

# Statistica

Statistica descrittiva

**GRAFICI e INDICI**

# La Statistica

“La statistica è una disciplina scientifica che si occupa della **raccolta**, dell’**analisi** e dell’**interpretazione dei dati** ottenuti da osservazioni o esperimenti. (...)” (E. Pearson)

# La Statistica

“La statistica è una disciplina scientifica che si occupa della **raccolta**, dell'**analisi** e dell'**interpretazione dei dati** ottenuti da osservazioni o esperimenti. (...)” (E. Pearson)

- **Insieme di dati prestabilito:** dati ufficiali, per esempio.
- **Dati raccolti secondo un esperimento progettato.**

# La Statistica

“La statistica è una disciplina scientifica che si occupa della **raccolta**, dell'**analisi** e dell'**interpretazione dei dati** ottenuti da osservazioni o esperimenti. (...)” (E. Pearson)

- **Insieme di dati prestabilito:** dati ufficiali, per esempio.
- **Dati raccolti secondo un esperimento progettato.**

Statistica **descrittiva**: organizza, descrive e riassume i dati, tramite grafici ed indici opportuni.

Statistica **inferenziale**: trae conclusioni dai dati tramite i metodi di stima e di verifica di ipotesi, estendendoli, **se possibile!**, ad una opportuna popolazione di riferimento.

# La Statistica

“La statistica è una disciplina scientifica che si occupa della **raccolta**, dell'**analisi** e dell'**interpretazione dei dati** ottenuti da osservazioni o esperimenti. (...)” (E. Pearson)

- **Insieme di dati prestabilito:** dati ufficiali, per esempio.
- **Dati raccolti secondo un esperimento progettato.**

Statistica **descrittiva**: organizza, descrive e riassume i dati, tramite grafici ed indici opportuni.

Statistica **inferenziale**: trae conclusioni dai dati tramite i metodi di stima e di verifica di ipotesi, estendendoli, **se possibile!**, ad una opportuna popolazione di riferimento.

# Le variabili

Le caratteristiche che vengono osservate sulle unità statistiche sono chiamate *caratteri* o **variabili statistiche**.

Variabile	Tipologia	Commento	Confronti
<b>Qualitativa o categorica</b>	<b>Nominale</b> (o sconnesso)	Ex.: Sesso, professione	= , ≠
	<b>Ordinale</b>	Esiste un ordinamento delle <b>modalità</b> : «sereno, poco nuvoloso, coperto»	= , ≠, <
<b>Quantitativa</b>	<b>Discreto</b>	“numero di ...” Ex. età	= , ≠, <
	<b>Continuo</b>	Ex. Altezza, età	= , ≠, <

# Un esempio semplice

	<b>Età (y)</b>	<b>Peso (kg)</b>	<b>Altezza (m)</b>	<b>Sesso</b>	<b>Malattia</b>
1	79	50.2	1.65	F	Cardiov.
2	22	75.6	1.78	M	Diabete
3	51	80.1	1.91	M	Diabete
4	83	56.7	1.72	M	Cardiov.
5	72	75.0	1.81	M	Polmonare
6	60	58.3	1.68	F	Ipert.

# Dati categorici

Che sia nominale o ordinale (“basso, medio, alto, molto alto”), non ci sono molte elaborazioni possibili:

- Frequenze **assolute**

Genere	$n_i$		
F	2		
M	4		
<b>tot.</b>	<b>6</b>		



# Dati categorici

Che sia nominale o ordinale (“basso, medio, alto, molto alto”), non ci sono molte elaborazioni possibili:

- Frequenze assolute, **relative**

Genere	$n_i$	$f_i = n_i/n$	
F	2	2/6	
M	4	4/6	
<b>tot.</b>	<b>6</b>	<b>1</b>	

# Dati categorici

Che sia nominale o ordinale (“basso, medio, alto, molto alto”), non ci sono molte elaborazioni possibili:

- Frequenze assolute, relative e **percentuali**:

Genere	$n_i$	$f_i = n_i/n$	% = $100f_i$
F	2	$2/6=0.333$	33.3%
M	4	$4/6=0.667$	66.7%
<b>tot.</b>	<b>6</b>	<b>1</b>	<b>100%</b>

**distribuzioni di frequenze**

# Dati categorici

permettono di fare confronti  
tra gruppi di diversa numerosità

Che sia nominale o ordinale ("basso, medio, alto, molto alto"), non ci sono molte elaborazioni possibili:

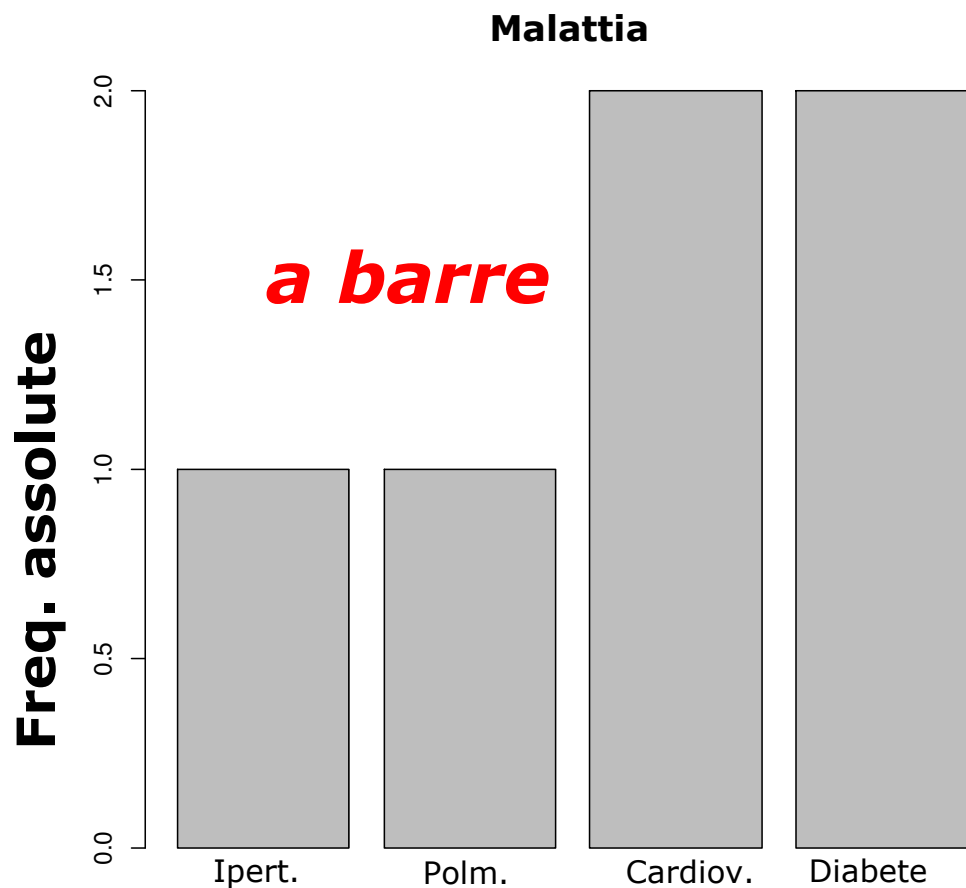
- Frequenze assolute, **relative e percentuali:**

Genere	$n_i$	$f_i = n_i/n$	% = $100f_i$
F	2	$2/6=0.333$	33.3%
M	4	$4/6=0.667$	66.7%
<b>tot.</b>	<b>6</b>	<b>1</b>	<b>100%</b>

**distribuzioni di frequenze**

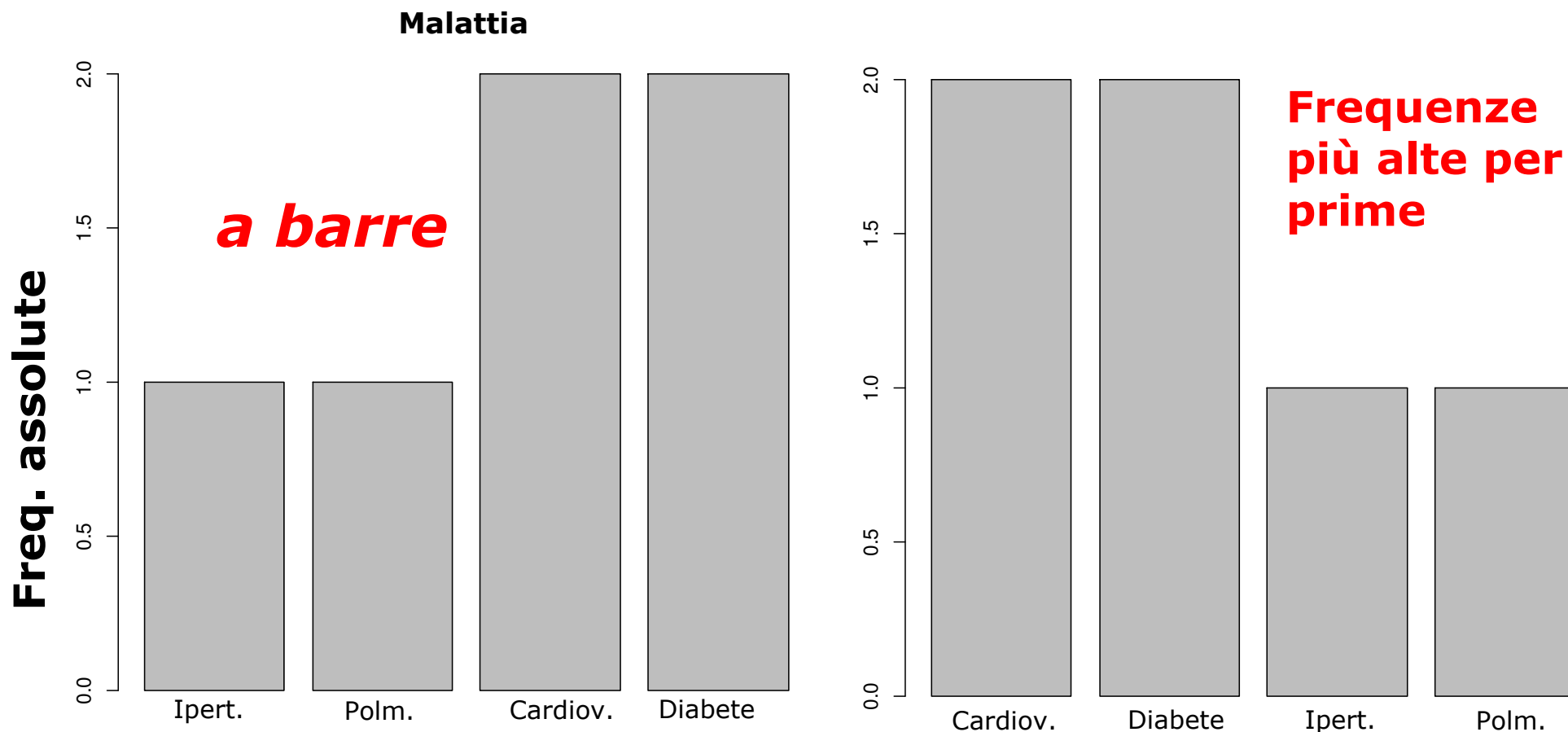
# Dati categorici

- Frequenze assolute, relative e percentuali;
- Diagrammi a barre (delle frequenze):



