

## *Variabilità*

- L'utilizzo di una *media* permette di sintetizzare efficacemente l'informazione contenuta in una distribuzione statistica dal punto di vista dell'*intensità* del carattere. Tuttavia la sintesi può essere eccessiva, nel senso si possono perdere informazioni su altre caratteristiche fondamentali come la *variabilità*.
- La *variabilità* è definibile come la tendenza delle unità di un collettivo ad assumere modalità diverse tra loro.

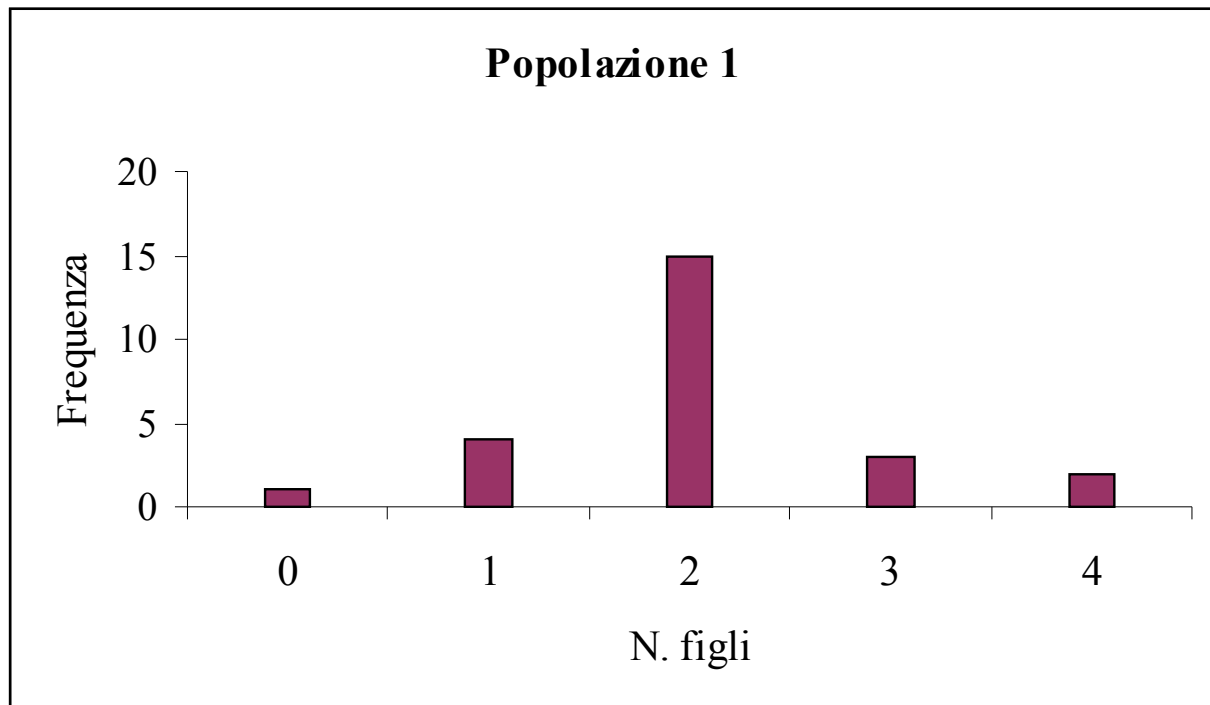
## *Esempio*

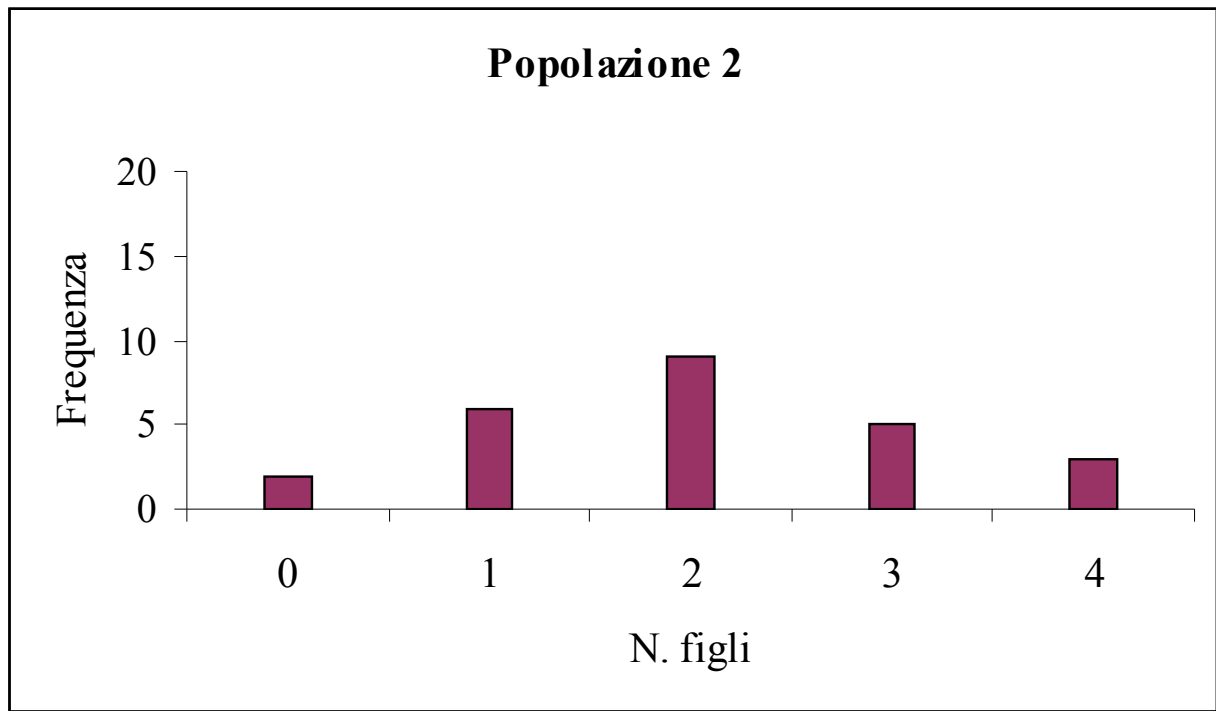
- Consideriamo due distribuzioni del numero di figli per un collettivo di 25 famiglie.

Popolazione 1	
N. figli ( $x_i$ )	Frequenze ( $n_i$ )
0	1
1	4
2	15
3	3
4	2
Totale	25

Popolazione 2	
N. figli ( $x_i$ )	Frequenze ( $n_i$ )
0	2
1	6
2	9
3	5
4	3
Totale	25

- Entrambe le distribuzioni hanno media  $\mu = 2,04$  ma, come è possibile dedurre dai grafici, sono molto diverse: la prima assume delle modalità molto più *concentrate* attorno alla media e quindi ha *minore variabilità*.





## ***Indici di variabilità***

- Per avere una descrizione più completa della distribuzione è quindi opportuno utilizzare, oltre a una media, un ***indice*** che misuri la variabilità della distribuzione.
- Un ***indice di variabilità*** deve:
  - assumere il valore minimo (tipicamente 0) se e solo se tutte le unità della distribuzione presentano la stessa modalità;
