

LEZIONI DI STATISTICA MEDICA

Prof. SIMONE ACCORDINI

Lezione n.3

- Distribuzioni di frequenza*
- Distribuzioni di frequenza
per una variabile qualitativa*



*Sezione di Epidemiologia & Statistica Medica
Università degli Studi di Verona*

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

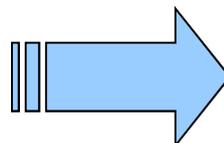
Insieme delle possibili **modalità o valori** (→ **intervalli di classe**) di una variabile con associata la **frequenza** con cui tali modalità o valori sono stati rilevati nel campione.

esempio: X = concentrazione di ozono ($\mu\text{g}/\text{m}^3$)

rilevato in 119 tempi diversi:

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	13
50-59	6
60-69	1
70-75	2
TOTALE	119

distribuzione di frequenza

COSTRUZIONE DI UNA DISTRIBUZIONE DI FREQUENZA

1. Definire un criterio di classificazione delle osservazioni

⇒ definizione delle modalità (variabile qualitativa)

⇒ definizione degli intervalli di classe (variabile quantitativa)

IL CRITERIO DI CLASSIFICAZIONE DEVE ESSERE:

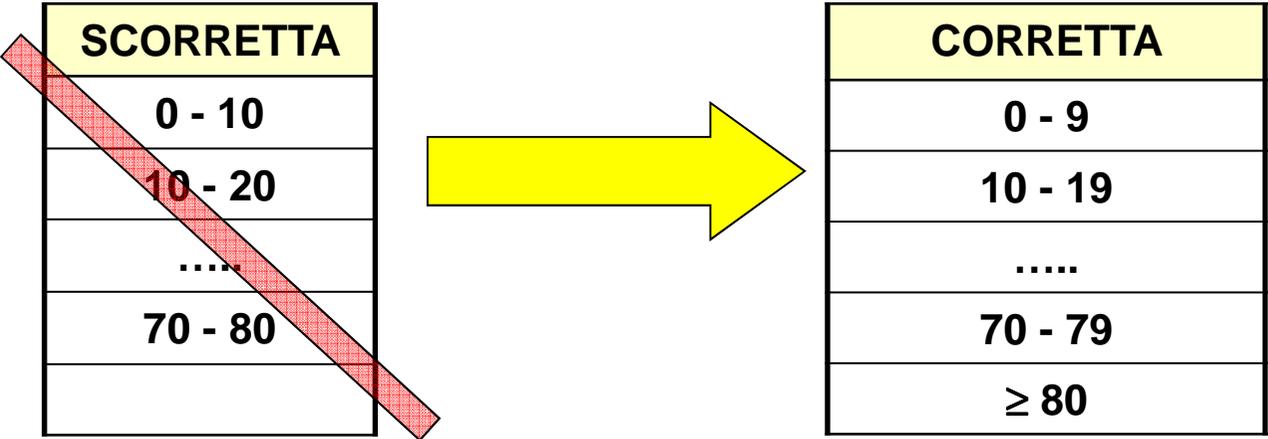
ESAUSTIVO: devono essere riportate tutte le modalità o i valori assunti dalla variabile

NON AMBIGUO: le modalità o gli intervalli di classe devono essere mutuamente esclusivi

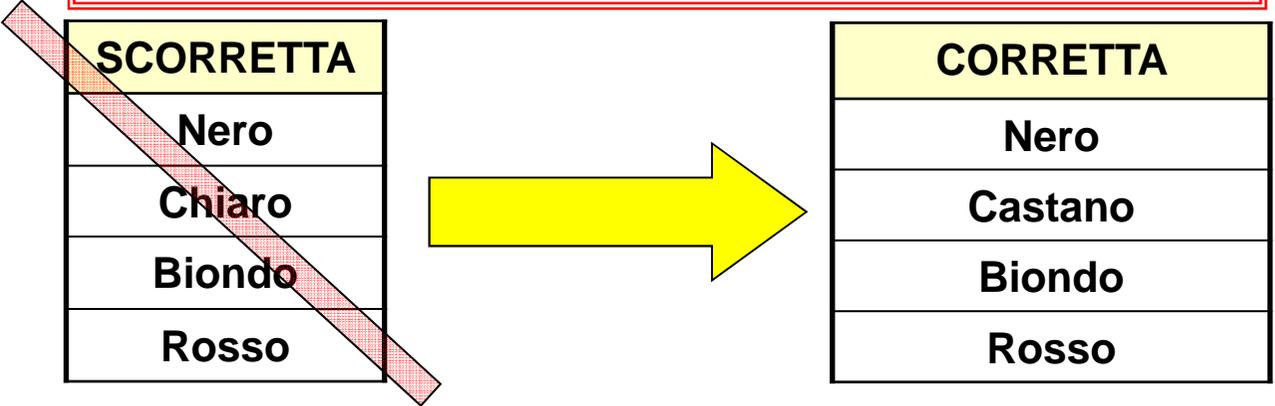
⇒ **ogni** unità statistica deve essere assegnata ad un'**unica** modalità o intervallo di classe

