

Statistica descrittiva

Monday, 20 March 2023

13:46

Prima di iniziare voglio dire questo: gli appunti del prof sono divini, DIVINI
Non ho mai visto appunti di un docente fatti così tanto bene

- Prendiamo i dati e li riassumiamo

Qui faccio subito degli esempi.

Dati x_1, x_2, \dots, x_N , N = numero dati

Mettendo caso che noi abbiamo un numero ridotto di valori distinti
E' possibile riassumerli attraverso una **tabella delle frequenze**.

Es.

Dati:

JUV, MIL, JUV, ATA, NES, NES, ATA, MIL, INT, NES,
ATA, ATA, NES, JUV, NES, INT, NES, JUV, NES, ATA

Proviamo a riassumerli:

Valori	Freq. Ass F_i	Freq. Rel p_i	%
$x = \text{juv}$	$X_{\text{num}} = 44$	$\frac{X_{\text{num}}}{N} = \frac{4}{20} = 0.20$	$0.20 * 100 = 20\%$

Ed ora si fa la stessa tabella per tutti i valori.

F_i = Numero volte numero compare

p_i = Frazione in cui compare

Questi sono **dati qualificativi**, dati quantitativi sono dati dove non abbiamo dei numeri ma delle stringhe/categorie

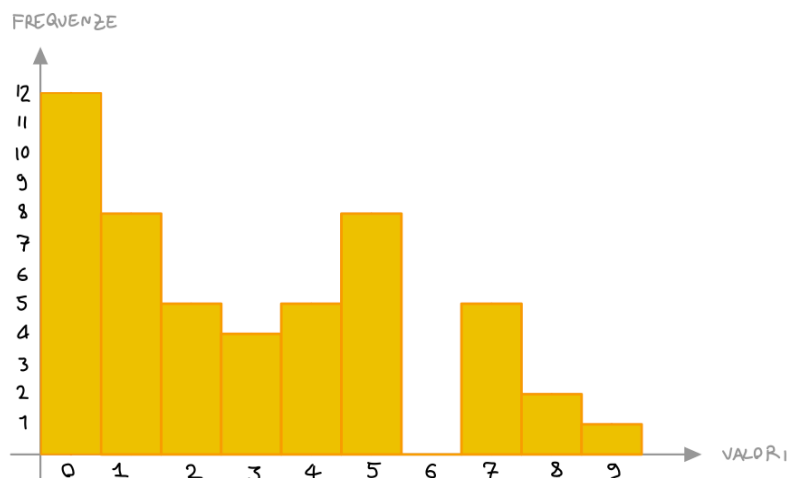
I **dati quantitativi** invece sono es:

2, 2, 0, 0, 5, 8, 3, 4, 1, 0, 0, 7, 1, 7, 1, 5,
4, 0, 4, 0, 1, 8, 9, 7, 0, 1, 7, 2, 5, 5, 4, 3,
3, 0, 0, 2, 5, 1, 3, 0, 1, 0, 2, 4, 5, 0, 5, 7, 5, 1

VALORI	FREQ. ASS. F_i	FREQ. REL. $p_i = F_i/N$
0	12	$12/50 = 0,24 = 24\%$
1	8	$8/50 = 0,16 = 16\%$
2	5	$5/50 = 0,1 = 10\%$
3	4	$4/50 = 0,08 = 8\%$

4	5	$5/50 = 0,1 = 10\%$
5	8	$8/50 = 0,16 = 16\%$
6	0	$0/50 = 0 = 0\%$
7	5	$5/50 = 0,1 = 10\%$
8	2	$2/50 = 0,04 = 4\%$
9	1	$1/50 = 0,02 = 2\%$

- Quando rappresentarli come tabella è bello, non lo è abbastanza
Quindi da questo si è creata la tabella delle frequenze, aka istogramma
Questo ci permette subito di comprendere qual è il valore più grande



Nota: l'istogramma è possibile: *istogramma* → *tabella frequenze*
e questo perché sono tutti e due equivalenti

- Mettiamo ora caso non abbiamo valori ridotti di numeri distinti
Ex.

213, 174, 193, 196, 220, 183, 194, 200, 192, 200,
200, 199, 178, 183, 188, 193, 187, 181, 193, 205,
196, 211, 202, 213, 216, 206, 195, 191, 171, 194,
184, 191, 221, 212, 221, 204, 204, 191, 183, 227

In questo caso è un bene lavorare per classi, aka raggruppamenti
Per farlo iniziamo a prendere min-max

$\min = 171$, $\max = 227$

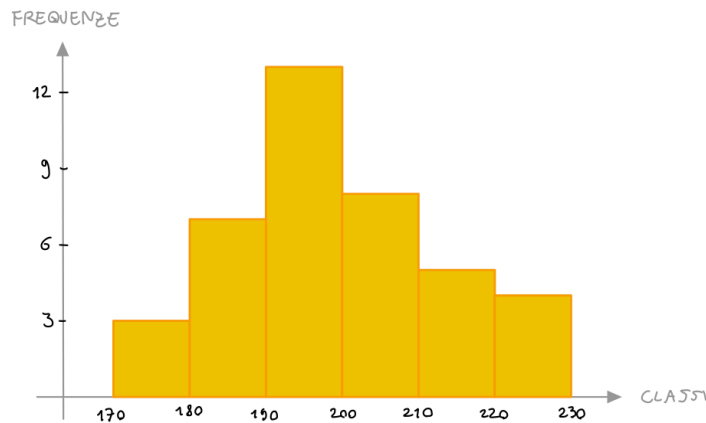
Ed ora da questi possiamo ricavare le classi

