Esercizio 1

Ad una gara podistica hanno partecipato 100 persone.

La tabella riporta i loro tempi di percorrenza (in minuti) raggruppati in classi:

classi x_i	frequenze n_i
0 |-- 30	60
30 |-- 50	30
50 |-- 80	10

(a) Di che tipo di carattere si tratta? Fornire una rappresentazione grafica opportuna

x_i classi	n_i freq. assolute	f_i freq. relative		
0 |-- 30	60	60/100 = 0.6		
30 |-- 50	30	30/100 = 0.3		
50 |-- 80	10	10/100 = 0.1		
	100	**1**		

Il totale deve risultare 1

numero totale partecipanti alla gara

X = "tempo impiegato"

Il carattere X

è quantitativo continuo

↓

istogramma

Con riferimento alle classi:

- la frequenza assoluta è il **numero** di partecipanti in ciascuna classe
- la frequenza relativa è la **proporzione** di partecipanti in ciascuna classe
- la frequenza relativa$\times$ 100% è la **percentuale** di partecipanti in ciascuna classe

x_i classi	n_i freq. assolute	f_i freq. relative	a_i ampiezza classi	
0 |-- 30	60	60/100 = 0.6	30-0 = 30	
30 |-- 50	30	30/100 = 0.3	50-30 = 20	
50 |-- 80	10	10/100 = 0.1	80-50 = 30	
	100	**1**		

X = "tempo impiegato"

Il carattere X è

quantitativo continuo

↓

istogramma

x_i **classi**	n_i **freq. assolute**	f_i **freq. relative**	a_i **ampiezza classi**	$d_i = f_i / a_i$ **densità**
0 |-- 30	60	60/100 = 0.6	30-0 = 30	0.6/30 = 0.020
30 |-- 50	30	30/100 = 0.3	50-30 = 20	0.3/20 = 0.015
50 |-- 80	10	10/100 = 0.1	80-50 = 30	0.1/30 = 0.003
	100	**1**		

X = “tempo impiegato”

Il carattere X

è quantitativo continuo

↓

istogramma

x_i classi	n_i freq. assolute	f_i freq. relative	a_i ampiezza classi	$d_i = f_i / a_i$ densità
0 |-- 30	60	60/100 = 0.6	30-0 = 30	0.6/30 = 0.020
30 |-- 50	30	30/100 = 0.3	50-30 = 20	0.3/20 = 0.015
50 |-- 80	10	10/100 = 0.1	80-50 = 30	0.1/30 = 0.003
	100	**1**		

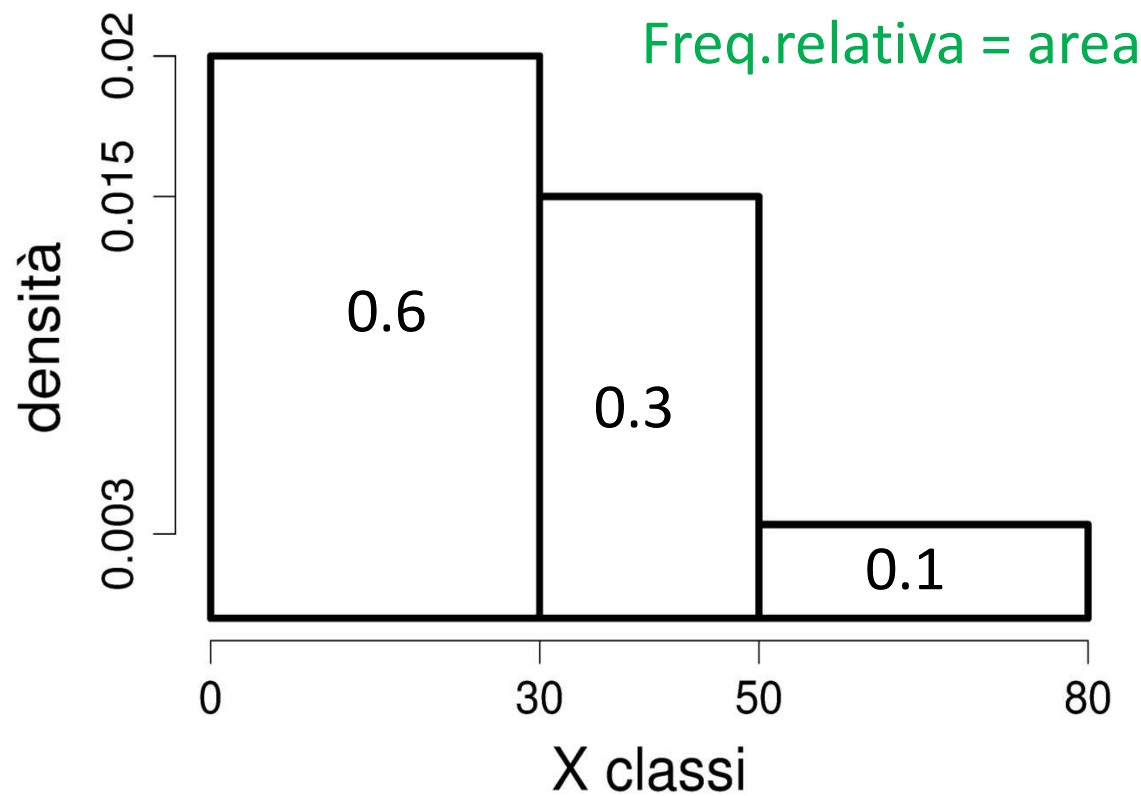
X = “tempo impiegato”

Il carattere X

è quantitativo continuo

↓

istogramma



x_i classi	n_i freq. assolute	f_i freq. relative	a_i ampiezza classi	$d_i = f_i / a_i$ densità
0 |-- 30	60	60/100 = 0.6	30-0 = 30	0.6/30 = 0.020
30 |-- 50	30	30/100 = 0.3	50-30 = 20	0.3/20 = 0.015
50 |-- 80	10	10/100 = 0.1	80-50 = 30	0.1/30 = 0.003
	100	**1**		

