

# Analisi dati cardiologici



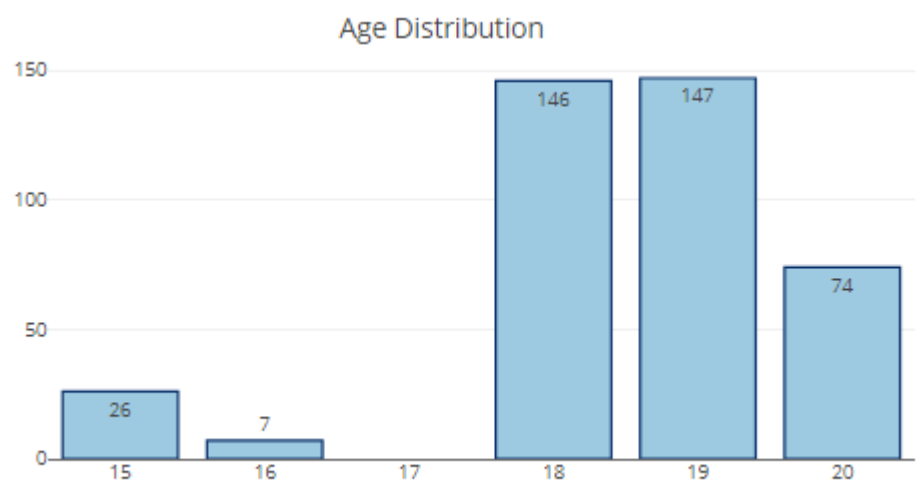
A cura di  
Marco Cortese



Analysis and Statistical models



Il dataset elaborato è caratterizzato da 400 osservazioni. Sono state registrate diverse variabili su pazienti con problemi cardiaci, i soggetti in esame presentavano un'età compresa tra i 15 e i 20 anni con la seguente ripartizione:

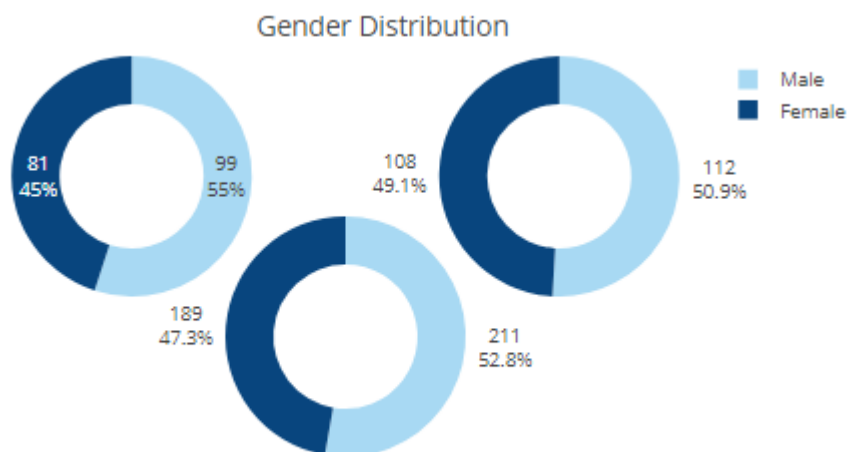


Age Distribution	
	Distribution
15	26
16	7
17	0
18	146
19	147
20	74

I pazienti verranno suddivisi in 2 gruppi di osservazione:

Group		
	Pubertal group	Post-pubertal group
Total students (n = 400)	220 (55%)	180 (45 %)

