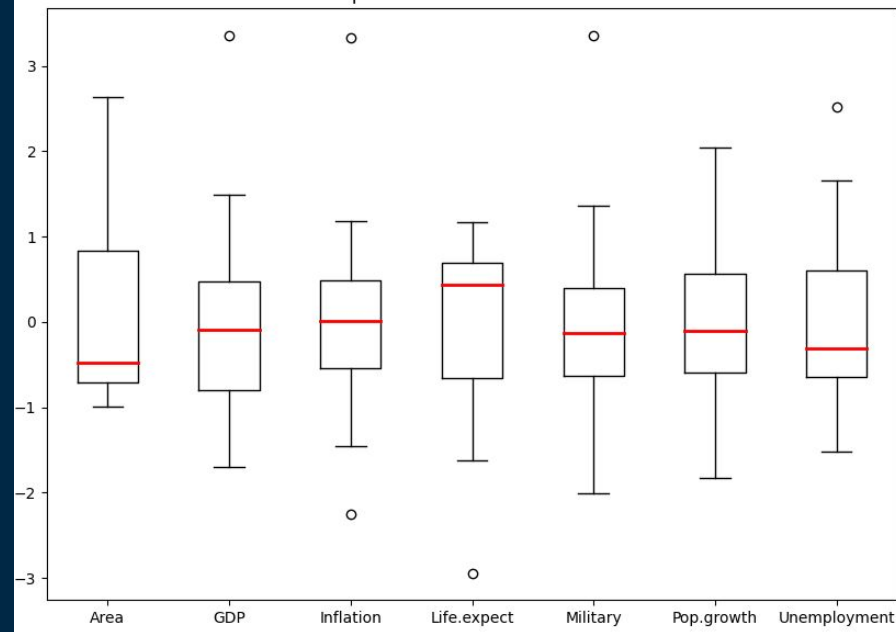
$$\bar{X}_i = \frac{1}{n} \sum_{j=1}^n X_i^j$$

$$s_i = \sqrt{\frac{1}{n-1} \sum_{j=1}^n (X_i^j - \bar{X}_i)^2}$$

Boxplot con datos no estandarizados



Boxplot con datos estandarizados

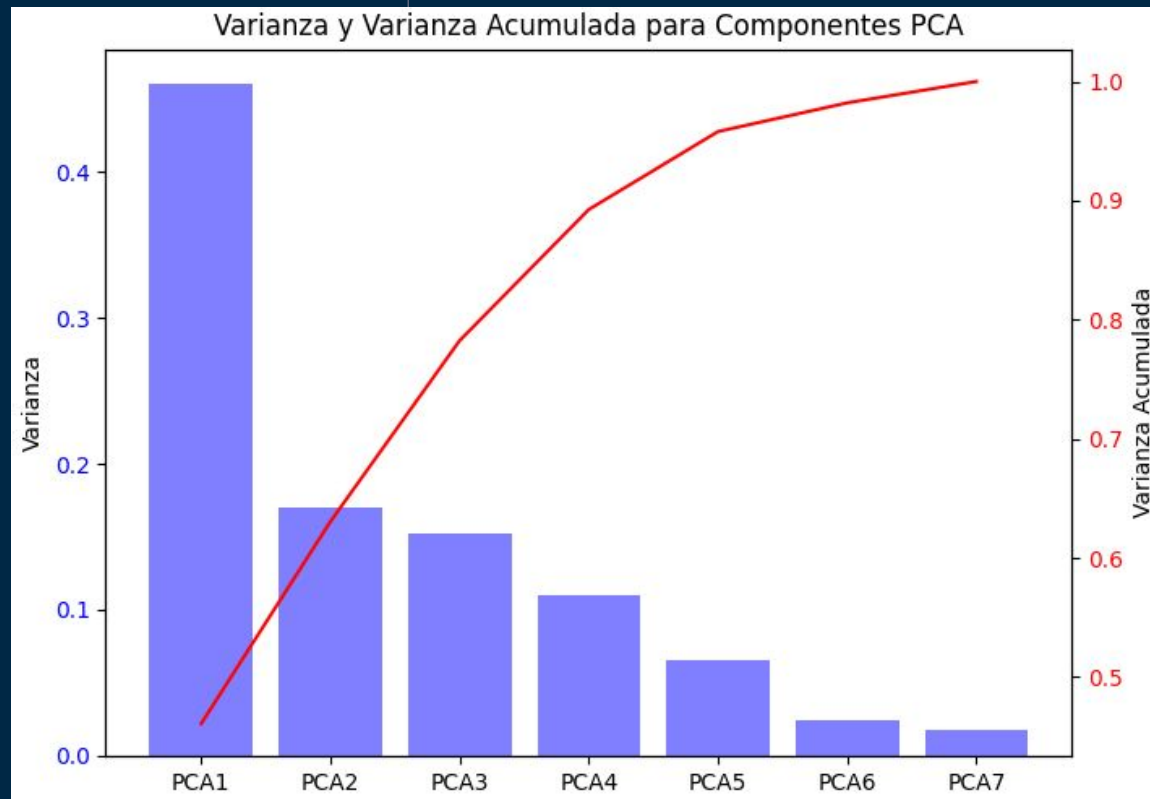


Podemos observar que con los datos no estandarizados parecería ser que el área y GDP son los parámetros que más condicionarán, mientras que el resto parecen tener poca o nula incidencia en comparación.

