

Técnicas de análisis cuantitativas y cualitativas

Sesión 4 — Análisis de conglomerados

Eduardo Corbelle Rico

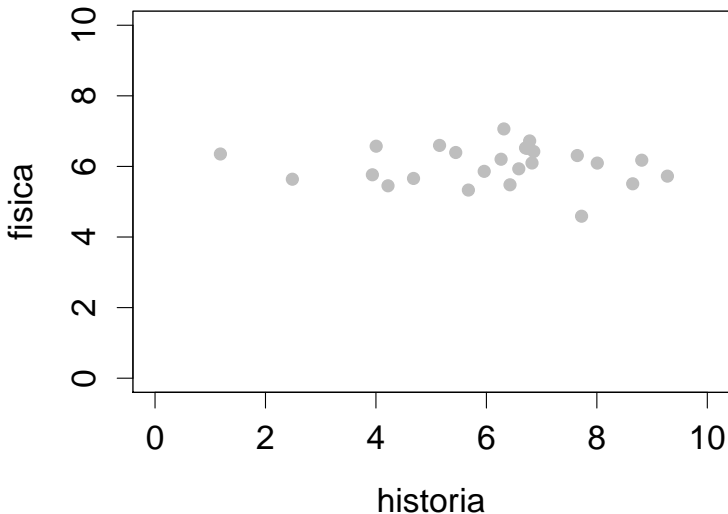
Máster Universitario en Xestión Sustentable da Terra e o Territorio
Universidade de Santiago de Compostela

Curso 2015–2016

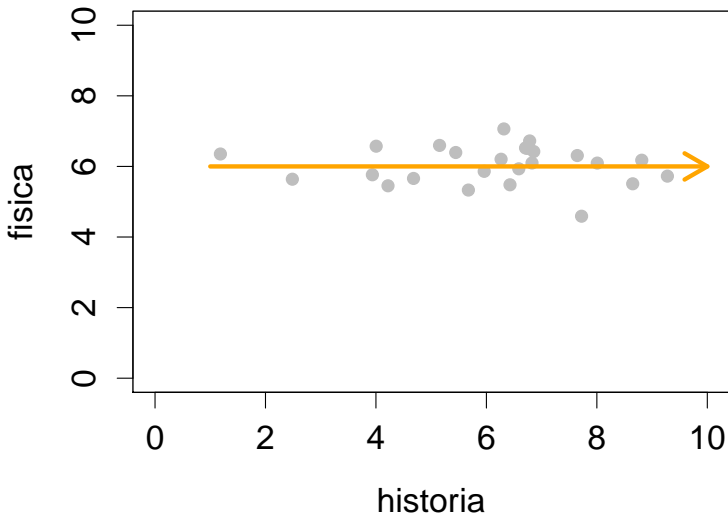
Contenidos

- 1 Análisis de componentes principales
- 2 Análisis de conglomerados
- 3 Árboles de regresión y clasificación
- 4 Práctica 7

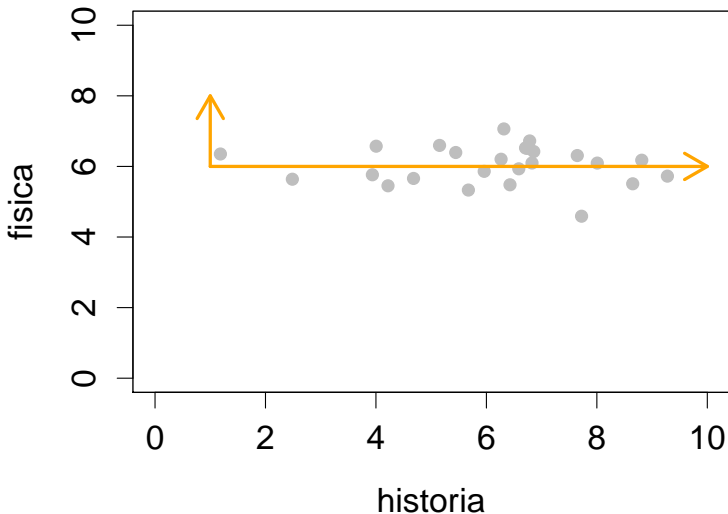
Resumir sin perder (demasiada) información



