

Carattere quantitativo discreto

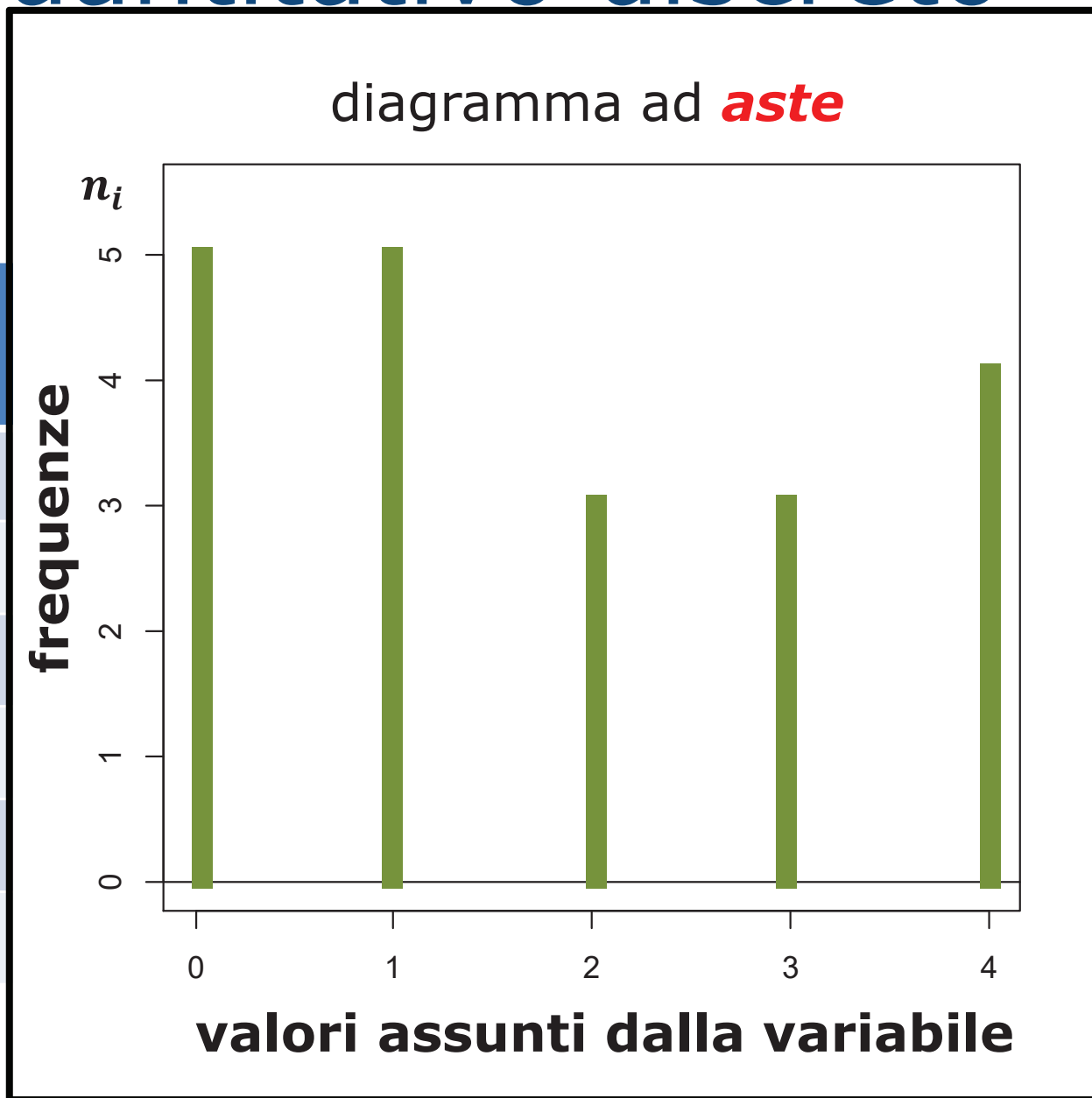
CARATTERE/
VARIABILE

FREQUENZE

num. figli	n_i				
0	5				
1	5				
2	3				
3	3				
4	4				
tot.	20				

Carattere quantitativo discreto

num. figli	n_i
0	5
1	5
2	3
3	3
4	4
tot.	20



Carattere quantitativo discreto

num. figli	n_i	f_i	$100f_i$	N_i	F_i	P_i (%)
0	5	0.25	25.0	5	0.25	25.0
1	5	0.25	25.0	10	0.50	50.0
2	3	0.15	15.0	13	0.65	65.0
3	3	0.15	15.0	16	0.80	80.0
4	4	0.20	20.0	20	1.00	100.0
tot.	20	1	100.0			

Dati quantitativi

Unità	Peso
1	118
2	151
3	143
4	172
5	147
6	146
7	138
8	175
9	134
10	172
11	118
12	151
13	155
14	155
15	146
16	135
17	127
18	178
19	136
20	180
21	151
22	186
23	122
24	132
25	114
26	171
27	140
28	187
29	106
30	159
31	127
32	191
33	192
34	181
35	143
36	153
37	144
38	139
39	148
40	179

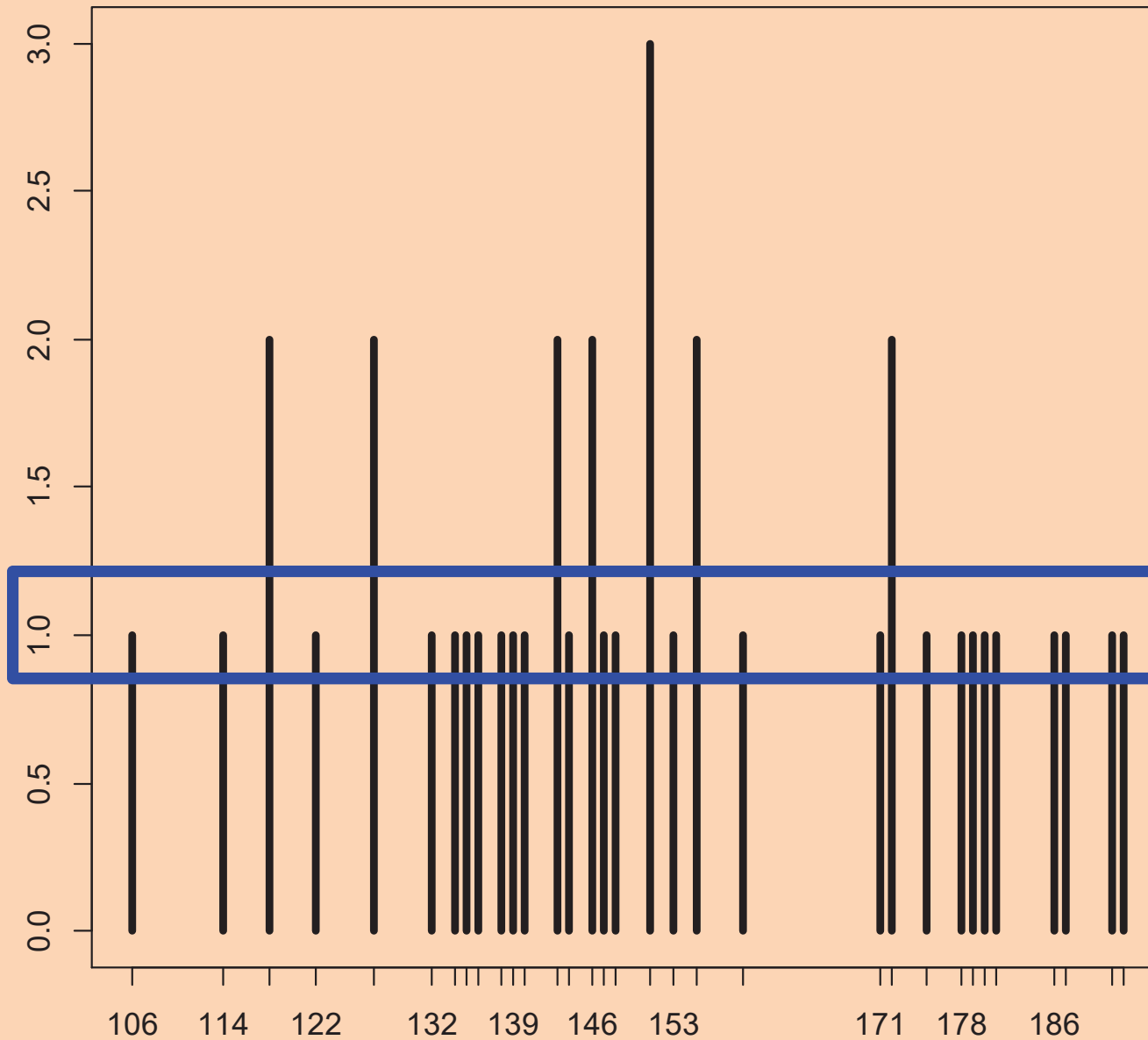
Dati quantitativi

Unità	Peso
1	118
2	151
3	143
4	172
5	147
6	146
7	138
8	175
9	134
10	172
11	118
12	151
13	155
14	155
15	146
16	135
17	127
18	178
19	136
20	180
21	151
22	186
23	122
24	132
25	114
26	171
27	140
28	187
29	106
30	159
31	127
32	191
33	192
34	181
35	143
36	153
37	144
38	139
39	148
40	179

ha senso
determinare
le frequenze?

