Esercizi (3)

ESERCIZIO 1. Ai 1000 abitanti di un piccolo comune viene richiesto di dare un giudizio sulla nuova biblioteca, assegnando un punteggio da 0 (0=pessimo) a 4 (4=ottimo). Le risposte sono riassunte nella tabella seguente

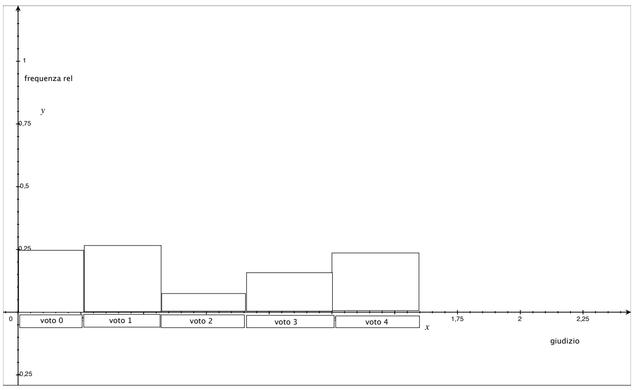
Giudizio	0	1	2	3	4
Frequenza	251	260	80	154	255

Cioe' 251 cittadini hanno dato voto 0, 260 voto 1 ecc.
In che percentuale gli abitanti sono molto contenti del servizio (voto 4) e contenti (voto 3)? Rappresentare in un istogramma delle frequenze relative le risposte dei cittadini.
Qual'e' il voto medio che ha ottenuto il servizio biblioteca? La moda e la mediana dei voti sono simili al voto medio? Commentare la risposta.
La media rappresenta con precisione i pareri dei cittadini?

Soluzione. La frequenza relativa del giudizio 4 (molto contento del servizio) si ottiene dal rapporto 255/1000=0.255$\approx$ 26%, mentre quella relativa del giudizio 3 (contento del servizio) si ottiene dal rapporto 154/1000=0.154$\approx$ 16%: si puo' concludere che un po' piu' di un quarto degli abitanti e' molto contento mentre un po' piu' del 15% e' contento.
In modo analogo si ottiene che da' un giudizio pessimo circa il 25% (0.251), un giudizio mediocre il 26%, un giudizio intermedio l'8%. (Si noti che la somma delle frequenze relative da' 1). La tabella delle giudizi e delle percentuali e' la seguente

Giudizio	0	1	2	3	4
Freq. relativa	25%	26	8%	15%	26%

L'istogramma delle frequenze e' il seguente



Il voto medio e'

$$M = \frac{0 \cdot 251 + 1 \cdot 260 + 2 \cdot 80 + 3 \cdot 154 + 4 \cdot 255}{1000} = 1.902 \approx 1.9.$$

La moda dei voti e' moda=1 visto che 260 cittadini (il numero piu' grande) ha espresso questo voto.
Per calcolare la mediana osserviamo che i dati sono ordinati in ordine crescente; sono in numero pari quindi la mediana e' la media dei dati che occupano la 500-esima e la 501-esima posizione. Dalla tabella si vede che la mediana e' 1.
La mediana e' uguale alla moda e entrambe sono diverse dalla media aritmetica: questo e' dovuto al fatto che circa la meta' dei dati e' concentrato nelle code della distribuzione (voto pessimo o voto molto buono).

Per sapere se la media riassume correttamente i pareri dei cittadini calcoliamo varianza e deviazione standard dei voti. Con un'approssimazione alla seconda cifra decimale la varianza vale

$$V = \frac{251(0-1.902)^2 + 260(1-1.902)^2 + 80(2-1.902)^2 + 154(3-1.902)^2 + 255(4-1.902)^2}{1000} \approx 2.43$$

e quindi la deviazione standard vale $\sigma = \sqrt{V} \approx 1.558$: questo significa che la votazione media approssima quella reale con un errore di 1.558 e quindi non e' molto fedele all'opinione dei votanti.

ESERCIZIO 2. Nello studio di un fenomeno, si rilevano le seguenti misure, in centimetri, con la frequenza riportata in corrispondenza sulla seconda riga

cm	5.1	2.3	4	1.2
F	7	4	1	3

Calcolare la media M, la moda e la mediana delle misure. Tra quelli trovati, qual'e' il miglior indice di tendenza centrale delle misure? (Giustificare la risposta)

Se alle precedenti misure viene aggiunta la misura m_5, che e' rilevata 3 volte, la media M diventa 3.45 cm. Quanto vale la misura m_5?

Dopo aver completato la tabella con m_5 e con la sua frequenza, aggiungere la riga con le frequenze relative delle misure e disegnare l'istogramma di queste ultime.