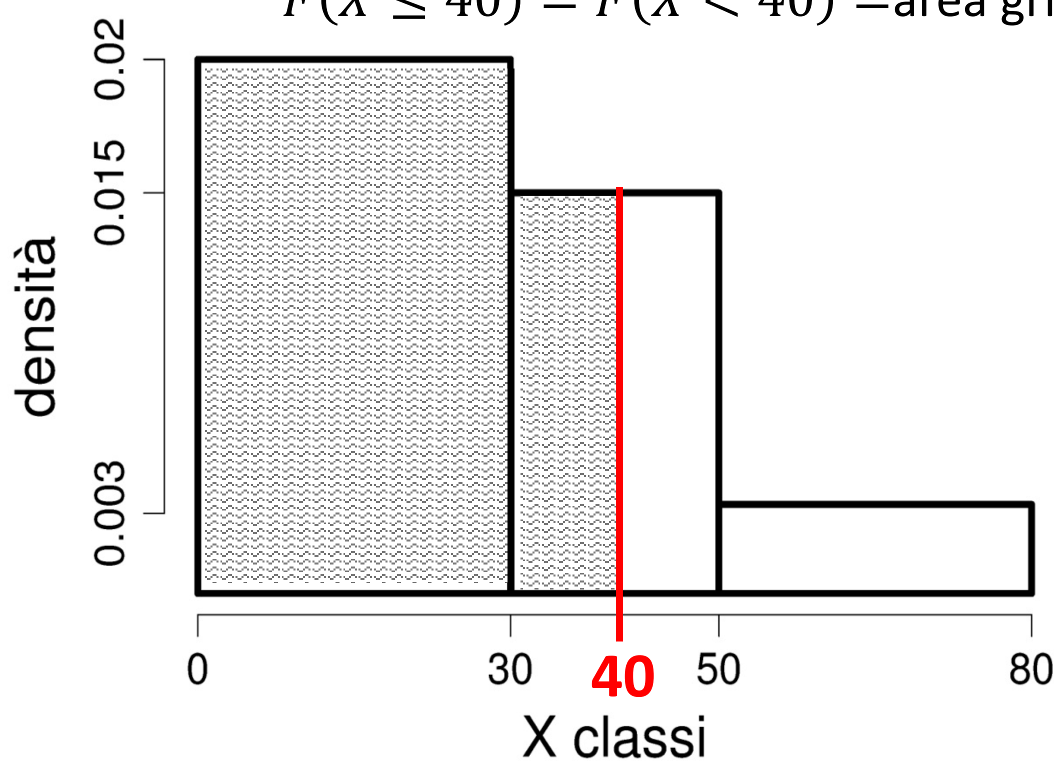
Proporzione di partecipanti che hanno impiegato meno di 40 minuti:

$$F(X < 40) =$$
$$= 0.6 + (40 - 30) \times 0.015 =$$
$$= 0.75$$

La proporzione è 0.75

La percentuale è 75%

$F(X \leq 40) = F(X < 40) =$area grigia



x_i classi	n_i freq. assolute	f_i freq. relative	a_i ampiezza classi	$d_i = f_i / a_i$ densità
0 |-- 30	60	60/100 = 0.6	30-0 = 30	0.6/30 = 0.020
30 |-- 50	30	30/100 = 0.3	50-30 = 20	0.3/20 = 0.015
50 |-- 80	10	10/100 = 0.1	80-50 = 30	0.1/30 = 0.003
	100	**1**		

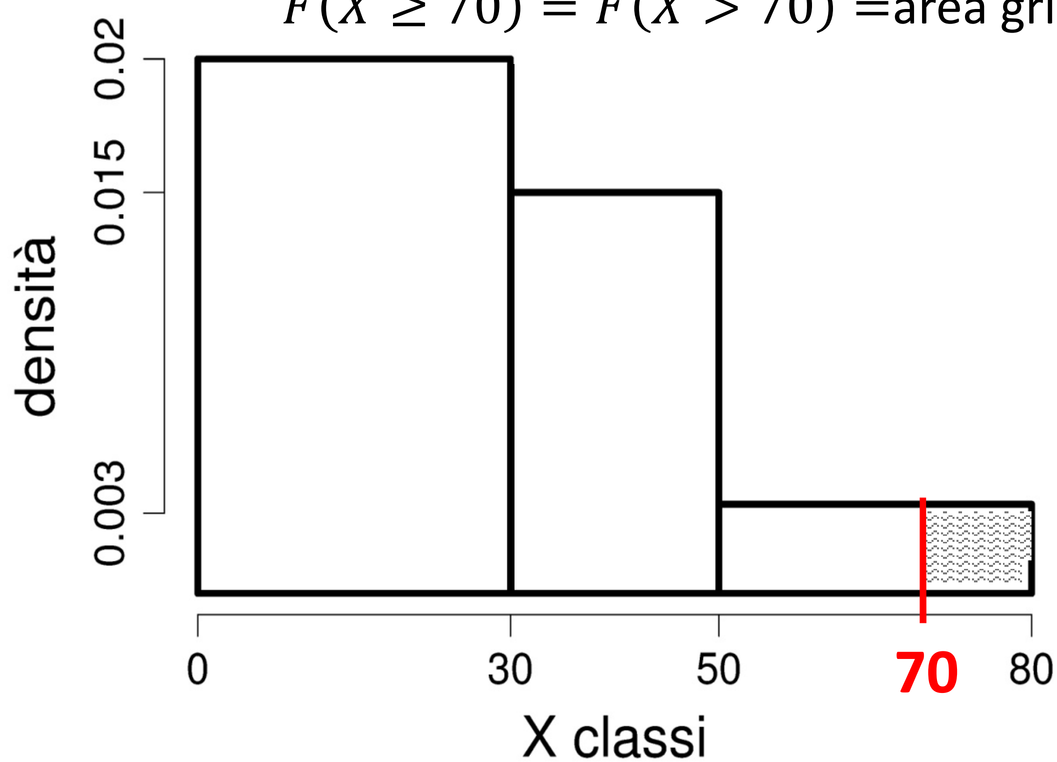
Proporzione di partecipanti che hanno impiegato più di 70 minuti:

$$F(X > 70) =$$
$$= (80 - 70) \times 0.003 =$$
$$= 0.03$$

La proporzione è 0.03

La percentuale è 3%

$F(X \geq 70) = F(X > 70) =$area grigia



QUIZ 1

L'istogramma è una rappresentazione grafica adatta per caratteri:

(a) categorici ordinali
(b) qualitativi nominali
(c) quantitativi raggruppati in classi
(d) nessuno dei precedenti

Sugli assi orizzontale e verticale, rispettivamente, di un istogramma si rappresentano:

(a) frequenze assolute e frequenze relative
(b) ampiezza delle classi e frequenze relative
(c) classi e frequenze assolute
(d) classi e densità di frequenza

