

Esercizio: calcolo di media e deviazione standard campionaria

Per verificare le prestazioni di un *hard disk* per PC, si osservano in misure ripetute i tempi di scrittura T_s di uno stesso *file* di grandi dimensioni.

In 5 prove ripetute si ottengono i seguenti valori, in secondi (s):

$T_{s,1}$	$T_{s,2}$	$T_{s,3}$	$T_{s,4}$	$T_{s,5}$
8.2	8.4	8.0	7.5	7.9

Si calcoli il **valor medio** di T_s nel campione di misure considerato e si ricavi la corrispondente **deviazione standard** campionaria $s(T_s)$.

Quanto vale la **dinamica** D di variazione dei valori misurati? e i rapporti D/\bar{T}_s e $s(T_s)/\bar{T}_s$ ("**dispersione relativa**")?

Esercizio: calcolo di media e deviazione standard campionaria

La media degli $n=5$ dati considerati (media delle misure) si calcola come

$$\bar{T}_s = \frac{x_1 + x_2 + \dots + x_n}{n} = \frac{1}{5} \sum_{i=1}^5 T_{s,i} = \frac{5 \times 8.0 + (0.2 + 0.4 + 0 - 0.5 - 0.1)}{5} \text{ s} = 8.0 \text{ s}$$

La varianza campionaria è invece

$$s^2(T_s) = \frac{\sum_{i=1}^5 (T_{s,i} - \bar{T}_s)^2}{5-1} = \frac{(0.2)^2 + (0.4)^2 + (0)^2 + (-0.1)^2 + (-0.5)^2}{4} \text{ s}^2 =$$
$$= \frac{0.46}{4} \text{ s}^2 = 0.115 \text{ s}^2$$

e dunque la deviazione standard del campione (dei dati raccolti in questo campione) vale

$$s(T_s) = \sqrt{s^2(T_s)} = 0.339 \text{ s} \cong 0.34 \text{ s}$$

Esercizio: calcolo di media e deviazione standard campionaria

Per calcolare la dinamica D , osserviamo prima i valori minimo (MIN) e massimo (MAX) dei dati raccolti:

$$\text{MIN}(T_i)=7.5 \text{ s}$$

$$\text{MAX}(T_i)=8.4 \text{ s}$$

e dunque la dinamica vale

$$D=[\text{MAX}(T_i)-\text{MIN}(T_i)]=[8.4 \text{ s} - 7.5 \text{ s}]=0.9 \text{ s}$$

I due rapporti cercati (dispersioni relative, in percentuale) sono:

$$D/\bar{T}_s = 11.25 \% \quad [\text{fascia di fluttuazione rapportata al valor medio}]$$

$$s(T_s)/\bar{T}_s = 4.24 \% \quad [\text{fluttuazione statistica rapportata al valor medio}]$$

Esercizi per casa (2.1 e 2.6)

Otto misure del diametro interno (in mm) di anelli forgiati per un pistone da automobile hanno portato ai valori:

2.1

1, 3, 15, 0, 5, 2, 5, 4

Si calcolino la media e la deviazione standard del campione.

Si disegni il corrispondente diagramma a punti (*dot plot*) e si commentino i dati.

I dati seguenti risultano da misure di intensità solare (W/m^2) in differenti giorni in una località della Spagna meridionale:

2.6

562, 869, 708, 775, 775, 704, 809, 856, 655, 806, 878, 909, 918, 558, 768, 870, 918, 940, 946, 661, 820, 898, 935, 952, 957, 693, 835, 905, 939, 955, 960, 498, 653, 730, 753

Si calcolino la media e la deviazione standard del campione.

Si disegni il corrispondente diagramma a ramo-e-foglia.

