

Dispersione



si cercano indici di dispersione che:

- utilizzino tutti i dati $\{x_1, x_2 \dots x_n\}$
- siano basati sulla nozione di *scarto* (*distanza*) dei dati
 - rispetto ad un centro $d_i = |x_i - C|$
ad es. rispetto alla media aritmetica $d_i = |x_i - \bar{x}|$
 - rispetto a un dato $d_i = |x_i - x_j|$

con alcune proprietà generali:

- non deve mai essere negativo
- assume valore 0 se i dati sono tutti uguali
- non cambia se aggiungiamo una costante ai dati



VARIANZA : è la media aritmetica (*semplice o ponderata*) dei quadrati degli scarti

- dato l'insieme di valori $\{x_1, x_2, \dots, x_n\}$

$$Var = s^2 = \frac{1}{n} \cdot \sum_{i=1}^n (x_i - \bar{x})^2$$

- dato l'insieme di valori $\{x_1, x_2, \dots, x_m\}$ con le rispettive frequenze assolute $\{f_1, f_2, \dots, f_m\}$

$$Var = s^2 = \frac{1}{n} \cdot \sum_{i=1}^m f_i \cdot (x_i - \bar{x})^2 \quad \left[n = \sum_{i=1}^m f_i \right]$$

Deviazione Standard



DEVIAZIONE STANDARD (*scarto quadratico medio*) :

è la radice quadrata positiva della varianza

$$s = \sqrt{\frac{1}{n} \cdot \sum_{i=1}^n (x_i - \bar{x})^2} \quad \text{o} \quad s = \sqrt{\frac{1}{n} \cdot \sum_{i=1}^m f_i \cdot (x_i - \bar{x})^2}$$

consente di avere un indice di dispersione espresso nella stessa unità di misura dei dati

NOTA applicando una trasformazione lineare ai dati

$$y_i = a x_i + b \Rightarrow s_y^2 = a^2 s_x^2 \quad s_y = |a| s_x$$



Spesso gli *indici statistici* vengono applicati non all'intera *popolazione* ma a un suo *campione*. Si cerca di stimare (*inferenza*) nel miglior modo possibile le caratteristiche dell'intera popolazione a partire dalle informazioni desunte da un *campione rappresentativo*. In questo caso si utilizzano le seguenti formule modificate:

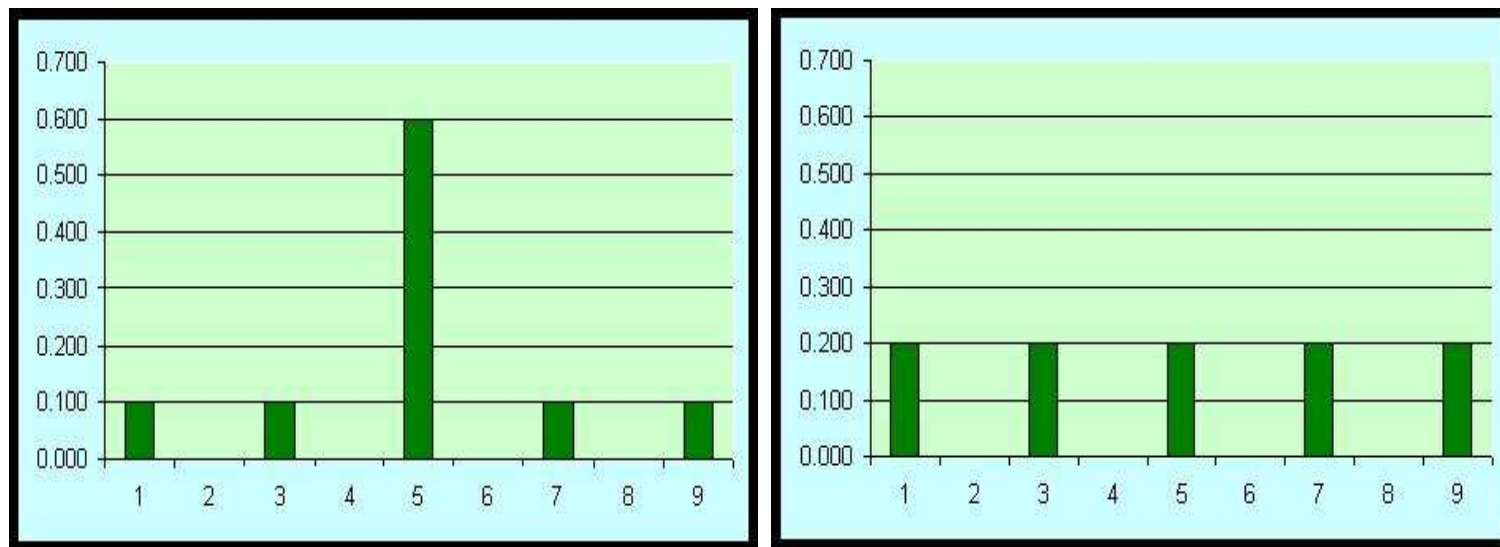
VARIANZA CAMPIONARIA (*stimata*) :

