



FUNDAMENTOS DE R MUNDO DE LA ANALÍTICA DE DATOS

Kenny Cárdenas Parra

ESTRUCTURAS DE DATOS

- Vectores
- Matrices
- Dataframes
- Listas
- Arrays



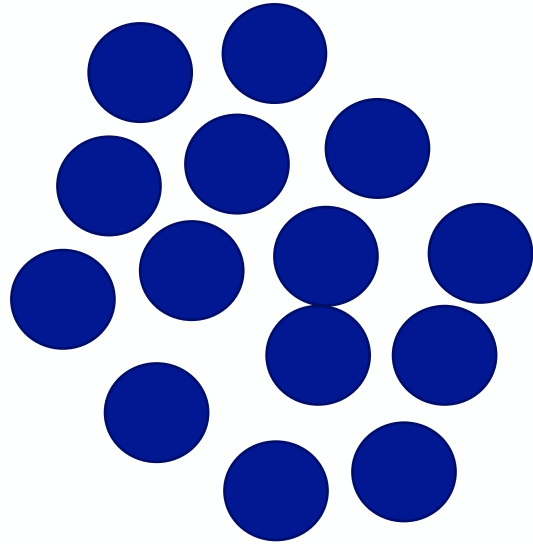
PAQUETES

R tiene la propiedad de ser expandible a través de paquetes. Los paquetes son un conjunto de funciones diseñadas para atender una necesidad específica.

Los paquetes se encuentran alojados en **CRAN**, pasando por un análisis riguroso antes de estar disponibles para su uso generalizado.

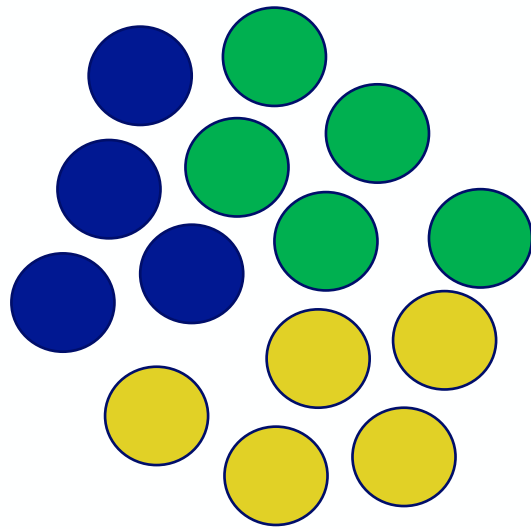


ANÁLISIS DE DATOS



¿Todos cumplen con las mismas características?

ANÁLISIS DE DATOS



Se puede evidenciar que hay un subgrupos, imperceptibles a simple vista

ALTA DIMENSIONALIDAD

Vari 1	Var 2	Var 3	Var 4	Var 5
X	X	X	X	X
X	X	X	X	X
X	X	X	X	X
X	X	X	X	X
X	X	X	X	X
X	X	X	X	X

¿Cómo analizar estos datos?

ALTA DIMENSIONALIDAD

Vari 1	Var 2	Var 3	Var 4	Var 5
X	X	X	X	X
X	X	X	X	X
X	X	X	X	X
X	X	X	X	X
X	X	X	X	X
X	X	X	X	X



