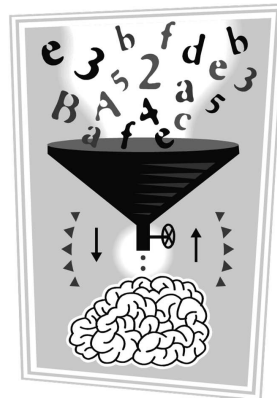


VARIABILI E DISTRIBUZIONI DI FREQUENZA



1

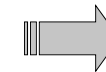
Il metodo più semplice e immediato per rappresentare in modo sintetico un insieme di osservazioni individuali relative ad una certa variabile è mediante la

DISTRIBUZIONE DI FREQUENZA

Insieme dei possibili **valori (modalità o intervalli di classe)** di una variabile con associata la frequenza con cui tali valori sono stati rilevati nel campione.

28	36	20	25	75	24	33
26	30	16	25	59	26	32
33	29	34	25	34	31	18
43	25	58	26	24	37	24
43	15	29	27	31	37	24
64	26	25	23	32	73	24
40	16	34	39	25	48	23
34	32	28	41	26	45	45
24	18	27	31	17	55	50
24	16	25	38	34	29	20
10	17	28	33	45	28	15
26	40	30	28	34	22	15
25	28	32	23	32	20	39
22	28	34	20	44	27	28
30	23	30	21	44	38	28
44	21	51	23	34	19	20
55	23	38	22	44	30	19

dati grezzi



concentrazione di ozono($\mu\text{g}/\text{m}^3$)	frequenza
10-19	13
20-29	53
30-39	31
40-49	12
50-59	6
60-69	1
70-80	4
TOTALE	120

distribuzione di frequenza

2

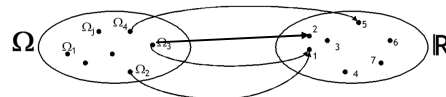
COSTRUZIONE DI UNA DISTRIBUZIONE DI FREQUENZA

1. definire un **critério di classificazione** delle osservazioni

- ⇒ definizione delle **modalità** (variabile qualitativa)
- ⇒ definizione degli **intervalli di classe** (variabile quantitativa)

IL CRITERIO DI CLASSIFICAZIONE DEVE ESSERE

- ESAUSTIVO**: devono essere riportate tutte le modalità o i valori assunti dalla variabile
- NON AMBIGUO**: le modalità/gli intervalli di classe devono essere **mutuamente esclusivi**



ogni unità statistica deve essere assegnata ad una **unica** modalità o intervallo di classe

3

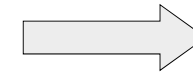
Esempio

Variabile quantitativa:
classificazione dell'età in anni compiuti



SCORRETTA

0-10
10-20
....
70-80



CORRETTA

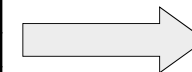
0-9
10-19
....
70-79
≥ 80

Variabile qualitativa:
classificazione del colore dei capelli



SCORRETTA

Nero
Chiaro
Biondo
Rosso




CORRETTA

Nero
Castano
Biondo
Rosso

4

2. Assegnare ad ogni valore (modalità/intervallo di classe) la **frequenza** (assoluta e/o relativa) corrispondente

FREQUENZA ASSOLUTA (n_i)

numero di osservazioni corrispondente ai diversi valori (modalità/intervalli di classe) della variabile  **ottenuta tramite un CONTEGGIO**

$\Rightarrow 0 \leq n_i \leq n$

n = numero totale delle osservazioni

$\Rightarrow \sum_{i=1}^K n_i = n_1 + n_2 + \dots + n_K = n$

K = numero dei valori/modalità/classi della variabile



FREQUENZA RELATIVA: ($p_i = n_i / n$)

rapporto tra il numero di osservazioni corrispondente ai diversi valori (modalità/intervalli di classe) della variabile e la dimensione campionaria

$\Rightarrow 0 \leq p_i \leq 1$

$\Rightarrow \sum_{i=1}^K p_i = p_1 + p_2 + \dots + p_K = 1$

FREQUENZA RELATIVA PERCENTUALE: ($p_i\% = n_i / n * 100$)

indica quanto volte un fenomeno si manifesta su una casistica di 100 osservazioni

$\Rightarrow 0\% \leq p_i\% \leq 100\%$

$\Rightarrow \sum_{i=1}^K p_i\% = p_1\% + p_2\% + \dots + p_K\% = 100\%$



PERCHÉ USARE LE FREQUENZE RELATIVE?



