METODI STATISTICI PER LA BIOINGEGNERIA

Università degli Studi di Padova

A.A. 2024-2025

Prof. Alessandra Bertoldo
Ing. Mattia De Francisci, Ing. Claudia Tarricone

- DATASET

Cosa trovate nel file *data_Lab5.mat* nella cartella *data*:

`elsa`: matrice di dimensioni 1000x16

Ogni riga corrisponde ad un diverso soggetto e contiene le realizzazioni delle variabili per quel soggetto.

Ogni colonna corrisponde ad una diversa variabile clinica (e.g., age, BMI (body mass index), heart rate…)

`elsa_labels`: array di celle di dimensioni 1x16

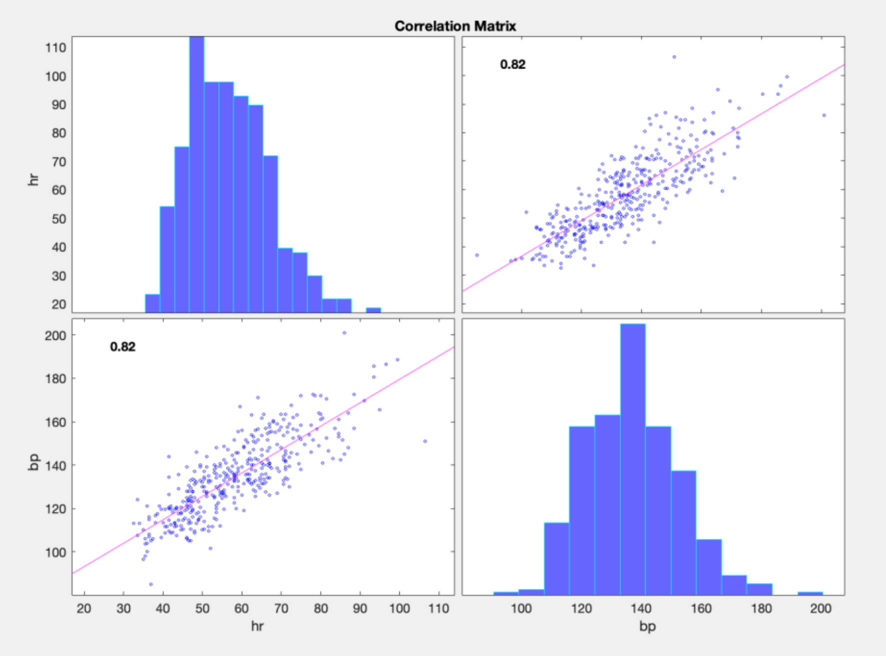
Ogni elemento corrisponde ad una diversa etichetta delle variabili cliniche.

ESERCIZIO 1.

A. Visualizzazione dei dati:

- 1) Caricare i dati
- 2) Eliminare eventuali soggetti che presentino in almeno una delle 16 variabili **negativi, NaN** o **inf** perché non fisiologici (`any, isnan, isinf`).
- 3) Caricare in due **vettori colonna** (`hr e bp`) i dati relativi alle due variabili del dataset ELSA dopo la pulizia corrispondenti rispettivamente ad heart_rate (colonna 4) e systolic_blood_pressure (colonna 15).
- 4) Utilizzare la funzione corrplot (Econometric toolbox) e osservare le distribuzioni e le relazioni tra le due variabili: sono correlate?

Le variabili possono essere considerate gaussiane?



UNIVERSITÀ DEGLI STUDI DI PADOVA

B. Impostazione del modello e stima dei parametri:

5) Impostare le matrici Y e X nell'ipotesi di utilizzare un modello lineare per stimare la frequenza cardiaca (Y) a partire dai dati di pressione sistolica (X).

$$Y = X\beta + q + \varepsilon$$

Visti i dati a disposizione, è necessaria la costante q nel modello? Cosa sarebbe cambiato se avessimo effettuato lo z-score dei dati?

[BONUS da fare a casa] ripetere l'esercizio con z-score dei dati.