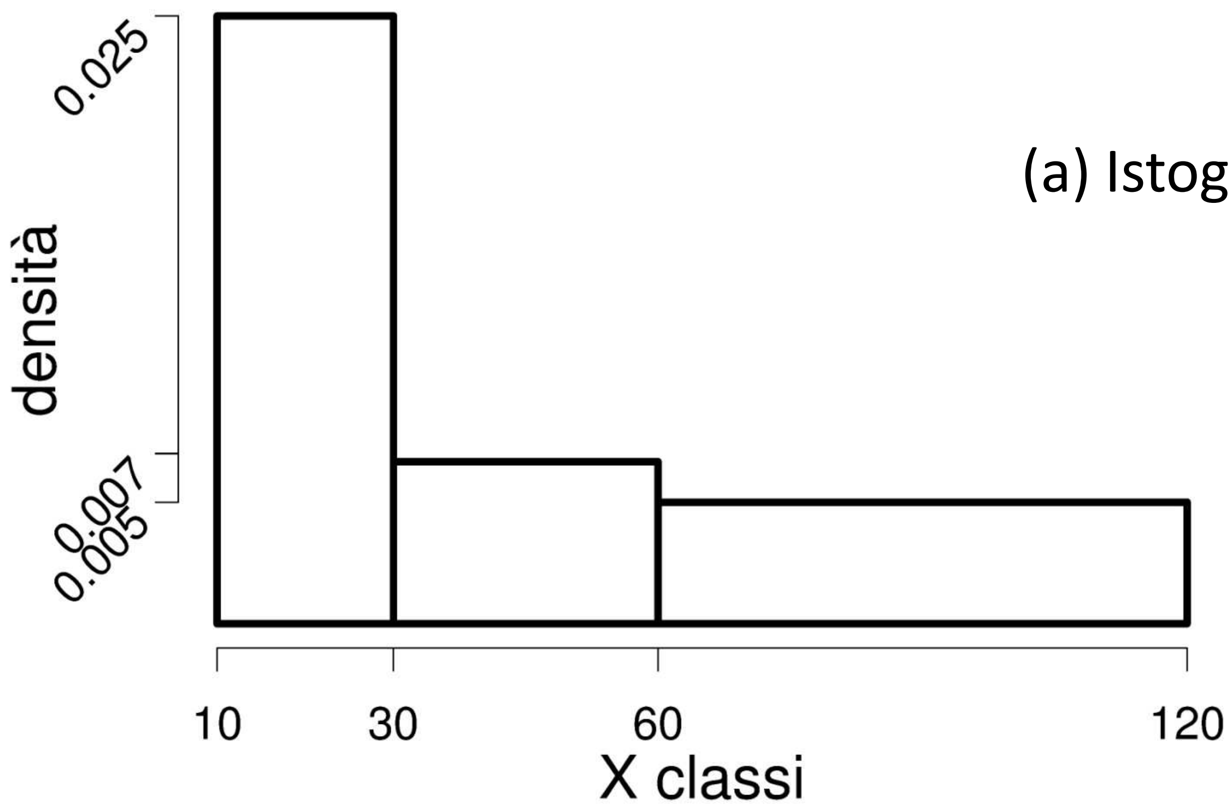
Esercizio 2

Alla stazione Porta Garibaldi sono stati intervistati 200 pendolari che hanno dichiarato la loro spesa mensile in abbonamenti a mezzi pubblici. I risultati sono i seguenti:

classi x_i	frequenze n_i
10 |-- 30	100
30 |-- 60	40
60 |-- 120	60

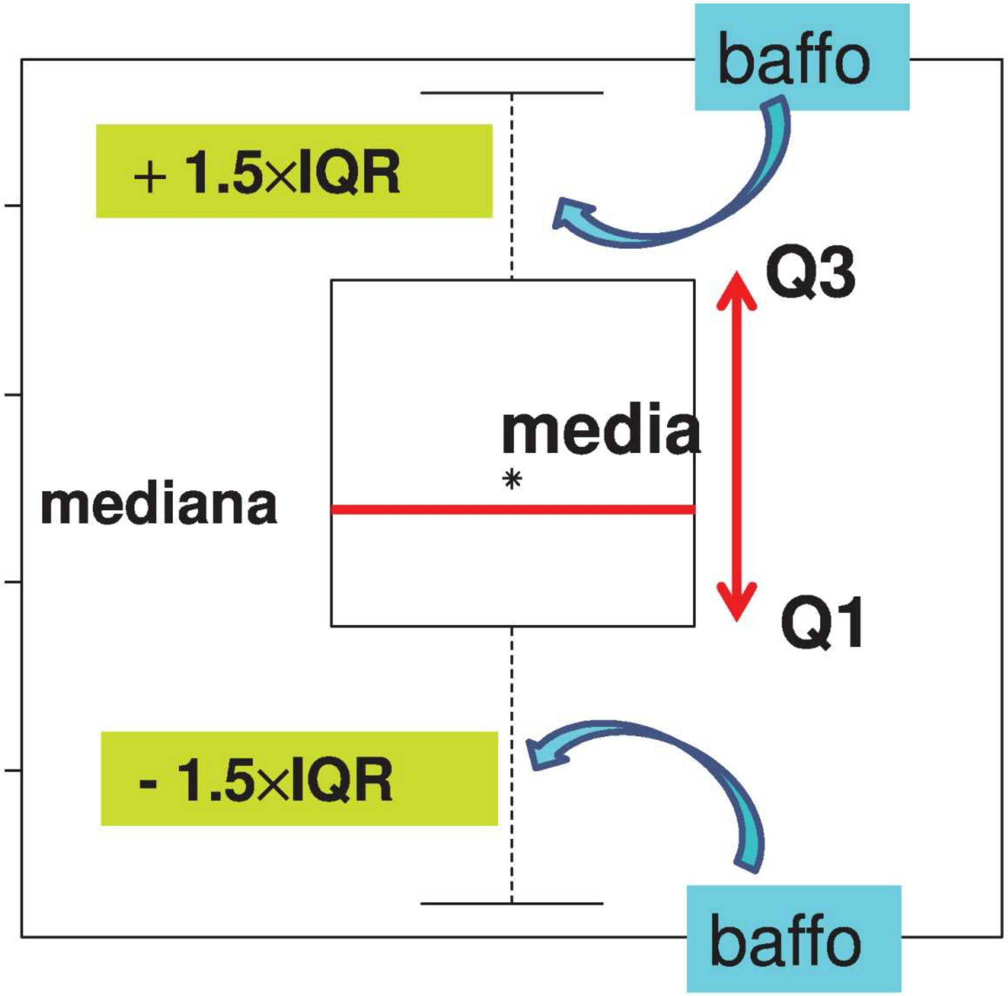
(a) Fornire una rappresentazione grafica opportuna
(b) Disegnare il box-plot
(c) Calcolare moda e coefficiente di variazione

x_i classi	n_i freq. assolute	f_i freq. relative	a_i ampiezza classi	$d_i = f_i / a_i$ densità
10 |-- 30	100	100/200 = 0.5	30-10 = 20	0.5/20 = 0.025
30 |-- 60	40	40/200 = 0.2	60-30 = 30	0.2/30 = 0.007
60 |-- 120	60	60/200 = 0.3	120-60 = 60	0.3/60 = 0.005
	200	**1**		