Dati quantitativi

Unità	Peso
1	118
2	151
3	143
4	172
5	147
6	146
7	138
8	175
9	134
10	172
11	118
12	151
13	155
14	155
15	146
16	135
17	127
18	178
19	136
20	180
21	151
22	186
23	122
24	132
25	114
26	171
27	140
28	187
29	106
30	159
31	127
32	191
33	192
34	181
35	143
36	153
37	144
38	139
39	148
40	179

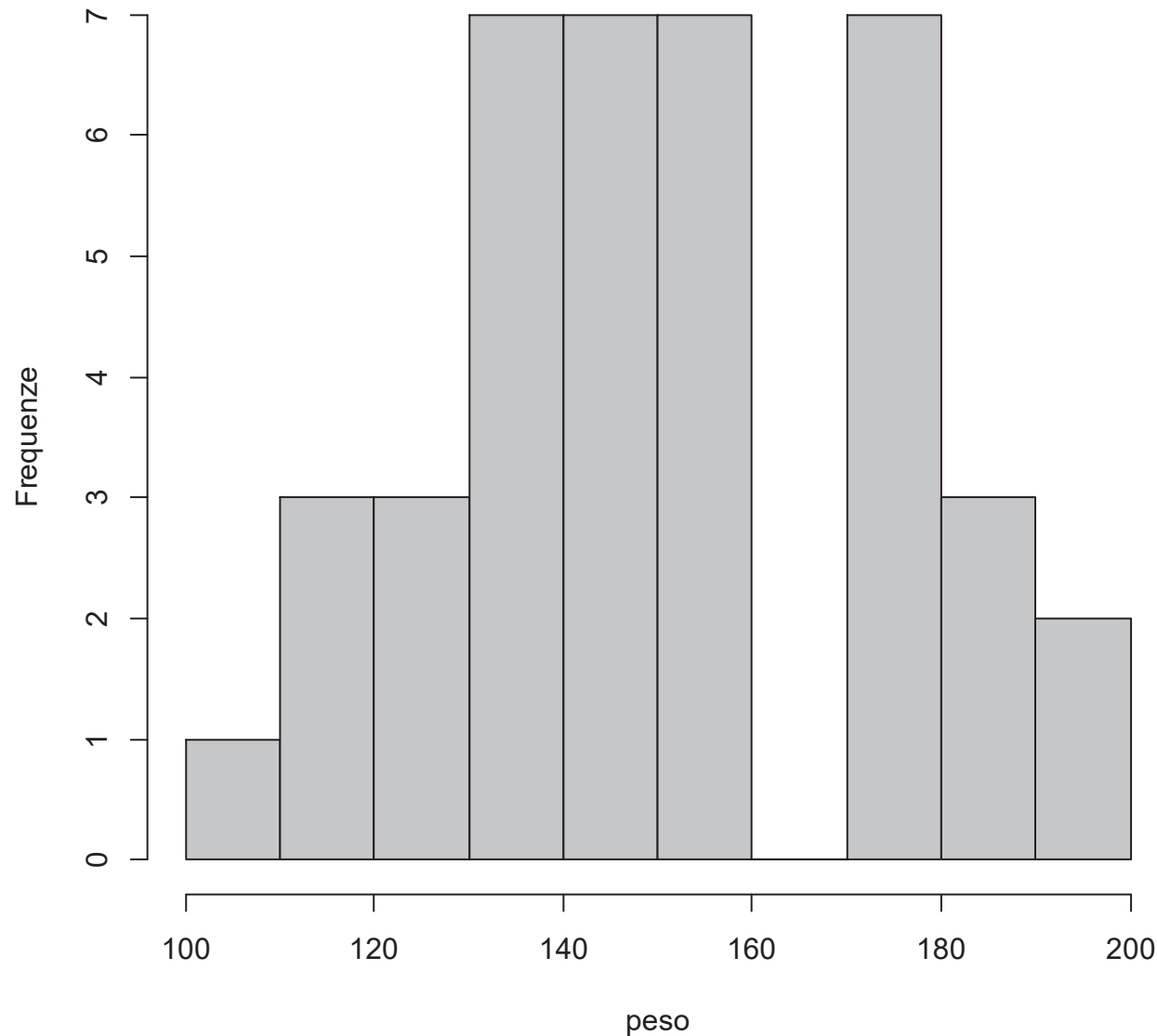


Dati quantitativi per classi

Classi	n_i (freq. ass.)	f_i (freq. rel.)	N_i (freq. cum.)	a_i	$l_i = \frac{n_i}{a_i}$	$l'_i = \frac{f_i}{a_i}$
100- 110	1	0.025	1	10	0.1	0.0025
110- 120	3	0.075	4	10	0.3	0.0075
120- 130	3	0.075	7	10	0.3	0.0075
130- 140	7	0.175	14	10	0.7	0.0175
140- 150	7	0.175	21	10	0.7	0.0175
150- 160	7	0.175	28	10	0.7	0.0175
160- 170	0	0.000	28	10	0.0	0.0
170- 180	7	0.175	35	10	0.7	0.0175
180- 190	3	0.075	38	10	0.3	0.0075
190- 200	2	0.050	40	10	0.2	0.0050
Tot.	40	1				

Istogramma

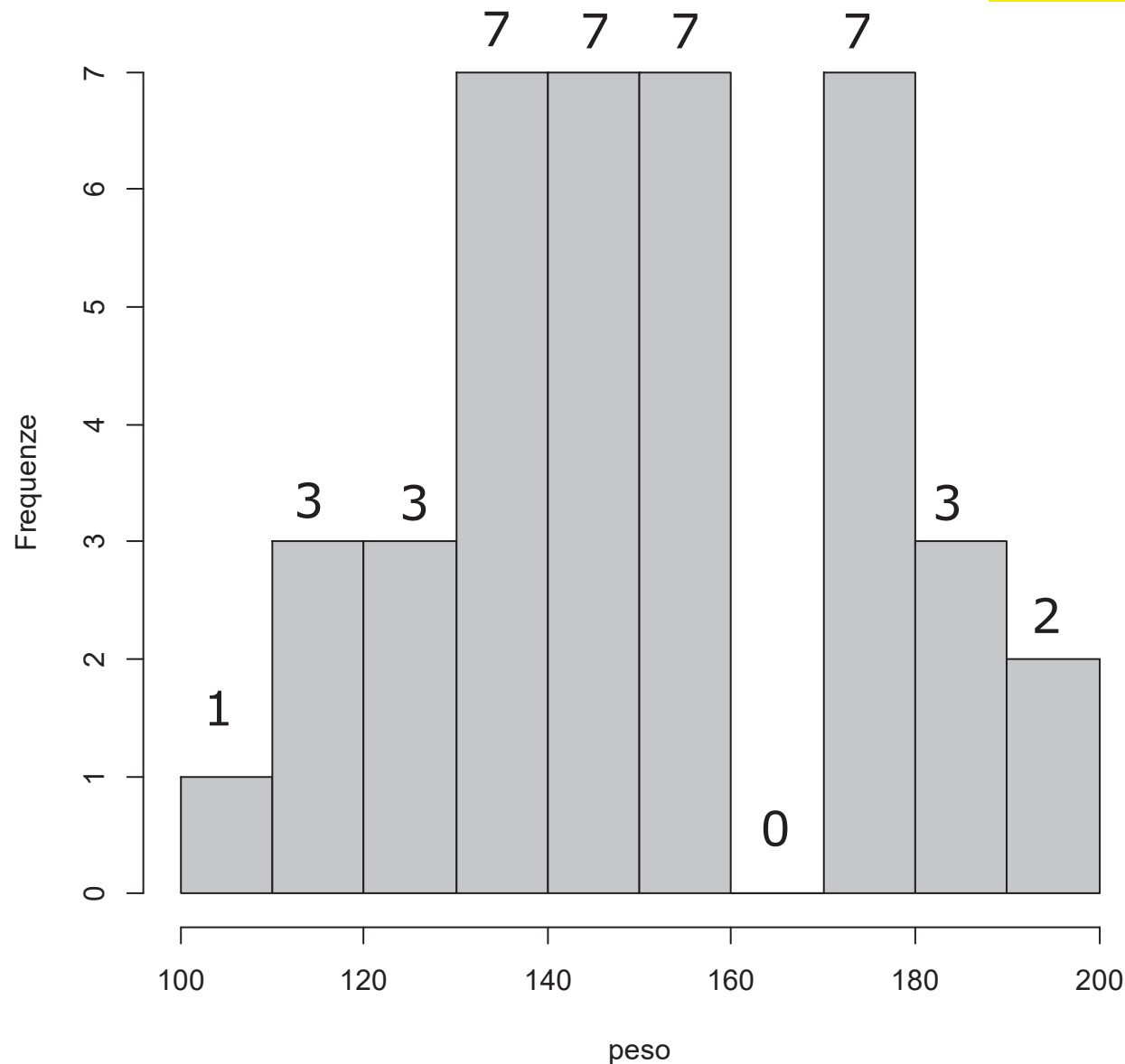
Rappresentazione
delle frequenze
tramite l'**area**



Suddividiamo il
range di valori in
intervalli
disgiunti e
contiamo quanti
dei dati cadono
in ciascuna
classe.