$$s^2 = \frac{1}{n-1} \cdot \sum_{i=1}^n (x_i - \bar{x})^2$$

DEVIAZIONE STANDARD CAMPIONARIA (*stimata*) :

$$s = \sqrt{\frac{1}{n-1} \cdot \sum_{i=1}^n (x_i - \bar{x})^2}$$

Esempio riassuntivo - 1



Esempio riassuntivo 2



Caso A

x_i	f_i	f_i/n
1	1	0.100
3	1	0.100
5	6	0.600
7	1	0.100
9	1	0.100
10	1.000	

media	5.00
mediana	5.00
varianza	4.00
varianza stimata	4.44
deviazione standard	2.00
deviazione standard stimata	2.11

Caso B

x_i	f_i	f_i/n
1	2	0.200
3	2	0.200
5	2	0.200
7	2	0.200
9	2	0.200
10	1.000	

media	5.00
mediana	5.00
varianza	8.00
varianza stimata	8.89
deviazione standard	2.83
deviazione standard stimata	2.98

Esercizio 1



ESERCIZIO – Trovare media, mediana, moda, varianza e deviazione standard dei seguenti dati non ordinati e non raggruppati. Tracciare l'istogramma delle frequenze.

7	4	10	9	15	12	7	8	11	4
14	10	5	14	1	10	8	12	6	5

Esercizio 1



SOLUZIONE – costruisco la tabella della distribuzione di frequenza

x	1	4	5	6	7	8	9	10	11	12	14	15	
f_{ass}	1	2	2	1	2	2	1	3	1	2	2	1	20

$$\bar{x} = \frac{1}{20}(1 + 8 + 10 + 6 + 14 + 16 + 9 + 30 + 11 + 24 + 28 + 15) = 8.6$$

$$s^2 = \frac{1}{20}(57.76 + 42.32 + 25.92 + 6.76 + 5.12 + 0.72 + 0.16 + 5.88 + \\ + 5.76 + 23.12 + 58.33 + 40.96) \approx 13.64$$

$$s \approx 3.69$$

$$\text{moda} = 10.0$$

$$\text{mediana} = 8.5$$

Esercizio 2



ESERCIZIO – Nel rilevare l'altezza in *cm* di un gruppo di reclute si è ottenuta la seguente tabella delle frequenze. Calcolare media, mediana e quartili.

<i>(cm)</i>	f_{ass}	f_{cum}
166	1	1
168	3	4
169	6	10
170	11	21
171	8	29
172	6	35
173	4	39
174	3	42
175	1	43
178	1	44

- $n = 44$ dimensione del campione
- $\bar{x} \approx 170.9$ media
- $M_e = \frac{x_{22} + x_{23}}{2} = 171$ mediana
- $q_1 = \frac{x_{11} + x_{12}}{2} = 170$ primo quartile
- $q_3 = \frac{x_{33} + x_{34}}{2} = 172$ terzo quartile
- $q_3 - q_1 = 2$ distanza interquartile

La distanza interquartile è un altro indice di dispersione, legato alla nozione di mediana. La mediana suddivide l'insieme dei dati ordinati x_i in due parti ugualmente numerose. I quartili si ottengono suddividendo i dati ordinati in quattro parti ugualmente numerose.