  - aumentare all'aumentare della diversità tra le modalità del carattere assunte dalle varie unità.
- Gli indici di variabilità possono essere basati:
  - sullo *scostamenti da una media*;
  - sulla *differenze tra statistiche d'ordine*.

<b>Scostamenti da una media</b>	<b>Differenze tra statistiche d'ordine</b>
Devianza	Campo di variazione
Varianza	Differenza interquartilica
Deviazione standard	
Coefficiente di variazione	
Scostamento semplice medio dalla mediana	

## ***Devianza e varianza***

- Per una *distribuzione unitaria* di un *carattere quantitativo*, la ***devianza*** è definita come

$$D = \sum_{i=1}^N (x_i - \mu)^2$$

- Per una *distribuzione di frequenza* non in classi

$$D = \sum_{i=1}^k (x_i - \mu)^2 n_i$$

- Se il *carattere è in classi* si utilizzano i *valori centrali*

$$x_i = \frac{c_{i-1} + c_i}{2}$$

al posto delle modalità.

- La ***varianza*** è normalmente preferita alla *devianza* e si ottiene come:

$$\sigma^2 = \frac{D}{N} = \frac{1}{N} \sum_{i=1}^k (x_i - \mu)^2 n_i = \sum_{i=1}^k (x_i - \mu)^2 f_i$$

## *Esempi*

- Per la *distribuzione unitaria* dei voti, che ha media  $\mu = 25,667$ , si ha:

Unità ( $i$ )	Voto ( $x_i$ )	$(x_i - \mu)^2$
1	27	1,7769
2	21	21,7809
3	28	5,4429
4	30	18,7749
5	21	21,7809
6	27	1,7769
Totale	154	71,3333

da cui  $D = 71,333$  e  $\sigma^2 = 71,333/6 = 11,889$ .

