

RAPPRESENTAZIONI GRAFICHE



SAPIENZA
UNIVERSITÀ DI ROMA

annarita.vestri@uniroma1.it

Rappresentazioni grafiche

- Diagrammi a barre (bar chart)
- Torte (pie chart)
- Diagrammi cartesiani
- Istogrammi di frequenza
- Box plots
- scatterplots

I grafici sono più veloci e più semplici da interpretare di una tabella.

Soddisfano ad almeno tre esigenze:

- dare una visione d'insieme efficace della serie di numeri che rappresentano
- rendere possibile il confronto tra fenomeni diversi, al fine di scoprire delle relazioni tra loro
- aumentare l'intelligibilità dei dati e delle loro elaborazioni, facilitandone una lettura di sintesi

I grafici sono molto usati anche nei media e sono uno strumento di grande forza comunicativa, pertanto è utile imparare ad usarli (leggerli e interpretarli, ma anche costruirli, correttamente).

QUALITATIVI

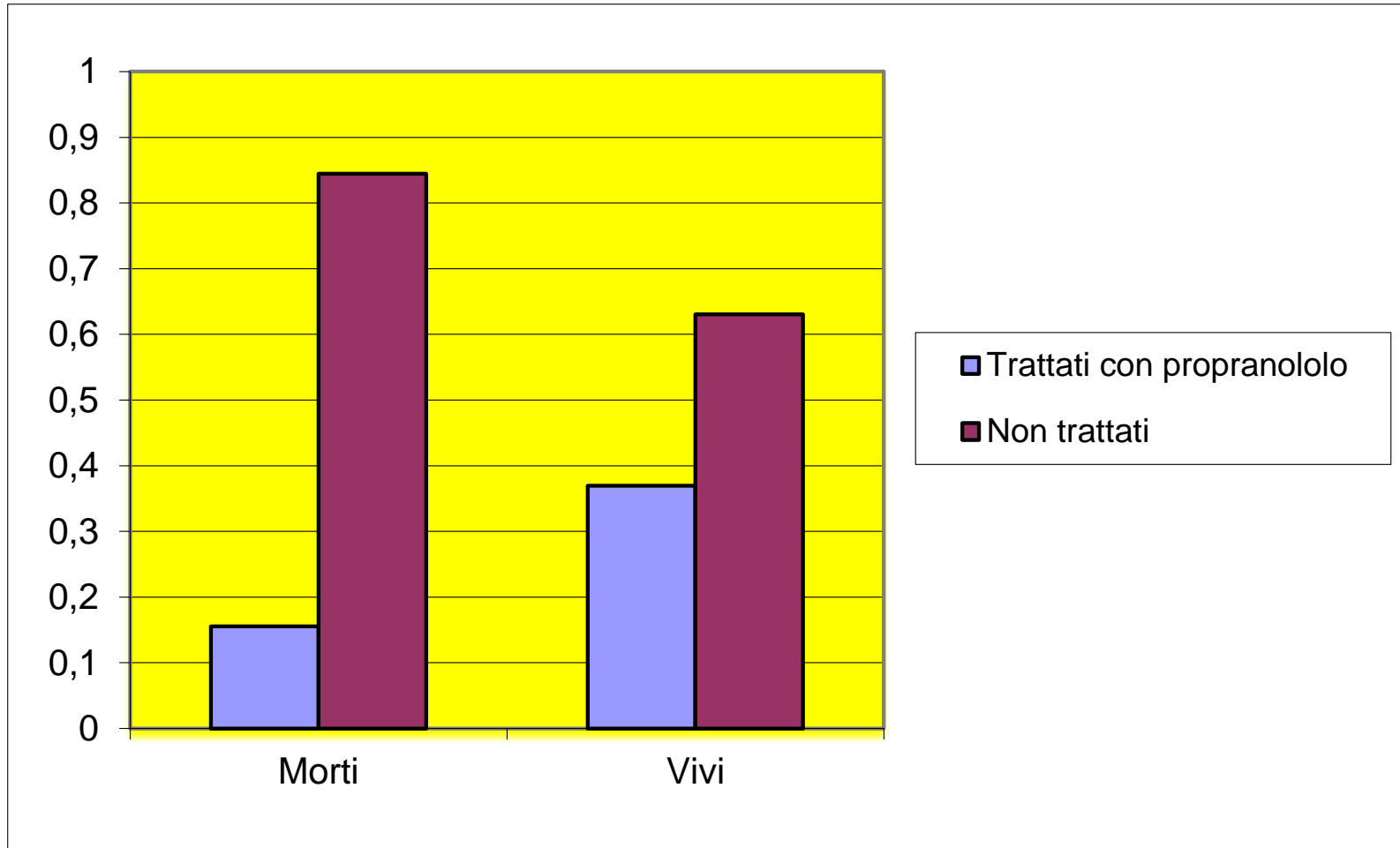
SCALA DICOTOMICA: dati più semplici consistono in osservazioni non ordinate dicotomiche o del tipo "tutto o nulla"; cioè: il paziente vive o muore, ha o non ha un particolare attributo.

Stato a 28 giorni dal ricovero	Trattati con propranololo	Non trattati
Morti	7	17
Vivi	38	29
Totale	45	46
% sopravvivenza	84%	63%

Con i dati espressi in scala nominale, l'ovvia ed intuitiva misura sintetica di descrizione è rappresentata dalla proporzione o percentuale dei soggetti che presentano l'attributo.

Lo scopo è di trarre una conclusione relativa all'effetto del farmaco, se usato in una vasta popolazione di pazienti affetti da infarto miocardico.

Bar chart



QUALITATIVI

Esempio:

Si consideri la variabile *esito di un esperimento di ototossicità* condotto su 120 cavie trattate, per due settimane consecutive, con un antibiotico che ha sia ototossicità sia tossicità generale. L'insieme delle coppie "**modalità, frequenza di comparsa**" viene anche indicato con il nome di **distribuzione** della variabile *esito*.

Rappresentazione tabellare di una serie qualitativa

Distribuzione degli esiti di un esperimento di ototossicità.

MODALITÀ	FREQUENZE	
	assolute	<i>relative</i>
morte nella 1a settimana	18	<i>0.15</i>
morte nella 2a settimana	12	<i>0.10</i>
sopravvissute e otolese	36	<i>0.30</i>
sopravvissute e non otolese	54	<i>0.45</i>

una frequenza relativa è il rapporto tra la frequenza assoluta con cui si manifesta una modalità e la numerosità totale del campione

GRAFICI PER SERIE QUALITATIVE

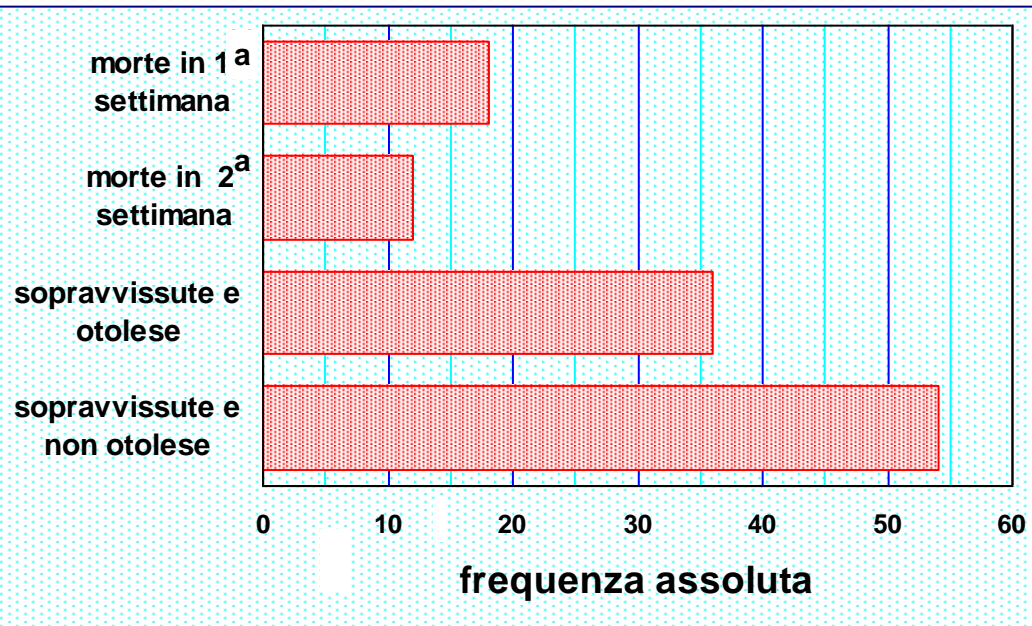
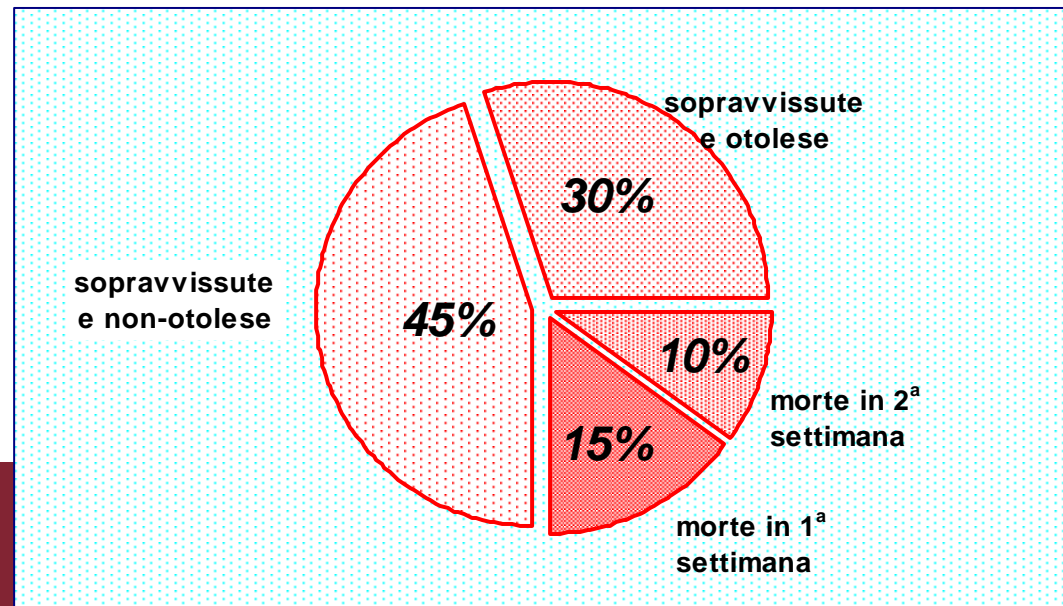


Diagramma a barre orizzontali

Diagramma areolare (o a torta)



QUALITATIVI

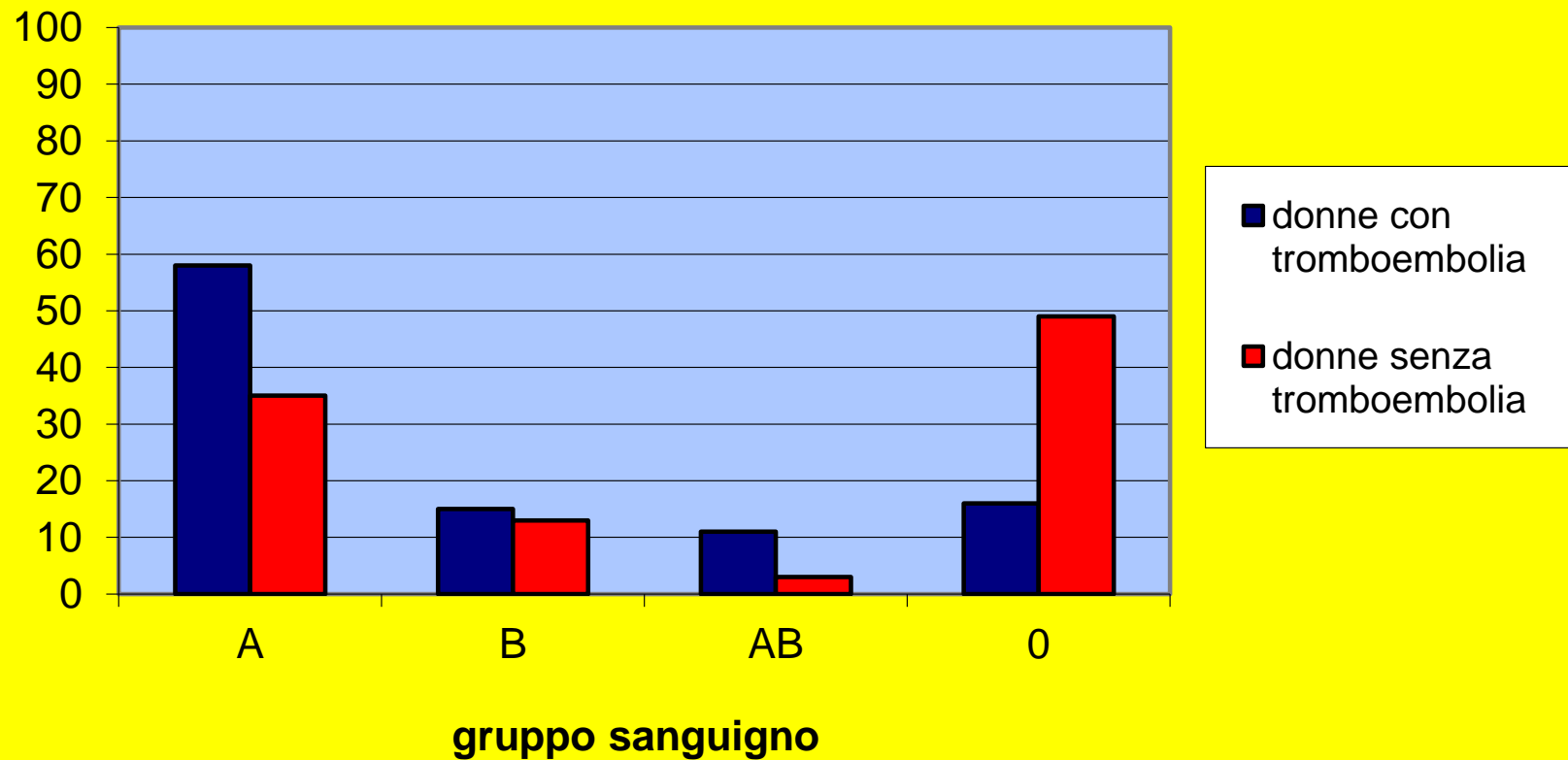
SCALA POLICOTOMICA: Non necessariamente una scala nominale deve essere dicotomica; spesso vi sono più di due alternative o criteri di classificazione. Per esempio i gruppi sanguigni illustrano una scala policotomica non ordinata.

