FISICA GENERALE

— Parte 3 —

Statistica

Link moodle: https://moodle2.units.it/course/view.php?id=16681

Codice Teams del corso: gz0wuf4

Programma delle lezioni

Lezione 1: Introduzione al corso, ai concetti generali e all'analisi degli errori; stima delle incertezze

Lezione 2: Errori casuali e sistematici, rappresentazione degli errori, cifre significative, discrepanza

Lezione 3: Errori assoluti e relativi, applicazioni particolari della propagazione degli errori, somma in quadratura

Lezione 4: Propagazione degli errori, funzioni di una o più variabili, formula generale; esempi ed esercizi

Lezione 5: Analisi statistica degli errori casuali; media, deviazione standard; errori sistematici

Lezione 6: Rappresentazione dei dati; istogrammi e distribuzioni, distribuzione limite

Lezione 7: Distribuzione normale o gaussiana (prima parte); livelli di confidenza

Lezione 8: Distribuzione gaussiana (seconda parte) e principio di massima verosimiglianza; rigetto dei dati

Lezione 9: Distribuzione binomiale

Lezione 10: Distribuzione di Poisson

Lezione 11: Metodo dei minimi quadrati; ripasso di eventuali argomenti a richiesta; esercizi

Distribuzioni ed istogrammi

Strumenti utili se si cerca un modo conveniente di trattare e mostrare i molti valori che si possono ottenere quando si effettuano o si ripetono misure.

Rappresentazione grafica dei dati

La statistica è fatta di numeri che misurano grandezze ed indicano frequenze di fenomeni.

Difficilmente i numeri offrono una immediata percezione della grandezza e della variabilità di un fenomeno.

Per questo motivo la statistica si avvale della **rappresentazione grafica** che consente una **più immediata** e facile **comprensione** delle caratteristiche dei fenomeni esposti.

I grafici non contengono dati diversi rispetto alle tabelle da cui spesso si ricavano, né aggiungono informazione, in quanto sono ricavati dalle stesse. Tuttavia, le rappresentazioni grafiche consentono di cogliere con evidenza visiva aspetti fondamentali di una distribuzione di frequenza come la tendenza, la variabilità, la forma, i valori anomali.

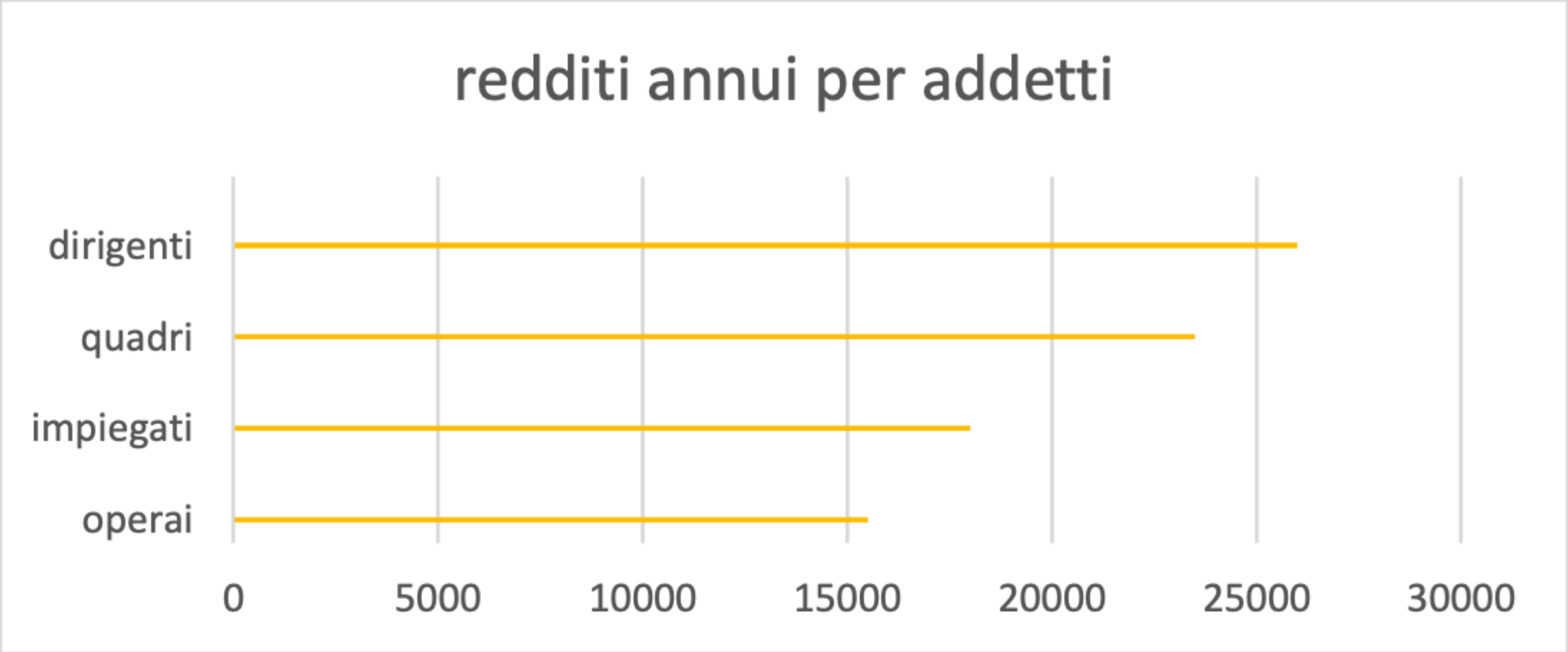
Questo perché siamo portati a dedurre informazioni dal grafico pressoché **istantaneamente**, mentre impieghiamo più tempo per leggere ad es. tutte le cifre esposte in una tabella e collegarle tra loro.

Ovviamente, metodo grafico e metodo numerico sono tra loro complementari e quindi devono essere usati congiuntamente per consentire a chi legge di **interpretare correttamente i dati**.

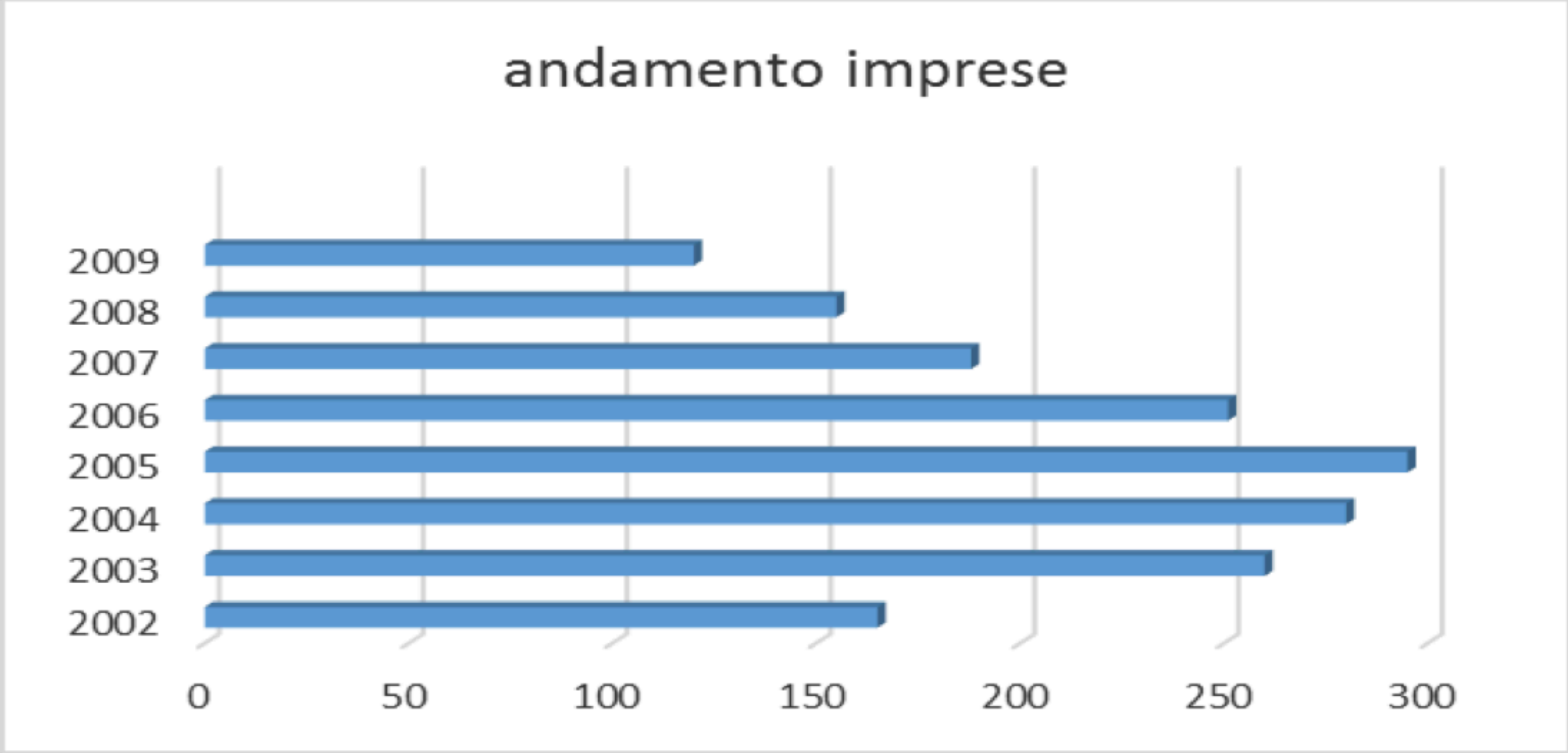
Rappresentazione grafica

Descrizione qualitativa tramite **diagrammi a barre**

addetti	redditi annui
operai	15.500
impiegati	18.000
quadri	23.500
dirigenti	26.000

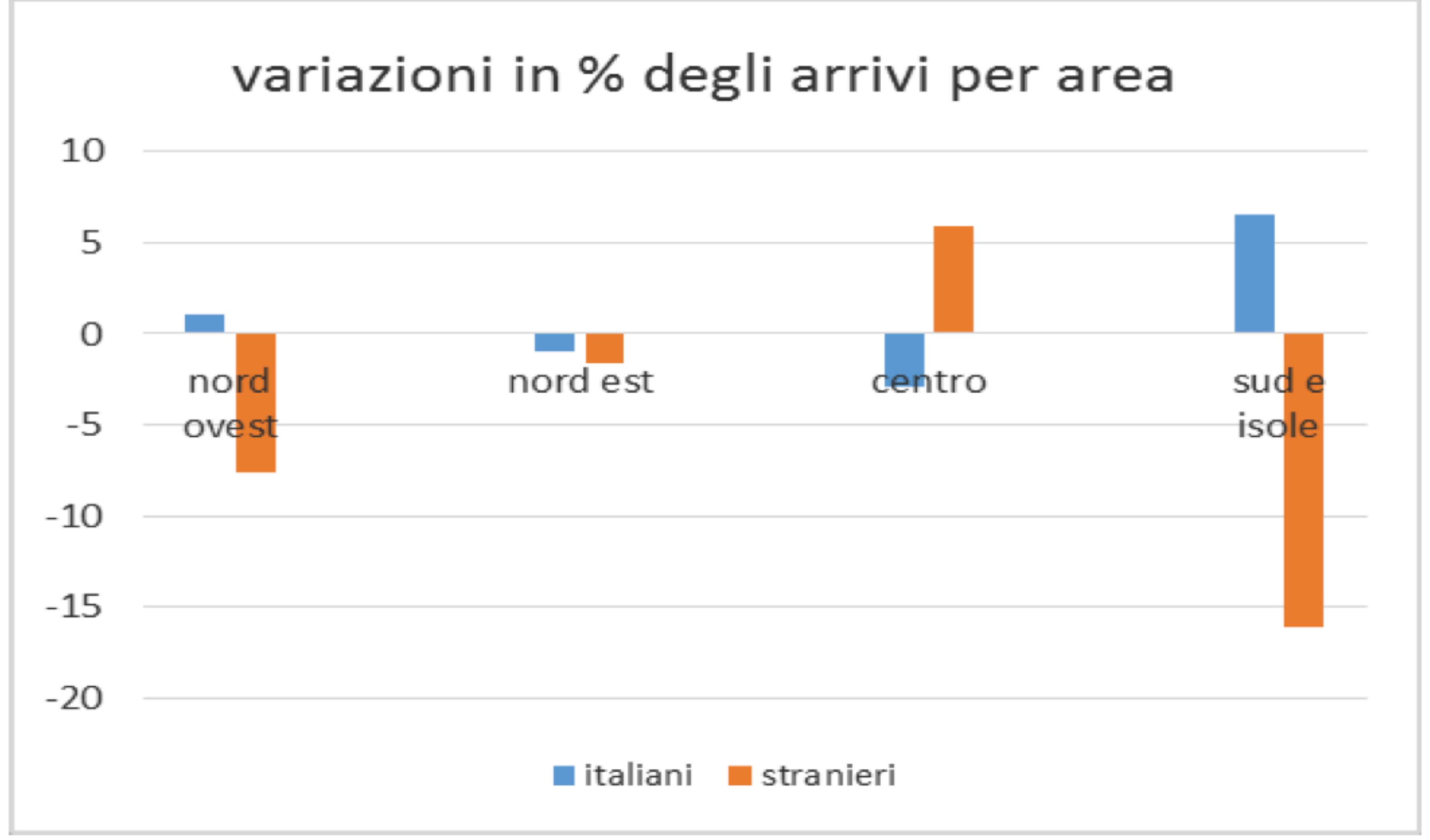


anni	imprese
2002	165
2003	260
2004	280
2005	295
2006	251
2007	188
2008	155
2009	120



Variazioni % arrivi

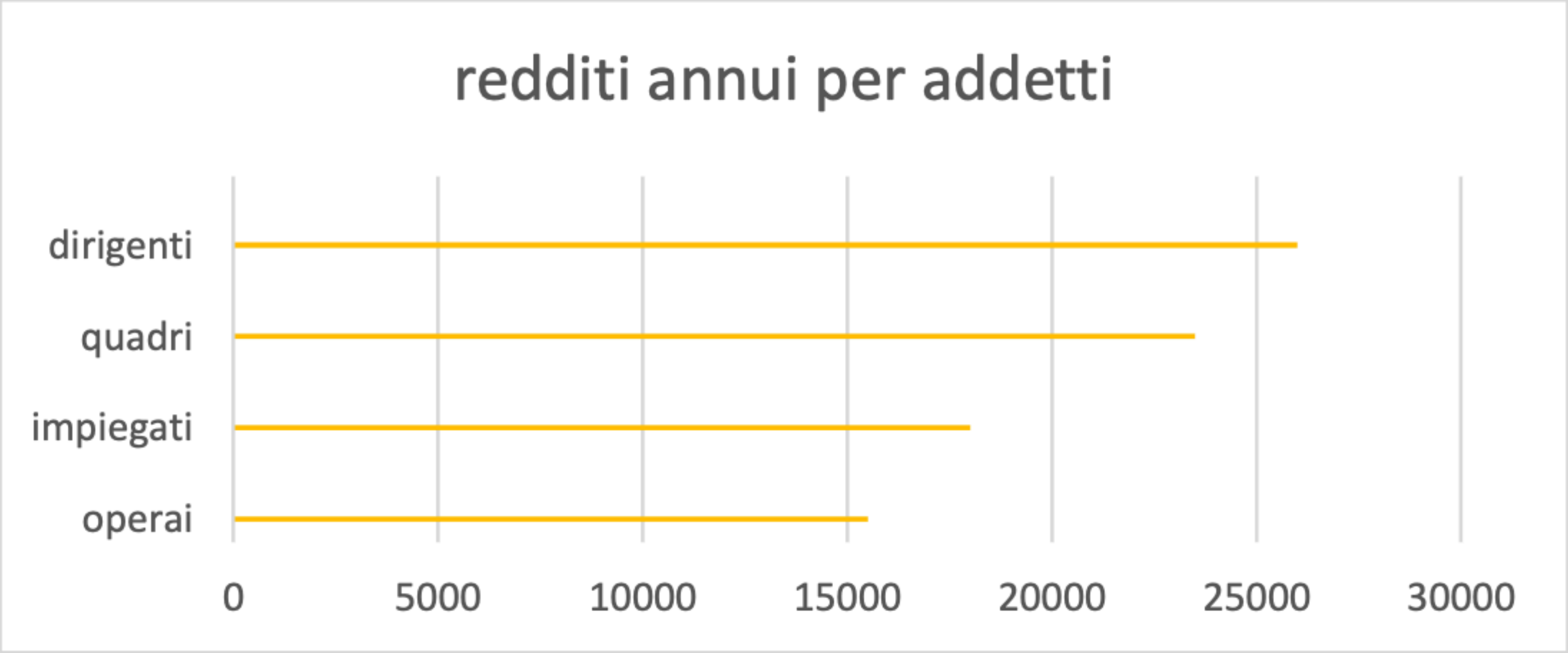
zona geog	italiani	stranieri
nord ovest	1,1	-7,6
nord est	-1,0	-1,6
centro	-2,9	5,9
sud e isole	6,5	-16,1