Istogramma

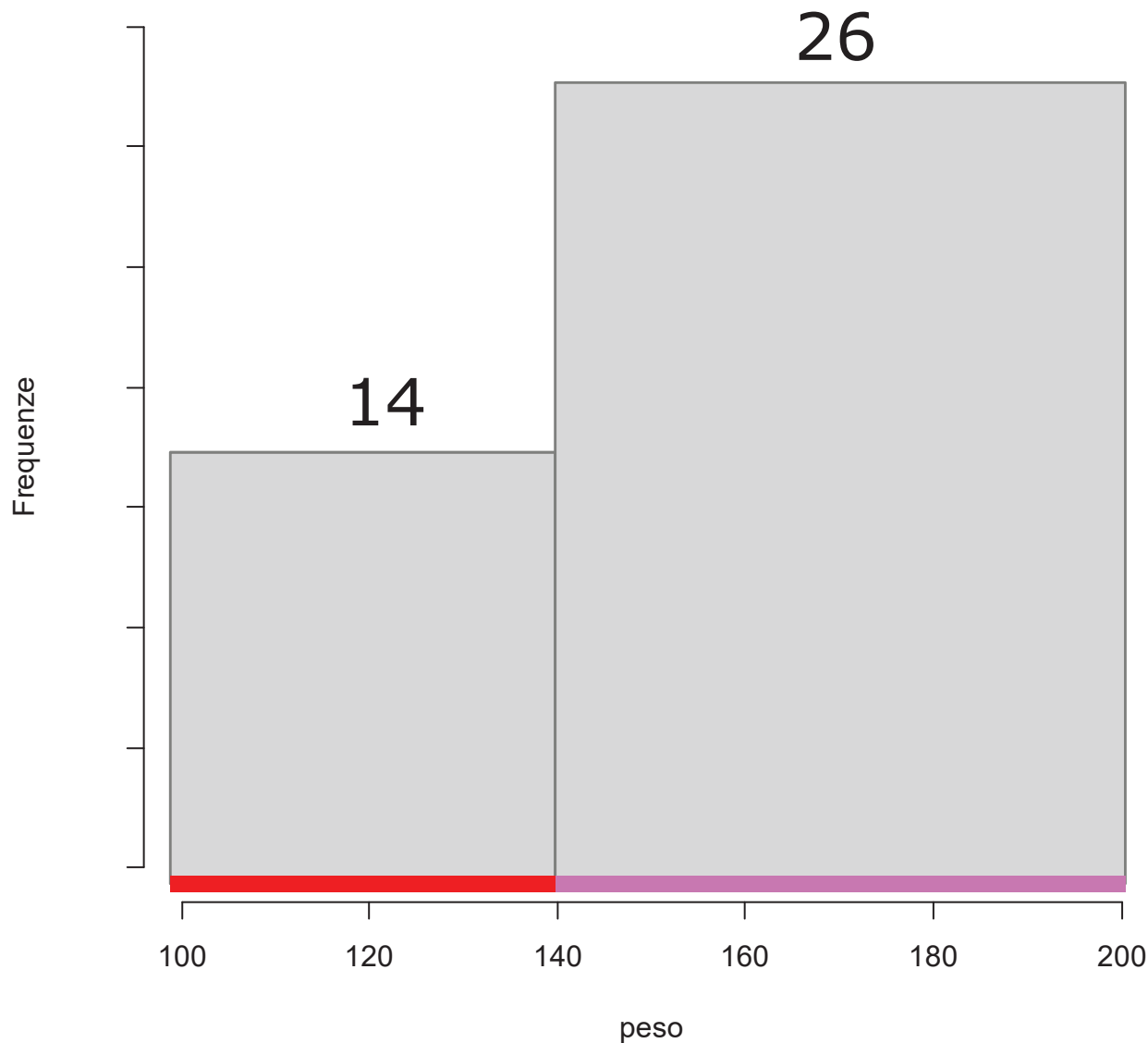
Rappresentazione
delle frequenze
tramite l'**area**



Suddividiamo il
range di valori in
intervalli
disgiunti e
contiamo quanti
dei dati cadono
in ciascuna
classe.

Istogramma

Rappresentazione
delle frequenze
tramite l'**area**



Suddividiamo il
range di valori in
intervalli
disgiunti e
contiamo quanti
dei dati cadono
in ciascuna
classe.

Istogramma

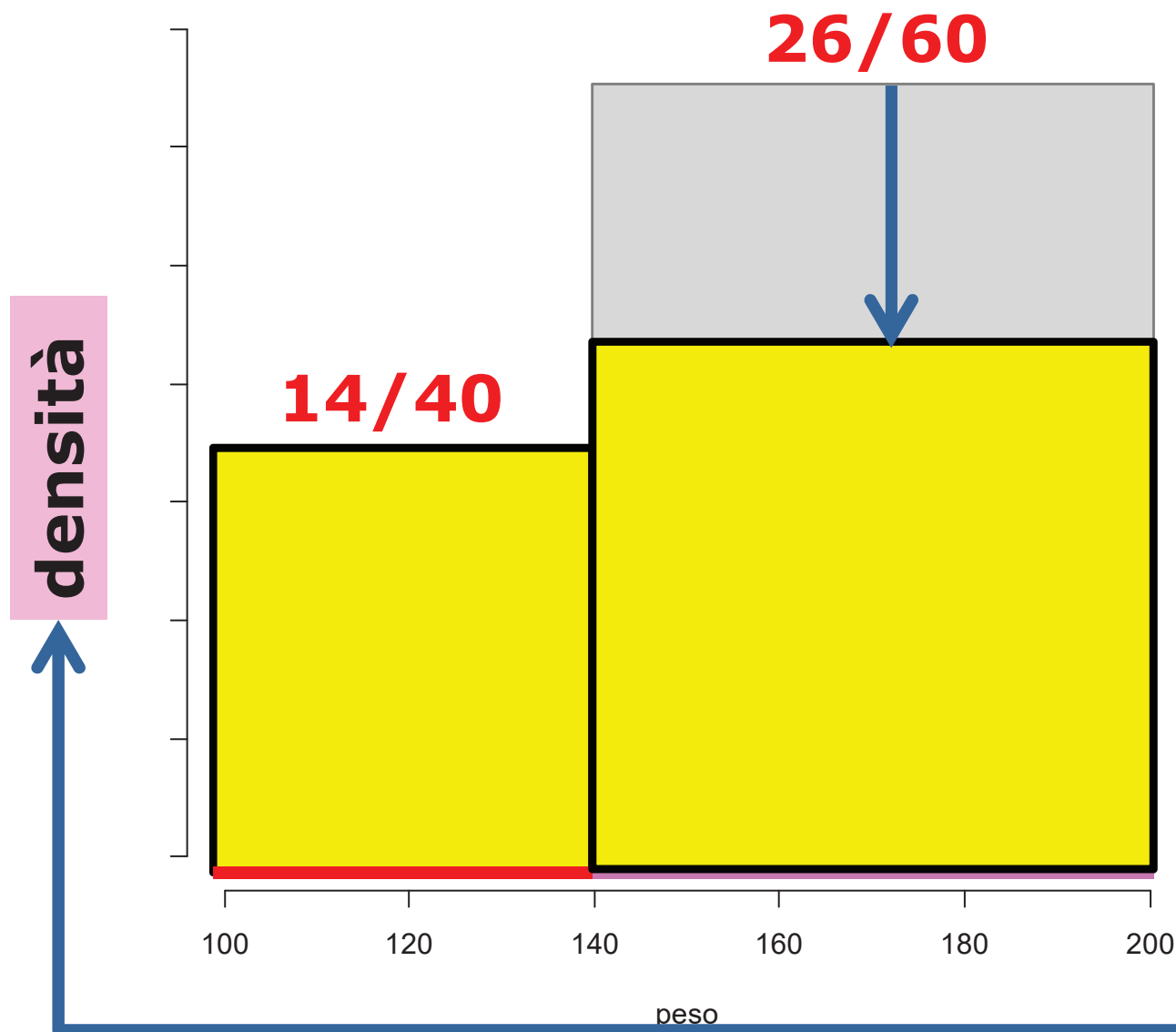
Rappresentazione
delle frequenze
tramite l'**area**



Suddividiamo il
range di valori in
intervalli
disgiunti e
contiamo quanti
dei dati cadono
in ciascuna
classe.