Gruppo sanguigno	donne con tromboembolia		donne senza tromboembolia	
	n.	%	n	%
A	32	58	51	35
B	8	15	19	13
AB	6	11	5	3
0	9	16	70	49
Totale	55	100	145	100

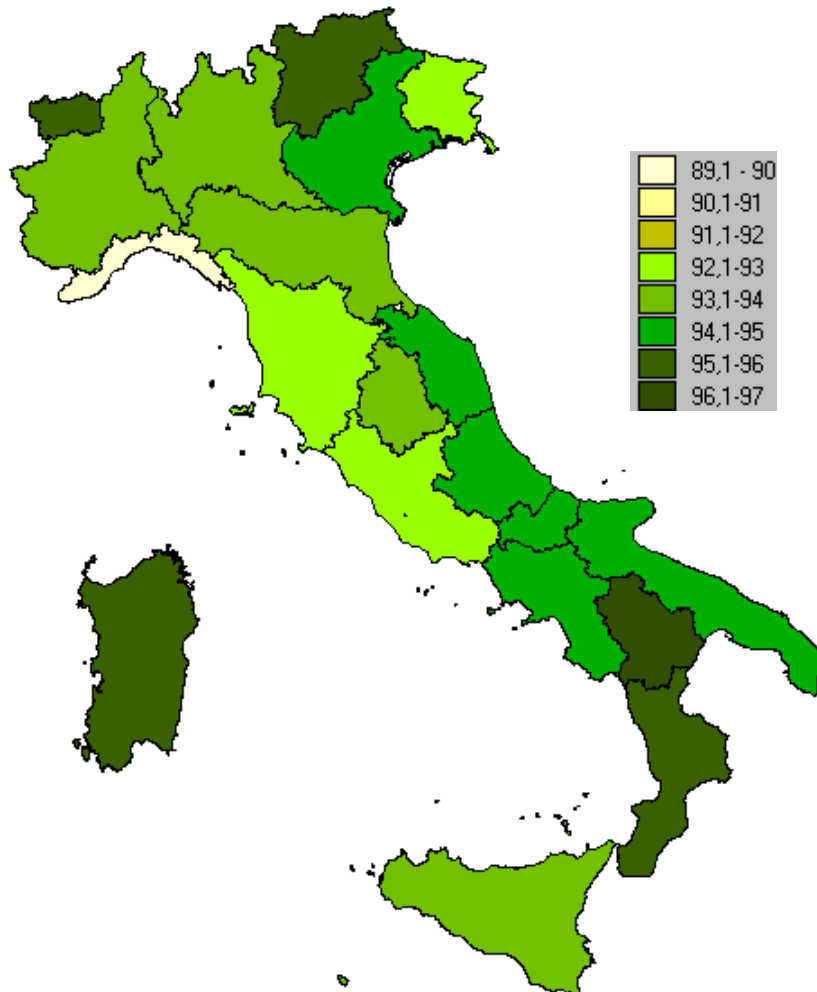
L'osservazione delle percentuali in funzione del gruppo sanguigno suggerisce un deficit del sangue di tipo 0 nel gruppo affetto da tromboembolia (16%) in confronto ai controlli (49%).

Bar chart

distribuzione del gruppo sanguigno in 200 donne



Rappresentazioni grafiche: Cartogramma



Cartogramma del rapporto di mascolinità
nelle regioni italiane - dati Censimento Istat
2001

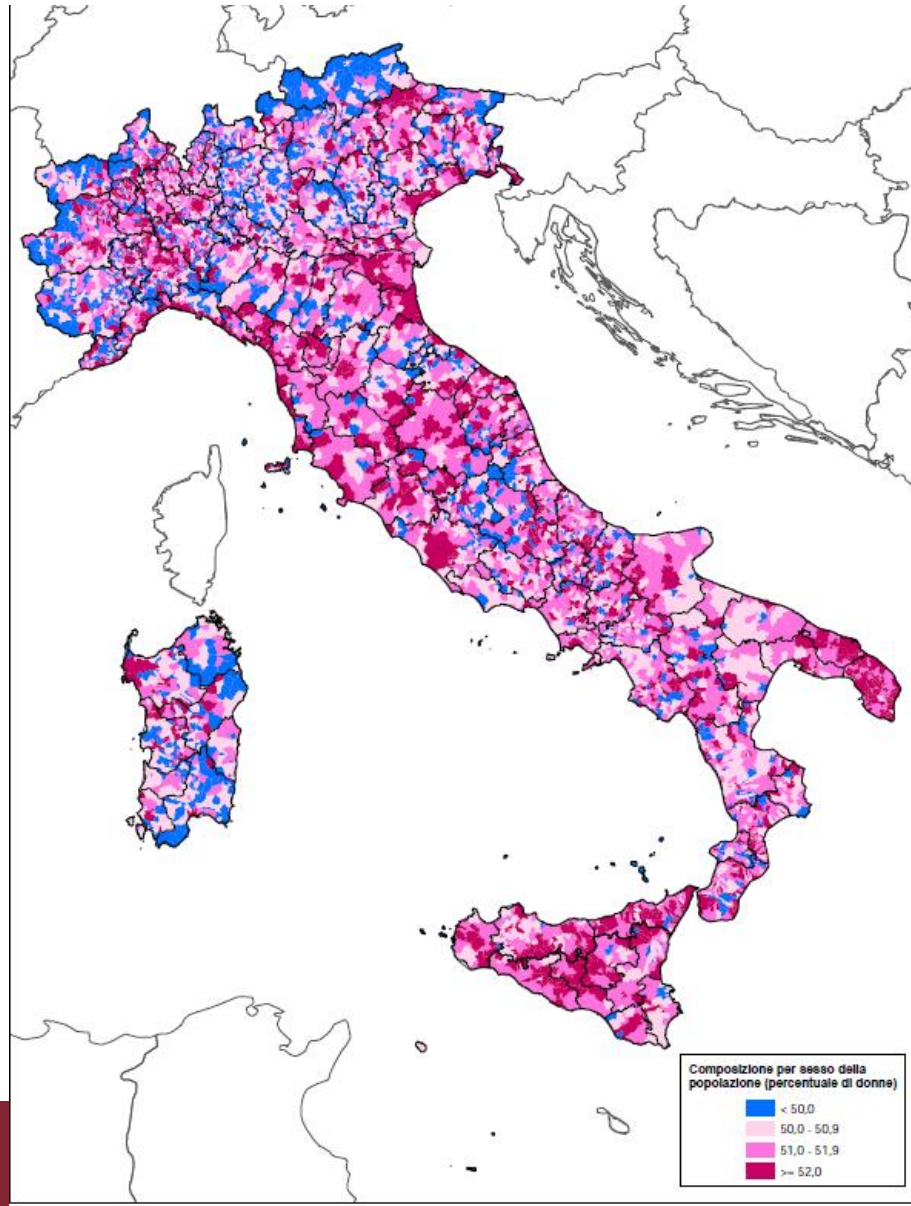
Mappe
geografiche

Suddivisione
dei dati in
classi

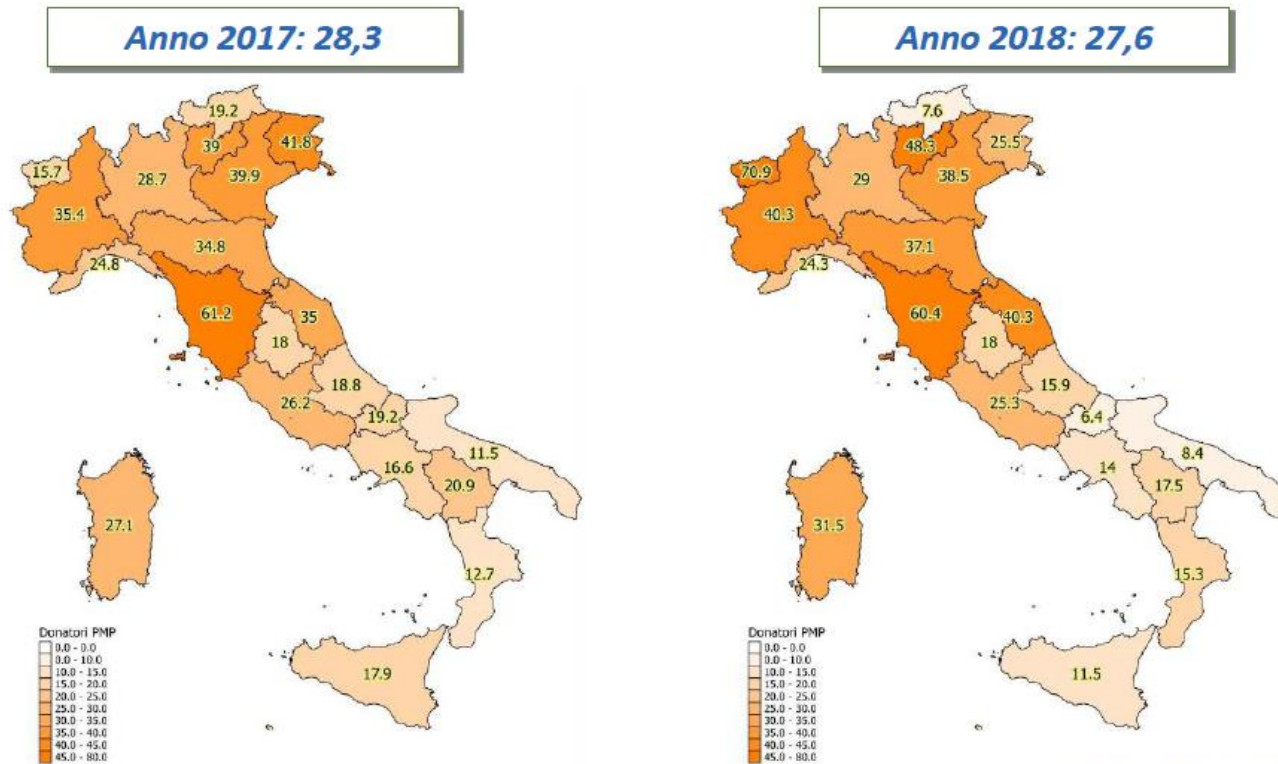
Creazione di una
scala cromatica

Colorazione
delle zone

Cartogramma 5 - Composizione per sesso della popolazione (percentuale di donne)



CONFRONTO DONATORI PMP - 2017 VS 2018*



* dati preliminari al 30 novembre 2018

Questo tipo di rappresentazione grafica sfrutta una carta geografica per evidenziare, nelle zone di interesse, le informazioni analizzate.

Talvolta possono essere inseriti barre di istogrammi in corrispondenza delle zone geografiche a cui si riferiscono

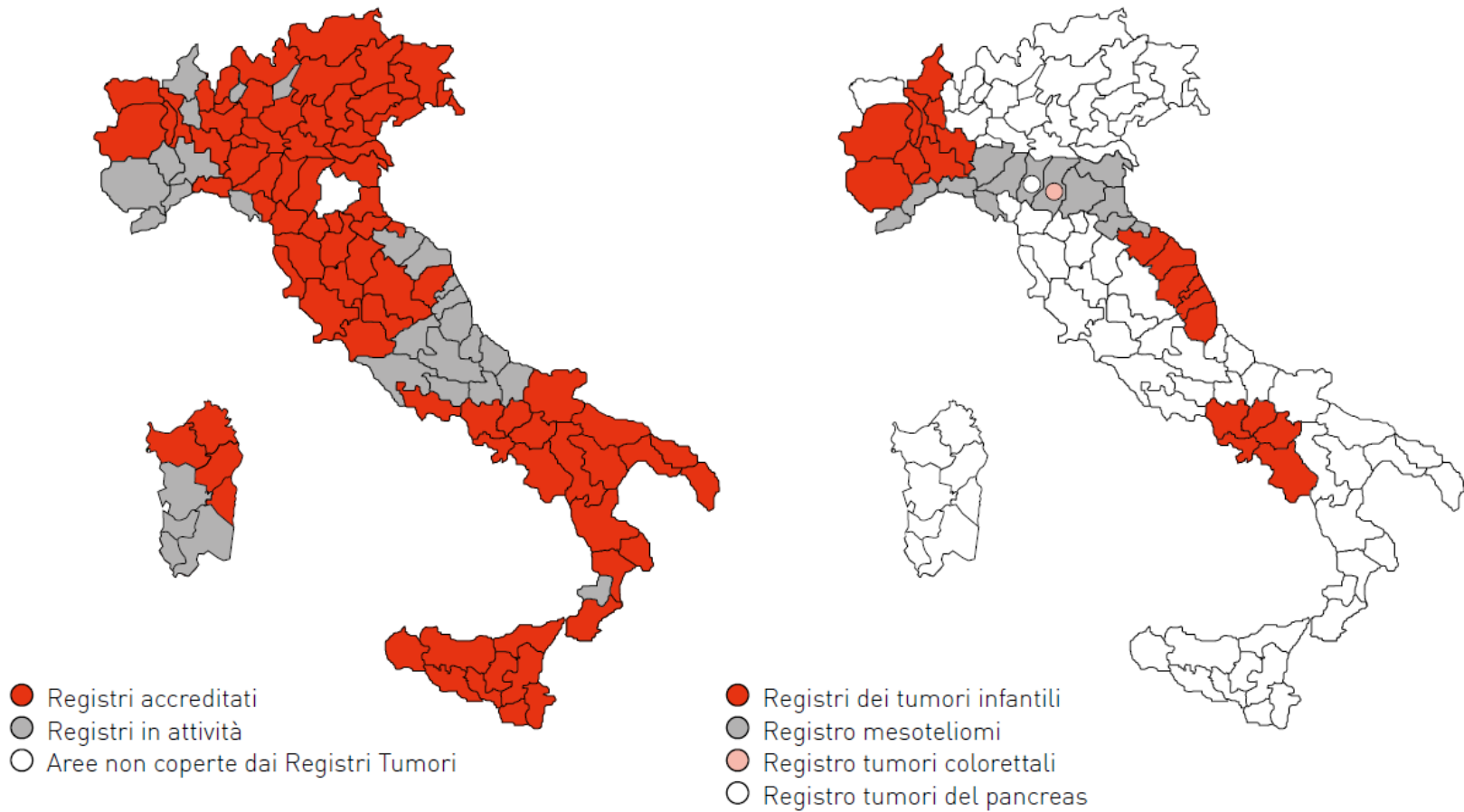


FIGURA 36. I Registri Tumori in Italia

Tutti i tumori (esclusi i tumori non melanomatosi della cute)
(ICD-10 C00-43, C45-96, D09.9, 30.3,41.4)

All sites except non-melanoma skin cancer

Complete prevalence by sex, age and macro-area 01.01.2006 (proportion per 100 000)					
Macro-area	Age class				
	0-44	45-59	60-74	75+	All ages
♂ MALE					
North West	565	2 820	10 817	20 086	4 325
North East	508	2 666	10 488	20 050	3 988
Centre	498	2 484	9 355	18 064	3 813
South	466	2 176	7 625	13 451	2 502
♀ FEMALE					
North West	721	5 030	10 385	13 548	5 158
North East	726	5 059	10 392	13 598	4 875
Centre	722	4 733	9 413	12 412	4 525
South	665	3 816	7 492	9 160	2 988
♂ ♀ MALE & FEMALE					
North West	641	3 959	10 580	15 799	4 760
North East	614	3 863	10 437	15 893	4 442
Centre	608	3 631	9 386	14 560	4 180
South	564	3 008	7 554	10 830	2 751