Vari 1	Var 2
X	X
X	X
X	X
X	X
X	X
X	X

PROMEDIO

Es un valor representativo de un conjunto de datos

$$\bar{X} = \frac{\sum_{i=1}^n x_i}{n}$$

- n = Número de observaciones
- x_i = Valor de la observación i



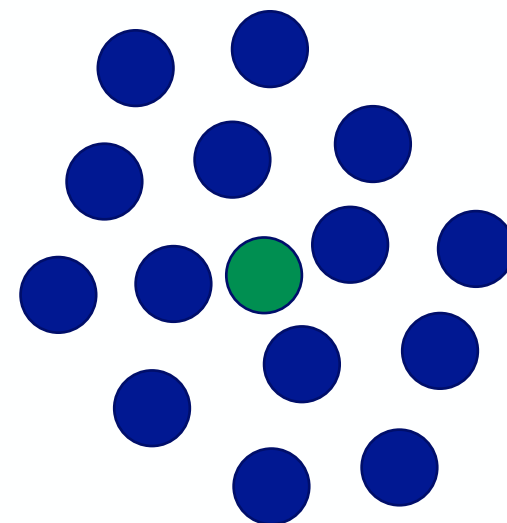
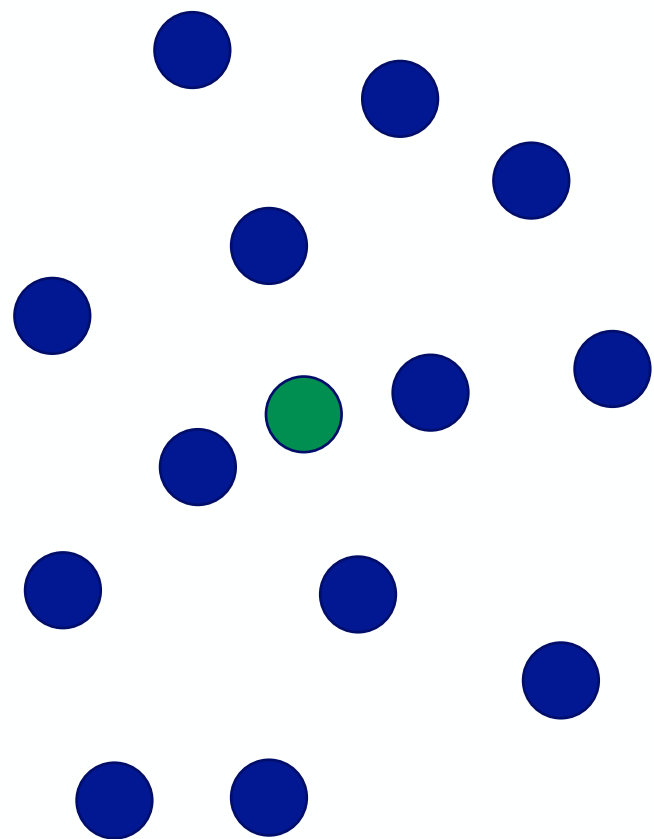
VARIANZA

Determina la variación de los datos respecto a su media

$$\sigma^2 = \frac{\sum_{i=1}^n (\bar{X} - x_i)^2}{n}$$

- n = Número de observaciones
- x_i = Valor de la observación i
- \bar{X} = Promedio de las observaciones

VARIANZA



MATRIZ DE COVARIANZA SIGMA

La covarianza es el valor que refleja en qué cuantía dos variables aleatorias varían de forma conjunta respecto a sus medias.

$$\Sigma = \begin{pmatrix} var(x) & cov(x, y) \\ cov(x, y) & var(y) \end{pmatrix}$$



PASO A PASO PARA CALCULAR

1. Calcular la matriz Σ .
2. Encontrar los valores y vectores propios de la matriz Σ .
3. Definir cuántas componentes principales.



MÉTODOS PARA RETENER PC

- Variabilidad explicada de las components
- Criterio de Kaiser (Para variables estandarizadas)
- Criterio de Sedimentación de Catell

VARIABILIDAD EXPLICADA DE LAS COMPONENTES

Se buscan las componentes que expliquen el mayor porcentaje de variación explicada.

$$\sum_i^V \frac{\lambda_i}{VT} \geq \textit{Porcentaje esperado}$$

CRITERIO DE KAISER (PARA VARIABLES ESTANDARIZADAS)

- En el caso de que las variables no estén estandarizadas, se retienen aquellas componentes que su valor lambda sea mayor al promedio de los valores lambda.

$$\sum_i^V \frac{\lambda_i}{V_T} \geq \textit{Porcentaje esperado}$$

- En el caso de que las variables estén estandarizadas, se retienen aquellas que sean mayor que 1.

$$\lambda_i \geq 1$$

CRITERIO DE SEDIMENTACIÓN DE CATELL