Soluzione. La moda delle misure è 5.1 perchè questo è il dato più frequente.
La media (aritmetica) dei dati è

$$M = \frac{5.1 \cdot 7 + 2.3 \cdot 4 + 4 + 1.2 \cdot 3}{15} = 3.5$$

mentre per calcolare la mediana dobbiamo ordinare i dati in ordine crescente

cm	1.2	2.3	4	5.1
F	3	4	1	7

Le misure sono in numero dispari, quindi al centro della distribuzione, al posto ottavo, c'è il valore 4, che è la mediana. Tra la media e la mediana c'è differenza di 0.5cm e tutti e due questi indici sono buoni indicatori della tendenza centrale della distribuzione visto che non c'è nessun dato molto grande o molto più piccolo degli altri. La misura 5.1 (la moda) non è invece un buon indice.

Se a quelle date si aggiunge la misura m_5 che ha frequenza 3, la tabella si modica nel modo seguente

cm	5.1	2.3	4	1.2	m_5
F	7	4	1	3	3

e la media è

$$M' = \frac{5.1 \cdot 7 + 2.3 \cdot 4 + 4 + 1.2 \cdot 3 + m_5 \cdot 3}{18}$$

Se deve essere

$$M' = 3.45 = \frac{5.1 \cdot 7 + 2.3 \cdot 4 + 4 + 1.2 \cdot 3 + m_5 \cdot 3}{18}$$

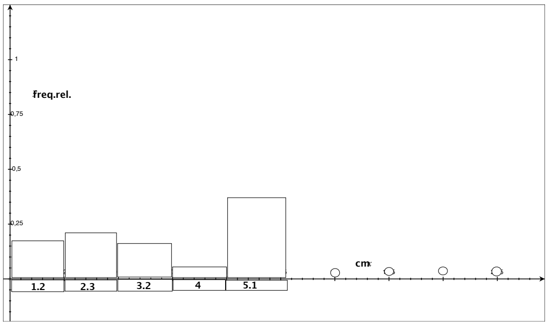
si può ricavare m_5 che deve valere

$$m_5 = \frac{62.1 - (5.1 \cdot 7 + 2.3 \cdot 4 + 4 + 1.2 \cdot 3)}{3} = 3.2.$$

La tabella con le frequenze relative, approssimate al decimale più vicino, sull'ultima riga diventa

cm	5.1	2.3	4	1.2	3.2
F	7	4	1	3	3
f	0.39	0.22	0.05	0.17	0.17

(**N.B**: la somma delle frequenze relative vale 0.39+0.22+0.05+2(0.17)=1). L'istogramma delle frequenze relative è



ESERCIZIO 3. Un'inchiesta giornalistica sull'occupazione riporta i seguenti dati sul numero di addetti per azienda manifatturiera:

n. addetti	n. aziende
[1,10]	50
[11,20]	60
[21,50]	45
[51,100]	30
[101,150]	15

cioe' 50 aziende hanno tra 1 e 10 addetti, 60 ne hanno tra 11 e 20 e cosi' via. Qual'e' il numero medio di addetti per azienda? Rappresentare il risultato nella tabella seguente

n. medio addetti	n. aziende
...	50
...	60
...	45
...	30
...	15

In media quanti addetti lavorano nelle aziende esaminate? Il dato medio e' rappresentativo del campione o no? (Giustificare la risposta)

Se nelle aziende considerate il fatturato medio e' di 500 000 euro con una deviazione standard del fatturato pari a 750 000 euro, riassumere a parole come si puo' interpretare il rapporto tra deviazione standard e media del fatturato. Che cosa si puo' dedurre dal confronto di questo rapporto tra con quello analogo tra deviazione standard del numero di addetti e numero medio di addetti?

Soluzione. Il valore medio (centrale) tra 1 e 10 e' 5 ((1+10)/2=5.5 che approssimato all'intero piu' vicino e' 5). Analogamente i valori medi tra 11 e 20, 21 e 50, 51 e 100, 101 e 150 sono, rispettivamente 15, 35,75, 125. Quindi la tabella si completa nel modo seguente

n. medio addetti	n. aziende
5	50
15	60
35	45
75	30
125	15

Il numero medio di addetti nelle 200 aziende considerate e'

$$M = \frac{50(5) + 60(15) + 45(35) + 30(75) + 15(125)}{200} = 34.25 \approx 34.$$

La varianza vale

$$V = \frac{50(5-34.25)^2 + 60(15-34.25)^2 + 45(35-34.25)^2 + 30(75-34.25)^2 + 15(125-34.25)^2}{200} \approx 1192.29$$

mentre la deviazione standard e' $\sigma = \sqrt{V} \approx 34.53$ e questo risultato dice che l'errore che commettiamo approssimando i dati con 34 e' di 34 unita' (il numero medio di addetti e' compreso tra 0 e 68 unita'). L'errore e' piuttosto grande e questo dipende dal fatto che non abbiamo il numero esatto di addetti per azienda ma il numero medio.
Il rapporto $\sigma_{fatt}/$(media fatturato), che vale 1.5 indica che l'oscillazione del fatturato e' uguale ad una volta e mezza il valore medio del fatturato. Si ha invece che $\sigma_{add}/$(media addetti) vale 34.53/34$\approx$ 1, e quindi l'oscillazione del numero addetti e' quasi uguale al numero medio di questi.
Dal confronto di questi due risultati possiamo desumere che, per le aziende considerate, la variazione del fatturato e' maggiore di quella del numero medio di addetti, in altre parole, la differenza tra le entrate delle aziende e' una variabile piu' importante di quella che studia la differenza del numero di dipendenti.

