

METODI STATISTICI PER LA BIOINGEGNERIA

Laboratorio 11

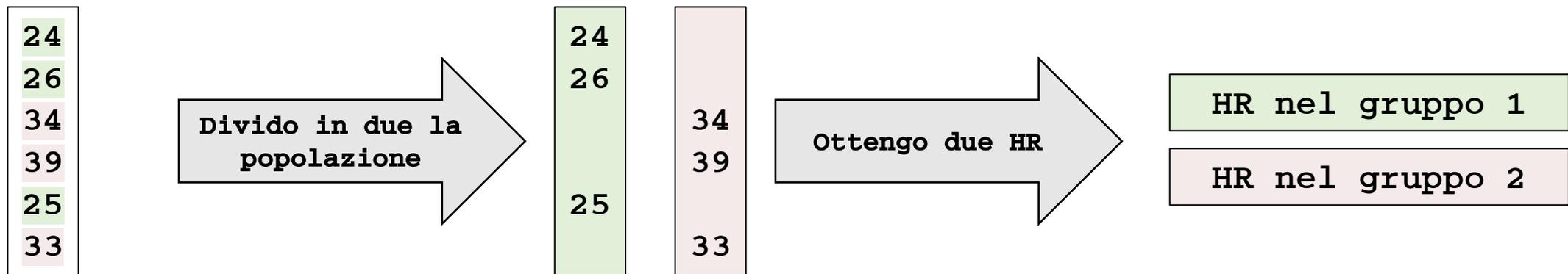
A.A. 2024-2025

Enrico Longato

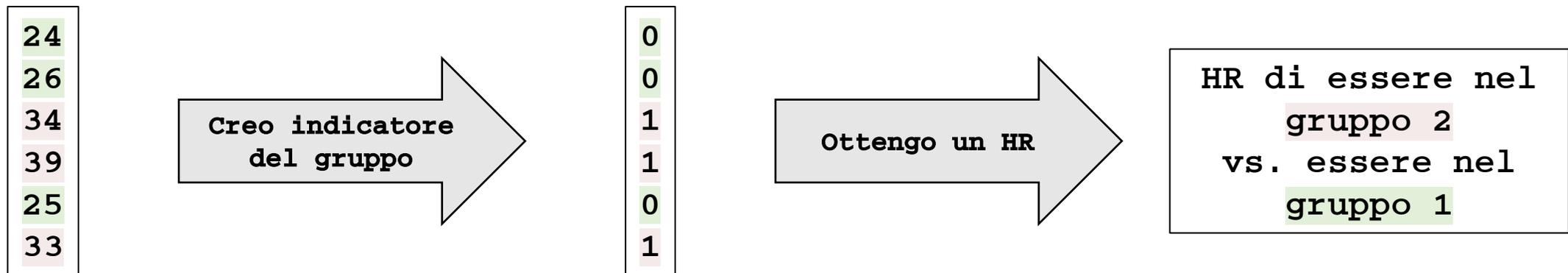


Dal lab 10: Cox per gruppi vs. tra i gruppi

Per stimare l'effetto sull'hazard dell'aumento di un'unità di BMI entro ciascun gruppo si fa così:



Per stimare l'effetto sull'hazard del cambiare gruppo (cioè per confrontare i due gruppi) si fa così:





Funzione **kmeans** (consultare l'help per maggiori informazioni)

Traccia di utilizzo (in rosso il codice; in nero il "testo libero")

```
[cluster, centroid] = kmeans(matrice di dati [N dati × N variabili], numero di cluster, ...  
                             'Distance', 'sqeuclidean', ...  
                             'Replicates', numero di inizializzazioni diverse da provare, ...  
                             'MaxIter', numero massimo di iterazioni di repliche);
```

Vettore dei cluster
assegnati a ciascun dato

Matrice dei centroidi
(uno per riga)

Funzione **silhouette** (consultare l'help per maggiori informazioni)

Traccia di utilizzo (in rosso il codice; in nero il "testo libero")

```
sil = silhouette(matrice di dati [N dati × N variabili], ...  
                vettore dei cluster corrispondenti (dimensione [N dati]))
```

Silhouette per ciascun dato.

