

Introduzione

In questo progetto si vuole analizzare i dati relativi a 3145 soggetti esaminati in base a 9 variabili (età, altezza, peso, pressione sistolica/diastolica, colesterolo e numero di sigarette fumate giornalmente). L'obiettivo è identificare i fattori di rischio e individuare una correlazione tra l'insorgenza di una **malattia coronaria (MC)** e le variabili disponibili.

Prima Analisi

Variabile	Min	I Quart	II Quart (Mediana)	Media	III Quart	Max
AGE	39	42	45	46.28	50	59
HEIGHT	60 (152 cm)	68 (172,7 cm)	70 (177,8 cm)	69.78 (177,5 cm)	72 (182,9 cm)	78 (198 cm)
WEIGHT	78 (35,4 kg)	155 (70 kg)	170 (77 kg)	170 (77 kg)	182 (82,6 kg)	320 (145 kg)
SBP	98	120	126	128.6	136	230
DBP	58	76	80	82.02	86	150
CHOL	103	197.2	223	226.4	253	645.0
NCIGS	0	0	0	11.6	20	99

In generale: da una prima analisi dei dati si nota che la maggior parte delle variabili hanno una media e una mediana molto vicine indicando una distribuzione pressoché simmetrica e centrale senza asimmetrie importanti.

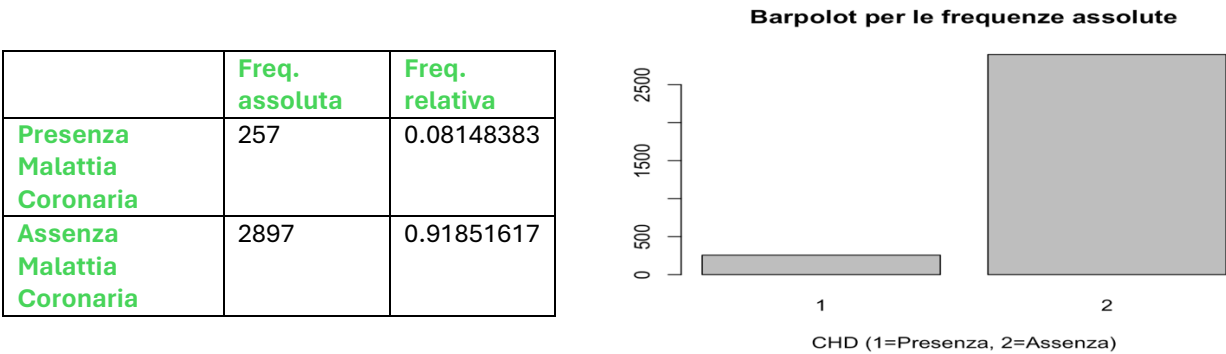
Variabilità per peso e colesterolo: Il peso e il colesterolo presentano una maggiore variabilità rispetto alle altre variabili (infatti i valori minimi e massimi sono molto distanti tra di loro), suggerendo una diversità maggiore nelle abitudini e nelle condizioni di salute dei partecipanti.

Abitudini di Fumo: la distribuzione del numero di sigarette fumate giornalmente è altamente asimmetrica con molti non fumatori e alcuni fumatori pesanti.