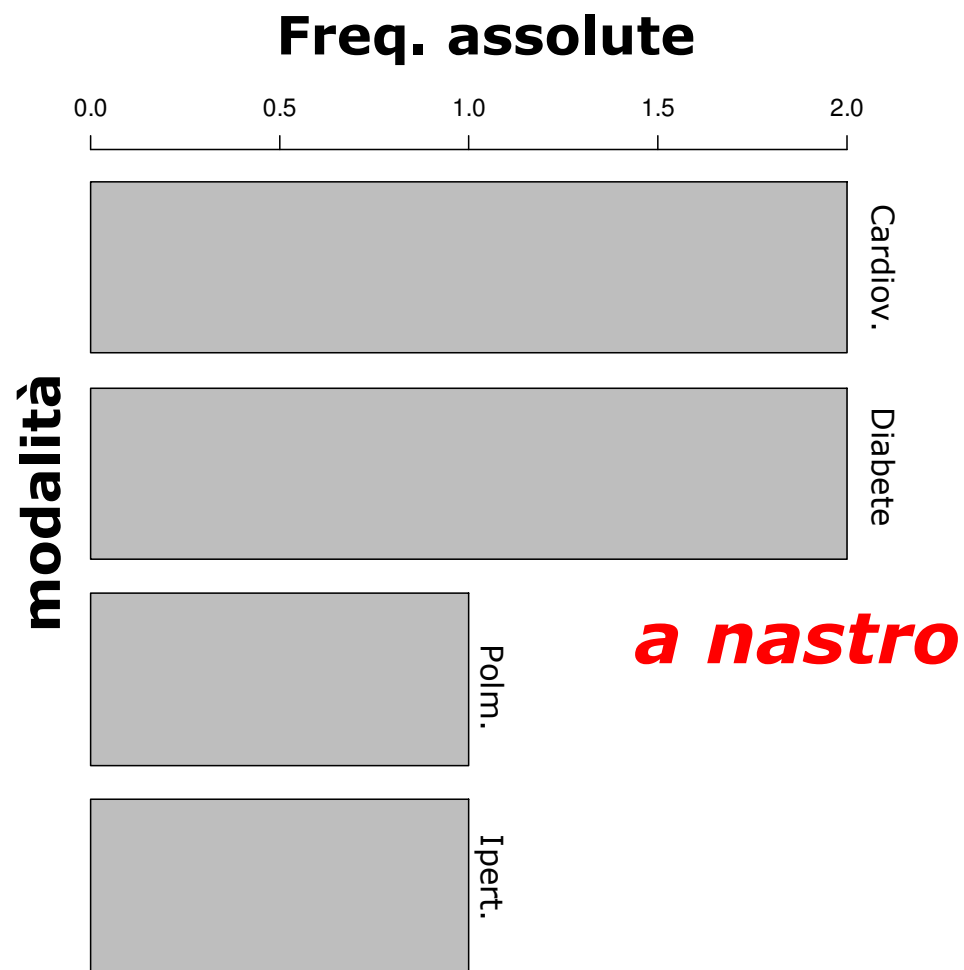
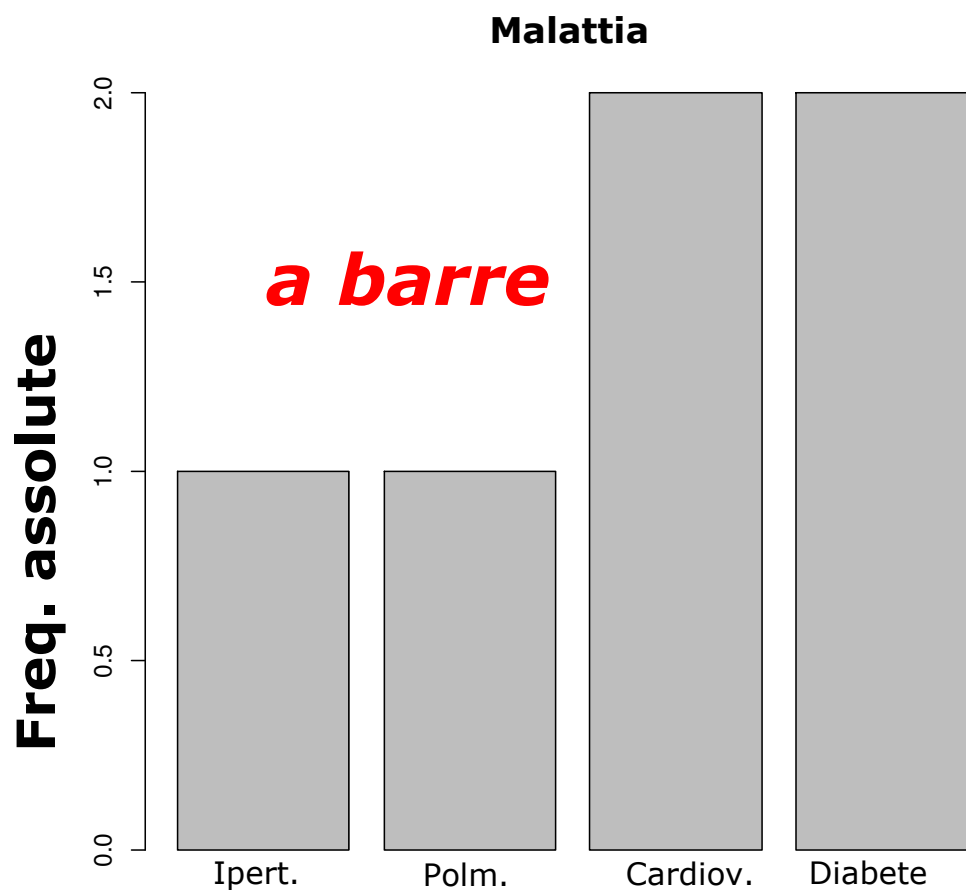
# Dati categorici

- Frequenze assolute, relative e percentuali;
- Diagrammi a barre (delle frequenze):

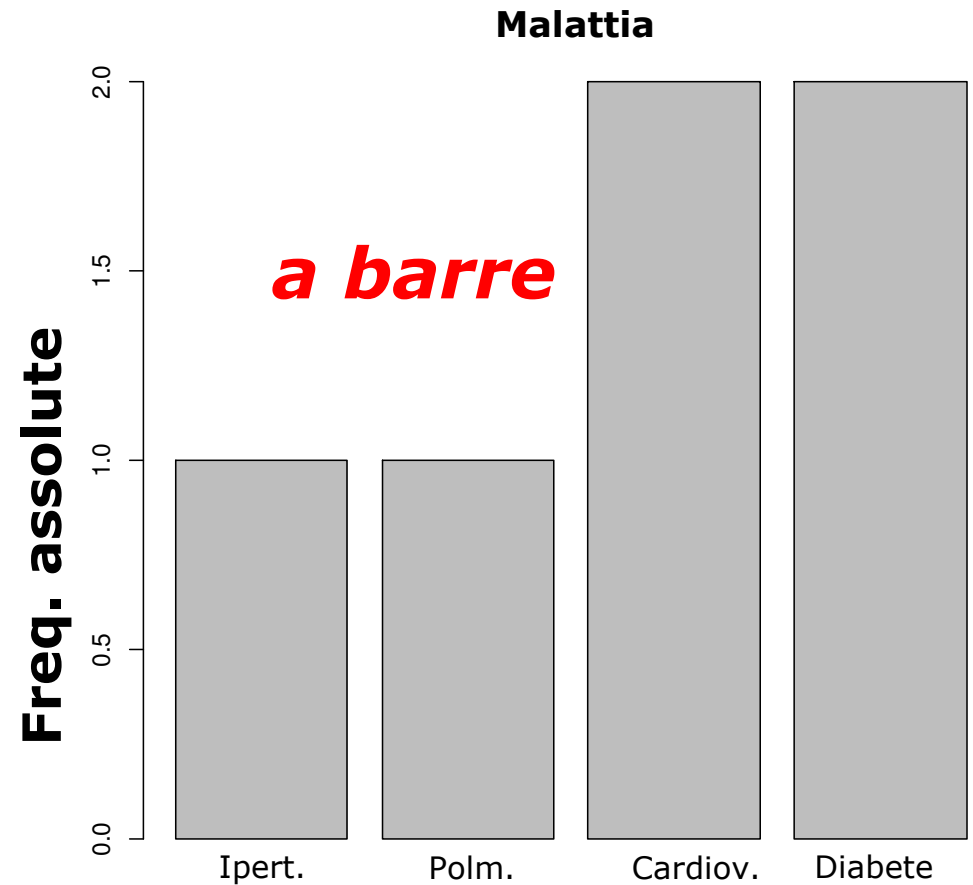
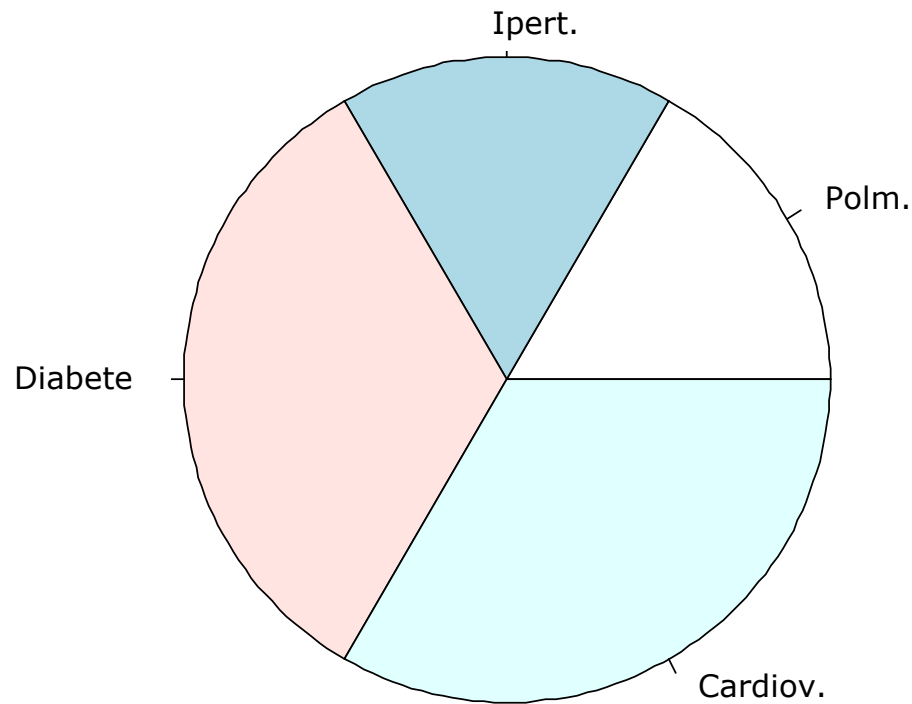


# Dati categorici

- Frequenze assolute, relative e percentuali;
- Diagrammi a barre (delle frequenze):

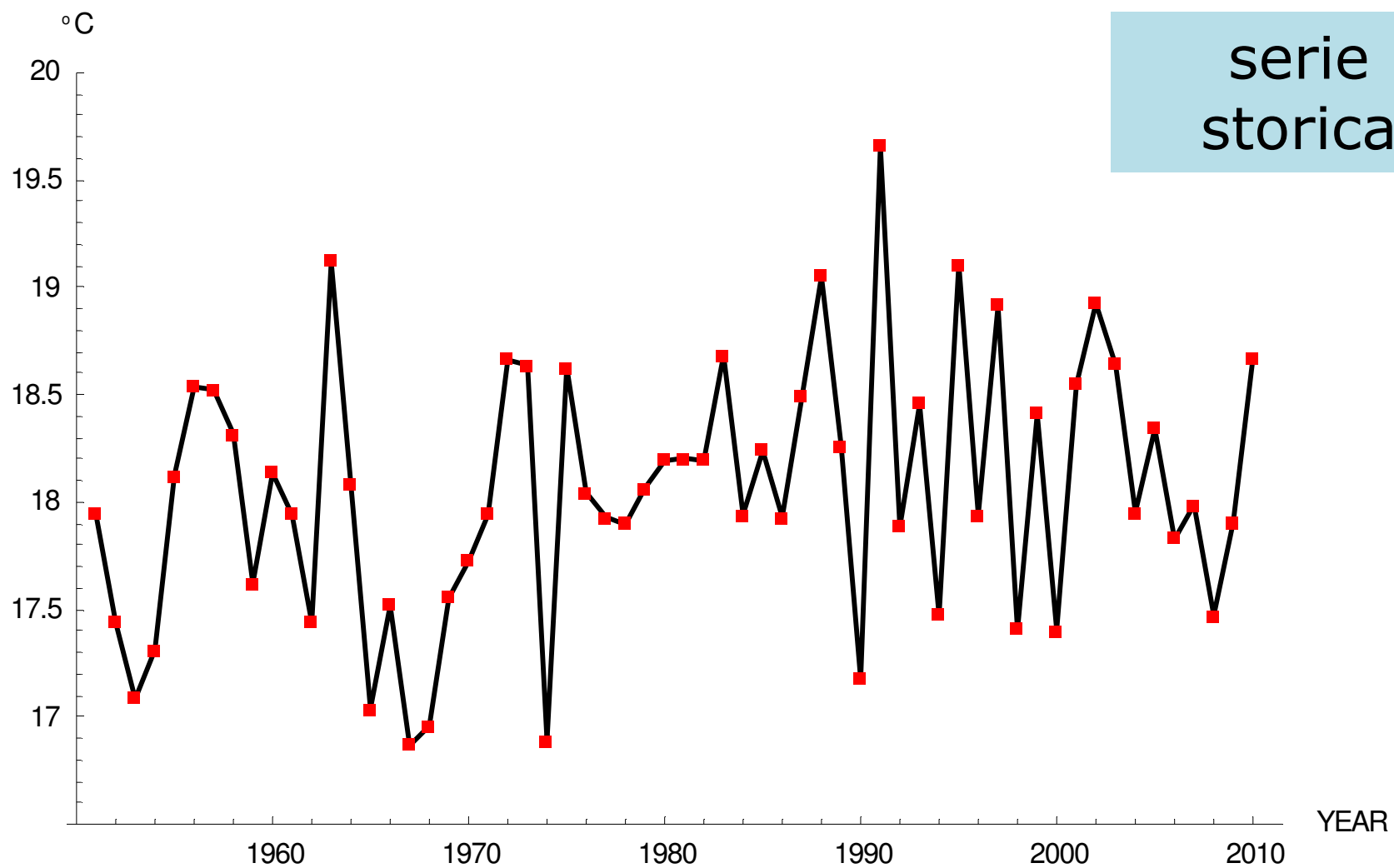


# Dati categorici



I diagrammi a torta sono una pessima rappresentazione, perché l'occhio è capace di giudicare bene le misure lineari, e meno bene le aree.