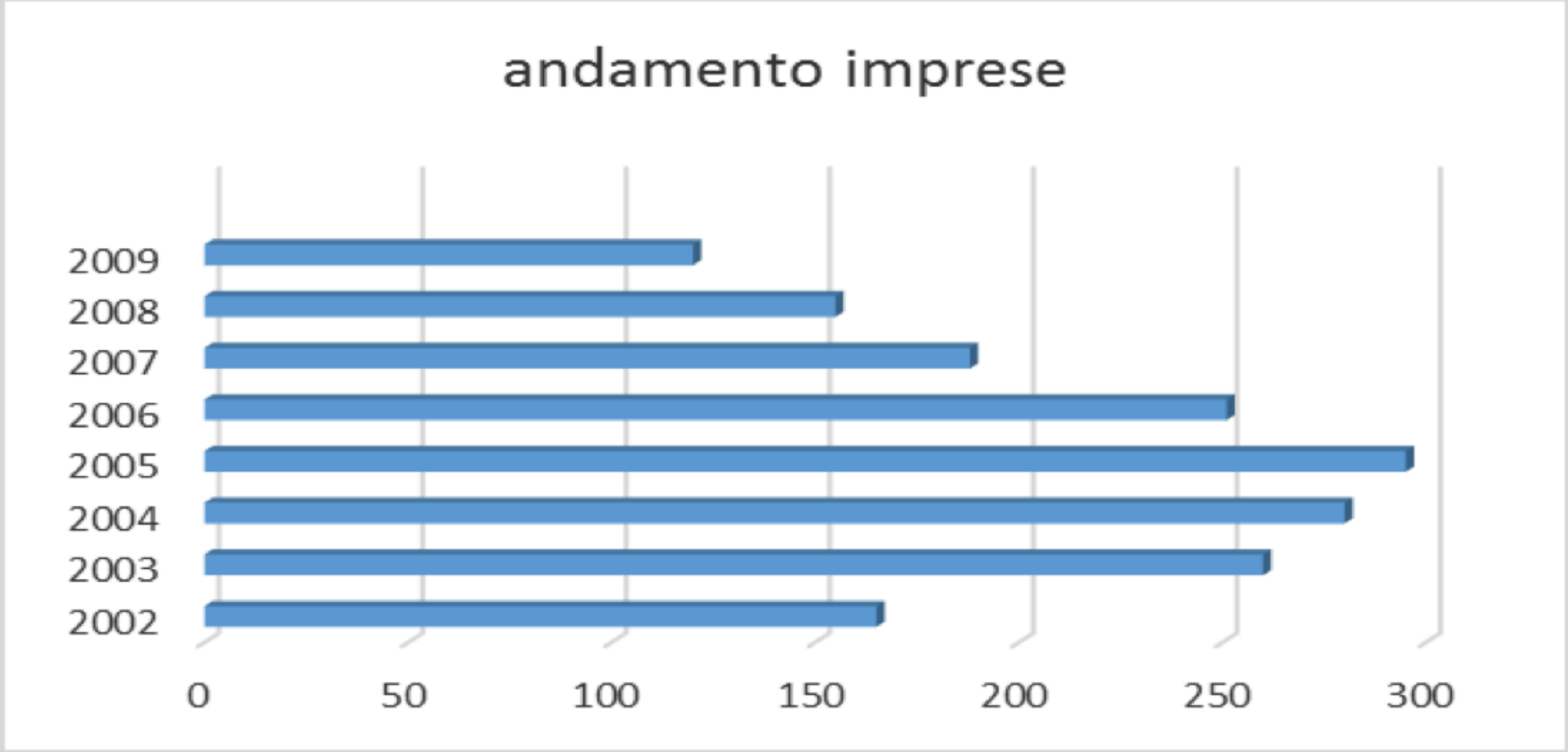
Rappresentazione grafica

Descrizione qualitativa tramite **diagrammi a barre**

addetti	redditi annui
operai	15.500
impiegati	18.000
quadri	23.500
dirigenti	26.000

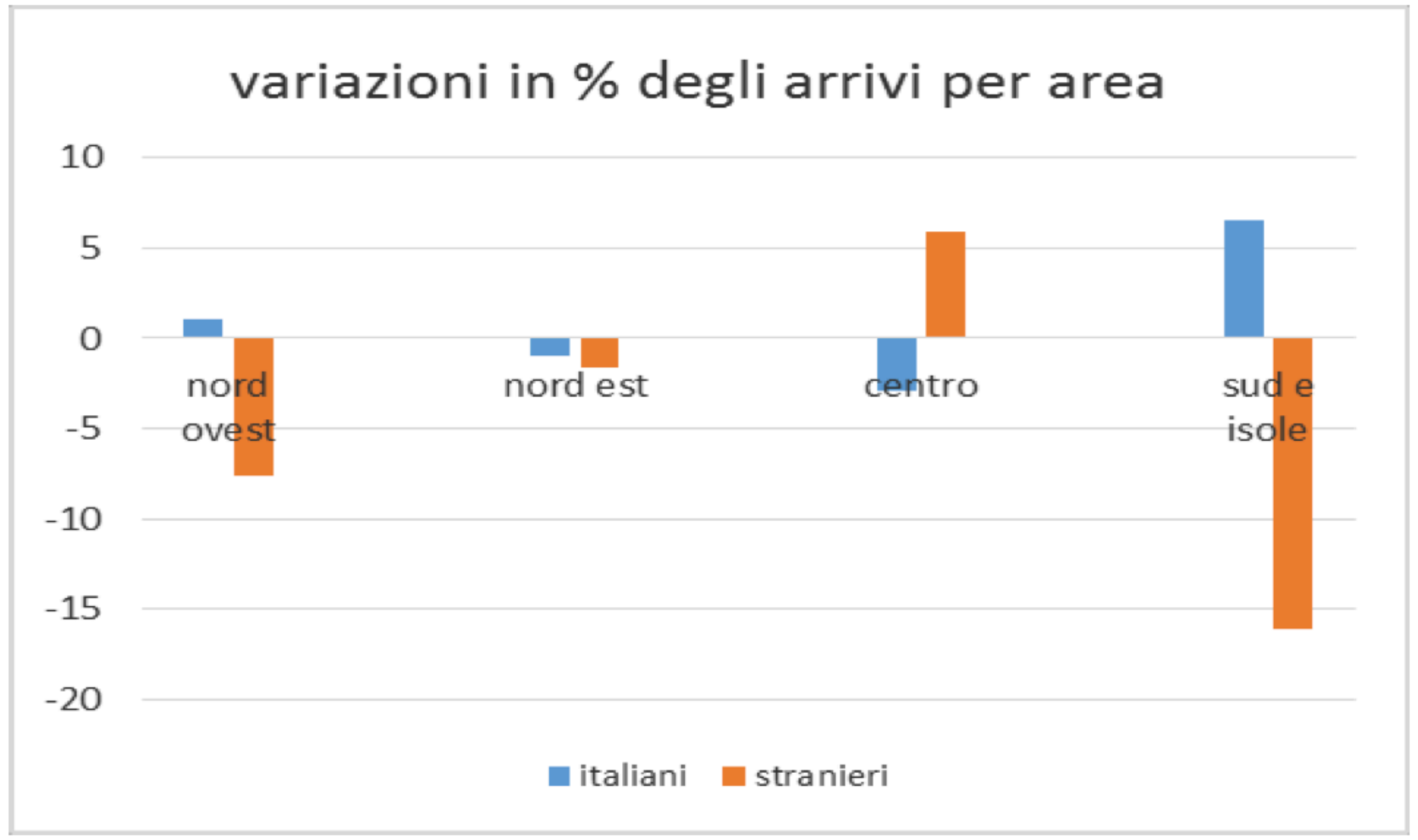


anni	imprese
2002	165
2003	260
2004	280
2005	295
2006	251
2007	188
2008	155
2009	120



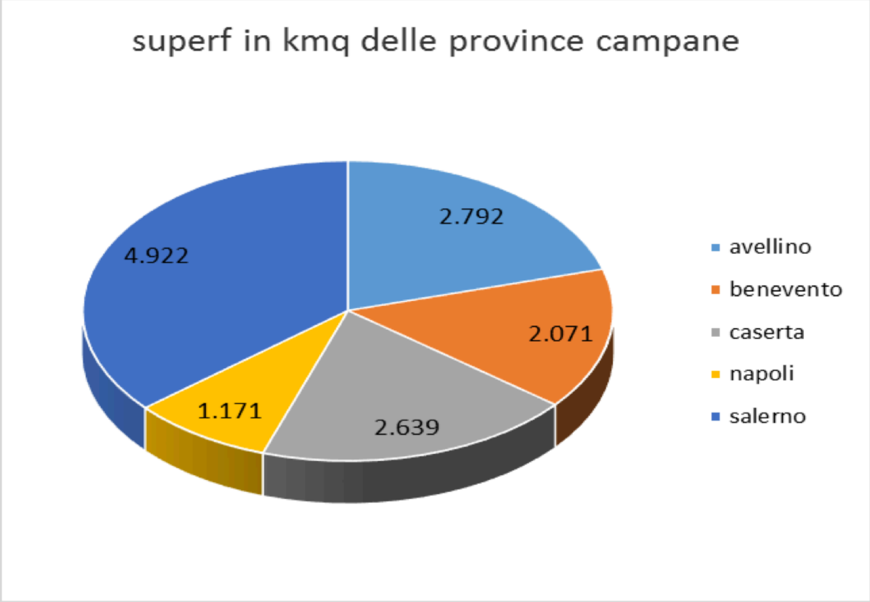
Variazioni % arrivi

zona geog	italiani	stranieri
nord ovest	1,1	-7,6
nord est	-1,0	-1,6
centro	-2,9	5,9
sud e isole	6,5	-16,1

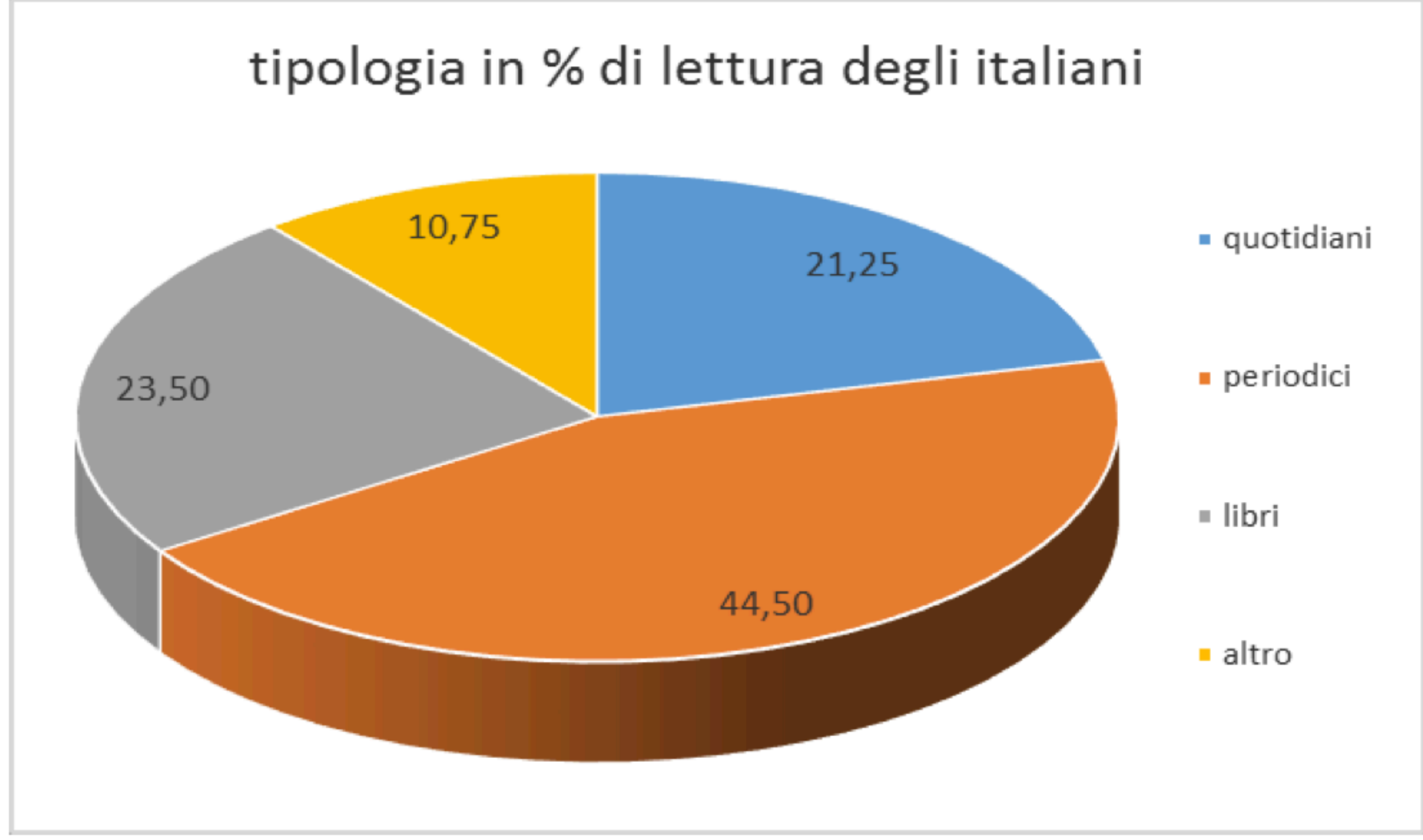


Areogrammi o grafici a torta

province	superf in kmq
avellino	2.792
benevento	2.071
caserta	2.639
napoli	1.171
salerno	4.922
totale	13.595