Per il **confronto** della distribuzione di una variabile in campioni di dimensioni diverse

Esempio: Si vuole valutare l'efficacia di uno psico-farmaco nel curare forme di balbuzie. L'esperimento coinvolge due gruppi randomizzati di pazienti (A e B): il farmaco viene somministrato a 150 pazienti nel gruppo A, mentre un placebo viene somministrato a 100 soggetti in B.

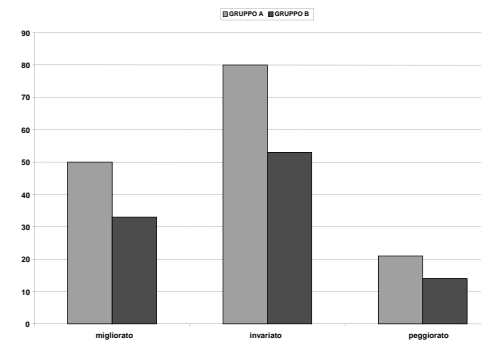


PERCHÉ USARE LE FREQUENZE RELATIVE?



FREQUENZE ASSOLUTE

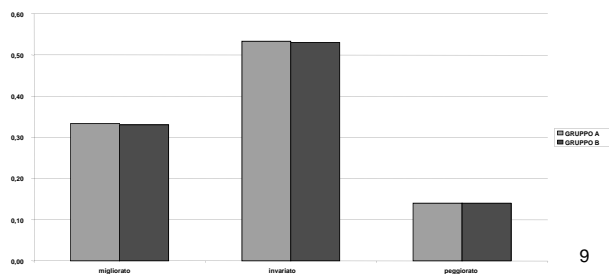
EFFETTO	n_i (A)	n_i (B)
migliorato	50	33
invariato	80	53
peggiorato	20	14
	150	100



PERCHÉ USARE LE FREQUENZE RELATIVE?

EFFETTO	$n_i(A)$	$n_i(B)$	$p_i(A)$	$p_i(B)$
migliorato	50	33	0,33	0,33
invariato	80	53	0,53	0,53
peggiolato	21	14	0,14	0,14
	151	100	1,00	1,00

FREQUENZE RELATIVE



FREQUENZA CUMULATA

FREQUENZA ASSOLUTA CUMULATA (F_i)
 numero di osservazioni il cui valore è **inferiore o uguale**
 ad una data modalità o a un dato valore x_i

$$F_i(-\infty)=0 \quad F_i(+\infty)=n$$

FREQUENZA RELATIVA CUMULATA
 $(P_i = F_i / n; P_i\% = F_i / n * 100\%)$

$$P_i(-\infty)=0 \quad P_i(+\infty)=1$$

COSTRUZIONE DELLA DISTRIBUZIONE DI FREQUENZA PER VARIABILI QUALITATIVE



Esempio: I dati seguenti si riferiscono al grado del trauma in 100 ricoverati al pronto soccorso:

```
0 2 1 1 1   2 0 0 1 0   1 1 0 0 0   3 1 2 0 1
1 0 0 1 0   1 1 0 2 0   0 0 1 0 1   0 2 1 2 0
0 2 0 1 0   1 0 1 0 3   1 2 0 0 0   0 1 0 0 0
1 0 1 0 1   0 2 0 1 2   1 2 0 1 0   2 2 1 0 1
0 0 0 0 4   0 1 1 2 0   0 2 1 0 2   0 0 2 1 0
```

X= grado del trauma

X_i :

0=assente 1=trauma lieve 2=trauma grave
 3=lesioni permanenti 4=decesso



Conteggio delle osservazioni...