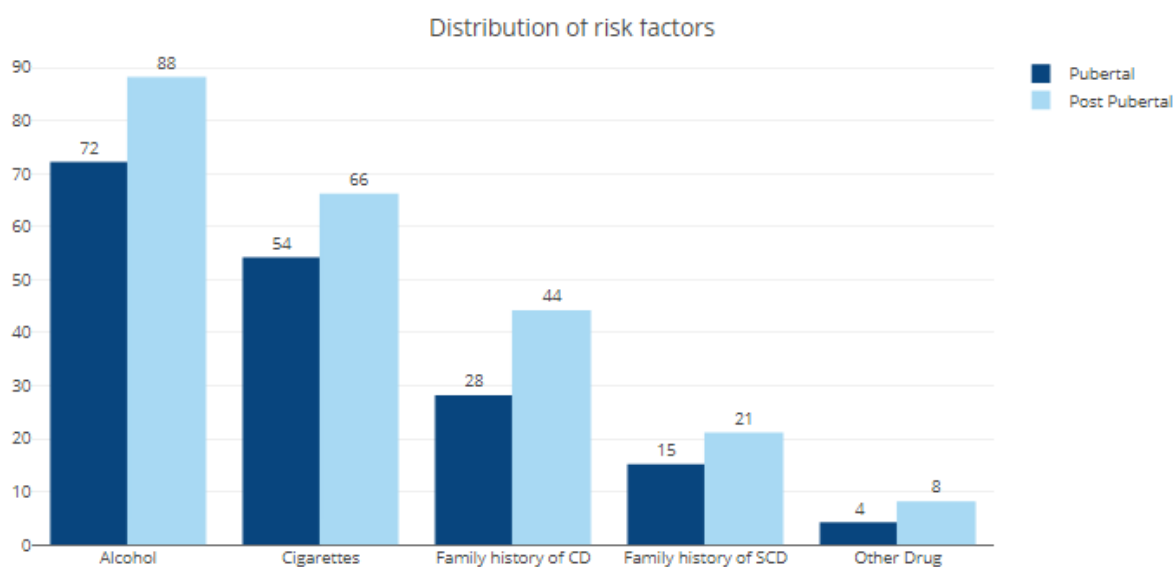




Gender Group Distribution			
	Pubertal group	Post-pubertal group	Total
Male, n (%)	112 (50.9 %)	99 (55 %)	211(52.8%)
Female, n(%)	108 (49.1 %)	81 (45%)	189 (47.3%)
	220	180	

I fattori di rischio considerati nel dataset sono:

- Alcohol
- Cigarettes
- Family history of CD
- Family history of SCD
- Other Drugs





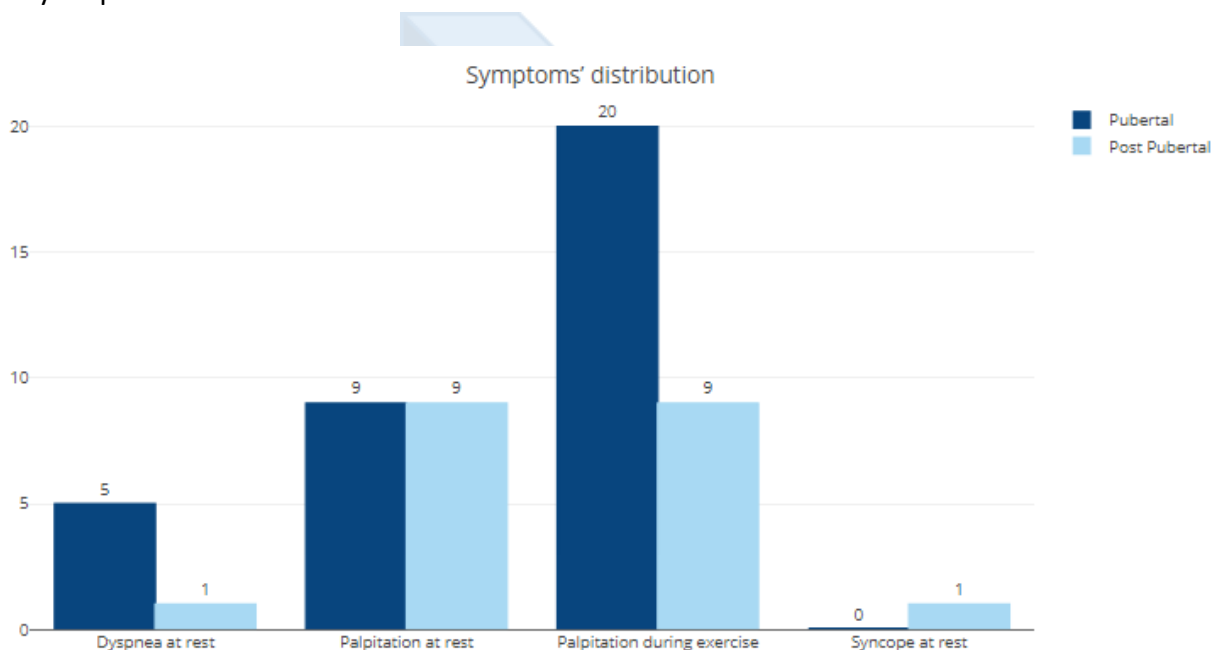
	Total (n = 400)	Pubertal (n = 220)	Post-Pubertal (n = 180)
Family history of CD, n (%)	72 (18 %)	28	44
Family history of SCD, n (%)	36 (5 %)	15	21
Alcohol use, n (%)	160 (40 %)	72	88
Cigarettes, n (%)	120 (30 %)	54	66
Other drugs, n (%)	12 (3 %)	4	8

