Università degli Studi di Cassino
Anno accademico 2003-2004
Corsi di Statistica 1, II (Prof. G. Prozio) e Statistica 1, IV (Dott. D. Vistocco)

Esercitazione del 2/2/2004
Dott. Claudio Conversano

Esercizio 1
Si consideri la matrice dei dati di seguito riportata, relativa a 15 lavoratori dell'azienda "X" sui quali sono stati rilevati 4 caratteri (Genere, Inquadramento, Annualità scolastiche e Stipendio annuo in migliaia di Euro).

Genere	Inquadramento	Annualità scolastiche	Stipendio annuo
uomo	Dirigente	15	57,0
uomo	Dirigente	16	40,20
donna	Impiegato	12	21,450
donna	Impiegato	8	21,90
uomo	Dirigente	15	45,0
uomo	Funzionario	15	32,10
uomo	Funzionario	15	36,0
donna	Impiegato	12	21,90
donna	Impiegato	15	27,90
donna	Impiegato	12	24,0
donna	Funzionario	16	30,30
uomo	Impiegato	8	28,350
uomo	Impiegato	15	27,750
donna	Funzionario	15	35,10
uomo	Impiegato	12	27,30

a) Per ognuno dei caratteri osservati costruire la distribuzione di frequenza e calcolare gli indici di posizione.
b) Relativamente al carattere "Stipendio annuo" suddividere la relativa distribuzione dapprima in 3 classi equiampie e poi in 3 classi equifrequenti. Calcolare gli indici di posizione sulle due distribuzioni in classi.

Svolgimento

Punto a)

Il primo carattere osservato +è il carattere "**genere**" Trattasi di un carattere qualitativo nominale. Per tali caratteri l'unico indice di posizione calcolabile è la moda.

Distribuzione di frequenza:

Modalità	Frequenze assolute	Frequenze relative
Uomo	8	0,53
Donna	7	0,47
Totale	**15**	**1,00**

La moda (modalità a cui corrisponde la frequenza più elevata) corrisa ponde alla modalità "uomo".

Il secondo carattere osservato è il carattere "**inquadramento**". Si tratta di un carattere qualitativo ordinale, in quanto è possibile ordinare le 3 modalità ad esso relative partendo dall'inquadramento inferiore ("impiegato") fino ad arrivare a quello principale ("dirigente"). Per i caratteri qualitativi ordinali gli indici che si possono calcolare sono la moda ed i quantili (quartili, decili, percentili).

Distribuzione di frequenza

Modalità	Frequenze assolute	Frequenze relative	Frequenze relative cumulate
Impiegato	8	0,53	0,53
Funzionario	4	0,27	0,80
Dirigente	3	0,20	1,00
Totale	15	1,00	

La moda è uguale alla modalità "Impiegato".

Per il calcolo dei quantili si può considerare la funzione di ripartizione empirica (corrispondente alla frequenza relativa cumulata). La regola da impiegare è la seguente:

- La mediana è uguale alla modalità in corrispondenza della quale la funzione di ripartizione empirica passa (anche idealmente) per il punto 0,5;
- Il primo quartile è uguale alla modalità in corrispondenza della quale la funzione di ripartizione empirica passa (anche idealmente) per il punto 0,25;
- Il terzo quartile è uguale alla modalità in corrispondenza della quale la funzione di ripartizione empirica passa (anche idealmente) per il punto 0,75;
- Il primo decile è uguale alla modalità in corrispondenza della quale la funzione di ripartizione empirica passa (anche idealmente) per il punto 0,10;
- Il primo percentile è uguale alla modalità in corrispondenza della quale la funzione di ripartizione empirica passa (anche idealmente) per il punto 0,01;
- …….. e così via.

Pertanto:

- Me = "Impiegato", perché in corrispondenza di tale modalità la funzione di ripartizione empirica passa da zero a 0,53 (quindi passa idealmente per il punto 0,5)
- Q_1 = "Impiegato", perché in corrispondenza di tale modalità la funzione di ripartizione empirica passa da zero a 0,53 (quindi passa idealmente per il punto 0,25)
- Q_3 = "Funzionario", perché in corrispondenza di tale modalità la funzione di ripartizione empirica passa da 0,53 a 0,80 (quindi passa idealmente per il punto 0,75)

Il terzo carattere osservato è il carattere "**Annualità Scolastiche**", inteso come numero di anni di studio (ad esempio per una persona che ha frequentato le scuole elementari (5 anni), le scuole medie inferiori (3 anni) e le scuole medie superiori (5 anni) il numero di annualità scolastiche sarà pari a

13). Si tratta di un carattere quantitativo discreto. Per tali caratteri è possibile calcolare tutti gli indici di posizione (moda, quantili e media aritmetica).

