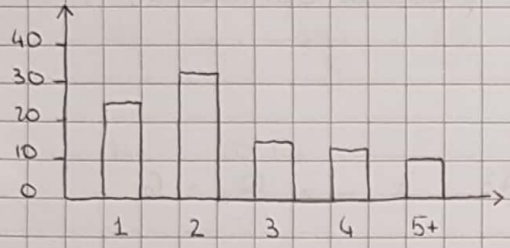


a) AMPIEZZA DELLE FAMIGLIE	FREQ. ASSOLUTE	FREQ. RELATIVE
1 comp.	30.1	0.27
2 comp.	37.0	0.33
3 comp.	17.8	0.16
4 comp.	15.3	0.14
5+ comp.	10.9	0.10
Tot.	111.1	1

b) Istogramma FREQ. RELATIVE



Distribuzione asimmetrica positiva
 ↳ coda di dx più lunga

c) Mediana?
 Moda?

$$Me \rightarrow \frac{n+1}{2} = \frac{111.1+1}{2} = 56.05 = 56\text{-esima osservazione}$$

↳ Me = MODALITA' 2 COMPONENTI

Moda => valore osservato con maggior frequenza

Moda = MODALITA' 2 COMPONENTI

Mediana e moda coincidono.

Es. 3.7 PAG. 65

Stato	N° di aborti
W	26
O	17
C	236
A	2
H	6

a) $\bar{x} = \frac{\sum x_i}{n} = \frac{26+17+236+2+6}{5} = 57.4$

b) $Me \Rightarrow \frac{n+1}{2} = \frac{5+1}{2} = 3$

ordino i dati 2 6 17 26 236

Me = 17

La media risente dei valori outlier, mentre la mediana è meno influenzata da valori outlier. In questo caso Me è più appropriata, perché \bar{x} non rappresenta bene la tipicità della distribuzione considerando il valore 236 come è outlier.

ULTERIORI RICHIESTE

Aggiungiamo i seguenti dati: NY (12), NJ (19), CO (29), V (24), M (1)

$\bar{x} = ?$ $Me = ?$

$$\bar{x} = \frac{\overset{\text{→ somma dei valori precedenti}}{287 + 12 + 19 + 29 + 24 + 1}}{10} = 37.2$$

$$Me \Rightarrow \frac{n+1}{2} = \frac{10+1}{2} = 5.5$$

ordinio i valori 1 2 6 12 (17) (19) 24 26 29 236
 media = $\frac{17+19}{2} = 18$

Me = 18 (di nuovo molto diversa da \bar{x} per la presenza dell'outlier)

SOMMARIO A 5 NUMERI

Min = 1
 Q1 = 6
 Me = 18
 Q3 = 26
 Max = 236

$$Q_1 \Rightarrow \frac{n+1}{4} = \frac{11}{4} = 2.75 = 3 \rightarrow Q_1 = 6$$

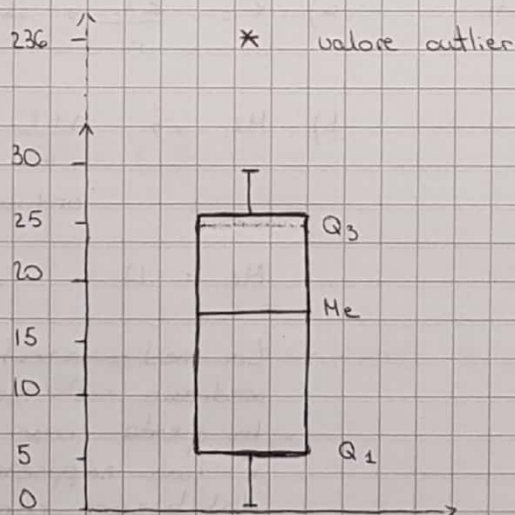
$$Q_3 \Rightarrow \frac{3 \cdot (n+1)}{4} = \frac{3 \cdot 11}{4} = 8.25 = 8 \rightarrow Q_3 = 26$$

Ci sono outlier? Sconto ^(IQR) interquartile = $Q_3 - Q_1 = 26 - 6 = 20$
 $Q_1 - 1.5 \cdot IQR = 6 - 1.5 \cdot 20 = \text{valore neg.}$
 ↳ no outlier nella parte inferiore della distribuzione
 $Q_3 + 1.5 \cdot IQR = 56$
 ↳ tutti i valori superiori a 56 sono outlier.