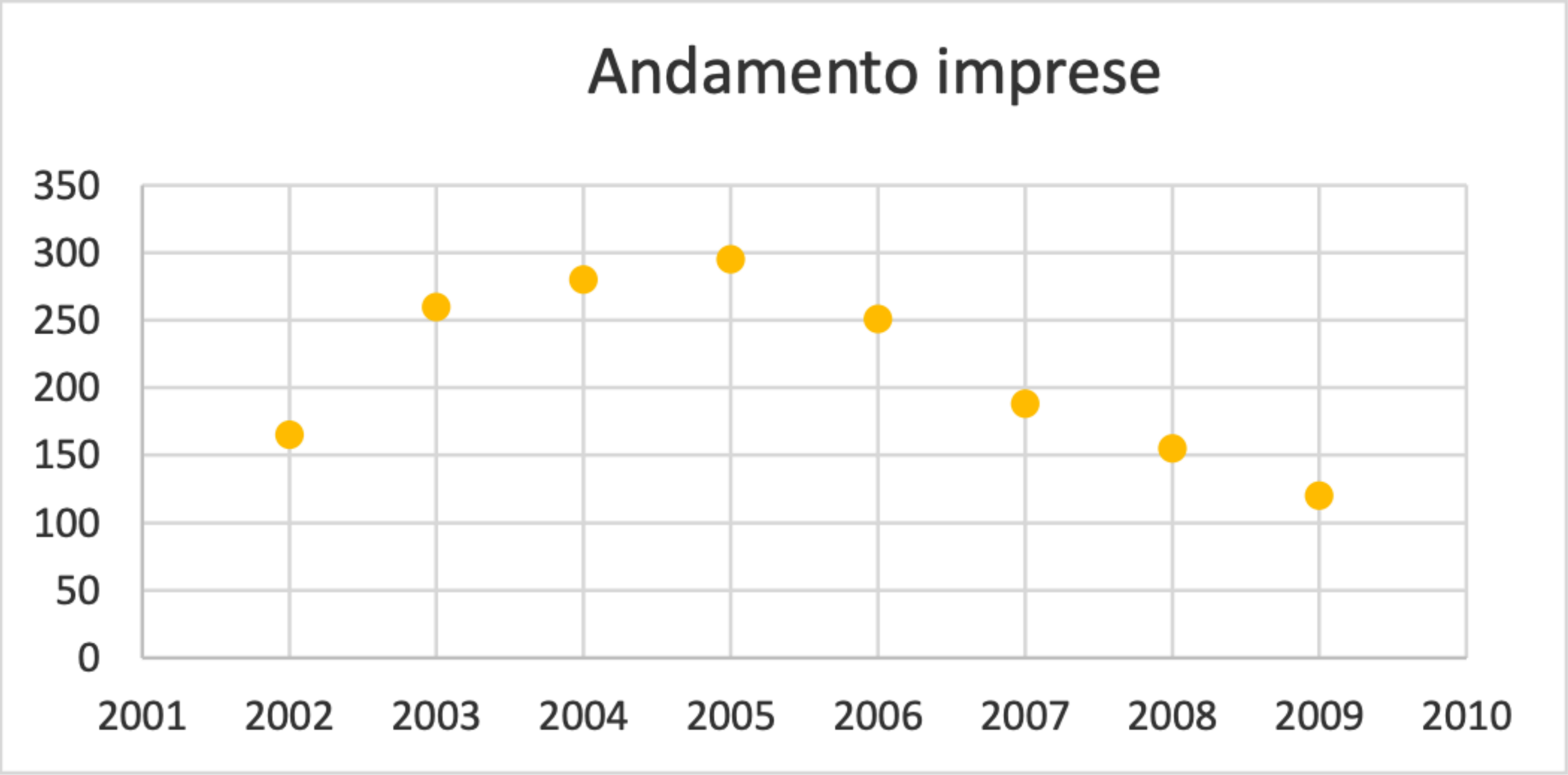
tipi di lettura	%
quotidiani	21,25
periodici	44,50
libri	23,50
altro	10,75
totale	100,00



Rappresentazione grafica

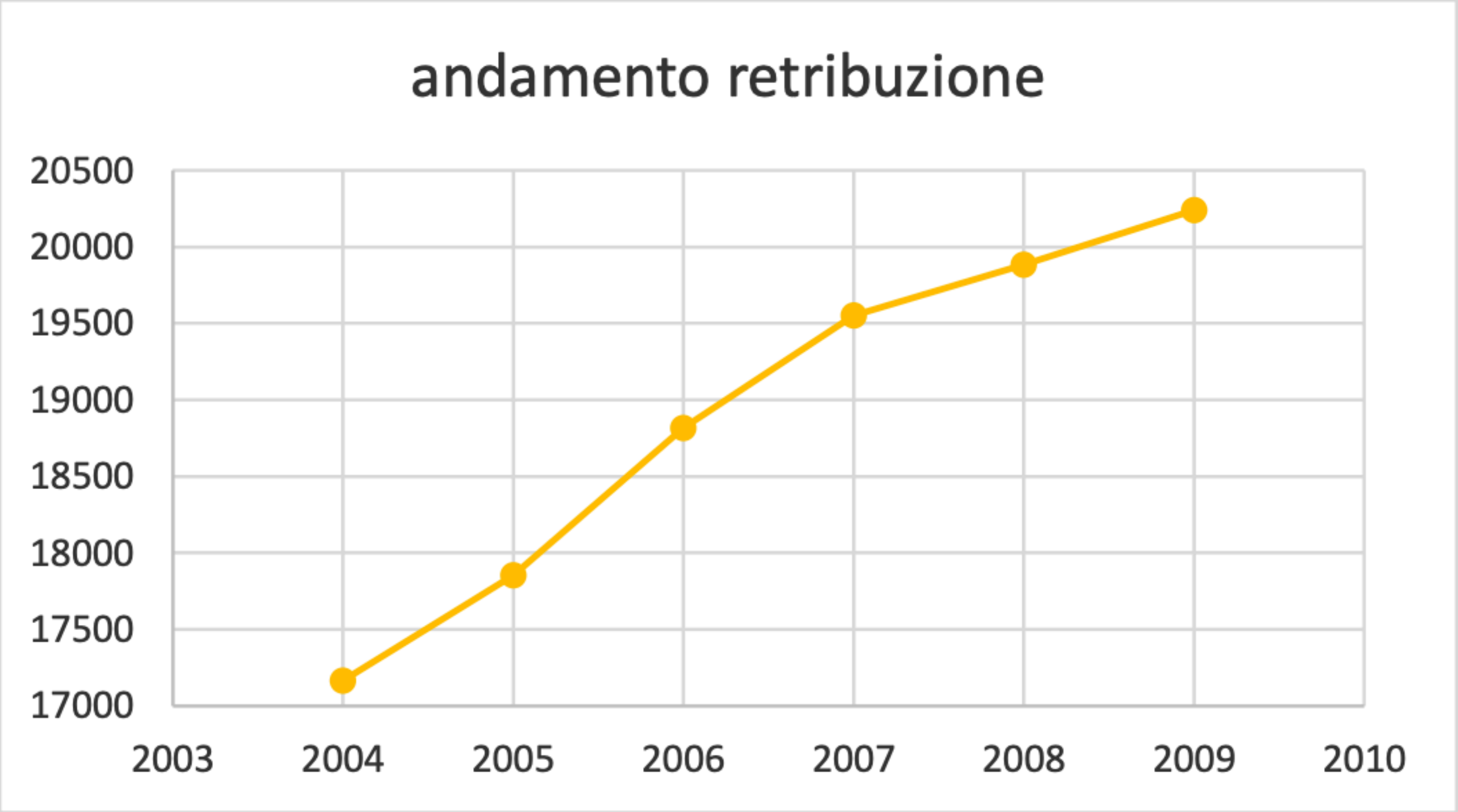
Diagrammi in coordinate cartesiane

anni	imprese
2002	165
2003	260
2004	280
2005	295
2006	251
2007	188
2008	155
2009	120

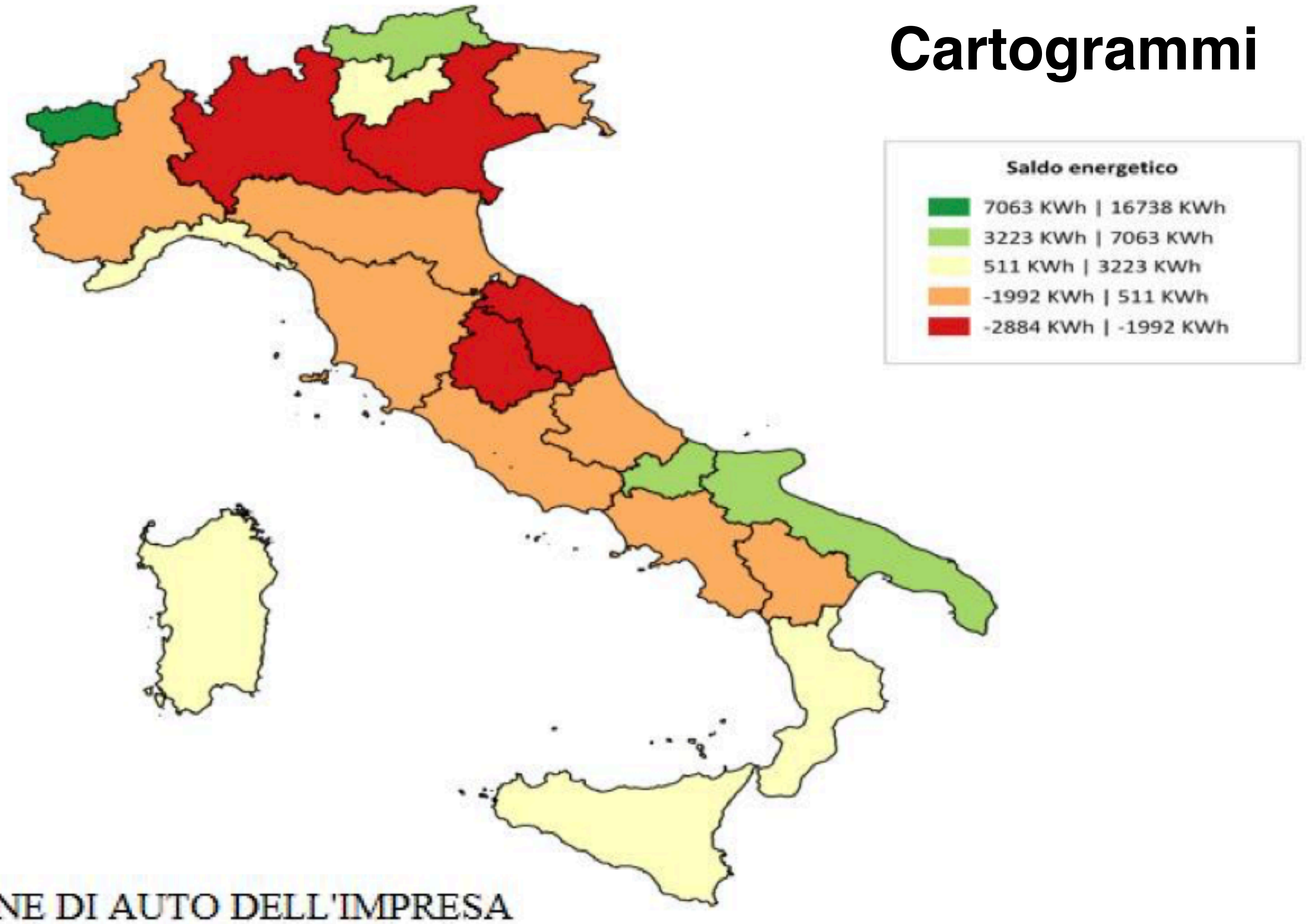


Retribuzione annua di una data categoria di lavoratori

anno	retribuzione
2004	17.166
2005	17.853
2006	18.818
2007	19.552
2008	19.884
2009	20.242

