

# 1-INDICI05-frequenze condizionate, dipendenza in media

domenica 31 gennaio 2021 15:05

Bologna 2020

Nel mese di agosto 200 persone sono state intervistate sul numero di videochiamate fatte giornalmente durante il periodo di quarantena. Nella tabella seguente si riporta la distribuzione secondo i caratteri  $X$ =età e  $Y$ = numero di videochiamate

Nel mese di agosto 200 persone sono state intervistate sul numero videochiamate fatte giornalmente durante il periodo di quarantena. Nella tabella seguente si riporta la distribuzione secondo i caratteri  $X$ =età e  $Y$ = numero di videochiamate:

X	Y		
	0 - 1	2 - 5	6 - 8
15- 20	20	15	5
20- 50	15	40	35
50- 60	35	25	10

- Considerando la distribuzione marginale di  $Y$ , disegnare il box plot e commentare sul grafico l'eventuale asimmetria del carattere.
- Costruire la tabella di frequenze relative  $Y|X$ . Selezionare un valore e interpretarlo.
- Esclusivamente sulla base dei valori ottenuti (nei punti precedenti a e b) è possibile affermare che esiste dipendenza in media di  $Y$  da  $X$ ? (sì, no, perché...) E dipendenza lineare di  $Y$  da  $X$ ? E connessione tra  $X$  e  $Y$ ?
- Considerando la distribuzione marginale di  $Y$ , calcolare media, moda e mediana. Confrontando questi indici a quale conclusione sulla asimmetria si arriva? E' coerente con il commento del punto a)? Se non lo è, spiegarne il motivo.

x	y		
	0 - 1	2 - 5	6 - 8
15- 20	20	15	5
20- 50	15	40	35
50- 60	35	25	10

- Considerando la distribuzione marginale di  $Y$ , disegnare il box plot e commentare sul grafico l'eventuale simmetria
- Costruire la tabella di frequenze relative  $Y|X$ . Selezionare un valore e interpretarlo.
- Esclusivamente sulla base dei valori ottenuti (nei punti precedenti a e b) è possibile affermare che esiste dipendenza in media di  $Y$  da  $X$ ? (sì, no, perché...) E dipendenza lineare

- di Y da X?
- D) Considerando la distribuzione marginale di Y, calcolare media, moda e mediana. Confrontando questi indici a quale conclusione si arriva? È coerente con il commento del punto A)? Se non lo è spiegarne il motivo.

Costruiamo prima i totali marginali

x	Y			tot
	0 - 1	2 - 5	6 - 8	
15- 20	20	15	5	40
20- 50	15	40	35	90
50- 60	35	25	10	70
tot	70	80	50	200

Distribuzione marginale delle y

Isoliamo la distribuzione per vederla meglio

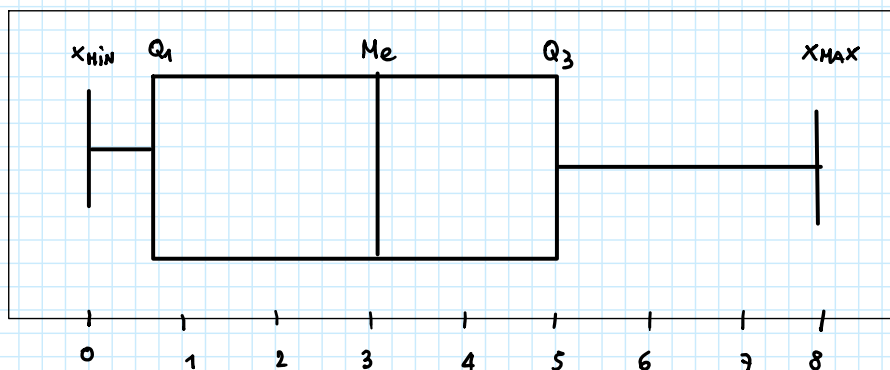
y	n <sub>j</sub>	N <sub>j</sub>	F <sub>j</sub>
0 - 1	70	70	0,35
2 - 5	80	150	0,75
6 - 8	50	200	1
tot	200		

$$Q_1 = x_{0,25} = 0 + \frac{0,25 - 0}{0,35 - 0} \cdot 1 = 0,714$$

$$Me = x_{0,50} = 2 + \frac{0,50 - 0,35}{0,75 - 0,35} \cdot 3 = 3,125$$

$$Q_3 = 5$$

BOX-PLOT



I dati vanno da un minimo di 0 ad un massimo di 8 chiamate.

La mediana è circa 3 ovvero il 50% degli individui effettua meno di 3 chiamate al giorno.

Il 50% degli intervistati effettua un numero di chiamate giornaliero tra approssimativamente tra 1 e 5.

A livello centrale sembra esserci una leggera asimmetria verso destra segnalata dal baffo a destra più lungo.

Il baffo corto del box-plot sulla sinistra indica che il 25% dei quelli che effettuano meno chiamate effettua meno di una chiamata al giorno. Mentre il baffo lungo a destra indica una maggior dispersione del numero delle chiamate effettuate dal 25% della popolazione che chiama di più.

A livello centrale sembra esserci un certo equilibrio.

