

METODI STATISTICI PER LA BIOINGEGNERIA

A.A. 2024-2025

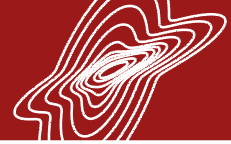
Prof. Alessandra Bertoldo
Ing. Simone Cauzzo



PRIMA PARTE: Analisi di Kaplan-Meier

Dati a disposizione e obiettivo del Laboratorio:

- I dati a disposizione per il laboratorio sono stati raccolti per uno studio atto a determinare se la variabile Body Mass Index (BMI) fosse un fattore indicativo per la sopravvivenza
- L'obiettivo è quindi verificare con l'analisi di sopravvivenza se un gruppo di soggetti diviso in base ad un valore soglia fornisce esiti statisticamente diversi sulla sopravvivenza
- I dati sono contenuti nel file ***data_lab10.mat***, sono presenti:
 - BMI***: BMI collezionato per 100 soggetti (100x1).
 - SurvTime***: vettore contenente il tempo di sopravvivenza dei soggetti (100x1).
 - threshold***: soglia per suddividere in due gruppi il dataset
 - C***: vettore booleano. 1 se il paziente è osservato, zero se è stato censurato.



PRIMA PARTE: Analisi di Kaplan-Meier

1. Caricamento dei dati e creazione delle variabili corrispondenti ai due gruppi:

GRUPPO 1 – **BMI > threshold**

GRUPPO 2 – **BMI <= threshold**

T1 = Tempo di sopravvivenza del gruppo 1

T2 = Tempo di sopravvivenza del gruppo 2

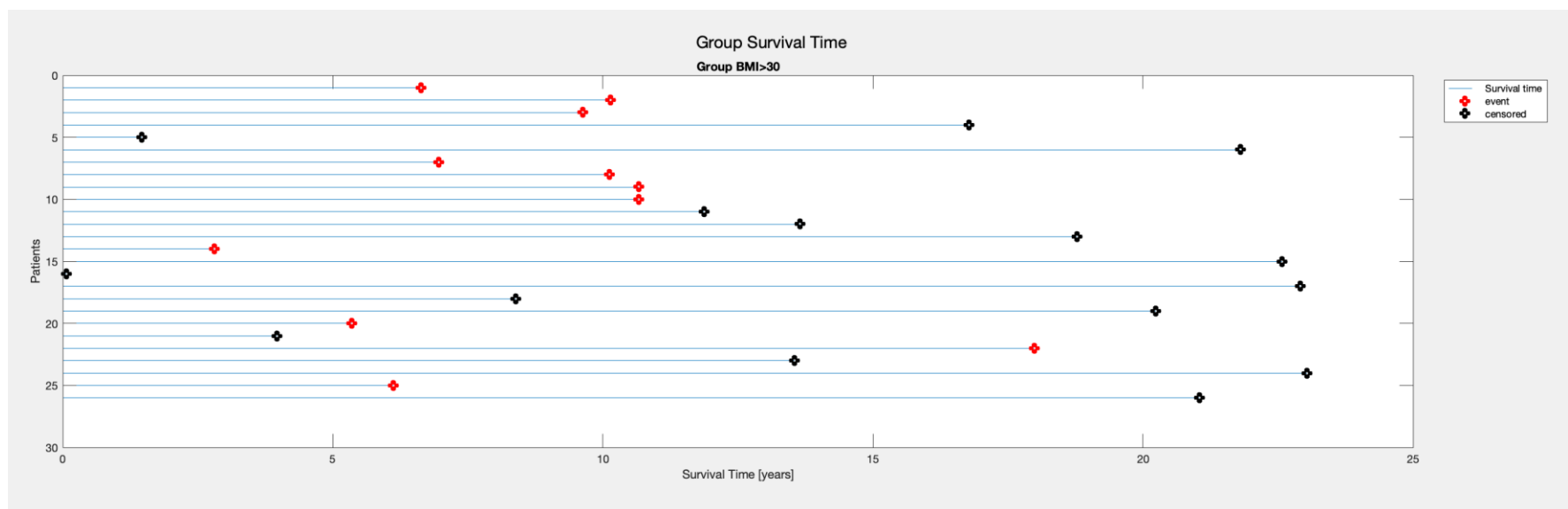
C1 = Pazienti censurati gruppo 1

C2 = Pazienti censurati gruppo 2

PRIMA PARTE: Analisi di Kaplan-Meier

2. Ricreare il plot visto a lezione (un subplot per ogni gruppo ottenuto al punto 1) e identificare:
 - a. I pazienti censurati (plottare l'evento di censura con marker nero ◊)
 - b. I pazienti osservati (plottare l'evento da monitorare con marker rosso ◊)
 - c. Sulle ascisse va il tempo di sopravvivenza, sulle ordinate tutti i pazienti

N.B. Utilizzare `stem` e `view(90,90)` per ruotare il grafico





PRIMA PARTE: Analisi di Kaplan-Meier

3. Inizializzazione

Impostare la griglia dei tempi di Kaplan-meier e inizializzare le variabili

T = gli istanti temporali unici dei pazienti osservati + tempo zero

Vettori delle stime della sopravvivenza

$S1$ = vettore di zeri, lungo come T .