Esempio

Variabile quantitativa:
classificazione dell'età in anni compiuti



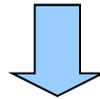
Variabile qualitativa:
classificazione del colore dei capelli



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

FREQUENZA ASSOLUTA: n_i

numero di osservazioni corrispondente alle diverse modalità o valori (→ intervalli di classe) della variabile



ottenuta tramite un **CONTEGGIO**

$$\Rightarrow 0 \leq n_i \leq n$$

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

n = numero totale delle osservazioni
(**numerosità campionaria**)

K = numero delle modalità, valori o intervalli di classe della variabile

FREQUENZA RELATIVA: $p_i = n_i / n$

rapporto tra il numero di osservazioni corrispondente alle diverse modalità o valori (→ intervalli di classe) della variabile e la numerosità 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: $p_i (\%) = p_i * 100$

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

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

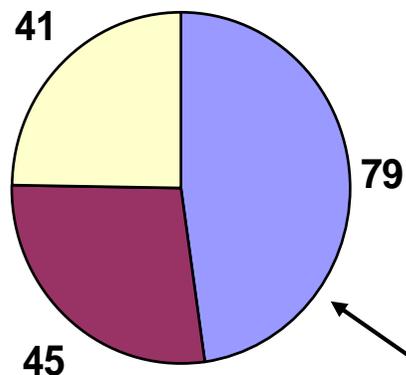
$$\Rightarrow \sum_{i=1}^K p_i (\%) = p_1 (\%) + p_2 (\%) + \dots + p_K (\%) = 100$$

DISTRIBUZIONE DI FREQUENZA DI UNA VARIABILE QUALITATIVA NOMINALE

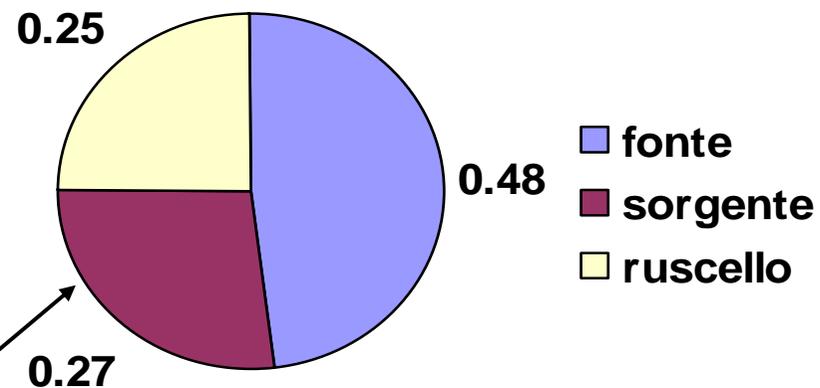
modalità X_i	frequenza assoluta n_i	frequenza relativa p_i	frequenza relativa percentuale p_i (%)
fonte	79	$79/165 = 0.48$	$(79/165)*100 = 48\%$
sorgente	45	0.27	27%
ruscello	41	0.25	25%
TOTALE	165	1.00	100%

RAPPRESENTAZIONE GRAFICA DELLA DISTRIBUZIONE DI UNA VARIABILE QUALITATIVA NOMINALE: **GRAFICO A TORTA**

FREQUENZE ASSOLUTE



FREQUENZE RELATIVE



■ fonte
■ sorgente
■ ruscello

area proporzionale
alla frequenza

DISTRIBUZIONE DI FREQUENZA CUMULATA

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

FREQUENZA ASSOLUTA CUMULATA: N_i

numero di osservazioni la cui modalità o il cui valore è inferiore o uguale ad una data modalità o a un dato valore x_i

valori ordinati in modo crescente	↓	x_1	n_1	$N_1 = n_1$
		x_2	n_2	$N_2 = n_1 + n_2$
	