B) Costruire la tabella di frequenze relative  $Y|X$ . Selezionare un valore e interpretarlo.

x	y			tot
	0 - 1	2 - 5	6 - 8	
15- 20	20	15	5	40
20- 50	15	40	35	90
50- 60	35	25	10	70
tot	70	80	50	200

x	0 - 1	2 - 5	6 - 8	tot
15- 20	0,50	0,38	0,13	1
20- 50	0,17	0,44	0,39	1
50- 60	0,50	0,36	0,14	1

Il 38% della popolazione nella fascia di età dai 15 ai 20 anni ha effettuato un numero di chiamate giornaliere da 2 a 5

C) Esclusivamente sulla base dei valori ottenuti (nei punti precedenti a e b) è possibile affermare che esiste dipendenza in media di Y da X? (sì, no, perché...) E dipendenza lineare di Y da X?

x	0 - 1	2 - 5	6 - 8	tot
15- 20	0,50	0,38	0,13	1
20- 50	0,17	0,44	0,39	1
50- 60	0,50	0,36	0,14	1

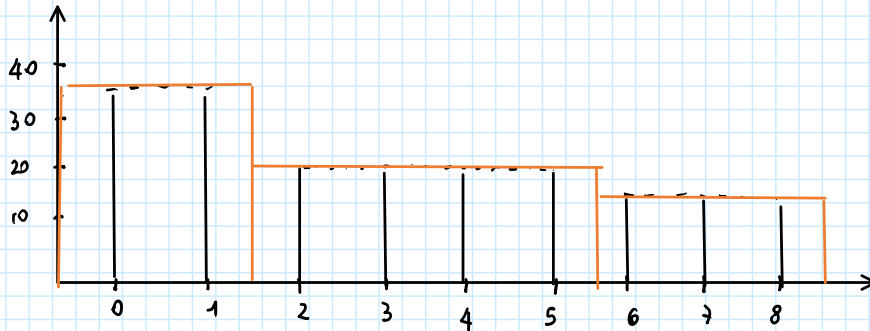
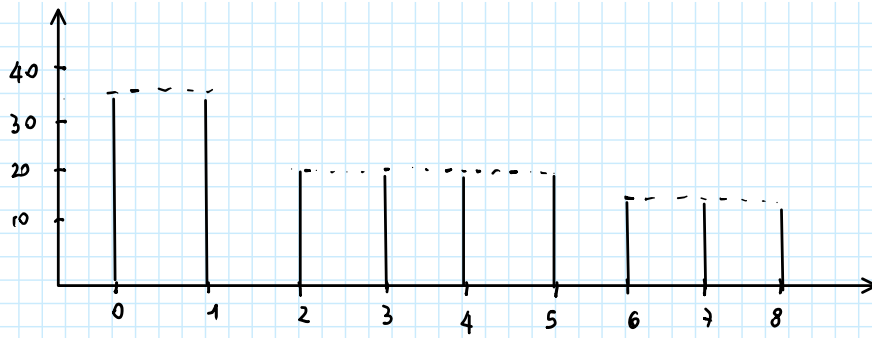
Si è possibile affermare che c'è una dipendenza in media del numero delle chiamate (y) rispetto all'età poiché le percentuali delle varie fasce del numero di chiamate cambia rispetto all'età degli individui considerati

D) Considerando la distribuzione marginale di Y, calcolare media, moda e mediana. Confrontando questi indici a quale conclusione si arriva? È coerente con il commento del punto A)? Se non lo è spiegarne il motivo.

y	ni	xmin	xmax	ci	ai	hi	fi	Fi
0 - 1	70	-0,5	1,5	0,5	2	35	0,35	0,35
2 - 5	80	1,5	5,5	3,5	4	20	0,4	0,75
6 - 8	50	5,5	8,5	7	3	16,667	0,25	1
totali	200							

Idea: siccome abbiamo fatto diventare il numero delle chiamate un carattere continuo possiamo trattarlo come tale.

L'escamotage è quello di ampliare di mezza unità le classi



y	ni	xmin	xmax	ci	ai	hi	fi	Fi
0   1	70	-0,5	1,5	0,5	2	35	0,35	0,35
2   5	80	1,5	5,5	3,5	4	20	0,4	0,75
6   8	50	5,5	8,5	7	3	16,667	0,25	1
totali	200							

Media:

$$\bar{x} = \frac{\sum c_i \cdot m_i}{N} = \frac{0,5 \cdot 70 + 3,5 \cdot 80 + 7 \cdot 50}{200} = 3,325$$

Moda

Guardiamo la classe con altezza relativa maggiore

Ovvero quella da 0 a 1.

Quindi assumiamo come mediana il valore centrale della classe

Mediana = 0,5

Mediana:

$$M_e = X_{\min M_e} + \frac{0,5 - F_{M_e-1}}{F_{M_e} - F_{M_e-1}} \cdot \omega_{M_e}$$

$$= 1,5 + \frac{0,5 - 0,35}{0,75 - 0,35} \cdot 4$$

Anche se i tre indici non coincidono sono molto vicini tra di loro.

Quindi si potrebbe concludere di trovarsi di fronte ad una distribuzione tendenzialmente simmetrica.