Nella prima posizione inseriamo 1

$S2$ = vettore di zeri, lungo come T .

Nella prima posizione inseriamo 1

Vettori degli eventi osservati

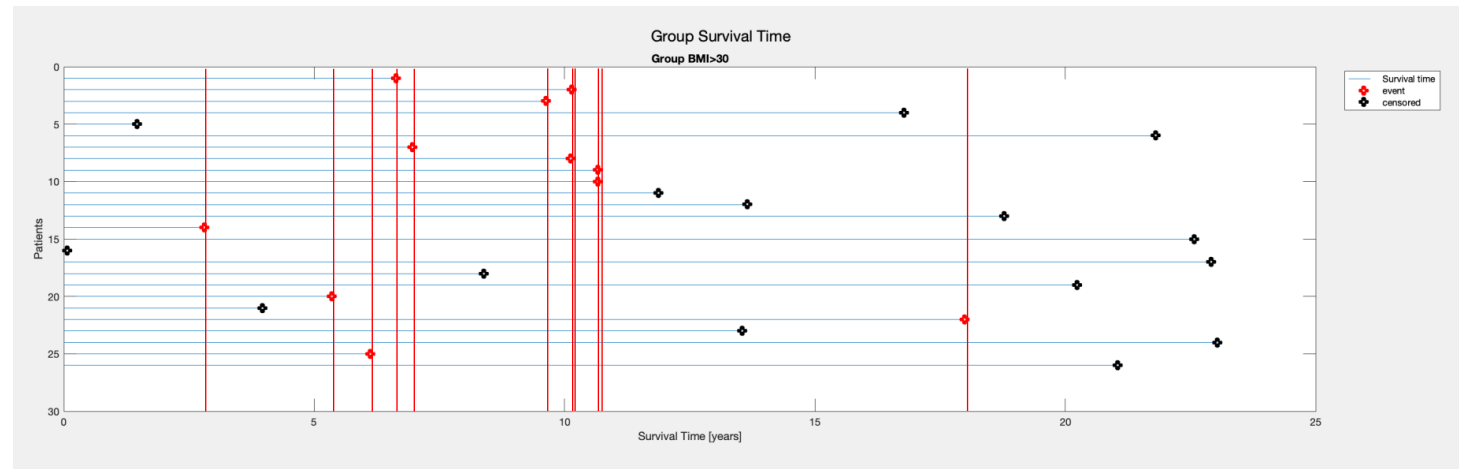
$O1$ = vettore di zeri, lunghezza pari al numero di intervalli

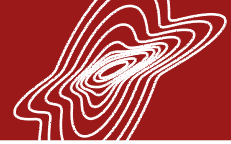
$O2$ = vettore di zeri, lunghezza pari al numero di intervalli

Vettori della mortalità attesa

$E1$ = vettore di zeri, lunghezza pari al numero di intervalli

$E2$ = vettore di zeri, lunghezza pari al numero di intervalli

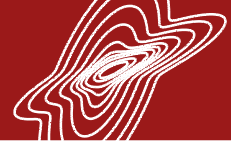




4. Implementazione: Pseudo-codice analisi di Kaplan-Meier

Per ogni **intervallo** (N.B. congiuntamente ai due gruppi) :

- a. calcolare la numerosità dei soggetti a rischio di ogni gruppo
- b. calcolare il numero di eventi osservati in ciascun gruppo
- c. stimare la sopravvivenza di ciascun gruppo
- d. stimare la mortalità attesa di ciascun gruppo



5. Log-rank test

Calcolare il p-value (N.B. coda destra rispetto a Z di una distribuzione χ^2 con 1 grado di libertà, `chi2cdf`)

$$z = \frac{(\text{mortalità attesa} - \text{mortalità osservata})^2}{\text{Somma prodotti delle mortalità attese nei due gruppi}}$$

6. Survival plot

Plottare le due curve di sopravvivenza stimate (`stairs`).

Nel titolo specificare la variabile considerata (BMI) e il p-value ottenuto col log-rank test.

