

UNIVERSITÀ
DEGLI STUDI
DI PADOVA

METODI STATISTICI PER LA BIOINGEGNERIA

Laboratorio 7

A.A. 2025-2026

Enrico Longato



Dal lab 6: Ma il codice è davvero così semplice?

Confermo: la regressione lineare si fa col "copia-incolla".

Il vero problema è mettere i dati giusti dentro alla matrice X e al vettore Y!

```
118 %% Regressione lineare di pp su SBP
119 % Scrivo la design matrix, ricordandomi di aggiungere l'intercetta alla
120 % fine
121 X = [sbp ones(length(sbp), 1)];
122
123 % Scrivo il vettore di outcome sincerandomi che sia una colonna
124 % (size(Y) deve essere [N_righe, 1])
125 Y = pp;
126
127 % Utilizzo la formula risolutiva per trovare i parametri
128 beta_hat = (X'*X)\X'*Y;
129
```

```
60 %% Regressione lineare di pp su SBP e glicemia
61 % Scrivo la design matrix, ricordandomi di aggiungere l'intercetta alla
62 % fine
63 X = [sbp gluc ones(length(sbp), 1)];
64
65 % Scrivo il vettore di outcome sincerandomi che sia una colonna
66 % (size(Y) deve essere [N_righe, 1])
67 Y = pp;
68
69 % Utilizzo la formula risolutiva per trovare i parametri
70 beta_hat = (X'*X)\X'*Y;
71
```

```
62 %% Regressione lineare
63 % Scrivo la design matrix, ricordandomi di aggiungere l'intercetta alla
64 % fine
65 X = [X_no_intercept ones(length(pp), 1)];
66
67 % Scrivo il vettore di outcome sincerandomi che sia una colonna
68 % (size(Y) deve essere [N_righe, 1])
69 Y = pp;
70
71 % Utilizzo la formula risolutiva per trovare i parametri
72 beta_hat = (X'*X)\X'*Y;
73
```



Dal lab 6: Stimatori e denominatori

Regola mnemonica: in tantissimi casi, i denominatori degli stimatori non polarizzati correggono per il **numero di campioni** meno il **numero di parametri già stimati** prima di poter stimare la quantità di interesse

$$\sigma_{campionaria}^2 = \frac{1}{N_{campioni} - 1} \sum_{i=1}^{N_{campioni}} (x_i - \mu_x)^2$$

Il parametro già stimato è 1:
la media μ_x

$$\sigma_{a\ posteriori}^2 = \frac{1}{N_{campioni} - N_{parametri}} \sum_{i=1}^{N_{campioni}} (y_i - \hat{y}_i)^2$$

I parametri già stimati sono tanti quanti sono i β_i , incluso, se presente nel modello, il β_0 dell'intercetta, perché sono quelli che servono per calcolare $\hat{y} = X\beta$

$$N_{parametri} = M_{variabili} + 1_{intercetta, se\ presente}$$



Laboratorio 7: Contenuti e obiettivi

1. Esercitazione "alla lavagna"
 - Analisi dei residui
2. Esercizi da svolgere in autonomia (per superare la "paura del file bianco")
 - Stima dei parametri e della loro variabilità
 - Calcolo di R^2
 - Calcolo dell'RMSE



Analisi dei residui: i residui devono avere 4 proprietà, che cerchiamo di verificare come segue

1. Normalità

- Come quando lo facevamo per il test di ipotesi (stesse identiche deduzioni): istogramma (**hist**), QQ-plot (**qqplot**), test di gaussianità (**lillietest**), skewness (**skewness**), curtosi (**kurtosis**).

2. Media nulla

- Semplicemente, guardiamo la media (**mean**), che deve essere vicina a 0.

3. Scorrrelazione

- Plot dell'autocorrelazione dopo aver ordinato secondo il valore predetto:
 1. **[Y_hat_sorted, i_sort] = sort(Y_hat)**
 2. **autocorr(residuals(i_sort))**
- Escluso il valore in 0 (che fa sempre 1), circa il 95% dei campioni dell'autocorrelazione ottenuta dai residui così ordinati deve stare entro le bande di confidenza.

4. Varianza omogenea

- Plot (**scatter**) dei residui (asse y) contro il valore predetto (asse x).
- Deve vedersi una "banda" o "nuvola" omogenea intorno alla riga orizzontale che passa per $y = 0$.
- **ATTENZIONE: NON fare il plot contro il valore vero** (che, per costruzione, è correlato ai residui).

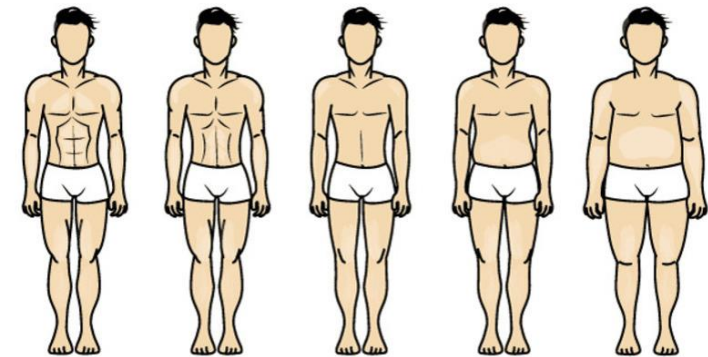


CONTESTO DELL'ESERCITAZIONE E DATI

Dataset di misure antropometriche per la predizione della % di grasso corporeo (**bodyfat.mat**).

Dati di 252 uomini descritti da 8 variabili

1. BodyFat (in %) sarà la nostra variabile dipendente
2. Age (età in anni, years)
3. Weight (peso in libbre, lbs)
4. Height (altezza in pollici, inches)
5. Neck (circonferenza del collo in cm)
6. Chest (circonferenza del petto in cm)
7. Hip (circonferenza dei fianchi in cm)
8. Thigh (circonferenza della coscia in cm)





Prima di svolgere l'esercitazione (oppure al bisogno), utilizzare il comando **help di MATLAB seguito dal nome delle seguenti function, utili allo svolgimento degli esercizi.**

Analisi dei residui

- v. slide precedente



ESERCIZIO 1 - PARTE 1: CARICAMENTO DATI E REGRESSIONE LINEARE (proposto)

- Caricare il file **bodyfat.mat**
- **Considerando BodyFat come variabile dipendente e tutte le altre come variabili indipendenti**
 - Stimare i parametri di un modello di regressione lineare, intercetta inclusa
 - Calcolare lo standard error e il CV percentuale dei parametri
 - **[NEW ma semplice]** Calcolare l'intervallo di confidenza delle stime dei parametri come l'intervallo di estremi $\hat{\beta} \pm 1.96 \times SE(\hat{\beta})$
 - Plottare il valore vero contro il valore predetto
 - Calcolare l'RMSE
 - Calcolare R^2



ESERCIZIO 1 - PARTE 2: ANALISI DEI RESIDUI (svolto)

- Verificare le quattro proprietà fondamentali dei residui
 1. Normalità
 2. Media nulla
 3. Scorrelazione
 4. Varianza omogenea
- Seguire la traccia di soluzione nella slide introduttiva, che comprende:
 - Calcoli da fare
 - Stima della media
 - Skewness e curtosi
 - Test di gaussianità
 - Plot da disegnare
 - Istogramma dei residui
 - QQ-plot dei residui
 - Autocorrelazione dei residui (dopo aver ordinato per valore predetto)
 - Scatter plot dei residui contro il valore predetto