# Altri grafici



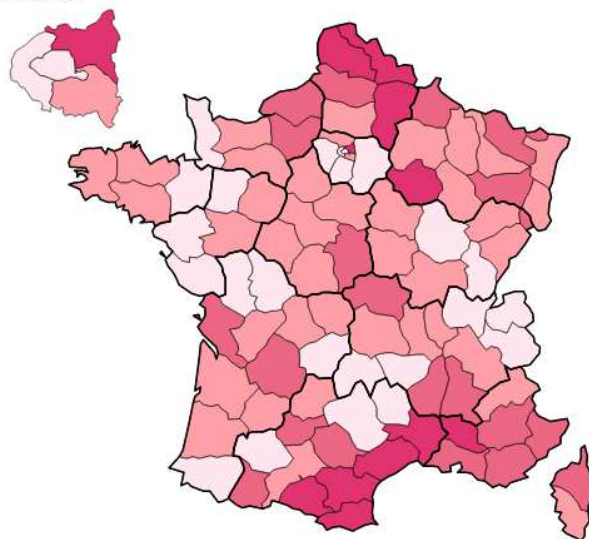
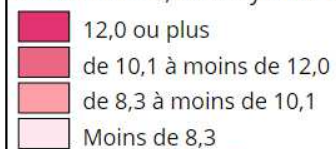


# Altri grafici

## Taux de chômage localisés au 4e trimestre 2016 : comparaisons départementales

4e trim. 2016 (p)

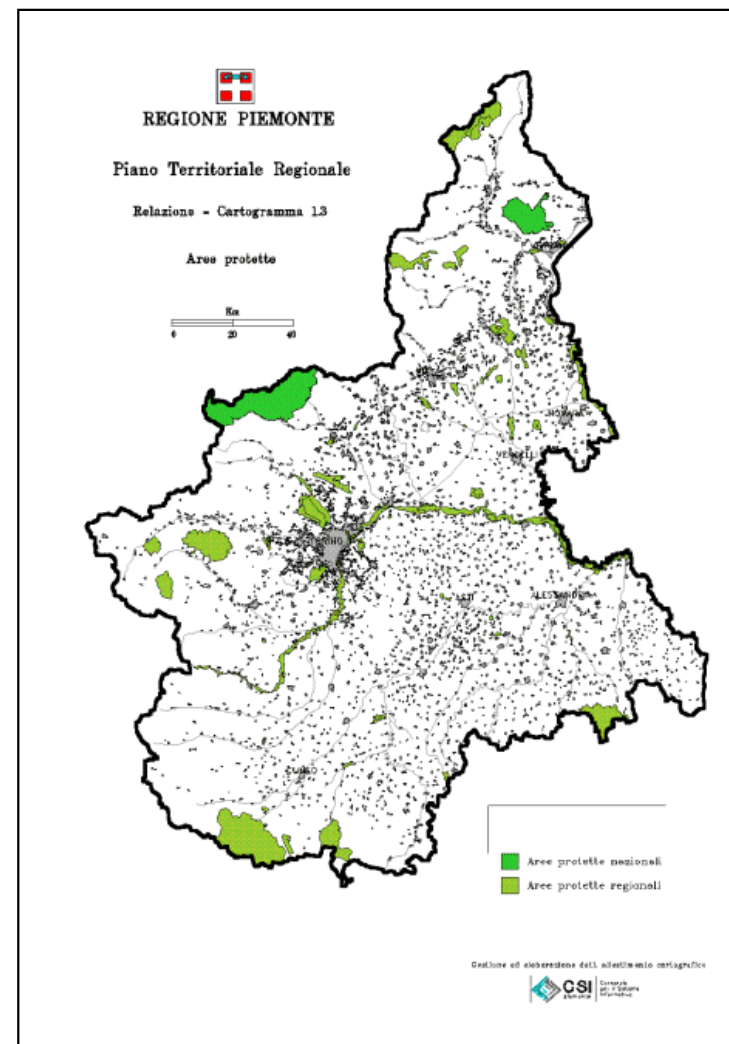
Données CVS, en moyenne trimestrielle (en %)



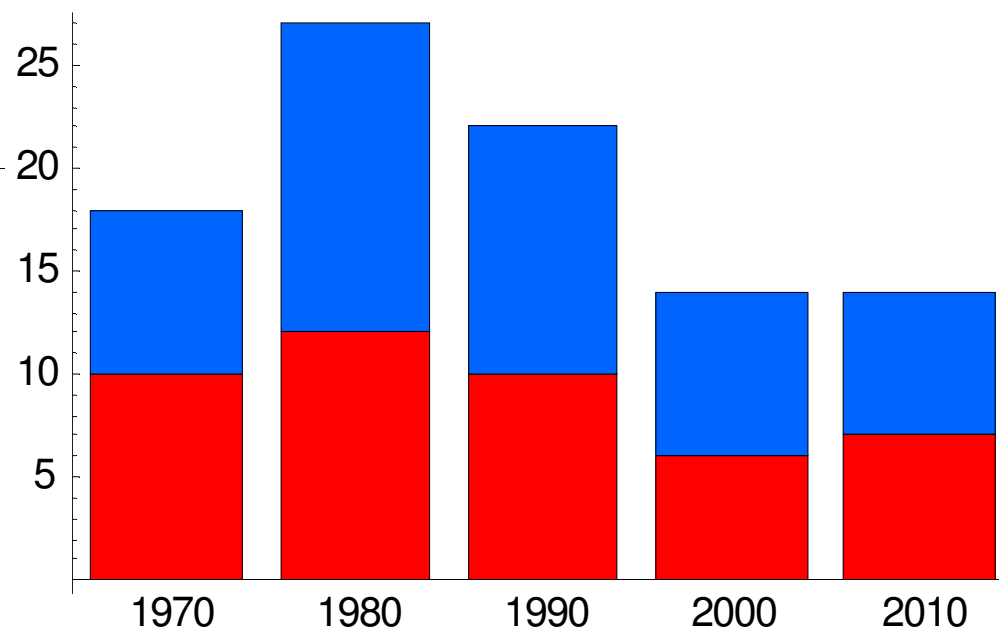
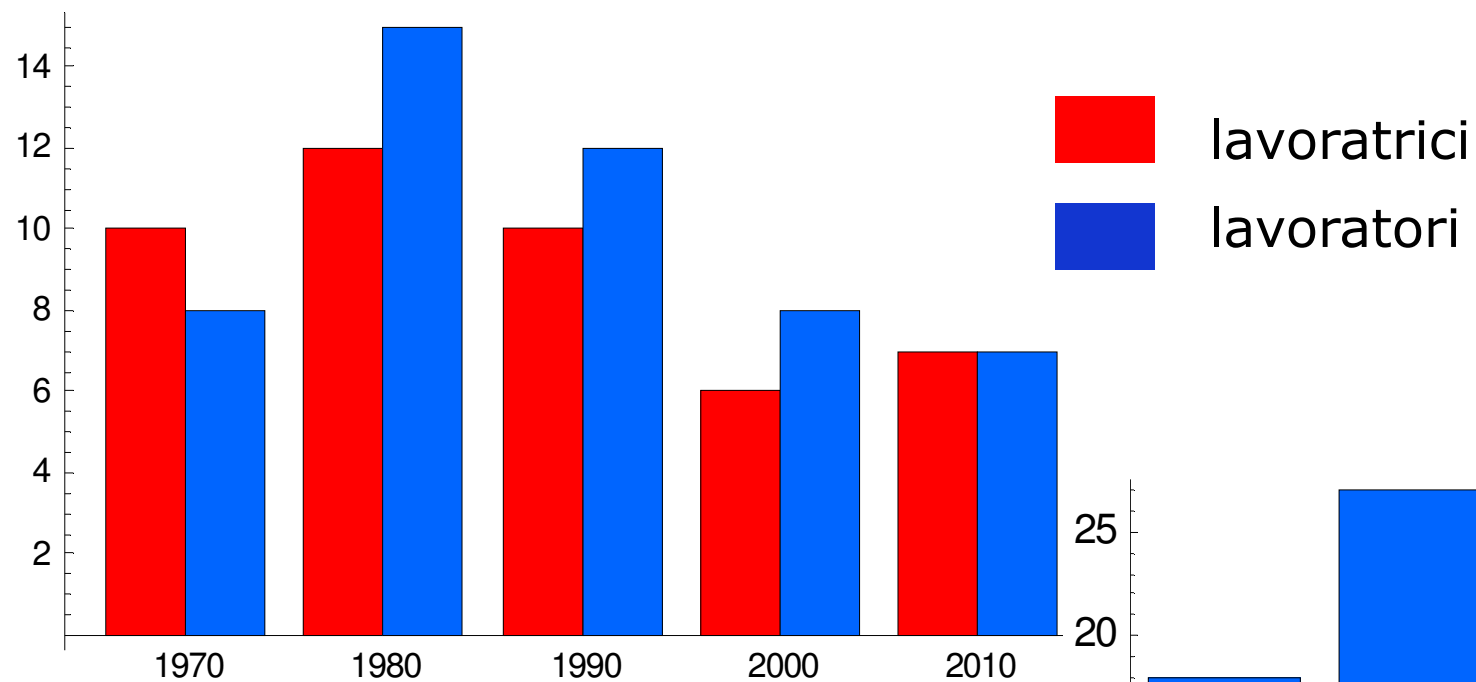
(p) Les données du dernier trimestre sont provisoires.

Champ : France métropolitaine

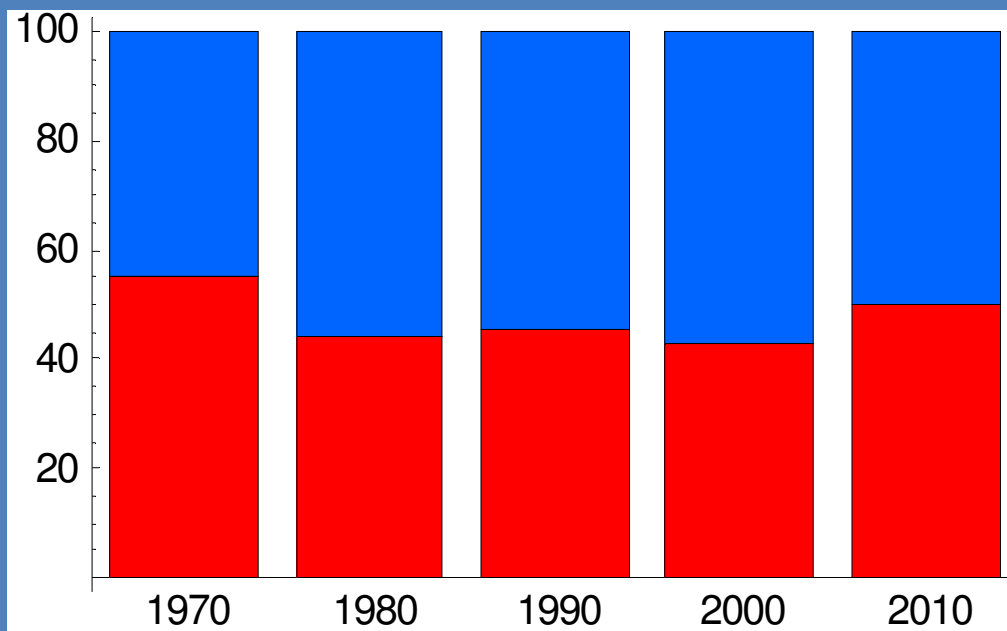
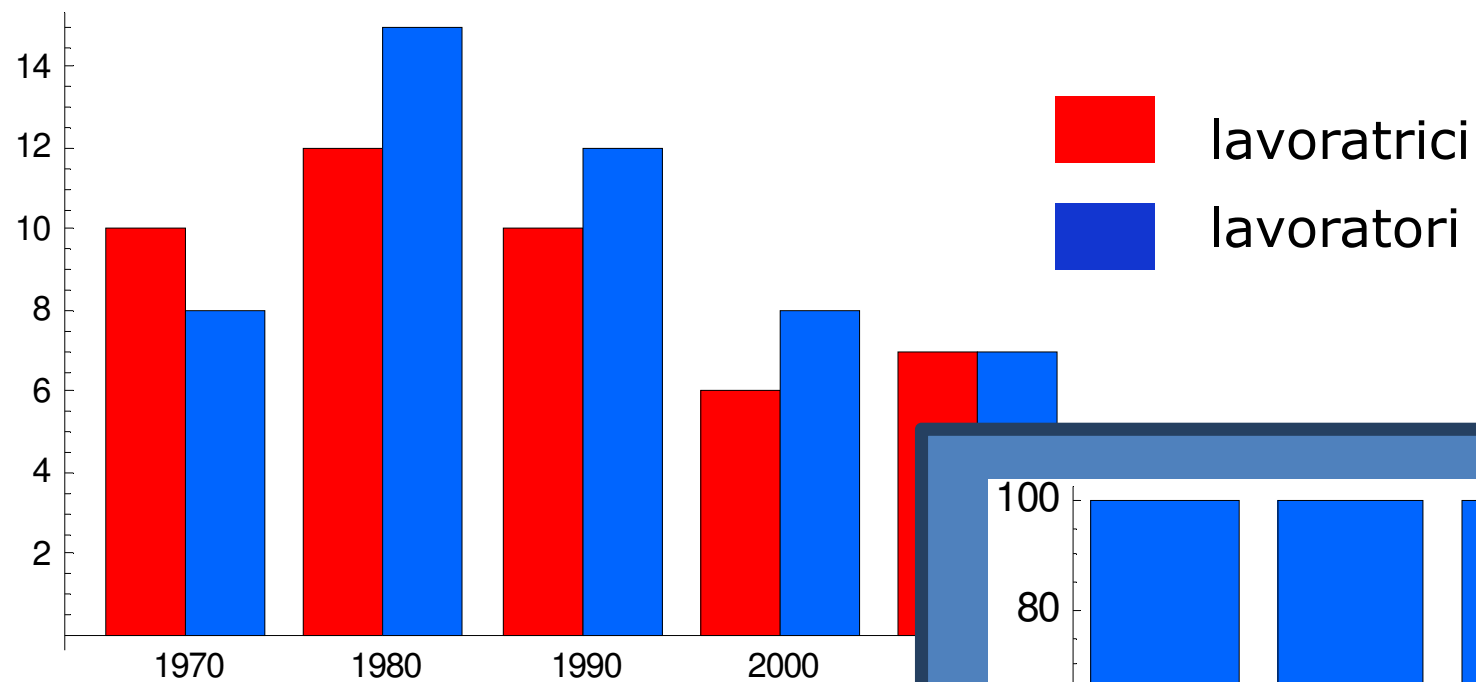
Source : Insee, taux de chômage localisés et taux de chômage au sens du BIT



