Stat. descrittiva per v. numeriche

$x_1 \dots x_n$ val. di X su n individui

Media

$$\mu_X = \frac{x_1 + \dots + x_n}{n}$$

una "misura di tendenza centrale"

Varianza

$$\sigma_X^2 = \frac{(x_1 - \mu)^2 + (x_2 - \mu)^2 + \dots + (x_n - \mu)^2}{n}$$

$$= \frac{x_1^2 + \dots + x_n^2}{n} - \mu^2$$

Deviazione standard

$$\sigma_X = \sqrt{\sigma_X^2}$$

$\mu - \sigma$ μ $\mu + \sigma$

Es 1 2 2 3 3 3 4 4 5

$$3 = \mu = \frac{27}{9} = 3$$

$$\sigma^2 = \frac{(1-3)^2 + (2-3)^2 + (2-3)^2 + \dots}{9}$$

$$= \frac{4 + 1 + 1 + 0 + 0 + 0 + 1 + 1 + 4}{9}$$

$$= \frac{12}{9} \approx 1.34$$

$$\sigma = \sqrt{1.34} \approx 1.16$$

Fatto

Almeno

Il 75% degli individui ha sempre

$$\mu - 2\sigma \leq x \leq \mu + 2\sigma$$

Varianza campionaria

$$s^2 = \frac{(x_1 - \mu)^2 + (x_2 - \mu)^2 + \dots + (x_n - \mu)^2}{n-1}$$

$$s^2 = \sigma^2 \frac{n}{n-1}$$

$$s = \sqrt{s^2}$$

E.g. per i dati precedenti:

$$s^2 = \sigma^2 \cdot \frac{9}{8} = \frac{12}{8} = 1.5$$

$$\sigma^2 = \frac{12}{9}$$

Mediana

$x_1 = 7 \quad x_2 = 1 \quad x_3 = 5 \quad x_4 = 8 \quad x_5 = 10$

↓ in ordine crescente.

$x_{(1)} = 1 \quad x_{(2)} = 5 \quad x_{(3)} = 7 \quad x_{(4)} = 8 \quad x_{(5)} = 10$

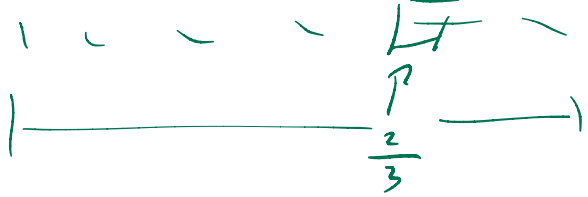
$$\tilde{x} = x_{(3)} = 7$$

$$\tilde{x} = \begin{cases} x_{\left(\frac{n+1}{2}\right)} & \text{se } n \text{ è dispari} \\ \frac{1}{2} x_{\left(\frac{n}{2}\right)} + \frac{1}{2} x_{\left(\frac{n}{2}+1\right)} & \text{pari} \end{cases}$$

Quantili

$\lfloor 5.21 \rfloor = 5$

$frac(5.21) = 0.21$



$k = 1 + (n-1)\,p$

$\lfloor k \rfloor$ p. intera

$frac(k)$ p. froz.

$$q_p = x_{(\lfloor k \rfloor)} + frac(k) \cdot \left(x_{(\lfloor k \rfloor + 1)} - x_{(\lfloor k \rfloor)} \right)$$

es. $\frac{2}{3}$

Esempio
Cerchiamo lo 0.75-esimo quantile di "terzo quartile"

$x_1 = 4.1 \quad x_2 = 6.2 \quad x_3 = 1.0 \quad x_4 = 2.7 \quad x_5 = 5.6 \quad x_6 = 4.1$