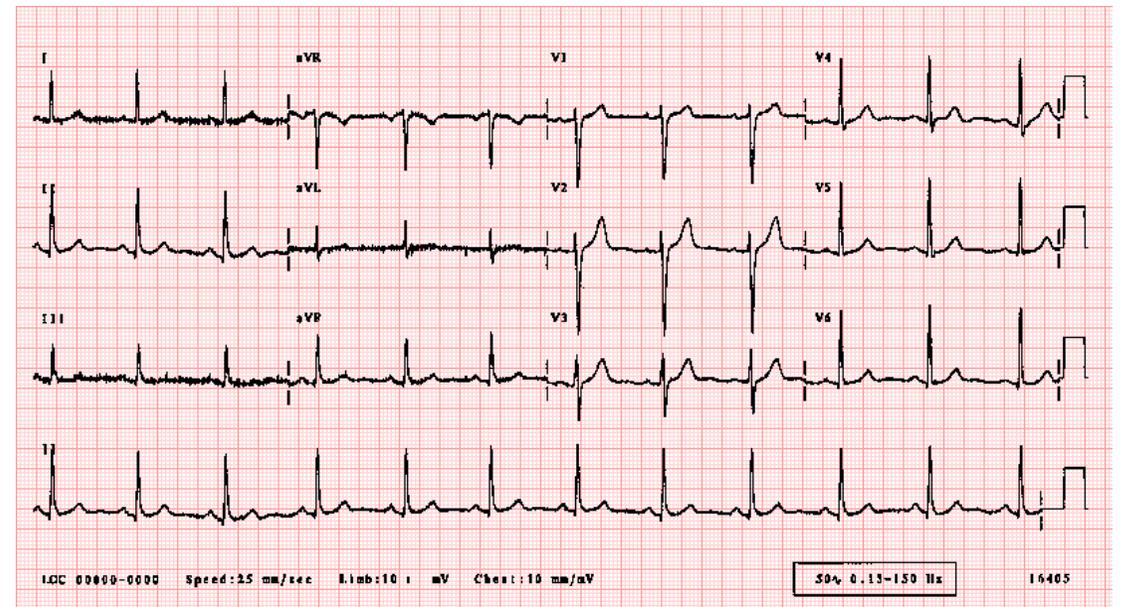
A noi, come indicatore globale della
bontà del clustering, serve la media!



DATI PER L'ESERCITAZIONE

Il file di dati in ingresso `data_lab10.mat` contiene

1. **ecg**: matrice di 100 acquisizioni di segnali elettrocardiografici, ciascuno da 300 campioni (i segnali sono stati resi a media nulla e riscaldati; quindi consideriamoli senza unità di misura).
2. **time**: vettore dei tempi a cui sono stati acquisiti i 300 campioni di ciascuna traccia ECG (unità di misura: secondi).





ESERCIZIO 1 – PARTE 1: CARICAMENTO E VISUALIZZAZIONE DATI (svolto)

- Caricare i dati.
- Rappresentare nella stessa figura, in due pannelli sovrapposti
 - Tutti i 100 segnali presenti nella matrice **ecg** (pannello superiore)
 - I soli segnali 1, 14 e 23, sovrapposti (pannello inferiore)

ESERCIZIO 1 – PARTE 2: K-MEANS CON K FISSATO (svolto)

- Fissato il numero di cluster $K = 3$, utilizzare la funzione **kmeans** per effettuare il clustering
 - Parametri di **kmeans** da utilizzare, oltre al numero di cluster fissato
 - 10 repliche (cioè inizializzazioni diverse)
 - 200 iterazioni massime per ciascuna replica
- Rappresentare in una figura, sovrapposti, i 3 centroidi (ovvero i tre segnali "prototipo" dei 3 cluster)
- Calcolare la silhouette di ciascuno dei 100 segnali originali e rappresentarla
 - Suggestimento: si usa la stessa chiamata alla funzione **silhouette**
 - Se non ci sono argomenti di uscita, viene disegnata la figura
 - Se c'è un argomento di uscita, quello sarà il vettore delle silhouette
- Calcolare la silhouette media (**mean**)

ESERCIZIO 1 – PARTE 3: RICERCA DEL NUMERO OTTIMO DI CLUSTER PER K-MEANS (proposto)

- Utilizzare gli stessi dati di prima
- Far variare il numero di cluster K da 2 a 10 (estremi inclusi) e
 1. Effettuare il clustering con **kmeans**
 2. Memorizzare la silhouette media ottenuta per ciascun K in un array
 3. Memorizzare i centroidi ottenuti per ciascun K in un cell array
 4. Trovare il numero di cluster ottimo come quello che massimizza la silhouette media (funzione **max**)
 5. Rappresentare l'andamento della silhouette al variare del numero di cluster
 6. Rappresentare, sovrapposti, nella stessa figura i centroidi corrispondenti al numero di cluster ottimo