...per ogni modalità

modalità	tally	frequenza
assente		48
lieve		32
grave		17
lesioni permanenti		2
decesso		1
		100

modalità	tally	frequenza
assente		48
lieve		32
grave		17
lesioni permanenti		2
decesso		1
		100

MODALITA'	frequenza assoluta n_i	frequenza relativa n_i/n
assente	48	$48/100 = 0,48$
lieve	32	0,32
grave	17	0,17
lesioni permanenti	2	0,02
decesso	1	0,01
TOTALE	100	

k=5

Costruzione della tabella e calcolo di frequenze relative

13



esempio (grado del trauma):

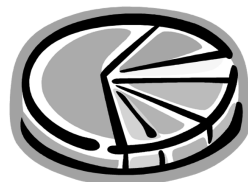
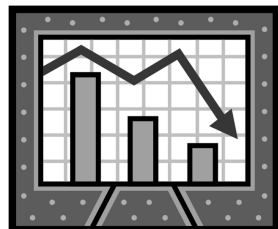
distribuzione di frequenza assoluta, relativa e cumulativa della variabile "grado del trauma"

valore x_i	assoluta n_i	relativa p_i	relativa percentuale $p_i (%)$	assoluta cumulata N_i	relativa cumulata P_i	relativa cumulata percentuale $P_i (%)$
assente	48	0.48	48%	48	$48 / 100 = 0.48$	$0.48 * 100 = 48%$
lieve	32	0.32	32%	$48 + 32 = 80$	$80 / 100 = 0.80$	$0.80 * 100 = 80%$
grave	17	0.17	17%	$80 + 17 = 97$	$97 / 100 = 0.97$	$0.97 * 100 = 97%$
lesioni permanenti	2	0.02	2%	$97 + 2 = 99$	$99 / 100 = 0.99$	$0.99 * 100 = 99%$
decesso	1	0.01	1%	$99 + 1 = 100$	$100 / 100 = 1$	$1 * 100 = 100%$
TOTALE	100	1	100%			

dimensione campionaria

14

RAPPRESENTAZIONI GRAFICHE DELLA DISTRIBUZIONE DI FREQUENZA PER VARIABILI QUALITATIVE



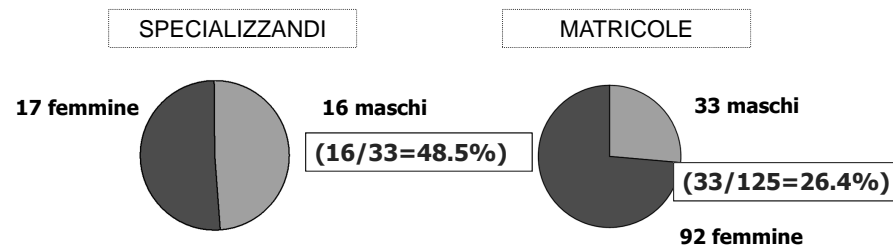
15



GRAFICO A TORTA

Esempio: ci sono 16 maschi tra gli specializzandi e 33 tra le matricole di Medicina (frequenze assolute, n_i).

Se consideriamo le frequenze assolute i maschi tra gli specializzandi sono la metà rispetto ai maschi tra le matricole di Medicina.



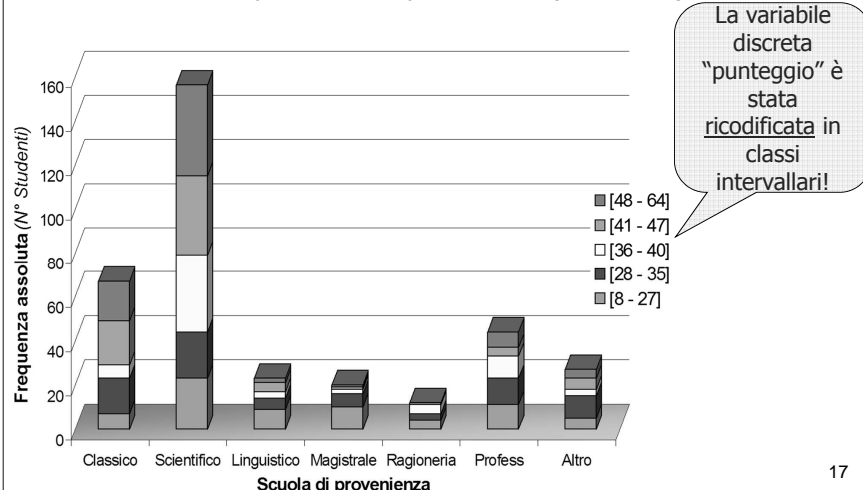
In realtà il sesso maschile è molto più presente tra gli specializzandi (48.5%) che non tra le matricole di medicina (26.4%)