(a) Istogramma

(b) Box-plot



- **Baffo superiore** = Q3 + 1.5 x IQR

- **Q3** = Terzo quartile
- **Q2** = Mediana
- ***** = Media
- **Q1** = Primo quartile

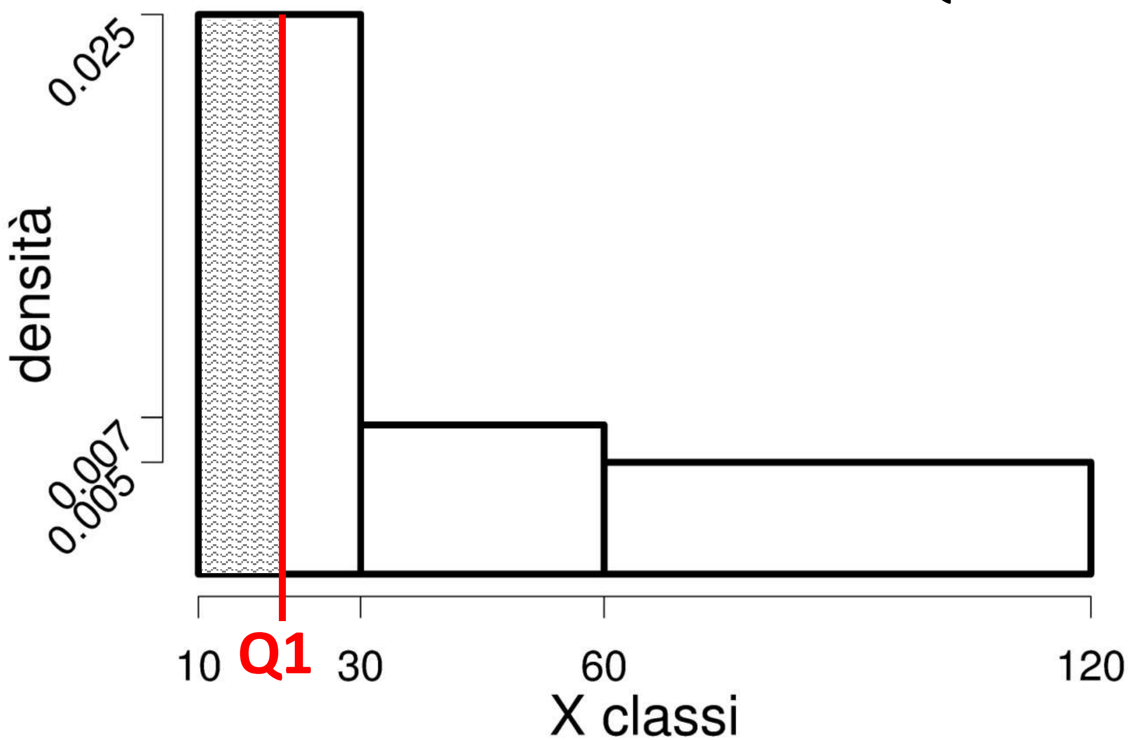
- **Baffo inferiore** = Q1 - 1.5 x IQR
- Eventuali **outliers**

dove **IQR** = Q3-Q1 è il range interquartile

x_i	n_i	f_i	a_i	d_i
10 |-- 30	100	0.5	20	0.025
30 |-- 60	40	0.2	30	0.007
60 |-- 120	60	0.3	60	0.005
	200	**1**		

Primo quartile $\mathbf{F(X < Q1) = 0.25}$

La prima colonna dell'istogramma ha area **0.5** e Q1 stacca un'area di **0.25**. Quindi Q1 sta nella prima classe.



Ragionando sull'area:

$$(Q1 - 10) \times 0.025 = 0.25$$

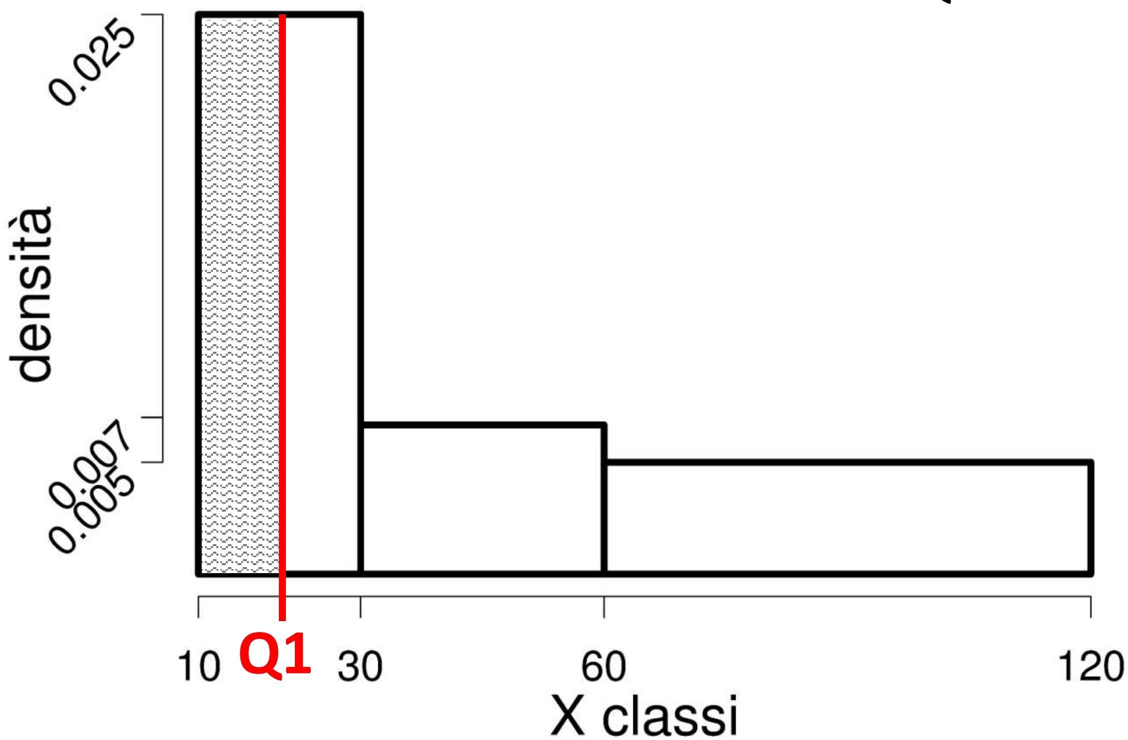
$$(Q1 - 10) = \frac{0.25}{0.025}$$

$$Q1 = \frac{0.25}{0.025} + 10 = 20$$

x_i	n_i	f_i	a_i	d_i
10 |-- 30	100	0.5	20	0.025
30 |-- 60	40	0.2	30	0.007
60 |-- 120	60	0.3	60	0.005
	200	**1**		

Primo quartile $\mathbf{F(X < Q1) = 0.25}$

La prima colonna dell'istogramma ha area **0.5** e Q1 stacca un'area di **0.25**. Quindi Q1 sta nella prima classe.