		x_K	n_K	$N_K = n_1 + n_2 + .. + n_K = n$

**FREQUENZA RELATIVA CUMULATA E RELATIVA CUMULATA
PERCENTUALE: P_i e $P_i(\%) = P_i * 100$**

frequenza relativa di osservazioni la cui modalità o il cui valore è inferiore o uguale ad una data modalità o a un dato valore x_i

valori ordinati in modo crescente	↓	x_1	p_1	$P_1 = p_1$
		x_2	p_2	$P_2 = p_1 + p_2$
	
		x_K	p_K	$P_K = p_1 + p_2 + .. + p_K = 1$

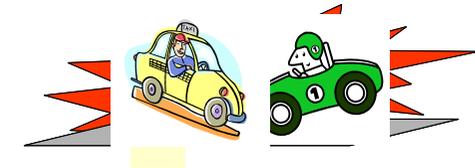
VARIABILE QUALITATIVA ORDINALE

I dati seguenti si riferiscono al grado del trauma in 95 pazienti accolti 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					

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

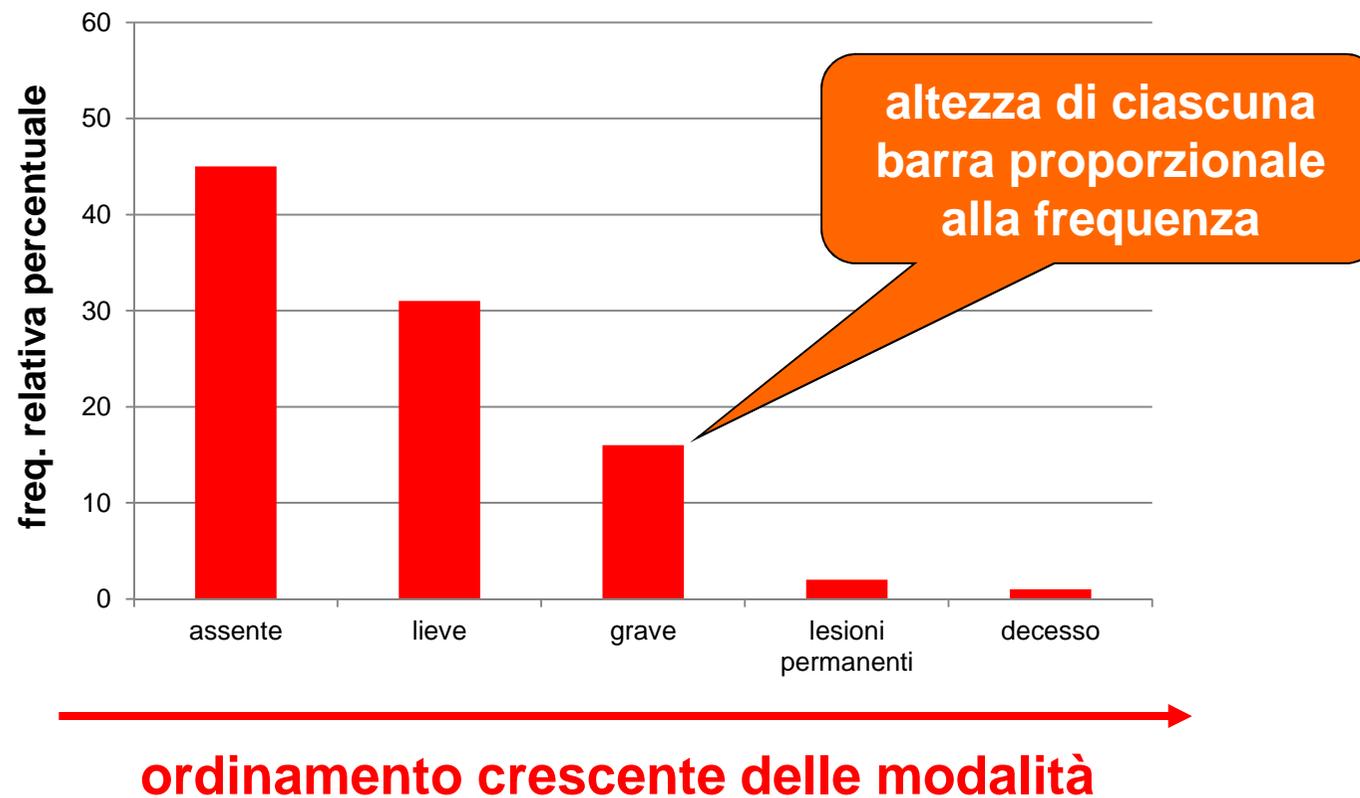


modalità	tally	frequenza
assente		45
lieve		31
grave		16
lesioni permanenti		2
decesso		1
		95