Fonte AIRTUM 2010

POPOLAZIONE RESIDENTE IN ITALIA, 1 GENNAIO 2006


58 751 711

NORTH WEST
4 760
per 100 000

NORTH EAST
4 442
per 100 000

CENTRE
4 180
per 100 000

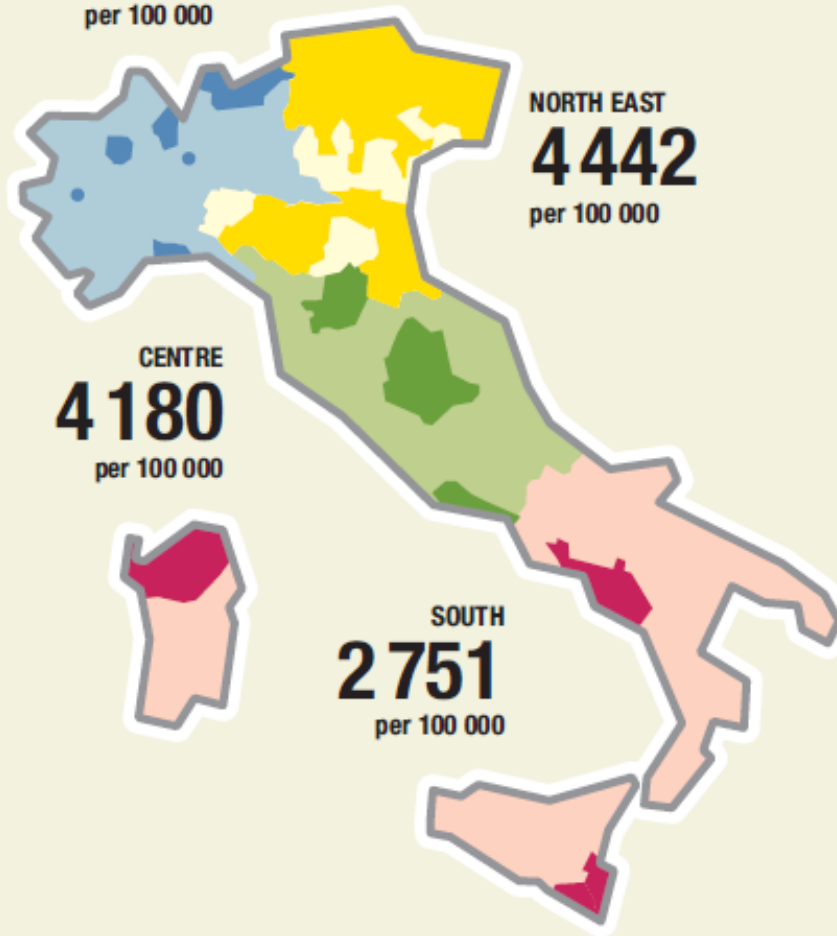
SOUTH
2 751
per 100 000

 **NORTH WEST**
CANCER REGISTRY AREAS

 **CENTRE**
CANCER REGISTRY AREAS

 **NORTH EAST**
CANCER REGISTRY AREAS

 **SOUTH**
CANCER REGISTRY AREAS



QUALITATIVO ORDINABILE

- Un paziente arruolato in uno studio di efficacia di un analgesico, può patire una qualunque intensità di dolore **senza potervi associare una quantità**. Egli, però, può classificare l'intensità del dolore nella **scala ordinale**:

nulla < lieve < moderata < forte

- Alle modalità si associa un **punteggio**
 - (**es.:** nulla=0, lieve=1, moderata=2, forte=3)
 - **che non ha significato quantitativo:**
 - 2 **non** è il doppio di 1,
 - 3 **non** è il triplo di 1,
 - la differenza tra 2 e 1 **non** è uguale a quella tra 3 e 2.

SCALA ORDINALE

- Una **variabile quantitativa** può anche essere misurata su **scala ordinale**.
 - **Esempi:** L'età dei pazienti può essere espressa nella scala ordinale
bambini < ragazzi < adulti < anziani

SCALA ORDINALE

Il numero di episodi anginosi alla settimana può essere espresso in scala ordinale

assenti < rari < frequenti

o, addirittura, in modo binario:

NO < SI

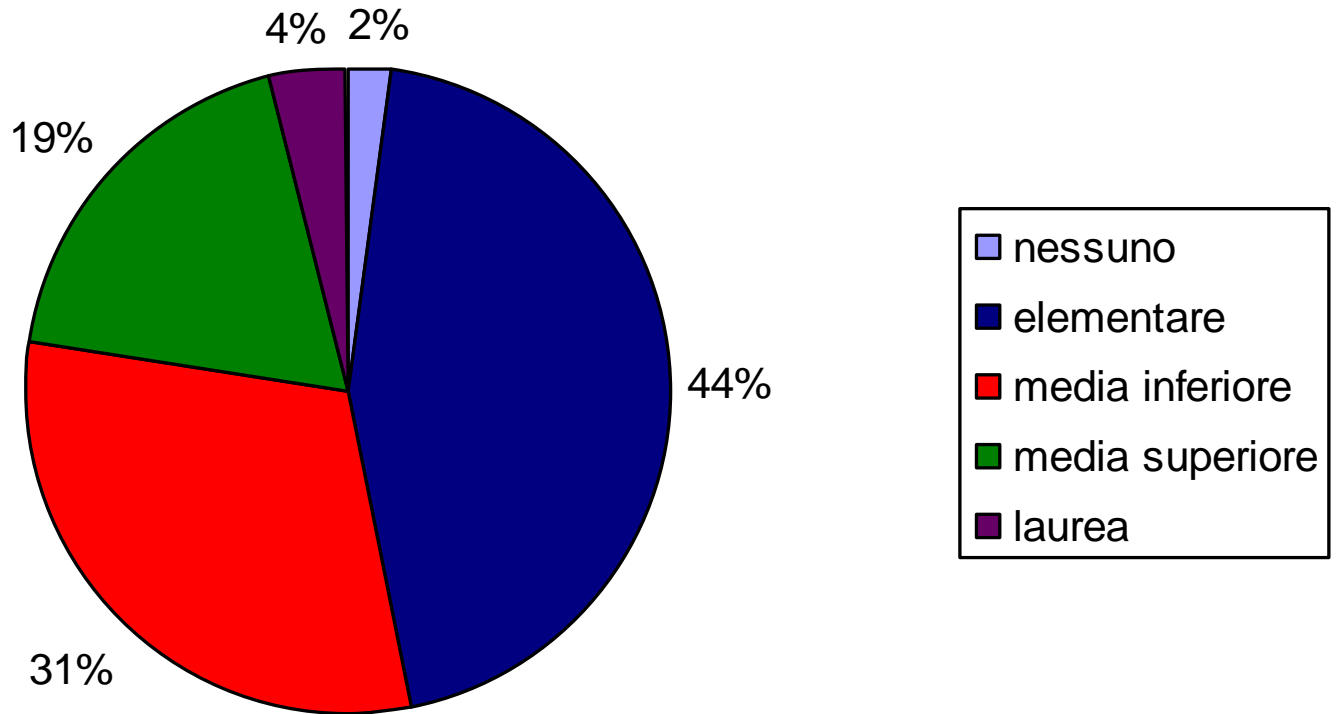
QUALITATIVO ORDINABILE

- **Distribuzione di frequenza** della variabile "Titolo di studio" (L'Italia in cifre, ISTAT 1996).

Titolo di studio	n° soggetti
nessuno	1.123
elementare	23.962
media inferiore	16.418
media superiore	9.947
laurea	2.032
Totale	53.482

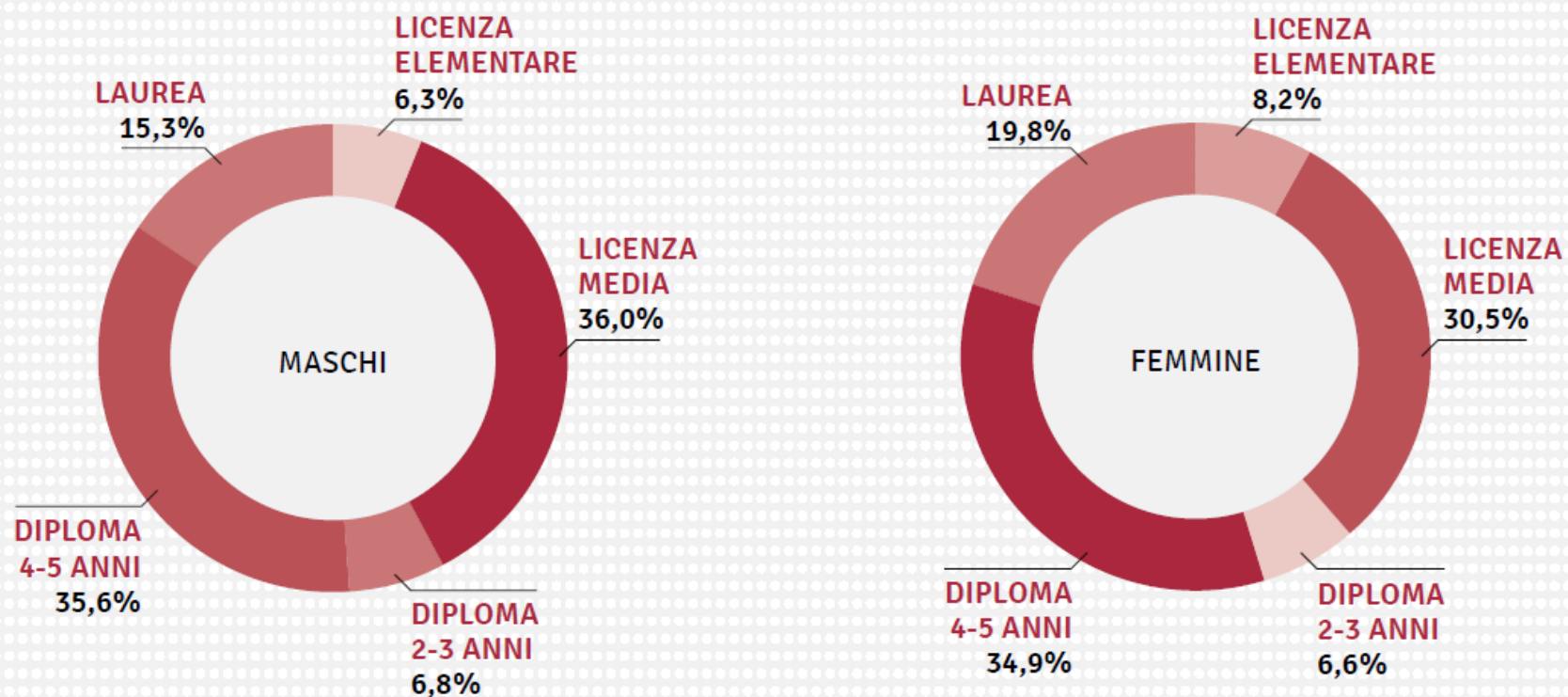
Pie chart

distribuzione secondo il titolo di studio ISTAT 1996



POPOLAZIONE DI 25-64 ANNI PER TITOLO DI STUDIO E SESSO

Anno 2015, composizioni percentuali



QUANTITATIVO DISCRETO

Morti per calcio di cavallo in 200 reggimenti di cavalleria prussiani. (Bortkiewicz. 1898)

numero di morti in un reggimento	frequenze semplici		frequenze cumulate	
	assolute	<i>relative</i>	assolute	<i>relative</i>
0	109	54,5	109	54,5
1	65	32,5	174	87,0
2	22	11,0	196	98,0
3	3	1,5	199	99,5
4	1	0,5	200	100,0
5+	0	0	200	100,0

Nota: la **frequenza cumulata assoluta** in corrispondenza di un valore x^* indica il numero di volte che la variabile x ha assunto valori pari o inferiori a x^* .

Ad esempio, la frequenza cumulata assoluta per il valore 2 è data dalla somma $109+65+22= 196$: ben 196 dei 200 reggimenti (pari al 98%, **frequenza cumulata relativa**) hanno presentato un numero di morti per calcio di cavallo inferiore oppure uguale a due.

RAPPRESENTAZIONE GRAFICA DI SERIE QUANTITATIVE

Diagramma ad aghi per frequenze semplici

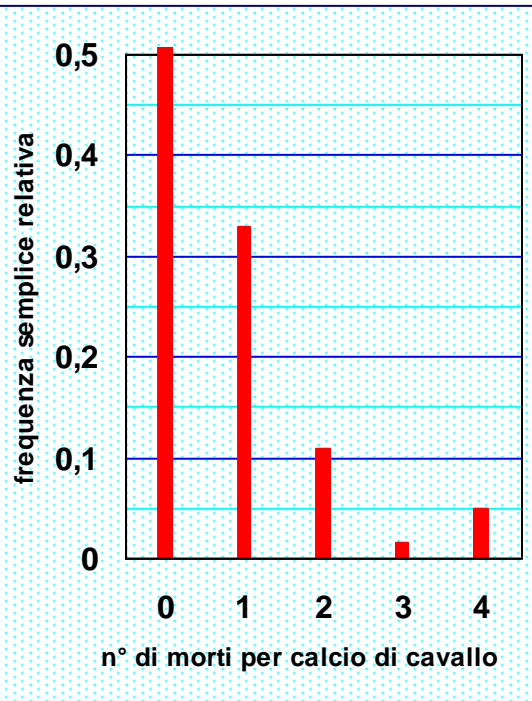


Diagramma a gradini per frequenze cumulate

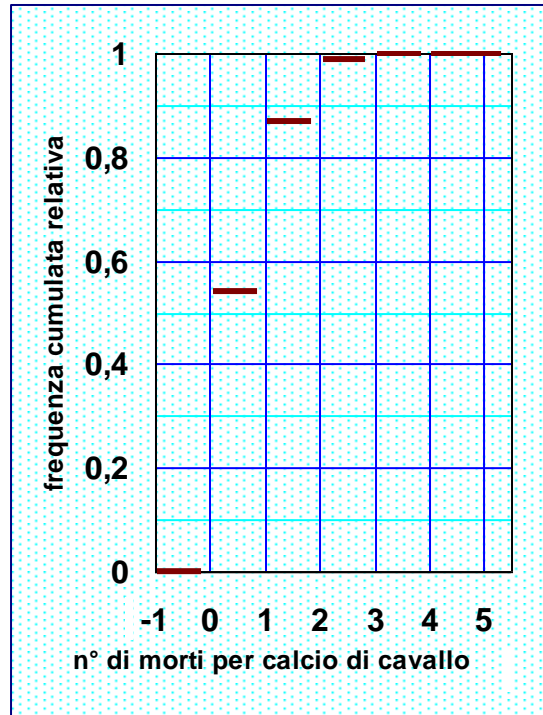
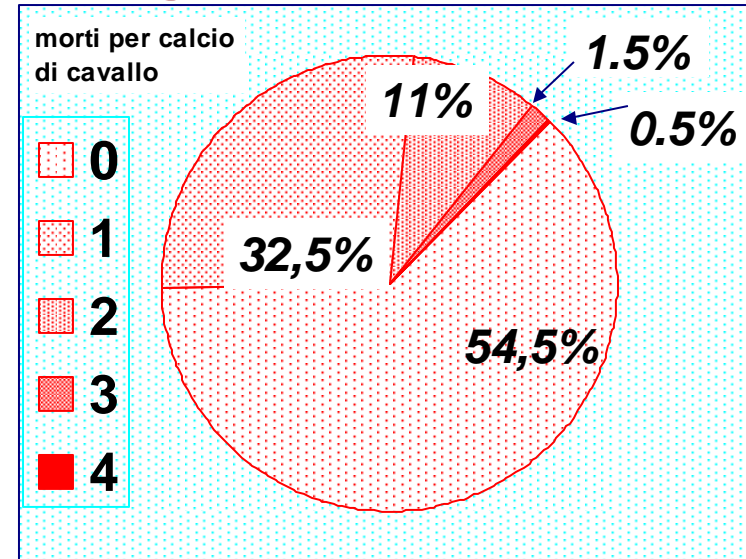