Istogramma

Rappresentazione
delle frequenze
tramite l'**area**



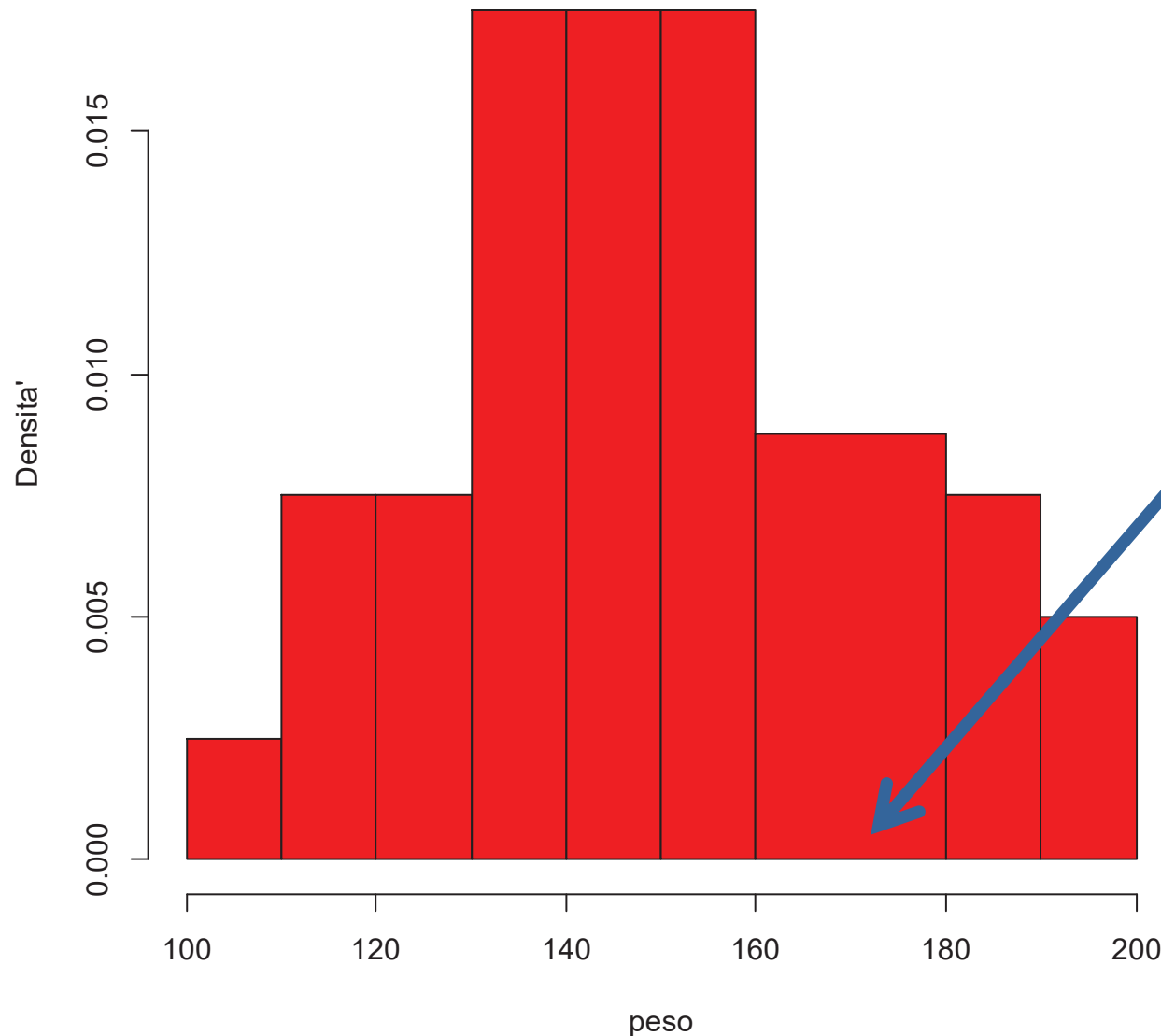
Suddividiamo il range di valori in **intervalli disgiunti** e contiamo quanti dei dati cadono in ciascuna classe, ***tenendo conto dell'ampiezza della classe.***

Dati quantitativi per classi

Classi	n_i (freq. ass.)	f_i (freq. rel.)	N_i (freq. cum.)	a_i	$l_i = \frac{n_i}{a_i}$	$l'_i = \frac{f_i}{a_i}$
100- 110	1	0.025	1	10	0.1	0.0025
110- 120	3	0.075	4	10	0.3	0.0075
120- 130	3	0.075	7	10	0.3	0.0075
130- 140	7	0.175	14	10	0.7	0.0175
140- 150	7	0.175	21	10	0.7	0.0175
150- 160	7	0.175	28	10	0.7	0.0175
160- 170	0	0.000	28	10	0.0	0.0
170- 180	7	0.175	35	10	0.7	0.0175
180- 190	3	0.075	38	10	0.3	0.0075
190- 200	2	0.050	40	10	0.2	0.0050
Tot.	40	1				

Istogramma

Rappresentazione
delle frequenze
tramite l'**area**



Suddividiamo il range di valori in **intervalli disgiunti** e contiamo quanti dei dati cadono in ciascuna classe, **tenendo conto dell'ampiezza della classe.**

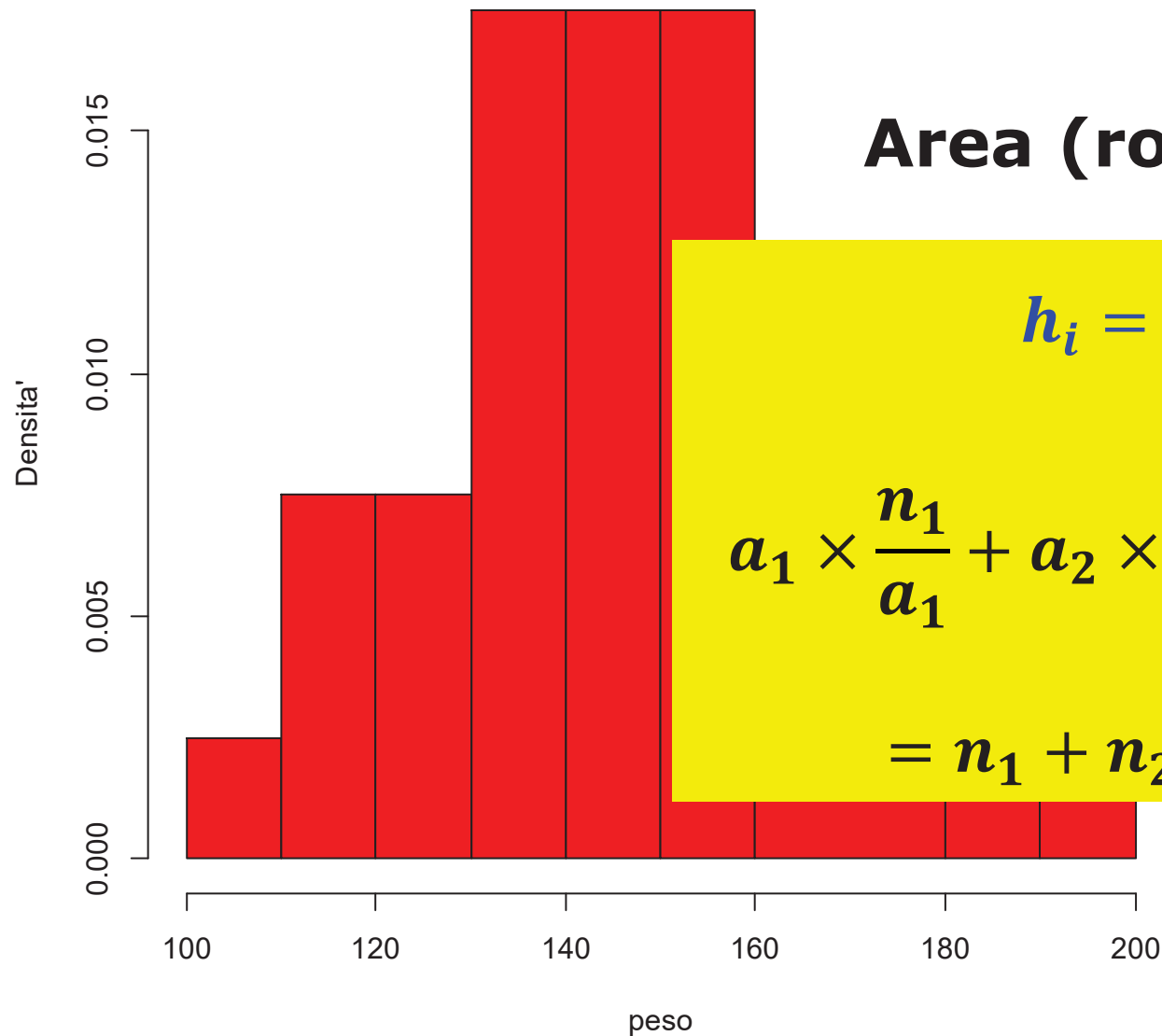
$$h_i = \frac{n_i}{a_i} = l_i$$

o

$$h_i = \frac{f_i}{a_i} = l'_i$$

Istogramma

Rappresentazione
delle frequenze
tramite l'**area**



Area (rossa) totale:

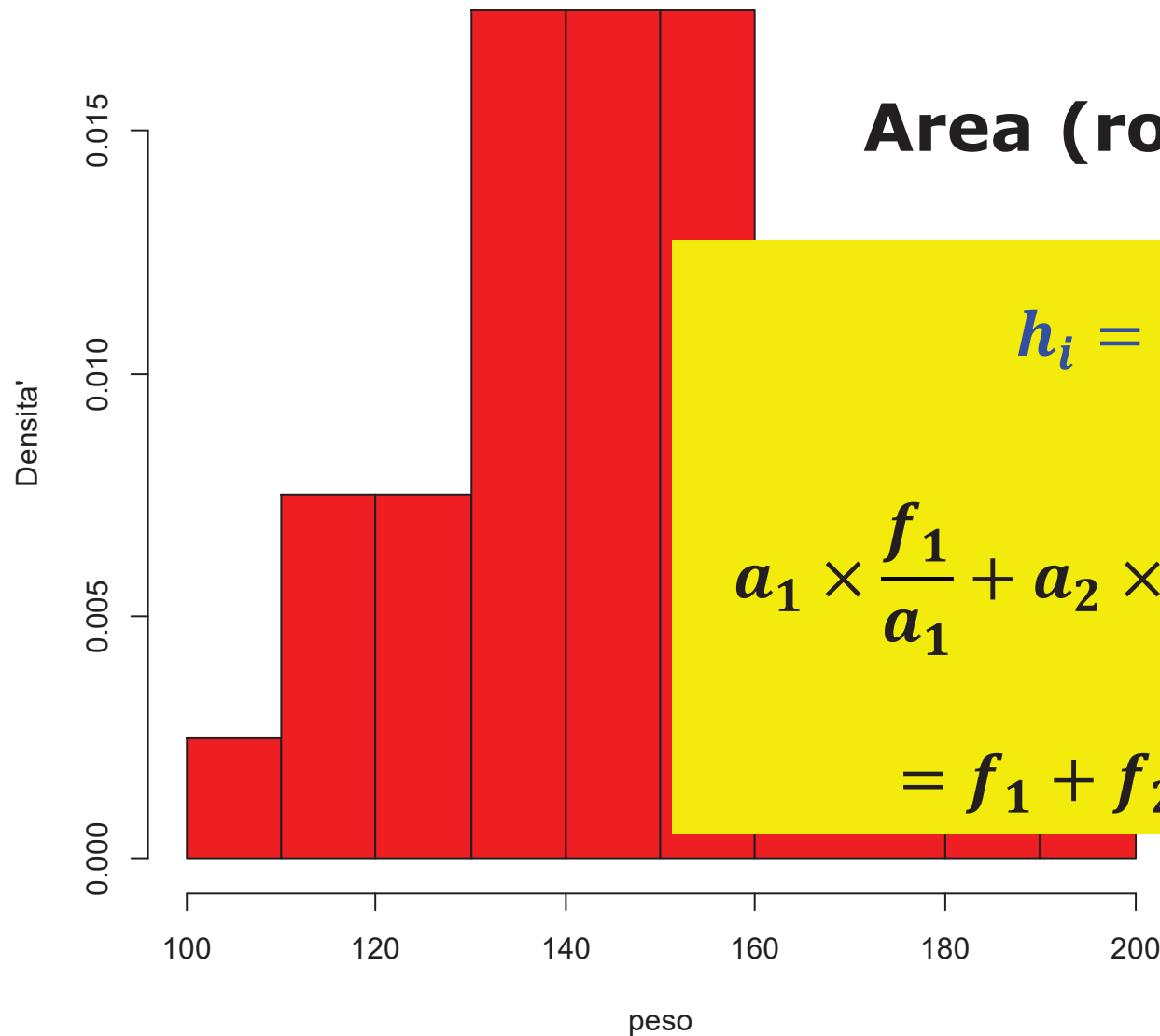
$$h_i = \frac{n_i}{a_i} = l_i \Rightarrow$$