Esercizio 2.1

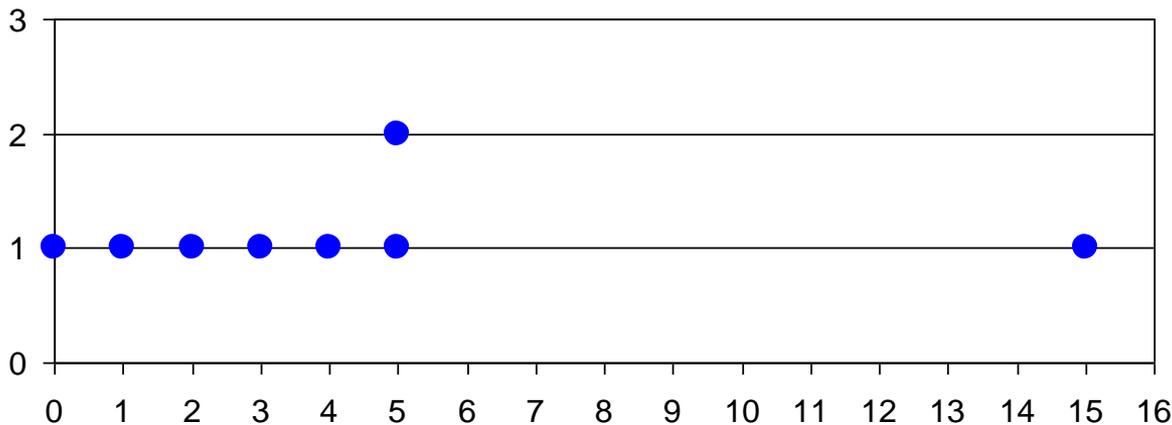
D_i (mm) = 1, 3, 15, 0, 5, 2, 5, 4 (diametro foro anelli)

2.1

FARE: media; dev. st.; *dot diagram*; commenti.

$$n=8 \text{ dati nel campione} \quad \bar{D} = \frac{1}{8} \sum_{i=1}^8 D_i = \frac{35}{8} \text{ mm} = 4.375 \text{ mm}$$

$$s_D^2 = \frac{1}{7} \sum_{i=1}^8 (D_i - \bar{D})^2 = 21.7 \text{ mm}^2 \quad s_D = \sqrt{s_D^2} = 4.658 \text{ mm}$$



c'è un punto
(valore $D_3=15$ mm)
decisamente “lontano”
da tutti gli altri

per i valori da 0 a 5 mm si
osserva una forte dispersione

la dev. st. circa uguale alla
media lascia da pensare...

Esercizio 2.6

2.6

Ci sono $n=35$ dati I_i (W/m^2) nel campione considerato
 562, 869, 708, 775, 775, 704, 809, 856, 655, 806, 878, 909,
 918, 558, 768, 870, 918, 940, 946, 661, 820, 898, 935, 952,
 957, 693, 835, 905, 939, 955, 960, 498, 653, 730, 753

Media e deviazione standard campionarie sono:

$$\bar{I} = \frac{1}{35} \sum_{i=1}^{35} I_i = \frac{28368}{35} \text{ W / m}^2 \cong 810.5 \text{ W / m}^2 \quad s_I = \sqrt{\frac{1}{34} \sum_{i=1}^{35} (I_i - \bar{I})^2} = 128.3 \text{ W / m}^2$$

I dati ordinati e il corrispondente diagramma a ramo-e-foglia sono:

498	558	562	653	655	661	693	704	708	730	753	768
775	775	806	809	820	835	856	869	870	878	898	905
909	918	918	935	939	940	946	952	955	957	960	

498												
558	562											
653	655	661	693									
704	708	730	753	768	775	775						
806	809	820	835	856	869	870	878	898				
905	909	918	918	935	939	940	946	952	955	957	960	

adesso appare evidente come la distribuzione dei dati non è simmetrica

Esercizio su calcolo di media, deviazione standard e box plot

Si studia il tempo di accensione a freddo di un motore di una automobile.

Da un insieme (campione) di misure sul veicolo si ottengono i seguenti dati (in secondi):

K	1	2	3	4	5	6	7	8
T_{acc}	1.75	1.92	2.62	2.35	3.09	3.15	2.53	1.91

- Calcolare la media campionaria e la deviazione standard campionaria
- Disegnare il box plot dei dati

Esercizio su calcolo di media, deviazione standard e box plot

La media degli $n = 8$ dati considerati si calcola come

$$\bar{T}_{\text{acc}} = \frac{1}{n} \sum_{i=1}^n T_{\text{acc},i} = \frac{19.32}{8} \text{ s} = 2.415 \text{ s}$$