Suggerimento: nel caso in cui sia necessario stimare la costante q del modello, in MATLAB occorre aggiungere una colonna di valori unitari alla matrice X.

$$Y = \begin{bmatrix} hr_1 \\ hr_1 \\ \vdots \\ hr_N \end{bmatrix} \qquad X = \begin{bmatrix} bp_1 & 1 \\ bp_2 & 1 \\ \vdots & 1 \\ bp_N & 1 \end{bmatrix} \qquad \beta = \begin{bmatrix} \beta_{bp} \\ q \end{bmatrix}$$

B. Impostazione del modello e stima dei parametri:

6) Stimare i parametri (β) del modello lineare con la formula chiusa vista a lezione:

$$\hat{\beta}^{LS} = (X^T X)^{-1} X^T Y$$

N.B. ricordarsi l'implementazione efficiente dell'inversa di una matrice vista nelle prime lezioni.

7) Ricavare la «predizione» della frequenza cardiaca $\hat{Y}$ a partire dai beta stimati e dai dati di pressione sistolica (X).

C. Verificare i risultati ottenuti in termini di errori di misura e precisione delle stime:

8) Ricavare e riportare a console (disp) l'indice di determinazione tra la variabile spiegata e la predizione del modello ottenuta al punto 7.

9) Calcolare i residui del modello $\left[Y - X\hat{\beta}^{LS}\right]$

10) Ricavare la stima della varianza dell'errore di misura della variabile spiegata

$$\hat{\sigma}^2 = \frac{\left[Y - X\hat{\beta}^{LS}\right]^T \left[Y - X\hat{\beta}^{LS}\right]}{n-m} = \frac{SSR}{n-m}$$

11) Calcolare l'errore standard delle misure a disposizione e successivamente i coefficienti di variazione dei parametri del modello (N.B. della matrice di covarianza siamo interessati agli elementi sulla diagonale)

$$Var\left(\hat{\beta}^{LS}\right) = \sqrt{\hat{\sigma}^2 (X^T X)^{-1}}$$

D. Visualizzazione grafica dei risultati:

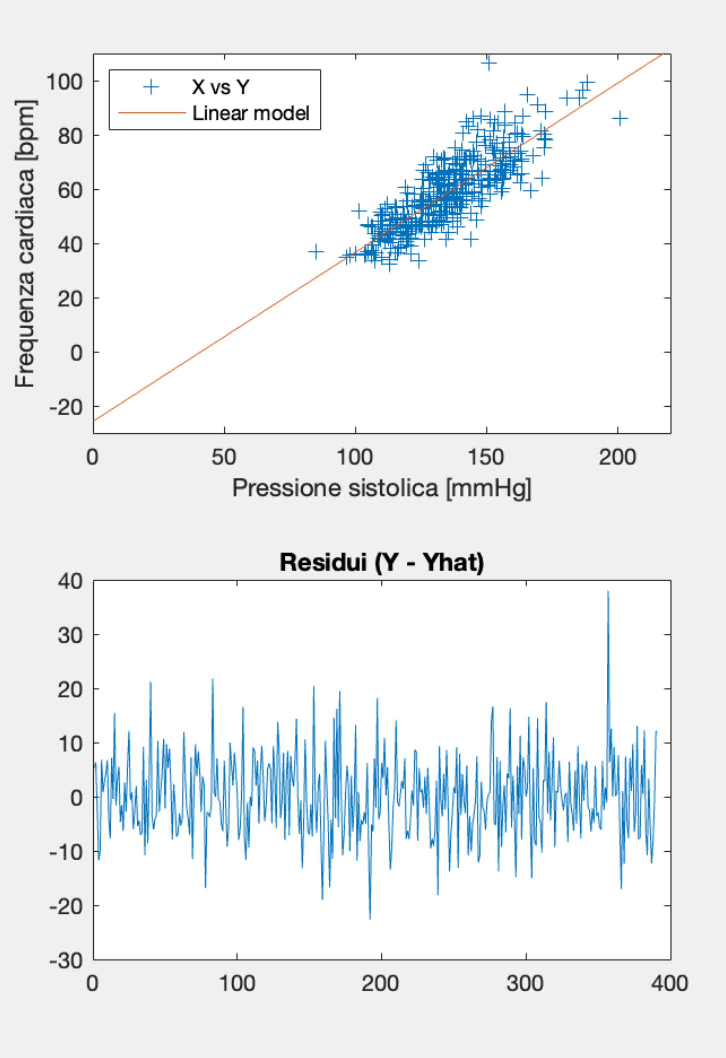
12) Visualizzare tramite scatter-plot le predizioni della frequenza cardiaca $\hat{Y}$ e i dati a disposizione Y

13) Ricavare la retta di regressione utilizzando i parametri appena stimati in un intervallo uniforme da 0 a 220 mmHg con 1000 elementi (linspace).