Ragionando sull'area:

$$(Q1 - 10) \times 0.025 = 0.25$$

$$(Q1 - 10) = \frac{0.25}{0.025}$$

$$Q1 = \frac{0.25}{0.025} + 10 = 20$$

Il 25% dei pendolari spende fino a 20 euro al mese

x_i	n_i	f_i	a_i	d_i
10 |-- 30	100	0.5	20	0.025
30 |-- 60	40	0.2	30	0.007
60 |-- 120	60	0.3	60	0.005
	200	**1**		

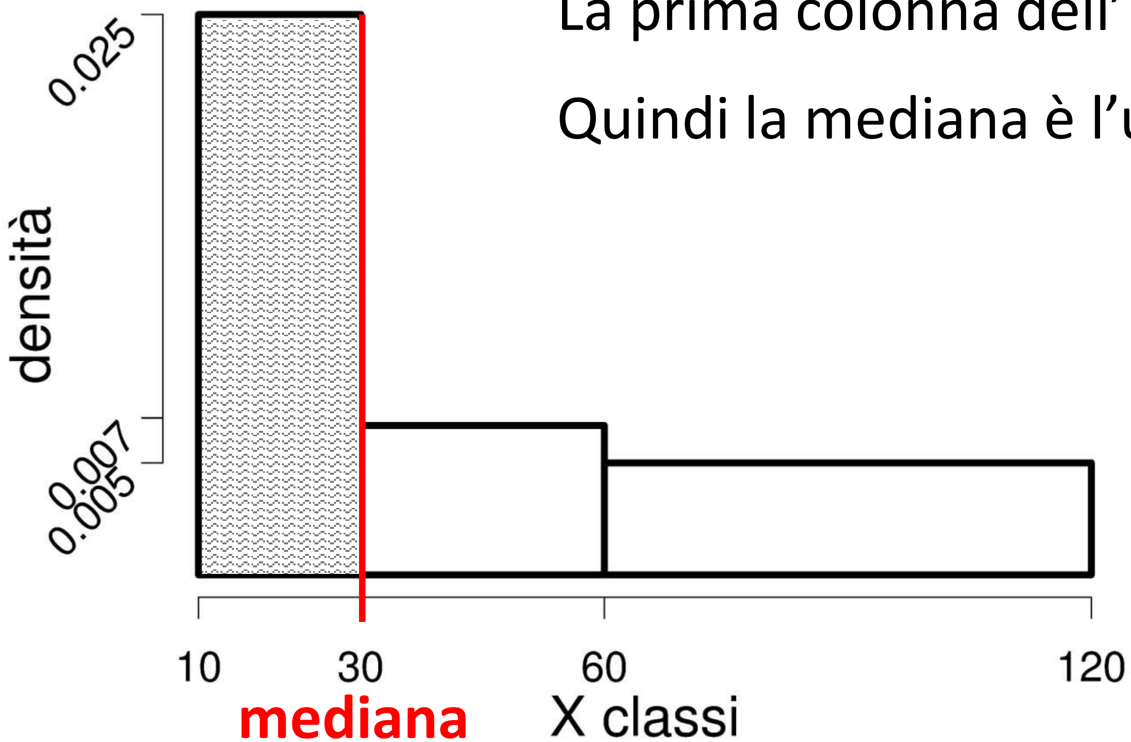
Secondo quartile o Mediana

$$F(X < \text{mediana}) = 0.50$$

La prima colonna dell'istogramma ha proprio area **0.5.**

Quindi la mediana è l'ultimo valore della prima classe:

$$\text{mediana} = 30$$

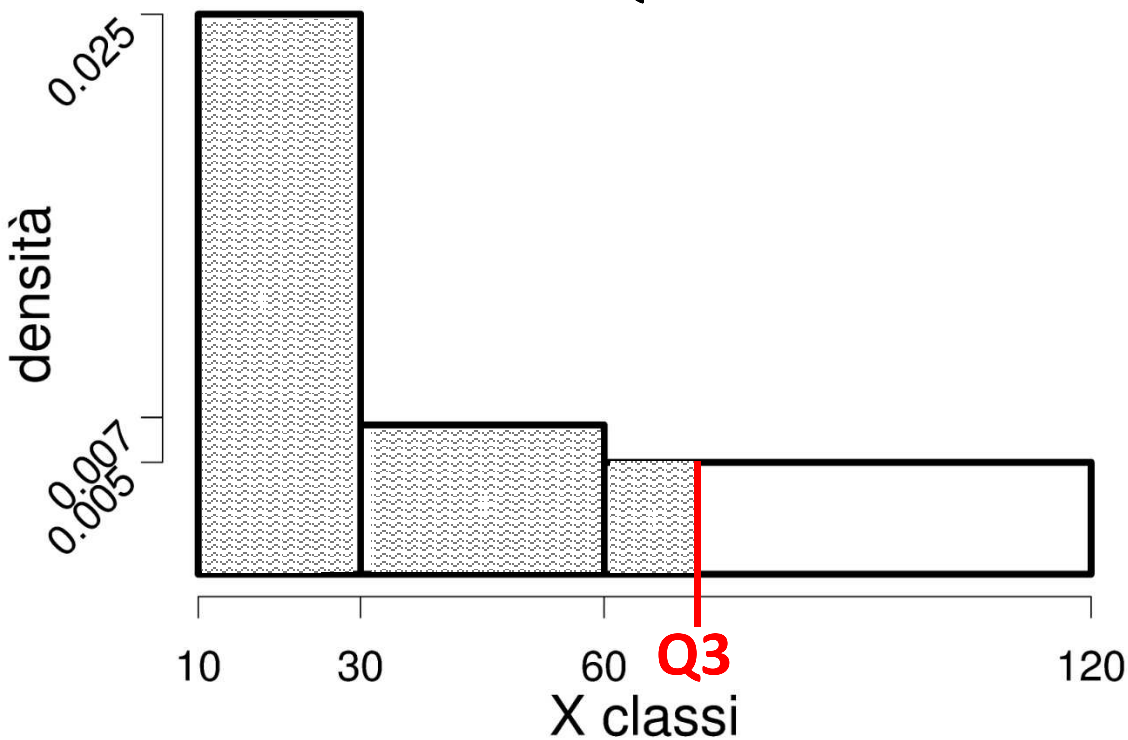


Il 50% dei pendolari spende fino a 30 euro al mese

x_i	n_i	f_i	a_i	d_i
10 |-- 30	100	0.5	20	0.025
30 |-- 60	40	0.2	30	0.007
60 |-- 120	60	0.3	60	0.005
	200	**1**		