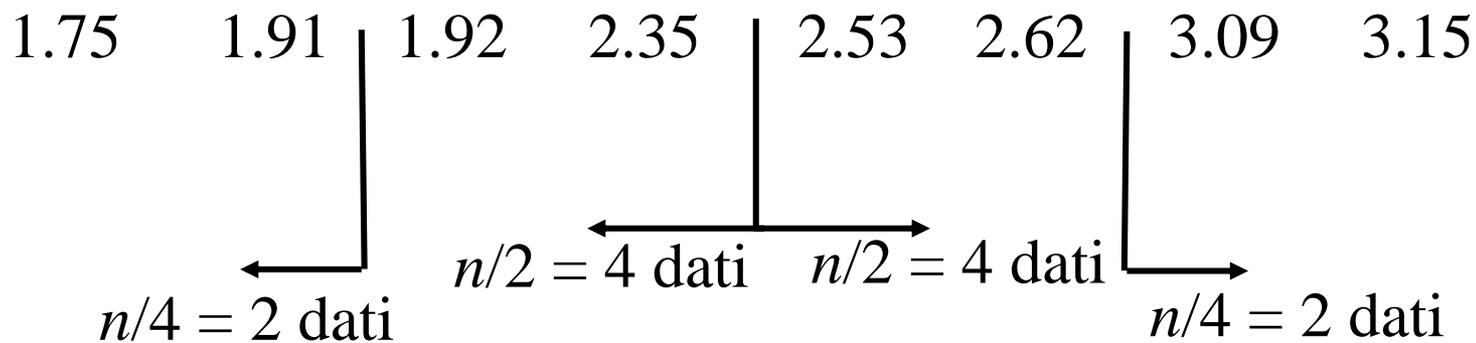
La deviazione standard dei dati nel campione vale

$$\begin{aligned} s(T_{\text{acc}}) &= \sqrt{s^2(T_{\text{acc}})} = \sqrt{\frac{1}{n-1} \sum_{i=1}^n (T_{\text{acc},i} - \bar{T}_{\text{acc}})^2} = \sqrt{\frac{1}{7} \left[\sum_{i=1}^8 T_{\text{acc},i}^2 - 8\bar{T}_{\text{acc}}^2 \right]} \cong \\ &\cong \sqrt{\frac{1}{7} [48.6554 - 8 \cdot 5.8322]} \text{ s}^2 \cong 0.534 \text{ s} \end{aligned}$$

Esercizio su calcolo di media, deviazione standard e box plot

Box plot (tutti i dati, $n=8$, sono in secondi)

Cominciamo col riordinare i dati (in ordine crescente):



Q_1
25-esimo percentile

Q_2 o *Me*
50-esimo percentile

Q_3
75-esimo percentile

Esercizio su calcolo di media, deviazione standard e box plot

$$I_{50}=(n+1)(50/100)=9\times 1/2=4.5 \quad I_{25}=9\times 1/4=2.25 \quad I_{75}=9\times 3/4=6.75$$

$$Me = Q_2 = (2.35 + 2.53)/2 = \mathbf{2.44 \text{ s}}$$

$$Q_2 = 2.35 + (2.53 - 2.35) \cdot (0.5) = \mathbf{2.44 \text{ s}}$$

$$Q_1 = 1.91 + (1.92 - 1.91) \cdot (0.25) = 1.91 + 0.0025 = \mathbf{1.9125 \text{ s}}$$

$$Q_3 = 2.62 + (3.09 - 2.62) \cdot (0.75) = 2.62 + 0.3525 = \mathbf{2.9725 \text{ s}}$$