$$a_1 \times \frac{n_1}{a_1} + a_2 \times \frac{n_2}{a_2} + \dots + a_9 \times \frac{n_9}{a_9} =$$

$$= n_1 + n_2 + \dots + n_9 = n$$

Istogramma

Rappresentazione
delle frequenze
tramite l'**area**



Area (rossa) totale:

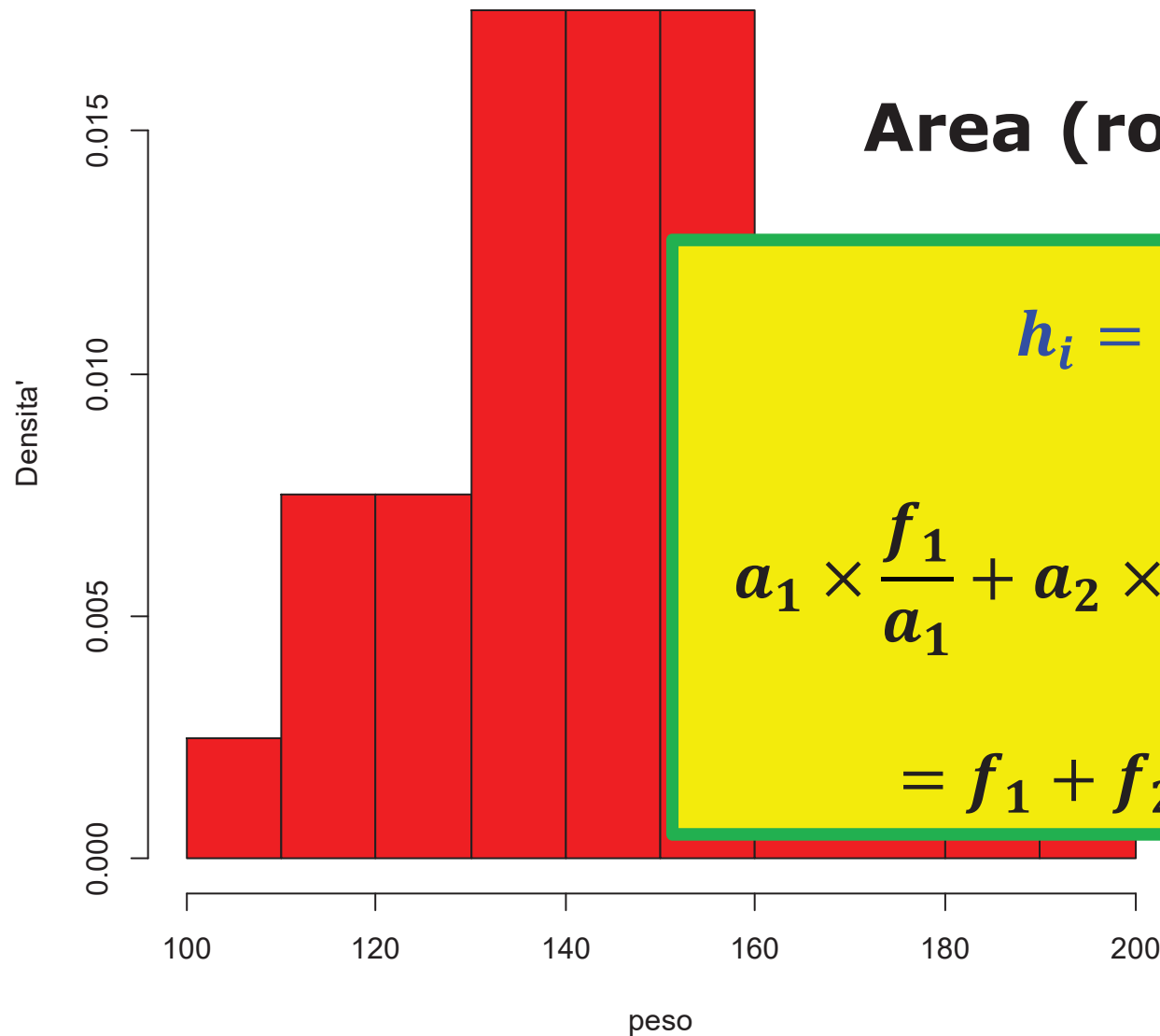
$$h_i = \frac{f_i}{a_i} = l'_i \Rightarrow$$

$$a_1 \times \frac{f_1}{a_1} + a_2 \times \frac{f_2}{a_2} + \dots + a_9 \times \frac{f_9}{a_9} =$$

$$= f_1 + f_2 + \dots + f_9 = 1$$

Istogramma

Rappresentazione
delle frequenze
tramite l'**area**



Area (rossa) totale:

$$h_i = \frac{f_i}{a_i} = l'_i \Rightarrow$$

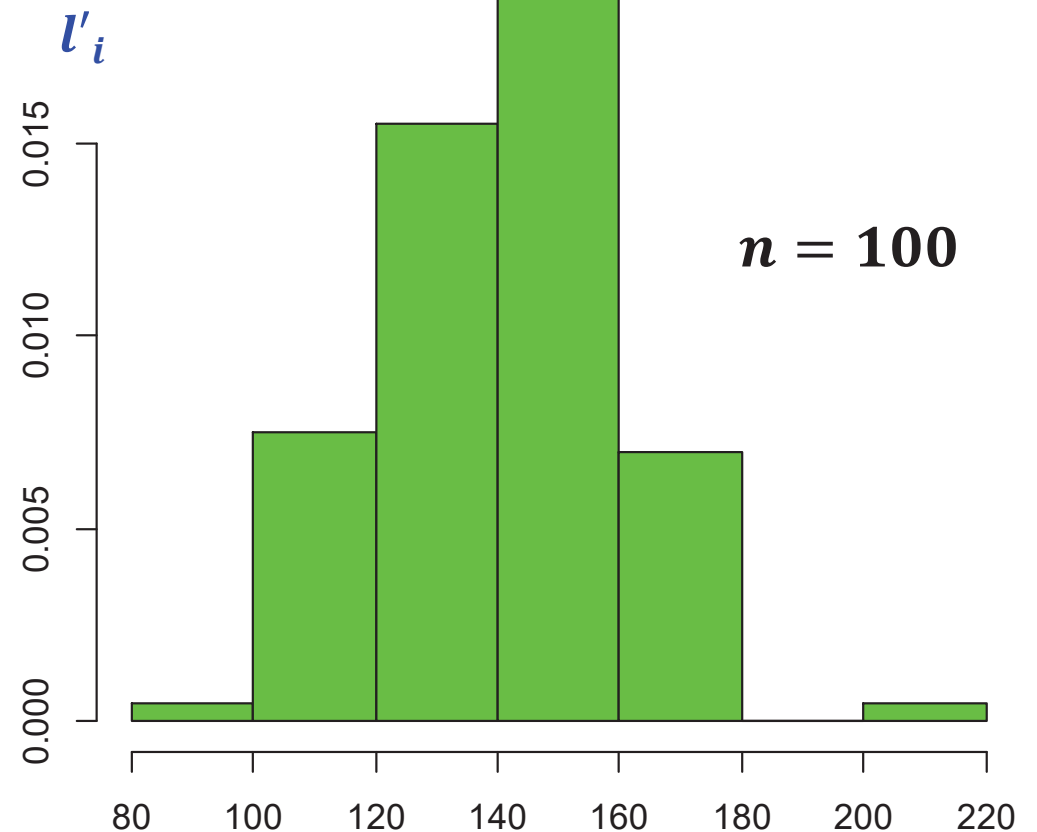
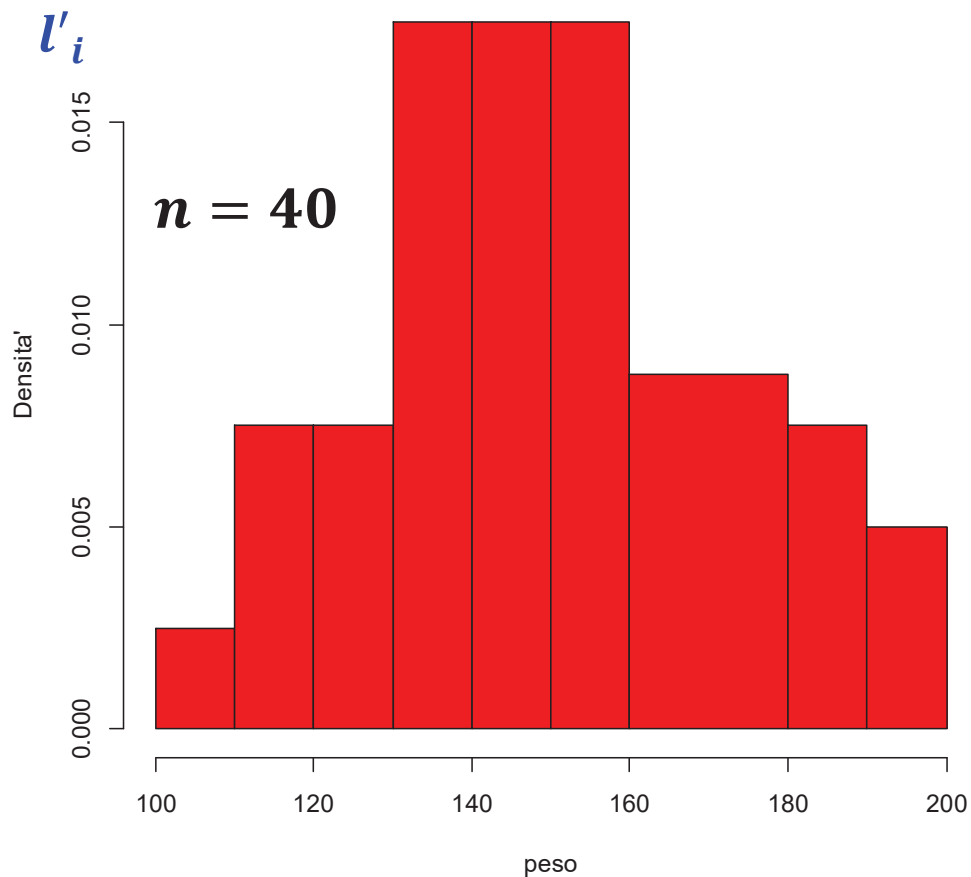
$$a_1 \times \frac{f_1}{a_1} + a_2 \times \frac{f_2}{a_2} + \dots + a_9 \times \frac{f_9}{a_9} =$$