Terzo quartile $\mathbf{F(X < Q3) = 0.75}$

Siccome le prime due colonne hanno area **0.7**, si deduce che Q3 sta nella terza classe. Ragionando sull'area:



$$0.7 + (Q3 - 60) \times 0.005 = 0.75$$

$$(Q3 - 60) = \frac{0.75 - 0.7}{0.005}$$

$$Q3 = \frac{0.75 - 0.7}{0.005} + 60 = 70$$

Il 75% dei pendolari spende fino a 70 euro al mese

Baffi

Abbiamo calcolato:

Q1 = 20 mediana = 30 Q3 = 70

Serve il range interquartile:

IQR = Q3 - Q1 = 70 – 20 = 50

1.5 x IQR = 1.5 x 50 = 75

Baffo superiore = Q3 + 1.5 x IQR = 70 + 75 = 145 $\longrightarrow$ = **120**

Baffo inferiore = Q1 - 1.5 x IQR = 20 - 75 = -55 $\longrightarrow$ = **10**

I valori dei baffi vengono posti uguali a max e min quando i valori calcolati **non** sono tra quelli delle classi.

In tal caso <u>non ci sono outliers</u>.

x_i classi	n_i	f_i Frequenze relative	a_i	d_i	m_i punto medio classi
10 |-- 30	100	0.5	20	0.025	(10+30)/2 = 20
30 |-- 60	40	0.2	30	0.007	(30+60)/2 = 45
60 |-- 120	60	0.3	60	0.005	(60+120)/2 = 90
	200	**1**			

Media

Servono i punti medi delle classi e le frequenze relative

$$\bar{x} = media = \sum_{\mathrm{i}} m_i \, \mathrm{f_i} =$$

$$= 20 \times 0.5 + 45 \times 0.2 + 90 \times 0.3 = 46$$

(In media i pendolari spendono 46 euro)

Baffo superiore = 120

Q3 = 70

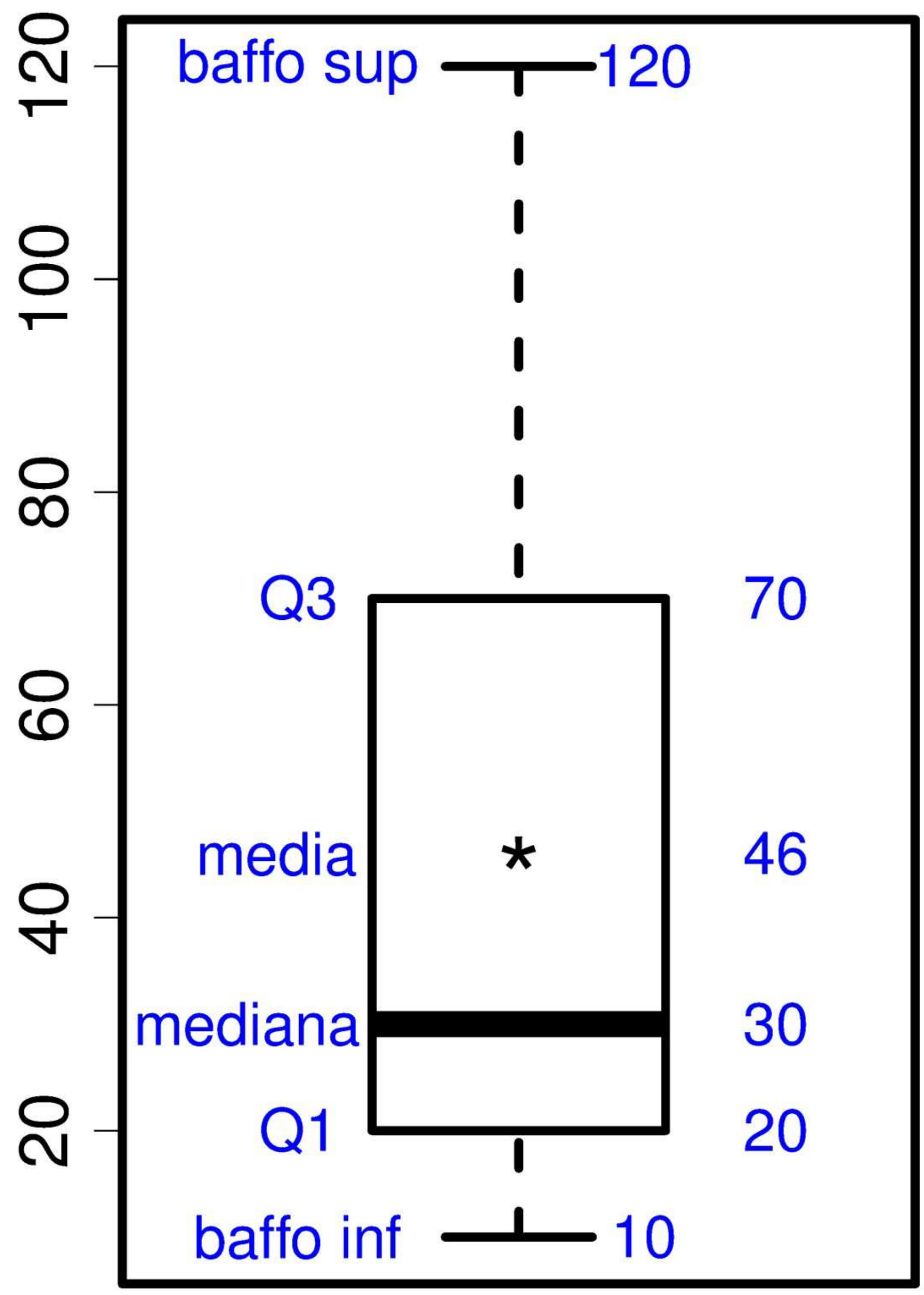
Mediana = 30

Q1 = 20

Baffo inferiore = 10

media = 46

Non ci sono outliers



x_i	n_i	f_i	a_i	d_i densità	m_i
10 |-- 30	100	0.5	20	0.025	20
30 |-- 60	40	0.2	30	0.007	45
60 |-- 120	60	0.3	60	0.005	90
	200	**1**			

(c) Moda

La classe modale è la classe a cui è associata la massima densità e la moda è il punto medio della classe modale.

