

Analisi dei dati del dataset WCGS

Questo documento propone un'analisi dei dati del dataset WCGS usando la statistica descrittiva, nell'intento di individuare i principali fattori di rischio dello sviluppo di malattie coronariche.

Il dataset WCGS contiene i dati empirici dello studio Western Collaborative Group Study (WCGS) epidemiologico prospettico su 3154 soggetti, per misurare il rischio di malattie coronariche (chd) basandosi su altri fattori di rischio, come il livello di colesterolo (cho_l), l'età (age), l'altezza (height), il peso (weight), la pressione arteriosa sistolica (sbp) e diastolica (dbp), e il numero di sigarette fumate giornalmente (ncigs).

Indici di posizione e di dispersione

Si riportano gli indici di posizione e dispersione relativi alle variabili del dataset WCGS.

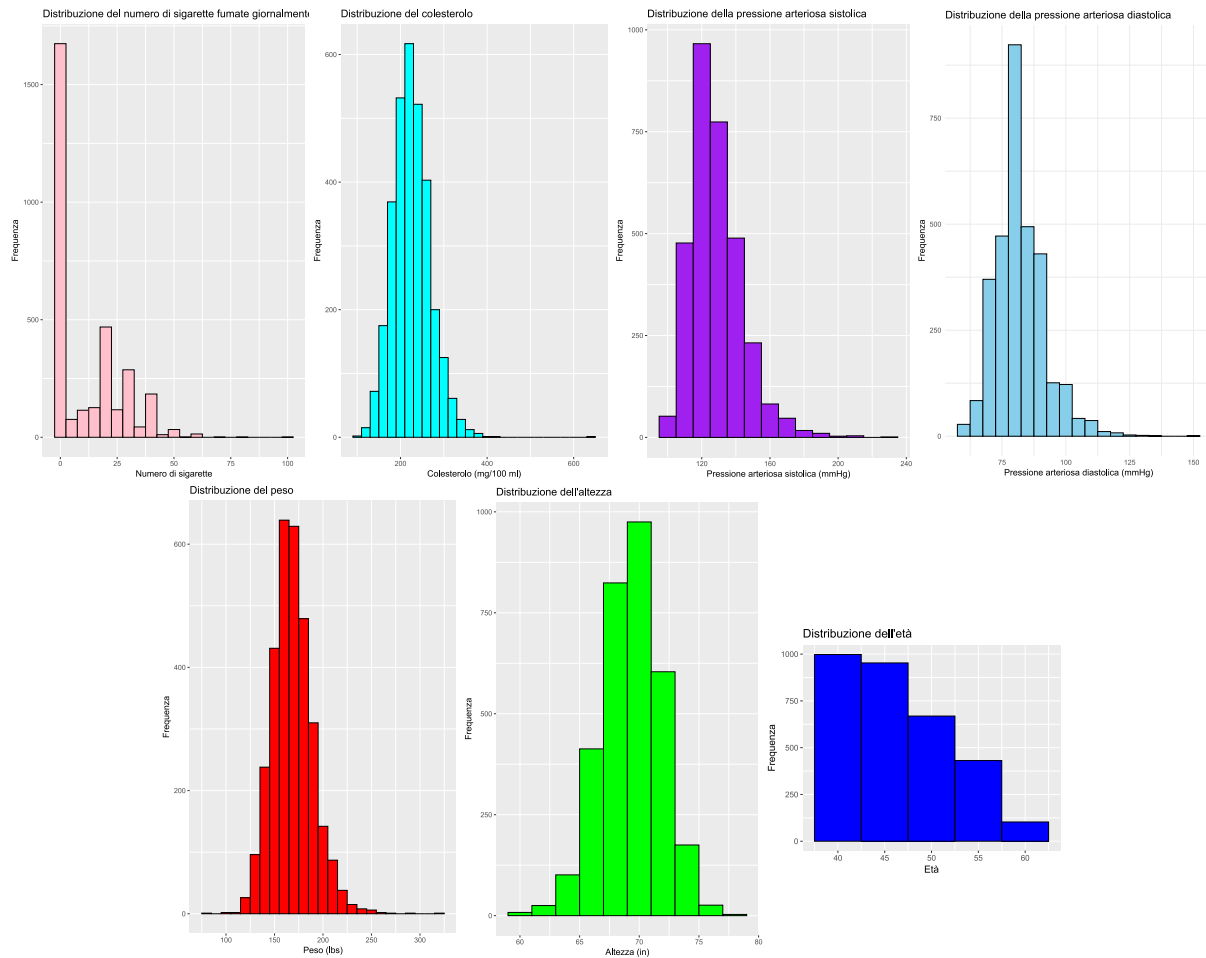
Var	Minimo	1° quadr.	Mediana	Media	Varianza	Dev. standard	3° quadr.	Massimo
Età	39.00	42.00	45.00	46.28	30.515	5.5240	50.00	59.00
Altezza (in pollici)	60.00	68.00	70.00	69.78	6.3942	2.5286	72.00	70.00
Peso (in libbre)	78	15	170	170	445.03	21.095	182	320
Pressione arteriosa sistolica (mmHg)	98.0	120.0	126.0	128.6	228.54	15.117	136.0	230.0
Pressione arteriosa diastolica (mmHg)	58.00	76.00	80.00	82.02	94.612	9.7268	86.00	150.00
Livello di colesterolo (mg/100 ml)	103.0	197.2	223.0	226.4	43.00	6.5574	253.0	645.0
Numero di sigarette fumate al giorno	0.0	0.0	0.0	11.60	210.76	14.517	20.0	99.0