16

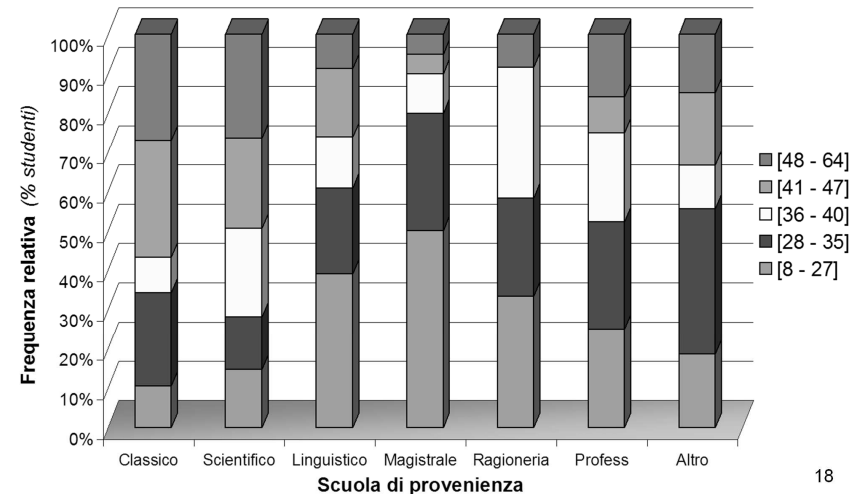


DIAGRAMMA A BARRE

Esempio: Distribuzione di **frequenza assoluta** dei punteggi al test degli studenti di Medicina, per scuola di provenienza (a.a. 95/96)



(continua...) Distribuzione di **frequenza relativa** dei punteggi al test degli studenti di Medicina, per scuola di provenienza (a.a. 95/96)



ESERCIZIO

I dati seguenti si riferiscono al tipo di parto di 50 neonati in Italia:



X = tipo di parto

- x_i = normale → 0
- forcipe → 1
- cesareo → 2

0	2	0	0	0	2	0	2	0	0
2	0	0	0	2	0	0	0	1	0
0	0	2	0	0	0	2	0	0	2
0	2	0	0	0	2	0	2	0	0
0	0	0	2	0	2	0	0	2	0

→ Determinare la distribuzione di frequenza



modalità x_i	frequenza assoluta n_i	frequenza relativa p_i	frequenza relativa percentuale p_i (%)
normale	35		
forcipe	1		
cesareo	14		
TOTALE	50		

COSTRUZIONE DELLA DISTRIBUZIONE DI FREQUENZA PER VARIABILI QUANTITATIVE



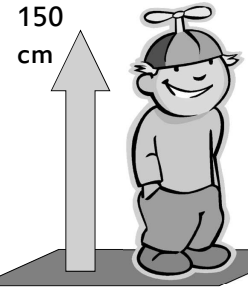
PESO, STATURA e SESSO delle MATRICOLE di MEDICINA dell'UNIVERSITA' di VERONA nell'A.A. 95/96

PESO STAT. SESSO	PESO STAT. SESSO	PESO STAT. SESSO
Kg cm	Kg cm	Kg cm
56 159 F	77 192 M	51 171 F
66 169 F	60 173 F	48 156 F
50 160 F	78 162 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	49 165 F
52 170 F	67 177 M	49 165 F
58 173 F	68 170 M	62 178 M
52 167 F	50 160 F	68 174 M
73 178 M	51 167 F	75 181 M
67 175 M	95 193 M	48 167 F
52 165 F	58 160 F	53 160 F
56 171 F	67 178 F	49 167 F
57 166 F	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	69 165 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	

serie grezza di dati



DISTRIBUZIONE DI FREQUENZA DELLA VARIABILE QUANTITATIVA CONTINUA STATURA



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

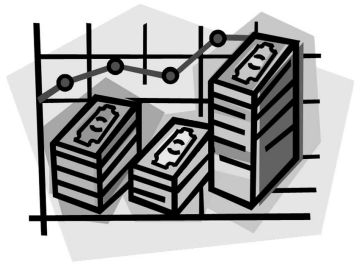
Costruiamo gli intervalli di frequenza:

