

Relazione di Statistica Descrittiva

Daniele Buser 894514 — <https://github.com/dbuser03/progetto-psi.git>

May 28, 2024

1 Introduzione

Il rapporto analizza statisticamente i dati di 3154 soggetti, includendo variabili come età, altezza, peso, pressione sistolica e diastolica, colesterolo e numero di sigarette fumate al giorno al fine di correlarli con la presenza di malattia coronarica (CHD).

2 Statistiche Descrittive

Prima dell'analisi, ho eseguito una pulizia dei dati gestendo i valori mancanti e gli outlier. I valori mancanti sono stati sostituiti con la moda, media o mediana in base alla distribuzione dei dati, mentre gli outlier sono stati individuati tramite il metodo IQR e sostituiti con il lower o upper bound. La variabile CHD è stata convertita in binaria (0, 1), mentre peso e altezza sono stati convertiti in cm e kg.

	CHD	Age	Height	Weight	SBP	DBP	Chol	NCigs
Min.	0.00000	39.00	152.4	35.38	98.0	58.00	103.0	0.00
1st Qu.	0.00000	42.00	172.7	70.31	120.0	76.00	198.0	0.00
Median	0.00000	45.00	177.8	77.11	126.0	80.00	223.0	0.00
Mean	0.08148	46.28	177.2	77.08	128.6	82.01	226.3	11.59
3rd Qu.	0.00000	50.00	182.9	82.55	136.0	86.00	252.8	20.00
Max.	1.00000	59.00	198.1	122.47	202.0	130.00	437.5	75.00

