Esercitazione 2
Corso di Statistica (Prof. M. Aria)

Consideriamo i prezzi per notte (in euro) di 15 alberghi a 4 stelle di Napoli.

Albergo	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
Prezzo per notte (€)	120	130	200	185	116	150	240	100	170	155	120	140	170	120	230

$$\text{Media: } \mu = \frac{\sum_{l=1}^{N} x_l}{N} = \frac{2346}{15} = 156{,}4$$

Per il calcolo della mediana, è necessario ordinare la distribuzione in base all'intensità crescente del carattere considerato:

Albergo	8	5	1	11	14	2	12	6	10	9	13	4	3	15	7
Prezzo per notte (€)	100	116	120	120	120	130	140	150	155	170	170	185	200	230	240

Essendo N=15, la mediana si troverà in corrispondenza dell'ottavo valore, ovvero quel valore che divide la distribuzione in due parti uguali (ci sono 7 osservazioni alla sua sinistra e altrettante alla sua destra). La mediana pertanto sarà:

Me $= x_{\left(\frac{N+1}{2}\right)} = x_{\left(\frac{15+1}{2}\right)} = x_{(8)} = 150$;

Per calcolare la moda, dobbiamo invece costruire la distribuzione di frequenza.

Prezzo	Frequenza
100	1
116	1
120	3
130	1
140	1
150	1
155	1
170	2
185	1
200	1
230	1
240	1

La moda è la modalità a cui corrisponde la massima frequenza; nel nostro caso, quindi, sarà M_0=120, poiché tale valore si presenta con frequenza 3.

Attraverso gli indici di posizione abbiamo un'indicazione circa la tendenza centrale di una distribuzione: tuttavia, questi indici non forniscono alcuna informazione circa la variabilità della distribuzione stessa.
Descriviamo quindi la distribuzione in esame attraverso gli indici di variabilità.

Per ottenere una misura di variabilità, si può quantificare l'entità degli scostamenti di ciascuna unità rispetto ad un centro, che di solito corrisponde alla media aritmetica.
Per trovare lo scarto quadratico medio (o deviazione standard) dobbiamo calcolare gli scarti al quadrato di ciascuna osservazione dalla media aritmetica.

Osservazioni	x_l	$(x_l - \mu)^2$	x_l^2
8	100	3180,96	10000
5	116	13456	13456
1	120	14400	14400
11	120	14400	14400
14	120	14400	14400
2	130	16900	16900
12	140	19600	19600
6	150	22500	22500
10	155	24025	24025
9	170	28900	28900
13	170	28900	28900
4	185	34225	34225
3	200	40000	40000
15	230	52900	52900
7	240	57600	57600
TOT	**2346**	**385387**	**392206**

Il totale degli scarti va poi diviso per le N osservazioni, per poi calcolare la radice quadrata di tale rapporto.

$$\sigma = \sqrt{\frac{\sum_{l=1}^{N}(X_l - \mu)^2}{N}} = 41{,}062$$

Per amplificare l'effetto delle variazioni (anche minime) rispetto al valore medio, è conveniente usare la varianza, ovvero il quadrato dello S.Q.M.

$$\sigma^2 = \frac{\sum_{l=1}^{N}(X_l - \mu)^2}{N} = 1686{,}1$$

L'unità di misura della varianza ha però il difetto di essere il quadrato della grandezza originaria.

FORMULA ALTERNATIVA PER IL CALCOLO DELLA VARIANZA

$$\sigma^2 = \frac{\sum_{l=1}^{N}(x_l - \mu)^2}{N} = \frac{\sum_{l=1}^{N}(x_l{}^2 - 2x_l\mu + \mu^2)}{N} = \frac{\sum_{l=1}^{N} x^2}{N} - \frac{2\mu\sum_{l=1}^{N} x_l}{N} + \frac{N}{N}\mu^2 = \frac{\sum_{l=1}^{N} x_i^2}{N} - 2\mu^2 + \mu^2 =$$

$$\boxed{\frac{\sum_{l=1}^{N} x_l^2}{N} - \mu^2}$$

Riprendendo i dati della tabella precedente abbiamo:

$$\sigma^2 = \frac{392206}{15} - 156{,}4^2 = 26147{,}07 - 24460{,}96 = 1686{,}1$$

Il vantaggio di questa formula è sicuramente la maggiore rapidità nelle operazioni di calcolo.
Proviamo ad aggiungere alla nostra distribuzione un valore pari alla media

Media: 156,4

Albergo	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16
Prezzo per notte (€)	120	130	200	185	116	150	240	100	170	155	120	140	170	120	230	156,4

Ricalcolando la media possiamo osservare come essa rimane invariata:

$$\mu = \frac{\sum_{l=1}^{N} X_l}{N} = \frac{2502,4}{16} = 156,4$$

Se calcoliamo nuovamente la deviazione standard vediamo che essa diminuisce:

$$\sigma = \sqrt{\frac{(\sum_{l=1}^{N}(x_l - \mu)^2) + (x_{N+1} - \mu)^2}{16}} = \sqrt{\frac{25291,6 + 0}{16}} = 39,75$$