- ↓ Trovo il valore minimo e il valore massimo → min = 150 cm
max = 193 cm
- ↓ Calcolo il campo di variazione (range):
 $X_{max} - X_{min}$ → r = 43
- ↓ Stabilire il numero degli intervalli → k = 9
- ↓ Calcolare l'ampiezza degli intervalli:
 $\delta_i = \text{Range} / k$ → $\delta_i = 43/9 = 4.8 \sim 5$
- ↓ Costruisco gli intervalli di classe (esclusivi ed esaustivi)
- ↓ Conto il numero di individui per ogni classe

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%

RAPPRESENTAZIONI GRAFICHE DELLA DISTRIBUZIONE DI FREQUENZA PER VARIABILI QUANTITATIVE

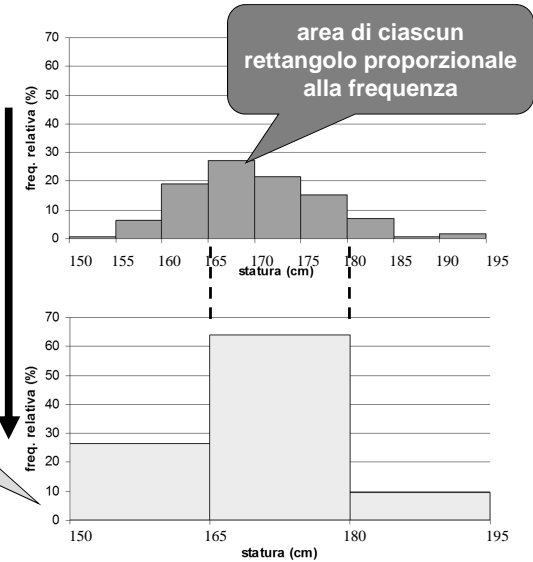


25

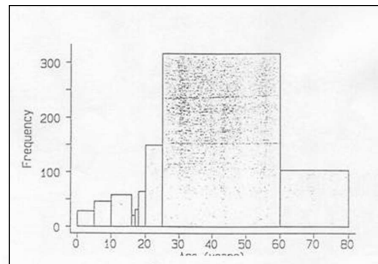
ISTOGRAMMA A CANNE D'ORGANO

perdita di informazione al diminuire del numero di intervalli

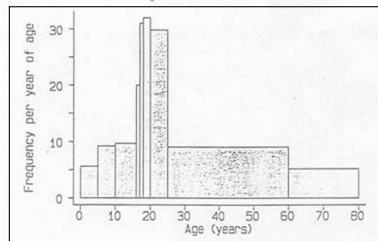
la base dei rettangoli è proporzionale all'ampiezza delle classi



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

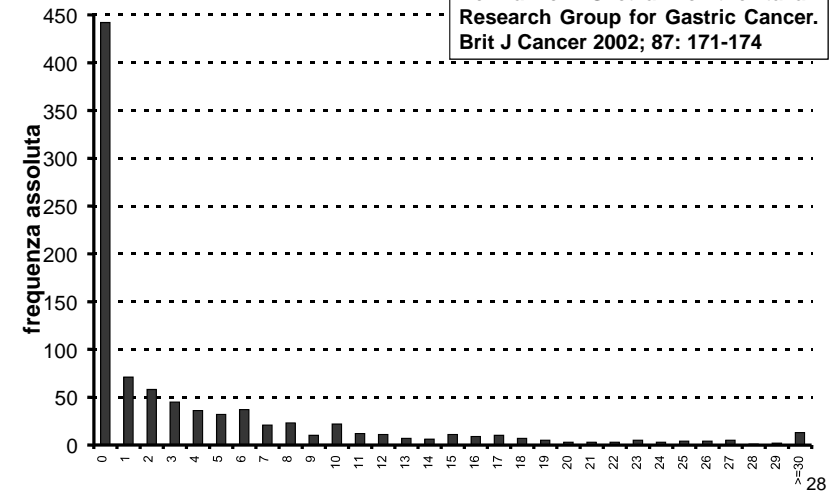
$\frac{316}{815} = 38.8\%$

$\frac{316}{35} = 9.0$

27

RAPPRESENTAZIONE GRAFICA DELLA DISTRIBUZIONE DI FREQUENZA DI UNA VARIABILE QUANTITATIVA DISCRETA: DIAGRAMMA A BARRE

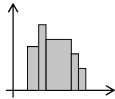
De Manzoni G et al. for the Italian Research Group for Gastric Cancer. Brit J Cancer 2002; 87: 171-174