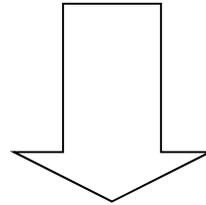
DISTRIBUZIONE DI FREQUENZA DI UNA VARIABILE QUALITATIVA ORDINALE

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	45	0.47	47%	45	$45 / 95 = 0.47$	$0.47 * 100 = 47\%$
lieve	31	0.33	33%	$45 + 31 = 76$	$76 / 95 = 0.80$	$0.80 * 100 = 80\%$
grave	16	0.17	17%	$76 + 16 = 92$	$92 / 95 = 0.97$	$0.97 * 100 = 97\%$
lesioni permanenti	2	0.02	2%	$92 + 2 = 94$	$94 / 95 = 0.99$	$0.99 * 100 = 99\%$
decesso	1	0.01	1%	$94 + 1 = 95$	$95 / 95 = 1$	$1 * 100 = 100\%$
TOTALE	95	1	100%			

RAPPRESENTAZIONE GRAFICA DELLA DISTRIBUZIONE DI FREQUENZA DI UNA VARIABILE QUALITATIVA ORDINALE: **DIAGRAMMA A BARRE**



PERCHÉ USARE LE FREQUENZE RELATIVE?

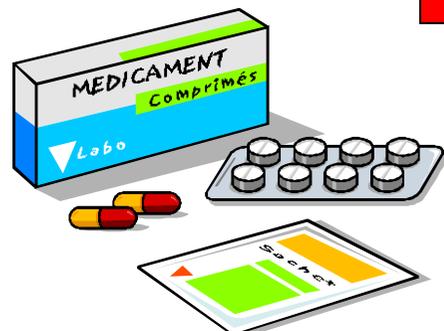


Per il **confronto** della distribuzione di una variabile in campioni di dimensioni diverse

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 nel gruppo B.



PERCHÉ USARE LE FREQUENZE RELATIVE?



FREQUENZE ASSOLUTE

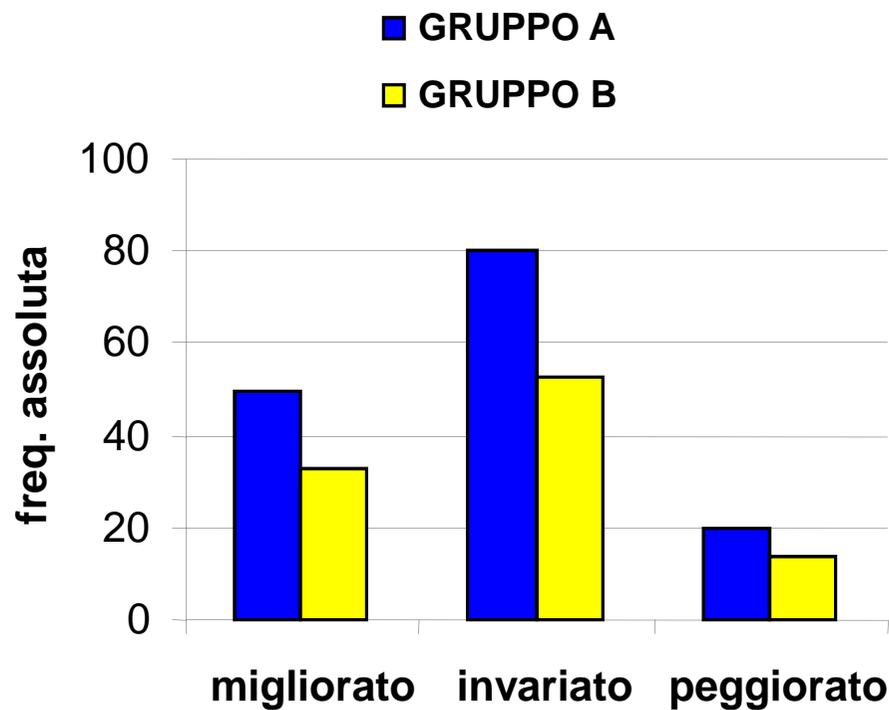
EFFETTO	$n_i(A)$	$n_i(B)$
---------	----------	----------

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
peggiorato	20	14	0.14	0.14
	150	100	1.00	1.00

FREQUENZE RELATIVE

