

# LEZIONE 5 - Distribuzioni doppie

RIVEDERE!

- Comportamento di due variabili insieme (x e y) → due o + variabili

**TABELLE di CONTINGENZA** o TABELLA a DOPPIA ENTRATA

- tabella con 2 RIGHE o 3 COLONNE  
I i                  J j

- FREQUENTE CONGIUNTE: sono freq. assolute di rilevazione di ogni coppia di modalità per le variabili x e y  
tali per cui si realizza contemporaneamente una certa modalità delle y e una certa modalità della x



- RIGA in BASSO: DISTRIBUZIONE MARGINALE della variabile x  
(somma di tutto quello che c'è sopra)



(rappresenta la freq. assoluta di rilevazioni raccolte in un certo sito)

- COLONNA A DESTRA: DISTRIBUZIONE MARGINALE della variabile y



(rappresenta i campioni contaminati e quelli no)

- N =  $\sum_{i=1}^I \sum_{j=1}^J n_{ij}$ : TOTALE OSSERVAZIONI (uguale <sup>summe</sup> colonne e <sup>summe</sup> righe)

**DISTRIBUZIONI CONDIZIONATE**: quando fissiamo una delle modalità dell'altra variabile

	x <sub>1</sub>	x <sub>2</sub>	...	x <sub>j</sub>	...	x <sub>J</sub>	TOTALE
y <sub>1</sub>	n <sub>11</sub>	n <sub>12</sub>	...	n <sub>1j</sub>	...	n <sub>1J</sub>	n <sub>1.</sub>
y <sub>2</sub>	n <sub>21</sub>	n <sub>22</sub>	...	n <sub>2j</sub>	...	n <sub>2J</sub>	n <sub>2.</sub>
...	...	...	...	...	...	...	...
y <sub>i</sub>	n <sub>i1</sub>	n <sub>i2</sub>	...	n <sub>ij</sub>	...	n <sub>iJ</sub>	n <sub>i.</sub>
...	...	...	...	...	...	...	...
y <sub>I</sub>	n <sub>I1</sub>	n <sub>I2</sub>	...	n <sub>Ij</sub>	...	n <sub>IJ</sub>	n <sub>I.</sub>
<b>TOTALE</b>	n <sub>.1</sub>	n <sub>.2</sub>	...	n <sub>.j</sub>	...	n <sub>.J</sub>	<b>N</b>

I mod. della y che definisce le I righe  
 distrib. MARGINALE della variabile x  
 distrib. MARGINALE della variabile y  
 FREQ. CONGIUNTE  
 TOT. OSSERVAZIONI

www.unidocs.it - Appunti e dispense per superare i tuoi esami universitari

www.unidocs.it - Appunti e dispense per superare i tuoi esami universitari

# ESERCITAZIONI 1°

1° Individuare numero delle unità statistiche, variabili e loro natura, scala, supporto

$N = 32$  (coppie di variabili)

Variabili  $X =$  numero di scarpe (taglia)  $\Rightarrow$  variabile quantitativa (xk è un numero)   
 $\Rightarrow$  variabile discreta (xk osserviamo le taglie)

supporto di questa variabile =  $\{34, 35, 36, \dots, 47\}$

Variabile  $Y =$  genere  $\Rightarrow$  variabile qualitativa su scala scemista

supporto M o F

• Di che tipo di variabile si tratta? Tratta di dati grezzi

Distribuzione statistica di aggregato

Matrice di dati ogni unità statistica primo passo x l'organizzazione di dati

Unità statistiche	x	y
1	42	M
2	.	.
3	.	.
...	.	.

$N \times 2$   
 $32 \times 2$

Dati dati nell'esempio

• **CRA:** iniziamo a sintetizzare: con riferimento all'informazione relative al numero di scarpe (la nostra  $X$ ), costruire la tabella di distribuzione di frequenza assoluta, relative, percentuale, cumulata assoluta, cumulata relative, cumulata %

considerazione

Spoglio dei dati

Modalità della variabile $x$	$n_i$	$f_i$	$f_i\%$	$N_i$ (freq. assoluta)	$N_i/N$	$F_i$ (cumulata ass)	$F_i\%$
34	1	0,03125		1			
35	1	0,03125		2			
36	1	0,03125		3			
37	2	0,0625		5			
38	2	0,0625		7			
39	2	0,0625		9			
40	4	0,125					
41	4	0,125					
42	5	0,15625					
43	4	0,125					
44	2	0,0625					
45	3	0,09375					
46	1	0,03125					
47	1	0,03125					