$$= f_1 + f_2 + \dots + f_9 = 1$$

Istogramma

Rappresentazione
delle frequenze
tramite l'**area**

In ciascuno l'area totale è 1: posso confrontarli!



Postilla

Qualche commento sulla tabella della cotinina... v. libro.

- decidere di che tipologia è la variabile: discreta (???), continua (!!!)
- aggiustare il discorso della larghezza delle classi per v. continua (.5)

Niente *poligono delle frequenze*

Niente *diagramma di Pareto*

Niente *dotplot*

Niente *diagramma ramo-foglia*

Sullo *scatter plot* torneremo

Statistica

Statistica descrittiva

INDICI

Indici di centralità: **moda**

	Età (y)	Peso (kg)	Altezza (m)	Sesso	Causa di morte
1	19	50.2	1.65	F	Nat.
2	22	75.6	1.78	M	Inc.
3	21	80.1	1.91	M	Inc.
4	23	56.7	1.72	M	Nat.
5	22	75.0	1.81	M	M.C.
6	20	58.3	1.68	F	Tum.

Indici di centralità: moda

Il genere M compare più spesso di F

Delle 4 cause considerate, due hanno frequenza maggiore (Nat. e Inc.)

Sesso	Causa di morte
F	Nat.
M	Inc.
M	Inc.
M	Nat.
M	M.C.
F	Tum.

Indici di centralità: moda

Nel Genere la **moda** è M

Nella Causa di morte ci sono **due mode**: Nat. e Inc. (**bi-modalità**)

Sesso	Causa di morte
F	Nat.
M	Inc.
M	Inc.
M	Nat.
M	M.C.
F	Tum.

Indici di centralità: moda

Voto all'esame: da 18 a 30L

Rendimento:

18-22: suff.

23-26: buono

27-30L: ottimo

**Il voto BUONO è
la moda del
rendimento**

Rendimento	n_i	$f_i = n_i/n$	% = $100f_i$
S	35	0.233	23.3 %
B	62	0.413	41.3 %
O	53	0.354	35.4 %
tot.	150	1	100 %

Indici di centralità: moda

CARATTERE/
VARIABILE

FREQUENZE

num. figli	n_i		
0	5		
1	5		
2	3		
3	3		
4	4		
tot.	20		

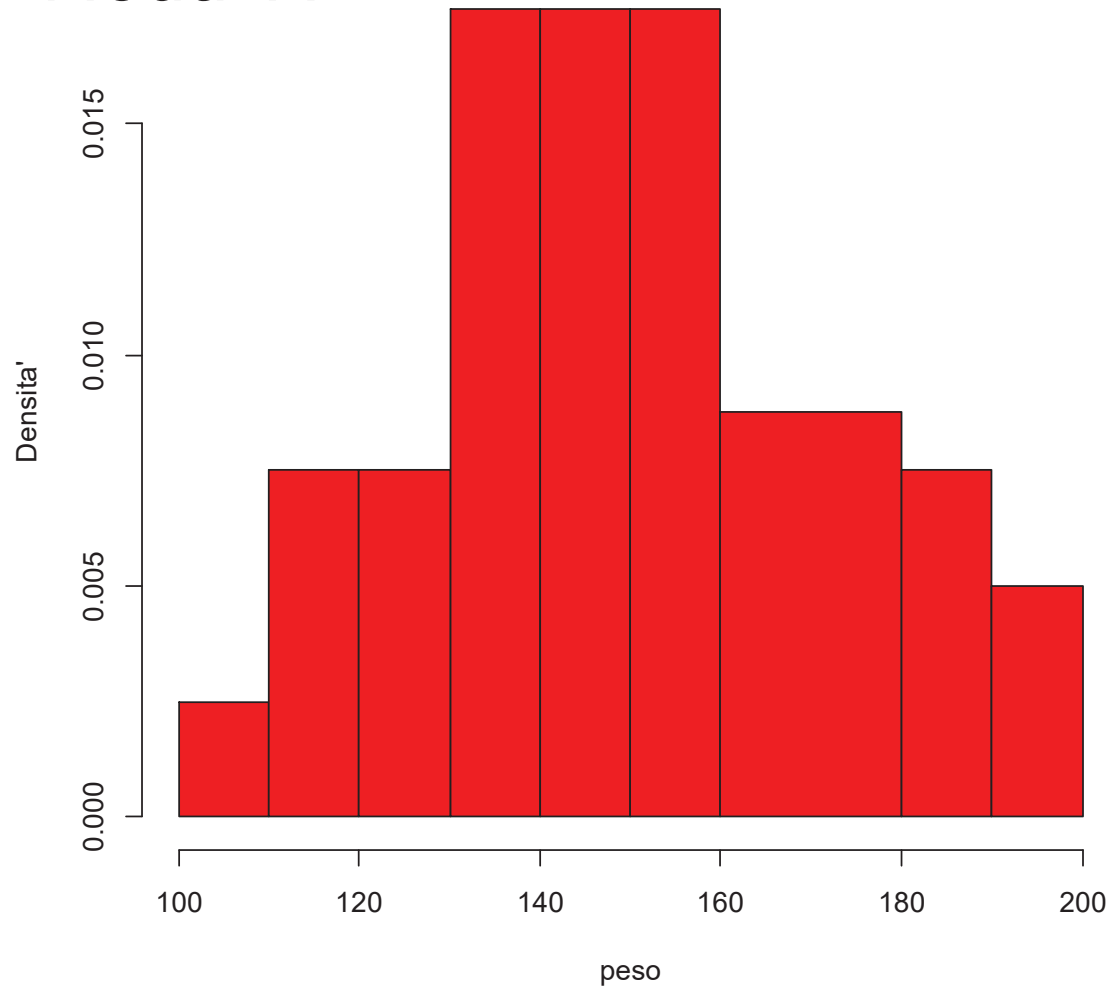
Entrambi i
valori 0 e 1
sono una moda:
bi-modalità

Indici di centralità: mo

Unità	Peso	Peso o.
1	118	106
2	151	114
3	143	118
4	172	118
5	147	122
6	146	127
7	138	127
8	175	132
9	134	134
10	172	135
11	118	136
12	151	138
13	155	139
14	155	140
15	146	143
16	135	143
17	127	144
18	178	146
19	136	146
20	180	147
21	151	148
22	186	151
23	122	151
24	132	151
25	114	153
26	171	155
27	140	155
28	187	159
29	106	171
30	159	172
31	127	172
32	191	175
33	192	178
34	181	179
35	143	180
36	153	181
37	144	186
38	139	187
39	148	191
40	179	192

Moda ??