Frequenze assolute e relative dell'insorgenza di malattie coronariche

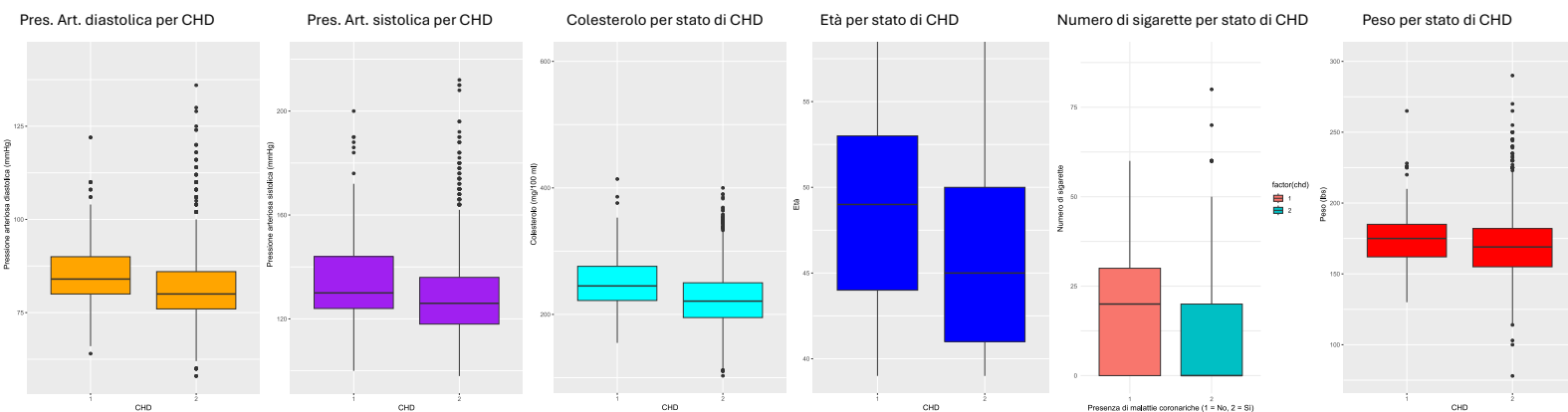
Si riportano le frequenze assolute e relative dei soggetti che hanno sviluppato una malattia coronarica.

Tipologia	Frequenza assoluta	Frequenza relativa
Insorgenza malattia coronarica	257	0.08148383
Assenza malattia coronarica	2897	0.9185162

Rappresentazione grafica distribuzioni



Box plot di confronto



Analizzando i valori disponibili e i grafici sopra mostrati, si può ipotizzare che il fumo di sigarette può contribuire allo sviluppo di malattie coronariche, come si evince dal box plot no. 5. Si può notare inoltre che altri fattori di rischio per lo sviluppo di malattie coronariche sono un'**alta pressione arteriosa**, un valore **alto di colesterolo**, e il **sovrappeso**. Questi risultati suggeriscono che, al fine di attuare strategie per prevenire il rischio di malattie coronariche, è importante mantenere questi fattori sotto controllo.