# Altri grafici



# Altri grafici



# Carattere quantitativo discreto

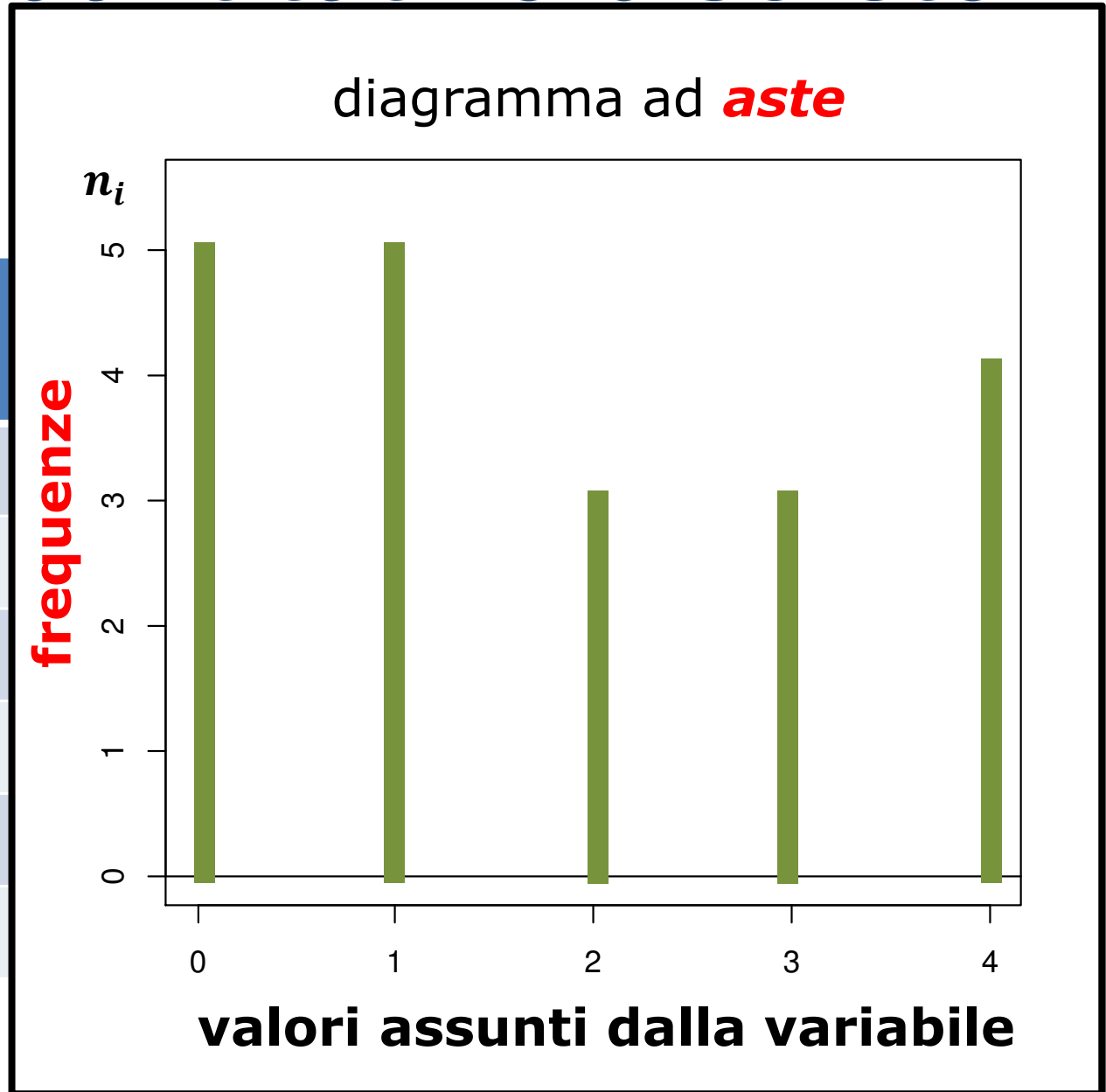
CARATTERE/  
VARIABILE

FREQUENZE

num. figli	$n_i$				
0	5				
1	5				
2	3				
3	3				
4	4				
<b>tot.</b>	<b>20</b>				

# Carattere quantitativo discreto

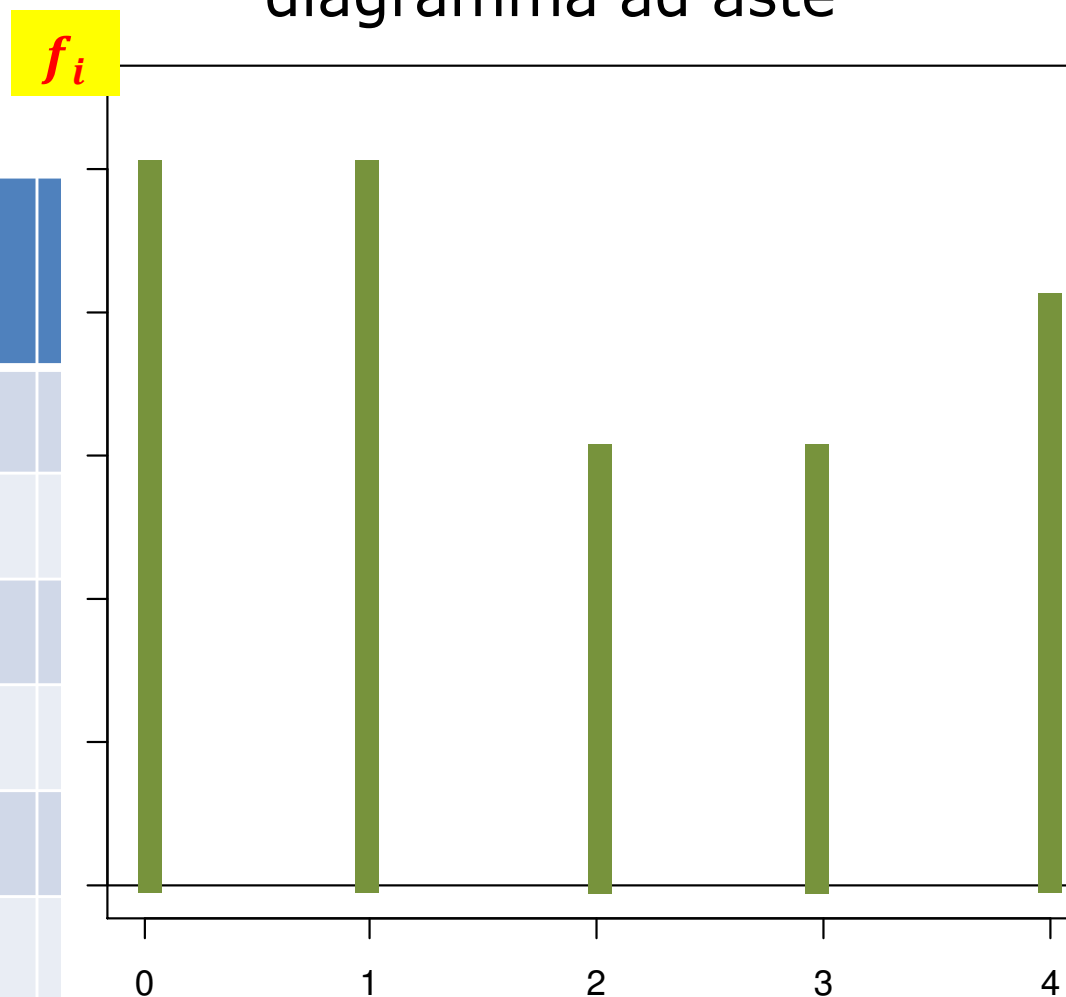
num. figli	$n_i$
0	5
1	5
2	3
3	3
4	4
<b>tot.</b>	<b>20</b>



# Carattere quantitativo discreto

diagramma ad aste

num. figli	$n_i$	$f_i$
0	5	<b>0.25</b>
1	5	<b>0.25</b>
2	3	<b>0.15</b>
3	3	<b>0.15</b>
4	4	<b>0.20</b>
<b>tot.</b>	<b>20</b>	<b>1</b>



**valori assunti dalla variabile**

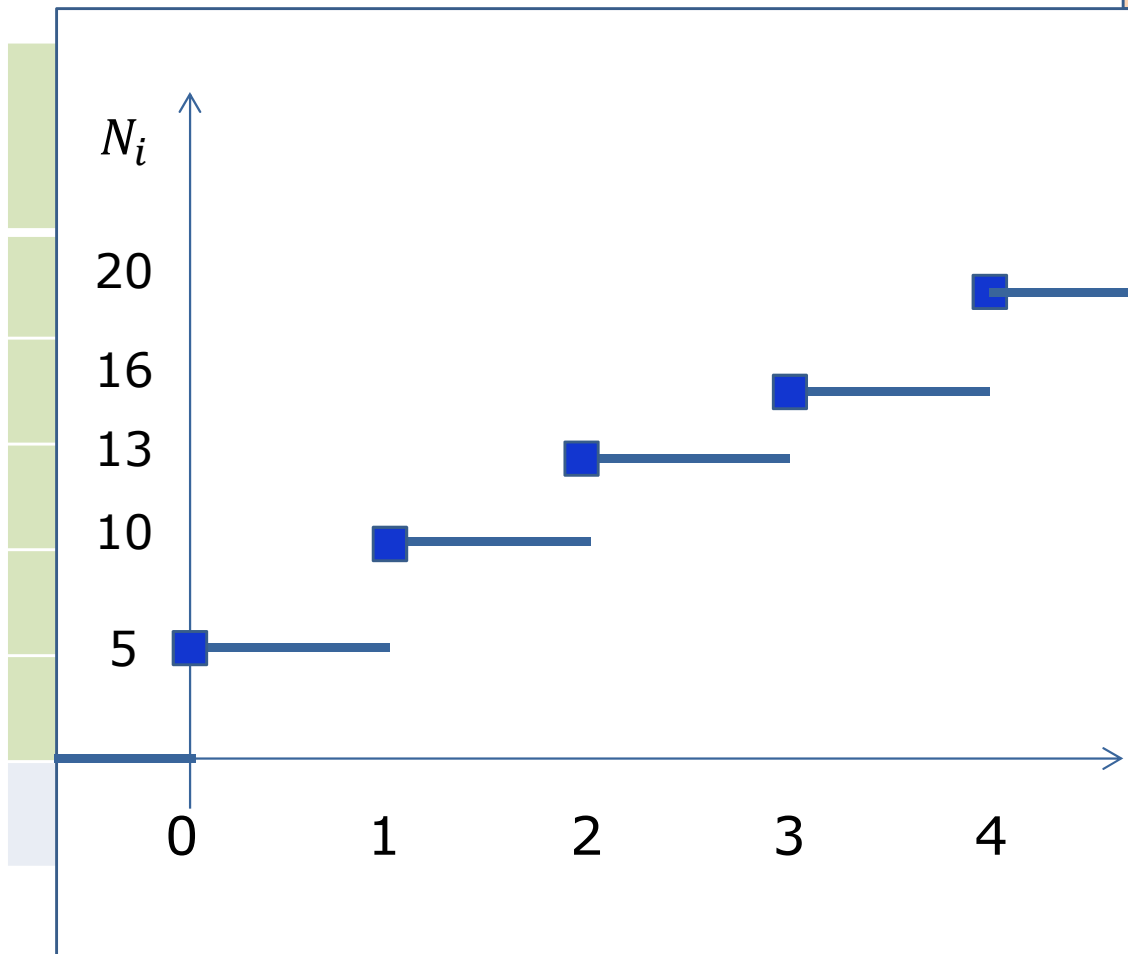
# Carattere quantitativo discreto

FREQUENZE CUMULATE

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	<b>20</b>	<b>1.00</b>	<b>100.0</b>
<b>tot.</b>	<b>20</b>	<b>1</b>	<b>100.0</b>			

# Carattere quantitativo discreto

FREQUENZE CUMULATE

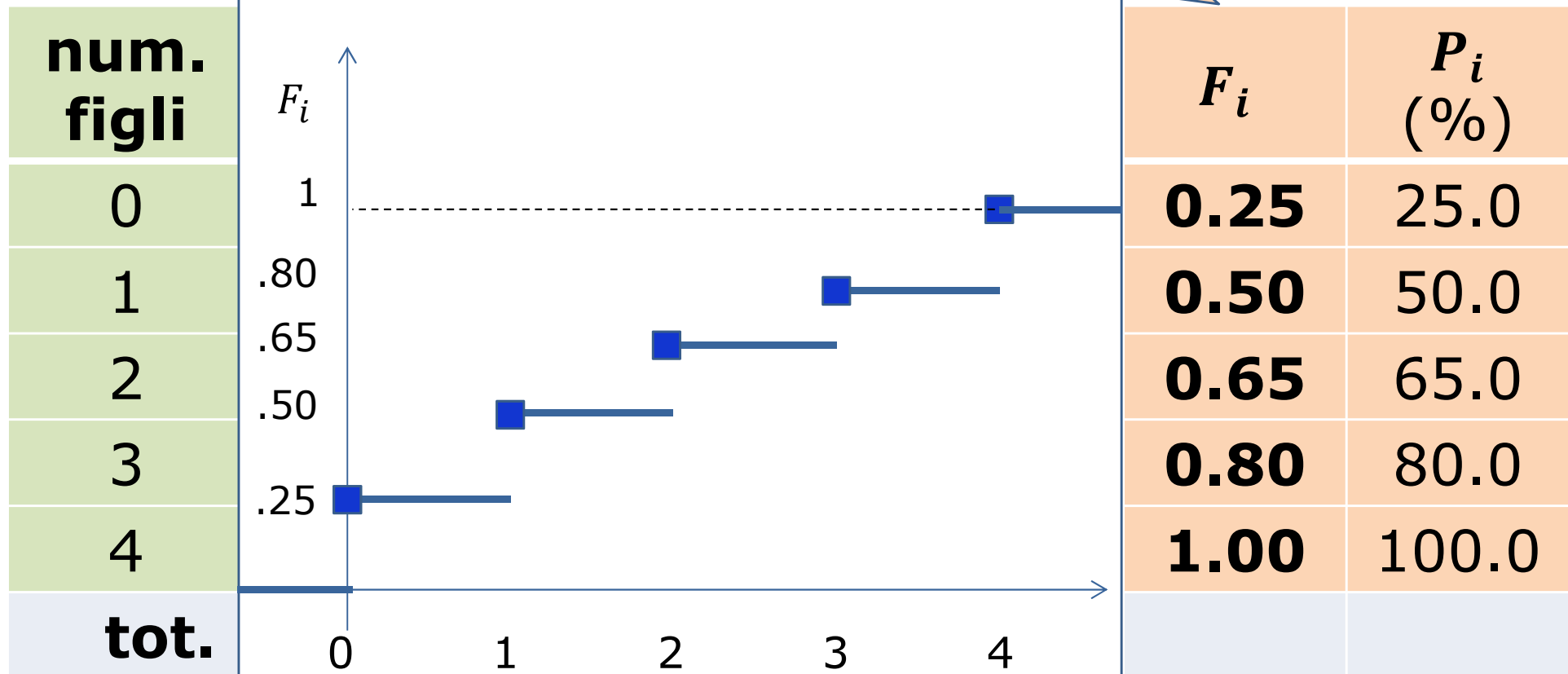


$N_i$	$F_i$	$P_i$ (%)
<b>5</b>	0.25	25.0
<b>10</b>	0.50	50.0
<b>13</b>	0.65	65.0
<b>16</b>	0.80	80.0
<b>20</b>	1.00	100.0