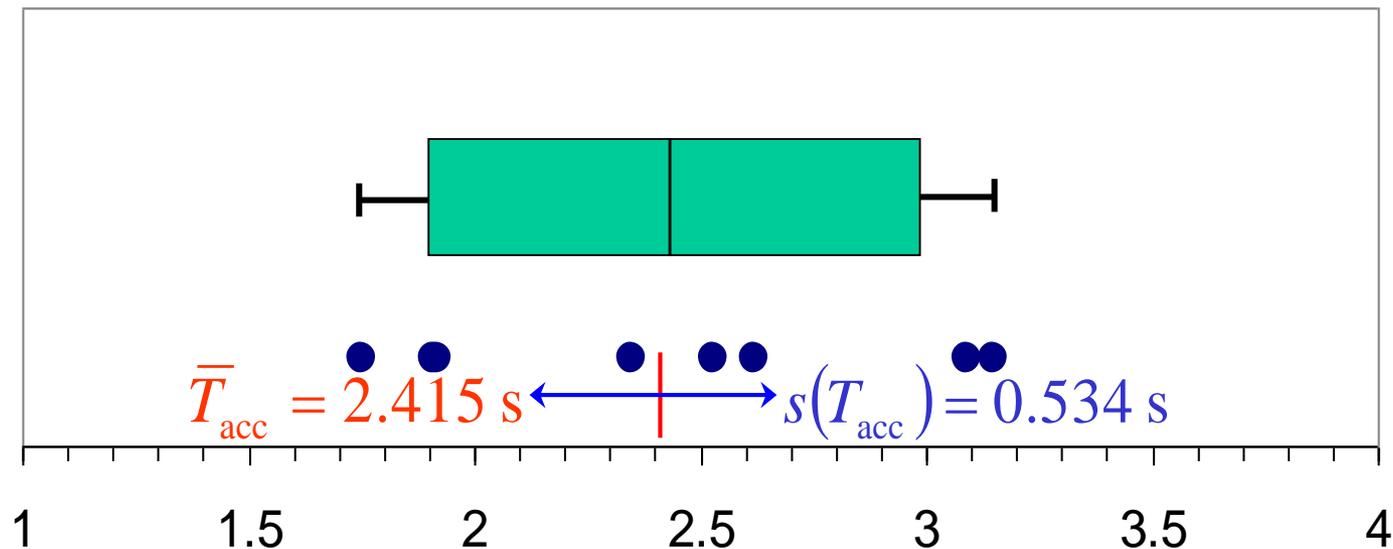
$$DIQ = Q_3 - Q_1 = \mathbf{1.06 \text{ s}}$$

$$W_H \leq Q_3 + 1.5 \cdot DIQ = 2.9725 + 1.5 \cdot 1.06 \cong \mathbf{4.56 \text{ s}}$$

$$W_L \geq Q_1 - 1.5 \cdot DIQ = 1.9125 - 1.5 \cdot 1.06 \cong \mathbf{0.32 \text{ s}}$$

Esercizio su calcolo di media, deviazione standard e box plot

$$Q_1 = 1.9125 \text{ s} \quad Me = 2.44 \text{ s} \quad Q_3 = 2.9725 \text{ s}$$



$$DIQ = 1.06 \text{ s}$$

$$W_L \geq 0.32 = 1.75 \text{ s}$$

$$W_H \leq 4.56 = 3.15 \text{ s}$$

Esercizio su istogrammi, cumulativa e box plot

Tema d'Esame di STATISTICA E MISURAZIONE lunedì 22 novembre 2004

1) Siamo interessati a controllare il nostro consumo domestico (giornaliero) di energia elettrica. Misuriamo quindi i valori del contatore per 13 giorni, ottenendo:

E_i [kWh]: 11.5 9.6 11.3 10.8 12.2 8.9 11.1 10.6 10.9 10.1 9.3 11.4 13.9

1a) Si calcolino media e varianza campionaria del consumo giornaliero di energia, descrivendo il significato e l'utilità di queste statistiche.

1b) Si costruisca un diagramma rami-e-foglie dai dati misurati e un istogramma della frequenza cumulativa.

1c) Si dia la definizione di mediana e di quartile (primo, secondo e terzo) e si rappresenti il box-plot dei dati.

Esercizio su istogrammi, cumulativa e box plot

1a)

$$\text{media campionaria} = \frac{1}{n} \sum_{i=1}^n E_i = 10.8923 \text{ kWh}$$

$$\text{varianza campionaria} = \frac{1}{n-1} \sum_{i=1}^n (T_i - \bar{T}_i)^2 = 1.7074 \text{ (kWh)}^2$$

La **media campionaria** è uno stimatore non polarizzato del valor medio della popolazione che si sta campionando. Stima la **tendenza centrale** della popolazione. Si può dimostrare che è uno stimatore più efficiente del singolo campione, in quanto presenta un errore quadratico medio inferiore (si veda l'inizio del capitolo 4).

La **varianza campionaria** è uno stimatore non polarizzato della varianza della popolazione. È una misura o stima del grado di **dispersione** della popolazione attorno alla media.