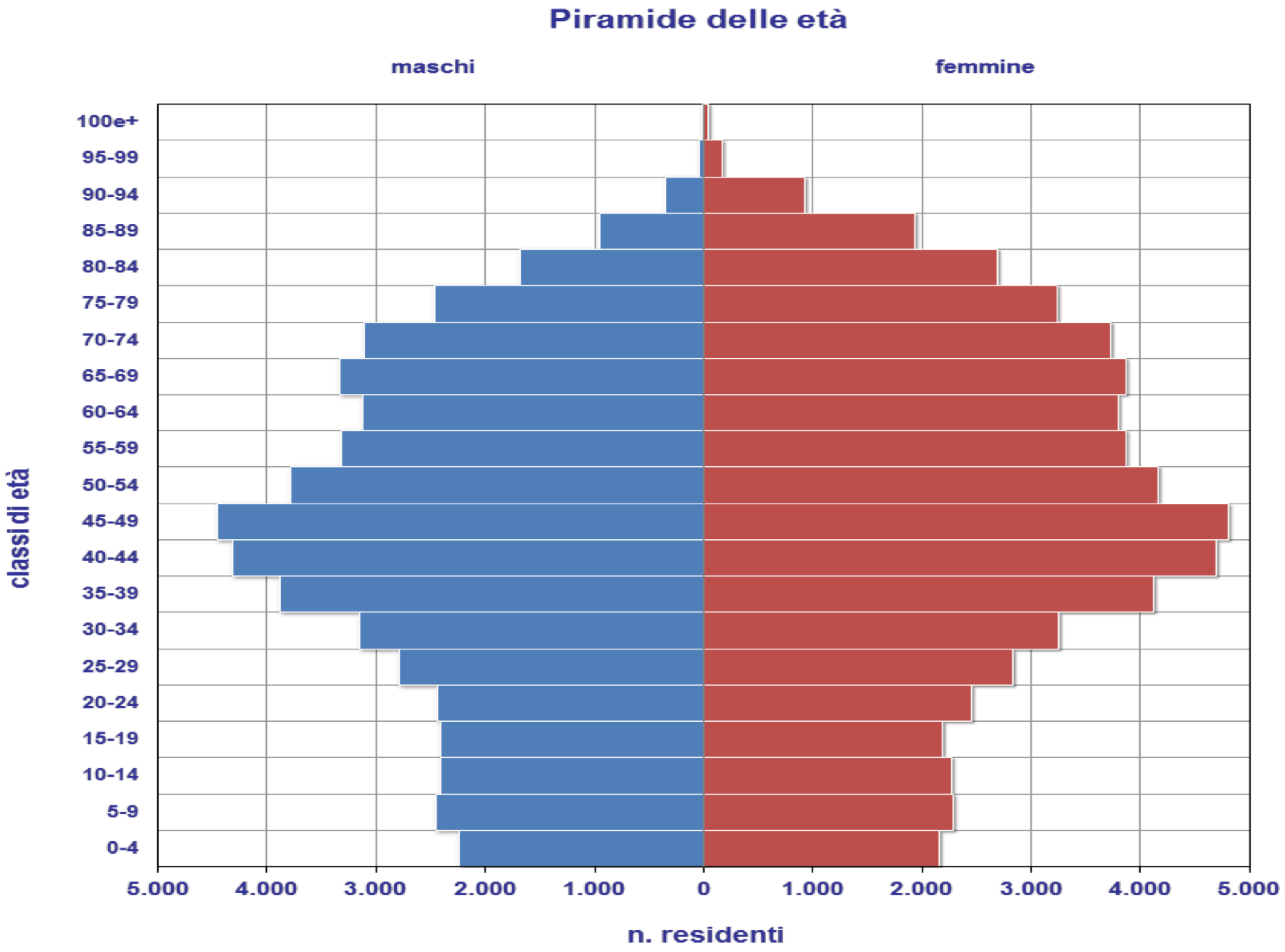


Cartogrammi



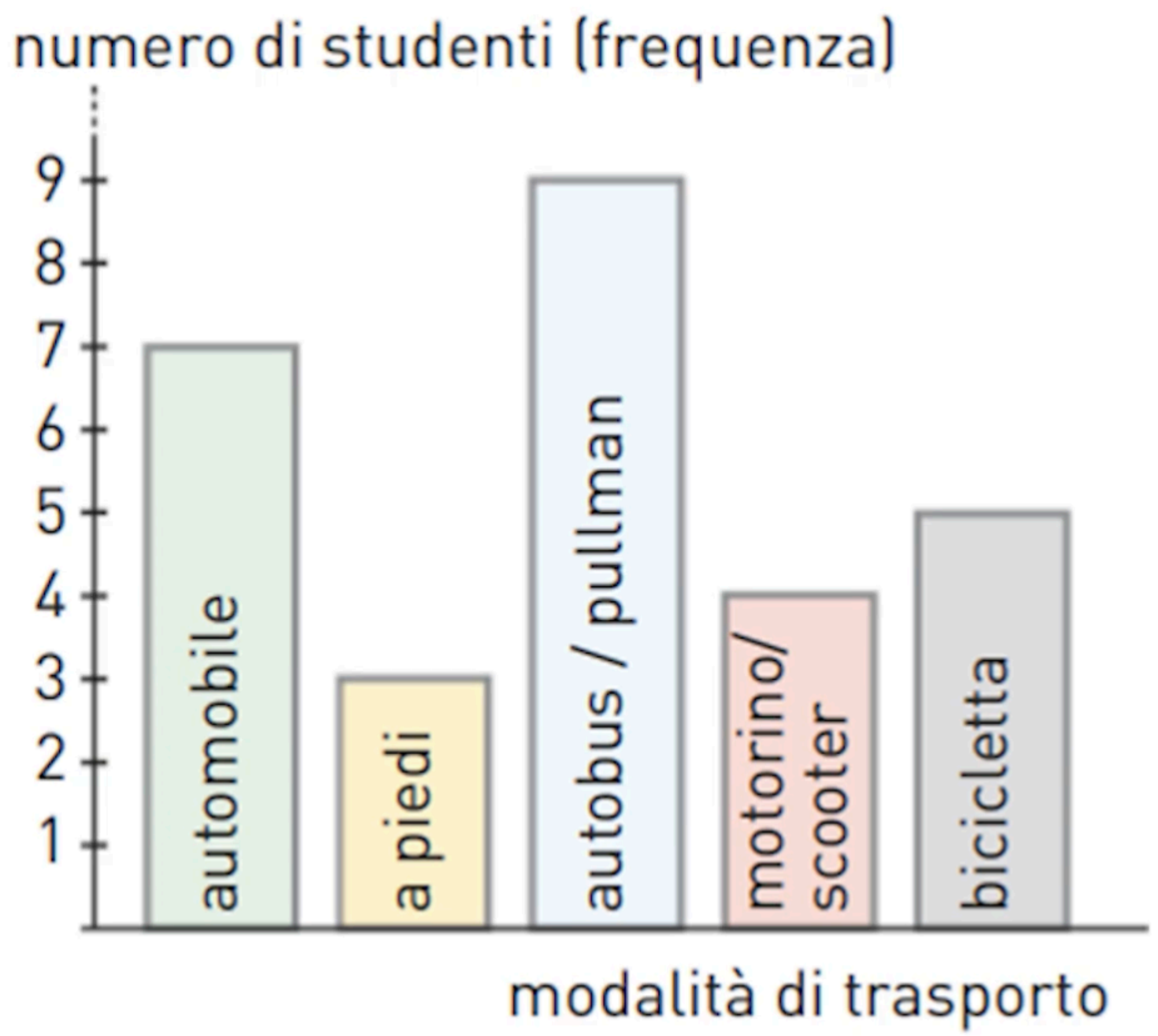
Ideogrammi

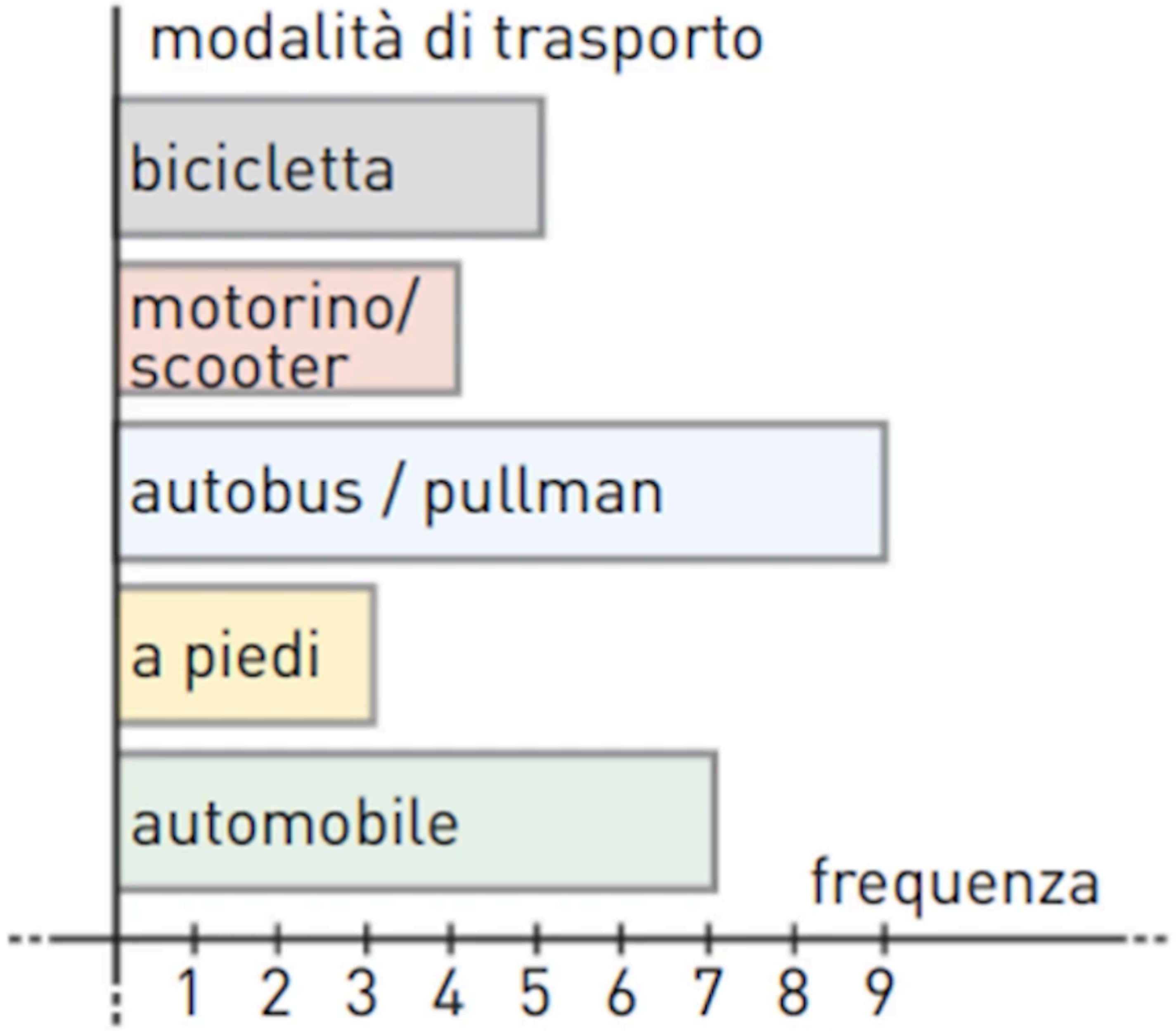
Rappresentazione grafica



Rappresentazione grafica

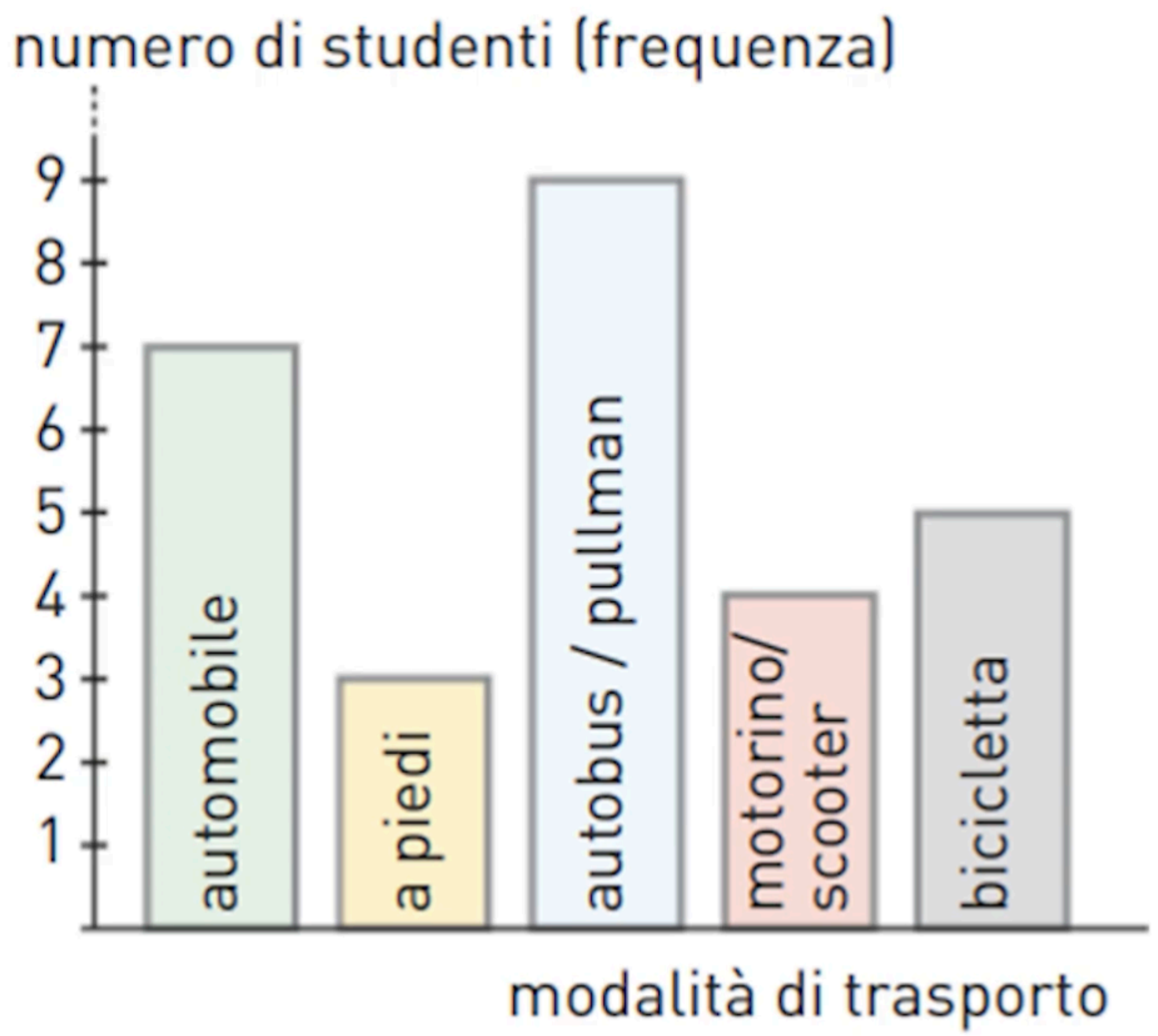
Modalità	Frequenza
automobile	7
a piedi	3
autobus/pullman	9
motorino/scooter	4
bicicletta	5
totale delle unità statistiche	*28*