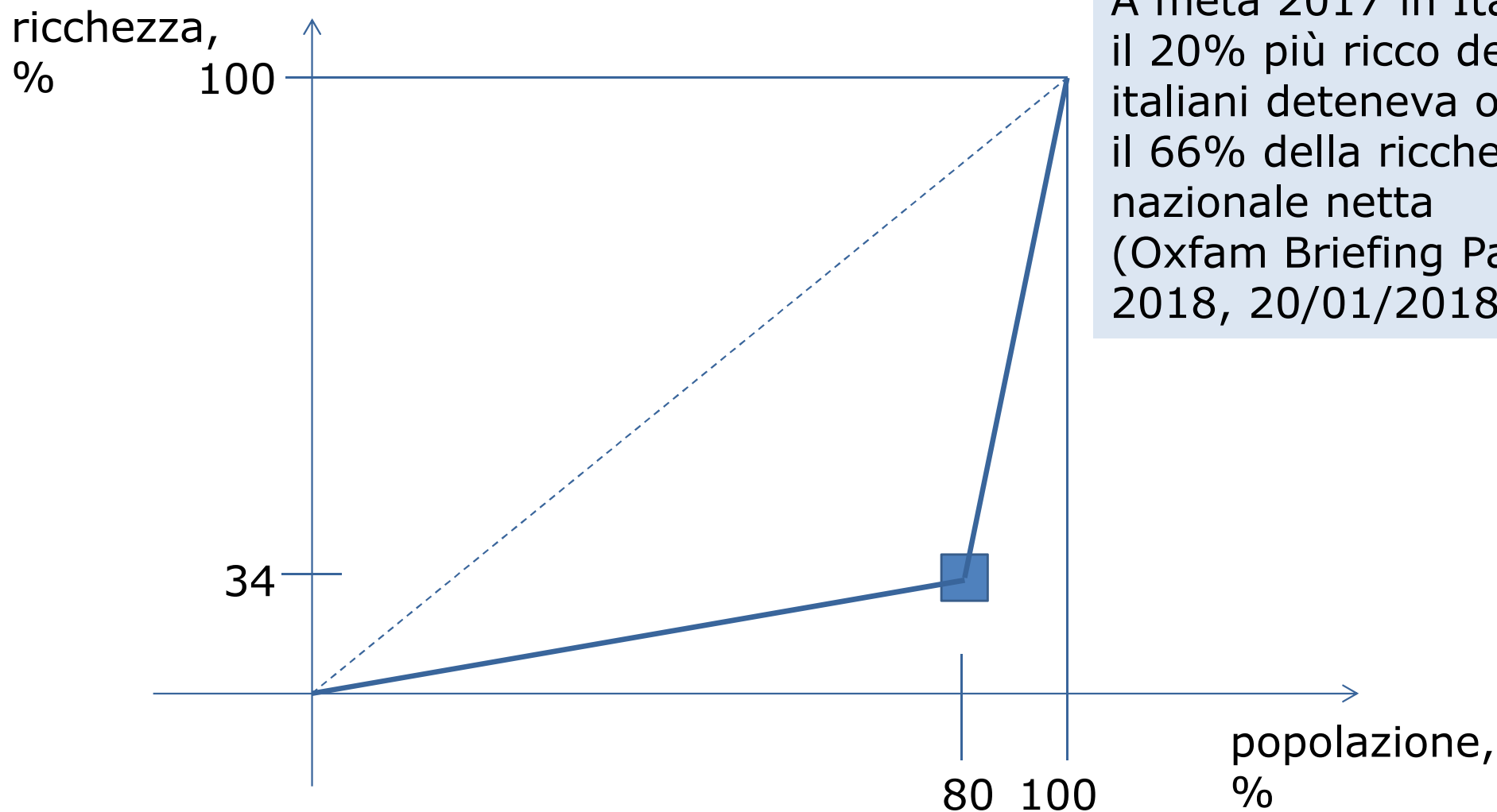


# Carattere quantitativo discreto

FREQUENZE CUMULATE

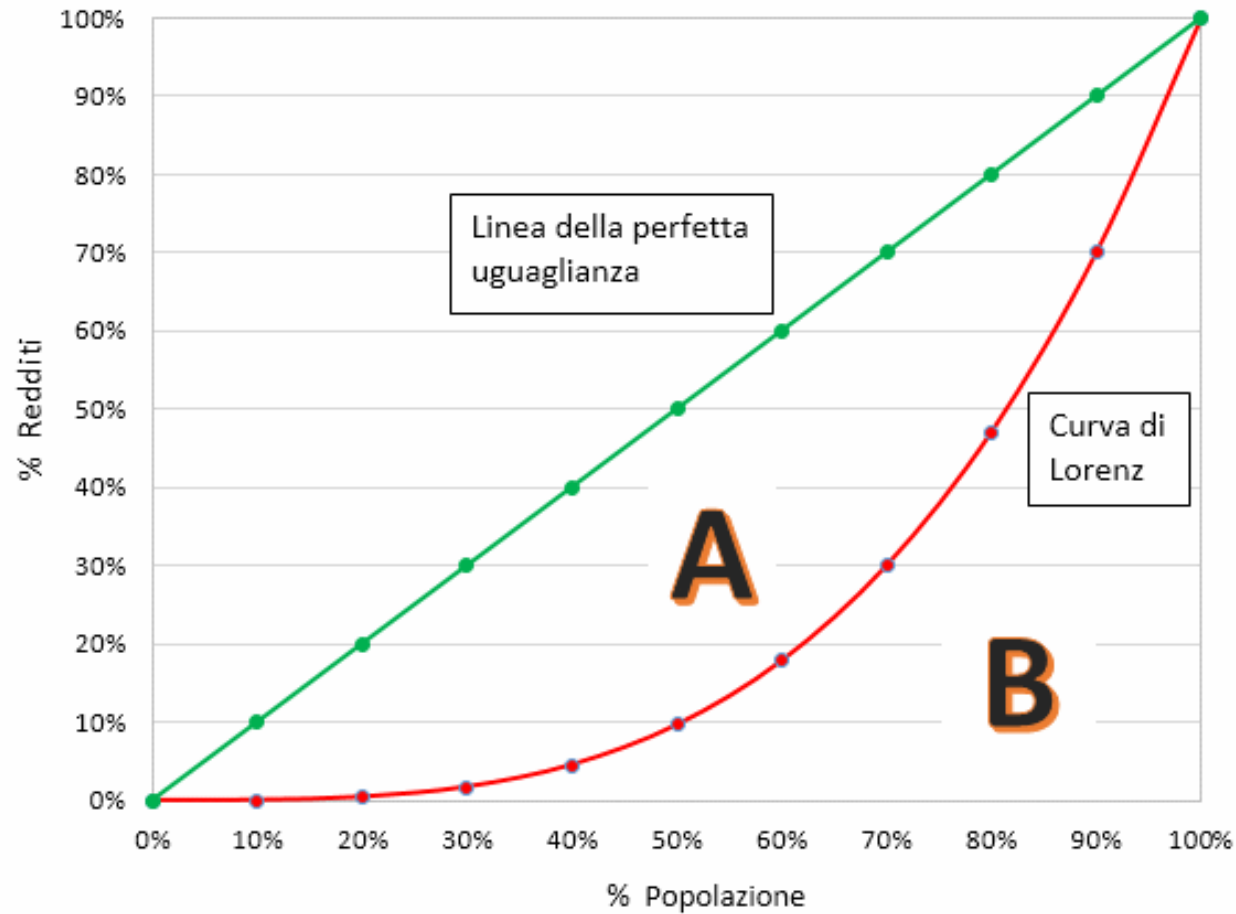


# Frequenze cumulate



A metà 2017 in Italia il 20% più ricco degli italiani deteneva oltre il 66% della ricchezza nazionale netta (Oxfam Briefing Paper 2018, 20/01/2018)

# Curva di concentrazione



$$\text{Coefficiente di Gini} = A / (A+B)$$

# Carattere quantitativo continuo

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

# Carattere quantitativo continuo

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

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

entrambi hanno valori interi

# Carattere quantitativo continuo

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

1.3 figli non ha senso, ma 146.3 (libbre) ha senso!  
→  
continuo!

entrambi hanno valori interi

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

# Carattere quantitativo continuo

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

1.3 figli non ha senso, ma 146.3 (libbre) ha senso!  
continuo!

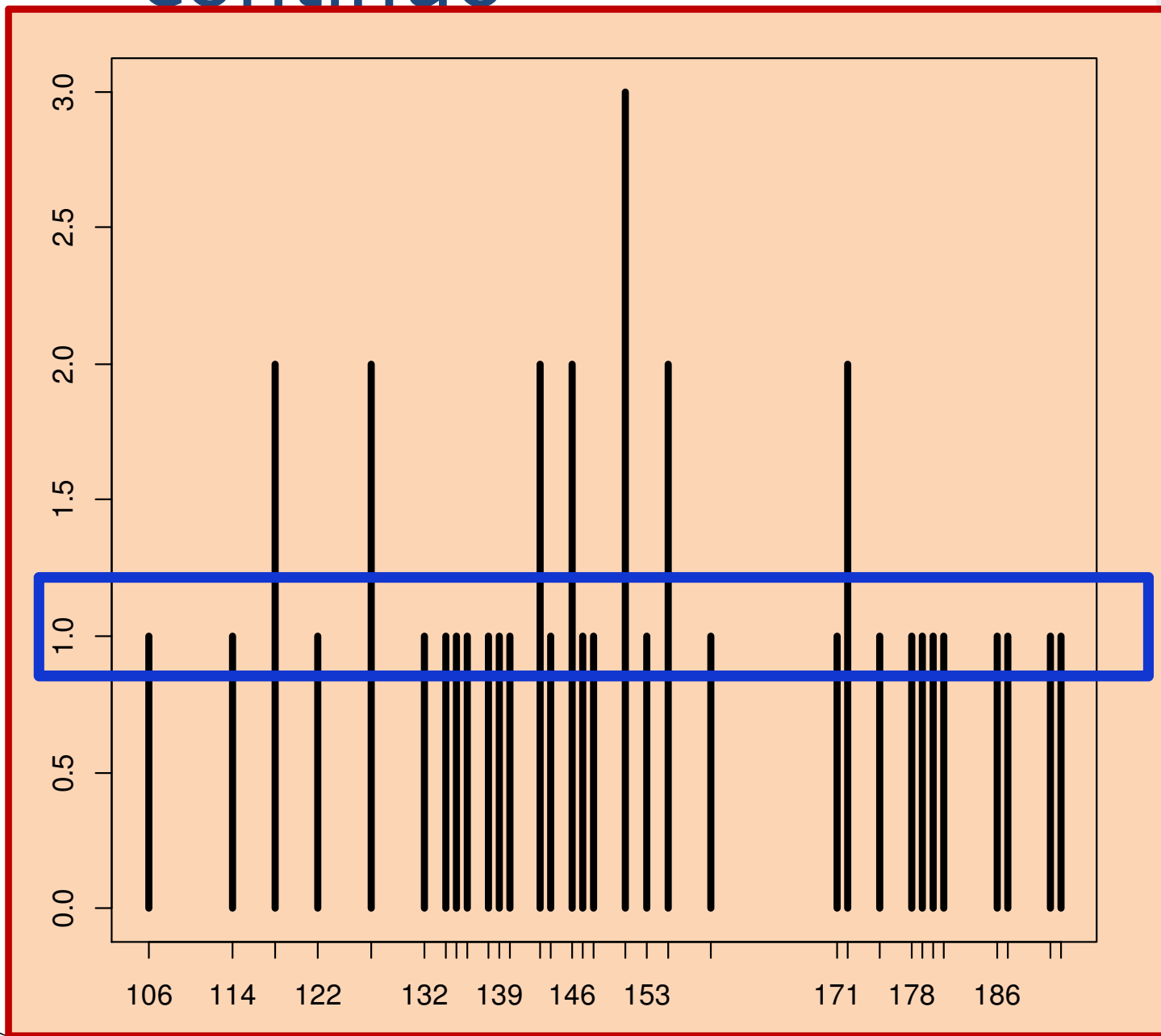
ha senso determinare le frequenze?