Allo stesso modo diminuisce anche la varianza:

$$\sigma^2 = \frac{(\sum_{l=1}^{N}(x_l - \mu)^2) + (x_{N+1} - \mu)^2}{16} = \frac{25291.6 + 0}{16} = 1580,72$$

Aggiungiamo ora alla nostra distribuzione una costante pari a 5

Albergo	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
Prezzo per notte (€)	120 +5	130 +5	200 +5	185 +5	116 +5	150 +5	240 +5	100 +5	170 +5	155 +5	120 +5	140 +5	170 +5	120 +5	230 +5
	125	135	205	190	121	155	245	105	175	160	125	145	175	125	235

Se calcoliamo la media vediamo che essa è pari alla precedente più il valore relativo alla costante:

$$\mu = \frac{\sum_{l=1}^{N} X_l}{N} = \frac{2421}{15} = 161,4 = (156,4+5)$$

Osserv.	x_l	$(x_l - \mu)$	$(x_l - \mu)^2$

1	125	-36,4	1324,96
2	135	-26,4	696,96
3	205	43,6	1900,96
4	190	28,6	817,96
5	121	-40,4	1632,16
6	155	-6,4	40,96
7	245	83,6	6988,96
8	105	-56,4	3180,96
9	175	13,6	184,96
10	160	-1,4	1,96
11	125	-36,4	1324,96
12	145	-16,4	268,96
13	175	13,6	184,96
14	125	-36,4	1324,96
15	235	73,6	5416,96
TOT	**2421**	**0**	**25291,6**

Se invece calcoliamo la deviazione standard essa rimane invariata:

$$\sigma = \sqrt{\frac{\sum_{l=1}^{N}(x_l - \mu)^2}{15}} = \sqrt{\frac{25291,6}{15}} = 41,062$$

Allo stesso modo rimane immutata anche la varianza:

$$\sigma^2 = \frac{\sum_{l=1}^{n}(x_l - \mu)^2}{N} = \frac{25291.6}{15} = 1686,1$$

Moltiplichiamo ai valori della nostra distribuzione una costante pari a 2

Albergo	Prezzo per notte (€)	$(x_l - \mu)^2$
1	240	5299,84
2	260	2787,84
3	400	7603,84
4	370	3271,84
5	232	6528,64
6	300	163,84
7	480	27955,80
8	200	12723,80
9	340	739,84
10	310	7,84
11	240	5299,84
12	280	1075,84
13	340	739,84
14	240	5299,84
15	460	21667,80
Totale	**4692**	**101166,30**

Se calcoliamo la media vediamo come essa sia pari alla precedente più il valore relativo alla costante:

$$\mu = \frac{\sum_{l=1}^{N} X_l}{N} = \frac{4692}{15} = 312{,}8 = (156{,}4*2)$$

Il valore della deviazione standard sarà pari a:

$$\sigma = \sqrt{\frac{\sum_{l=1}^{N}(x_l - \mu)^2}{15}} = \sqrt{\frac{101166{,}30}{15}} = 82{,}124 = (41{,}062 * 2)$$

Allo stesso modo la varianza:

$$\sigma^2 = \frac{\sum_{l=1}^{N}(x_l - \mu)^2}{N} = \frac{101166{,}30}{15} = 6744{,}42 = (1686{,}1 * 2^2)$$

Consideriamo la distribuzione del carattere **età** relativa ad un collettivo di turisti, suddivisa in classi.

	Frequenza n_i	Frequenze assolute cumulate N_i	Percentuale f_i	Percentuale cumulata F_i	Valore centrale c_i	$(c_i - \mu)^2 * n_i$
Fino a 18 anni	6	6	7,3	7,3	9,0	6029,34
Da 19 a 26 anni	21	27	25,6	32,9	22,5	6956,04
Da 27 a 40 anni	13	40	15,9	48,8	33,5	673,92
Da 41 a 60 anni	29	69	35,4	84,1	50,5	2785,16
Oltre 60 anni	13	82	15,9	100,0	70,0	11160,37
Totale	82		100,0			27604,83

Calcoliamo la media:

$$\mu = \frac{(6 * 9) + (21 * 22{,}5) + (13 * 33{,}5) + (29 * 50{,}5) + (13 * 70)}{82} =$$

$$= \frac{54 + 472{,}5 + 435{,}5 + 1464{,}5 + 910}{82} = \frac{3336{,}5}{82} = 40{,}7$$

Calcoliamo la varianza:

$$\sigma^2 = \frac{\sum_{i=1}^{K} (c_i - \mu)^2 * n_i}{N} = \frac{27604{,}83}{82} = 336.64$$

Calcoliamo la deviazione standard:

$$\sigma = \sqrt{\frac{\sum_{i=1}^{K} (c_i - \mu)^2 * n_i}{N}} = \sqrt{336.64} = 18{,}35$$