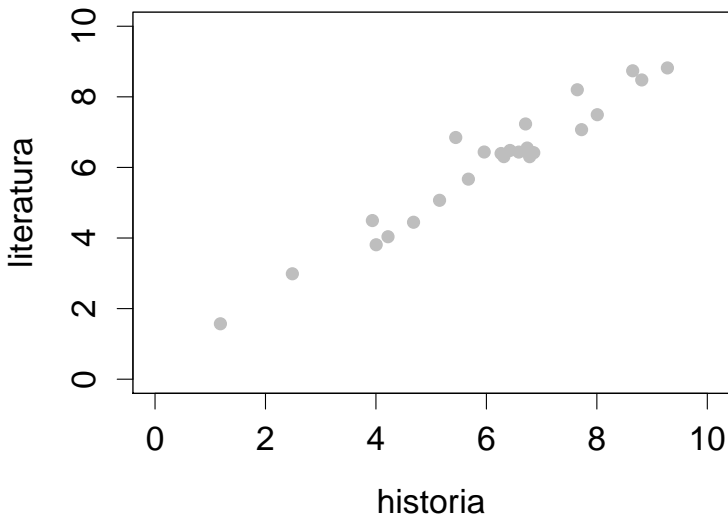
Resumir sin perder (demasiada) información



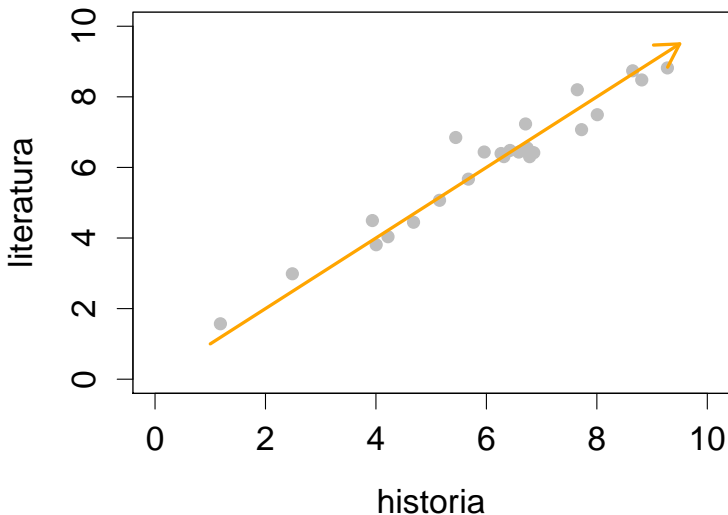
Resumir sin perder (demasiada) información



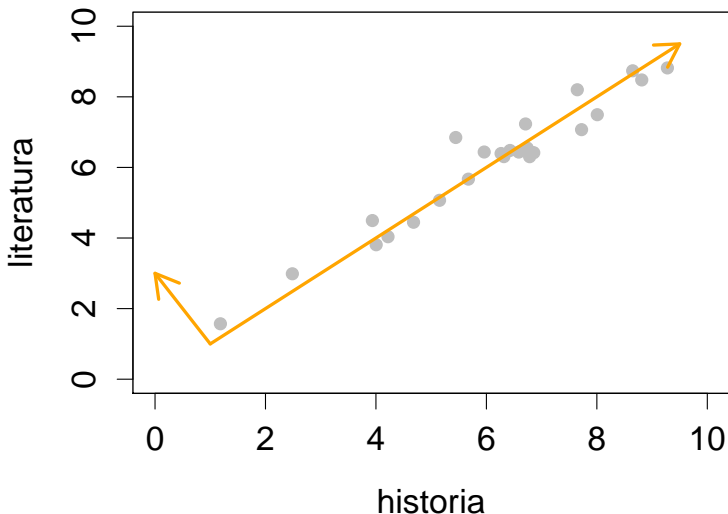
Resumir sin perder (demasiada) información



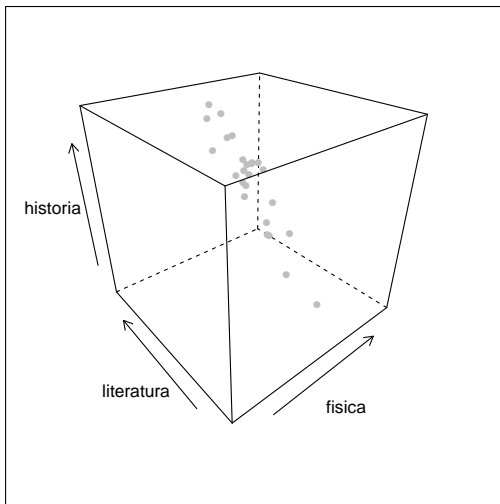
Resumir sin perder (demasiada) información



