

## Analisi dei gruppi

### dist()

La funzione `dist(x, method = ,...)` costruisce la matrice di dissomiglianze calcolando la distanza specificata tra coppie di righe di una matrice dati.

<code>x</code>	matrice o data frame contenente le osservazioni;
<code>method</code>	permette di specificare la distanza da utilizzare: "euclidean": distanza euclidea; "maximum": massimo scarto tra coppie di elementi dei due vettori; "manhattan": distanza assoluta; "binary": indice di dissomiglianza semplice.

### hclust()

La funzione `hclust(d, method = ,...)` effettua un'analisi dei gruppi gerarchica aggregativa partendo da una matrice di dissomiglianze

<code>d</code>	matrice di dissomiglianze prodotta tramite la funzione <code>dist()</code> ;
<code>method</code>	criterio di aggregazione utilizzato: "ward": metodo di Ward; "single": legame singolo; "complete": legame completo; "average": legame medio; "centroid": metodo del centroide.

Gli oggetti prodotti dalla funzione `hclust()` descrivono il diagramma ad albero prodotto dal metodo gerarchico aggregativo specificato. Tali oggetti sono liste i cui elementi principali sono:

<code>merge</code>	matrice $(n-1) \times 2$ le cui righe descrivono le aggregazioni avvenute a ciascun passo dell'intero procedimento. Gli elementi negativi indicano singole unità, mentre quelli positivi indicano gruppi già formati;
<code>height</code>	vettore di $n-1$ numeri reali non-decrescenti che indicano i livelli di dissomiglianza ai quali avvengono le aggregazioni;
<code>labels</code>	vettore di etichette delle osservazioni;
<code>method</code>	il criterio di aggregazione utilizzato.

### plot()

La rappresentazione grafica del diagramma ad albero o dendrogramma si ottiene mediante la funzione `plot()` applicata ad oggetti creati con la funzione `hclust()`. In tal caso gli argomenti principali della funzione `plot()` sono:

<code>clus</code>	oggetto prodotto dalla funzione <code>hclust()</code> ;
<code>hang</code>	parametro per la rappresentazione delle etichette delle osservazioni nel grafico;
<code>labels</code>	vettore delle etichette delle osservazioni.

### rect.hclust()

La funzione `rect.hclust()`, applicata dopo la funzione `plot()` ad oggetti creati con la funzione `hclust()`, permette di tracciare dei rettangoli attorno ai rami del dendrogramma, mettendone in evidenza i gruppi corrispondenti. I suoi argomenti principali sono:

<code>clust</code>	oggetto prodotto dal comando <code>hclust</code> ;
<code>k, h</code>	scalari che indicano il numero esatto di gruppi da creare ( $k$ ) ovvero il livello di dissomiglianza ( $h$ ) a cui tagliare il dendrogramma;
<code>which</code>	vettori che indicano attorno a quali gruppi vanno tracciati i rettangoli (per default vengono selezionati tutti i gruppi individuati);
<code>border</code>	permette di specificare il colore dei rettangoli;

La funzione `print()` applicata ad un oggetto prodotto tramite `rect.hclust()` permette di ottenere gli elementi appartenenti ai gruppi individuati.

### identify()

La funzione `identify()`, applicata dopo la funzione `plot()` ad oggetti creati con la funzione `hclust()`, permette di tagliare il dendrogramma individuando i gruppi corrispondenti ai suoi rami semplicemente cliccando con il mouse sul grafico dello stesso.

Il comando produce un oggetto contenente una lista di vettori aventi per elementi gli indicatori (ed eventualmente le etichette) delle osservazioni appartenenti ai cluster individuati.

La funzione `print()` applicata ad un oggetto prodotto tramite `identify()` permette di ottenere gli elementi appartenenti ai gruppi individuati.

### kmeans()

La funzione `kmeans()` effettua l'analisi dei gruppi non gerarchica con il metodo k-means. I suoi argomenti principali sono:

<code>x</code>	matrice dati a valori numerici;
<code>centers</code>	scalare che indica il numero di gruppi o matrice avente per righe i centroidi della partizione iniziale;
<code>iter.max</code>	massimo numero di iterazioni permesse per l'ottimizzazione.

Il comando produce una lista contenente i seguenti elementi:

<code>cluster</code>	vettore di numeri interi che indicano l'appartenenza di ciascuna osservazione ad uno dei gruppi individuati;
<code>centers</code>	matrice dei centroidi dei gruppi ottenuti;
<code>withinss</code>	devianza di ciascun cluster;
<code>size</code>	numero di osservazioni in ciascun gruppo.