- Per la prima distribuzione del numero di figli per un collettivo di 25 famiglie, che ha media  $\mu = 2,04$ , si ha:

N. figli ( $x_i$ )	Frequenze ( $n_i$ )	$(x_i - \mu)^2$	$(x_i - \mu)^2 n_i$
0	1	4,1616	4,1616
1	4	1,0816	4,3264
2	15	0,0016	0,0240
3	3	0,9216	2,7648
4	2	3,8416	7,6832
Totale	25	-	18,960

da cui  $D = 18,96$  e  $\sigma^2 = 18,96/25 = 0,758$ .

- Per la seconda distribuzione, che ha sempre media  $\mu = 2,04$ , si ha:

N. figli ( $x_i$ )	Frequenze ( $n_i$ )	$(x_i - \mu)^2$	$(x_i - \mu)^2 n_i$
0	2	4,1616	8,3232
1	6	1,0816	6,4896
2	9	0,0016	0,0144

3	5	0,9216	4,6080
4	3	3,8416	11,5248
<b>Totale</b>	25	10,008	30,9600

da cui

$$D = 30,96 \quad \text{e} \quad \sigma^2 = 30,96 / 25 = 1,238$$

che indica che la seconda distribuzione ha maggiore variabilità della prima.

- Per la distribuzione dell'altezza per un collettivo di 50 soggetti, che ha media  $\mu = 173,8$ , si ha:

Altezza ( $c_{i-1} - c_i$ )	Frequenze ( $n_i$ )	Valori centrali ( $x_i$ )	$(x_i - \mu)^2 n_i$
150-160	1	155	353,44
160-170	10	165	774,40
170-180	35	175	50,40
180-200	4	190	1049,76
<b>Totale</b>	50	-	2228,00

da cui  $D = 2228$  e  $\sigma^2 = 2228 / 50 = 44,56$ .

## ***Deviazione standard e coefficiente di variazione***

- La ***deviazione standard*** (o ***scostamento quadratico medio***) è l'indice di variabilità più utilizzato in quanto è espresso nella stessa unità di misura del carattere. Si ottiene come

$$\sigma = \sqrt{\sigma^2} = \sqrt{\frac{1}{N} \sum_{i=1}^k (x_i - \mu)^2 n_i}$$

- Nel caso in cui la distribuzione abbia media aritmetica ( $\mu$ ) positiva, il ***coefficiente di variazione*** si calcola come (normalmente in percentuale)

$$CV = \frac{\sigma}{\mu}$$

### ***Esempi***

- Per la ***distribuzione dei voti***:  $\sigma = \sqrt{11,889} = 3,448$ ,  $CV = \frac{3,448}{25,667} = 13,43\%$

- Per le ***distribuzione del numero di figli***:

popolazione 1:  $\sigma = \sqrt{0,7584} = 0,871$ ,  $CV = \frac{0,871}{2,04} = 42,69\%$

popolazione 2:  $\sigma = \sqrt{1,2384} = 1,113$ ,  $CV = \frac{1,113}{2,04} = 54,55\%$

- Per la *distribuzione dell'altezza*:  $\sigma = \sqrt{44,56} = 6,675$ ,  $CV = \frac{6,675}{173,8} = 3,84\%$

## *Proprietà*

- **Proprietà 1:** gli indici  $D$ ,  $\sigma^2$  e  $\sigma$  sono sempre non negativi e assumono il valore minimo (0) se e solo se tutte le modalità della distribuzione sono uguali tra loro.

- **Proprietà 2:** la devianza può essere calcolata come (*formula semplificata*)

$$D = \sum_{i=1}^N x_i^2 - N\mu^2 \quad (\text{distribuzioni unitarie})$$

$$D = \sum_{i=1}^k x_i^2 n_i - N\mu^2 \quad (\text{distribuzioni di frequenza})$$

che ha vantaggi nel calcolo anche della *varianza* e della *deviazione standard*.