numero di linfonodi metastatici in 921 pazienti con cancro gastrico

DIAGRAMMI IN SINTESI

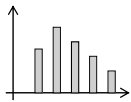
ISTOGRAMMA
A CANNE
D'ORGANO



**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**

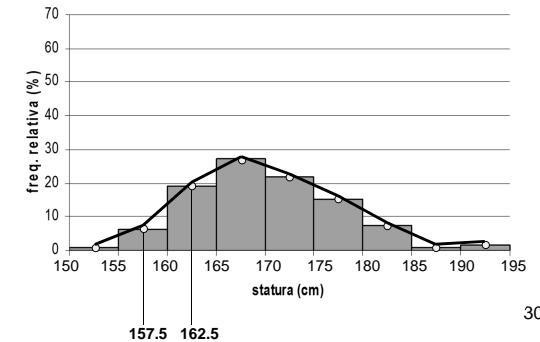
1. barre separate (per evidenziare la non continuità dei valori / la distinzione tra le modalità)
2. le basi delle barre sono tutte di uguale ampiezza
3. frequenza proporzionale alla **ALTEZZA** delle barre

29

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



30

TABELLA DI FREQUENZA

	n_i	p_i	F_i	P_i
statura in classi	Freq. Assoluta	Freq. Relativa (%)	Freq. Assoluta Cumulata	Freq. Relativa Cumulata (%)
[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	

31

ESERCIZIO

Nella tabella seguente sono riportati i dati relativi ad uno studio sulla crescita condotto su 40 soggetti:

Distanza in mm fra il centro della ghiandola pituitaria e la fossa pterigo-mascellare:

16	19	19	20	20	20	20	21	21	21
21	21	21	22	22	22	22	22	22	23
23	23	23	23	23	23	23	24	24	24
24	24	24	25	25	25	25	26	26	27



1. Costruire 4 intervalli di frequenza
2. Costruire la tabella di frequenza riportando frequenze assolute, frequenze relative e frequenze cumulate relative.

32

SOLUZIONE ESERCIZIO



- ✚ Valore minimo=16 mm valore massimo =27 mm
- ✚ Campo di variazione (*range*): **27-16=11**
- ✚ Numero degli intervalli: **k=4**
- ✚ Ampiezza degli intervalli: $\delta_i = 11/4 = 2.75 \sim 3$

TABELLA DI FREQUENZA:

classe	n_i	p_i	F_i	P_i
16-18	1	2,5%	1	
19-21	12	30,0%		
22-24	20	50,0%		
25-27	7	17,5%		
	40	100,0%		