Esercizio 3



ESERCIZIO – Considero la seguente tabella relativa alle frequenze dei pesi in *Kg* di 100 individui adulti.

<i>Peso p in Kg</i>	<i>f_{ass}</i>
$50 \leq p < 55$	20
$55 \leq p < 60$	15
$60 \leq p < 65$	18
$65 \leq p < 70$	22
$70 \leq p < 75$	18
$75 \leq p < 80$	7

- le classi sono di uguale ampiezza
- suppongo che i dati siano uniformemente distribuiti all'interno di ogni classe
- posso definire per ogni classe un rappresentante r_i (*class mark*)

Esercizio 3



<i>Peso p in Kg</i>	f_i	F_i	r_i
$50 \leq p < 55$	20	20	52.5
$55 \leq p < 60$	15	35	57.5
$60 \leq p < 65$	18	53	62.5
$65 \leq p < 70$	22	75	67.5
$70 \leq p < 75$	18	93	72.5
$75 \leq p < 80$	7	100	77.5

- calcolo la media utilizzando i valori dei rappresentanti
- calcolo lo scarto quadratico medio utilizzando i valori dei rappresentanti

calcolo il peso medio

$$\bar{p} = \frac{1}{100} (20 \cdot 52.5 + 15 \cdot 57.5 + 18 \cdot 62.5 + 22 \cdot 67.5 + 18 \cdot 72.5 + 7 \cdot 77.5) = 63.7$$

Esercizio 3



calcolo la varianza e lo scarto quadratico medio (*deviazione standard*)

r_i	$r_i - \bar{p}$	$(r_i - \bar{p})^2$	f_i
52.5	-11.2	125.44	20
57.5	-6.2	38.44	15
62.5	-1.2	1.44	18
67.5	3.8	14.44	22
72.5	8.8	77.44	18
77.5	13.8	190.44	7

$$s^2 = \frac{1}{100} (20 \cdot 125.44 + 15 \cdot 38.44 + 18 \cdot 1.44 + 22 \cdot 14.44 + 18 \cdot 77.44 + 7 \cdot 190.44) \approx 61.56 \text{ Kg}^2$$

$$s \approx 7.85 \text{ kg}$$

Media - Varianza - Deviazione Standard



$\bar{x}$ media	$\frac{1}{n} \cdot \sum_{i=1}^n x_i$	$\frac{1}{n} \cdot \sum_{i=1}^m f_i x_i$
s^2 varianza	$\frac{1}{n} \cdot \sum_{i=1}^n (x_i - \bar{x})^2$	$\frac{1}{n} \cdot \sum_{i=1}^m f_i \cdot (x_i - \bar{x})^2$
s dev. standard	$\sqrt{\frac{1}{n} \cdot \sum_{i=1}^n (x_i - \bar{x})^2}$	$\sqrt{\frac{1}{n} \cdot \sum_{i=1}^m f_i \cdot (x_i - \bar{x})^2}$
s^2 campionaria	$\frac{1}{n-1} \cdot \sum_{i=1}^n (x_i - \bar{x})^2$	$\frac{1}{n-1} \cdot \sum_{i=1}^m f_i \cdot (x_i - \bar{x})^2$
s^2 campionaria	$\sqrt{\frac{1}{n-1} \cdot \sum_{i=1}^n (x_i - \bar{x})^2}$	$\sqrt{\frac{1}{n-1} \cdot \sum_{i=1}^m f_i \cdot (x_i - \bar{x})^2}$