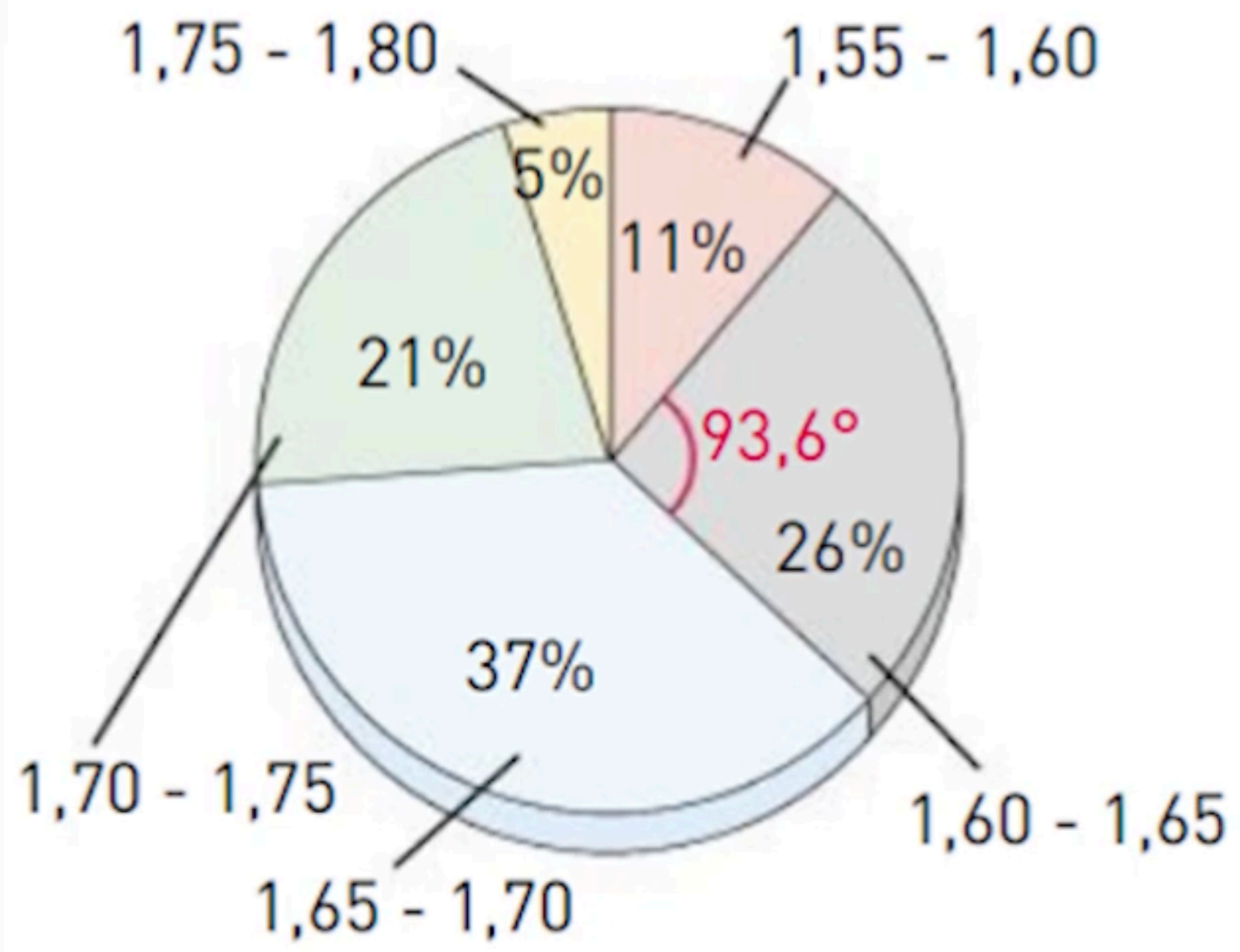


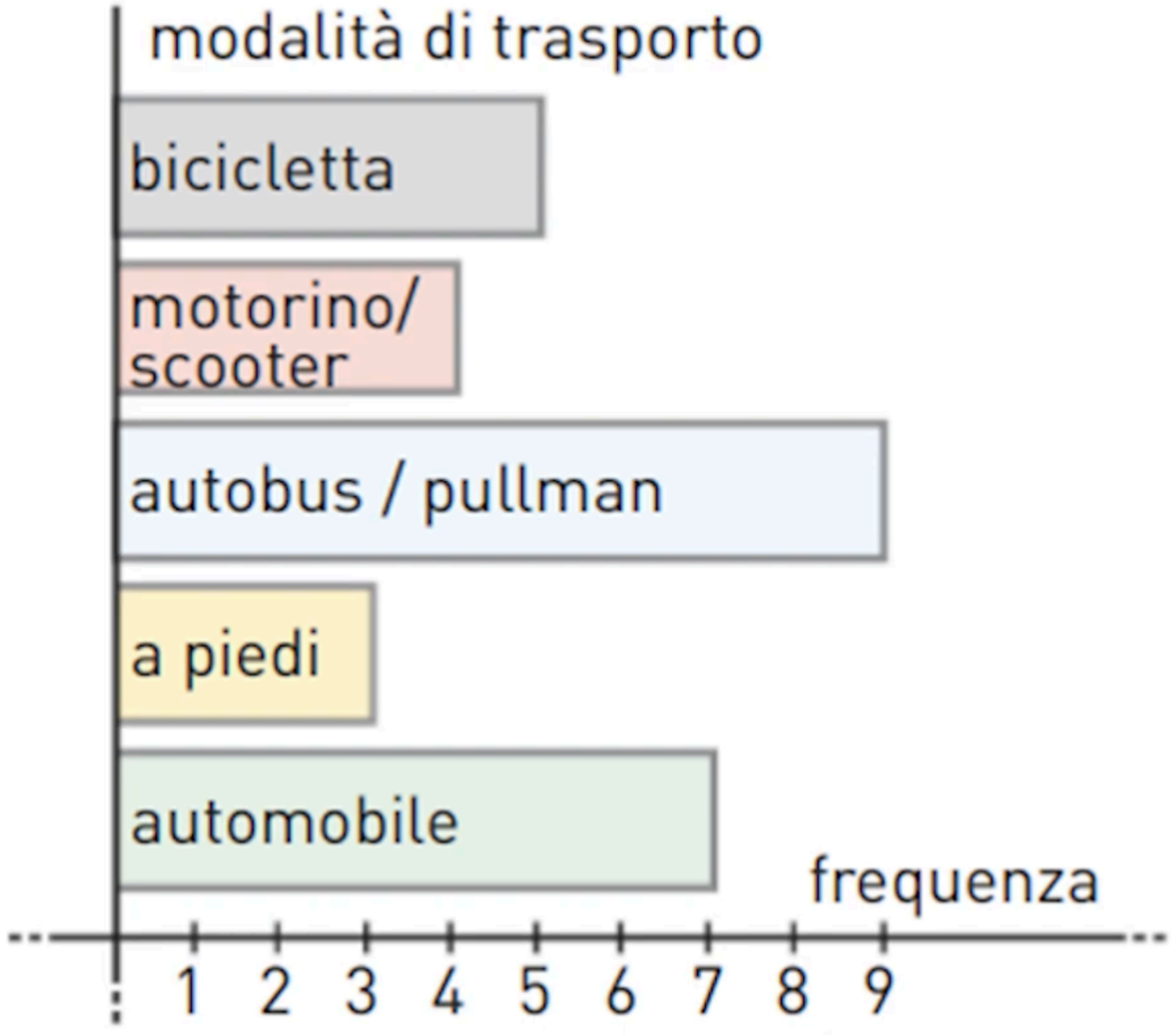


Rappresentazione grafica

Modalità	**Frequenza**
automobile	7
a piedi	3
autobus/pullman	9
motorino/scooter	4
bicicletta	5
totale delle unità statistiche	*28*

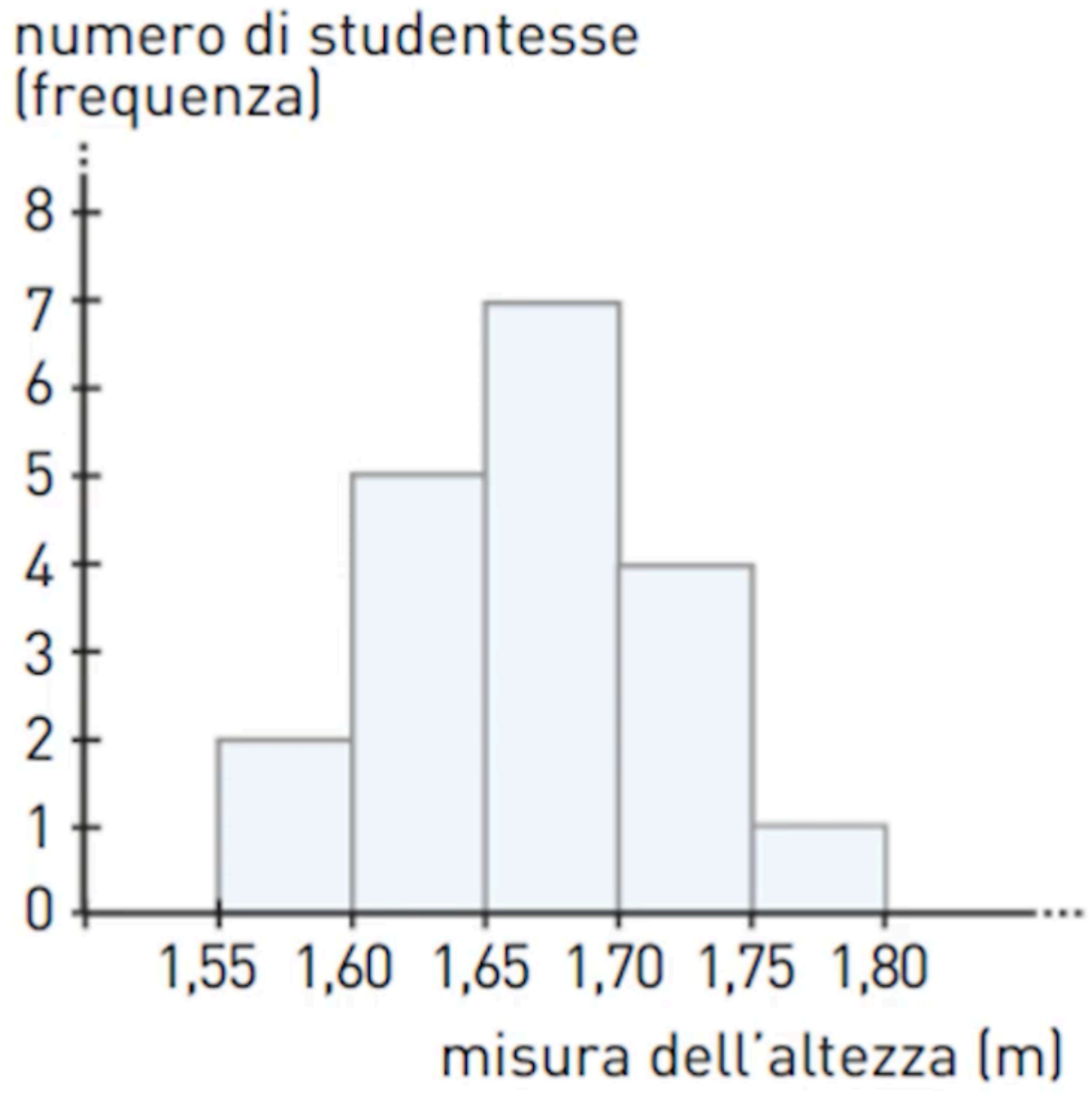






Classe	**Frequenza**	**Frequenza relativa percentuale**
1,55-1,60	2	11%
1,60-1,65	5	26%
1,65-1,70	7	37%
1,70-1,75	4	21%
1,75-1,80	1	5%

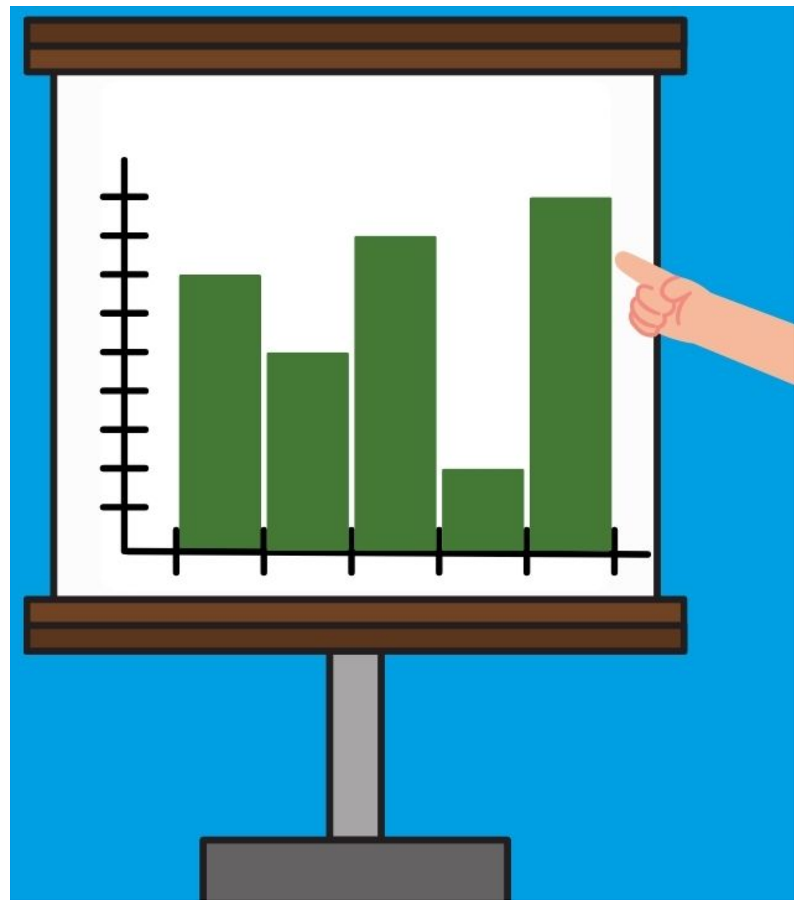
$$x : 360^\circ = f_{\%} : 100 \rightarrow x = \frac{360^\circ \cdot f_{\%}}{100} \qquad x = \frac{360^\circ \cdot 26}{100} = 93{,}6^\circ$$



La statistica raccoglie, analizza e presenta grandi quantità di dati.

Analisi statistica di misure ripetute → modo conveniente di **mostrare** i valori ottenuti → formalismo corretto per **interpretare** i dati

Dal campionamento dei dati → all'interpretazione dei risultati



Un modo per evidenziare graficamente
una distribuzione di misure
è quindi tramite **istogrammi**.

Istogrammi

In un esperimento, bisogna spesso registrare e mettere in evidenza un gran numero di valori misurati prima di analizzarli.

Faccio ad esempio 10 misure di una lunghezza ed ottengo:

26, 24, 26, 28, 23, 24, 25, 24, 26, 25

Per meglio comprendere l'esito del mio esperimento, potrei riorganizzare i numeri in **ordine** crescente:

23, 24, 24, 24, 25, 25, 26, 26, 26, 28

Posso pensare di registrare i **diversi valori ottenuti**, insieme al **numero di volte** in cui ciascun numero è stato trovato.

Diversi valori x_k	23	24	25	26	27	28
Numero di volte	1	3	2	3	0	1

Istogrammi

Introducendo la seguente notazione:

x_k (con k=1, 2, 3, …) per denotare i vari diversi valori trovati — es. x_1=23, x_2=24, x_3=25, …

n_k (con k=1, 2, 3, …) per denotare il numero di volte in cui il corrispondente x_k è stato trovato — es. n_1=1, n_2=3, …

Registrando le misure come nella precedente tabella, allora è possibile riscrivere la **definizione della media**

$$\bar{x} = \frac{\sum_i x_i}{N} = \frac{23 + 24 + 24 + \ldots + 28}{10}$$

sommo su tutte le misure fatte

come

$$\bar{x} = \frac{23 + (24 \times 3) + (25 \times 2) + \ldots + 28}{10}$$

sommo i diversi valori ottenuti, moltiplicando ogni valore per l'occorrenza (somma pesata per l'occorrenza n_k)

o in generale come

$$\bar{x} = \frac{\sum_k x_k n_k}{N}$$

Istogrammi

Sommando le **occorrenze**, si ottiene il numero totale di misure fatte: $$\sum_k n_k = N$$

Invece di usare le occorrenze n_k (= il numero di volte in cui ottengo un valore), posso pensare in termini di **frazione delle N misure totali che ha dato il risultato** x_k

$$F_k = \frac{n_k}{N}$$

Le frazioni F_k specificano la distribuzione dei nostri risultati, dal momento che specificano come le nostre misure sono distribuite tra i diversi possibili valori.

È possibile riscrivere la formula per la media in termini delle frazioni F_k : $$\bar{x} = \sum_k x_k F_k$$

somma pesata dei diversi valori x_k, ciascuno pesato per F_k

Questo implica che: $$\sum_k F_k = 1$$ **condizione di normalizzazione**

ovvero che sommando le frazioni F_k di tutti i possibili risultati x_k dobbiamo ottenere 1.
Qualunque insieme di numeri la cui somma è 1 è detto essere **normalizzato**.