interi

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

# Carattere quantitativo continuo

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



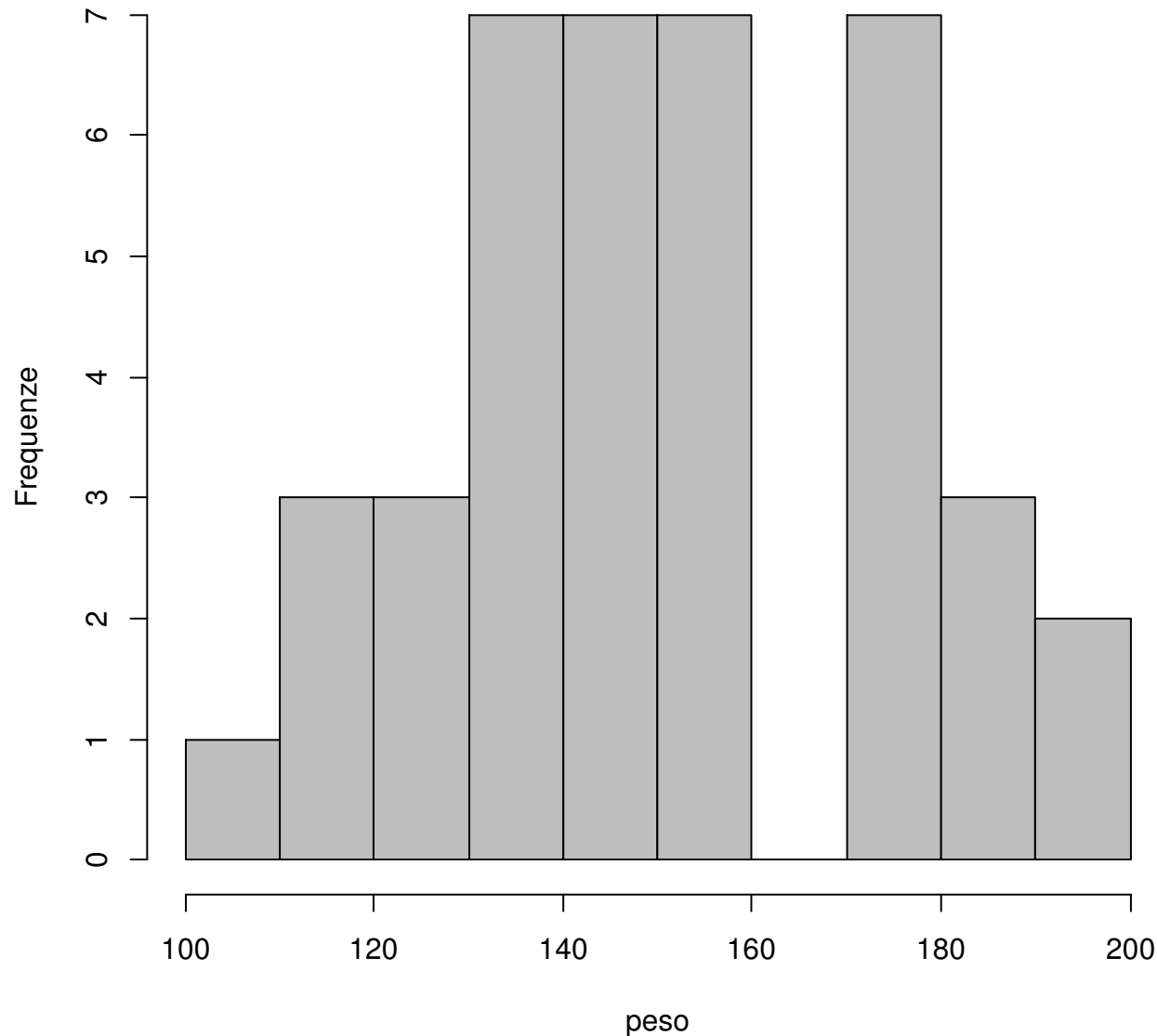


# Dati quantitativi per classi

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
<b>Tot.</b>	<b>40</b>	<b>1</b>				

# Istogramma

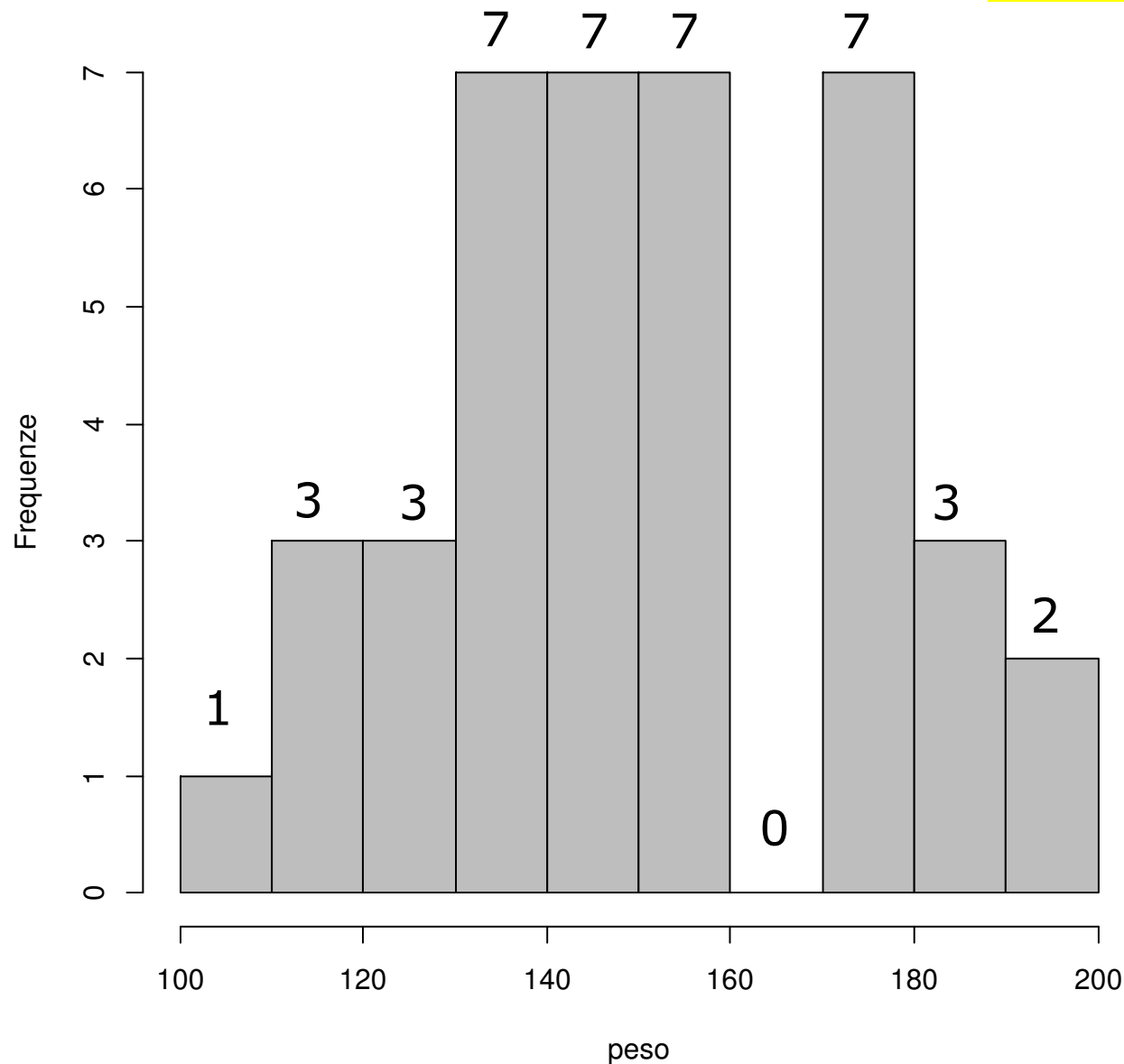
Rappresentazione  
delle frequenze  
tramite l'**area**



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

# Istogramma

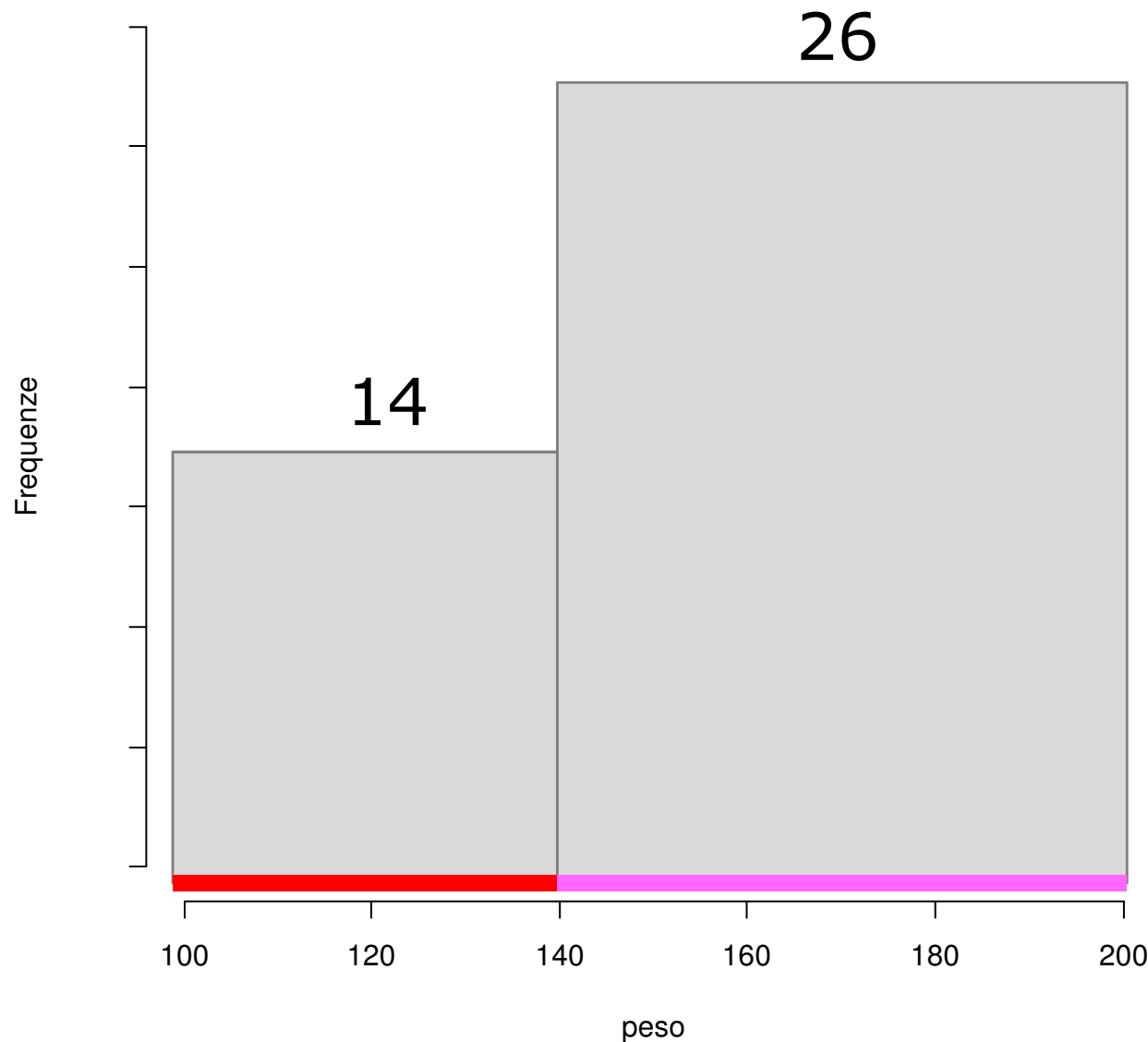
Rappresentazione  
delle frequenze  
tramite l'**area**



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

# Istogramma

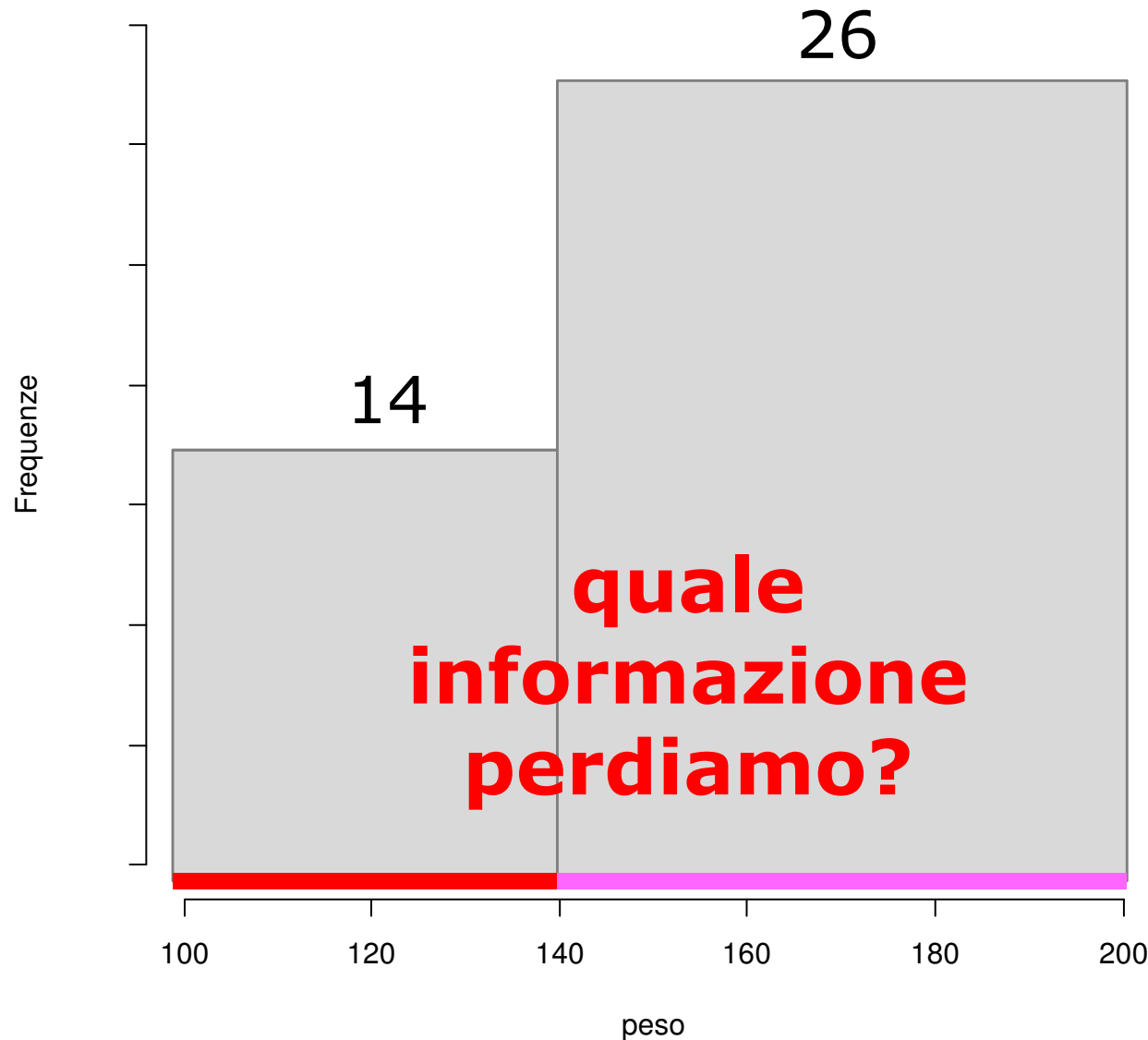
Rappresentazione  
delle frequenze  
tramite l'**area**



