METODI STATISTICI PER LA BIOINGEGNERIA

A.A. 2024-2025

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

- DATASET

Cosa trovate nel file `data_SUVR.mat` nella cartella `data`:

`SUVR`: matrice di dimensioni 218x1

Ogni riga corrisponde ad un diverso soggetto e contiene un valore quantificato a partire da un'immagine acquisita con la tecnica della tomografia a emissione di positroni (PET)

`G_bivar`: matrice di dimensioni 218x51

Ogni riga corrisponde ad un soggetto, ogni colonna rappresenta una diversa variabile indipendente da utilizzare nel modello di regressione lineare.

Pulizia:

- Eliminare eventuali soggetti che presentino in almeno una delle 51 variabili valori **negativi, NaN** o **inf** perché non fisiologici (`any, isnan, isinf`).
- Effettuare lo z-score per le variabili indipendenti e dipendenti per il modello $Y = X\beta + \varepsilon$ con X → G_bivar e Y → *SUVR*

Stima ai minimi quadrati:

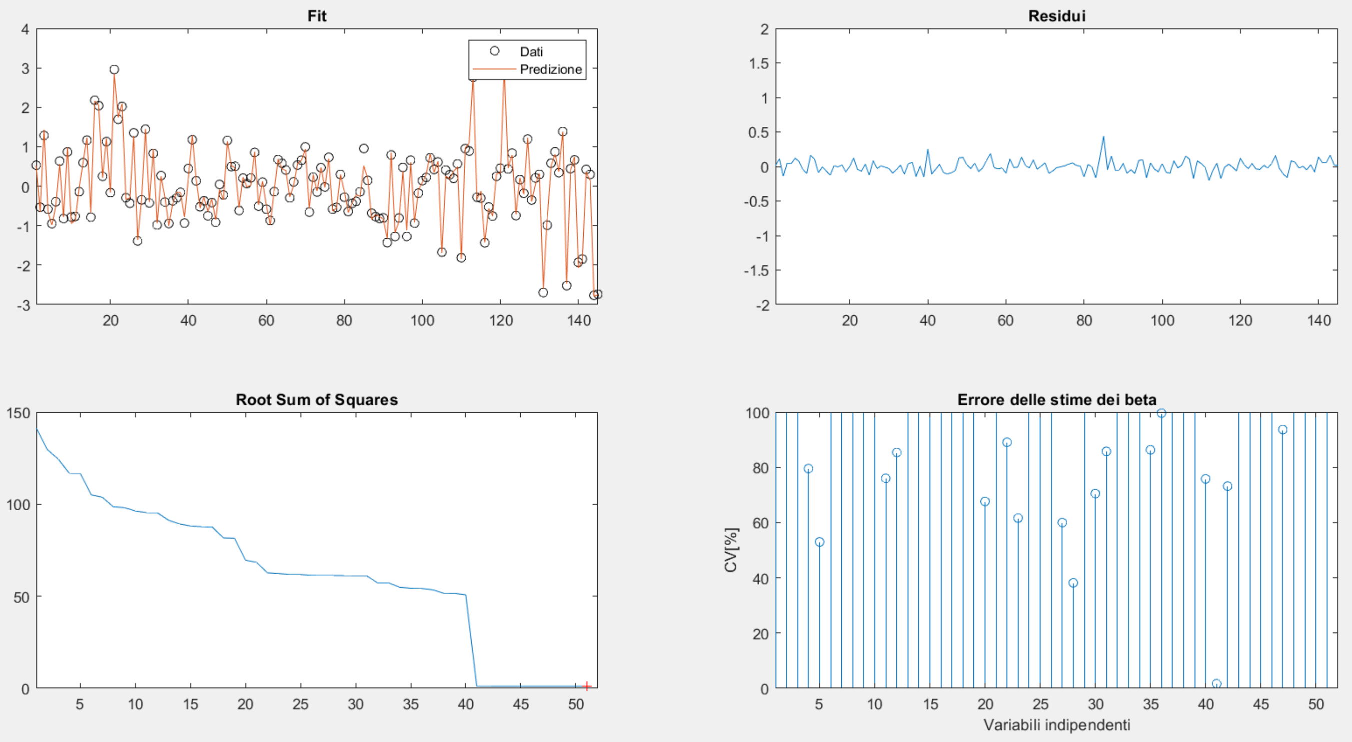
- Costruire il modello di regressione lineare prendendo in considerazione un sottoinsieme crescente di variabili indipendenti (colonne di X da 1 a 51)

  N.B. definire un ciclo all'interno del quale viene ottenuta una matrice G con le prime `i` colonne di X
- Per ogni matrice G stimare i beta con la formula dei minimi quadrati $\hat{\beta}^{LS} = (G^T G)^{-1} G^T Y$
- Calcolare i residui e la somma dei residui al quadrato (RSS)

