STATISTICA 1 ESERCITAZIONE 4

Dott. Giuseppe Pandolfo

21 Ottobre 2013

Percentili: i valori che dividono la distribuzione in cento parti di uguale numerosità.

<u>Esercizio 1</u>

La seguente tabella riporta la distribuzione della variabile *Numero di libri* letti quest'anno per 319 unità statistiche:

Numero di libri	n_i
1	29
2	31
3	33
4	56
5	55
6	44
7	30
8	22
9	19
Totale	

Calcolare la moda, la mediana, il primo e il non decile, il 73-esimo centile della distribuzione.

<u>Soluzione</u>

Calcoliamo le frequenze assolute cumulate:

Numero di libri	n_i	N_i
1	29	29
2	31	60
3	33	93
4	56	149
5	55	204
6	44	248
7	30	278
8	22	300
9	19	319
Totale	319	

La moda è 4.

N è dispari:

$$Mediana = X_{\left(\frac{n+1}{2}\right)} = X_{\left(\frac{319+1}{2}\right)} = X_{160}$$

Quindi in base alle frequenze cumulate la mediana è 5.

La posizione del primo decile è data da:

$$D_1 = X_{\left(\frac{n+1}{10}\right)} = X_{32}$$

Quindi $D_1 = 2$.

La posizione del nono decile è data da:

$$D_9 = X_{\left(9 \times \frac{n+1}{10}\right)} = X_{288}$$

Quindi $D_9 = 8$.

La posizione del 73-esimo centile è data da:

$$C_{73} = X_{\left(73 \times \frac{n+1}{100}\right)} = X_{233,6}$$

Quindi $C_{73} = 6$.

MEDIA ARITMETICA

Utilizzata per variabili quantitative.

$$\mu_X = \mu = \frac{\sum_{i=1}^{n} x_i}{n}$$

Proprietà:

1) La somma dei valori x_i assunti da n unità statistiche è uguale al valor medio moltiplicato per il numero di unità:

$$\sum_{i=1}^{n} x_i = n\mu$$

2) *(Baricentricità)* La somma delle differenze tra i valori delle x_i e la loro media aritmetica μ è uguale a zero:

$$\sum_{i=1}^{n}(x_i - \mu) = 0$$

3) La somma degli scarti al quadrato dei valori x_i da una costante *c* è minima quando c = μ:

$$\sum_{i=1}^{n}(x_i - c)^2 \text{ è } minima\ per\ c = \mu$$

4) *(Associatività)* Se un collettivo *n* unità statistiche viene diviso in *M* sottoinsiemi disgiunti di numerosità $n_1,\ n_2,\ \ldots, n_M$, tali che

$$\sum_{h=1}^{n} n_h = n \qquad con\ media\ \mu_1, \mu_2, \ldots, \mu_M$$

allora la media aritmetica generale μ è una media ponderata delle medie dei sottoinsiemi con pesi uguali alla loro numerosità:

$$\mu = \frac{1}{n}\sum_{h=1}^{M}\mu_h\, n_h$$

5) *(Internalità)* La media aritmetica è sempre compresa tra il valore minimo e il valore massimo assunto dalle modalità della distribuzione:

$$x_{min} \leq \mu \leq x_{max}$$

6) *(Linearità)* Data la distribuzione di una variabile *X* con media μ, se moltiplichiamo ogni modalità per una costante *a* e aggiungiamo una costante *b*, la media della distribuzione sarà uguale a $a\mu + b$.

Esercizio 2

Di seguito è riportata la tabella dei dati relativi al numero di vani di 100 appartamenti:

Numero di vani	n_i
1	20
2	19
3	11
4	22
5	15
6	13
Totale	100

Calcoliamo la media aritmetica per la variabile *Numero di vani*.

$$\mu = \frac{\sum_{i=1}^{n} x_i n_i}{n}$$

$$\mu = \frac{(1 \times 20) + (2 \times 19) + (3 \times 11) + (4 \times 22) + (5 \times 15) + (6 \times 13)}{100} = \frac{332}{100} = 3{,}32$$

La media è 3,32.

Verifichiamo le proprietà di internalità, linearità e baricentricità.