La **massima densità** è 0.025, quindi

la **classe modale** è 10|--30 da cui

$$\mathbf{moda = 20}$$

x_i	n_i	f_i	a_i	d_i	m_i
10 |-- 30	100	0.5	20	0.025	20
30 |-- 60	40	0.2	30	0.007	45
60 |-- 120	60	0.3	60	0.005	90
	200	**1**			

(c) Coefficiente di Variazione

$$\mathrm{CV} = \frac{\sigma_x}{|\bar{x}|} = \frac{30.315}{46} = 0.659$$

$$\bar{x} = \text{media} = 46$$

$$\sigma_x{}^2 = \text{varianza} = \left(\sum_i m_i{}^2 \mathrm{f_i}\right) - \bar{x}^2 =$$

$$= (20^2 \times 0.5 + 45^2 \times 0.2 + 90^2 \times 0.3) - 46^2 =$$

$$= 200 + 405 + 2430 - 2116 = 919$$

$$\sigma_x = \text{scarto quadratico medio} = \sqrt{919} = 30.315$$

x_i	n_i	f_i	a_i	d_i	m_i
10 |-- 30	100	0.5	20	0.025	20
30 |-- 60	40	0.2	30	0.007	45
60 |-- 120	60	0.3	60	0.005	90
	200	**1**			

$$\bar{\mathrm{x}} = \mathrm{media} = 46$$

La varianza si può calcolare anche così:

$$\sigma_x{}^2 = \mathrm{varianza} = \sum_i (m_i - \bar{x})^2\, \mathrm{f_i} =$$

$$= (20-46)^2 \times 0.5 + (45-46)^2 \times 0.2 + (90-46)^2 \times 0.3 =$$

$$= 919$$

QUIZ 2

Sia X una variabile con media $\mu_x = 4$ e varianza ${\sigma_x}^2 = 9$.

Media e varianza di $Y = -2X + 7$ sono:

(a) $\mu_y = 8$, ${\sigma_y}^2 = 1$

(b) $\mu_y = -1$, ${\sigma_y}^2 = 36$

(c) $\mu_y = -1$, ${\sigma_y}^2 = 12$

(d) $\mu_y = -11$, ${\sigma_y}^2 = -6$

$$Y = -2X + 7$$

Media di Y

$$\boldsymbol{\mu_y} = -2\mu_x + 7 = -2 \times 4 + 7 = -8 + 7 = \mathbf{-1}$$

Varianza di Y

$$\boldsymbol{{\sigma_y}^2} = (-2)^2{\sigma_x}^2 = 4 \times 9 = \mathbf{36}$$

QUIZ 3

Sia X una variabile con media $\mu_x = -2$ e deviazione standard σ_x =5.

Media e varianza di $Y = -X + 2$ sono:

(a) $\mu_y = 4\,,\qquad \sigma_y{}^2 = 1$

(b) $\mu_y = 0\,,\qquad \sigma_y{}^2 = -1$

(c) $\boxed{\mu_y = 4\,,\qquad \sigma_y{}^2 = 25}$

(d) $\mu_y = 0,\qquad \sigma_y{}^2 = 25$

$$Y = -X + 2$$

Media di Y

$$\boldsymbol{\mu_y} = -\mu_x + 2 = (-1) \times \mu_x + 2 =$$

$$= (-1) \times (-2) + 2 = 2 + 2 = \mathbf{4}$$

Varianza di Y

$$\boldsymbol{\sigma_y{}^2} = (-1)^2 \sigma_x{}^2 = 1 \times 25 = \mathbf{25}$$

Esercizio 3

In un paese montano 50 dei 200 abitanti sono in pensione.

(a) Supponendo di estrarre 2 abitanti diversi di quel paese, calcolare la probabilità che siano entrambi in pensione

(b) Supponendo di estrarre 2 abitanti con reimmissione, calcolare la probabilità che solo uno sia in pensione

(c) Estraendo con reimmissione 9 abitanti, qual è la probabilità che esattamente 4 siano in pensione?

(d) Estraendo con reimmissione 9 abitanti, qual è la probabilità che almeno uno sia in pensione?

(e) Calcolare la probabilità che, estraendo con reinserimento 80 abitanti, almeno 20 siano in pensione.

Esercizio 3

In un paese montano 50 dei 200 abitanti sono in pensione.

(a) Supponendo di estrarre 2 abitanti diversi di quel paese, calcolare la probabilità che siano entrambi in pensione

Abitanti **diversi** ⟶ Estrazioni **senza** reinserimento (dipendenti)

$$\boldsymbol{P(X1 = 1 \cap X2 = 1)} = P(X1 = 1) \times P(X2 = 1|X1 = 1) = \ ?$$

Estrazione n.1 200 abitanti = 50 pensionati + 150 non pensionati

$$P(X1 = 1) = 50/200 = \mathbf{0.250}$$

Estrazione n.2 199 abitanti = 49 pensionati + 150 non pensionati

$$P(X2 = 1|X1 = 1) = 49/199 = \mathbf{0.246}$$

$$\boldsymbol{P(X1 = 1 \cap X2 = 1)} = P(X1 = 1) \times P(X2 = 1|X1 = 1) =$$

$$= 0.250 \times 0.249 = \mathbf{0.062}$$

Esercizio 3

In un paese montano 50 dei 200 abitanti sono in pensione.

(b) Supponendo di estrarre 2 abitanti con reimmissione, calcolare la probabilità che solo uno sia in pensione

Soluzione 1

Estrazioni con reimmissione (indipendenti)

$$P(X_1 = 1 \cap X_2 = 0) \quad + \quad P(X_1 = 0 \cap X_2 = 1) =$$

$$= P(X_1 = 1) \times P(X_2 = 0) + P(X_1 = 0) \times P(X_2 = 1) =$$

$$= \quad 0.25 \quad \times \quad 0.75 \quad + \quad 0.75 \quad \times \quad 0.25 \quad =$$

$$= 0.375$$

Esercizio 3

In un paese montano 50 dei 200 abitanti sono in pensione.

(b) Supponendo di estrarre 2 abitanti con reimmissione, calcolare la probabilità che solo uno sia in pensione

Soluzione 2

Estrazioni con reimmissione (indipendenti)

$$Y = X_1 + X_2 = \textit{numero di pensionati su 2 estrazioni}$$

Y ha distribuzione Binomiale(n=2, p=0.25)

$$P(Y = 1) = \binom{2}{1} 0.25^1 (1 - 0.25)^{2-1} = 2 \times 0.25 \times 0.75 = 0.375$$

Esercizio 3

In un paese montano 50 dei 200 abitanti sono in pensione.

(c) Estraendo con reimmissione 9 abitanti, qual è la probabilità che esattamente 4 siano in pensione?