Stima ai minimi quadrati:

- Mostrare 4 sub-plot in una figura ad ogni iterazione, che contenga:

1. Dati (linea nera tratteggiata) e predizione del modello considerato (linea rossa)
2. I residui del modello considerato
3. Un plot con l'andamento dell'RSS (che comprenda anche i valori dei modelli precedenti)
4. Una rappresentazione dell'errore delle stime (CV %) (si veda `stem`)



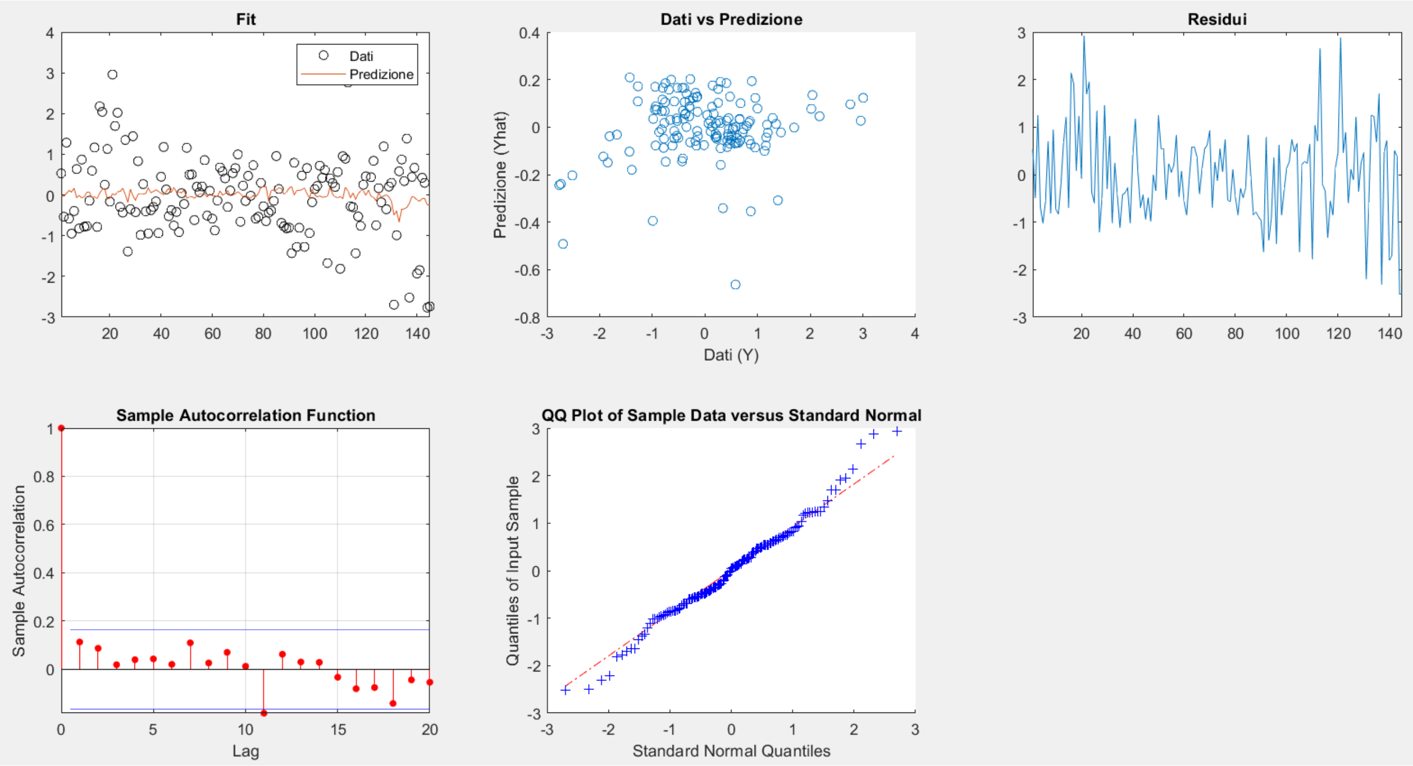
Stima ai minimi quadrati:

- Stimare la varianze dell'errore di misura per ogni modello considerato $$\hat{\sigma}^2 = \frac{\left[Y - X\hat{\beta}^{LS}\right]^T \left[Y - X\hat{\beta}^{LS}\right]}{n-m} = \frac{SSR}{n-m}$$
- Calcolare i coefficienti di variazione dei beta stimati $$Var\left(\hat{\beta}^{LS}\right) = \hat{\sigma}^2 (X^T X)^{-1}$$
- Calcolare il coefficiente di determinazione R^2 (R2) utilizzando la correlazione
- Calcolare e salvare in una variabile (p_val_R2) il p–value ottenuto dalla funzione corr

Per il modello con una colonna (prima iterazione del ciclo for) mostrare una figura con 5 grafici:

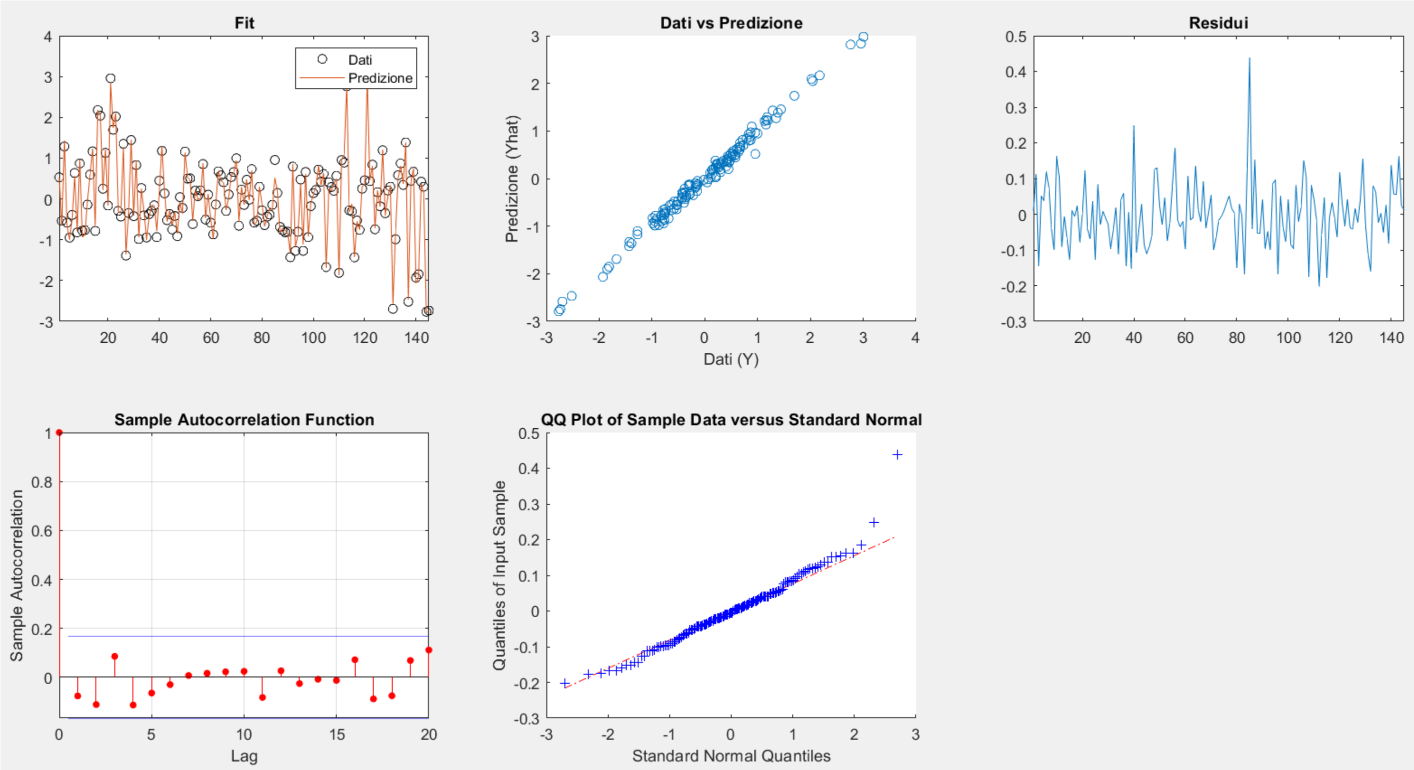
1. Dati (cerchi neri) e predizione del modello considerato (linea rossa)
2. Uno scatter plot tra dati e predizione
3. I residui del modello considerato
4. Il grafico dell'autocorrelazione
5. Il qq-plot

Notiamo qualcosa dal fit e dall'autocorrelazione?



Si ripetano i 5 grafici per il caso con G=X (ultima iterazione)

1. Dati (cerchi neri) e predizione del modello considerato (linea rossa)
2. Uno scatter plot tra dati e predizione
3. I residui del modello considerato
4. Il grafico dell'autocorrelazione
5. Il qq-plot



Notiamo qualcosa dai residui?

Preparare una figura in cui si mostrino:

1. L'andamento di RSS in funzione del numero di colonne della matrice G: *notiamo qualche caratteristica nell'andamento del grafico? Ci sono «discontinuità» marcate?*
2. I coefficienti beta stimati con l'ultimo modello considerato
3. I coefficienti di variazione dei suddetti beta: *che considerazioni si possono fare? Qualche CV dei beta stimati coglie la nostra attenzione?*

Si mostrino a video i valori del coefficiente di determinazione R2 con G contenente una singola variabile e con G contente tutte le colonne di X: che differenze si notano?