Esercizio su istogrammi, cumulativa e box plot

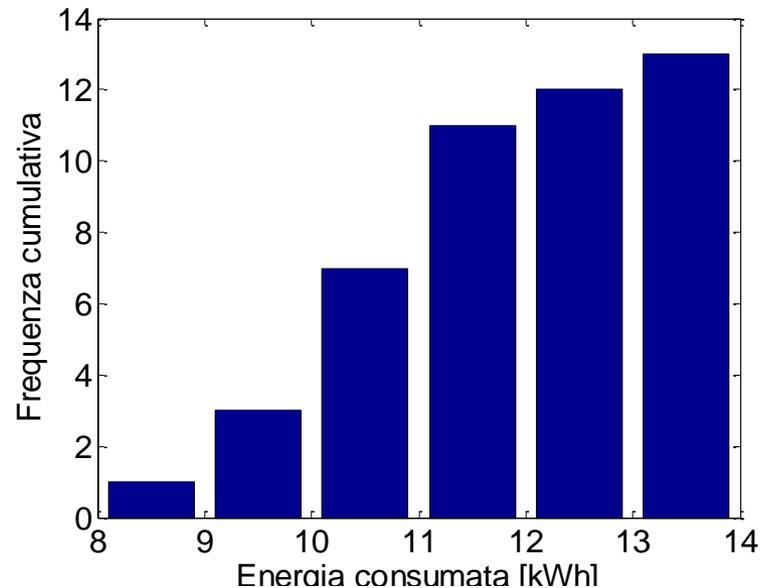
E_i [kWh]: 8.9 9.3 9.6 10.1 10.6 10.8 10.9 11.1 11.3 11.4 11.5 12.2 13.9

1b) Diagramma a ramo e foglia

con $N = 13$

**ISTOGRAMMA
CUMULATIVO**

Rami	Foglie	Freq.	F.cum.
8	9	1	1
9	36	2	3
10	1689	4	7
11	1345	4	11
12	2	1	12
13	9	1	13



F. Relative: F/n

Esercizio su istogrammi, cumulativa e box plot

1c) Mediana: è una misura del centro del campione che divide i dati in due parti uguali (tante misure al di sopra quante al di sotto). Se il numero di dati è pari, la mediana è a metà tra i due valori centrali. Se dispari, la mediana coincide con il valore centrale. Fornisce un'idea della tendenza centrale dei dati.

Primo quartile (quartile basso o di sx): valore tale che un quarto delle osservazioni abbia un valore inferiore; corrisponde al 25-esimo percentile. (tre quarti delle osservazioni hanno invece un valore superiore)

Secondo quartile: valore tale che due quarti delle osservazioni abbiano un valore inferiore; corrisponde al 50-esimo percentile e dunque alla mediana. (due quarti delle osservazioni hanno un valore superiore)

Terzo quartile (quartile alto o di dx): valore tale che tre quarti delle osservazioni abbiano un valore inferiore; corrisponde al 75-esimo percentile. (un quarto delle osservazioni ha invece un valore superiore)

Esercizio su istogrammi, cumulativa e box plot

Per disegnare il box-plot dei dati dobbiamo innanzitutto calcolare la mediana e il primo e terzo quartile. La formula generale per ricavare l'indice di un generico k -esimo percentile è: $I_k = (n+1) \times k / 100$. Dall'indice si ricava quindi il valore esatto con un'interpolazione lineare tra i due dati (con indici uguali all'intero precedente e successivo a I_k).

I 13 dati ordinati sono E_i [kWh]: 8.9 9.3 9.6 10.1 10.6 10.8 10.9 11.1 11.3 11.4 11.5 12.2 13.9.

La mediana è il 50-esimo percentile, per cui $I_{\text{mediana}} = (13+1) \times 50 / 100 = 7$, quindi la mediana coincide con il settimo dato (ricordiamo infatti che per un numero dispari di dati corrisponde al valore centrale). Quindi, **mediana = 10.9 kWh = $Me = Q_2$** .

Il primo quartile è il 25-esimo percentile, per cui $I_{1Q} = (13+1) \times 25 / 100 = 3.5$, quindi il primo quartile si trova tra il terzo ed il quarto dato, per cui: primo quartile = $9.6 + (10.1 - 9.6) \cdot (0.5) = \mathbf{9.85 \text{ kWh} = Q_1}$.