Diagramma areolare



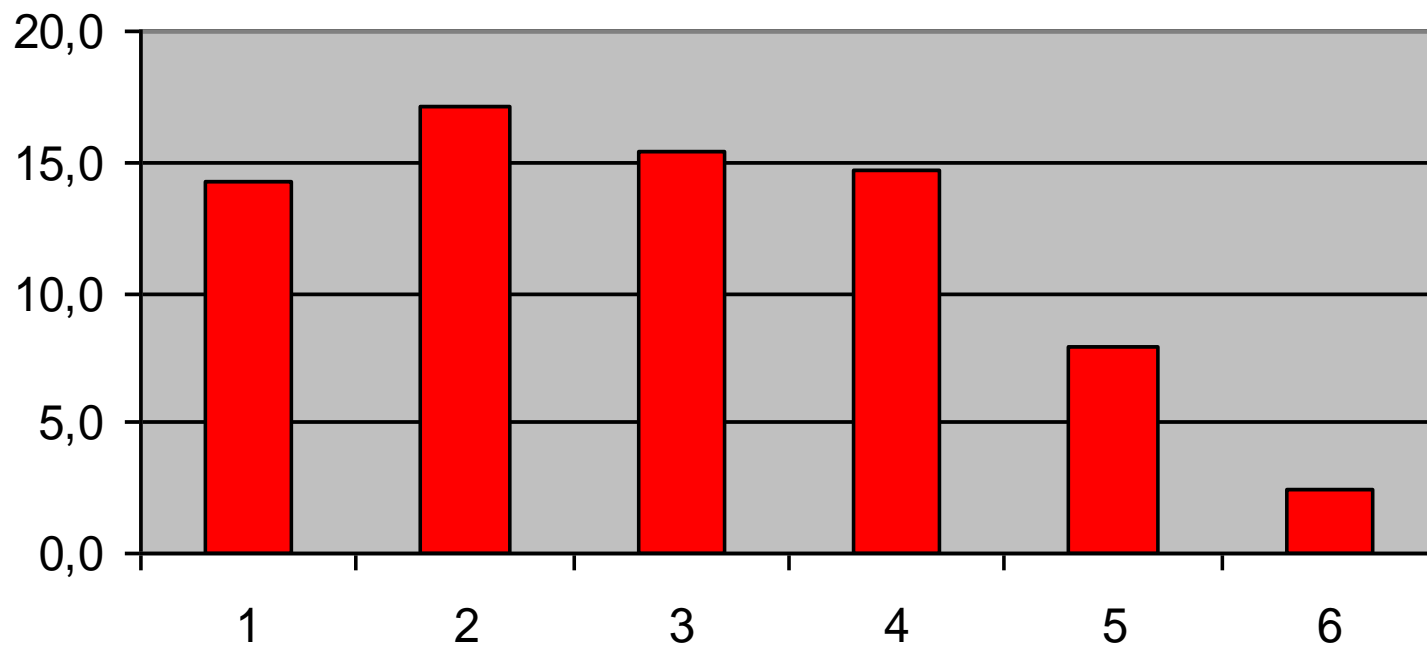
QUANTITATIVO DISCRETO

- **Distribuzione di frequenza** del numero di componenti per famiglia (L'Italia in cifre, ISTAT 1996).

Componenti modalità	n° famiglie(Frequenze)			
	assolute	relative	cumulate	relative
1	4.101	0.206	4 101	0.206
2	4.917	0.247	9 018	0.453
3	4.419	0.222	13 437	0.675
4	4.220	0.212	17 657	0.887
5	1.572	0.079	19 229	0.966
6	477	0.024	19 706	0.990
o più	203	0.010	19 909	1.000
Totale	19.909	1.000		



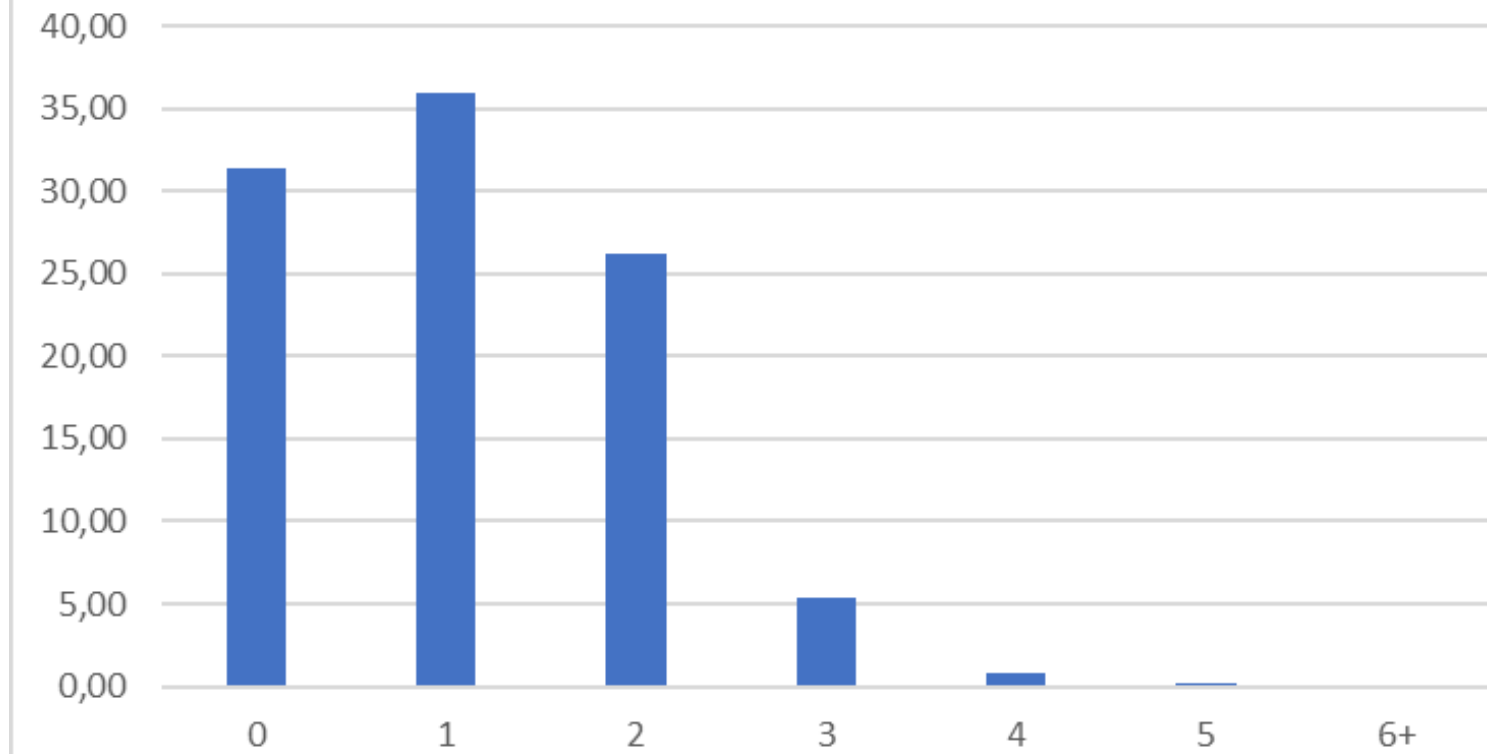
Distribuzione di frequenza del numero di componenti per famiglia (L'Italia in cifre, ISTAT 1996).



Distribuzione di frequenza del numero di figli per famiglia (L'Italia in cifre, ISTAT 2011).

figli	N assoluto	%
0	5 230 296	31,41543
1	5 984 691	35,94665
2	4 368 312	26,23798
3	901 874	5,417047
4	133 281	0,800544
5	22 712	0,136418
6 e +	7 647	0,045931
totale	16648813	100

numero dei figli per famiglia



Serie storica

Tavola 1 - Popolazione italiana residente alle date dei censimenti generali, riportata ai confini attuali - Anni 1861-2001 (migliaia di unità)

ANNI	Popolazione residente
1861	22.176
1871	27.300
1881	28.952
1901	32.963
1911	35.842
1921	39.397
1931	41.043
1936	42.398
1951	47.516
1961	50.624
1971	54.137
1981	56.557
1991	56.778

Fonte: Istituto nazionale di statistica. Censimenti della popolazione.
<http://www.istat.it/Censimenti/index.htm>

Diagramma cartesiano

Figura 1 – Evoluzione della popolazione italiana. Anni 1861-2001 (migliaia di unità)

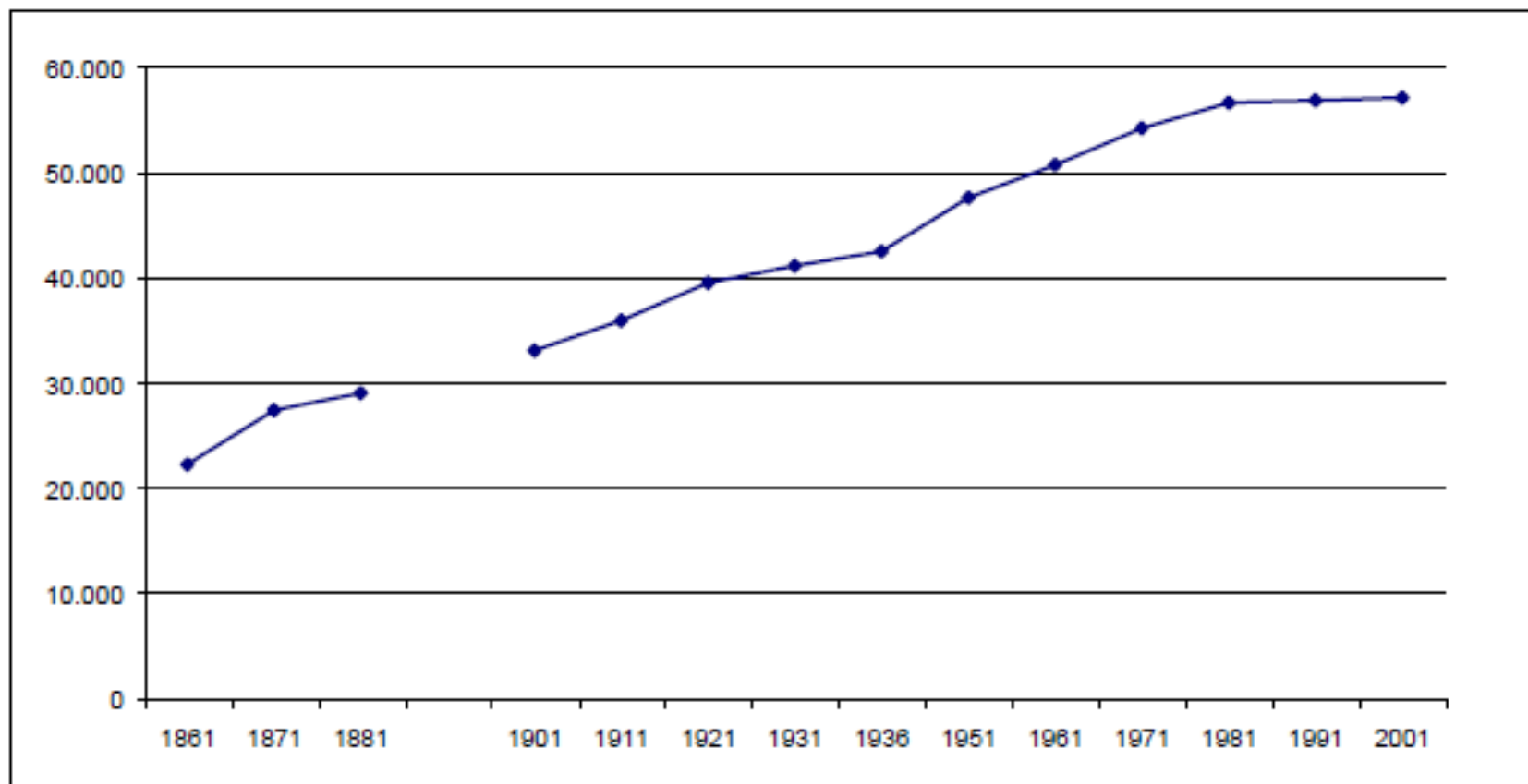
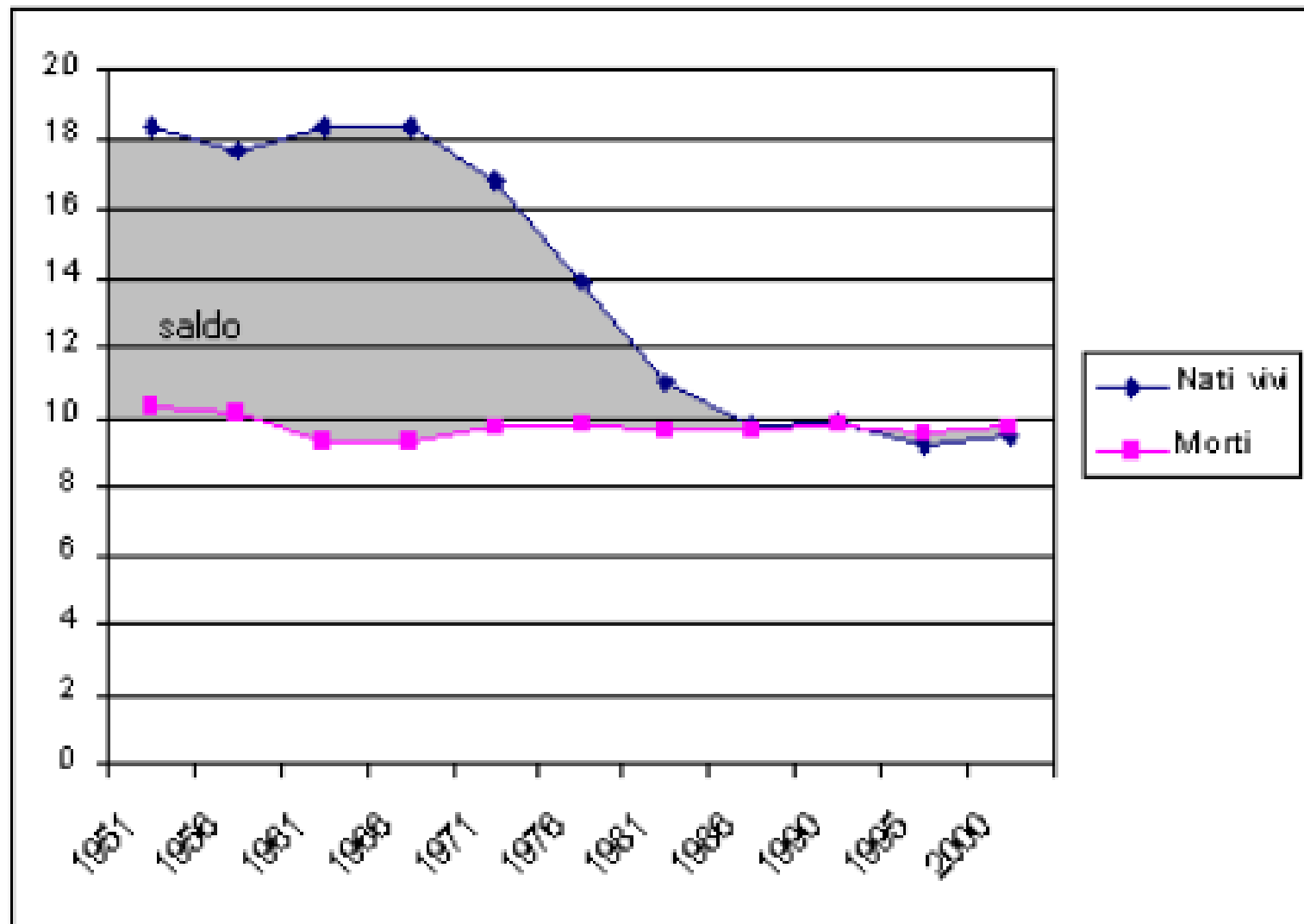


Tavola 3 - Nati vivi, morti e saldo naturale della popolazione italiana in alcuni anni fra il 1951 e il 2000 (per 1.000 abitanti)

ANNI	Nati vivi	Morti	Differenza: saldo naturale
1951	18,4	10,3	8,1
1956	17,7	10,1	7,6
1961	18,4	9,3	9,1
1966	18,4	9,3	9,1
1971	16,8	9,7	7,1
1976	13,9	9,8	4,1
1981	11,0	9,6	1,4
1986	9,8	9,6	0,3
1990	9,9	9,8	0,1
1995	9,2	9,5	-0,3
2000	9,4	9,7	-0,3

Fonte (per gli anni dal 1955 al 1999): Istituto nazionale di statistica. Annuario statistico italiano; (per l'anno 2000) Istituto nazionale di statistica. Popolazione e statistiche demografiche. Indicatori strutturali 2001. <http://demo.istat.it>

Figura 2 - Nati vivi, morti e saldo naturale della popolazione italiana -
Anni 1951 - 2000 (per 1.000 abitanti)



VARIABILI QUANTITATIVE CONTINUE

Si dice **continua** una variabile che può virtualmente assumere un **qualsiasi valore reale**, in un certo ambito.