ESERCIZIO 4. Considerate le coppie di dati (18,30), (22,45), (14,28), (25,49), (21,38), discutere la possibilita' che esista una relazione tra i primi e i secondi elementi delle coppie. Scritta l'equazione della retta di regressione, si completino le seguenti coppie (20,...) e (23, ...).

Soluzione. Si deve discutere l'andamento congiunto dei dati

$$X: \quad 18 \; 22 \; 14 \; 25 \; 21$$

$$Y: \quad 30 \; 45 \; 28 \; 49 \; 38$$

La media dei dati X e' $M_X = 20$, quella dei dati Y e' $M_Y = 38$. La varianza dei dati X e' $V_X = 14$ e quindi la deviazione standard e' $\sigma_X = \sqrt{V_X} \approx 3.74$, mentre per i dati Y si ha $V_Y = 66.8$ e quindi la deviazione standard e' $\sigma_Y = \sqrt{V_Y} \approx 8.17$

La covarianza e'

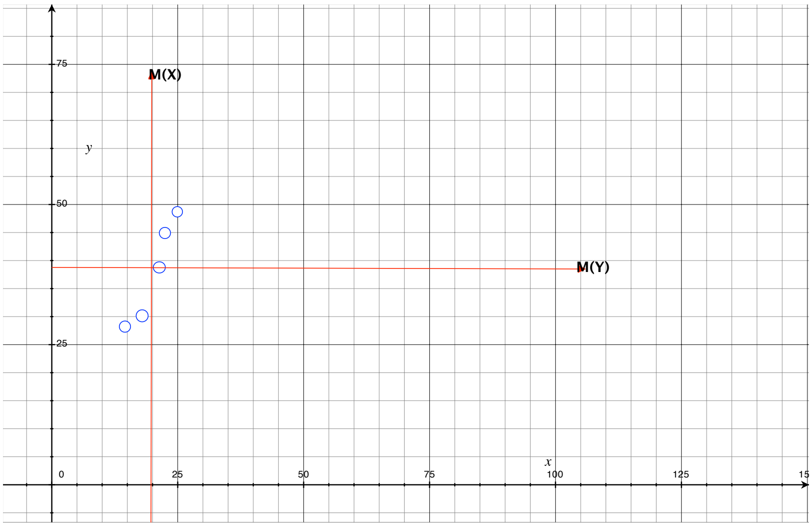
$$\sigma_{XY} = \frac{1}{5}[(18-20)(30-38) + (22-20)(45-38) + (14-20)(28-38) +$$

$$+ (21-20)(49-38) + (21-20)(38-38)] = 20.2 > 0$$

quindi i dati Y aumentano al crescere dei dati X. Questo risultato e' confermato se si osserva che $\sigma_X \sigma_Y \approx 30.56$ e quindi

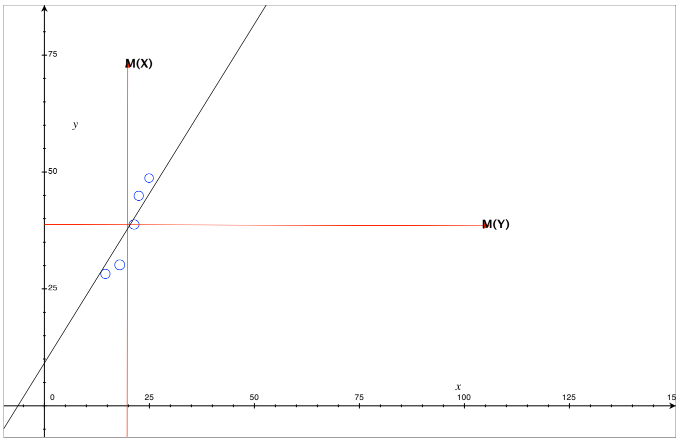
$$\rho = \frac{\sigma_{XY}}{\sigma_X \sigma_Y} \approx 0.66.$$

Una rappresentazione grafica dei dati conferma i precedenti risultati



Visto che $\sigma_X^2 = V_X = 14$, e che $\frac{\sigma_{XY}}{\sigma_X^2} = \frac{20.2}{14} \approx 1.44$, l'equazione della retta di regressione e'

$$y = \frac{\sigma_{XY}}{\sigma_X^2}x + \Big(M_Y - \frac{\sigma_{XY}}{\sigma_X^2}M_X\Big) \approx 1.44x + (38 - 28.86) = 1.44x + 9.14.$$



(si noti che la retta passa per il punto $(M_X, M_Y) = (20, 38)$).
I punti devono avere coordinate (20, 37.94) e (23, 42.26).