Istogrammi a barre

La distribuzione delle nostre misure può essere evidenziata graficamente in un **istogramma a barre**:

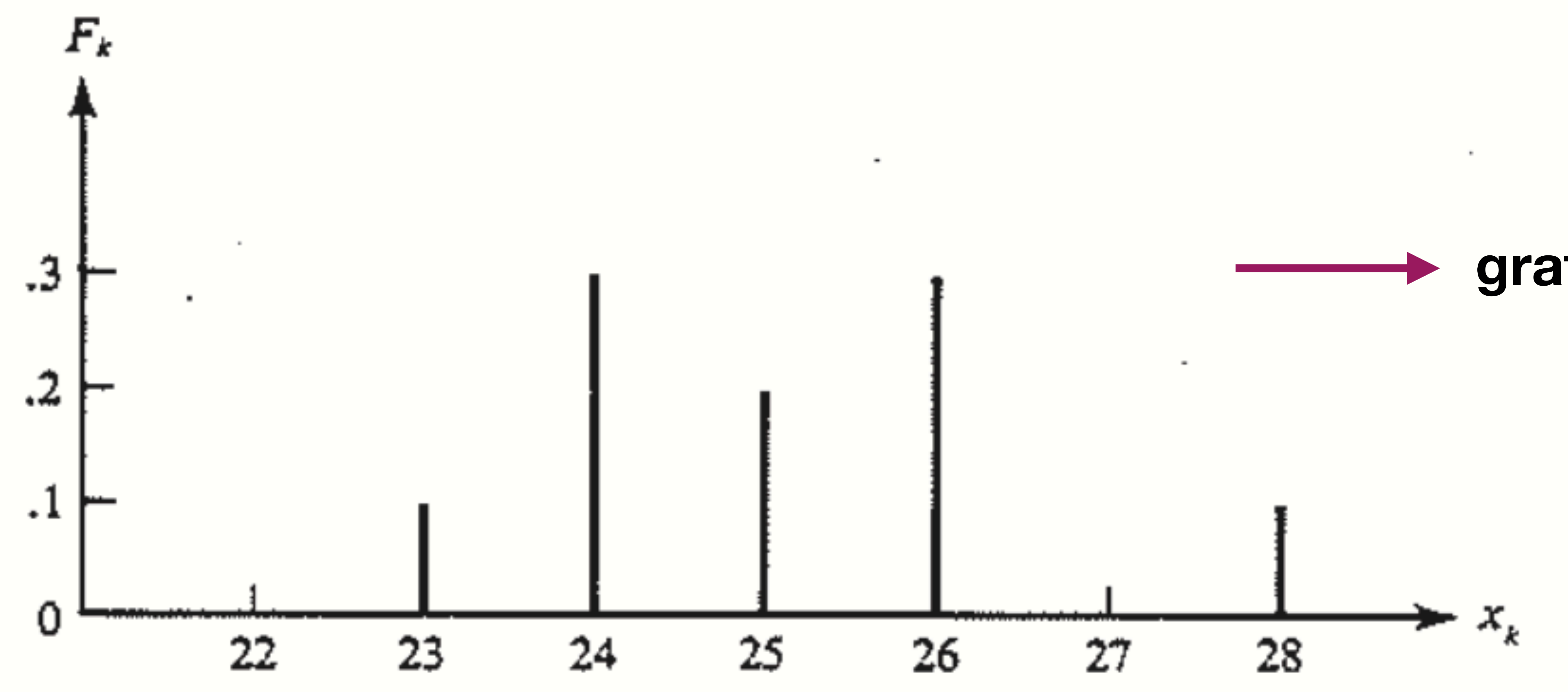


Figura 5.1. *Istogramma per le dieci misure di una lunghezza x. L'asse verticale mostra la frazione di volte F_k in cui ciascun valore x_k è stato osservato.*

grafico di F_k in funzione di x_k

l'altezza della barra verticale disegnata sopra x_k indica la frazione di volte F_k in cui ciascun x_k è stato ottenuto

Istogrammi a barre sono appropriati quando i valori di x_k sono ordinatamente spaziati, con valori interi.

La maggior parte delle misure, però, non forniscono risultati precisi interi perché le grandezze fisiche hanno spesso un intervallo continuo di valori possibili.

Istogrammi a intervalli

La maggior parte delle misure, però, non forniscono risultati precisi interi perché le grandezze fisiche hanno spesso un intervallo continuo di valori possibili.
Ad esempio 10 misure di una lunghezza mi danno:

26.4, 23.9, 25.1, 24.6, 22.7, 23.8, 25.1, 23.9, 25.3, 25.4

Un istogramma a barre di questi 10 valori consisterebbe di 10 barre separate, tutte della stessa altezza, e fornirebbe poca informazione.

Date misure come sopra, **è conveniente dividere la serie di valori in un numero opportuno di intervalli, e poi contare quanti valori** cadono **in ciascun intervallo**.

Tabella 5.2.

Intervallo	22-23	23-24	24-25	25-26	26-27	27-28
Eventi per intervallo	1	3	1	4	1	0

Istogrammi a intervalli

10 misure di una lunghezza mi danno: **26.4, 23.9, 25.1, 24.6, 22.7, 23.8, 25.1, 23.9, 25.3, 25.4**

Intervallo	22-23	23-24	24-25	25-26	26-27	27-28
Eventi per intervallo	1	3	1	4	1	0

Divido la serie di valori in un numero opportuno di intervalli

Conto quanti valori cadono in ciascun intervallo

Se una misura cade esattamente sul limite tra due intervalli, bisogna decidere dove porla (tutta nel bin inf, sup, oppure metà e metà)

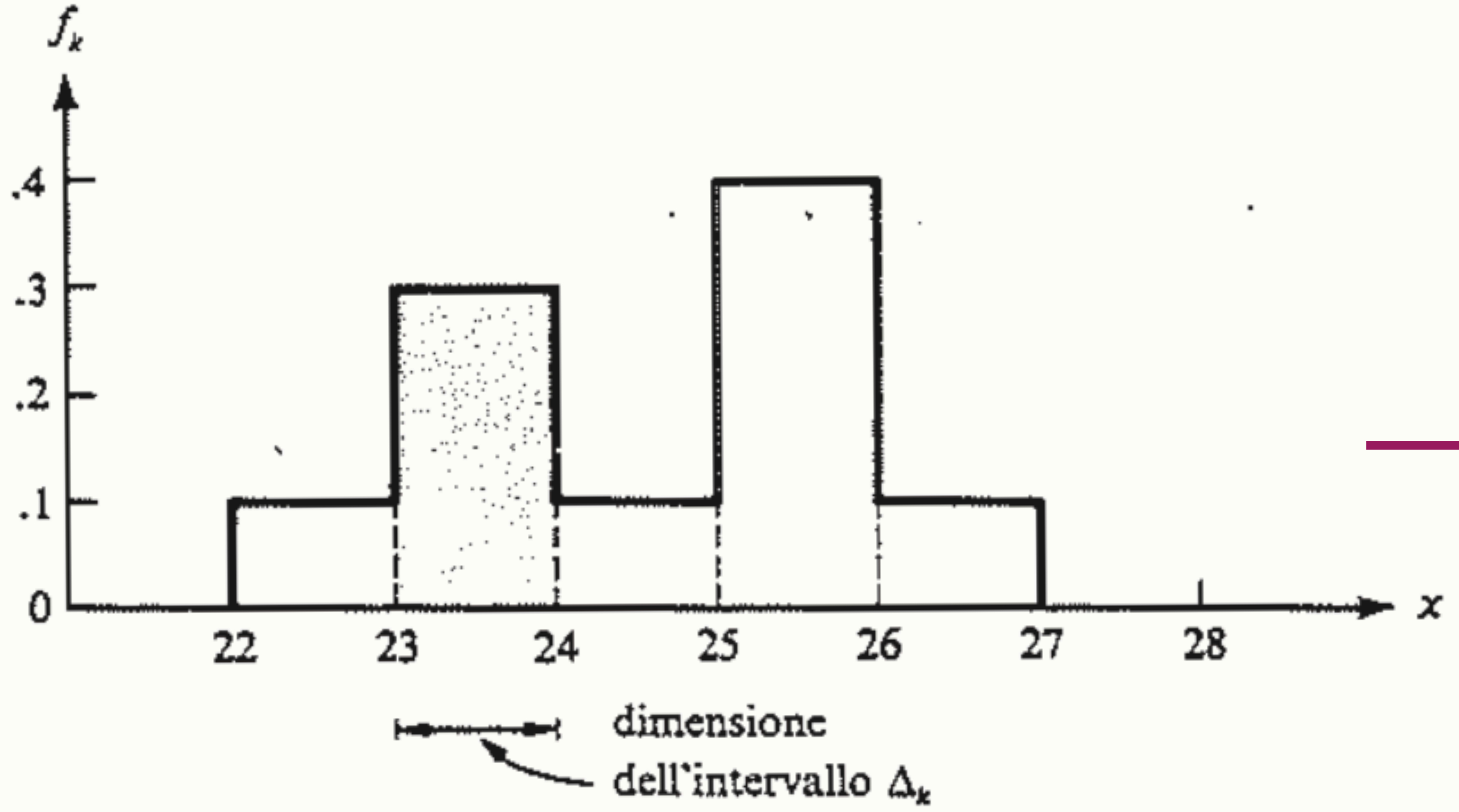


Figura 5.2. *L'istogramma ad intervalli mostra la frazione di misure di x che cadono negli "intervalli" 22-23, 23-24, e così via. L'area del rettangolo sopra ciascun intervallo dà la frazione di misure che cadono in quell'intervallo. Così l'area del rettangolo ombreggiato è 0.3, indicando che 3/10 di tutte le misure giacciono tra 23 e 24.*

Istogramma a intervalli:
grafico di f_k in funzione di x_k

Istogrammi a intervalli

Altezza del rettangolo f_k disegnato sopra l'intervallo Δ_k è scelta in modo tale che l'area $f_k\ \Delta_k$ sia uguale alla frazione di misure nell'intervallo k-esimo

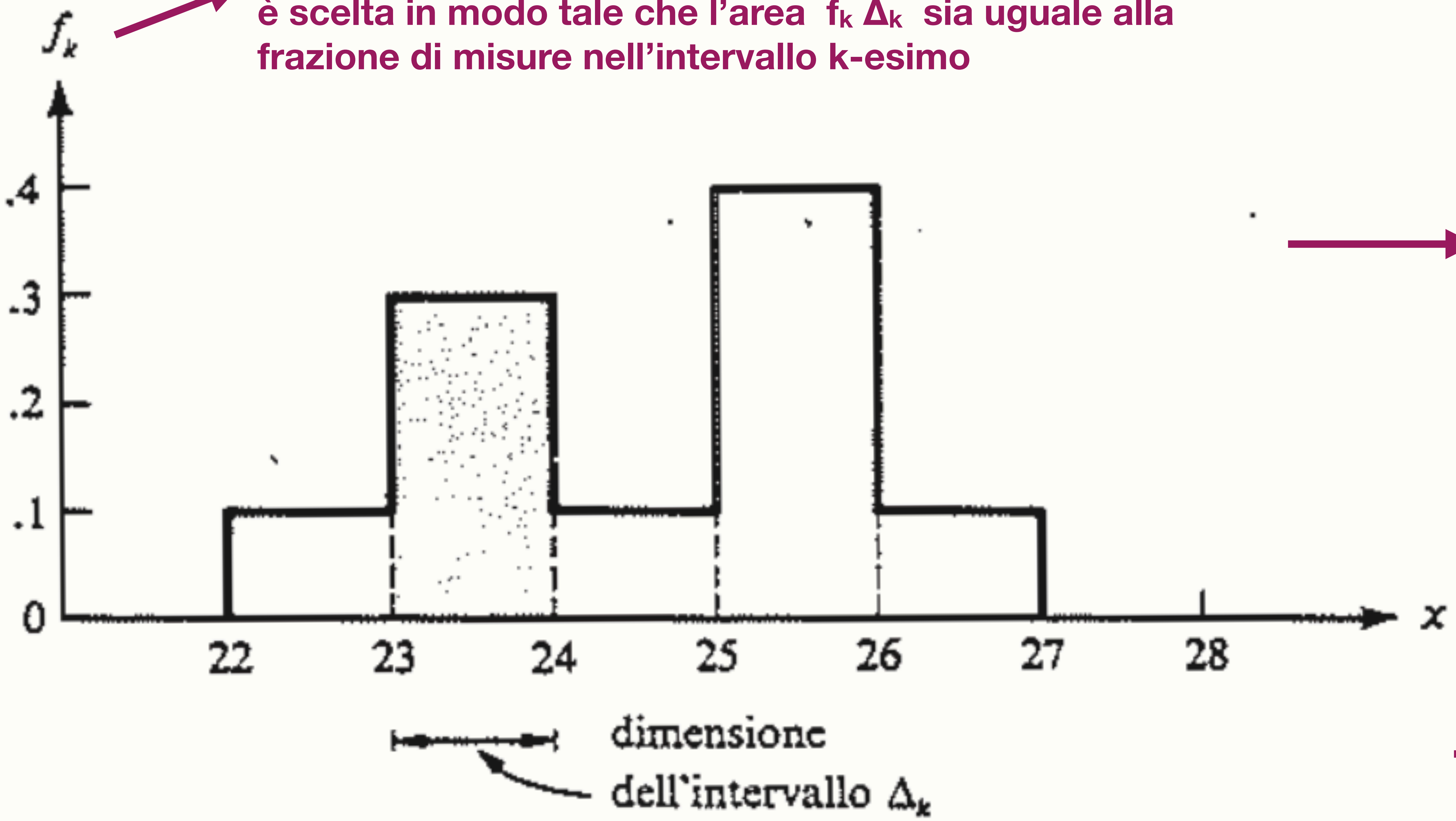


Figura 5.2. *L'istogramma ad intervalli mostra la frazione di misure di x che cadono negli "intervalli" 22-23, 23-24, e così via. L'area del rettangolo sopra ciascun intervallo dà la frazione di misure che cadono in quell'intervallo. Così l'area del rettangolo ombreggiato è 0.3, indicando che 3/10 di tutte le misure giacciono tra 23 e 24.*

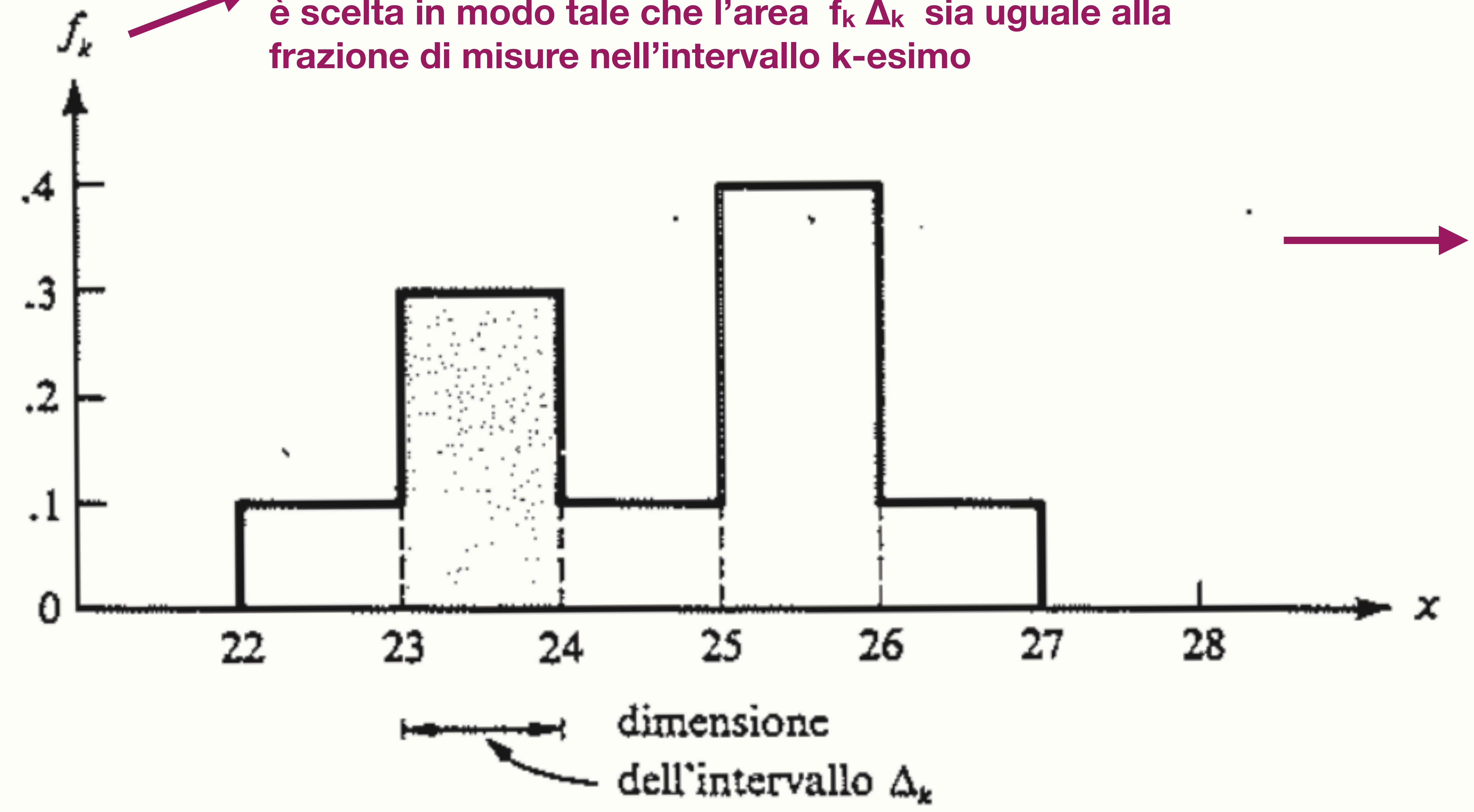
La frazione di misure che cadono in ciascun intervallo è indicata dall'area del rettangolo disegnato sopra l'intervallo.