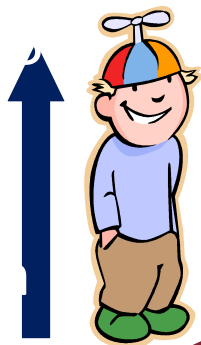
- **Ad esempio**, Achille può essere alto esattamente metri, cioè 1.7724538509... m.
- In pratica, tuttavia, le **misure** di una variabile quantitativa **possono assumere solo certi valori**, in relazione al potere di risoluzione dello strumento di misura.
- **Ad esempio**, l'altezza di Achille è 1.77 m, se misurata con un metro da sarto; è invece 1.772 m se misurata con lo stadio-metro Harpenden.

PESO, STATURA e SESSO delle MATRICOLE di MEDICINA

PESO Kg	STAT. cm	SESSO	PESO Kg	STAT. cm	SESSO	PESO Kg	STAT. cm	SESSO
56	159	F	77	192	M	51	171	F
66	169	F	60	173	F	48	156	F
50	160	F	78	182	M	55	167	F
53	170	F	52	167	F	60	177	M
54	168	F	47.5	164	F	58	170	F
53	161	F	64	166	F	67	167	F
63	172	M	52	160	F	50	172	F
53	170	F	72	184	M	58	169	F
62	161	F	48	169	F	77	179	M
56	163	F	66	170	M	52	162	M
50	160	F	55	172	F	49	160	F
52	170	F	67	177	M	49	165	F
58	173	F	66	170	M	62	178	M
52	167	F	50	160	F	68	174	M
73	178	M	51	167	F	75	181	M
57	166	F	95	193	M	48	167	F
52	165	F	58	160	F	53	160	F
56	171	F	67	178	F	49	167	F
67	175	M	67	175	M	52	165	F
63	182	F	60	160	F	55	155	F
55	169	F	56	165	F	84	188	M
58	165	F	50	165	F	56	170	F
55	175	M	52	170	F	60	171	F
66	176	M	58	172	F	52	176	M
55	164	F	60	170	F	62	180	F
47	160	F	54	166	F			
47	155	F	60	165	F			
63	169	M	74	172	M			
61	177	F	53	173	F			
53	170	F	72	183	M			
55	168	M	52	168	F			
53	162	F	51	164	F			
62	162	F	81	176	M			
45	160	F	50	160	F			
57	167	F	51	171	F			
45	158	F	64	180	F			
53	168	F	82	183	M			
50	160	F	47	156	F			
55	162	F	70	175	M			
70	177	M	58	168	F			
64	178	F	59	173	F			
52	164	F	68	165	F			
75	175	M	63	177	F			
75	178	M	50	159	F			
70	165	F	65	150	F			
58	167	F	60	170	F			
45	160	F	51	167	F			
50	167	F	75	182	M			
56	156	F	62	170	M			
59	165	F	85	174	M			



**DISTRIBUZIONE DI
FREQUENZA DELLA
VARIABILE
QUANTITATIVA
CONTINUA STATURA**



STATURA	n_i	f_i
Value	Frequency	Percent
150	1	.8
155	2	1.6
156	3	2.4
158	1	.8
159	2	1.6
160	13	10.4
161	2	1.6
162	4	3.2
163	1	.8
164	4	3.2
165	10	8.0
166	3	2.4
167	11	8.8
168	5	4.0
169	5	4.0
170	12	9.6
171	4	3.2
172	5	4.0
173	4	3.2
174	2	1.6
175	5	4.0
176	3	2.4
177	5	4.0
178	5	4.0
179	1	.8
180	2	1.6
181	1	.8
182	3	2.4
183	2	1.6
184	1	.8
188	1	.8
192	1	.8
193	1	.8
Total	125	100.0

TABELLA DI FREQUENZA

statura in classi	Freq. Assoluta	Freq. Relativa (%)
[150-155)	1	$1/125=$ 0,8%
[155-160)	8	$8/125=$ 6,4%
[160-165)	24	$24/125=$ 19,2%
[165-170)	34	27,2%
[170-175)	27	21,6%
[175-180)	19	15,2%
[180-185)	9	7,2%
[185-190)	1	0,8%
[190-195)	2	1,6%
TOTALE	125	100,0%

Istogramma di frequenza

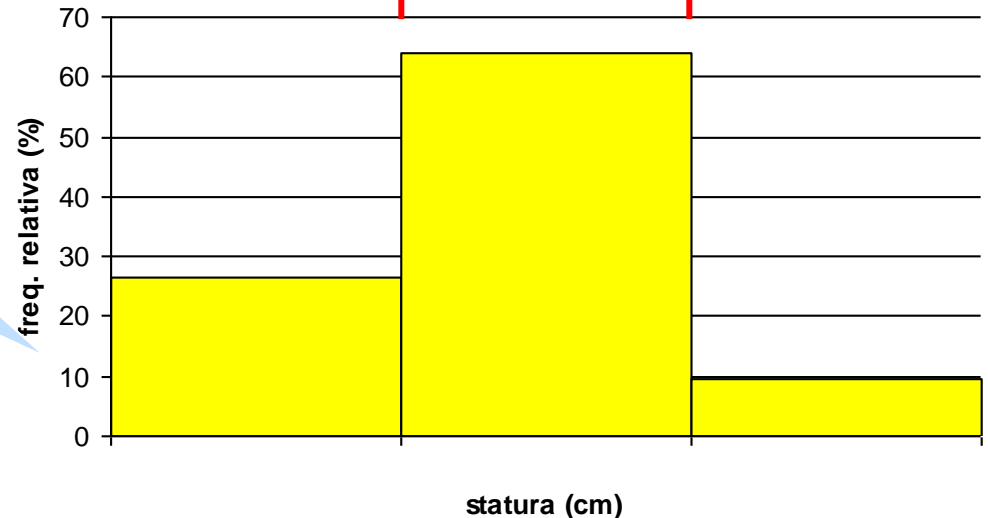
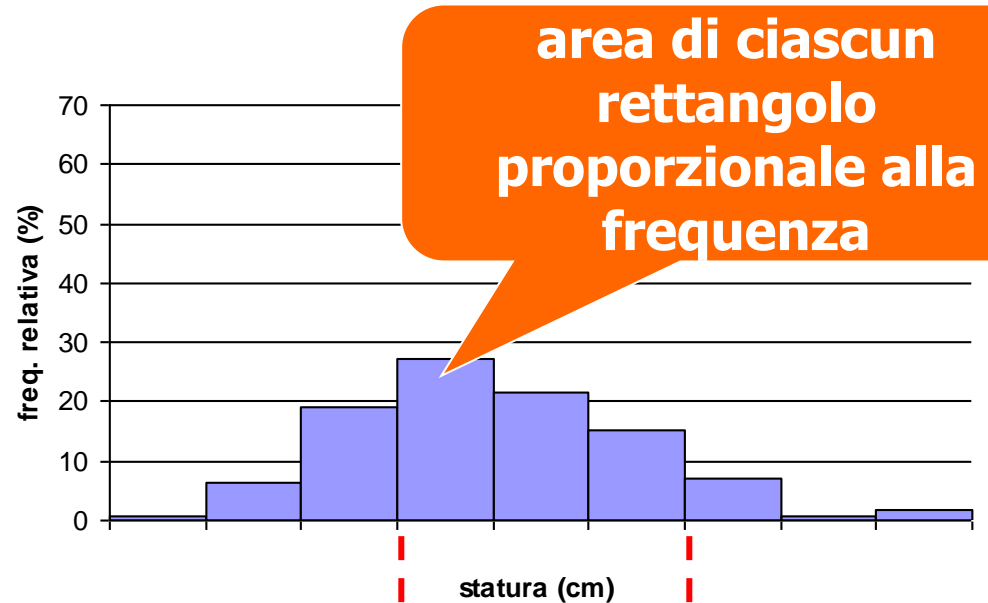
In un **istogramma di frequenza** ad ogni classe è associato un **rettangolo**:

- la **base** del rettangolo è pari all'**ampiezza** di classe
- l'**altezza** del rettangolo è pari alla **densità** di frequenza, ovvero al rapporto tra la frequenza e l'ampiezza di una classe
- l'**area** del rettangolo è per costruzione la **frequenza** (assoluta o relativa) associata alla classe

ISTOGRAMMA A CANNE D'ORGANO

**perdita di
informazione al
diminuire del
numero di intervalli**

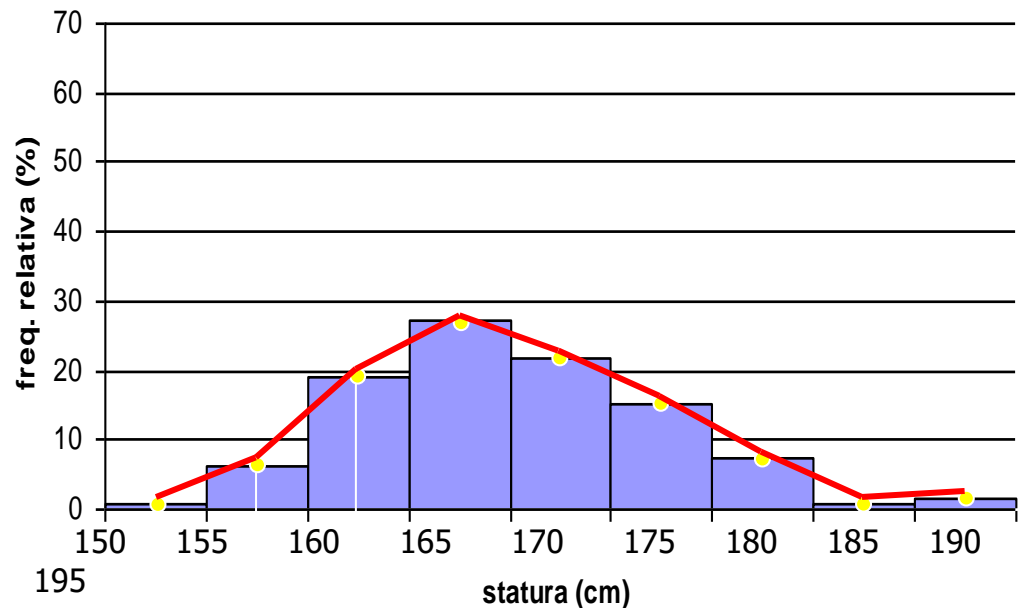
la base dei rettangoli è
proporzionale
all'ampiezza delle classi



statura in classi	Freq. Assoluta	Freq. Relativa (%) n_i	Freq. Cumulata Relativa (%)
[150-155)	1	0.8%	0.8%
[155-160)	8	6.4%	7.2%
[160-165)	24	19.2%	26.4%
[165-170)	34	27.2%	53.6%
[170-175)	27	21.6%	75.2%
[175-180)	19	15.2%	90.4%
[180-185)	9	7.2%	97.6%
[185-190)	1	0.8%	98.4%
[190-195)	2	1.6%	100.0%
TOTALE	125	100.0%	

POLIGONO o CURVA DI FREQUENZA

si ottiene unendo i **PUNTI CENTRALI** dei lati superiori dei rettangoli di un istogramma



157.5 162.5

TABELLA DI FREQUENZA

n_i

p_i

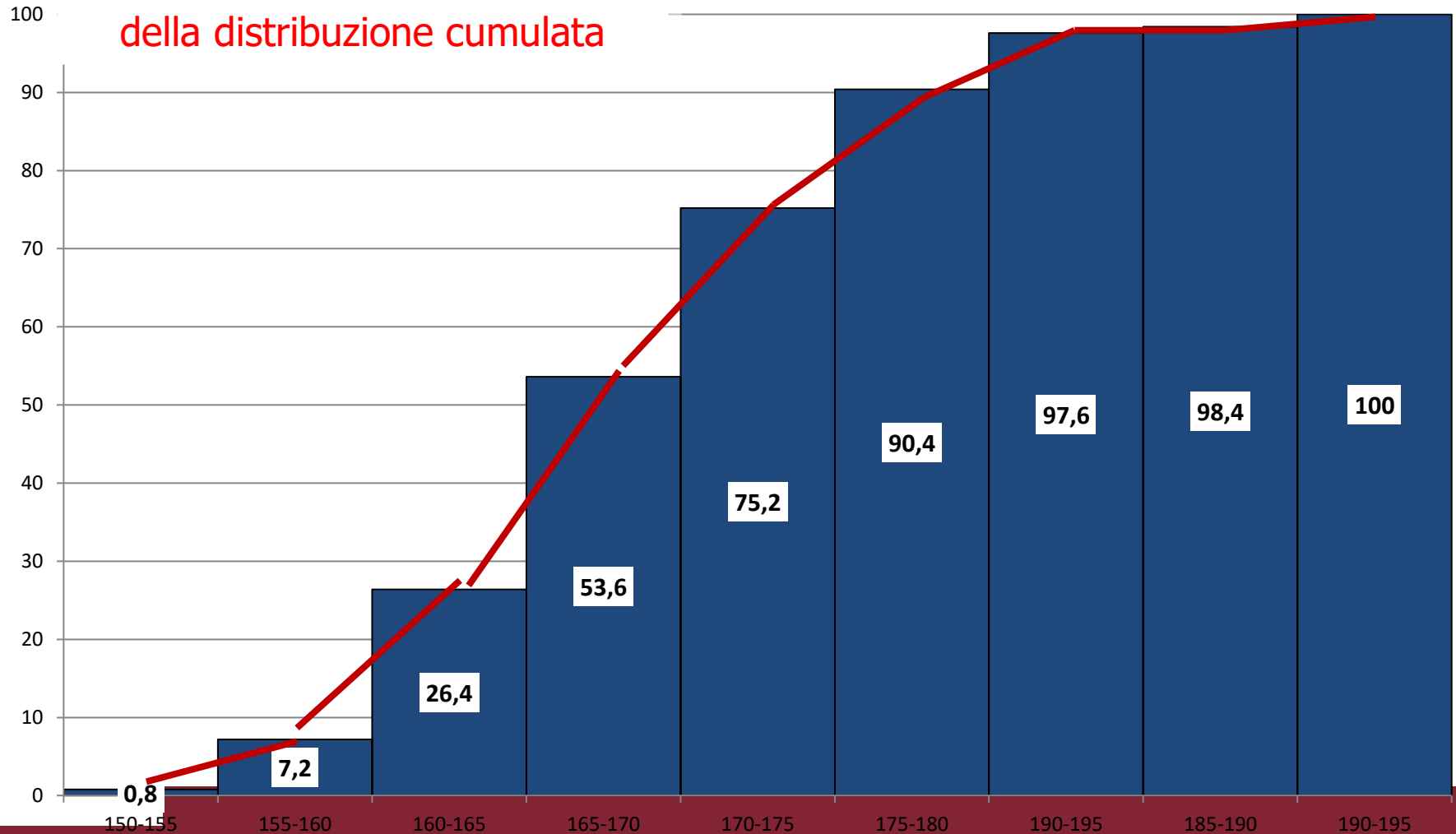
F_i

P_i

statura in classi	<i>Freq. Assoluta</i>	<i>Freq. Relativa (%)</i>	<i>Freq. Assoluta Cumulata</i>	<i>Freq. Relativa Cumulata (%)</i>
[150-155)	1	1/125= 0.8%	1	0.8%
[155-160)	8	8/125= 6.4%	1+8 9	0,8+6,4 7.2%
[160-165)	24	24/125= 19.2%	1+8+24 33	0,8+6,4+19,2 26.4%
[165-170)	34	27.2%	1+8+24+34 67	0,8+6,4+19,2+27,2 53.6%
[170-175)	27	21.6%	94	75.2%
[175-180)	19	15.2%	113	90.4%
[180-185)	9	7.2%	122	97.6%
[185-190)	1	0.8%	123	98.4%
[190-195)	2	1.6%	125	100.0%
TOTALE	125	100.0%	125	