Sono 197 i pazienti che presentano un solo fattore di rischio, mentre 100 ne presentano 2, 1 paziente evidenzia la presenza di 3 fattori di rischio:

#### Family history of SCD Cigarettes e other drugs.

Tra i 400 pazienti, 54 hanno fatto registrare i seguenti sintomi:

- Dyspnea at rest
- Palpitation at rest
- Palpitation during exercise
- Syncope at rest



In nessun caso osserviamo la presenza di sintomi contemporaneamente.

Symptoms' distribution			
	Total (n = 400)	Pubertal (n = 220)	Post-pubertal (n = 180)
Syncope at rest, n (%)	1 (0.25 %)	0	1
Palpitation at rest, n (%)	18 (4.5 %)	9	9



Palpitation during exercise, n (%)	29 (7.25 %)	20	9
Dyspnea at rest, n (%)	6 (1.5 %)	5	1

Per eseguire le analisi che seguiranno è stata effettuata una codifica delle variabili per renderle numeriche:

### Echocardiographic abnormalities

mitral valve prolapse	→	1
bicuspid aortic valve	→	2
tricuspidal regurgitation	→	3

### CMR abnormal

anomal origin of coronary	→	1
not performed	→	2

### Pathological at the end of screening

X	→	1
NA	→	0

### abnormal T wave

0	→	0
1	→	1
abnormal T wave	→	2

É stata calcolata la matrice di correlazione totale considerando tutte le variabili del dataset con il rho di Spearman essendo le variabili numeriche ma discrete a poche modalità differenti.

Il coefficiente di correlazione è una misura specifica usata nell'analisi della correlazione per quantificare la forza della relazione lineare tra due variabili.

Si distingue in:

- diretta (o positiva): la variazione di un elemento interessa - in via diretta - anche l'altro; valori elevanti di una variabile corrispondono a valori elevanti di un'altra variabile;
- indiretta (anche inversa o negativa): alla variazione di un elemento corrisponde, in senso contrario, quella dell'altro.

Di seguito riportiamo la tabella ridotta delle sole correlazioni che sono  $> 0.5$  o  $< -0.5$

Alcohol use	Many Risk	0.6237464
Palpitation during exercise	Many Symtoms	0.7077065
Palpitation at rest	Many Symtoms	0.5494722
Pathological ECG	Ventricular ectopy	0.5744417
Ventricular ectopy	Pathological ECG	0.5744417
<b>Echocardiographic abnormalities</b>	<b>Pathological at the end of screening</b>	<b>0.8429948</b>
<b>CMR abnormal</b>	<b>Pubertal</b>	<b>0.9989845</b>



Le correlazioni maggiormente interessanti sono:

Echocardiographic abnormalities	Pathological at the end of screening	0.8429948
CMR abnormal	Pubertal	0.9989845

Queste correlazioni positive, molto elevate indicano come ad una modalità elevata della variabile 1 corrisponda una modalità elevata della variabile 2.