Internalità

min(Numero di vani) = 1

max(Numero di vani) = 6

$$1 \leq 3{,}32 \leq 6$$

Linearità

Moltiplichiamo le modalità della variabile *Numero di vani* per $a = 2$ e aggiungiamo poi una costante $b = 3$:

Numero di vani	n_i
5	20
7	19
9	11
11	22
13	15
15	13
Totale	100

$$\mu(aX + b) = \frac{(5 \times 20) + (7 \times 19) + (9 \times 11) + (11 \times 22) + (13 \times 15) + (15 \times 13)}{100} = 9{,}64$$

$$a\mu + b = 2(3{,}32) + 3 = 9{,}64$$

Baricentricità

$$20 \times (1 - 3{,}32) + 19 \times (2 - 3{,}32) + 11 \times (3 - 3{,}32) + 22 \times (4 - 3{,}32) + 15 \times (5 - 3{,}32) +$$

$$13 \times (6 - 3{,}32) = -46{,}4 - 25{,}08 - 3{,}52 + 14{,}96 + 25{,}2 + 34{,}84 = 0$$

<u>Esercizio 3</u>

La seguente tabella riporta la spesa bimestrale in euro per energia elettrica per 73 appartamenti:

Classe	n_i
[0, 100)	22
[100, 200)	13
[200, 400)	24
[400, 600)	11
[600, 1000)	3
Totale	73

Calcoliamo la media aritmetica per la spesa bimestrale di energia elettrica.

<u>Soluzione</u>

Per dati raggruppati in classi la media aritmetica viene calcolata ipotizzando che per ogni classe la frequenza si concentri nel valore centrale della classe stessa $\hat{x}_i = \frac{min+max}{2}$.

Classe	n_i	$\hat{x}_i$
[0, 100)	22	50
[100, 200)	13	150
[200, 400)	24	300
[400, 600)	11	500
[600, 1000)	3	800
Totale	73	

$$\mu = \frac{(22 \times 50) + (13 \times 150) + (24 \times 300) + (11 \times 500) + (3 \times 800)}{73} = \frac{18150}{73} = 248{,}6$$

La media è 248,6 euro.

Esercizio 4

In base ai dati riportati nella tabella seguente, determiniamo il rango percentile corrispondente al valore 450 utilizzando le frequenze cumulate e le densità di frequenza.

Classe	n_i
[0, 100)	10
[100, 200)	15
[200, 400)	24
[400, 600)	31
[600, 1000)	37
Totale	117

La classe Mediana è [400, 600).

Soluzione

Calcoliamo le frequenze assolute cumulate e le densità di frequenza:

Classe	$\boldsymbol{n_i}$	$\boldsymbol{N_i}$	$\boldsymbol{d_i}$
[0, 100)	10	10	0,1
[100, 200)	15	25	0,25
[200, 400)	24	49	0,12
[400, 600)	31	80	0,155
[600, 1000)	37	117	0,0925
Totale	117		

$$N_{i-1} + \frac{n_i}{x_i - x_{i-1}}(450 - x_{i-1}) = 49 + \frac{31}{600 - 400}(450 - 400) = 56{,}75$$

La posizione è ≈ 57.

In termini percentili

$$\frac{N_{450}}{n} \times 100 = \frac{56{,}75}{117} \times 100 = 48{,}5 \approx 48$$

Quarantottesimo percentile.

Usando le densità di frequenza:

$$(450 - 400) \times d_4 = 50 \times \frac{31}{600 - 400} = 50 \times 0{,}155 = 7{,}75 \approx 8$$

La posizione è $\approx 57\ (49 + 8)$.

MEDIANA

Proprietà:

1) La mediana è applicabile anche a variabili qualitative ordinali poiché per definizione richiede solamente che i termini siano ordinabili.
   Esempio: fascia di reddito (bassa, medio - bassa, media, medio - alta, alta).

2) La mediana non varia se i valori minori (maggiori) di essa vengono sostituiti da altri valori comunque minori (maggiori). La media aritmetica, invece, è maggiormente influenzata da valori anomali (non è robusta).