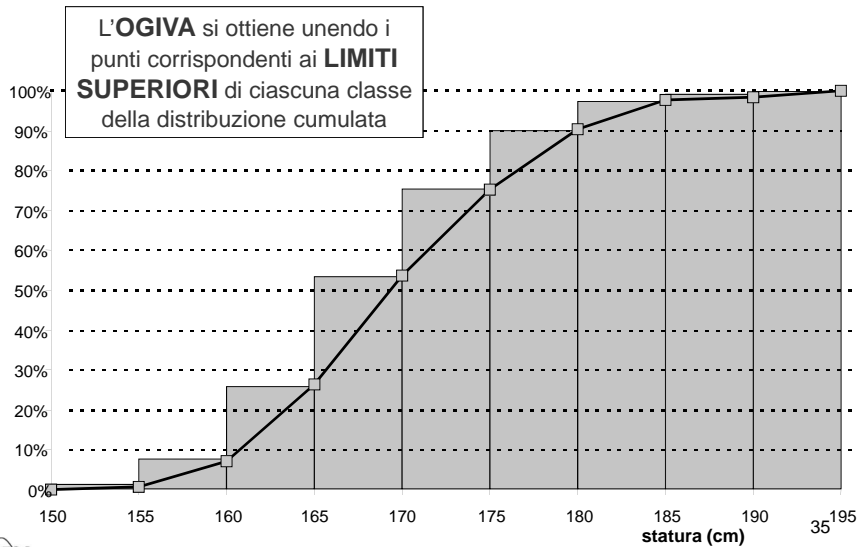


TABELLA DI FREQUENZA

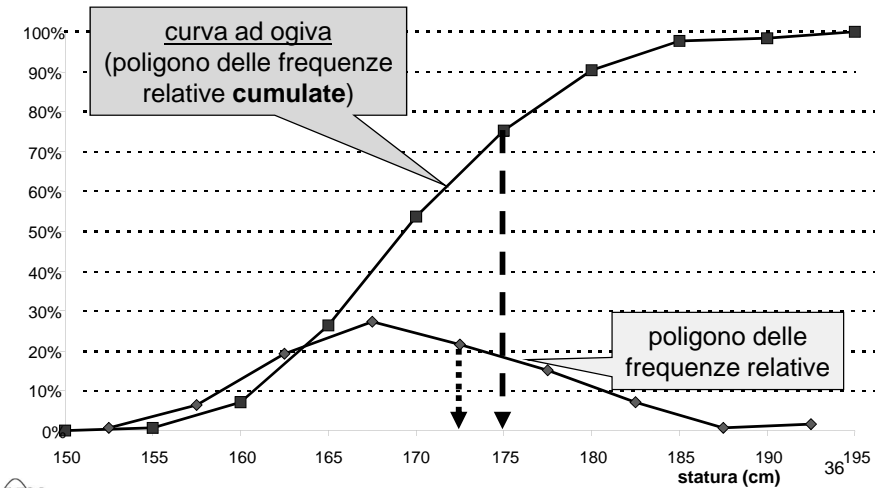
n_i	p_i	F_i	P_i
statura in classi	Freq. Assoluta	Freq. Relativa (%)	Freq. Assoluta Cumulata / Freq. Relativa Cumulata (%)
[150-155)	1	1/125= 0.8%	1 / 0.8%
[155-160)	8	8/125= 6.4%	1+8 9 / 0,8+6,4 7.2%
[160-165)	24	24/125= 19.2%	1+8+24 33 / 0,8+6,4+19,2 26.4%
[165-170)	34	27.2%	1+8+24+34 67 / 0,8+6,4+19,2+27,2 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



ISTOGRAMMA e POLIGONO DELLE FREQUENZE CUMULATE (CURVA AD OGIVA)



RAPPRESENTAZIONE GRAFICA DELLA DISTRIBUZIONE DI FREQUENZA DI UNA VARIABILE QUANTITATIVA: POLIGONO DELLE FREQUENZE SEMPLICI E CUMULATE



MISURE D'ORDINE IN UNA DISTRIBUZIONE

SCOPO: descrivere la posizione di un dato individuale nell'ambito di una distribuzione

RANGO: posizione di un'osservazione x_i in una serie di dati ordinati in modo crescente

RANGO PERCENTILICO: sia x_i la i -esima osservazione di un campione di n unità ordinate in modo crescente. Il rango percentilico corrispondente è dato da:

$$R_p = \frac{\text{rango}(x_i)}{n+1} * 100\%$$

37

Esempio:

nelle seguenti tabelle si riportano le osservazioni del peso per n soggetti

$n = 6$

PESO (kg)	53	55	60	61	63	65
-----------	----	----	----	----	----	----

Rango = 3
 $R_p = 3 / (6+1) * 100 = 43\%$

$n = 60$

PESO (kg)	53	55	60	61	63	65	92
-----------	----	----	----	----	----	----	-------	----

Rango = 3
 $R_p = 3 / (60+1) * 100 = 5\%$



Se due o più individui (unità statistiche) hanno lo stesso valore, si assegna ad esso il **rango medio** delle posizioni da essi occupati.

rango	1	2	3	4	5
valori	3	4	4	5	6
		2,5	2,5		

rango	1	2	3	4	5
valori	3	4	4	4	5
		3	3	3	

39