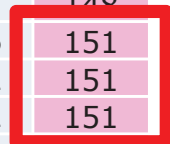
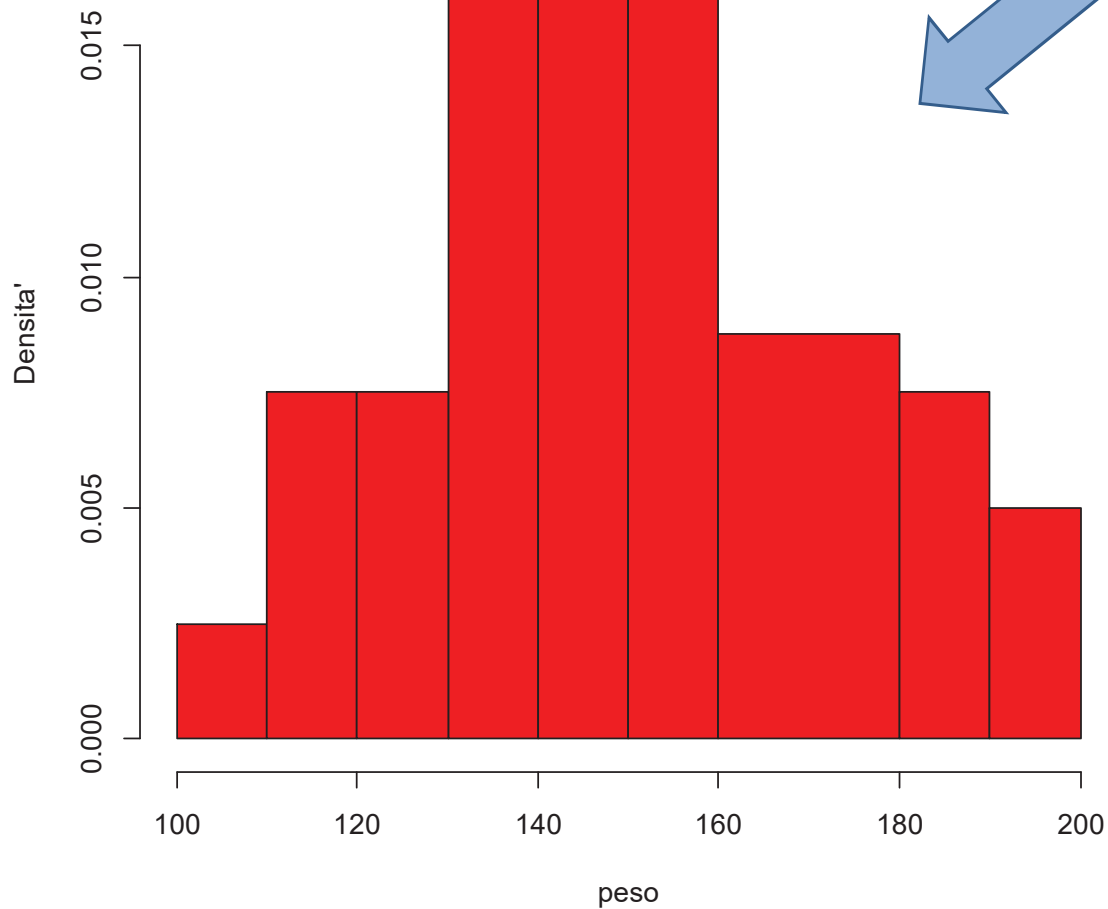
Moda ??



Indici di centralità: mo

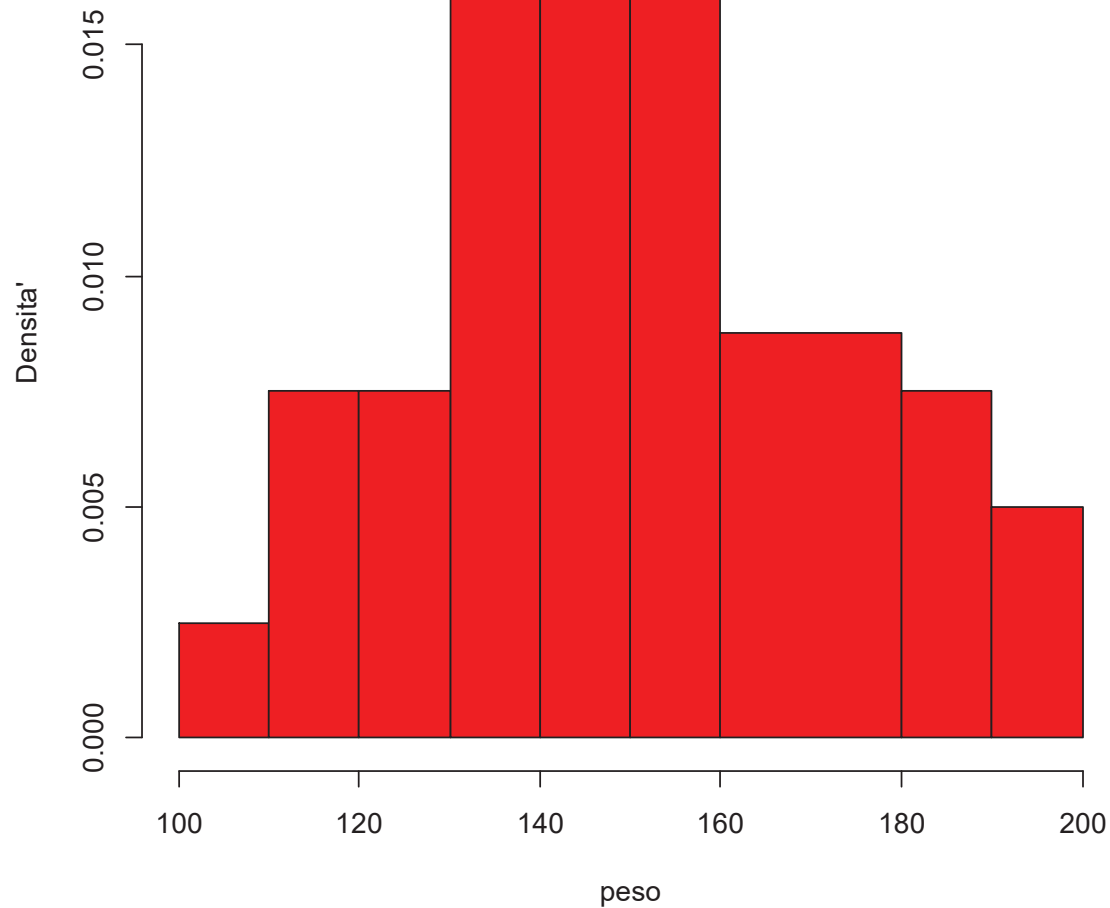
Unità	Peso	Peso o.
1	118	106
2	151	114
3	143	118
4	172	118
5	147	122
6	146	127
7	138	127
8	175	132
9	134	134
10	172	135
11	118	136
12	151	138
13	155	139
14	155	140
15	146	143
16	135	143
17	127	144
18	178	146
19	136	146
20	180	147
21	151	148
22	186	151
23	122	151
24	132	151
25	114	153
26	171	155
27	140	155
28	187	159
29	106	171
30	159	172
31	127	172
32	191	175
33	192	178
34	181	179
35	143	180
36	153	181
37	144	186
38	139	187
39	148	191
40	179	192

Moda



Indici di centralità: mo

Tre classi
modali



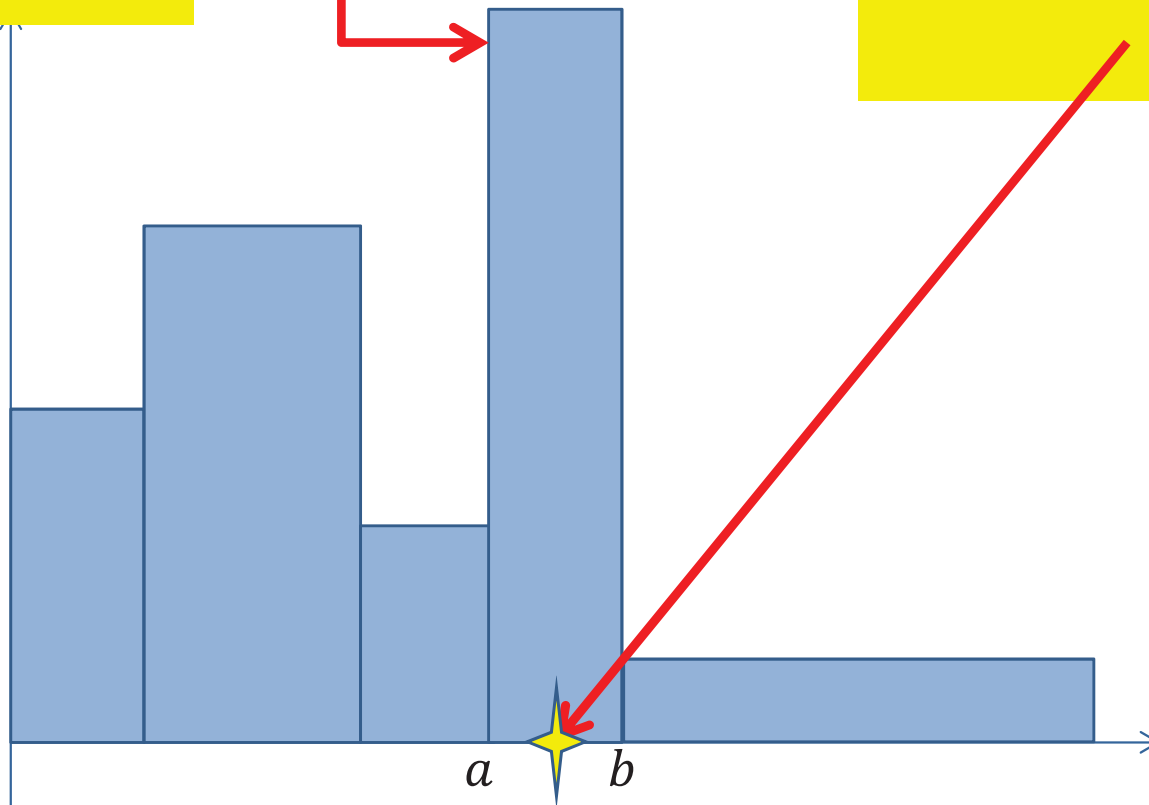
Unità	Peso	Peso o.
1	118	106
2	151	114
3	143	118
4	172	118
5	147	122
6	146	127
7	138	127
8	175	132
9	134	134
10	172	135
11	118	136
12	151	138
13	155	139
14	155	140
15	146	143
16	135	143
17	127	144
18	178	146
19	136	146
20	180	147
21	151	148
22	186	151
23	122	151
24	132	151
25	114	153
26	171	155
27	140	155
28	187	159
29	106	171
30	159	172
31	127	172
32	191	175
33	192	178
34	181	179
35	143	180
36	153	181
37	144	186
38	139	187
39	148	191
40	179	192