Resumir sin perder (demasiada) información



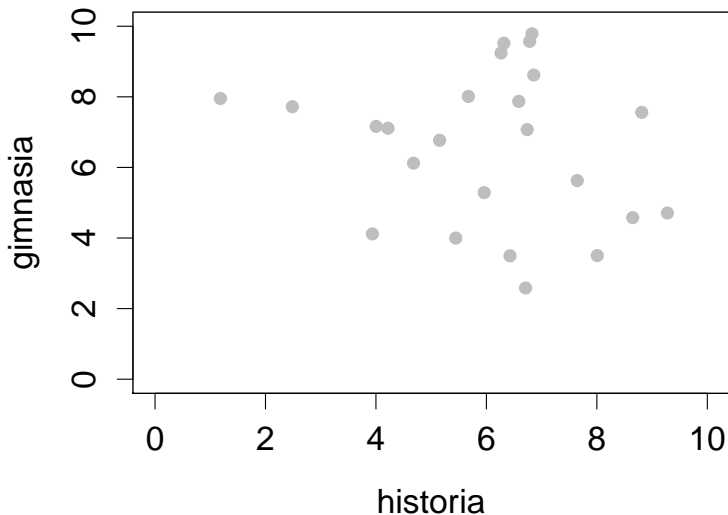
Resumir sin perder (demasiada) información



Resumen

- Transformación del espacio de coordenadas
- El número de dimensiones no varía...
- ...pero podemos quedar con las primeras

En ausencia de correlación



1 Análisis de componentes principales

2 Análisis de conglomerados

3 Árboles de regresión y clasificación

4 Práctica 7

Presentación

Objetivos

- Formar grupos (homogéneos) de observaciones
- Carácter exploratorio — generación de hipótesis

Presentación

Objetivos

- Formar grupos (homogéneos) de observaciones
- Carácter exploratorio — generación de hipótesis

Principales variantes

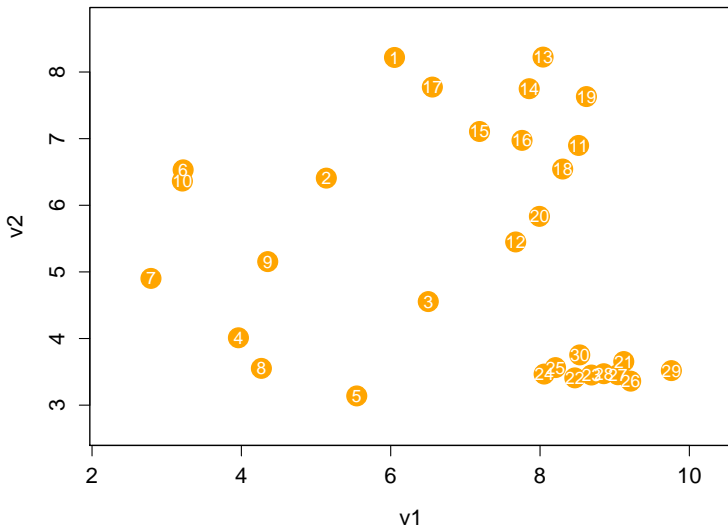
- Agrupamiento jerárquico
- Métodos iterativos

Métodos de agrupamiento jerárquico

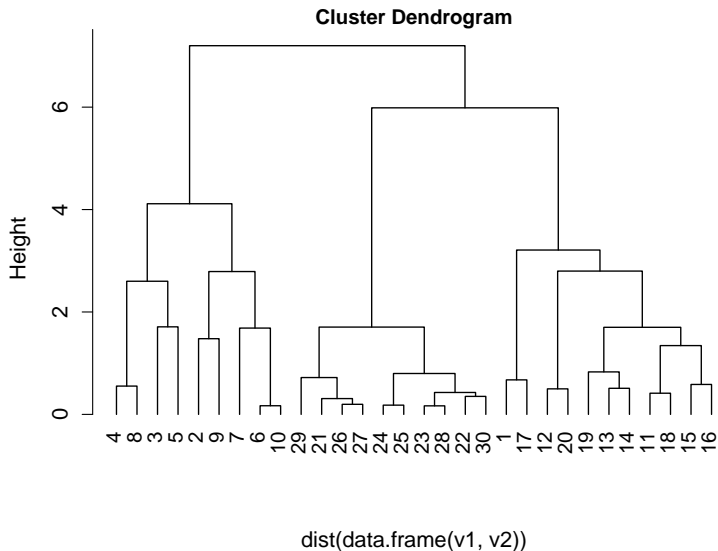
Fases principales

- 1 Cálculo de distancias (matriz de distancias)
- 2 Agrupamiento por distancia
- 3 Decisión sobre el número de grupos