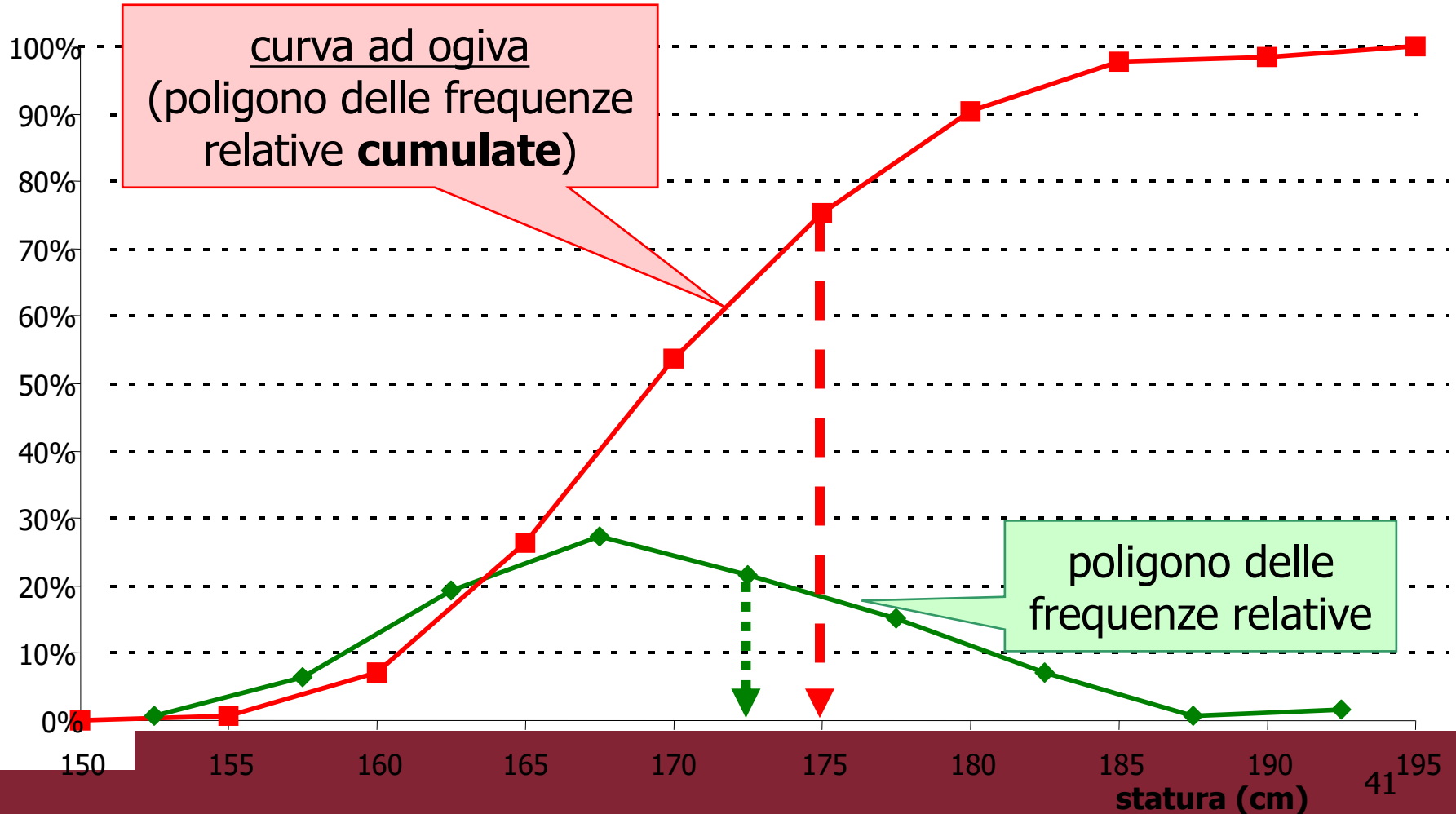
Ogiva di Galton per le frequenze cumulate

L'**OGIVA** si ottiene unendo i punti corrispondenti ai **LIMITI SUPERIORI** di ciascuna classe della distribuzione cumulata



RAPPRESENTAZIONE GRAFICA DELLA DISTRIBUZIONE DI FREQUENZA DI
UNA VARIABILE QUANTITATIVA:

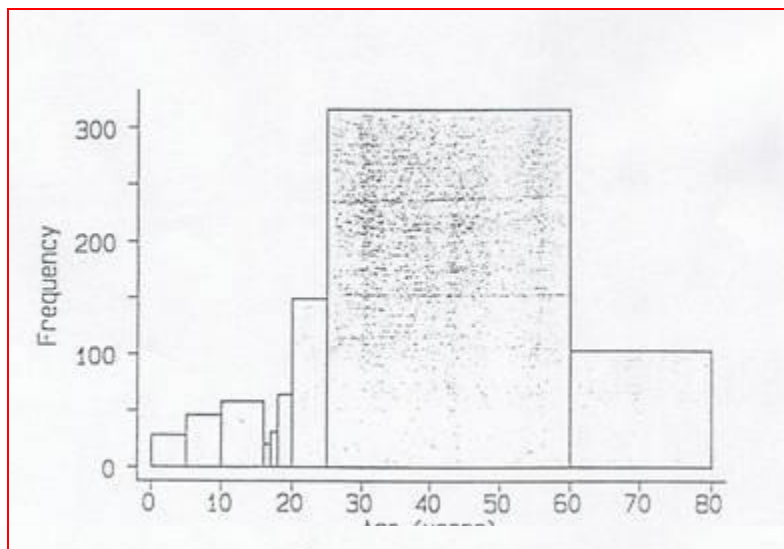
POLIGONO DELLE FREQUENZE SEMPLICI E CUMULATE



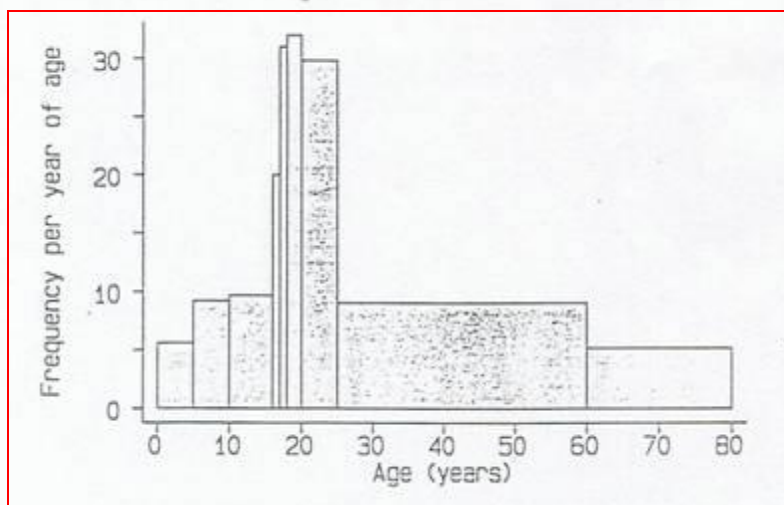
Esempio: Vittime di incidenti stradali nel London Borough of Harrow nel 1985.



scorretto



corretto



ETA'	FREQUENZA
0-4	28
5-9	46
10-15	58
16	20
17	31
18-19	64
20-24	149
25-59	316
60+	103
TOTALE	815

$316 / 815 = 38.8\%$

$316 / 35 = 9.0$

Istogramma Variabili Quantitative

TABELLA 2-4

Intervallo	Centro della classe	Frequenza	Frequenza cumulativa
0-4	2	1	1
5-9	7	4	5
10-14	12	9	14
15-19	17	11	25
20-24	22	8	33
25-29	27	15	48
30-34	32	12	60
35-39	37	14	74
40-44	42	9	83
45-49	47	5	88
50-54	52	6	94
55-59	57	4	98
60-64	62	1	99
65-69	67	1	100

Una tabella riassuntiva dei dati in tabella 2-3, che mostra gli intervalli, i valori al centro dell'intervallo, le frequenze assolute e le frequenze cumulative.

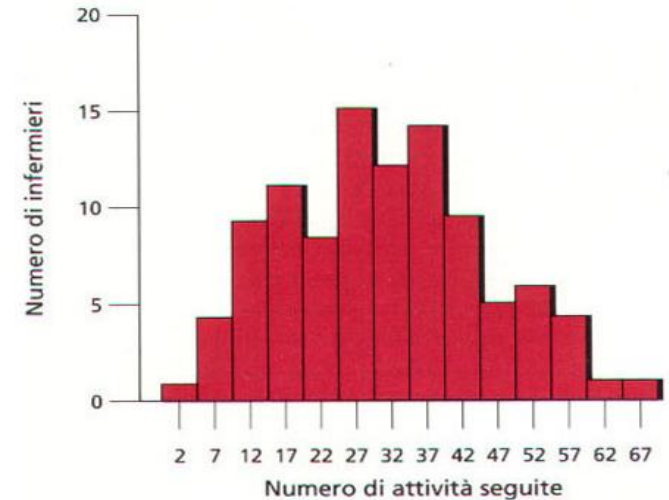
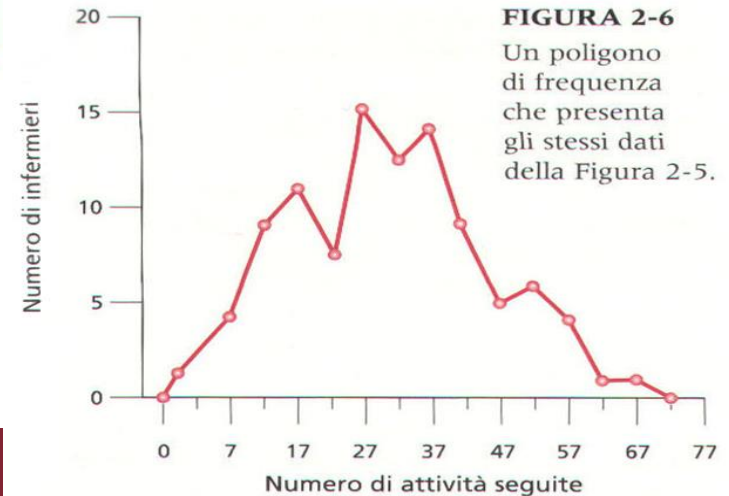


FIGURA 2-6

Un poligono di frequenza che presenta gli stessi dati della Figura 2-5.

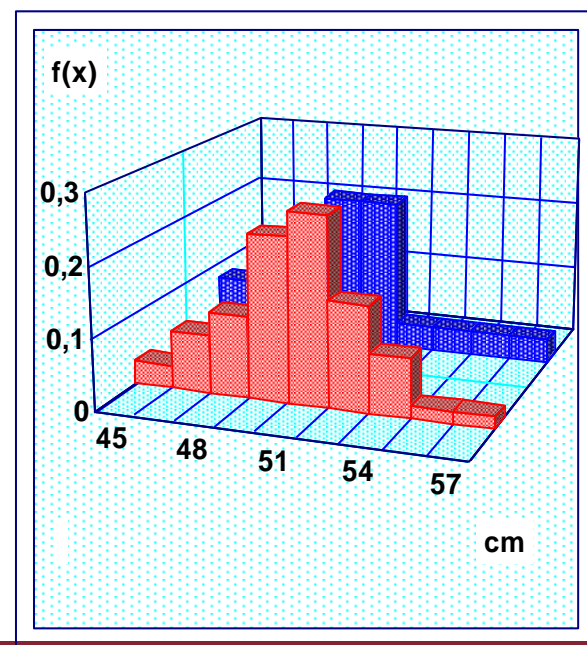
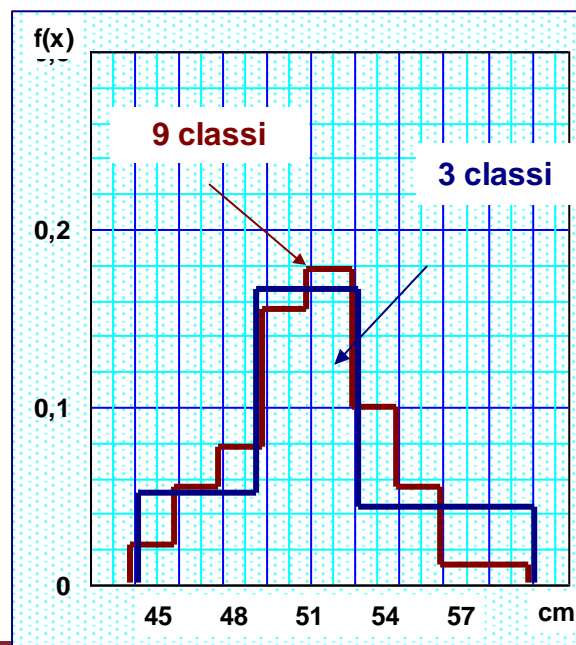
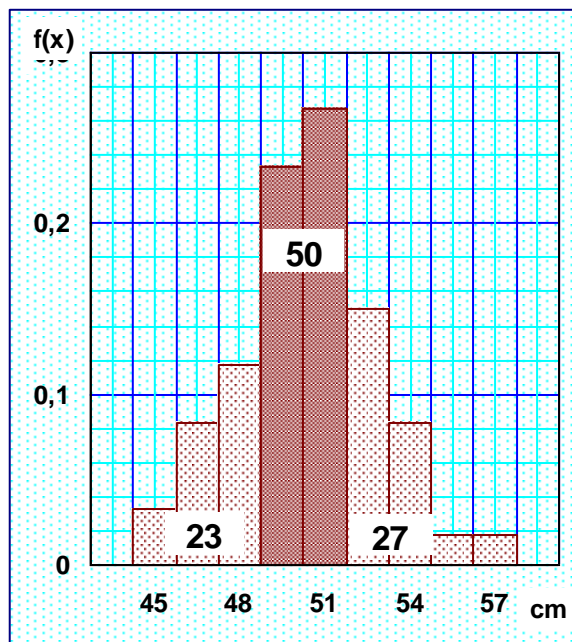


- Barre Contigue
- Poligono → Valori Continui

ISTOGRAMMI E POLIGONI DI DENSITÀ DI FREQUENZA

Negli istogrammi e nei poligoni di “densità” le **frequenze** sono **proporzionali all'area** (delimitata dalla spezzata che li costituisce e inclusa tra due valori reali sull'asse orizzontale), e **non all'altezza** della figura.