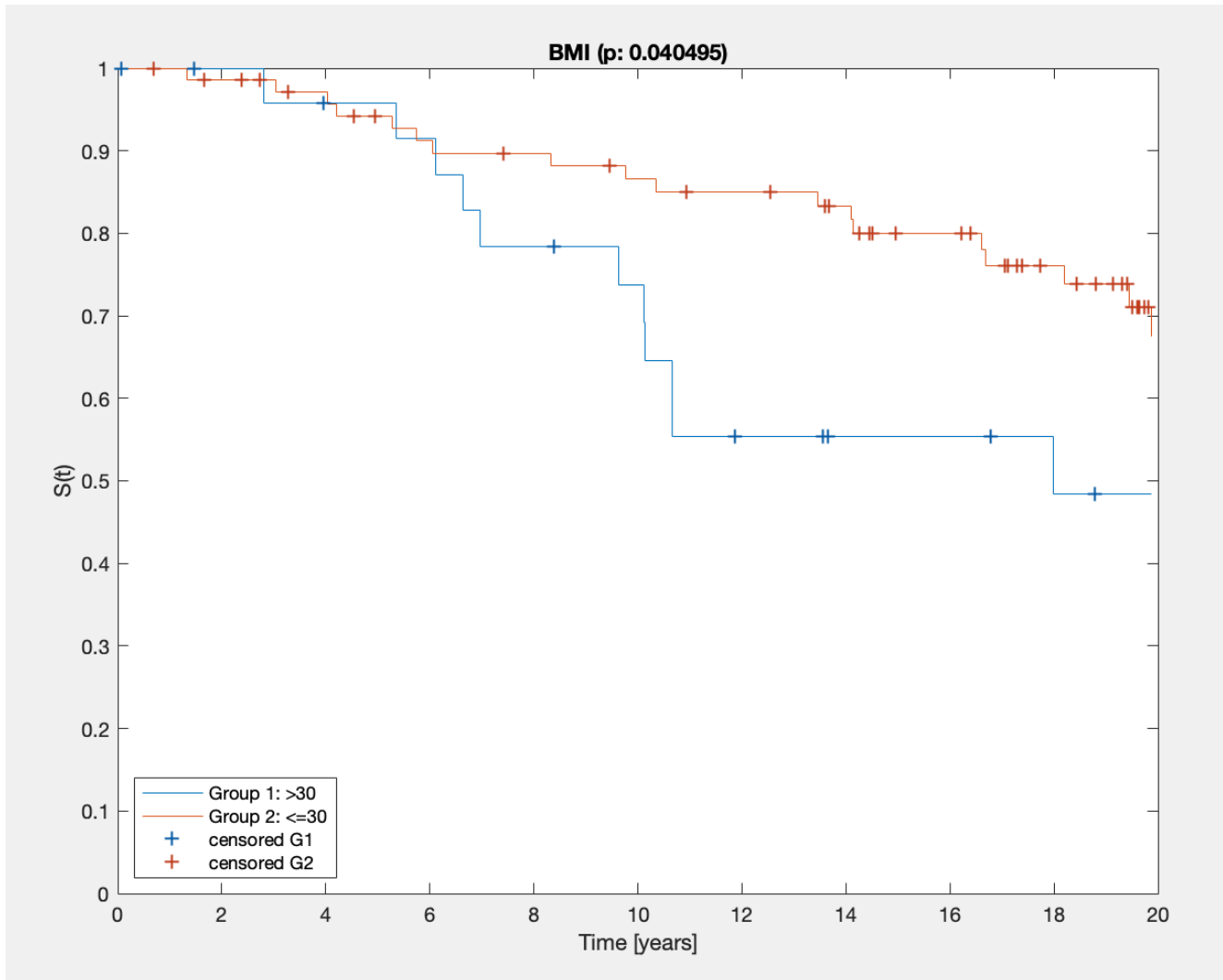
Come labels degli assi utilizzare 'Time [years]' e 'S(t)' rispettivamente per ascisse e ordinate.

N.B. per poter aggiungere i soggetti censurati bisogna interpolare la survival rate al tempo della censura:

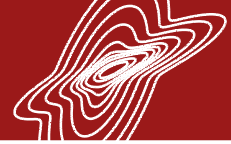
```
hold on
pos_cens1 = interp1(T, S1, T1(C1==0), 'previous');
pos_cens2 = interp1(T, S2, T2(C2==0), 'previous');
scatter(T1(C1==0), pos_cens1, 'k+')
scatter(T2(C2==0), pos_cens2, 'k+')
```

Infine aggiungere la legenda al grafico.

6. Survival Plot



Plot delle curve di sopravvivenza ottenute con l'analisi di Kaplan-Meier e il relativo p-value ottenuto dal confronto Log-rank



SECONDA PARTE: Analisi con modello di Cox

Usare la funzione `coxphfit` in modo univariato per ottenere il coefficiente beta relativo alla variabile binaria ottenuta dalla sogliatura del BMI.

Calcolare l'hazard ratio di questa variabile (si consulti l'help per capire la relazione tra gli output della funzione e l'hazard ratio).

È maggiore o minore di uno?

Come si può interpretare?