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

# Istogramma

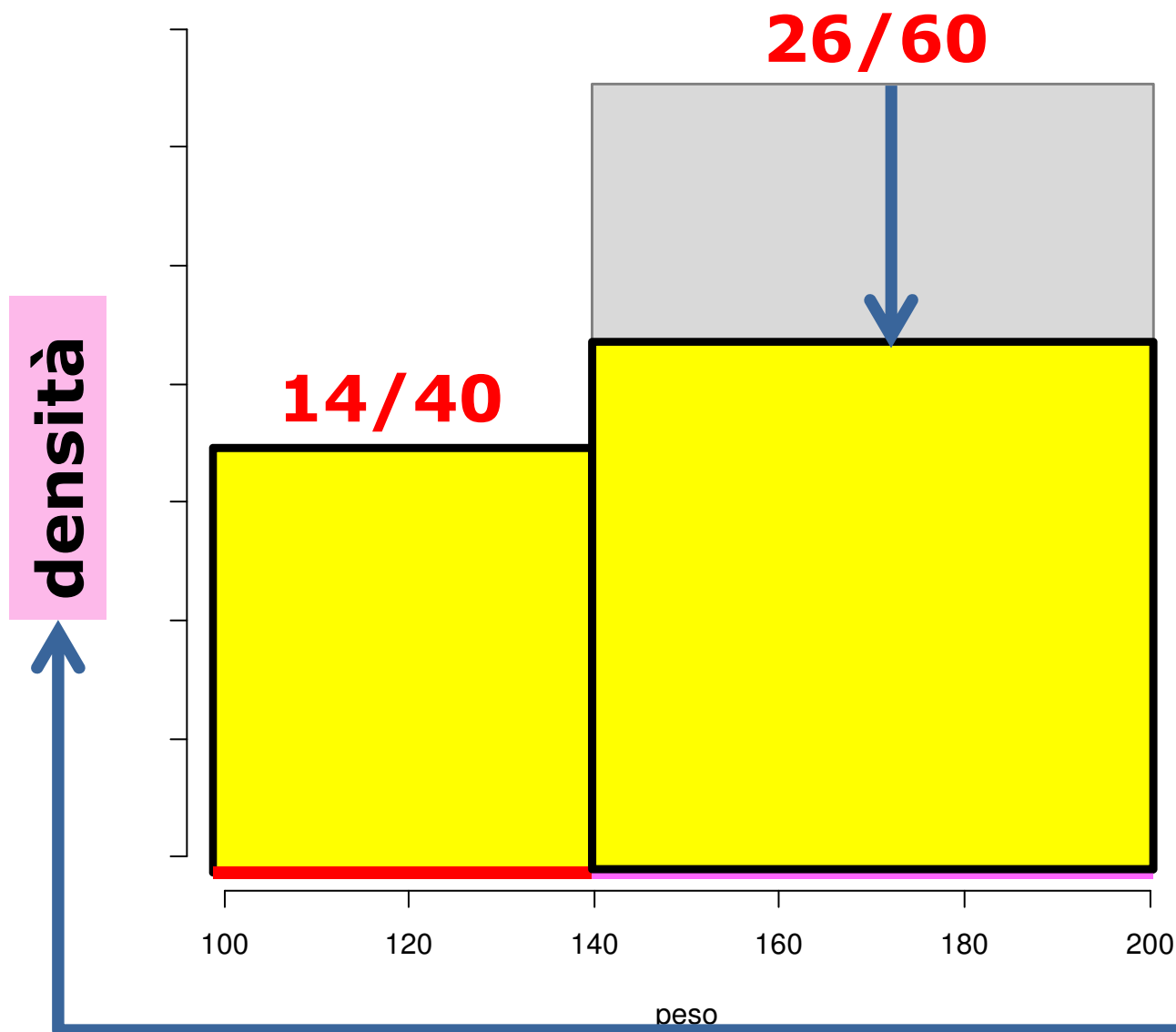
Rappresentazione  
delle frequenze  
tramite l'**area**



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

# Istogramma

Rappresentazione  
delle frequenze  
tramite l'**area**



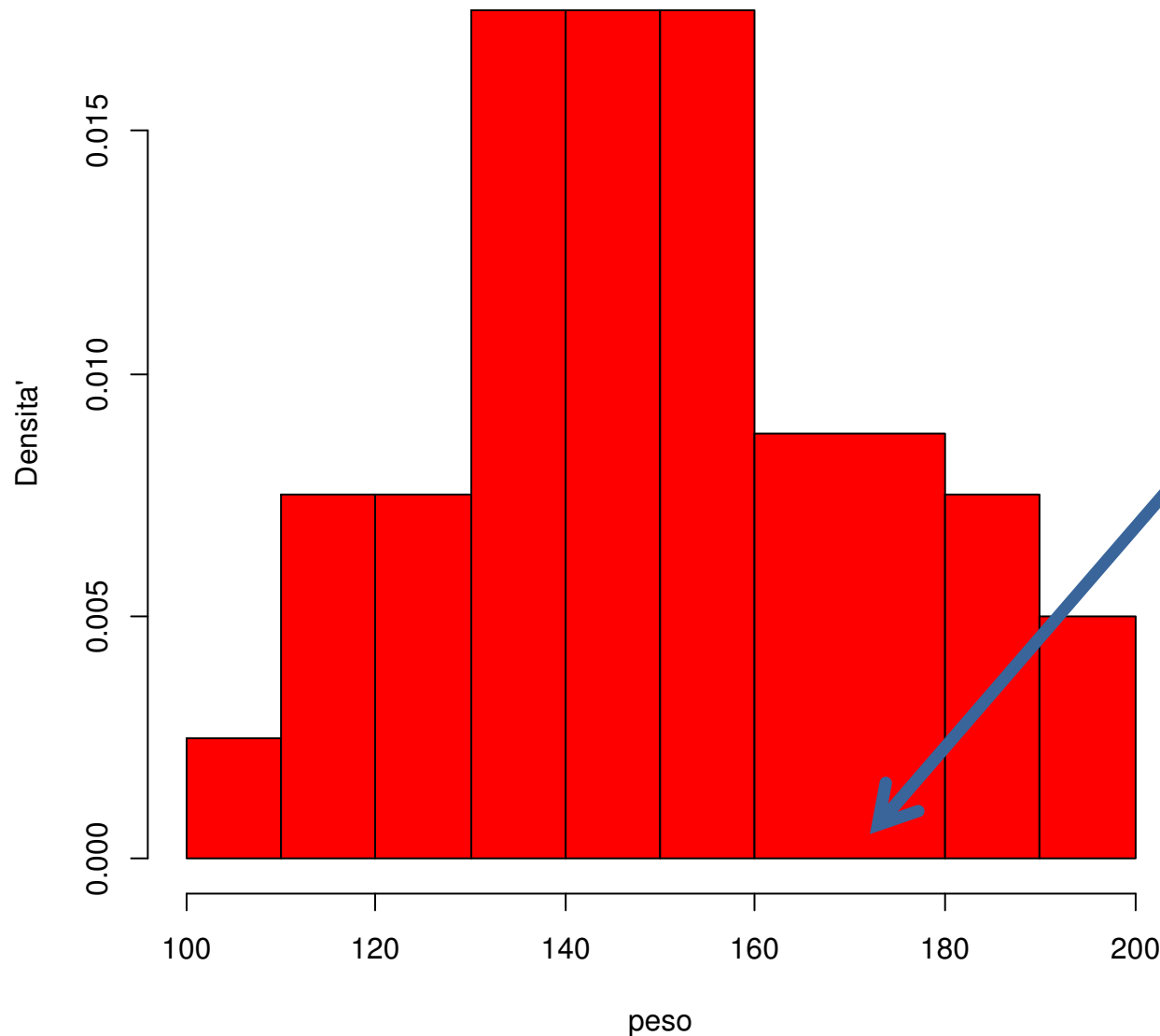
Suddividiamo il range di valori in **intervalli disgiunti** e contiamo quanti dei dati cadono in ciascuna classe, **tenendo conto dell'ampiezza della classe.**

# 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	<b>0.1</b>	0.0025
110- 120	3	0.075	4	10	<b>0.3</b>	0.0075
120- 130	3	0.075	7	10	<b>0.3</b>	0.0075
130- 140	7	0.175	14	10	<b>0.7</b>	0.0175
140- 150	7	0.175	21	10	<b>0.7</b>	0.0175
150- 160	7	0.175	28	10	<b>0.7</b>	0.0175
160- 170	0	0.000	28	10	<b>0.0</b>	0.0
170- 180	7	0.175	35	10	<b>0.7</b>	0.0175
180- 190	3	0.075	38	10	<b>0.3</b>	0.0075
190- 200	2	0.050	40	10	<b>0.2</b>	0.0050
<b>Tot.</b>	<b>40</b>	<b>1</b>				