Ovviamente, quando le classi hanno tutte la **stessa ampiezza**, l'area è proporzionale anche all'altezza. I valori riportati sull'asse verticale indicano la **densità di frequenza** per una prefissata ampiezza di classe.

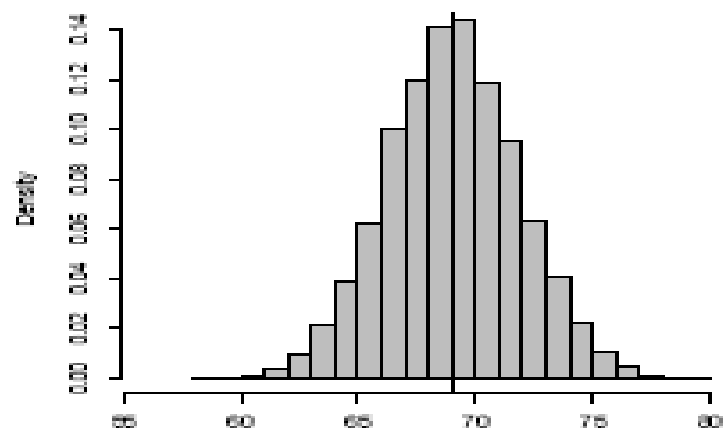
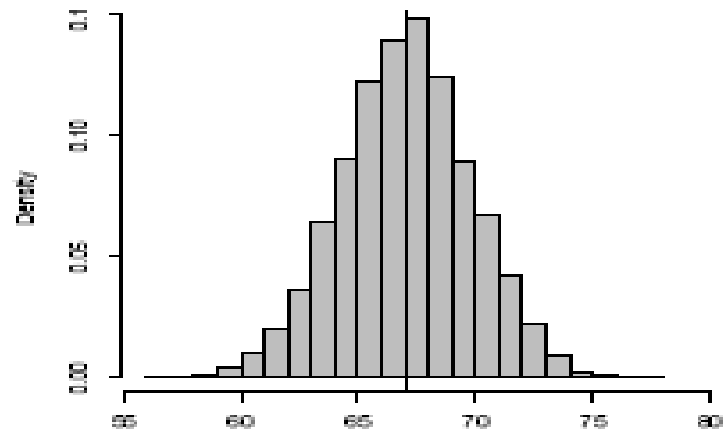


Modificazione della forma degli istogrammi in funzione dell'ampiezza delle classi.

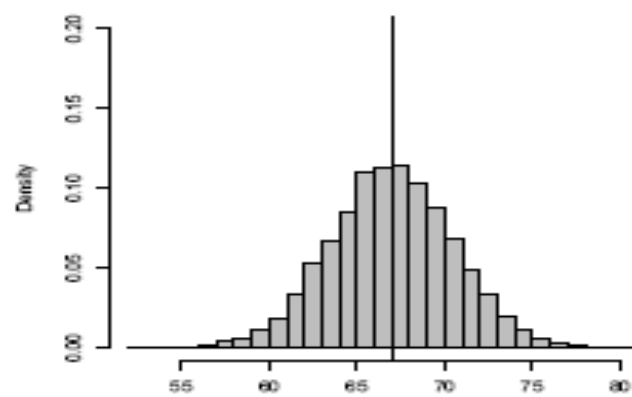
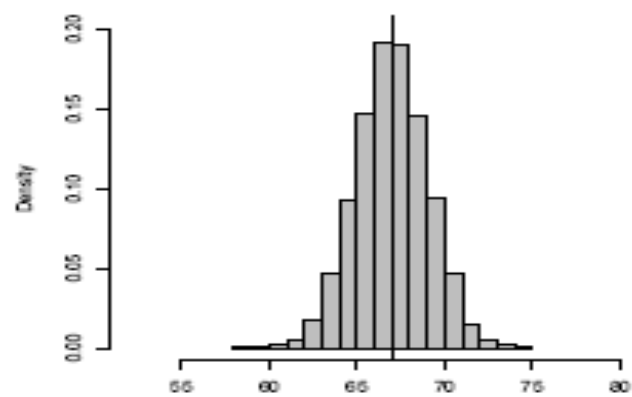
Da un istogramma è possibile desumere alcune rilevanti caratteristiche del fenomeno, per esempio:

- tendenza centrale
- dispersione
- grado di simmetria della distribuzione

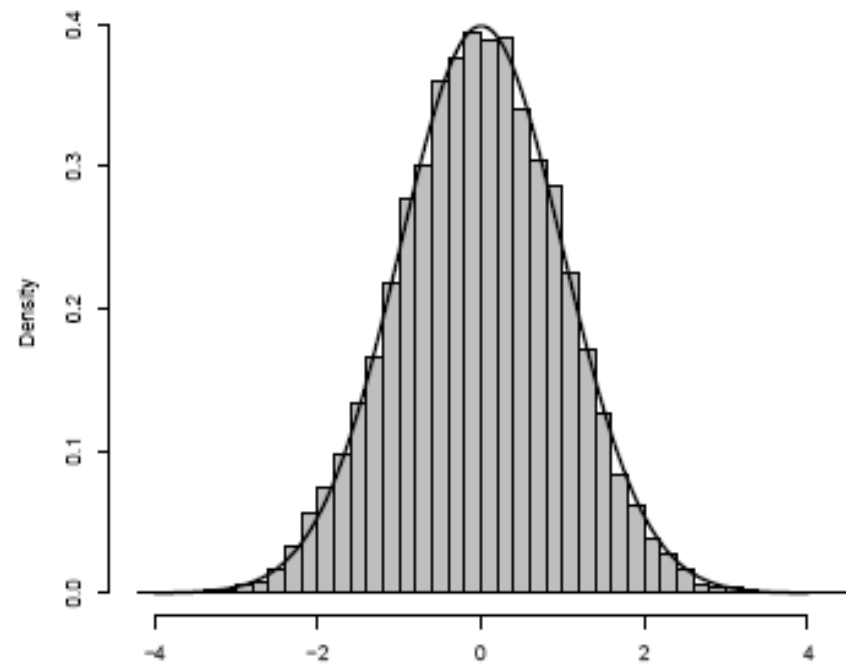
La tendenza centrale



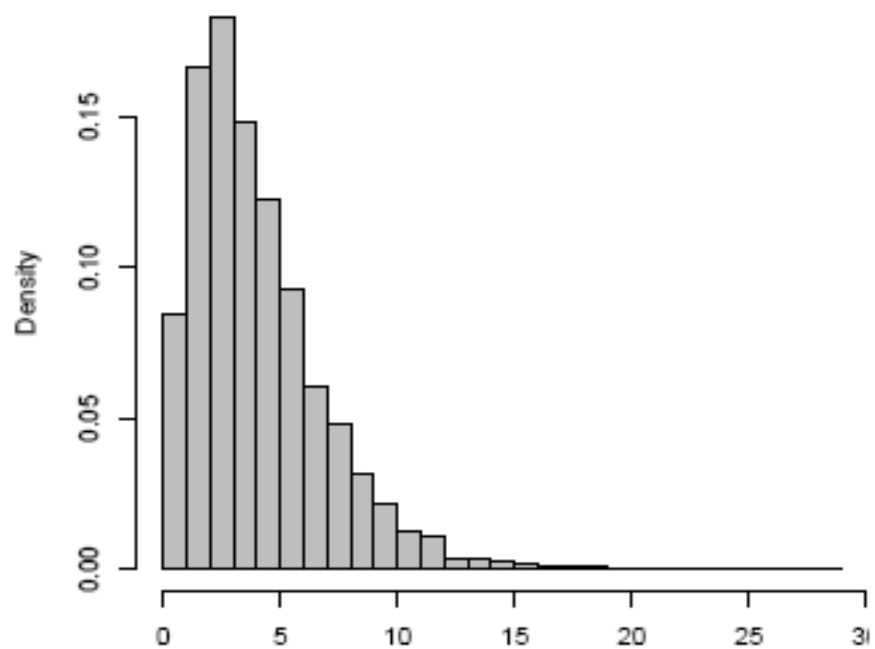
Il grado di dispersione



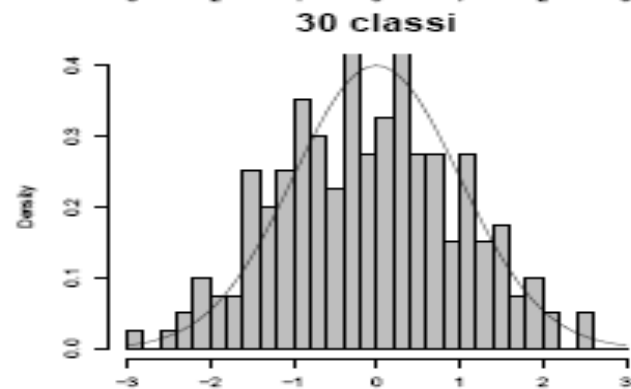
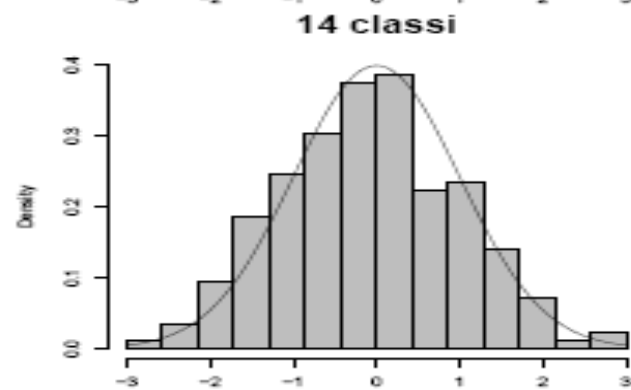
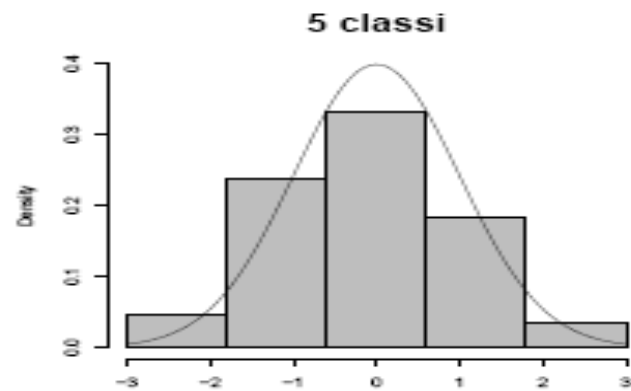
Simmetria ...



e asimmetria di una distribuzione

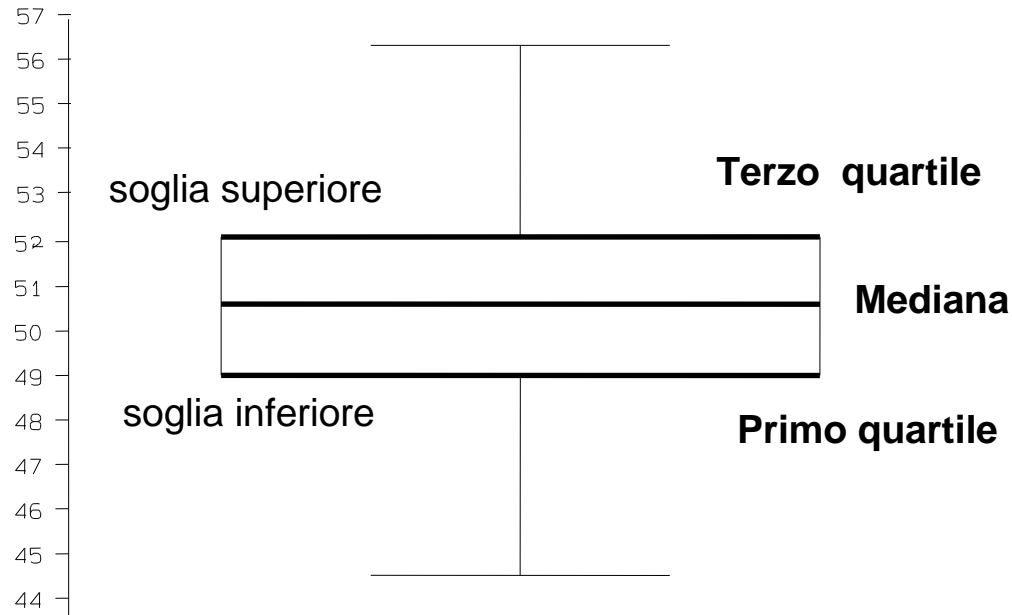


L'effetto dell'ampiezza di classe



BOXPLOT

Boxplot della *Lunghezza supina (cm)* in un campione di *60 neonati*



La linea orizzontale entro la scatola è indica il valore della *mediana* della distribuzione delle *Lunghezze supine*.

I bordi superiore ed inferiore della scatola sono **le soglie superiore ed inferiore** (ovvero i quartili) della distribuzione delle *Lunghezze supine*.

Le linee verticali agli estremi della box connettono i punti estremi delle rispettive soglie.

BOXPLOT

Sono stati rilevate le pulsazioni cardiache in un minuto di un gruppo di studenti.

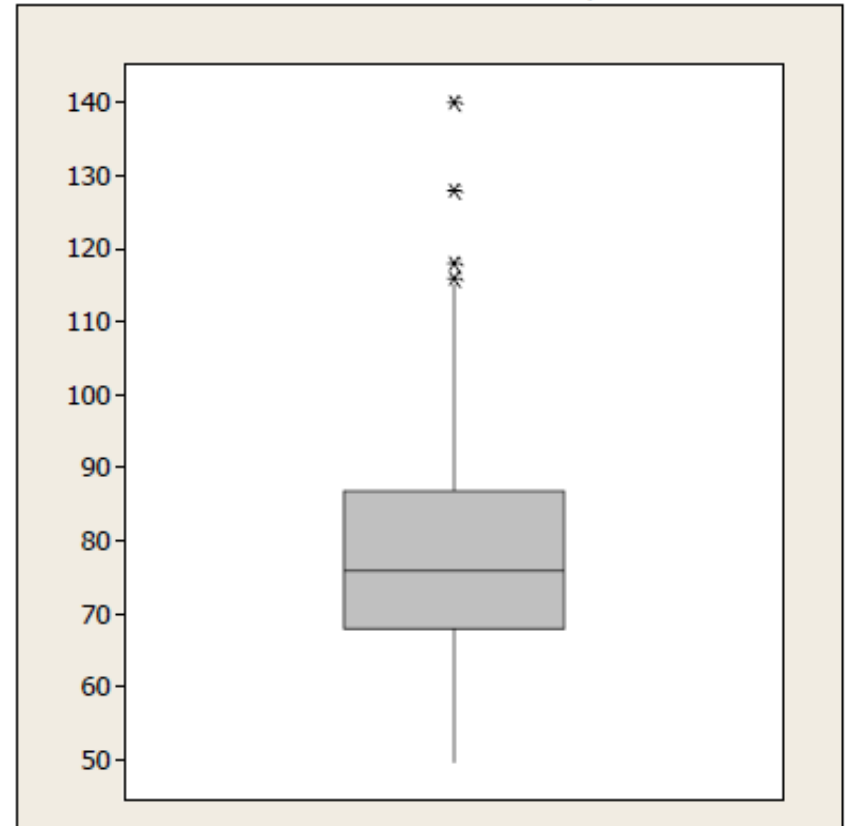
Alcuni di questi prima della rilevazione hanno effettuato un minuto di corsa, altri no.