Indici di centralità: moda

**Classe/i
modale:** con
la **densità**
maggiore:
 $(a, b]$

Moda: valore
centrale della/e
classe modale:

$$\frac{a + b}{2}$$



Indici di centralità: **media**

	Età (y)	Peso (kg)	Altezza (m)	Sesso	Causa di morte
1	19	50.2	1.65	F	Nat.
2	22	75.6	1.78	M	Inc.
3	21	80.1	1.91	M	Inc.
4	23	56.7	1.72	M	Nat.
5	22	75.0	1.81	M	M.C.
6	20	58.3	1.68	F	Tum.

Indici di centralità: media

$$\frac{19 + 22 + 21 + 23 + 22 + 20}{6} = 21.17 \text{ anni}$$

	Età (y)	Peso (kg)	Altezza (m)	Sesso	Causa di morte
1	19	50.2	1.65	F	Nat.
2	22	75.6	1.78	M	Inc.
3	21	80.1	1.91	M	Inc.
4	23	56.7	1.72	M	Nat.
5	22	75.0	1.81	M	M.C.
6	20	58.3	1.68	F	Tum.

Indici di centralità: media

$$\bar{x} = \frac{x_1 + x_2 + \dots + x_n}{n} = \frac{1}{n} \sum_{i=1}^n x_i$$

	Età (y)	Peso (kg)	Altezza (m)	Sesso	Causa di morte
1	19	50.2	1.65	F	Nat.
2	22	75.6	1.78	M	Inc.
3	21	80.1	1.91	M	Inc.
4	23	56.7	1.72	M	Nat.
5	22	75.0	1.81	M	M.C.
6	20	58.3	1.68	F	Tum.

La media aritmetica

$$\bar{x} = \frac{1}{n} \sum_{i=1}^n x_i$$

$$\bar{x} = 151.05$$

Unità	Peso
1	118
2	151
3	143
4	172
5	147
6	146
7	138
8	175
9	134
10	172
11	118
12	151
13	155
14	155
15	146
16	135
17	127
18	178
19	136
20	180
21	151
22	186
23	122
24	132
25	114
26	171
27	140
28	187
29	106
30	159
31	127
32	191
33	192
34	181
35	143
36	153
37	144
38	139
39	148
40	179

X : n. figli

x_i	n_i
0	4
1	6
2	4
3	1

$$n = 15$$

Tabella di
frequenze

La media aritmetica

$$\bar{x} = \frac{1}{n} \sum_{i=1}^n x_i$$

$$\bar{x} = 151.05$$

Unità	Peso
1	118
2	151
3	143
4	172
5	147
6	146
7	138
8	175
9	134
10	172
11	118
12	151
13	155
14	155
15	146
16	135
17	127
18	178
19	136
20	180
21	151
22	186
23	122
24	132
25	114
26	171
27	140
28	187
29	106
30	159
31	127
32	191
33	192
34	181
35	143
36	153
37	144
38	139
39	148
40	179

X : n. figli

x_i	n_i
0	4
1	6
2	4
3	1

$n = 15$

**Tabella di
frequenze**



Unità	n. figli
1	0
2	0
3	0
4	0
5	1
6	1
7	1
8	1
9	1
10	1
11	2
12	2
13	2
14	2
15	3

La media aritmetica

$$\bar{x} = \frac{1}{n} \sum_{i=1}^n x_i$$

X : n. figli

x_i	n_i
0	4
1	6
2	4
3	1

$n = 15$



Unità	n. figli
1	0
2	0
3	0
4	0
5	1
6	1
7	1
8	1
9	1
10	1
11	2
12	2
13	2
14	2
15	3

$$\bar{x} = \frac{0 + 0 + 0 + 0 + 1 + 1 + \dots + 2 + 2 + 3}{15}$$

$$= \frac{0 \times 4 + 1 \times 6 + 2 \times 4 + 3 \times 1}{15} =$$

$$= \frac{1}{n} \sum_{i=1}^K x_i \times n_i = \sum_{i=1}^K x_i \times f_i$$

$$\left(f_i = \frac{n_i}{n} \right)$$

$$\bar{x} = 1.13$$

La media aritmetica

$$\bar{x} = \frac{1}{n} \sum_{i=1}^n x_i$$

x_i	n_i
Inc.	2
Nat.	2
M.C.	1
Tum.	1

$$n = 6$$

Tabella di
frequenze

Inc.+Inc.+....

non senso!

La **moda** si può **sempre** determinare, perchè si calcola usando solo le frequenze;

la **media aritmetica** si calcola solo per variabili **quantitative.**

Indici di centralità: **mediana**

quale dato sta *in mezzo*?

	Età (y)	Peso (kg)	Altezza (m)	Sesso	Causa di morte
1	19	50.2	1.65	F	Nat.
2	22	75.6	1.78	M	Inc.
3	21	80.1	1.91	M	Inc.
4	23	56.7	1.72	M	Nat.
5	22	75.0	1.81	M	M.C.
6	20	58.3	1.68	F	Tum.


Indici di centralità: mediana

quale dato sta *in mezzo*?

	Età (y)	Età ordinata			
1	19	19			
2	22	20			
3	21	21			
4	23	22			
5	22	22			
6	20	23			

Indici di centralità: mediana

quale dato sta *in mezzo*?

	Età (y)	Età ordinata			
1	19	19		metà dei dati minore di 21.5 e metà maggiore di 21.5	
2	22	20			
3	21	21			
4	23	22			
5	22	22			
6	20	23			

Indici di centralità: mediana

quale dato sta *in mezzo*? **mediana: 21.5**

	Età (y)	Età ordinata			
1	19	19	metà dei dati minore di 21.5 e metà maggiore di 21.5		
2	22	20			
3	21	21			
4	23	22			
5	22	22			
6	20	23			

Indici di centralità: me

dati : x_1, x_2, \dots, x_n

$$n = 40$$

dati **riordinati**:

$$x_{(1)} \leq x_{(2)} \leq \dots \leq x_{(n)}$$

$$x_{(20)} = 147$$

$$x_{(21)} = 148$$

Unità	Peso	Peso o.
1	118	106
2	151	114
3	143	118
4	172	118
5	147	122
6	146	127
7	138	127
8	175	132
9	134	134
10	172	135
11	118	136
12	151	138
13	155	139
14	155	140
15	146	143
16	135	143
17	127	144
18	178	146
19	136	146
20	180	147
21	151	148
22	186	151
23	122	151
24	132	151
25	114	153
26	171	155
27	140	155
28	187	159
29	106	171
30	159	172
31	127	172
32	191	175
33	192	178
34	181	179
35	143	180
36	153	181
37	144	186
38	139	187
39	148	191
40	179	192



Indici di centralità: me

dati : x_1, x_2, \dots, x_n

$$n = 40$$

dati **riordinati**:

$$x_{(1)} \leq x_{(2)} \leq \dots \leq x_{(n)}$$

$$x_{(20)} = 147$$

$$x_{(21)} = 148$$

n pari :

La **posizione** della **mediana** è

$$\frac{n + 1}{2} \quad (20.5)$$

Il **valore** della **mediana** è
(**LA mediana**)

$$\frac{x_{(n/2)} + x_{(n/2+1)}}{2} \quad 147.5$$

Unità	Peso	Peso o.
1	118	106
2	151	114
3	143	118
4	172	118
5	147	122
6	146	127
7	138	127
8	175	132
9	134	134
10	172	135
11	118	136
12	151	138
13	155	139
14	155	140
15	146	143
16	135	143
17	127	144
18	178	146
19	136	146
20	180	147
21	151	148
22	186	151
23	122	151
24	132	151
25	114	153
26	171	155
27	140	155
28	187	159
29	106	171
30	159	172
31	127	172
32	191	175
33	192	178
34	181	179
35	143	180
36	153	181
37	144	186
38	139	187
39	148	191
40	179	192

147.5



Indici di centralità: mediana

dati : x_1, x_2, \dots, x_n

$$n = 11$$

dati **riordinati**:

$$x_{(1)} \leq x_{(2)} \leq \dots \leq x_{(n)}$$

n dispari :

La **posizione** della **mediana** è

$$\frac{n + 1}{2} \quad (6)$$

Il **valore** della **mediana** è
(**LA mediana**)

$$x_{\left(\frac{n+1}{2}\right)}$$

Unità	Peso	Peso o.
1	118	118
2	151	118
3	143	134
4	172	138
5	147	143
6	146	146
7	138	147
8	175	151
9	134	172
10	172	172
11	118	175

146