$32 = N$

2) Sintesi della tabella appena fatta: costruire la distribuzione in classi associata alla variabile numero di scarpe, dopo una scelta opportuna del numero delle classi.  
+ considerazioni

$$k(\text{numero di classi}) = \log_2 N + 1 = 6$$

X	$n_i$	$f_i\%$	$F_i\%$	$\bar{x}_i$ <small>x medio, valore centrale della classe</small>	$d_i$ <small>ampiezza delle classi</small>	$h_i$ <small>densità</small>
33,5 - 36,5	3	9,375%	9,375%	35	3	1
36,5 - 39,5	6	18,75%	28,125%	38	3	2
39,5 - 41,5	7	21,875%	50%	40,5	2	3,5
41,5 - 43,5	9	28,125%	78,125%	42,5	2	4,5
43,5 - 45,5	5	15,625%	93,75%	44,5	2	2,5
45,5 - 47,5	2	6,25%	100%	46,5	2	1

Sfruttiamo questa distribuzione in classi: Rispondiamo

A) Quale % di persone ha il numero di scarpe  $\leq 41$ ?  $F_3 = f_1 + f_2 + f_3 = 50\%$

B) Quale % di pax ha il numero di scarpe tra 41 escluso 45?

$$F_5 - F_3 = f_4 + f_5 = 43,75\%$$

C) Quale % di pax ha il numero di scarpe oltre il 45?

$$F_6 - F_5 = f_6 = 6,25\%$$

D) Quale % di pax ha il numero di scarpe almeno pari a 37?

$$100\% - 9,375 = 90,625\%$$

## ESERCIZIO 2

Mese dell'anno 1985	Incidenti stradali
Gennaio	17.401
Febbraio	17.568
Marzo	21.759
Aprile	22.235
Maggio	23.828
Giugno	24.645
Luglio	25.520
Agosto	20.968
Settembre	24.262
Ottobre	25.771
Novembre	21.594
Dicembre	22.173
<b>Totale</b>	<b>267.722</b>

① Unità statistica è il mese

Variabile è il numero di incidenti (quantitativa discreta)

$$N = 12$$

② Unità statistica sono gli incidenti

Variabile è il mese dell'incidente (qualitativa ordinale)

$$N = 267.722$$

## • Rappresentazioni grafiche (parte 1)

strumento indispensabile per visualizzare le principali caratteristiche delle distribuzioni di frequenza

Rendono possibile il confronto tra più distribuzioni in maniera efficace

mettono in risalto dati anomali, trend, picchi, ...

\* Rappresentano lo strumento preferibile per **diffondere i risultati** (specialmente a non statistici) deve essere **corretta e completa leggibile**, deve permettere di essere **chiaro ai lettori**

2 TIPI

x variabili qualitative

$$n_i = N \cdot f_i \cdot 360^\circ$$

$$x_i = \frac{n_i \cdot 360^\circ}{N} = f_i \cdot 360^\circ$$

Diagramma a nastri o barre

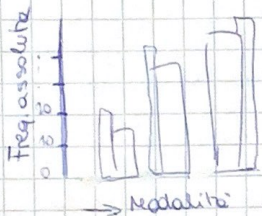
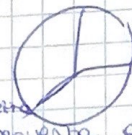


Grafico a settori circolari



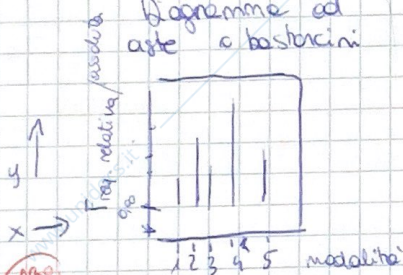
\* più usato

ogni settore ha un'ampiezza  $x_i$  proporzionale alle frequenze assolute della modalità  $x_i$  rilevata  $\rightarrow NB: + \text{unità} = + \text{cerchi}$

← **ATTENZIONE: NO ASSE DELLE ASCISSE!**

2 TIPI per variabili quantitative discrete

Diagrammi ad aste o bastoncini



\* attenzione agli assi dividere bene

Funzione di ripartizione empirica

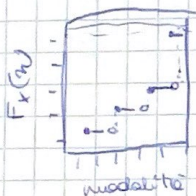
$$F_x(x) = \frac{\text{Numero di osservazioni} \leq x}{N}$$

- Frazione delle unità statistiche con modalità minore o uguale a un certo valore  $x$

- Dominio =  $\mathbb{R}$

- Codominio =  $[0, 1]$   $x_k$  è una frazione di unità statistiche

\* le modalità hanno posto in ordine crescente



SI DISSEGNA IL GRAFICO

x	$n_i$	$N_i$
modalità		

$$F_x(x) = \begin{cases} 0, & x < 1 \\ ; \\ 1, & x > 5 \end{cases}$$

- Funzione a scalari ovvero costante a tratti

- Scalari di altezza pari a  $n_i/N$  in corrispondenza dei valori del supporto di  $x$  salti tra un gradino e l'altro