$x_{(1)} = 1.0 \qquad x_{(2)} = 2.7 \qquad x_{(3)} = 4.1$

$x_{(4)} = 4.1 \qquad x_{(5)} = 5.6 \qquad x_{(6)} = 6.2$

$k = 1 + 0.75\,(n-1) = 1 + \frac{3}{4} \cdot 5 = 4.75$

$\lfloor k \rfloor = 4 \qquad frac(k) = 0.75$

$x_{(4)} = 4.1$

$x_{(5)} = 5.6$

$q_{0.75} = x_{(4)} + 0.75 \cdot (x_{(5)} - x_{(4)})$

$= 4.1 + 0.75 \cdot 1.5$

$= 5.225$

Decili (Percentili) Quartili

$Q_1 = q_{1/4}$

$Q_2 = q_{1/2} = \tilde{x}$

$Q_3 = q_{3/4}$

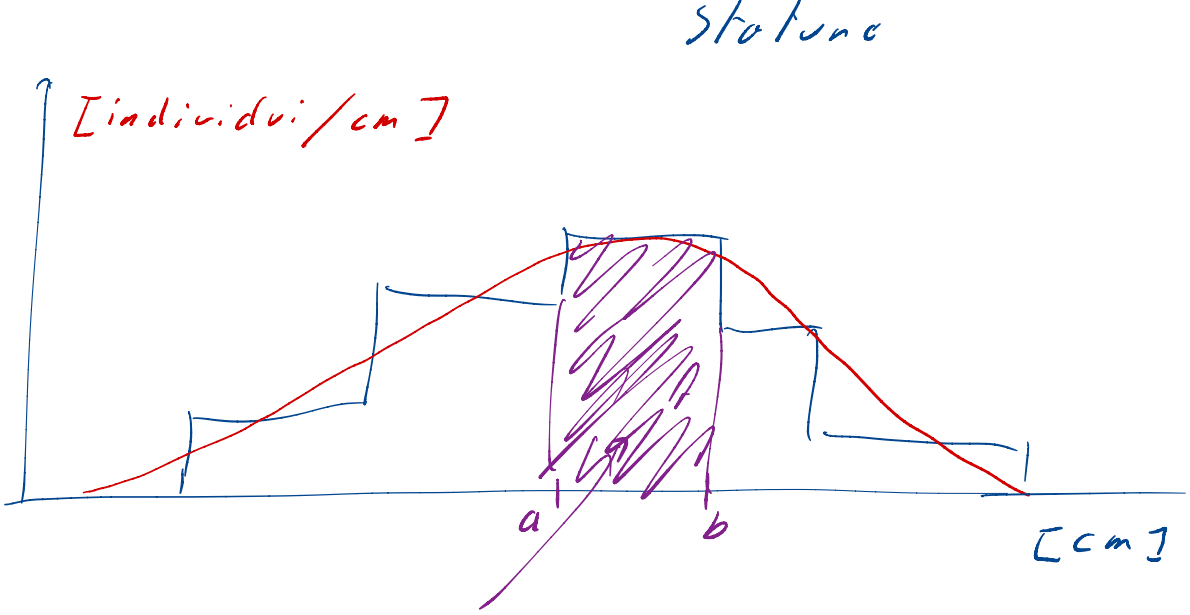
$IQR = Q_3 - Q_1$

"differenza interquartile"