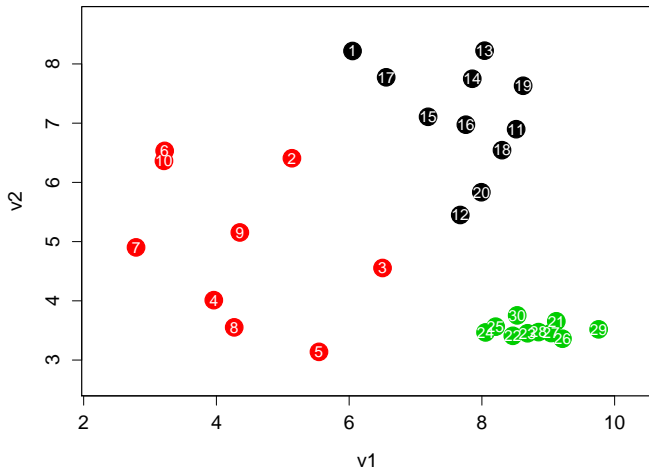
Ejemplo (2D)



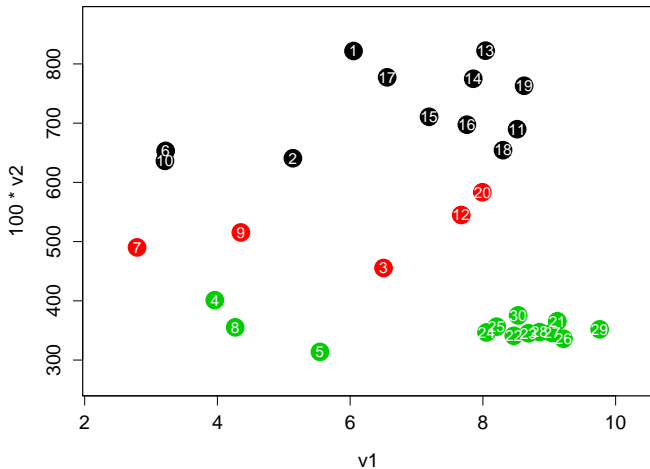
Dendrograma



3 grupos



Influencia de la escala



Estandarización de variables

Posibilidades

- 1 V. numéricas: $\frac{x_i - \bar{x}}{S_x}$
- 2 V. numéricas con asimetría: corregir asimetría y aplicar (1)
- 3 V. categóricas: asignar valores entre 0–1

Agrupamiento jerárquico: Resumen

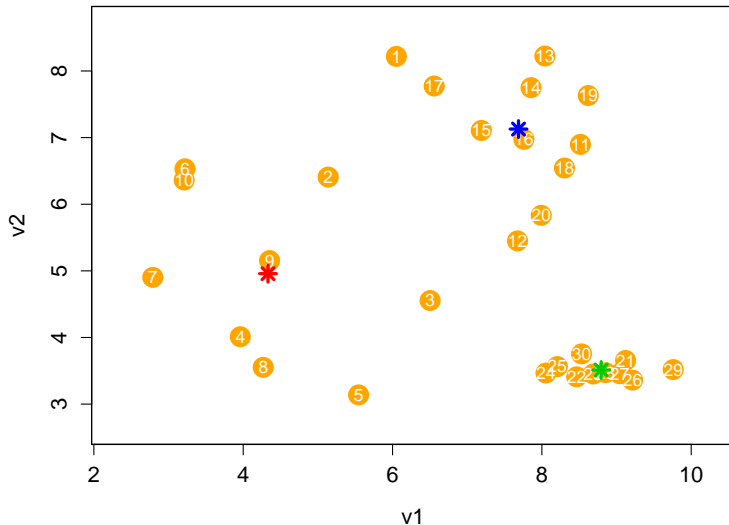
- 0 Selección de variables y observaciones de control
- 1 Estandarización
- 2 Cálculo de distancias entre observaciones
(Euclídea / Máxima / Manhattan...)
- 3 Agrupamiento
(Completo / Promedio / Centroide / Ward...)
- 4 Decisión del número de grupos
- 5 Revisión de resultados