Sin embargo, al estandarizar se nota claramente la variabilidad de dichas características y cómo se las pueden comparar entre sí.



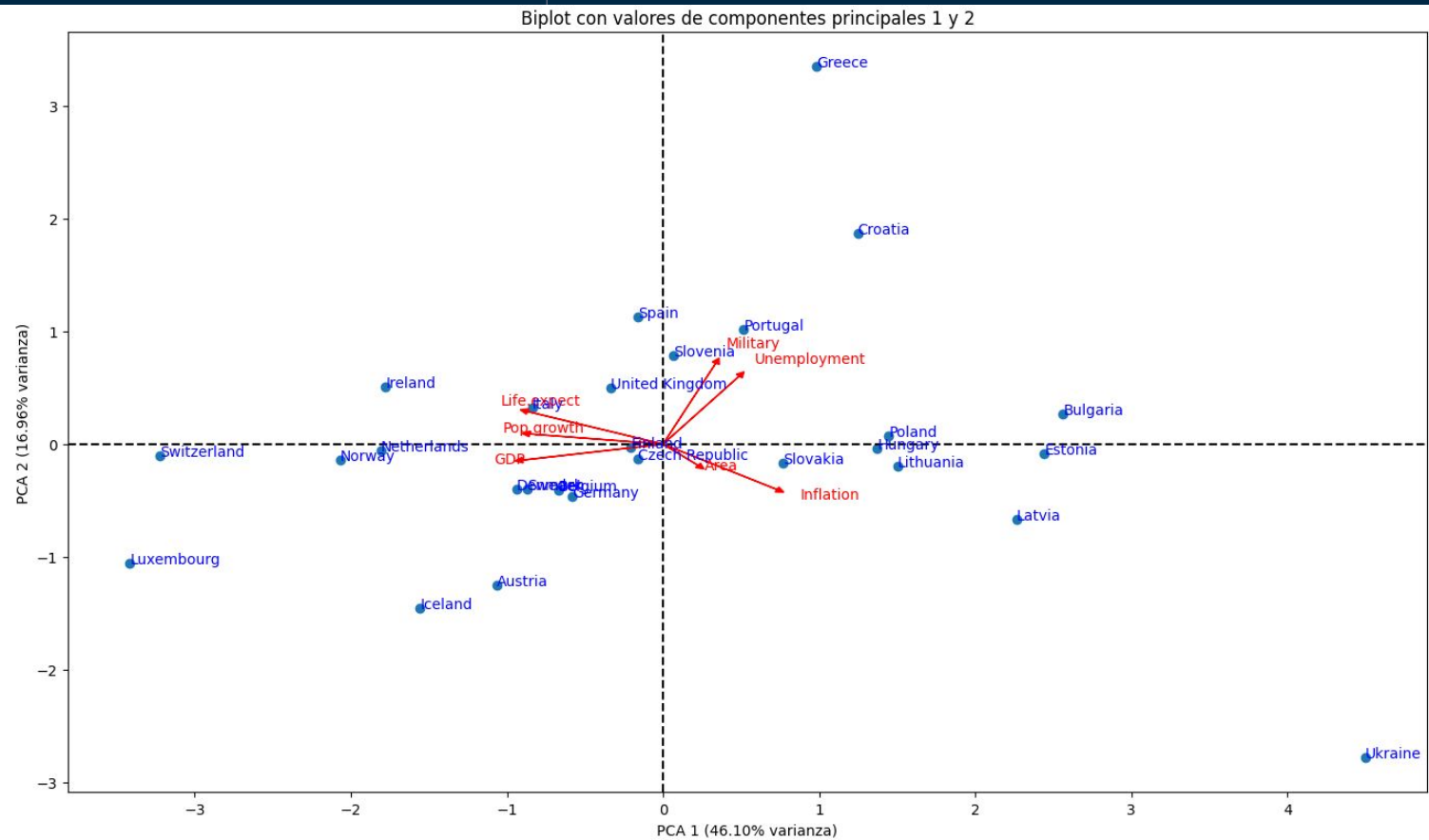
# Análisis de PCA con Sklearn



Varianza de cada PC: [0.4610 0.1696 0.1519 0.1101 0.0654 0.0241 0.018]