Gli intervalli sono in genere tutti uguali, ma non devono necessariamente esserlo.

In un istogramma ad intervalli l'area $f_k\ \Delta_k$ ha lo stesso significato che l'altezza F_k ha in un istogramma a barre.

Istogrammi a intervalli

Altezza del rettangolo f_k disegnato sopra l'intervallo Δ_k è scelta in modo tale che l'area $f_k \Delta_k$ sia uguale alla frazione di misure nell'intervallo k-esimo



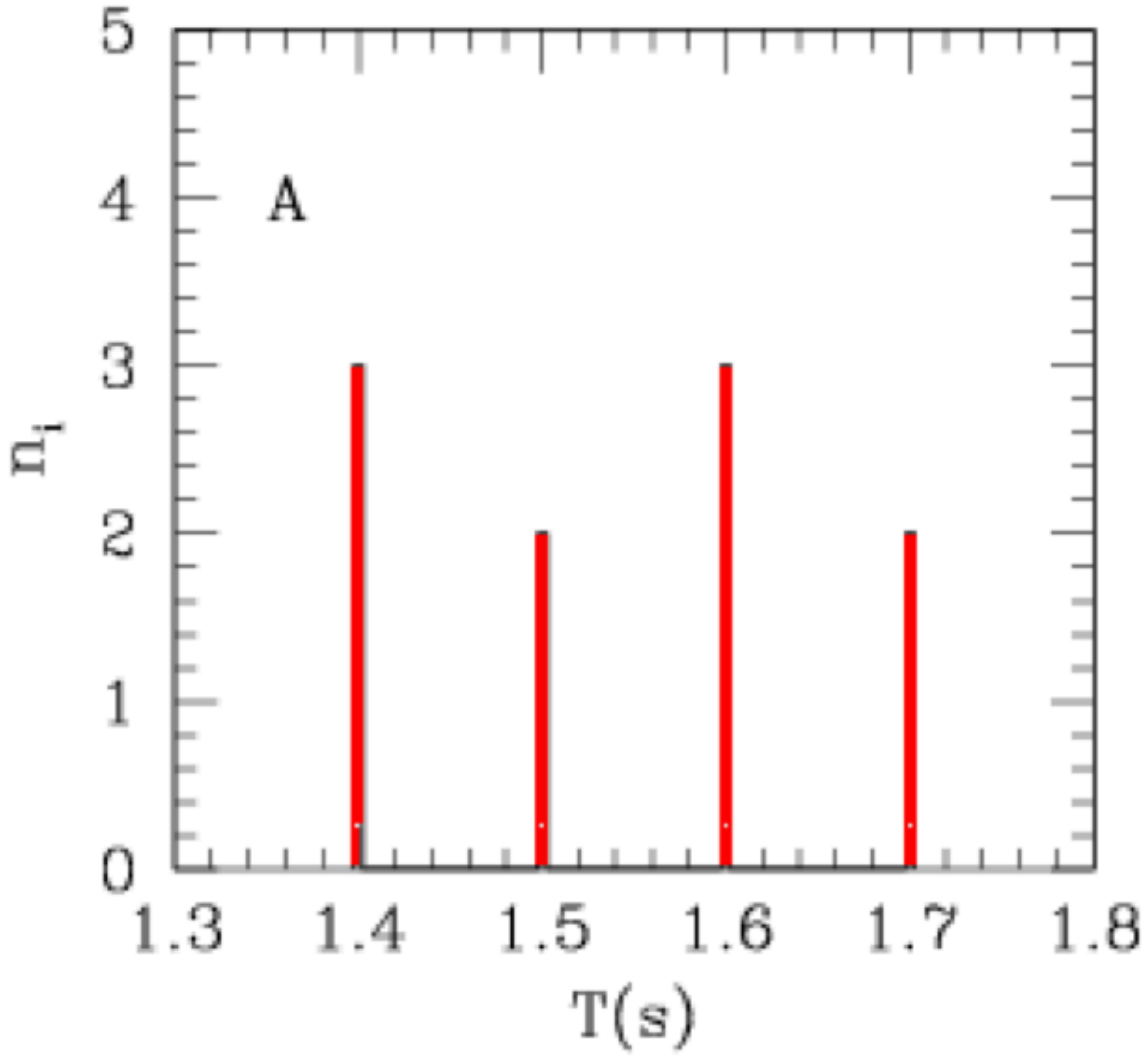
La frazione di misure che cadono in ciascun intervallo è indicata dall'area del rettangolo disegnato sopra l'intervallo.

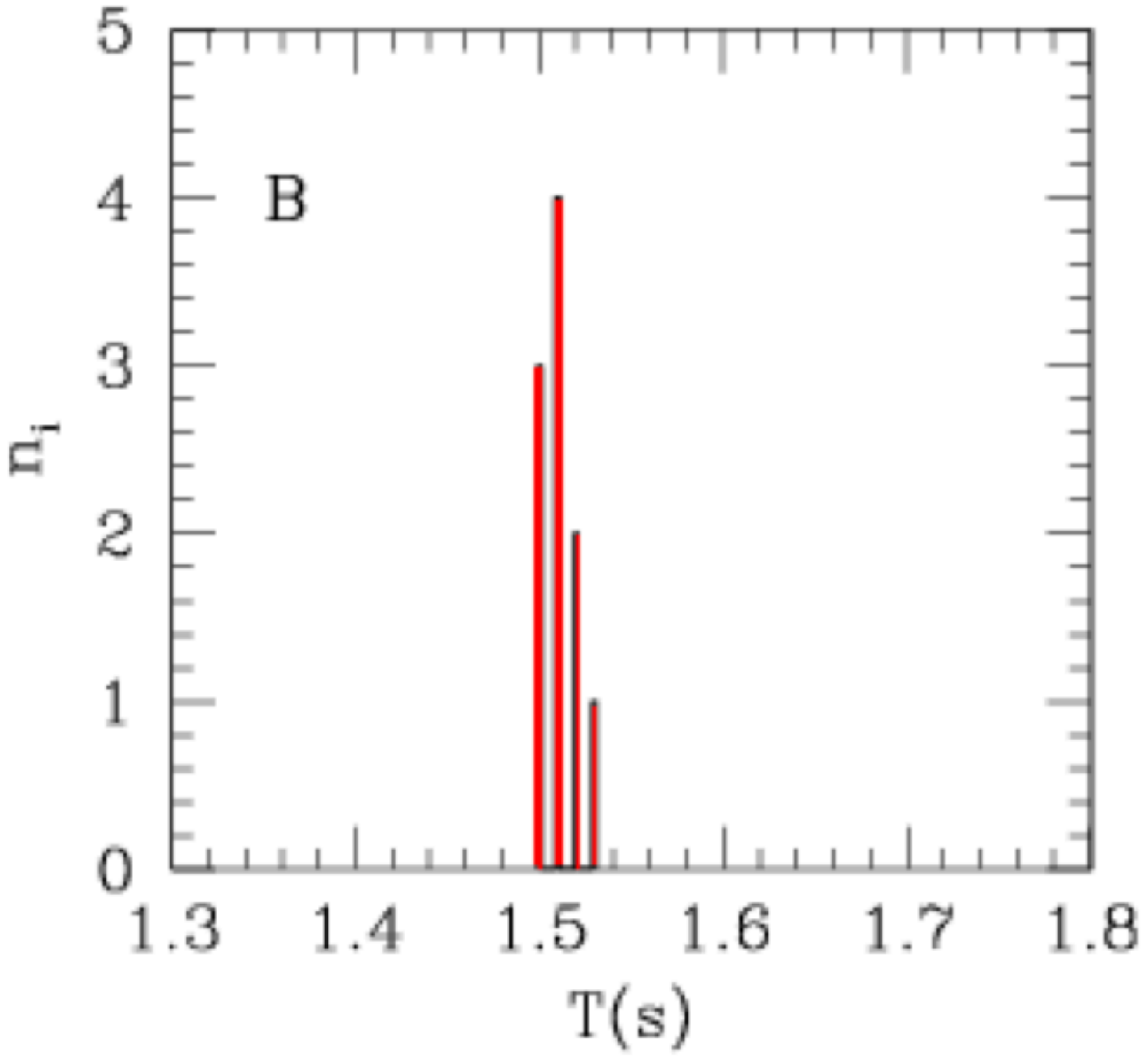
In un istogramma ad intervalli l'area $f_k \Delta_k$ ha lo stesso significato che l'altezza F_k ha in un istogramma a barre.

Generalmente si presta attenzione perché l'area totale dell'istogramma sia unitaria (istogramma normalizzato a 1).

Scelta dell'ampiezza degli intervalli (bin) è cruciale.

Istogrammi, frequenze, normalizzazione





Misure Andrea	Frequenza	Misure Barbara	Frequenza
1.4	3	1.50	3
1.5	2	1.51	4
1.6	3	1.52	2
1.7	2	1.53	1

Diagrammi a barre in frequenze assolute, relativi alle misure di A e B.

La **frequenza assoluta** di un certo dato è il numero di volte che lo stesso si presenta in una certa serie di misure.

Per **frequenza relativa** si definisce il rapporto fra la frequenza assoluta di un certo valore ottenuto ed il numero totale delle misure effettuate, N: $f_i = \frac{n_i}{N}$

$$\sum_{i=1}^{N} n_i = N \quad \longrightarrow \quad \sum_{i=1}^{N} \frac{n_i}{N} = 1$$

Istogrammi, frequenze, normalizzazione

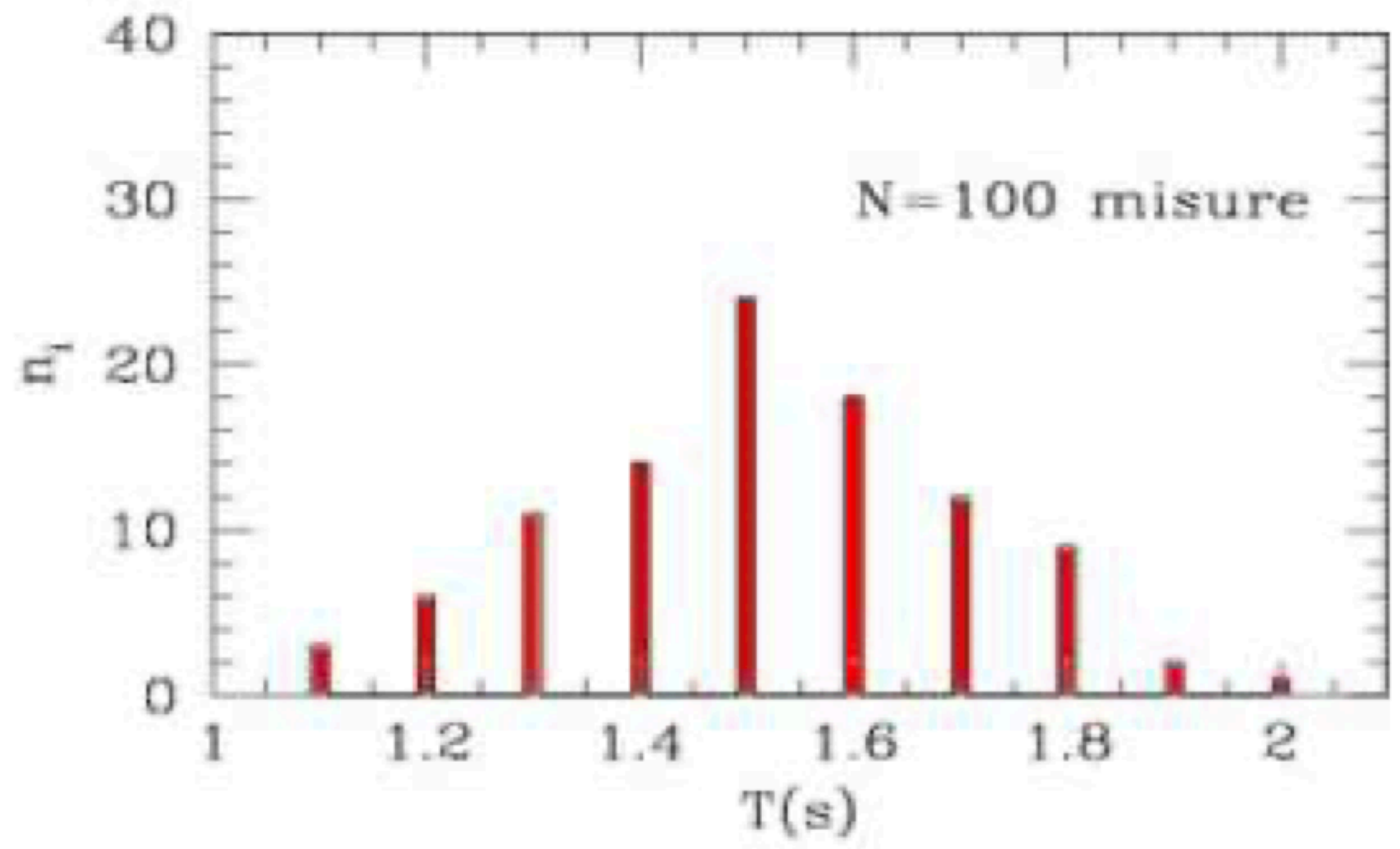


Figura 4.2: Diagramma a barre in frequenze assolute, relativo a alle 100 misure effettuate da A.