Varianza - Deviazione Standard



le espressioni della varianza (e della deviazione standard) possono essere riscritte come segue

$$s^2 = \frac{1}{n} \cdot \left[\sum_{i=1}^n x_i^2 - n \bar{x}^2 \right] \quad s^2 = \frac{1}{n} \cdot \left[\sum_{i=1}^m f_i x_i^2 - n \bar{x}^2 \right]$$

infatti

$$\begin{aligned} \sum_{i=1}^n (x_i - \bar{x})^2 &= \sum_{i=1}^n (x_i^2 - 2 x_i \bar{x} + \bar{x}^2) = \sum_{i=1}^n x_i^2 - 2 \bar{x} \sum_{i=1}^n x_i + \sum_{i=1}^n \bar{x}^2 = \\ &= \sum_{i=1}^n x_i^2 - 2 \bar{x} (n \bar{x}) + n \bar{x}^2 = \sum_{i=1}^n x_i^2 - n \bar{x}^2 \end{aligned}$$

Esercizio Voti - 1



Un'indagine su un campione di $n = 100$ studenti, che ha sostenuto la prova scritta di matematica, ha prodotto il seguente risultato. Le votazioni in *centesimi* sono state raggruppate in quattro *classi*.

<i>classe</i>	f_i	f_i/n
20 – 40	10	0.10
40 – 60	20	0.20
60 – 80	50	0.50
80 – 100	20	0.20
	100	1.00

Le classi sono di uguale ampiezza e contigue

Esercizio Voti - 2



Nell'ipotesi di *distribuzione uniforme*, è naturale associare a ciascuna classe, come *rappresentante*, il valore centrale r_i della classe stessa.

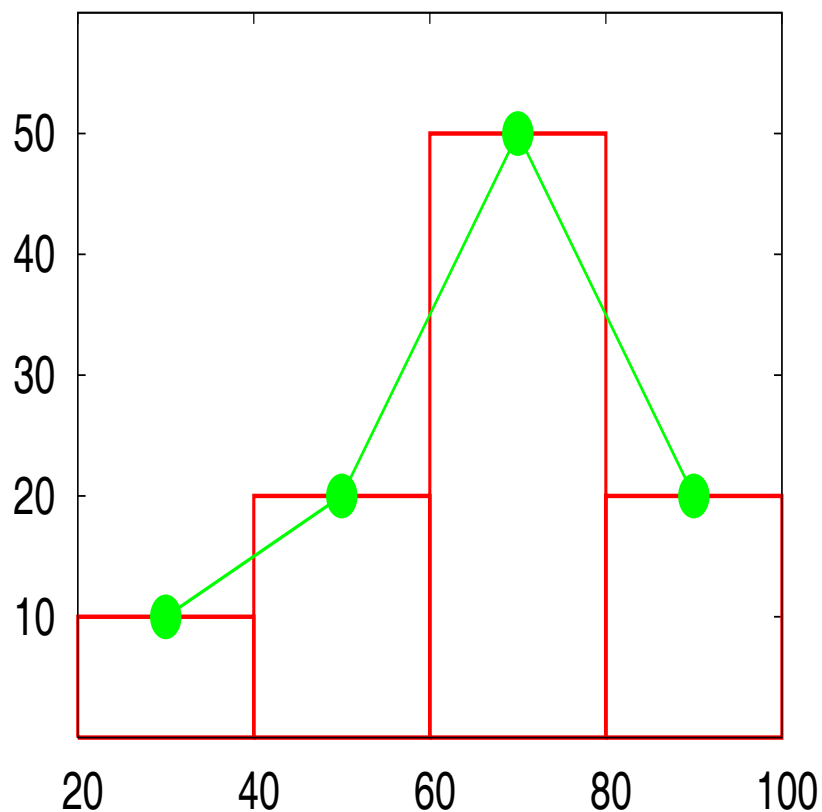
<i>classe</i>	r_i	f_i	F_i
20 – 40	30	10	10
40 – 60	50	20	30
60 – 80	70	50	80
80 – 100	90	20	100

$$media = \frac{1}{100} \cdot (10 \cdot 30 + 20 \cdot 50 + 50 \cdot 70 + 20 \cdot 90) = 66$$