Table 1: Statistiche descrittive delle variabili

3 Grafici dei dati

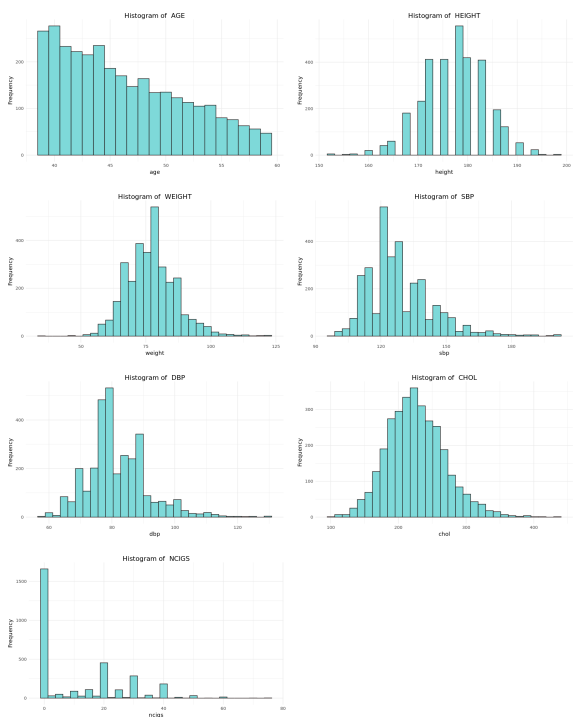


Figure 1: Distribuzione delle variabili continue

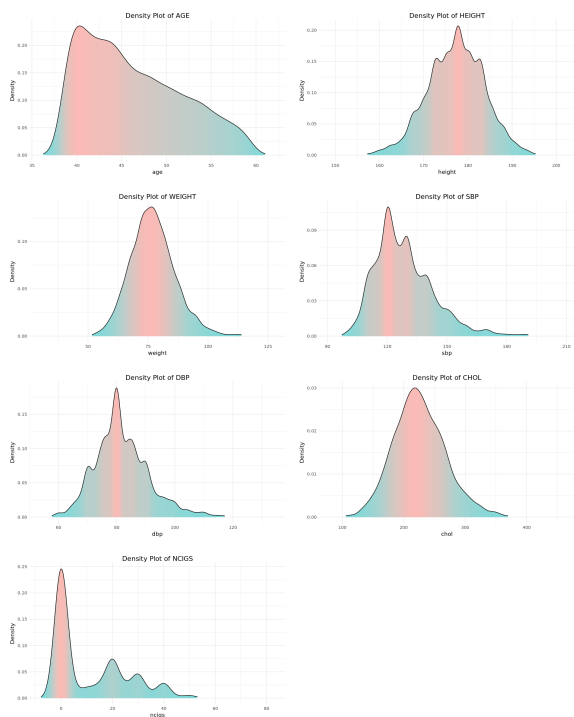


Figure 2: Grafici a densità delle variabili continue

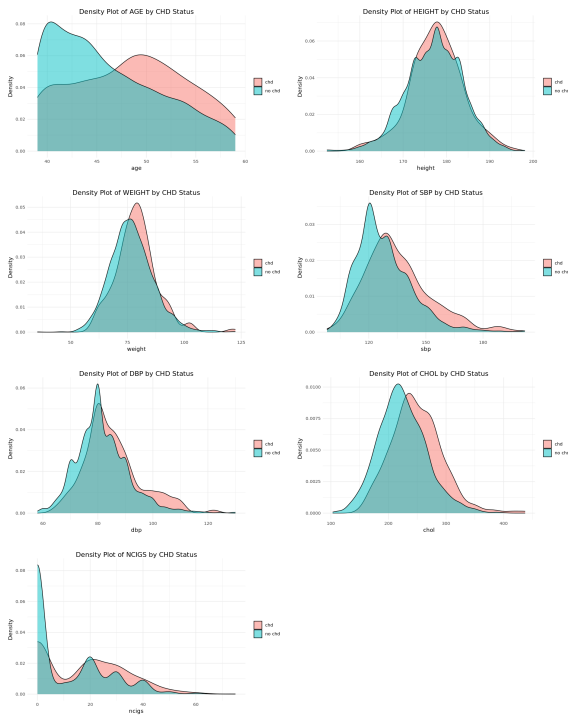


Figure 3: Grafici a densità divisi per CHD

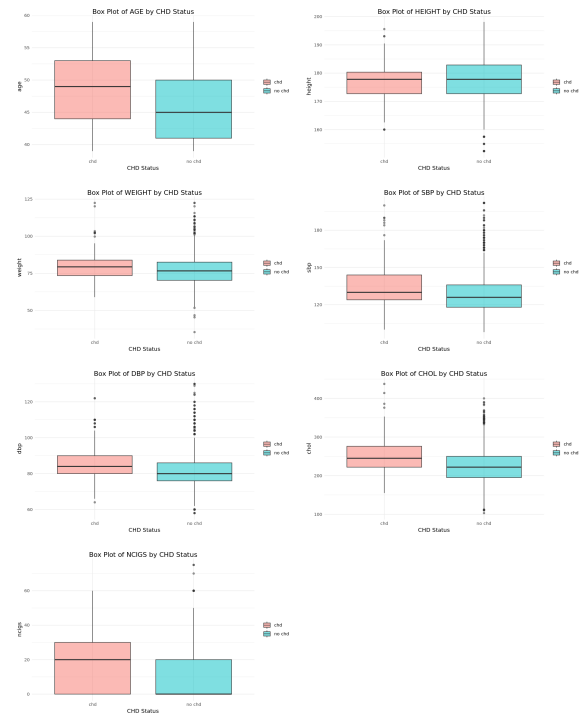


Figure 4: Boxplot delle variabili divise per CHD

4 Correlazioni

Variable	Pearson Correlation	t-value	p-value	CI Lower	CI Upper
Age	0.1192632	6.7439	1.827×10^{-11}	0.08471402	0.15352589
Height	0.01884893	1.0584	2.899×10^{-1}	-0.01606341	0.05371536
Weight	0.06161473	3.4658	5.357×10^{-4}	0.02677053	0.09630939
SBP	0.1343514	7.6119	3.547×10^{-14}	0.09991812	0.16846319
DBP	0.101808	5.7456	1.003×10^{-8}	0.06714486	0.13622577
Chol	0.1596526	9.0798	$<2.2 \times 10^{-16}$	0.1254499	0.1934763
NCigs	0.104454	5.8966	4.104×10^{-9}	0.06980667	0.13884953

Table 2: Pearson's product-moment correlation results

L'analisi dei dati evidenzia distinzioni tra gli individui affetti da malattia coronarica e quelli senza la condizione. In primo luogo, risulta significativo il ruolo dell'età: gli individui con CHD sono generalmente più anziani, con una concentrazione della densità intorno ai 50 anni. Questa relazione è supportata da una correlazione positiva e debole tra età e CHD. Inoltre, si osserva che il peso è leggermente più elevato tra coloro con CHD, con una concentrazione della densità tra gli 80-85 kg. Analogamente, la pressione sistolica raggiunge un picco intorno ai 130-140 mmHg tra gli individui con CHD, evidenziando la correlazione tra questo parametro e la malattia cardiaca. Per quanto riguarda il colesterolo, si registra un aumento dei livelli tra le persone affette da CHD, con una concentrazione intorno ai 250 mg/dL, indicando una correlazione debole e positiva più significativa rispetto ad altri fattori (Pearson Correlation = 0.1596 | p-value = $<2.2 \times 10^{-16}$). Queste tendenze sono confermate dai boxplot, che mostrano mediane più elevate per età, peso, pressione sistolica, colesterolo e numero di sigarette fumate al giorno tra gli individui con CHD rispetto a quelli senza la malattia.

5 Conclusioni

In base a questi risultati, possiamo concludere che tra i vari fattori studiati, il colesterolo, la pressione sistolica e l'età sono quelli la cui correlazione alla presenza di malattia coronarica (CHD) risulta statisticamente più significativa. Altri fattori come il peso e il numero di sigarette fumate al giorno rappresentano comunque potenziali fattori di rischio e richiedono quindi attenzione e un monitoraggio continuo. Non è escluso che altri fattori non considerati possano avere un ruolo significativo nella determinazione della malattia coronarica.