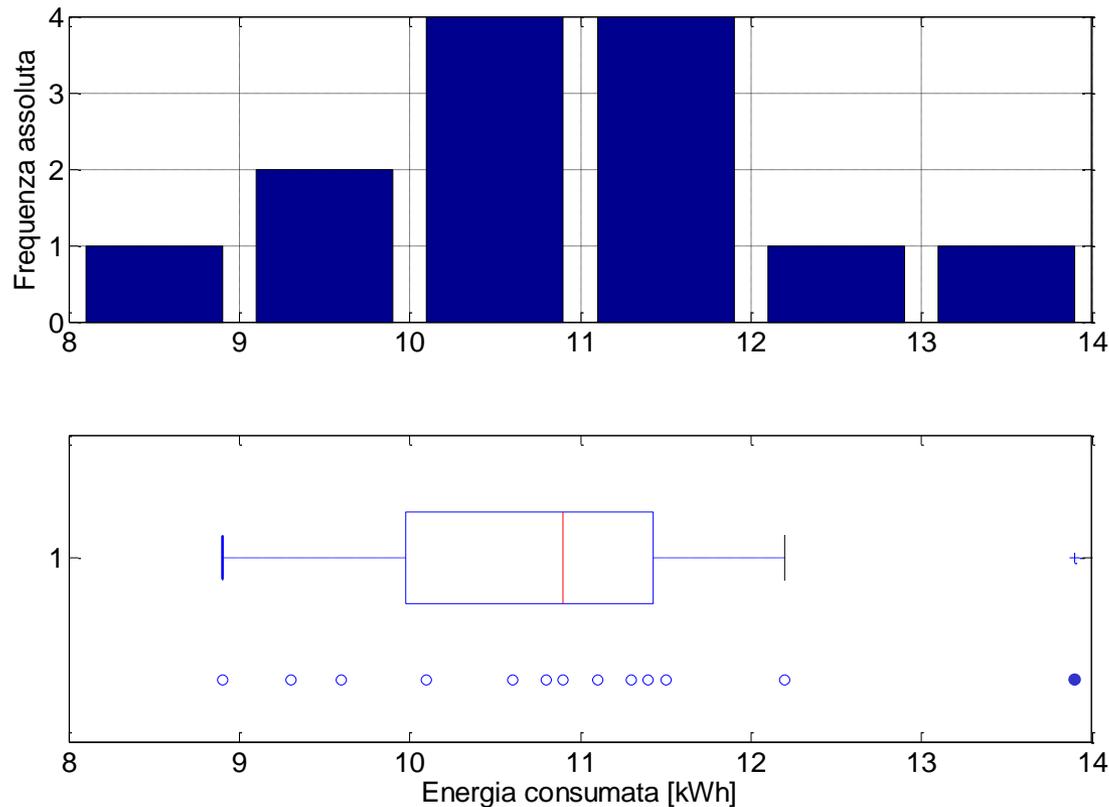
Il terzo quartile è il 75-esimo percentile, per cui $I_{3Q} = (13+1) \times 75 / 100 = 10.5$, quindi il terzo quartile si trova tra il decimo e l'undicesimo dato: terzo quartile = $11.4 + (11.5 - 11.4) \cdot (0.5) = \mathbf{11.45 \text{ kWh} = Q_3}$.

La dinamica interquartile vale **$DIQ = Q_3 - Q_1 = 1.6 \text{ kWh}$** . I baffi si possono estendere fino a $1.5 \times DIQ = 2.4 \text{ kWh}$ dai quartili, quindi fino a $9.85 - 2.4 = 7.45 \text{ kWh}$ e $11.45 + 2.4 = 13.85 \text{ kWh}$.
Minimo dato = 8.9 kWh. Massimo dato = 13.9 kWh.

Sulla parte inferiore del box-plot tutti i dati sono contenuti nel baffo, che dunque termina a 8.9 kWh, ma nella parte superiore non tutti i dati sono contenuti nel baffo, che dunque termina a 12.2 kWh.

Ricordiamo che i baffi si fermano sempre all'ultimo dato contenuto nel campione considerato.

Esercizio su istogrammi, cumulativa e box plot



In figura sono riportati anche i singoli punti sperimentali e (sopra) l'istogramma della frequenza assoluta