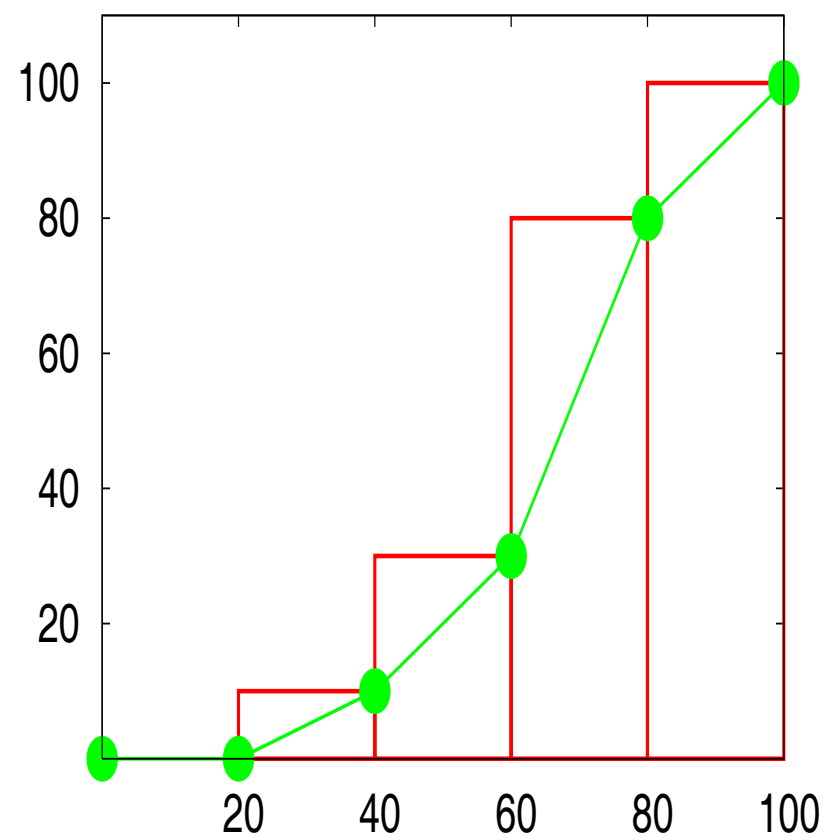
$mediana = 70$ guardando solo i rappresentanti

$$varianza = \frac{1}{100} \cdot (10 \cdot 36^2 + 20 \cdot 16^2 + 50 \cdot 4^2 + 20 \cdot 24^2) = 21.44$$