↳ 236 è valore outlier

BOXPLOT

Valore adiacente inferiore = 1
 " superiore = 29



STATO	SALARIO MIN	Me = ?
A	10.00	$\bar{x} = ?$
N	10.25	campo di variazione ?
F	10.46	s = ?
UK	10.01	
USA	5.15	

a) Escludendo gli USA (n=4)

Me $\Rightarrow \frac{n+1}{2} = 2.5 \rightarrow$ dati ordinati 10.00 10.01 10.25 10.46

media = 10.13

Me = 10.13

molto simili

$\bar{x} = \frac{40.72}{4} = 10.18$

Range di variazione = max - min = 10.46 - 10.00 = 0.46

$s = \sqrt{\frac{\sum (x_i - \bar{x})^2}{n-1}} = \sqrt{\frac{0.144}{4-1}} = \sqrt{0.048} = 0.22$

$(10.00 - 10.18)^2 = (-0.18)^2 = 0.032$
 $(10.25 - 10.18)^2 = (0.07)^2 = 0.005$
 $(10.46 - 10.18)^2 = (0.28)^2 = 0.078$
 $(10.01 - 10.18)^2 = (-0.17)^2 = 0.029$

 0.144

b) Includendo gli USA (n=5)

Me = 10.01
 $\bar{x} = \frac{45.87}{5} = 9.17$

Range di variazione = 10.46 - 5.15 = 5.31

$s = \sqrt{\frac{20.385}{5-1}} = \sqrt{5.096} = 2.26$

Includere gli USA comporta un abbassamento della media e un aumento considerevole del range di variazione e della dev. st.

Stato	N° medio gg. vacanza
IT	42
F	37
G	35
B	34
UK	28
C	26
GiAP.	25
USA	13

a) $\bar{x} = \frac{42+37+\dots+13}{8} = 30$

$s = \sqrt{\frac{568}{8}} = \sqrt{71} = 8.43$

La media è 30 e la dev. st. 8.43, quindi mi aspetto abbastanza variabilità nei dati.

b) Sintesi a 5 numeri

- Min = 13
- Q₁ = 25
- Me = 31
- Q₃ = 37
- Max = 42

analisi i dati

- 13 25 26 28 34 35 37 42

Q₁ → $\frac{n+1}{4} = 2.25 = 2 \rightarrow Q_1 = 25$

Me → $\frac{n+1}{2} = 4.5 \rightarrow$ faccio la media tra i valori in posizione 4 e 5.

→ Me = $\frac{28+34}{2} = 31$

Q₃ → $\frac{3 \cdot (n+1)}{4} = 6.75 = 7 \rightarrow Q_3 = 37$

ES N.3.26 PAG. 68

$\bar{x} = 76$

- 20 e' impossibile perche' s ≥ 0
- 0 e' improbabile perche' s = 0 se tutti i valori sono uguali
- 50 e' improbabile perche' il range dei voti e' compreso tra 0 e 100

⇒ il valore più plausibile e' 10

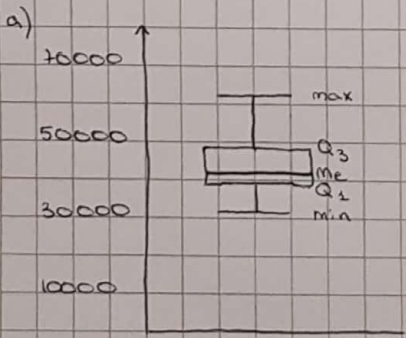
ES N. 3.31 PAG. 68

- Max = 61800
- Q₃ = 48850
- Me = 42700
- Q₁ = 39250
- Min = 33100

a) Campo di variazione = max - min = 61800 - 33100 = 28700
 ↓
 differenza tra il valore max e quello min, dal min se aggiungo 28700 arrivo al valore max.

b) IQR = Q₃ - Q₁ = 48850 - 39250 = 9600
 ↓
 mi indica la dispersione dei valori centrali della distribuzione.
 La parte centrale dei dati ricade in un range di ampiezza pari a 9600.

ES N. 3.32 PAG. 68



b) La distribuzione e' asimmetrica positiva (coda di dx più lunga)

c) Se distribuz campanulone il valore più plausibile per s sarebbe 7000

- 100 : troppo piccolo (*)
- 1000 : il 68% dei dati dovrebbe essere compreso tra 41700 - 43700 → improbabile
- 25000 : troppo grande (*)
- (*) Mi baso sulla regola empirica