N.B. Ricordare di includere anche la colonna della costante q

14) Visualizzare in una figura con due sub-plot

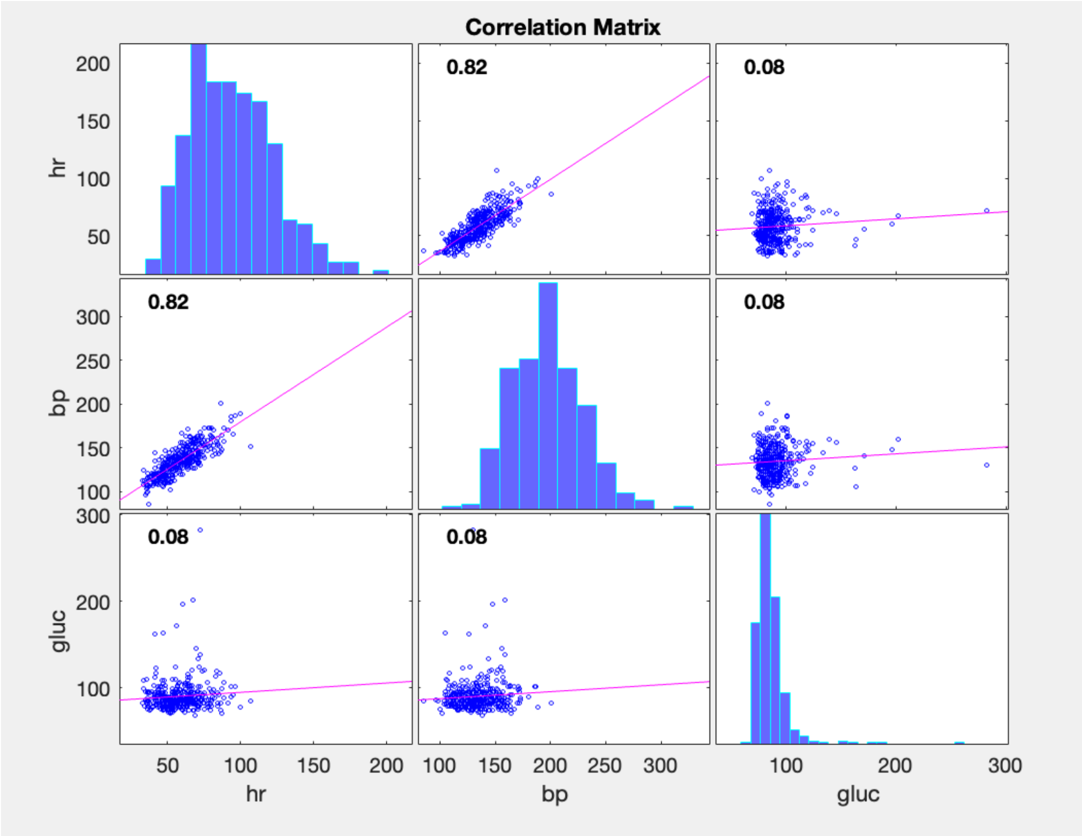
- dati, la predizione del modello $\hat{Y}$ e la retta di regressione
- residui del modello



ESERCIZIO 2.

Ripetere l'esercizio 1 aggiungendo una variabile (una colonna) ad X. In altre parole si vuole utilizzare sia la Pressione sistolica che la concentrazione di glucosio (colonna 9) per stimare la frequenza cardiaca.

Quando si ricavano le precisioni di stima dei parametri le si confrontino con il modello che utilizza la sola pressione sistolica: che apporto da la variabile concentrazione di glucosio in questo modello?



ESERCIZIO 2.

[BONUS] L'ultimo punto in cui si chiede di visualizzare la predizione del modello insieme a dati e retta di regressione nel caso di due variabili, si può interpretare come un piano.

Per visualizzare il piano si deve calcolare la predizione del modello per coppie di valori (X1 e X2), si può fare ciò utilizzando meshgrid.