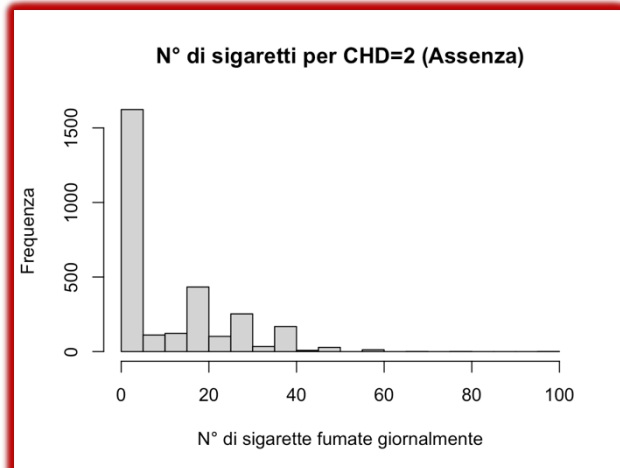
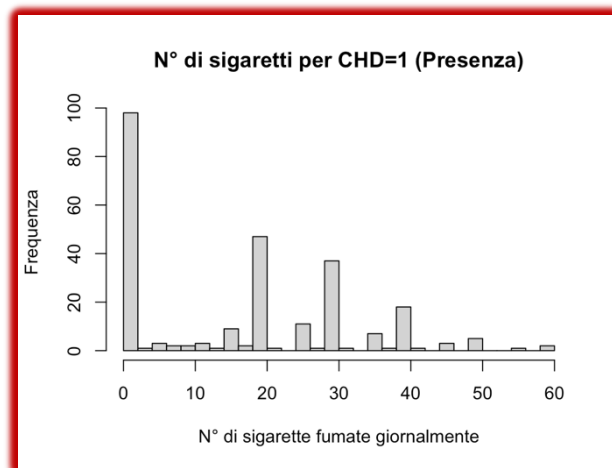
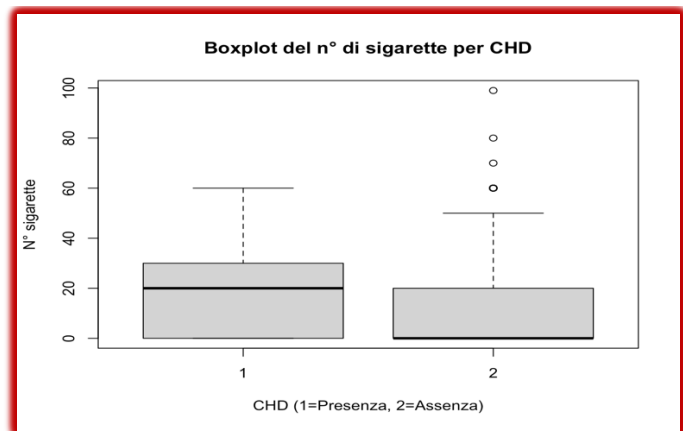
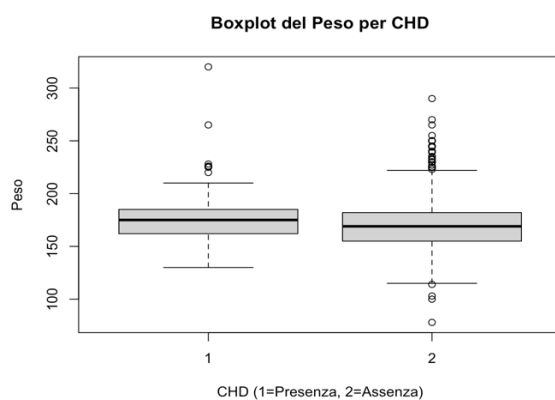
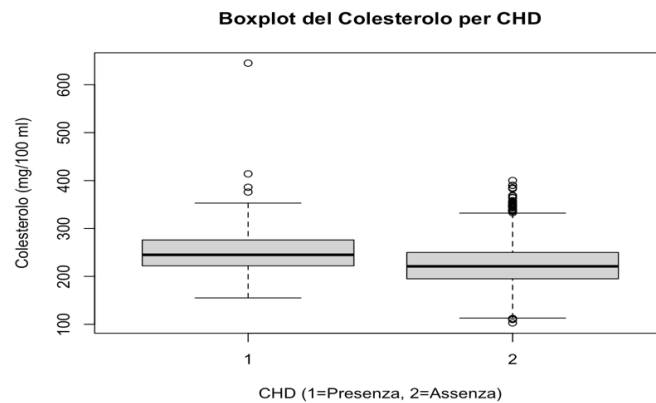
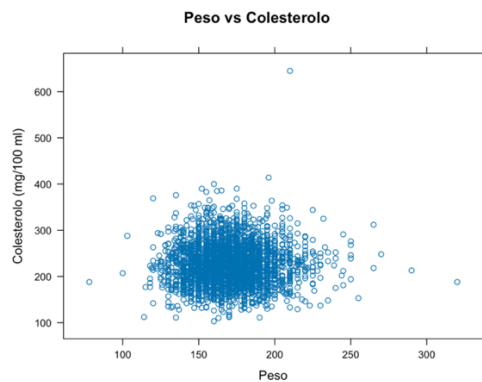
Correlazioni

	age	height	weight	sbp	dbp	chol	ncigs
age	1.000000000	-0.091924057	-0.033078516	0.17005253	0.143979992	0.089188510	-0.006111181
height	-0.091924057	1.000000000	0.533288106	0.01556533	0.006985365	-0.088937779	0.017304324
weight	-0.033078516	0.533288106	1.000000000	0.25127827	0.293968705	0.008537442	-0.082673038
sbp	0.170052531	0.015565329	0.251278272	1.000000000	0.770044438	0.123061297	0.029284852
dbp	0.143979992	0.006985365	0.293968705	0.77004444	1.000000000	0.129597108	-0.062670918
chol	0.089188510	-0.088937779	0.008537442	0.12306130	0.129597108	1.000000000	0.096031834
ncigs	-0.006111181	0.017304324	-0.082673038	0.02928485	-0.062670918	0.096031834	1.000000000

In questa tabella invece possiamo vedere le varie correlazioni tra le variabili. La correlazione più significativa è tra SBP e DBP (correlazione molto significativa dello 0,77 circa); ci sono poi PESO e ALTEZZA, PESO e SBP, PESO e DBP (che invece hanno una correlazione più debole che si aggira attorno ai 0,25 e 0,5).



Grafici



Conclusioni:

Le variabili individuali riportate (età, peso, altezza, pressione sanguigna, colesterolo) **non sembrano essere cause dirette di un'insorgenza di una MC**. Questo potrebbe indicare che nessuna di queste variabili, **presa singolarmente**, ha un impatto significativo sulla probabilità di sviluppare una MC. Nonostante la mancanza di correlazioni forti tra le singole variabili e CHD, è probabile che l'insorgenza di una MC sia influenzata da una combinazione di più fattori. Ad esempio, la combinazione di obesità, alta pressione sanguigna e alto livello di colesterolo potrebbe aumentare il rischio di insorgenza di una MC.

*** L'unica correlazione rilevante trovata riguarda il numero di sigarette fumate giornalmente. Come mostrano i grafici sopra, il numero di sigarette fumate giornalmente può essere associato all'insorgenza di una MC, infatti, **risulta che i soggetti con una MC fumano in media (giornalmente) più sigarette rispetto ai soggetti senza una MC**.