x-esimo decile = $q_{x/10}$

x-esimo centile = $q_{x/100}$

Istogrammi



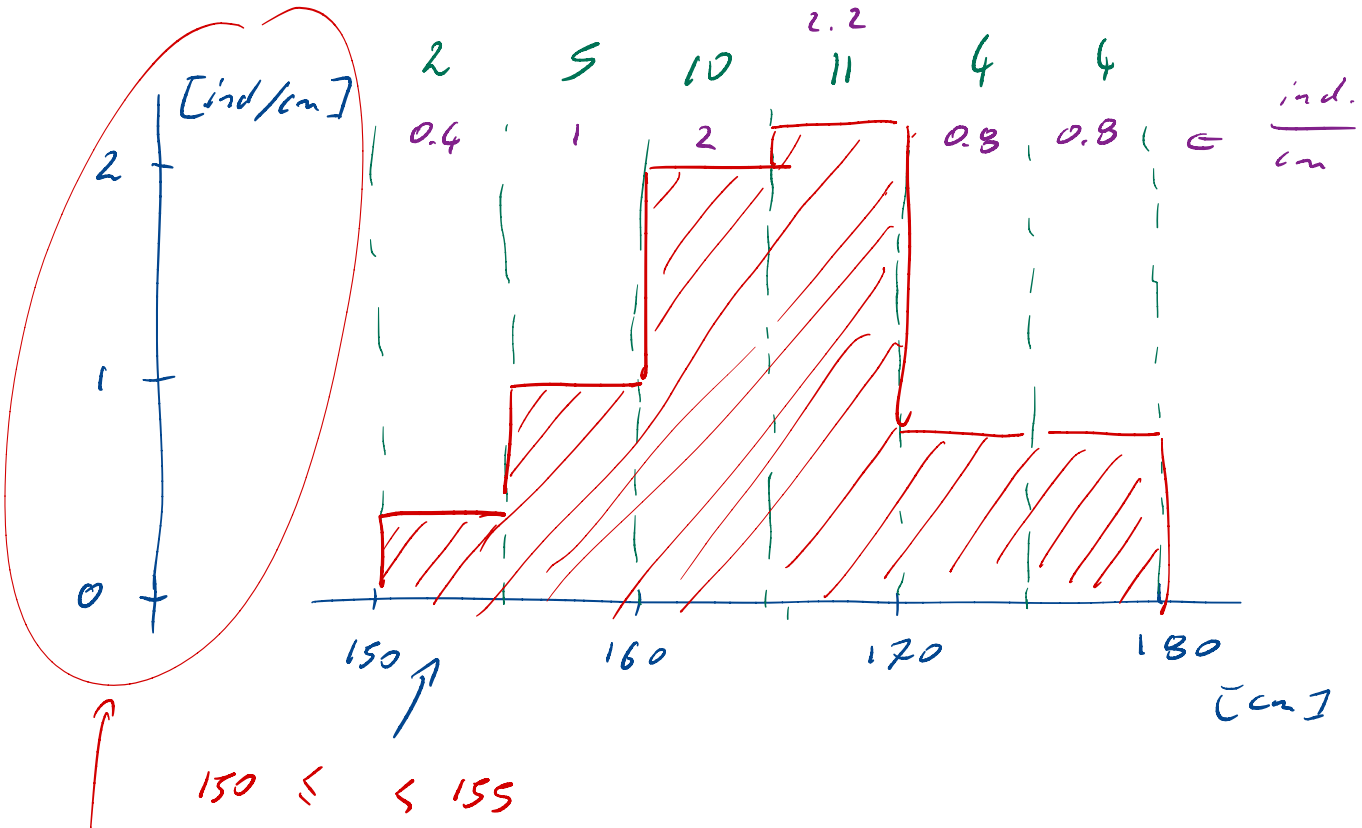
A = n. di individui fra a e b

decine

	[cm]
15	0468899
16	00001333445555778889
17	00115588

150, 154, 156, 158, ...

