



# **METODI STATISTICI PER LA BIOINGEGNERIA**

## **Laboratorio 10**

A.A. 2025-2026

**Enrico Longato**



## Dal lab 9: Generatore di numeri casuali

Vi sarete accorte ed accorti che, facendo ripartire alcuni script, come quelli del laboratorio scorso, si ottengono risultati leggermente diversi ogni volta.

**Questo è vero per tutti gli script che si basano direttamente o attraverso qualche function su un generatore di numeri (pseudo)casuali.**

**Per far sì che il codice esegua sempre identico, basta utilizzare le istruzioni**

```
rng default  
rng(SEED)
```

Dove **SEED** è un numero qualunque; allo stesso seed corrisponde la stessa esecuzione (ovvero la stessa sequenza di numeri pseudo-casuali)



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

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

```
[km_values, km_t] = ecdf(vettore dei tempi di sopravvivenza, ...  
                        'Censoring', vettore "strano" in cui c'è 1 se il dato è censored, ...  
                        'Function', 'survivor');
```

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

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

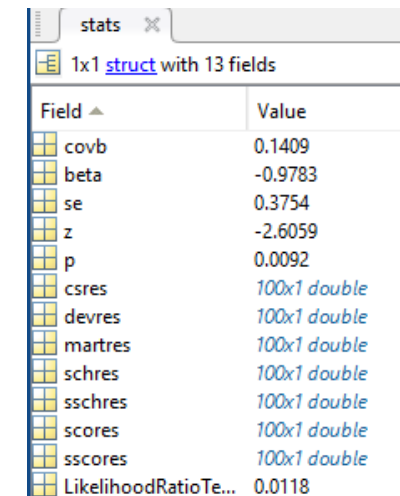
```
[b, logl, H, stats] = coxphfit(X, vettore dei tempi di sopravvivenza, ...  
                              'Censoring', vettore "strano" in cui c'è 1 se il dato è censored);
```

Beta stimato (il  
logaritmo dell'HR!)

Log-likelihood  
(non ci interessa)

Baseline cumulative hazard  
(non ci interessa)

Il p-value del test sul  
coefficiente è in  
**stats.p**



Field	Value
covb	0.1409
beta	-0.9783
se	0.3754
z	-2.6059
p	0.0092
csres	100x1 double
devres	100x1 double
martres	100x1 double
schres	100x1 double
sschres	100x1 double
scores	100x1 double
sscores	100x1 double
LikelihoodRatioTe...	0.0118



### DATI PER L'ESERCITAZIONE

Il file di dati in ingresso **data\_lab10.mat** contiene

1. **BMI**: vettore di valori di BMI, uno per soggetto, in  $\text{kg}/\text{m}^2$ .
2. **threshold**: scalare pari a 30 da usare per dividere il database in due gruppi nelle due parti proposte dell'esercizio
3. **SurvTime**: vettore di tempi di sopravvivenza o di censoring
4. **C**: indicatore del censoring (1 = evento; 0 = censored)

L'esercitazione è molto semplice (e all'esame non ci sarà niente di più), ma è istruttiva anche al di là dei comandi per eseguire delle semplici survival analysis.

**Dobbiamo ricondurre le quantità contenute nel file di dati:**

1. prima, alla teoria ("Chi sono  $T$  e  $\delta$ ?");
2. poi, alla sintassi di MATLAB ("Come si specificano in MATLAB?").



### ESERCIZIO 1 - PARTE 1: KAPLAN-MEIER (svolto)

- Caricare i dati.
- Usare la funzione **ecdf** ("empirical cumulative distribution function") per disegnare la curva di Kaplan-Meier dell'intero dataset.
- Usare intelligentemente la funzione **interp1** per sovrapporre alla curva le crocette che tipicamente indicano i tempi di censoring.
  - Suggerimento: usare come quarto argomento la stringa **'previous'**.

### ESERCIZIO 1 - PARTE 2: COX PROPORTIONAL HAZARDS MODEL (svolto)

- Usando la funzione **coxphfit**, stimare l'hazard ratio associato al BMI.
  - **Attenzione!** Il censoring è specificato "al contrario" rispetto alla teoria.
- Risalire al p value corrispondente a tale hazard ratio: cosa si può concludere?



### ESERCIZIO 1 - PARTE 3: KAPLAN-MEIER PER GRUPPI (proposto)

- Dividere i dati in due gruppi, uno in cui il BMI è minore di  $30 \text{ kg/m}^2$  e l'altro complementare.
- Usare la funzione **ecdf** ("empirical cumulative distribution function") per disegnare le due curve di Kaplan-Meier, una per gruppo, sovrapposte sullo stesso plot.
- Usare intelligentemente la funzione **interp1** come nella parte 1.

### ESERCIZIO 1 - PARTE 4: COX PROPORTIONAL HAZARDS MODEL SULL'INDICATORE DEL GRUPPO (proposto)

- Usando la funzione **coxphfit**, stimare l'hazard ratio associato al gruppo di appartenenza (BMI minore o maggiore di  $30 \text{ kg/m}^2$ ).
- Trarre le conclusioni del caso.