[170, 180), [180, 190), [190, 200), [200, 210), [210, 220), [220, 230)

E da questo è possibile fare ciò che abbiamo fatto prima

CLASSI	FREQ. ASS. f_i	FREQ. REL. $p_i = f_i/N$
[170, 180)	3	$3/40 = 0,075 = 7,5\%$

$[180, 190)$	7	$7/40 = 0,175 = 17,5\%$
$[190, 200)$	13	$13/40 = 0,325 = 32,5\%$
$[200, 210)$	8	$8/40 = 0,2 = 20\%$
$[210, 220)$	5	$5/40 = 0,125 = 12,5\%$
$[220, 230)$	4	$4/40 = 0,10 = 10\%$



Nota: lavorando per classi perdiamo delle informazioni

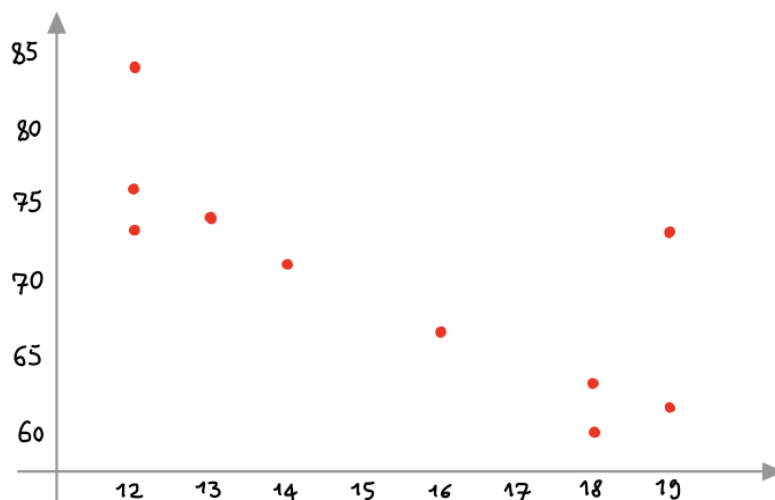
- Dati a coppie

Quando abbiamo più dati in coppia, ex $(x_1, y_1) \dots (x_N, y_N)$

Abbiamo 2 modi per gestirli:

- 1) Ignorare completamente uno dei due dati
Non una brutta idea, diciamo la scelta pigra, eh? fa per me!
- 2) Cercare una associazione, o correlazione, tra di esse

Parlando del caso 2, un metodo grafico è utilizzare il diagramma di dispersione

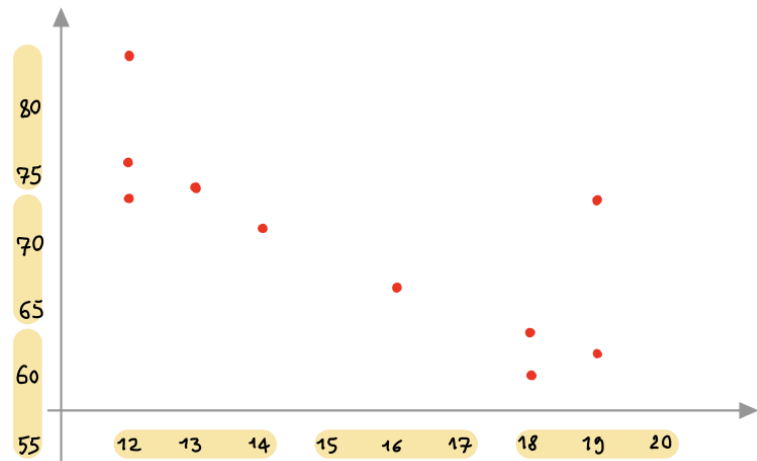


Guardandolo, ad occhio possiamo notare una correlazione negativa quindi, possiamo supporre che dopo 19 la probabilità è probabile che il numero continuerà a scendere

Parola chiave: è probabile, non è detto che succederà.

Detto questo, la frequenza assoluta/relativa per le coppie esiste

Per farlo dobbiamo immaginare di prendere in classi i nostri dati



Qui abbiamo evidenziato delle possibili classi x-y

X	[12, 15)	[15, 17)	[18, 21)
Y	[55, 65)	[65, 75)	[75, 85)

Facendo questo possiamo contare il numero di valori che ci sono dentro queste classi

(X, Y)	[55, 65)	[65, 75)	[75, 85)
[12, 15)	0	3	2
[15, 18)	0	1	0
[19, 21)	3	1	0

Detto questo, in futuro scopriremo come risolvere questo, questa ultima parte è stata totalmente inutile per il momento e mi sento onorato di avervi fatto perdere 1 minuto della vostra vita... Io ne ho perso 5 scrivendo questa parte