15	0468899
16	000013334455555778889
17	00115588

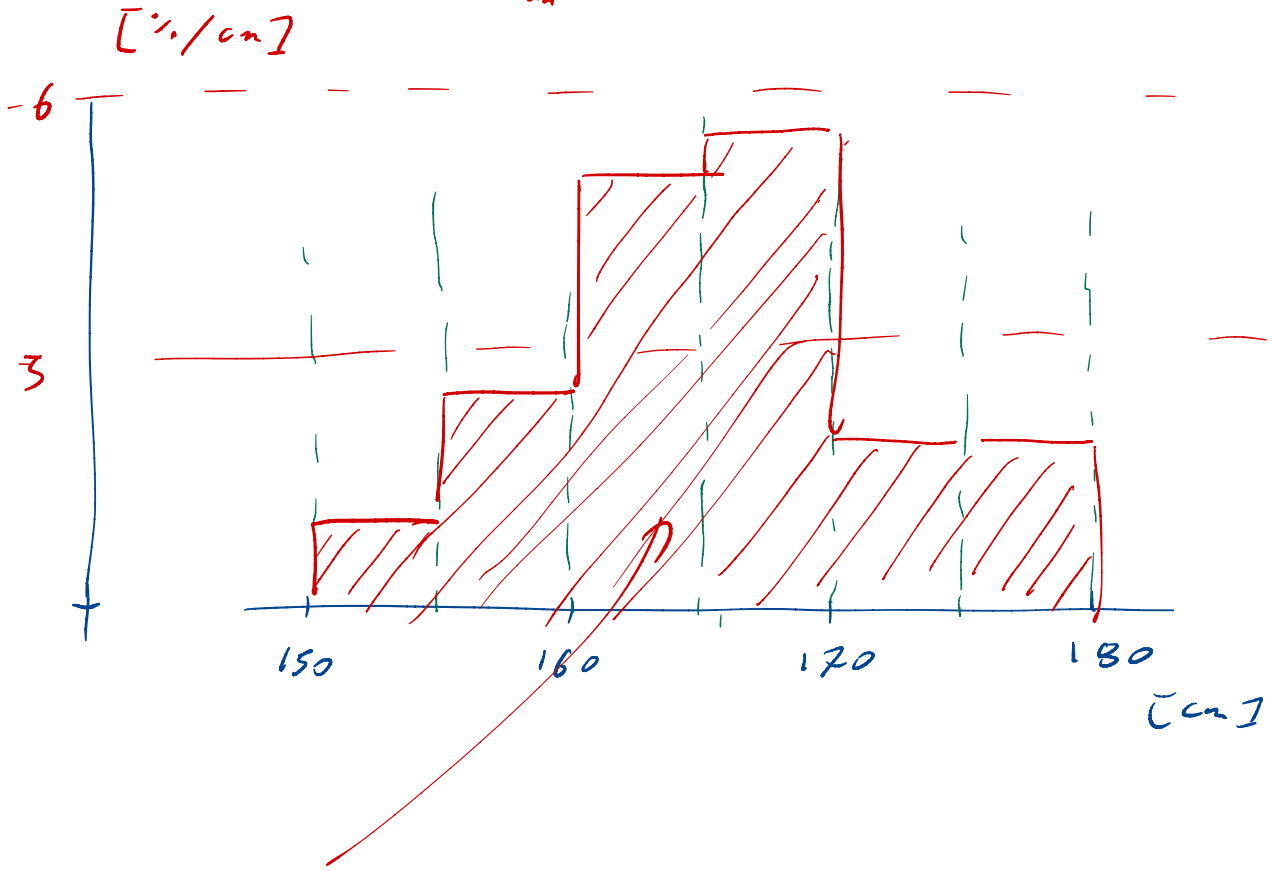


$150 \leq \quad < 155$

"densità di frequenza"

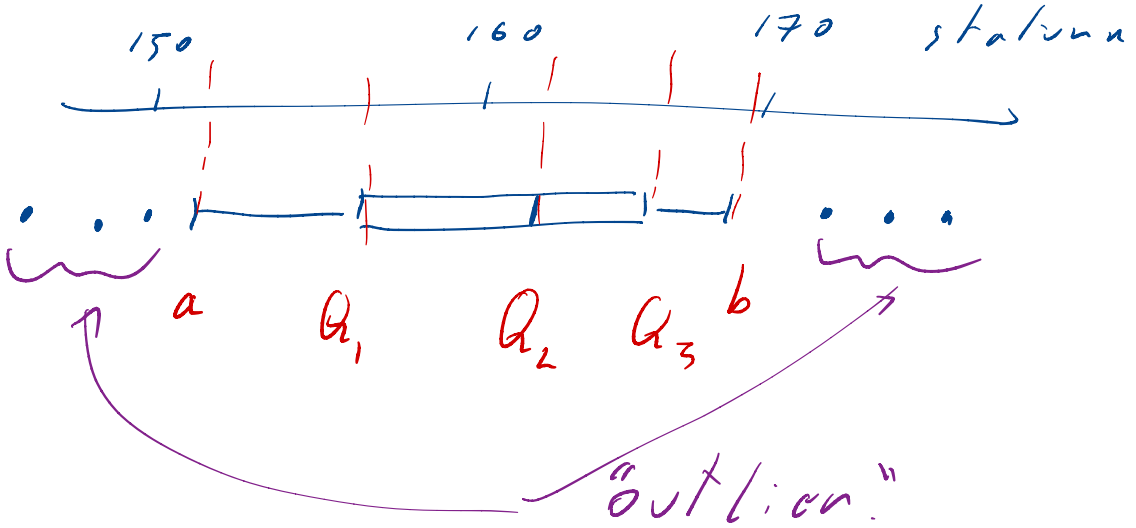
"densità di probabilità"

2	5	10	11	4	4	
0.4	1	2	2.2	0.8	0.8	← ind./cm
	↓	↓				
	14%	28%				
	2.8 %/cm	5.6 %/cm				



A = 100 %

Box plot



b = max dei valori sotto

$Q_3 + 1.5 \cdot IQR$

a = min ——— sopra

$Q_1 - 1.5 \cdot IQR$