Métodos de agrupamiento no jerárquicos

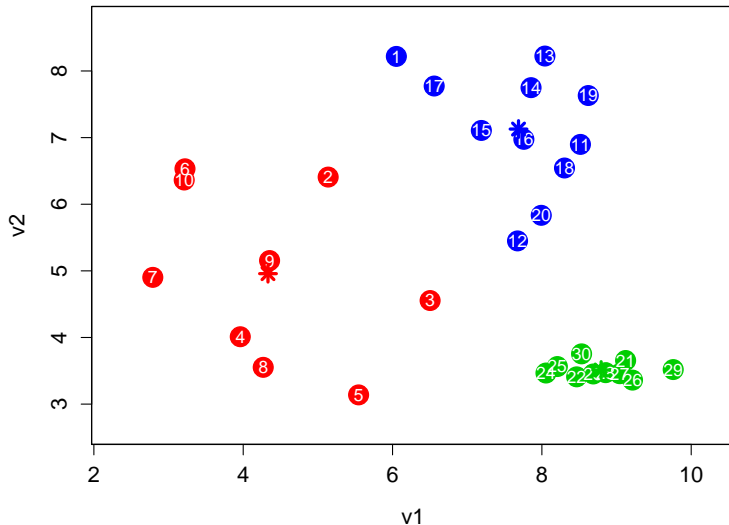
K-medias y variantes (ISODATA...)

- 1 División arbitraria de las observaciones en k grupos
- 2 Cálculo de la media de cada variable para cada grupo
- 3 Asignación de observaciones a la media más cercana
- 4 Volver a (2)

Clasificación con *k-medias*, 3 clases



Clasificación con k -medias, 3 clases



1 Análisis de componentes principales

2 Análisis de conglomerados

3 Árboles de regresión y clasificación

4 Práctica 7

Árboles de regresión y clasificación

Classification and Regression Trees (CART)

Permiten establecer relaciones entre...

- Una variable dependiente (numérica / categórica)
- Un conjunto de variables independientes

...mediante división sucesiva (*recursive partitioning*)

Árboles de regresión y clasificación

Classification and Regression Trees (CART)

Permiten establecer relaciones entre...

- Una variable dependiente (numérica / categórica)
- Un conjunto de variables independientes

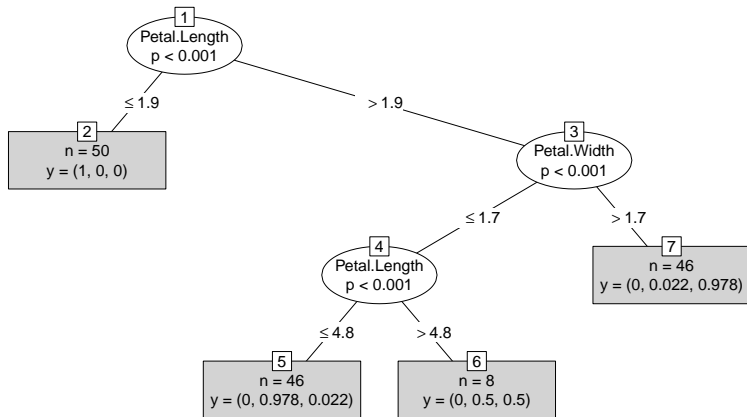
...mediante división sucesiva (*recursive partitioning*)

En función de la variable dependiente

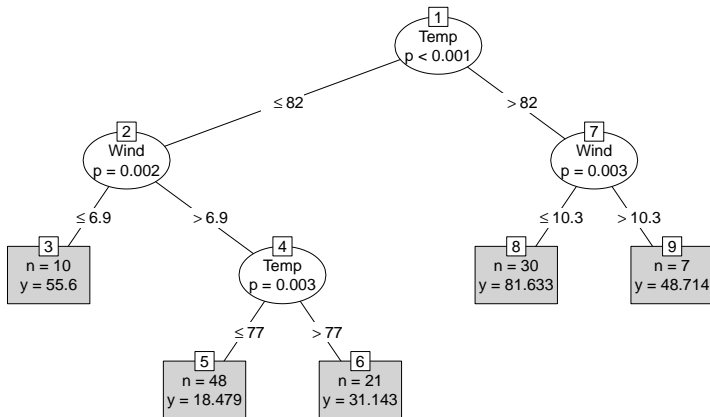
Categórica Árboles de clasificación

Numérica Árboles de regresión

Ejemplo: árbol de clasificación



Ejemplo: árbol de regresión



Práctica 7

- Análisis de componentes principales
- Análisis de conglomerados