Il boxplot a fianco rappresenta i dati dell'intero gruppo di studenti.

Si può osservare che 4 studenti hanno pulsazioni non sono comprese nei baffi.

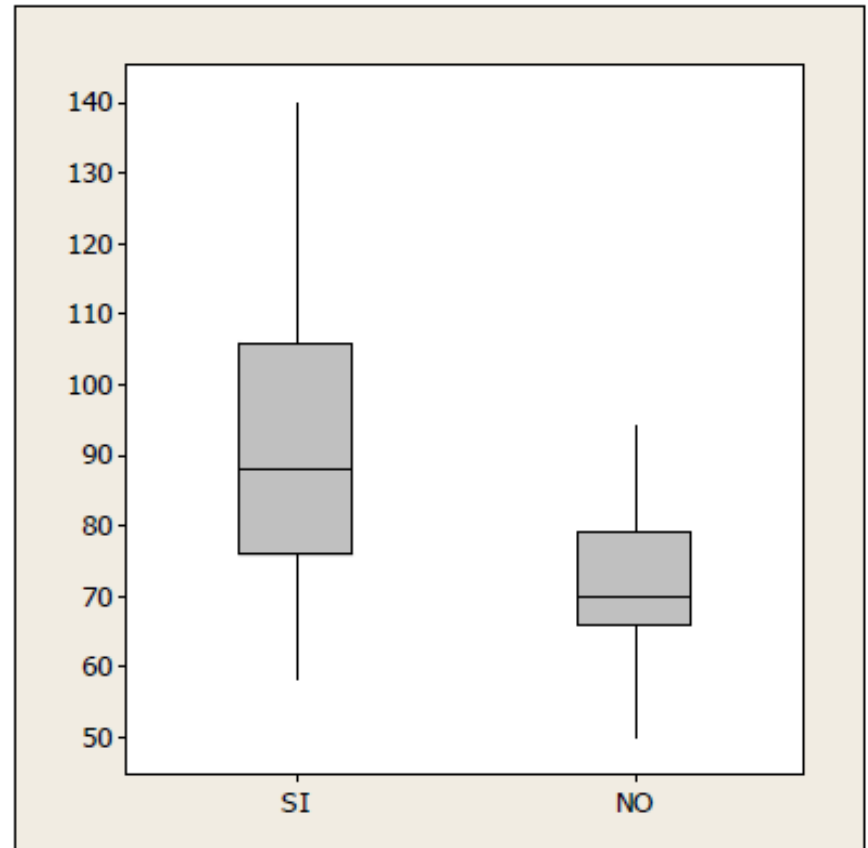
I dati così rappresentati però non sono omogenei; ovviamente le pulsazioni variano molto tra chi ha corso e chi non ha corso.

Il grafico a fianco rappresenta i dati dell'intero gruppo di studenti.



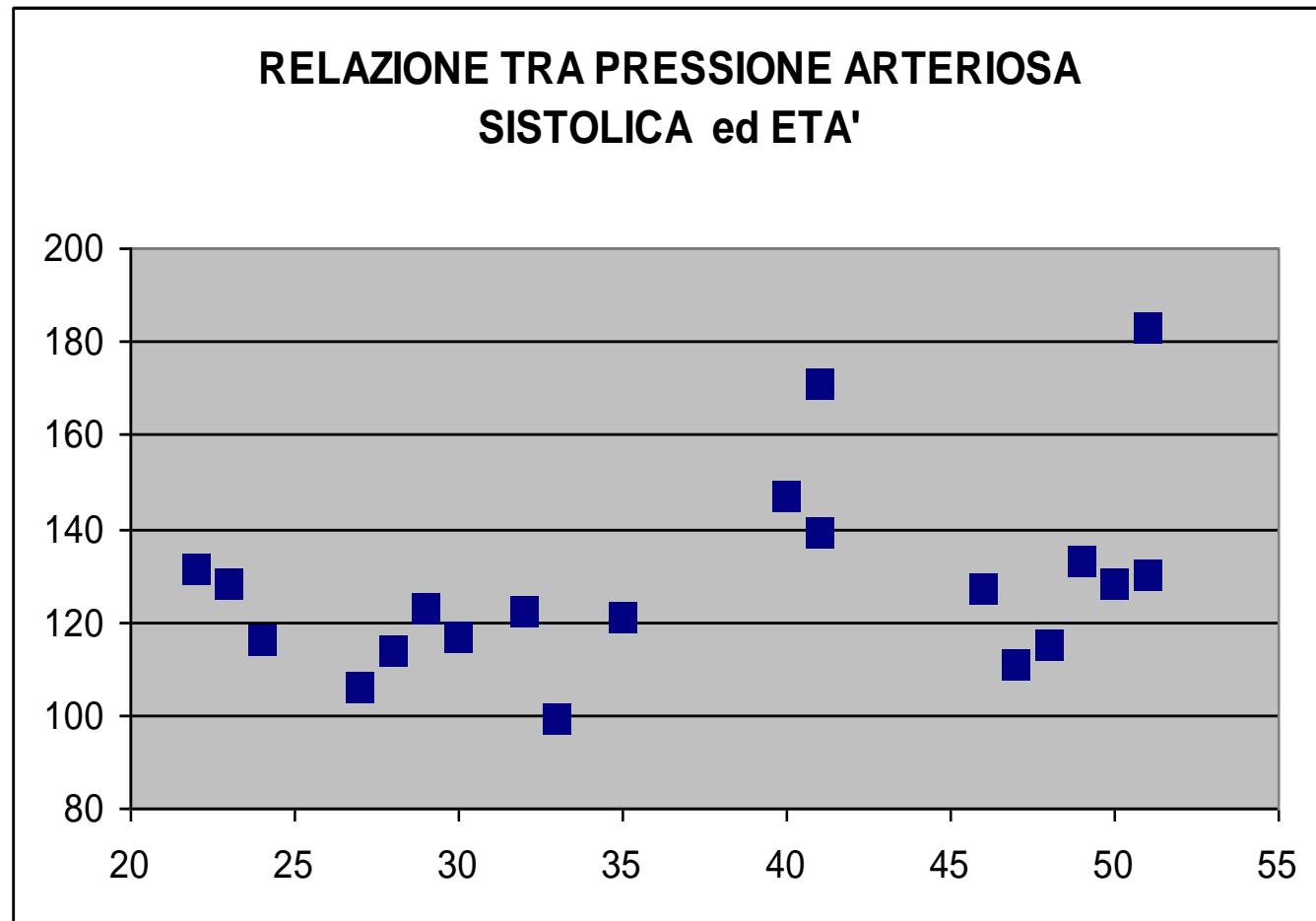
BOXPLOT

Si può osservare che le due distribuzioni assumono valori su intervalli diversi e che ciascun quartile delle pulsazioni di chi non ha corso è più basso del corrispondente quartile di chi ha corso. Inoltre le pulsazioni di chi non ha corso sono più concentrate sia nella parte centrale che nelle code. Anche le simmetrie sono diverse, in particolare i baffi sono più asimmetrici per chi ha corso. Questo e la maggiore dispersione della distribuzione si possono spiegare pensando che la reazione alla corsa varia molto da individuo a individuo: la dispersione aumenta e questo avviene soprattutto per valori alti.



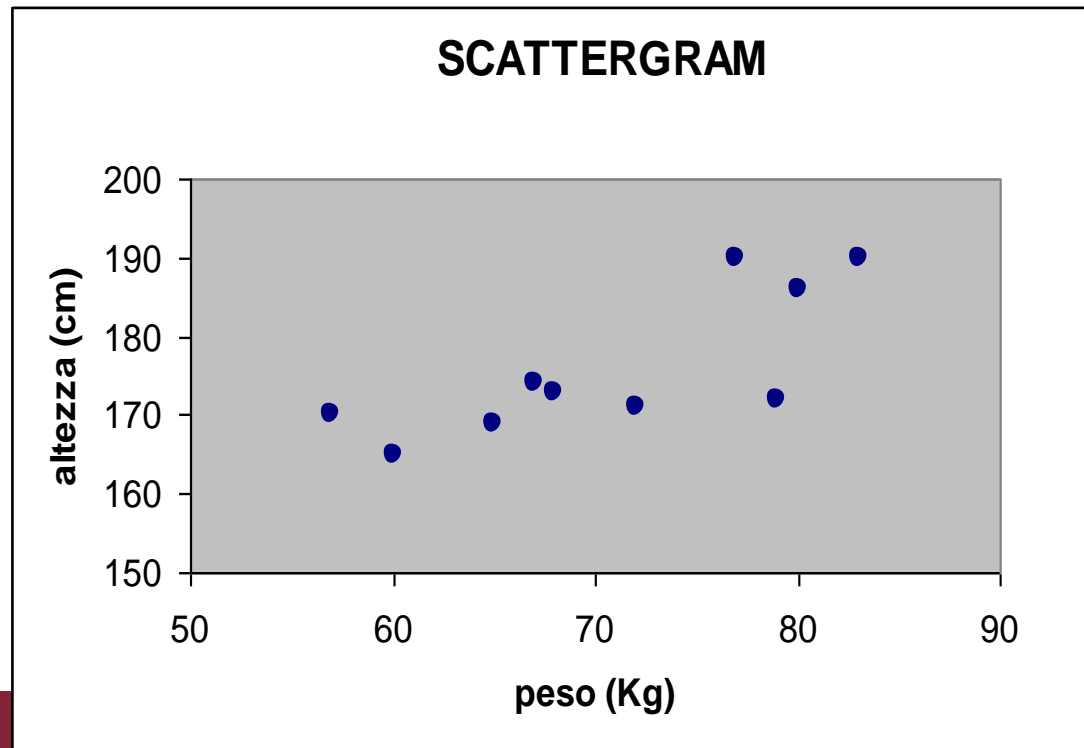
Distribuzione doppia

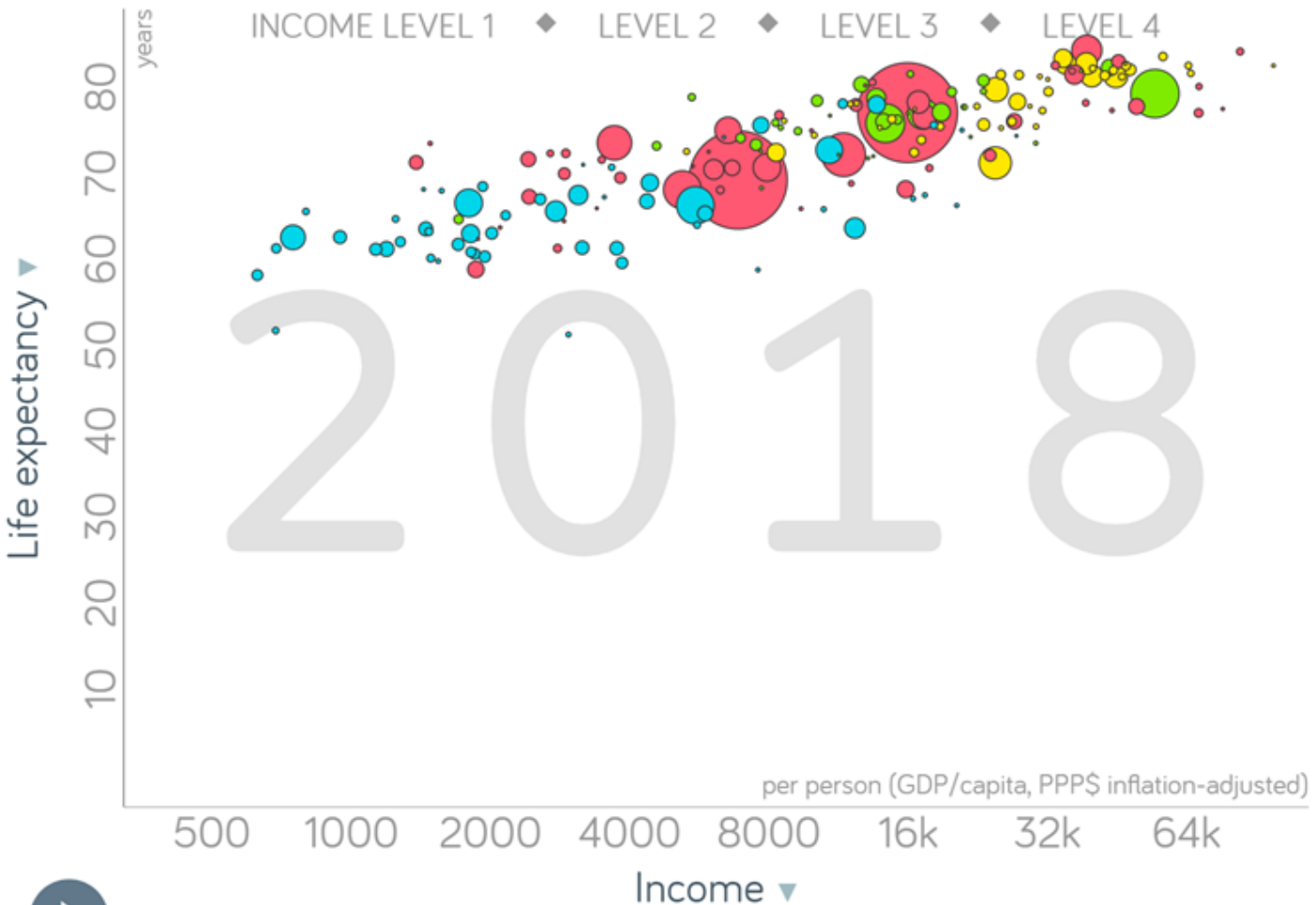
età	PAS
22	131
23	128
24	116
27	106
28	114
29	123
30	117
32	122
33	99
35	121
40	147
41	139
41	171
46	127
47	111
48	115
49	133
50	128
51	183
51	130



DISTRIBUZIONE DOPPIA

	Peso (kg)	Altezza (cm)
marco	72	171
antonell	65	169
luca	80	186
marina	57	170
gianna	60	165
luigi	77	190
francesco	83	190
michele	79	172
stefania	67	174
claudia	68	173





Color World Regions ▼



Select Search...

- Afghanistan
- Albania
- Algeria
- Andorra
- Angola
- Antigua and Barbuda
- Argentina
- Armenia
- Australia
- Austria
- Azerbaijan
- ...

Size Population ▼ ?

Zoom 100%

OPTIONS PRESENT RESTORE

DIAGRAMMI IN SINTESI

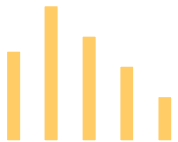
ISTOGRAMMA A CANNE D'ORGANO



***variabile
continua***

1. rettangoli adiacenti
2. le basi dei rettangoli possono essere diverse
3. frequenza sempre proporzionale all'**AREA** dei rettangoli (*anche all'altezza se basi uguali*)

DIAGRAMMA A BARRE



***variabile
discreta***

***variabile
qualitativa***

1. barre separate (per evidenziare **la non continuità dei valori / la distinzione tra le modalità**)
2. le basi delle barre sono tutte di uguale ampiezza
3. frequenza proporzionale alla **ALTEZZA** delle barre