   Esempio: dati relativi al peso di 5 persone

   70, 65, 88, 90, 52

   serie ordinata → 52, 65, 70, 88, 90

   La mediana è 70

   La media è 73

   Se sostituiamo i valori estremi con altri valori:

   60, 65, 70, 88, 98

   La mediana è 70

   La media è 76,2

3) La somma degli scarti in valore assoluto delle modalità della variabile da una costante *c* è minima quando *c* è la mediana:

$$\sum_{i=1}^{n} |x_i - c| \quad \text{è } minima\ per\ c = Me.$$

INDICI DI VARIABILITA' PER VARIABILI QUANTITATIVE

Indice di mutua variabilità

Senza ripetizione:

$$\frac{\sum_{i=1}^{n} \sum_{j=1}^{n} d(x_i - x_j)}{n(n-1)}$$

Con ripetizione:

$$\frac{\sum_{i=1}^{n} \sum_{j=1}^{n} d(x_i - x_j)}{n^2}$$

Due possibili scelte:

- Differenza semplice $\quad d(x_i, x_j) = |x_i - x_j|$
- Differenza quadratica $\quad d(x_i, x_j) = (x_i - x_j)^2$

Da cui

differenza semplice media

$$\Delta_r = \frac{\sum_{i=1}^{n} \sum_{j=1}^{n} |x_i - x_j|}{n^2}$$

differenza quadratica media

$$\Delta_r{}^2 = \sqrt{\frac{\sum_{i=1}^{n} \sum_{j=1}^{n} \left(x_i - x_j\right)^2}{n^2}}$$

sostituendo n^2 con $n(n-1)$ otteniamo la differenza media semplice e la differenza quadratica media senza ripetizione.

<u>**Esercizio 5**</u>

La seguente tabella di frequenze per singole modalità riporta la variabile *Peso* rilevata per 80 unità.

Peso	n_i
70	27
90	52
100	1
Totale	80

Calcolare la differenza semplice media e la differenza media quadratica.

<u>Soluzione</u>

La variabilità di un fenomeno può essere studiata in termini di differenza di ciascun dato da tutti gli altri. Tutte le differenze possibili compresa la differenza dell'i-esimo valore da se stesso è una matrice *n x n*. Nel nostro caso:

	c_j		
c_i	**70**	**90**	**100**
70	0	- 20	- 30
90	20	0	- 10
100	30	10	0

La seguente tabella riporta i prodotti $n_i \times n_j$

	n_j		
n_i	**27**	**52**	**1**
27	729	1404	27
52	1404	2704	52
1	27	52	1

<u>Per tabelle di frequenza</u>

differenza semplice media

$$\Delta_r = \frac{\sum_{i=1}^{n} \sum_{j=1}^{n} \left| x_i - x_j \right| n_i n_j}{n^2}$$

differenza quadratica media

$$\Delta_r{}^2 = \sqrt{\frac{\sum_{i=1}^{n} \sum_{j=1}^{n} \left(x_i - x_j \right)^2 n_i n_j}{n^2}}$$

Sostituendo n^2 con $n(n-1)$ otteniamo la differenza media semplice e la differenza quadratica media senza ripetizione. Nel caso di dati in classi sostituiamo x_i con i corrispondenti centri $\widehat{x_i}$.

Differenza semplice media con ripetizione

$$\Delta_r = \frac{(20 \times 1404) + (30 \times 27) + (20 \times 1404) + (10 \times 52) + (30 \times 27) + (10 \times 52)}{80^2} = \frac{58820}{6400} = 9{,}19$$

Differenza quadratica media con ripetizione

$$\Delta_r{}^2 = \sqrt{\frac{[(-20)^2 \times 1404] + [(-30)^2 \times 27] + [20^2 \times 1404] + [(-10)^2 \times 52] + [30^2 \times 27] + [10^2 \times 52]}{80^2}}$$

$$= \sqrt{\frac{1182200}{6400}} = 13{,}59$$

Differenza semplice media senza ripetizione

$$\Delta_{sr} = \frac{(20 \times 1404) + (30 \times 27) + (20 \times 1404) + (10 \times 52) + (30 \times 27) + (10 \times 52)}{80(80-1)} = \frac{58820}{6320} = 9,3$$

Differenza semplice media senza ripetizione

$$\Delta_{sr}{}^2 = \sqrt{\frac{[(-20)^2 \times 1404] + [(-30)^2 \times 27] + [20^2 \times 1404] + [(-10)^2 \times 52] + [30^2 \times 27] + [10^2 \times 52]}{80(80-1)}}$$

$$= \sqrt{\frac{1182200}{6320}} = 13,67$$