- **Proprietà 3:** se a ogni termine della distribuzione viene applicata la trasformazione  $aX + b$ , allora gli indici di variabilità cambieranno nel modo seguente

Devianza	----->	$a^2 D$
Varianza	----->	$a^2 \sigma^2$
Deviazione standard	----->	$a \sigma$

*Esempi*

- Per la *distribuzione unitaria* dei voti, che ha media  $\mu = 25,667$ , si ha:

Unità ( $i$ )	Voto ( $x_i$ )	$x_i^2$
1	27	729
2	21	441
3	28	784
4	30	900
5	21	441
6	27	729
Totale	154	4024

da cui

$$D = 4024 - 6 \cdot 25,667^2 = 71,23 \approx 71,333$$

- Per la prima distribuzione del numero di figli, che ha media  $\mu = 2,04$ , si ha:

N. figli ( $x_i$ )	Frequenze ( $n_i$ )	$x_i^2 n_i$
0	1	0
1	4	4
2	15	60
3	3	27
4	2	32
Totale	25	123

da cui

$$D = 123 - 25 \cdot 2,04^2 = 18,96$$

## ***Scostamenti semplici medi***

- Per una *distribuzione unitaria* di un *carattere quantitativo*, lo **scostamento semplice medio dalla media aritmetica** è definito come

$$S_{\mu} = \frac{1}{N} \sum_{i=1}^N |x_i - \mu|$$

- Per una *distribuzione di frequenza* non in classi

$$S_{\mu} = \frac{1}{N} \sum_{i=1}^k |x_i - \mu| n_i = \sum_{i=1}^k |x_i - \mu| f_i.$$

- Se il *carattere è in classi* si utilizzano i *valori centrali*

$$x_i = \frac{c_{i-1} + c_i}{2}$$

al posto delle modalità.

- Lo **scostamento semplice medio dalla mediana** si ottiene sostituendo la mediana alla media aritmetica:

$$S_m = \frac{1}{N} \sum_{i=1}^N |x_i - m| \quad \text{(distribuzione unitaria)}$$

$$S_m = \frac{1}{N} \sum_{i=1}^k |x_i - m| n_i = \sum_{i=1}^k |x_i - m| f_i \quad \text{(distribuzione di frequenza)}$$

## *Esempi*

- Per la *distribuzione unitaria* dei voti ( $\mu = 25,667$ ,  $m = 27$ ) si ha:

Unità ( $i$ )	Voto ( $x_i$ )	$ x_i - \mu $	$ x_i - m $
1	27	1,333	0
2	21	4,667	6
3	28	2,333	1
4	30	4,333	3
5	21	4,667	6
6	27	1,333	0
<b>Totale</b>	154	18,666	16

da cui  $S_\mu = 18,666/6 = 3,111$  e  $S_m = 16/6 = 2,667$ .

- Per la prima distribuzione del *numero di figli* ( $\mu = 2,04$ ,  $m = 2$ ) si ha:

N. figli ( $x_i$ )	Frequenze ( $n_i$ )	$ x_i - \mu n_i$	$ x_i - m n_i$
0	1	2,04	2
1	4	4,16	4
2	15	0,60	0
3	3	2,88	3
4	2	3,92	4
<b>Totale</b>	25	13,60	13

da cui  $S_\mu = 13,6/25 = 0,544$  e  $S_m = 13/25 = 0,52$ .

- Per la seconda distribuzione del *numero di figli* ( $\mu = 2,04$ ,  $m = 2$ ) si ha:

$$S_\mu = 0,854 \quad \text{e} \quad S_m = 0,84$$

che confermano che la distribuzione presenta maggiore variabilità della prima.

## *Altri indici di variabilità*

- Per una distribuzione con modalità ordinate,  $y_1, \dots, y_N$ , il ***campo di variazione*** è definito come

$$\Delta_c = y_N - y_1$$

- E' l'indice di variabilità più semplice da calcolare, ma non è molto efficace nel misurare la variabilità della distribuzione.
- La ***differenza interquartilica*** si basa sul primo quartile ( $q_1$ ) e il terzo quartile ( $q_3$ ) ed è definita come

$$\Delta_q = q_3 - q_1$$

## *Esempi*

- Per la *distribuzione dei voti*, si ha

$$\Delta_c = 30 - 21 = 9$$

- Per entrambe le *distribuzioni del numero dei figli*, si ha

$$\Delta_c = 4 - 0 = 4$$

- Per la *distribuzioni dell'altezza*, si ha

$$\Delta_q = 177,57 - 170,43 = 7,14$$