Esercizio Voti - 3

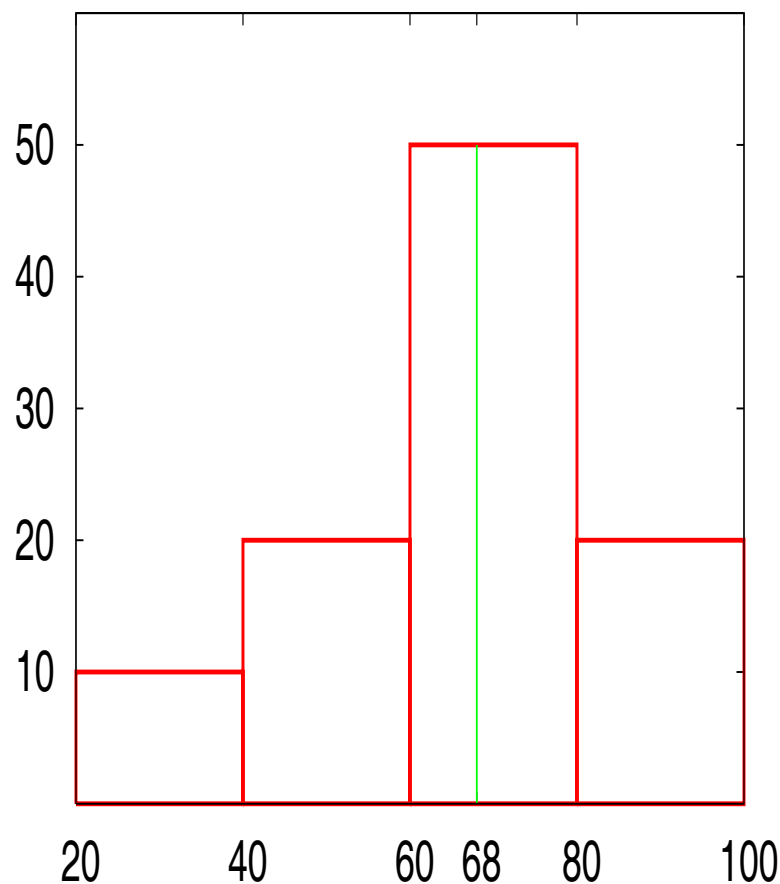


POLIGONO DELLE FREQUENZE



OGIVA DI FREQUENZA

Esercizio Voti - 4



CALCOLO DELLA MEDIANA

area totale

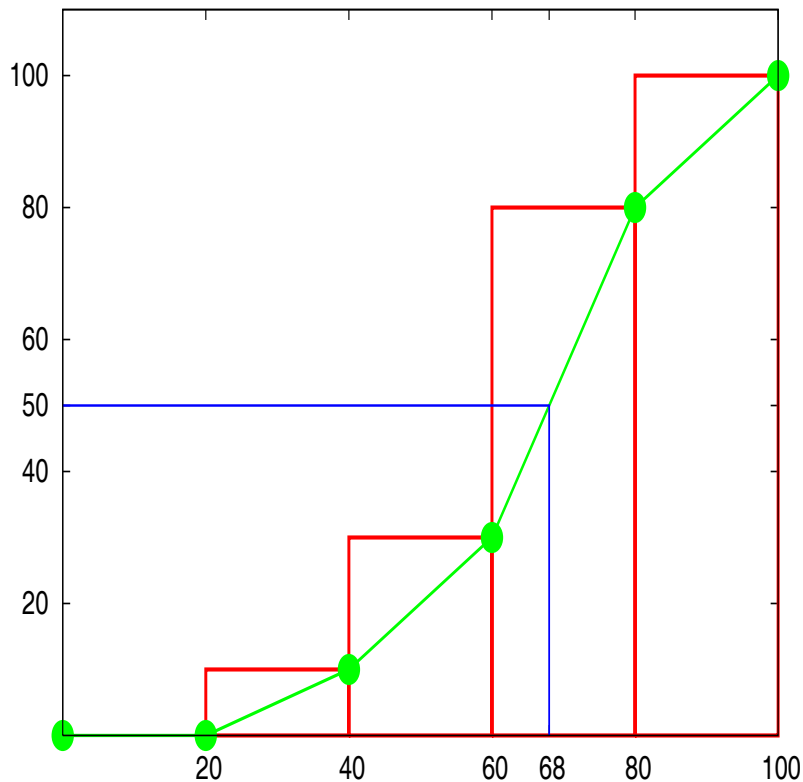
$$20 \cdot (10 + 20 + 50 + 20) = 2000$$

cerco il valore $x = M_e$ tale che

$$20 \cdot 10 + 20 \cdot 20 + (x - 60) \cdot 50 = 1000$$

$$\Rightarrow x = 68 \Rightarrow M_e = 68$$

Esercizio Voti - 5



CALCOLO DELLA MEDIANA

interpolazione lineare sui punti

$$A = (60, 30) \quad B = (80, 80)$$

$$\begin{cases} y &= 50 \\ y - 30 &= \frac{5}{2} \cdot (x - 60) \end{cases}$$

$$\Rightarrow (50 - 30) = \frac{5}{2} \cdot (x - 60)$$

$$\Rightarrow x = 68 \Rightarrow M_e = 68$$

DOMANDE

- qual è il voto massimo preso dal 10% degli studenti ? $[40]$
- calcolare i quartili dall'ogiva di frequenza $[q_1 = 55, q_2 = 78]$

Esercizio Voti - 6