Le variabili identificate assumono i seguenti valori:

Echocardiographic abnormalities	→	0,1,2,3
Pathological at the end of screening	→	0,1
CMR abnormal	→	0,1,2
Pubertal	→	0,1

Perciò

Echocardiographic abnormalities	Pathological at the end of screening	0.8429948
---------------------------------	--------------------------------------	-----------

Implica che al tendere del valore 3 per la variabile Echocardiographic abnormalities la variabile Pathological at the end of screening assumerà valore 1; mentre assumerà valore 0 al tendere verso 0 della variabile Echocardiographic abnormalities.

CMR abnormal	Pubertal	0.9989845
--------------	----------	-----------

Implica che al tendere del valore 2 per la variabile CMR abnormal la variabile Pubertal assumerà valore 1; ed entrambi tenderanno a 0 contemporaneamente.

Sul dataset analizzato sono stati effettuati 2 test:

Test Chiquadro con correzione di Fisher e Test Binomiale.

Il primo analizza la distribuzione di una variabile condizionata ai gruppi di una seconda variabile. Il secondo ci consente di valutare la differenza di probabilità di una variabile in 2 gruppi.

In allegato al presente elaborato sono stati forniti 2 file csv dei quali mostrerò un estratto.



## Test\_Binomiale:

A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T	U
Variabile	valore_es	test	Gruppo_c	Gruppo_A	probabili	probabilit	p_value	Commento												
Family_hi	0	binomiale Post Pube	Pre Puber	0.883333	0.9318181	0.0266210	La probabilità che la variabile Family_history_of_SCD assum	0	nel gruppo Post Puberal	è pari a 0.8833	mentre la probabilità che									
Family_hi	1	binomiale Post Pube	Pre Puber	0.1166666	0.0681818	0.0266210	La probabilità che la variabile Family_history_of_SCD assum	1	nel gruppo Post Puberal	è pari a 0.1167	mentre la probabilità che									
Family_hi	0	binomiale Post Pube	Pre Puber	0.7555555	0.8727272	2,00E+07	La probabilità che la variabile Family_history_of_CD assum	0	nel gruppo Post Puberal	è pari a 0.7556	mentre la probabilità che									
Family_hi	1	binomiale Post Pube	Pre Puber	0.2444444	0.1272727	2,00E+07	La probabilità che la variabile Family_history_of_CD assum	1	nel gruppo Post Puberal	è pari a 0.2444	mentre la probabilità che									
alcohol_u	0	binomiale Post Pube	Pre Puber	0.5111111	0.6727272	1,41E+08	La probabilità che la variabile alcohol_use assum	0	nel gruppo Post Puberal	è pari a 0.5111	mentre la probabilità che assum									
alcohol_u	1	binomiale Post Pube	Pre Puber	0.4888888	0.3272727	1,41E+08	La probabilità che la variabile alcohol_use assum	1	nel gruppo Post Puberal	è pari a 0.4889	mentre la probabilità che assum									
other_dru	0	binomiale Post Pube	Pre Puber	0.9555555	0.9818181	0.0688310	La probabilità che la variabile other_drugs assum	0	nel gruppo Post Puberal	è pari a 0.9556	mentre la probabilità che assum									
other_dru	1	binomiale Post Pube	Pre Puber	0.0444444	0.0181818	0.0688310	La probabilità che la variabile other_drugs assum	1	nel gruppo Post Puberal	è pari a 0.0444	mentre la probabilità che assum									
cigarettes	1	binomiale Post Pube	Pre Puber	0.6333333	0.7545454	0.0001479	La probabilità che la variabile cigarettes assum	1	nel gruppo Post Puberal	è pari a 0.6333	mentre la probabilità che assum									
cigarettes	0	binomiale Post Pube	Pre Puber	0.3666666	0.2454545	0.0001479	La probabilità che la variabile cigarettes assum	0	nel gruppo Post Puberal	è pari a 0.3667	mentre la probabilità che assum									
Syncope_	0	binomiale Post Pube	Pre Puber	0.9944444		1	0.6391843	La probabilità che la variabile Syncope_at_rest assum	0	nel gruppo Post Puberal	è pari a 0.9944	mentre la probabilità che ass								
Syncope_	1	binomiale Post Pube	Pre Puber	0.0055555		0	0.6391843	La probabilità che la variabile Syncope_at_rest assum	1	nel gruppo Post Puberal	è pari a 0.0056	mentre la probabilità che ass								