Distribuzione di frequenza:

Modalità	**Frequenze assolute**	**Frequenze relative**	**Frequenze relative cumulate**
8	2	0,13	0,13
12	4	0,27	0,40
15	7	0,47	0,87
16	2	0,13	1,00
Totale	15	1,00	

Moda= 15

Me= 15 *Q_1=12* *Q_3=15*

$$\mu_X = \frac{1}{n}\sum_{i=1}^{k} x_i n_i = \frac{1}{15}\sum_{i=1}^{4} x_i n_i = \frac{1}{15}\left(8\cdot 2+12\cdot 4+15\cdot 7+16\cdot 2\right)=13,4$$

Il quarto carattere osservato è lo stipendio annuo in migliaia di Euro. Si tratta di un carattere quantitativo continuo. Anche per tale carattere è possibile calcolare tutti gli indici di posizione. A tal fine, è conveniente ordinare in senso non decrescente i dati grezzi.

21,45; 21,90; 21,90; 24,00; 27,30; 27,75; 27,90; 28,35; 30,30; 32,10; 35,10; 36,00; 40,20; 45,00; 57,00

Per tale distribuzione la Moda è uguale a 21,90 (unica intensità che si presenta 2 volte nella successione ordinata).

Inoltre:

$$\mu_X = \frac{1}{n}\sum_{i=1}^{n} x_i = \frac{1}{15}\sum_{i=1}^{15} x_i = \frac{1}{15}(21{,}45+21{,}90+21{,}90+24{,}00+27{,}30+27{,}75+27{,}90+$$
$$+28{,}35+30{,}30+32{,}10+35{,}10+36{,}00+40{,}20+45{,}00+57{,}00) = 31{,}75$$

$$Me = x_{\left(\frac{n+1}{2}\right)} = x_{\left(\frac{15+1}{2}\right)} = x_{(8)} = 28{,}35$$

Il primo quartile può essere calcolato come la mediana della prima metà delle osservazioni (dalla prima alla settima), ossia:

$$Q_1 = Me\{21{,}45;\ 21{,}90;\ 21{,}90;\ 24{,}00;\ 27{,}30;\ 27{,}75;\ 27{,}90\} =$$
$$= x_{\left(\frac{n+1}{2}\right)} = x_{\left(\frac{7+1}{2}\right)} = x_{(4)} = 24{,}00$$

Il terzo quartile può essere calcolato come la mediana della seconda metà delle osservazioni (dalla ottava alla quindicesima), ossia:

$$Q_3 = Me\{30{,}30;\ 32{,}10;\ 35{,}10;\ 36{,}00;\ 40{,}20;\ 45{,}00;\ 57{,}00\} =$$
$$= x_{\left(\frac{n+1}{2}\right)} = x_{\left(\frac{7+1}{2}\right)} = x_{(4)} = 36{,}00$$

Punto b)

Come è noto, le modalità di un carattere quantitativo continuo possono essere suddivise in classi, per cui si può considerare la distribuzione in classi e calcolare su tale distribuzione gli indici di posizione. Il risultato ottenuto per tali indici non sarà un risultato preciso bensì approssimato, in quanto al variare del numero delle classi e del criterio di suddivisione in classi cambierà il valore degli indici di posizione.

3 classi equiampie:

Ampiezza delle classi: $d = \frac{x_{\max} - x_{\min}}{k} = \frac{57 - 21{,}45}{3} = 11{,}85$

Distribuzione in classi

Classe	Frequenze assolute	Frequenze relative	Frequenze relative cumulate	Densità di frequenza	Valori centrali
21,45-|33,30	10	0,66	0,66	0,056	27,375
33,30-|45,15	4	0,27	0,93	0,022	39,225
45,15-|57,00	1	0,07	1,00	0,006	51,075
Totale	15	1,00			

La moda corrisponde alla classe che presenta la densità di frequenza più elevata, ossia la classe 21,45-| 33,30.