# 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.***

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

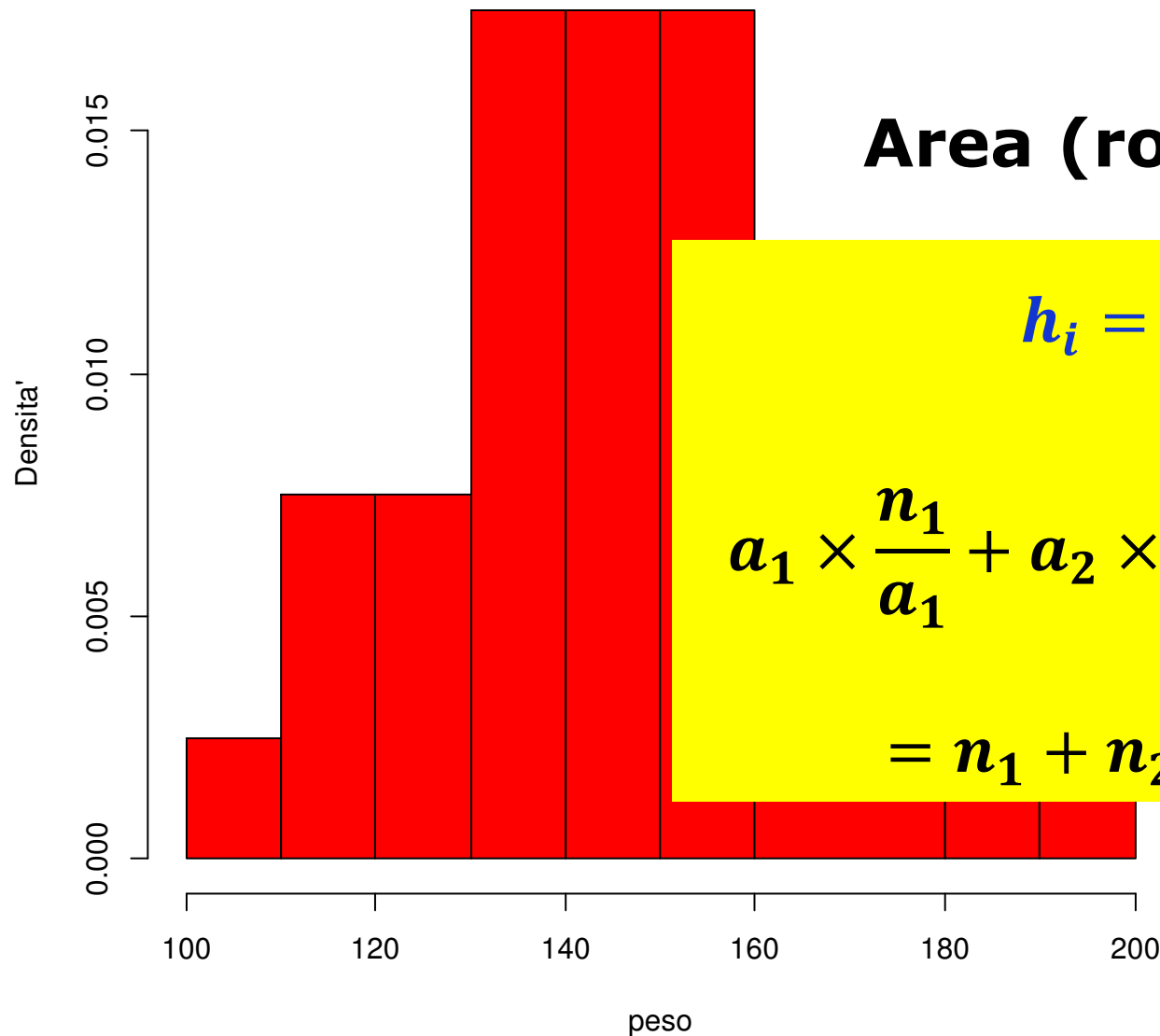
o

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



# Istogramma

Rappresentazione  
delle frequenze  
tramite l'**area**



**Area (rossa) totale:**

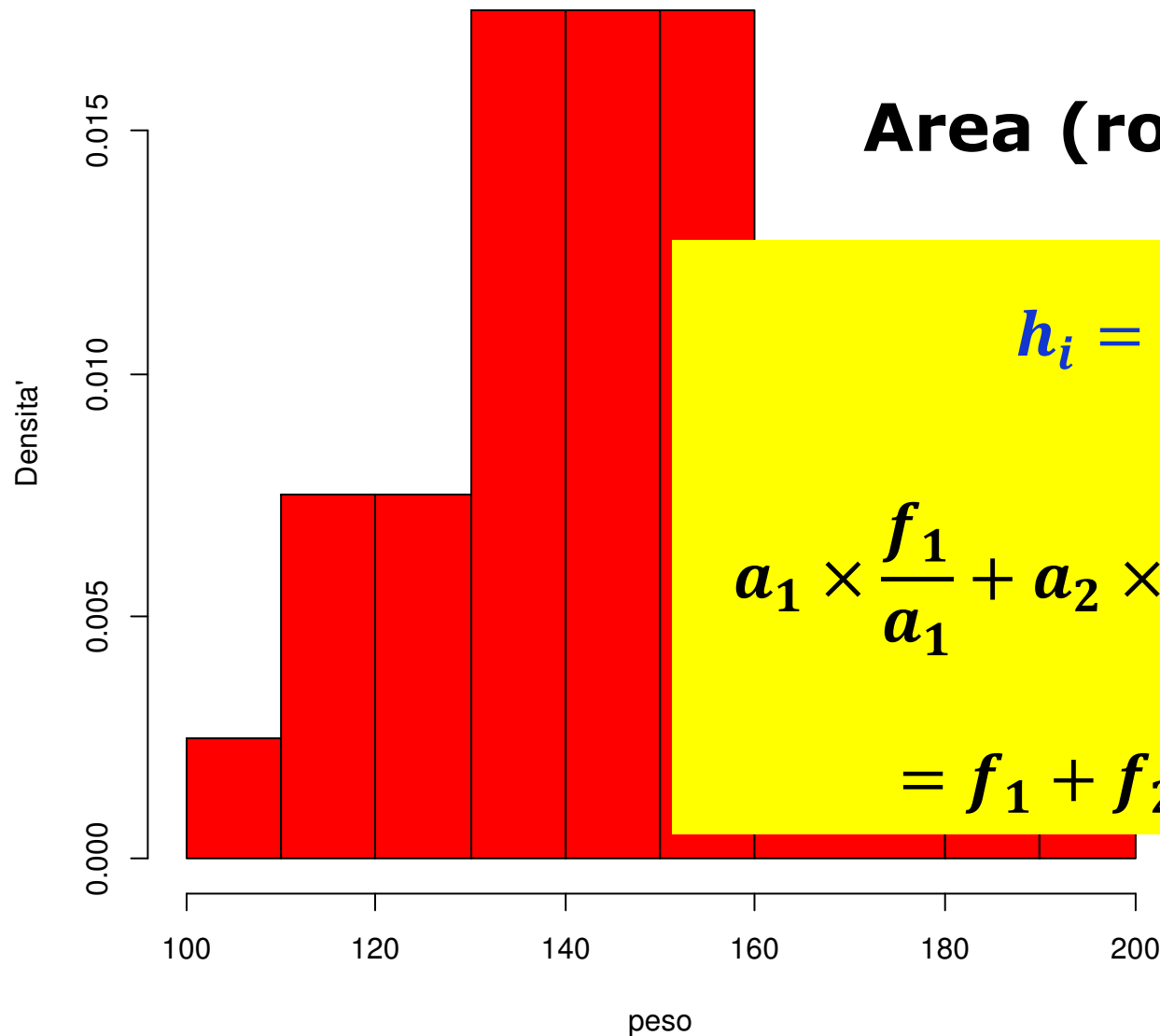
$$h_i = \frac{n_i}{a_i} = l_i \Rightarrow$$

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

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

# Istogramma

Rappresentazione  
delle frequenze  
tramite l'**area**



**Area (rossa) totale:**

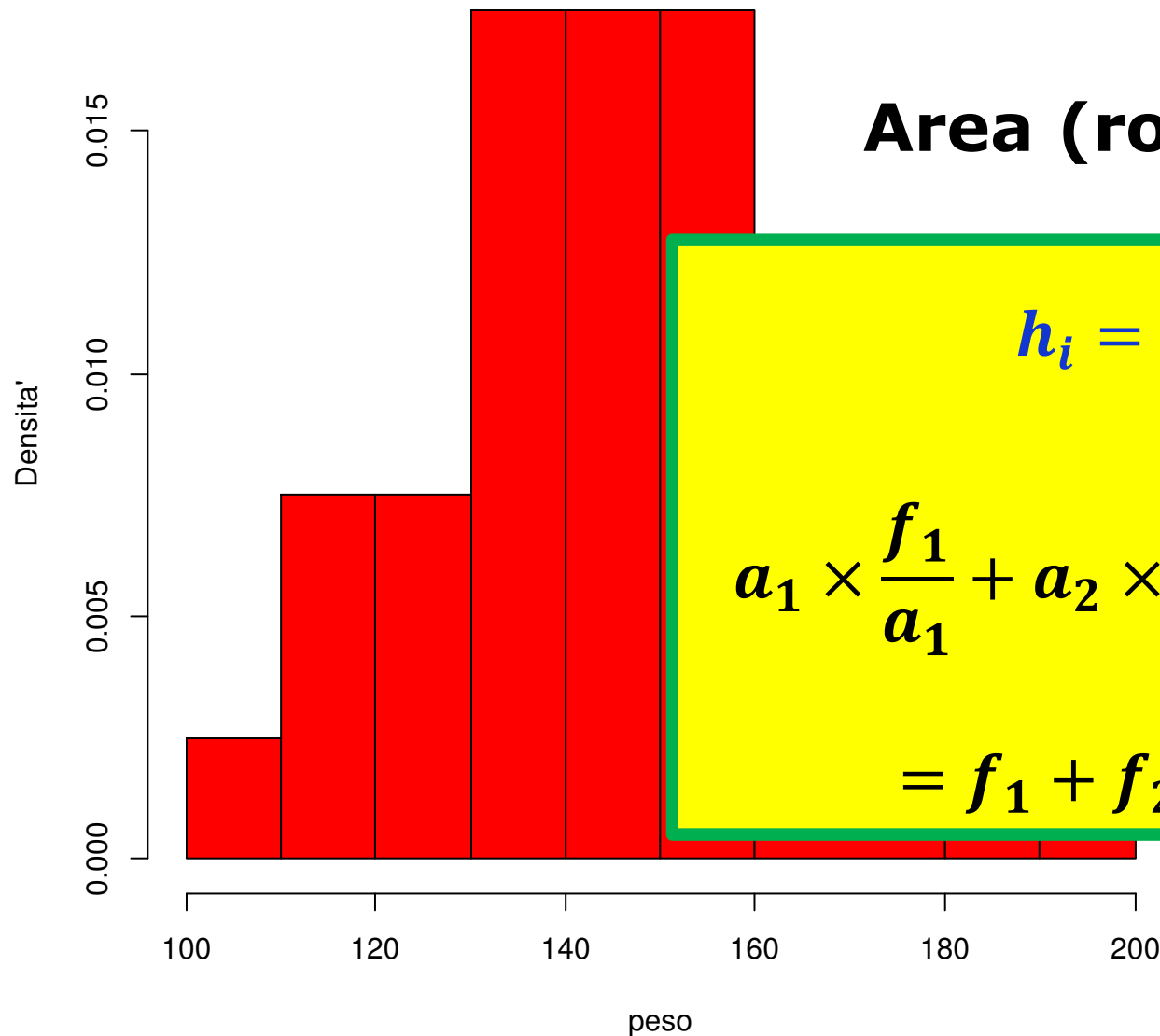
$$h_i = \frac{f_i}{a_i} = l'_i \Rightarrow$$

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

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

# Istogramma

Rappresentazione  
delle frequenze  
tramite l'**area**



**Area (rossa) totale:**

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

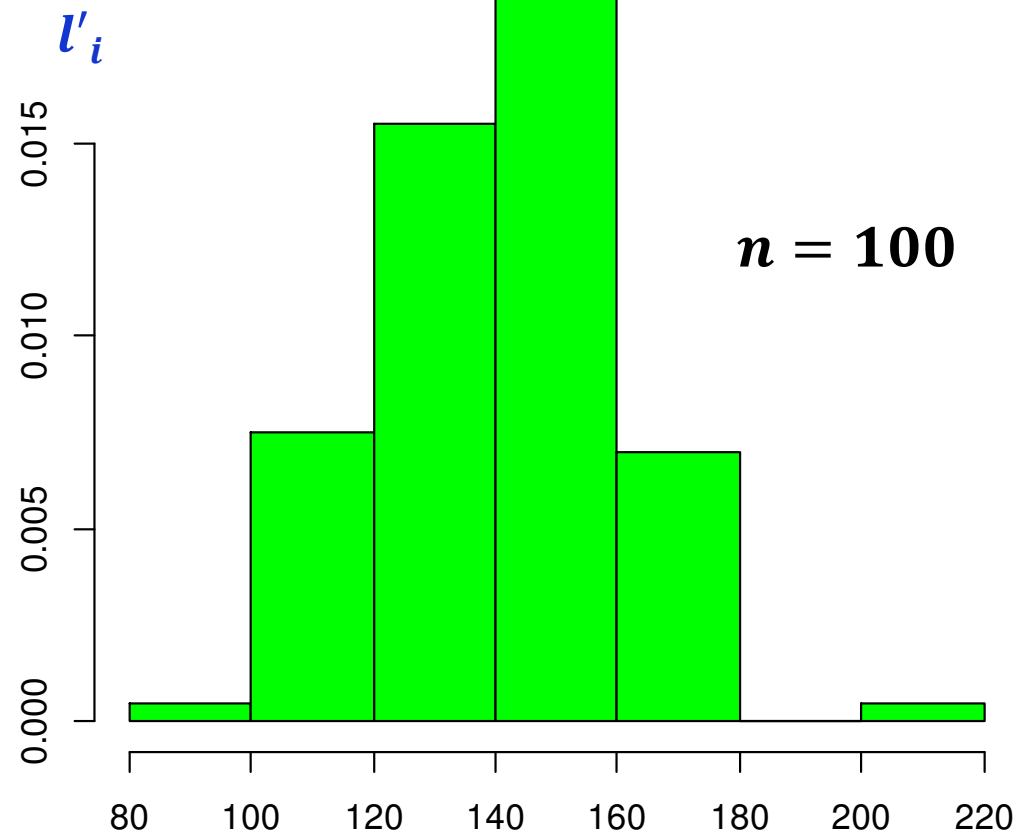
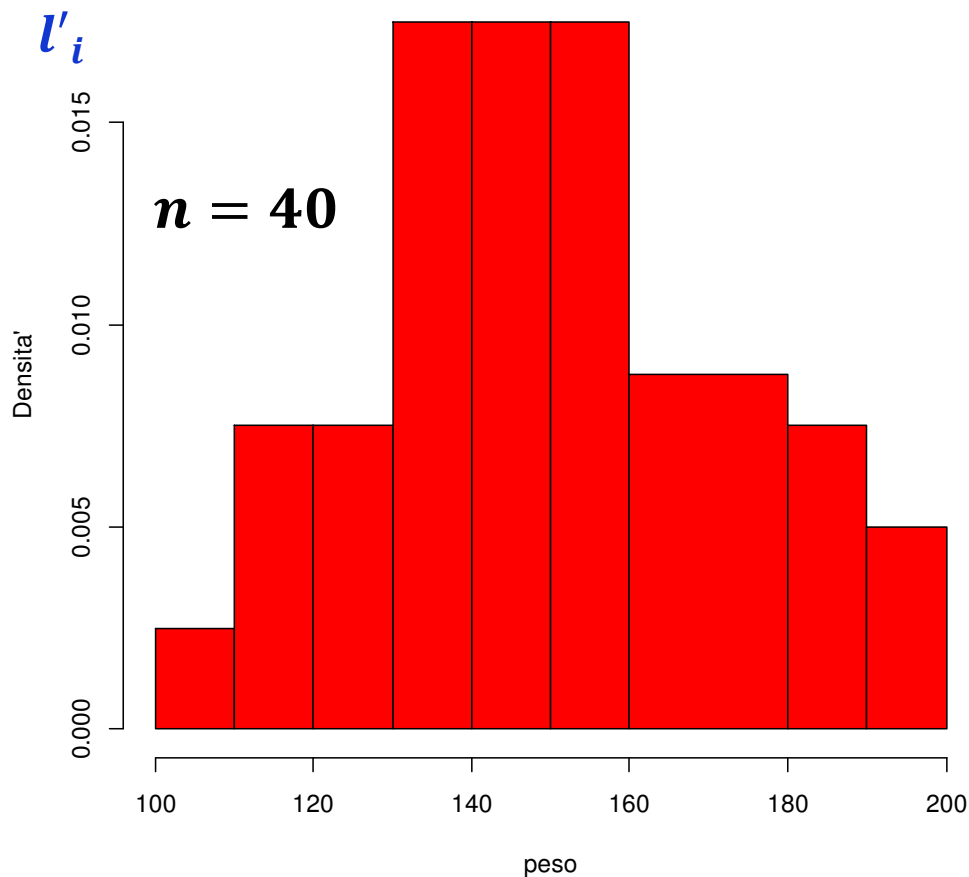
$$a_1 \times \frac{f_1}{a_1} + a_2 \times \frac{f_2}{a_2} + \dots + a_9 \times \frac{f_9}{a_9} =$$

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

# Istogramma

Rappresentazione  
delle frequenze  
tramite l'**area**

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



# Istogramma

Rappresentazione  
delle frequenze  
tramite l'**area**

**Proporzione** di u.s. con peso  $< 140$