Variabile	→	Variabile utilizzata per il test
Valore esaminato	→	modalità della variabile considerata nel test
Test	→	tipologia di test eseguito
Gruppo Check	→	Pazienti del gruppo Post Pubertal (età>16)
Gruppo Analisi	→	Pazienti del gruppo Pubertal (età<=16)
Probabilità check	→	Probabilità dei verificarsi del valore esaminato nel gruppo Post Pubertal
Probabilità analisi	→	Probabilità dei verificarsi del valore esaminato nel gruppo Pubertal
p-value	→	Probabilità di sbagliare rifiutando l'ipotesi nulla
Commento	→	Ipotesi di commento basato sulla totalità delle informazioni

## Test Chiadrato con correzione di Fisher

Variabile	test	p_value	Cp_value	Fodds_ratio	Commento
Family_hi	Chiadrato	0.1310150	0.1137882	0.554844727355402	Con un p-value 0.13102 per il test Chi-quadro e 0.11379 per la correzione di Fisher non possiamo rifiutare l'ipotesi nulla, perciò la differenza di distribuz
Family_hi	Chiadrato	0.0036869	0.0026489	0.451670634958383	Con un p-value 0.00369 per il test Chi-quadro e 0.00265 per la correzione di Fisher rifiutiamo l'ipotesi nulla, perciò la differenza di distribuzione della ve
alcohol_u	Chiadrato	0.0014734	0.0014365	0.509475978673441	Con un p-value 0.00147 per il test Chi-quadro e 0.00144 per la correzione di Fisher rifiutiamo l'ipotesi nulla, perciò la differenza di distribuzione della ve
other_dru	Chiadrato	0.2159966	0.1481031	0.399050881794724	Con un p-value 0.216 per il test Chi-quadro e 0.1481 per la correzione di Fisher non possiamo rifiutare l'ipotesi nulla, perciò la differenza di distribuzione
cigarettes	Chiadrato	0.0116640	0.0114949	0.56269815351803	Con un p-value 0.01166 per il test Chi-quadro e 0.01149 per la correzione di Fisher rifiutiamo l'ipotesi nulla, perciò la differenza di distribuzione della ve
Syncope_	Chiadrato	0.9198444	0.45		0 Con un p-value 0.91984 per il test Chi-quadro e 0.45 per la correzione di Fisher non possiamo rifiutare l'ipotesi nulla, perciò la differenza di distribuzion
Palpitation	Chiadrato	0.1688546	0.1256918	189.708.681.359.461	Con un p-value 0.16885 per il test Chi-quadro e 0.12569 per la correzione di Fisher non possiamo rifiutare l'ipotesi nulla, perciò la differenza di distribuz

Variabile	→	Variabile utilizzata per il test
Test	→	tipologia di test eseguito
p-value_Chi_quadrate	→	Probabilità di sbagliare rifiutando l'ipotesi nulla
p-value_Fisher	→	Probabilità di sbagliare rifiutando l'ipotesi nulla
odds_ratio	→	Il rapporto tra il numero di volte in cui l'evento si verifica e il numero di volte in cui l'evento non si verifica.
Commento	→	Ipotesi di commento basato sulla totalità delle informazioni

In entrambi i file il secondo foglio riporta la tabella da condividere.