La media aritmetica si calcola considerando i valori centrali

$$\mu_X = \frac{1}{n}\sum_{i=1}^{k} c_i n_i = \frac{1}{15}\sum_{i=1}^{4} c_i n_i = \frac{1}{15}\left(27{,}375 \cdot 10 + 39{,}225 \cdot 4 + 51{,}075 \cdot 1\right) = 32{,}115$$

I quantili si calcolano applicando le formule per le distribuzioni in classi.
Per la mediana, bisogna individuare dapprima la classe mediana, ossia la classe in corrispondenza della quale la funzione di ripartizione empirica passa (anche idealmente) per il punto 0,5. In questo caso la classe mediana è la prima classe, perché passando dai valori minori di 21,45 fino a 33,30 la funzione di ripartizione empirica passa da zero a 0,66, e quindi passa idealmente per il punto 0,5. Successivamente, si applica la formula seguente che considera gli estremi della classe mediana ed i valori della funzione di ripartizione empirica della classe mediana e della classe ad essa precedente:

$$Me = x_{i-1} + (x_i - x_{i-1})\frac{0,5 - F_{i-1}}{F_i - F_{i-1}} = 21,45 + (33,30 - 21,45)\frac{0,5-0}{0,66-0} = 30,34$$

Per il calcolo del primo quartile si procede allo stesso modo, tenendo conto che bisogna individuare dapprima la classe che contiene il primo quartile, ossia la classe in corrispondenza della quale la funzione di ripartizione empirica passa (anche idealmente) per il punto 0,25. In questo caso la classe che contiene il primo quartile è la prima classe, perché passando dai valori minori di 21,45 fino a 33,30 la funzione di ripartizione empirica passa da zero a 0,66, e quindi passa idealmente per il punto 0,25. Successivamente, si applica la formula seguente che considera gli estremi della classe che contiene il primo quartile ed i valori della funzione di ripartizione empirica della classe che contiene il primo quartile e della classe ad essa precedente:

$$Q_1 = x_{i-1} + (x_i - x_{i-1})\frac{0,25 - F_{i-1}}{F_i - F_{i-1}} = 21,45 + (33,30 - 21,45)\frac{0,25-0}{0,66-0} = 25,89$$

Per il calcolo del terzo quartile si procede allo stesso modo, tenendo conto che bisogna individuare dapprima la classe che contiene il terzo quartile, ossia la classe in corrispondenza della quale la funzione di ripartizione empirica passa (anche idealmente) per il punto 0,75. In questo caso la classe che contiene il terzo quartile è la seconda classe classe, perché passando dai valori minori di 33,30 fino a 45,15 la funzione di ripartizione empirica passa da 0,66 a 0,93, e quindi passa idealmente per il punto 0,75. Successivamente, si applica la formula seguente che considera gli estremi della classe che contiene il terzo quartile ed i valori della funzione di ripartizione empirica della classe che contiene il terzo quartile e della classe ad essa precedente:

$$Q_3 = x_{i-1} + (x_i - x_{i-1})\frac{0,75 - F_{i-1}}{F_i - F_{i-1}} = 33,30 + (45,15 - 33,30)\frac{0,75-0,66}{0,93-0,66} = 37,00$$

3 classi equifrequenti:

Frequenza delle classi: $n_i = \frac{n}{k} = \frac{15}{3} = 5$

Distribuzione in classi

Classe	Frequenze assolute	Frequenze relative	Frequenze relative cumulate	Ampiezza della classe	Densità di frequenza	Valori centrali
21,45-|27,30	5	0,33	0,33	5,85	0,057	24,375
27,30-|32,10	5	0,33	0,67	4,80	0,069	29,70
32,10-|57,00	5	0,33	1,00	24,90	0,013	44,55
Totale	15	1,00				

La moda corrisponde alla classe che presenta la densità di frequenza più elevata, ossia la classe 27,30-| 32,10.

$$\mu_X = \frac{1}{n}\sum_{i=1}^{k} c_i n_i = \frac{1}{15}\sum_{i=1}^{4} c_i n_i = \frac{1}{15}(24,37 \cdot 5 + 29,70 \cdot 5 + 44,55 \cdot 5) = 32,875$$

$$Me = x_{i-1} + \left(x_i - x_{i-1}\right)\frac{0,5 - F_{i-1}}{F_i - F_{i-1}} = 27,30 + \left(32,10 - 27,30\right)\frac{0,5 - 0,33}{0,67 - 0,33} = 29,70$$

$$Q_1 = x_{i-1} + \left(x_i - x_{i-1}\right)\frac{0,25 - F_{i-1}}{F_i - F_{i-1}} = 21,45 + \left(27,30 - 21,45\right)\frac{0,25 - 0}{0,33 - 0} = 25,84$$

$$Q_3 = x_{i-1} + \left(x_i - x_{i-1}\right)\frac{0,75 - F_{i-1}}{F_i - F_{i-1}} = 32,10 + \left(57 - 32,10\right)\frac{0,75 - 0,67}{1 - 0,67} = 38,32$$