Y = numero di pensionati su 9 estrazioni con reimmissione

Y ha distribuzione Binomiale(n = 9, p = 0.25)

$$P(Y = i) = \binom{n}{i} p^i (1-p)^{n-i} \qquad \mathrm{i} = 0,1,2,\dots,\mathrm{n}$$

Quindi si tratta di calcolare:

$$P(Y = 4) = \binom{9}{4} 0.25^4 (1-0.25)^{9-4}$$

$$\boldsymbol{P}(\boldsymbol{Y} = \mathbf{4}) = \binom{9}{4} 0.25^4 (1 - 0.25)^{9-4} =$$

$$= \frac{9!}{4!\,(9-4)!} \cdot 0.25^4 \cdot 0.75^5 =$$

$$= \frac{1 \times 2 \times 3 \times 4 \times 5 \times 6 \times 7 \times 8 \times 9}{(1 \times 2 \times 3 \times 4) \times (1 \times 2 \times 3 \times 4 \times 5)} \cdot 0.25^4 \cdot 0.75^5 =$$

$$= \frac{\overset{\mathbf{1}}{\cancel{1 \times 2 \times 3 \times 4 \times 5}} \times \overset{\mathbf{7x2x9}}{\cancel{6 \times 7 \times 8 \times 9}}}{\underset{\mathbf{1}}{\cancel{(1 \times 2 \times 3 \times 4)}} \times \underset{\mathbf{1}}{\cancel{(1 \times 2 \times 3 \times 4 \times 5)}}} \times 0.004 \times 0.237 =$$

$$= \mathbf{0.119}$$

La probabilità di selezionare esattamente 4 pensionati su 9 estrazioni con reimmissione è 0.119 (ossia 11.9%)

Esercizio 3

In un paese montano 50 dei 200 abitanti sono in pensione.

(d) Estraendo con reimmissione 9 abitanti, qual è la probabilità che almeno uno sia in pensione?

Y = numero di pensionati su 9 estrazioni con reimmissione

Y ha distribuzione Binomiale(n = 9, p = 0.25)

$$P(Y = i) = \binom{n}{i} p^i (1-p)^{n-i} \qquad \mathrm{i} = 0,1,2,\dots,\mathrm{n}$$

Quindi si tratta di calcolare:

$$P(Y \geq 1) = 1 - P(Y = 0) = 1 - \binom{9}{0} 0.25^0 (1 - 0.25)^{9-0}$$

$$\boldsymbol{P}(\boldsymbol{Y} = \mathbf{0}) = \binom{9}{0} 0.25^0 (1 - 0.25)^{9-0} =$$

$$= \frac{9!}{0!\,(9-0)!} \cdot 0.25^0 \cdot 0.75^9 =$$

$$= \frac{1 \times 2 \times 3 \times 4 \times 5 \times 6 \times 7 \times 8 \times 9}{(1) \times (1 \times 2 \times 3 \times 4 \times 5 \times 6 \times 7 \times 8 \times 9)} \times 1 \times 0.075 =$$

$$= 1 \times 1 \times 0.075 =$$

$$= \mathbf{0.075}$$

La probabilità di selezionare alemno un pensionato su 9 estrazioni con reimmissione è 0.075 (ossia 7.5%)

ATTENZIONE $\mathbf{0! = 1}$ $\mathbf{1! = 1}$ $\boldsymbol{x^0 = 1}$ $\mathbf{0^n = 0}$

Esercizio 3

In un paese montano 50 dei 200 abitanti sono in pensione.

(e) Calcolare la probabilità che, estraendo con reinserimento 80 abitanti, almeno 20 siano in pensione.

Y = numero di pensionati su 80 abitanti estratti con reimmissione

Y ha distribuzione Binomiale(n = 80, p = 0.25)

Quindi si tratterebbe di calcolare:

$$P(Y \geq 20) = P(Y = 20) + P(Y = 21) + \ \dots + P(X = 80) =$$

$$= \sum_{i=20}^{80} \binom{80}{i} 0.25^i (1 - 0.25)^{80-i}$$

Si deve approssimare !

Esercizio 3

In un paese montano 50 dei 200 abitanti sono in pensione.

(e) Calcolare la probabilità che, estraendo con reinserimento 80 abitanti, almeno 20 siano in pensione.

Y = numero di pensionati su 80 abitanti estratti con reimmissione

Y ha distribuzione Binomiale(n = 80, p = 0.25)

$$P(Y \geq 20) = ?$$

Si può ricorrere all'approssimazione della distribuzione Binomiale con la distribuzione Normale?

$$\mathrm{X} \sim \mathrm{Binomiale}(\mathrm{n}, \mathrm{p}) \quad \rightarrow \quad \mathrm{Z} = \frac{\mathrm{X} - \mathrm{np}}{\sqrt{\mathrm{np}(1-\mathrm{p})}} \sim \mathrm{Normale}(0,1)$$

qualora $\mathrm{n} \geq 30, \quad \mathrm{np} \geq 5, \quad \mathrm{n}(1-\mathrm{p}) \geq 5$

Si può perché per n=80 e p=0.25 risulta

$$\boldsymbol{n = 80} \quad \geq 30$$

$$\boldsymbol{np} = 80 \times 0.25 = \boldsymbol{20} \quad \geq 5,$$

$$\boldsymbol{n(1-p)} = 80 \times (1 - 0.25) = \boldsymbol{60} \quad \geq 5$$

Quindi si applica questa approssimazione:

$$\boldsymbol{X \sim Binomiale(n, p)} \quad \rightarrow \quad \boldsymbol{Z} = \frac{\boldsymbol{X - np}}{\sqrt{\boldsymbol{np(1-p)}}} \sim \boldsymbol{Normale(0, 1)}$$

$$\boldsymbol{np} = \boldsymbol{20}$$

$$\sqrt{\boldsymbol{np(1-p)}} = \sqrt{80 \times 0.25 \times 0.75} = \sqrt{15} = \boldsymbol{3.87}$$

$$\boldsymbol{P(X \geq 20)} = 1 - P(X < 20) = 1 - P\left(Z < \frac{\boldsymbol{20 - 20}}{\boldsymbol{3.87}}\right) =$$

$$= 1 - P(Z < 0) = 1 - 0.5 = \boldsymbol{0.5}$$

QUIZ 4

X ha distribuzione Normale con media $\mu_x = 1$ e deviazione standard $\sigma_x = 5$. La sua mediana vale:

(a) 1

(b) 25

(c) 5

(d) 0

Simmetria rispetto alla media !

La probabilità dell'unione di eventi **disgiunti** è data da:

(a) $P(A \cup B) = P(A) + P(B)$

(b) $P(A \cup B) = P(A)P(B) - P(A \cap B)$

(c) $P(A \cup B) = P(A) + P(B) - P(A \cap B)$

(d) $P(A \cup B) = P(A)P(B)$

(c) vale in generale per tutti gli eventi A e B

(a) vale per eventi A e B disgiunti : $P(A \cap B) = \mathrm{P}(\emptyset) = 0$