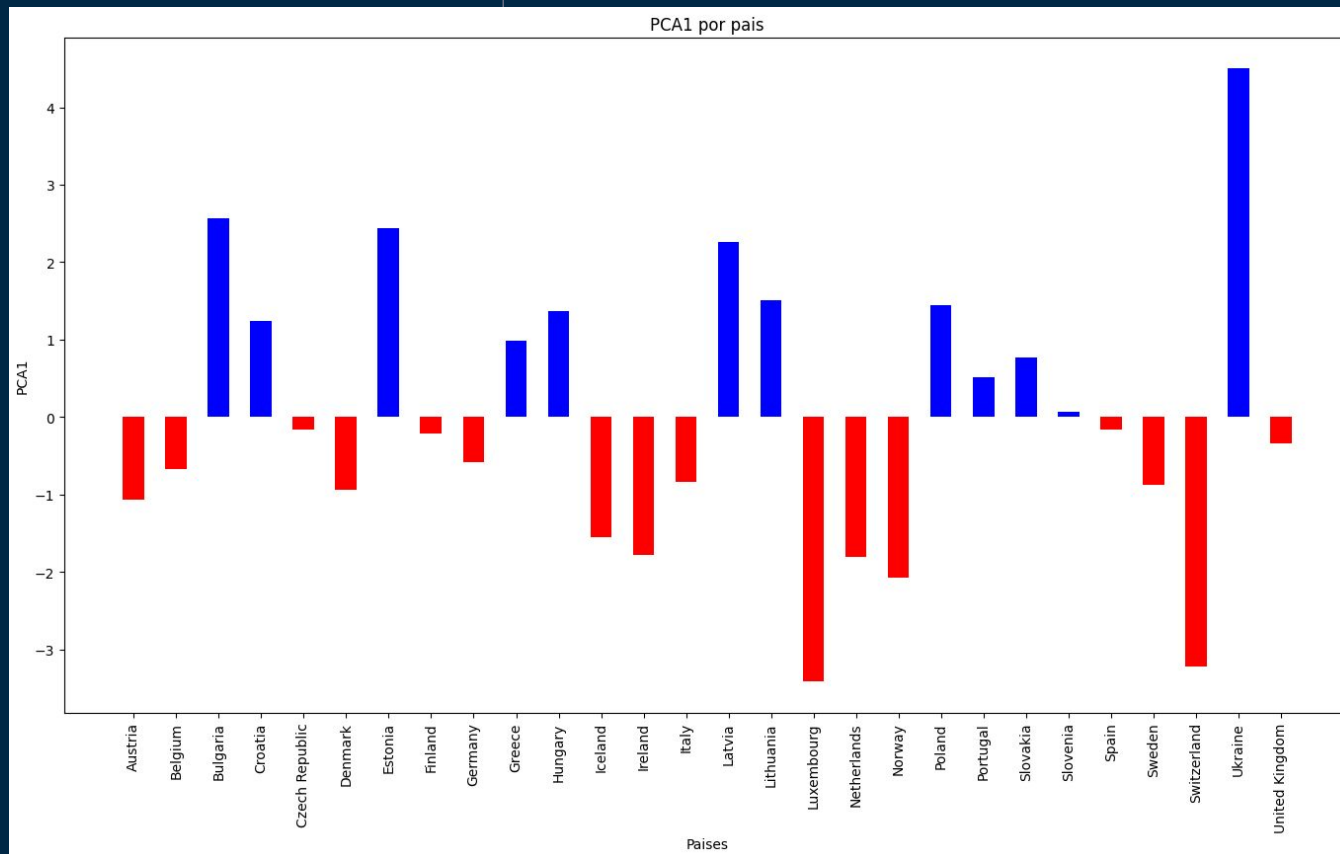
Varianza acumulada: [0.4610 0.6306 0.7825 0.8925 0.9579 0.9821 1]

Se alcanza una variabilidad mayor al 95% a partir de la PC5



PCA1 separa las variables en dos grupos:

- PBI, crecimiento poblacional y expectativa de vida.
- Área, inflación, desempleo y nivel militar.



- Si  $PCA1 > 0$ , el país se ve más afectado por área, inflación, desempleo y nivel militar.
- Si  $PCA1 < 0$ , el país se ve más afectado por PBI, crecimiento poblacional y expectativa de vida.

**Vector de cargas:** [ 0.1249 -0.5005 0.4065 -0.4829 0.1881 -0.4757 0.2717]

The background is a dark navy blue. It is decorated with various geometric elements: thin white vertical lines of varying lengths, and small squares in teal, light pink, and orange. Some squares are solid, while others are just outlines. These elements are scattered across the frame, creating a modern, minimalist aesthetic.

# ¡GRACIAS!