Indici di Dispersione

Si cercano indici di dispersione che:

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

con alcune proprietà generali:

- l'indice di dispersione non deve mai essere negativo
- assume il valore 0 se i dati sono tutti uguali
- non cambia se si aggiunge una costante ai dati

Varianza

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

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

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

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

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

Deviazione Standard

Deviazione standard *(o scarto quadratico medio)*: è la radice quadrata della varianza.

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

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

Nota: applicando una trasformazione lineare ai dati

$$y_i = ax_i + b \qquad \Rightarrow \qquad s_y^2 = a^2 s_x^2 \,, \qquad s_y = |a|\, s_x$$

Statistiche Campionarie

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:

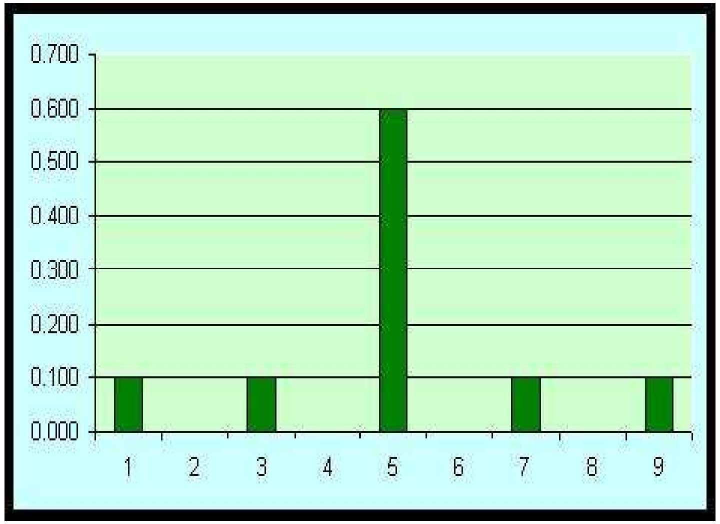
Varianza campionaria (*stimata*):

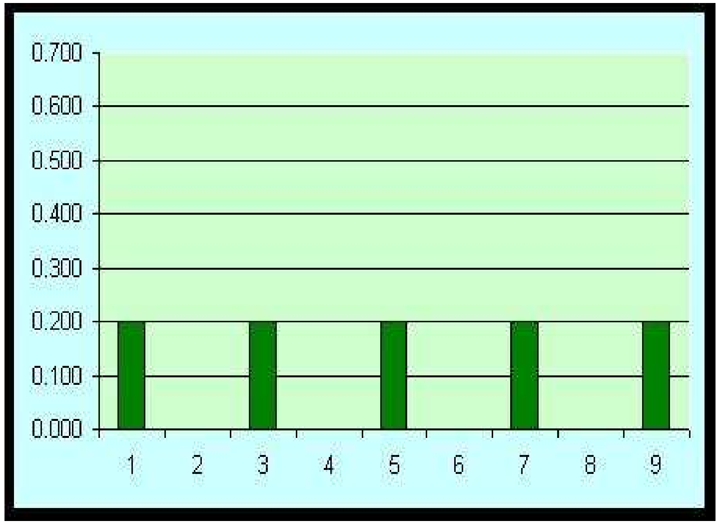
$$s^2 = \frac{1}{n-1} \cdot \sum_{i=1}^{n} (x_i - \bar{x})^2$$

Deviazione standard campionaria (*stimata*):

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

Esempio Riassuntivo





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

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

Esempio Riassuntivo

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

Esercizi

Esercizio 1. Si consideri la seguente tabella relativa alle frequenze dei pesi in Kg di 100 individui adulti.

Peso p in Kg	f_{ass}
$50 \leq p < 55$	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
- supponiamo che i dati siano uniformemente distribuiti all'interno di ogni classe
- possiamo definire per ogni classe un rappresentante r_i *(class mark)*

Calcolare il peso medio e lo scarto quadratico medio.

Esercizi

Soluzione: calcoliamo la media e lo scarto quadratico medio utilizzando i valori dei rappresentanti.

Peso p in Kg	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

Calcoliamo il peso medio:

$$\begin{aligned} \bar{p} &= \frac{1}{100}(20 \cdot 52.5 + 15 \cdot 57.5 + 18 \cdot 62.5 + 22 \cdot 67.5 + \\ &\quad + 18 \cdot 72.5 + 7 \cdot 77.5) = 63.7\,\text{Kg} \end{aligned}$$

Esercizi

Calcoliamo la varianza e lo scarto quadratico medio:

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

$$
\begin{aligned}
s^2 &= \frac{1}{100}(20 \cdot 125.44 + 15 \cdot 38.44 + 18 \cdot 1.44 + 22 \cdot 14.44 + \\
&\quad + 18 \cdot 77.44 + 7 \cdot 190.44) \simeq 61.56\,\mathrm{Kg}^2
\end{aligned}
$$

$s \simeq 7.85\,\mathrm{Kg}$