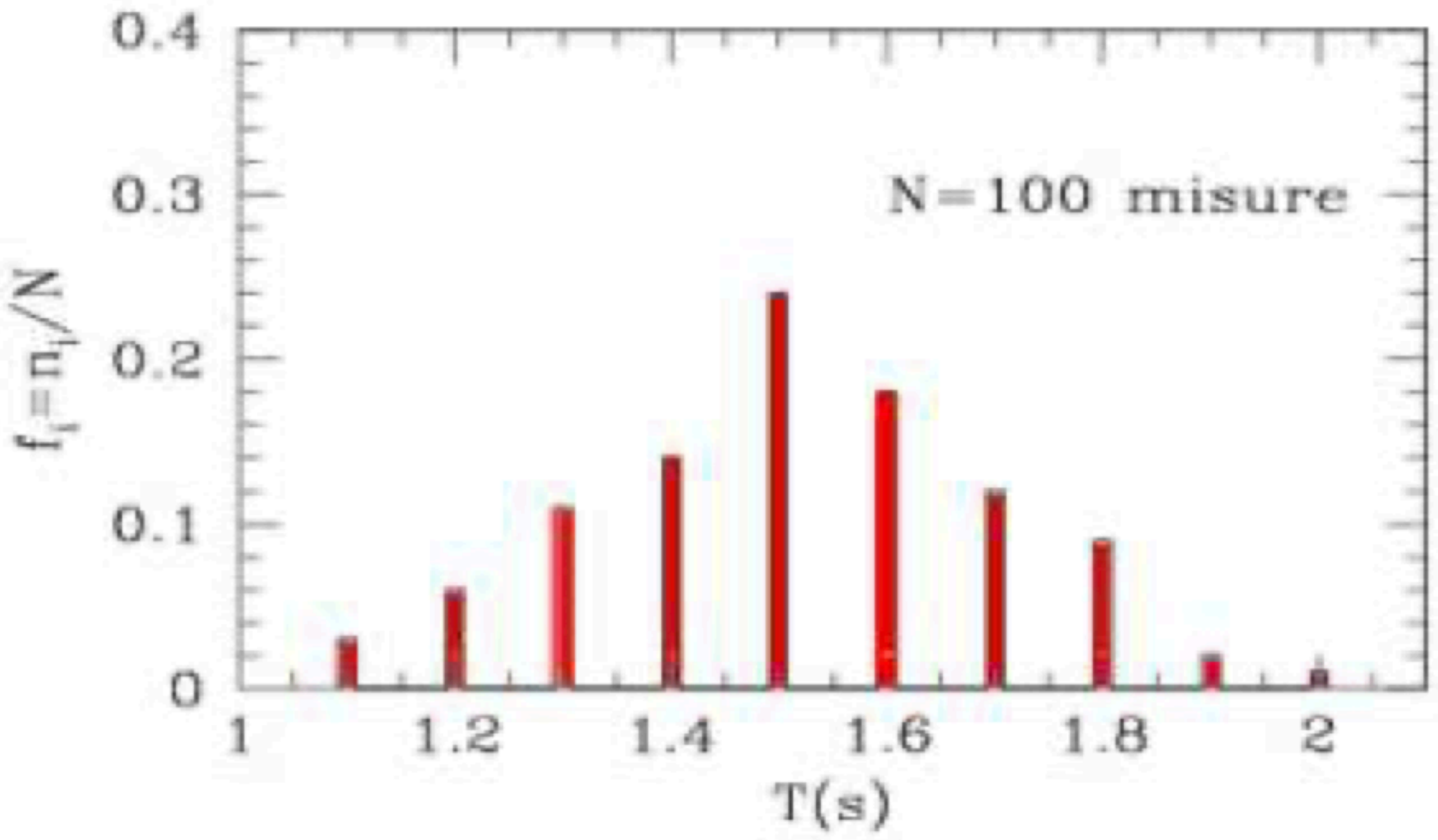


Figura 4.3: Diagramma a barre in frequenze relative, relativo alle 100 misure effettuate da A.

Sui diagrammi delle **frequenze relative** occorre sempre **indicare il numero totale** di misure fatte in quanto questa informazione non è più direttamente ricavabile da esso.

$$\sum_{i=1}^{N} n_i = N \quad \longrightarrow \quad \sum_{i=1}^{N} \frac{n_i}{N} = 1$$

La distribuzione limite

Nella maggior parte degli esperimenti, se si aumenta il numero di misure, l'istogramma inizia ad assumere qualche semplice forma definita. Nel processo, cambiano spesso la posizione e il numero dei picchi e le proprietà di simmetria della figura.

Ad esempio, se si pensa di ripetere 10, 100, oppure 1000 volte una misura di lunghezza di un'asticella lunga circa 25 cm, la situazione potrebbe facilmente essere la seguente:

La distribuzione limite

Ad esempio, se si pensa di ripetere 10, 100, o 1000 volte una misura di lunghezza di un'asticella lunga circa 25 cm, la situazione potrebbe facilmente essere la seguente:

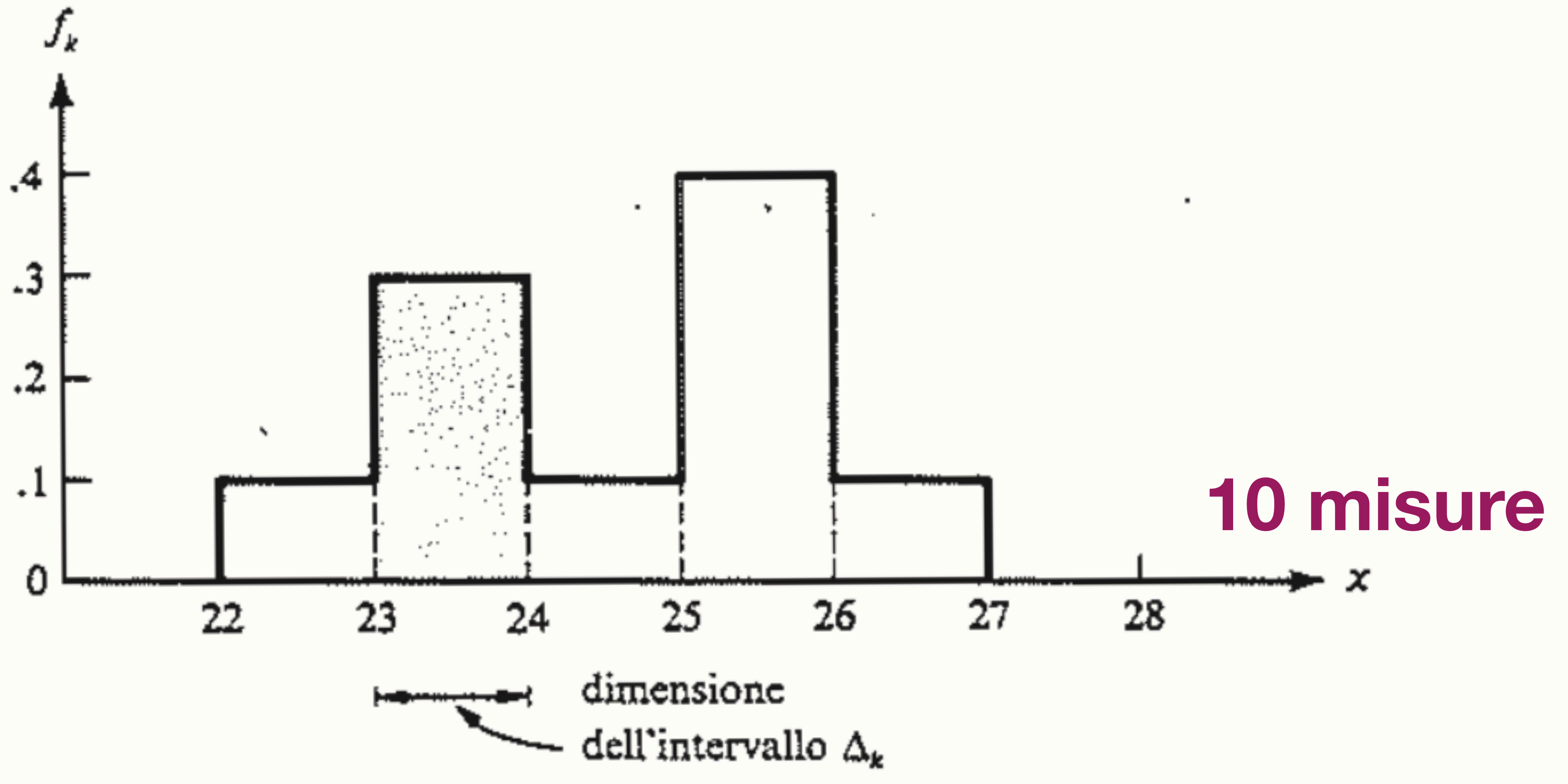


Figura 5.2. *L'istogramma ad intervalli mostra la frazione di misure di x che cadono negli "intervalli" 22-23, 23-24, e così via.*

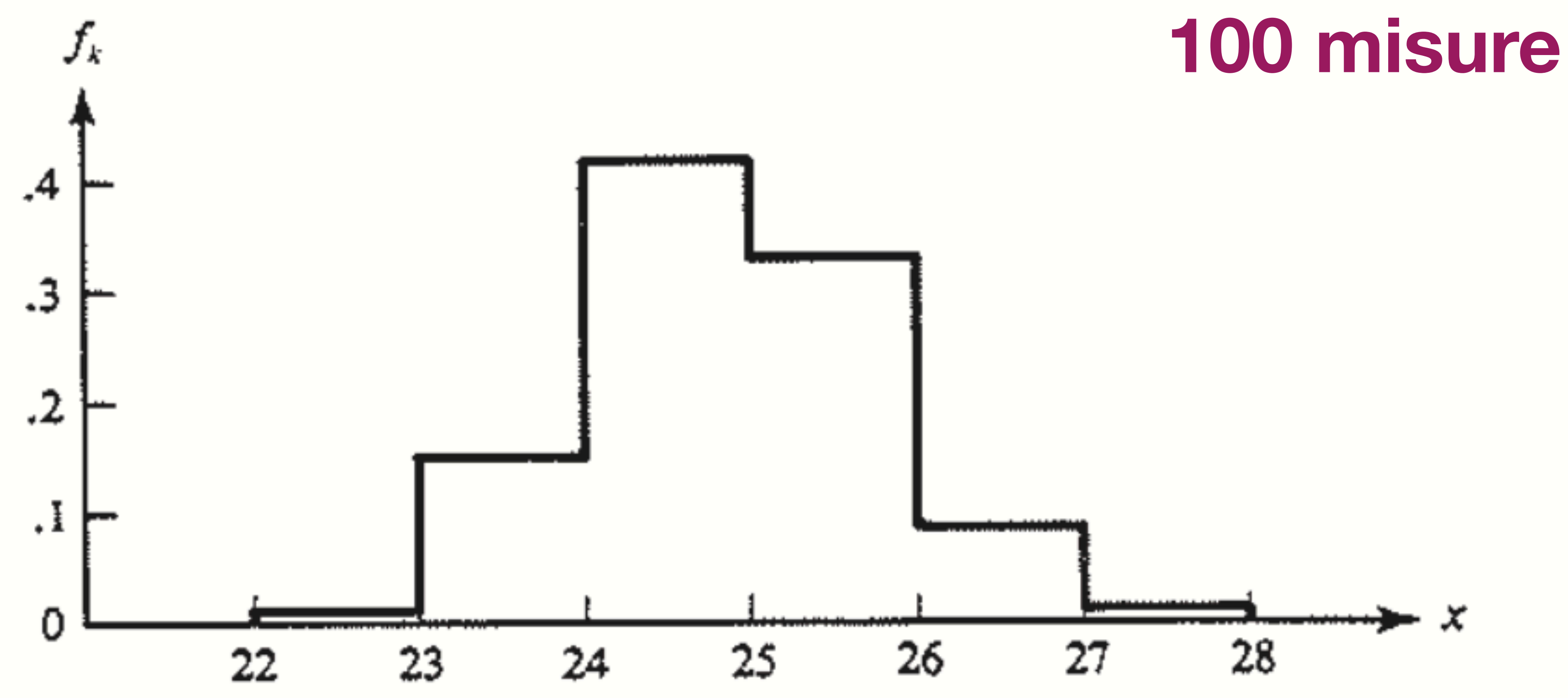


Figura 5.3. *Istogramma per 100 misure della stessa grandezza di Figura 5.2.*

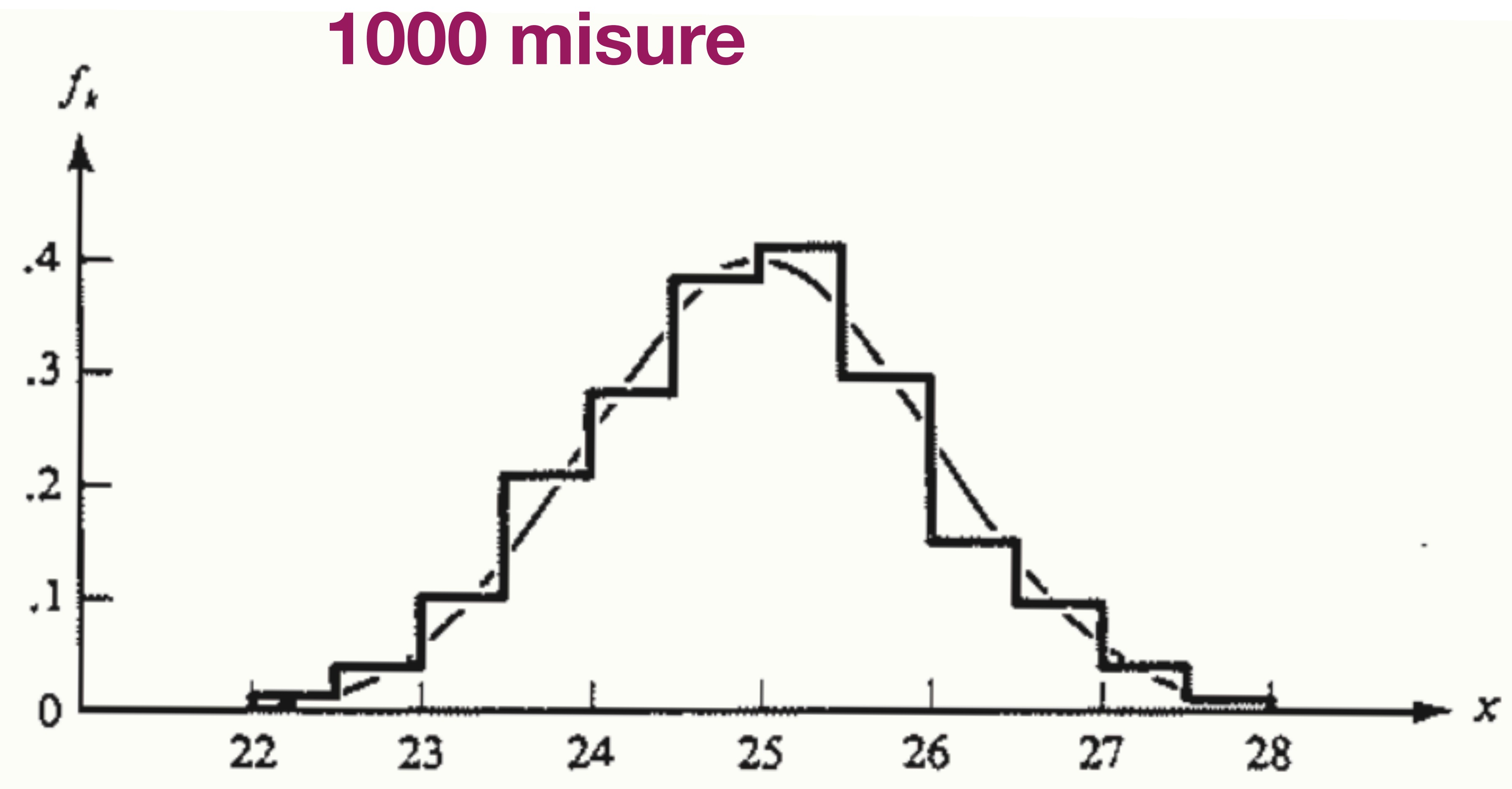


Figura 5.4. *Istogramma per 1000 misure della stessa grandezza di Figura 5.3. La curva tratteggiata è la distribuzione limite.*

La distribuzione limite

Ad esempio, se si pensa di ripetere 10, 100, o 1000 volte una misura di lunghezza di un'asticella lunga circa 25 cm, la situazione potrebbe facilmente essere la seguente:

10 misure

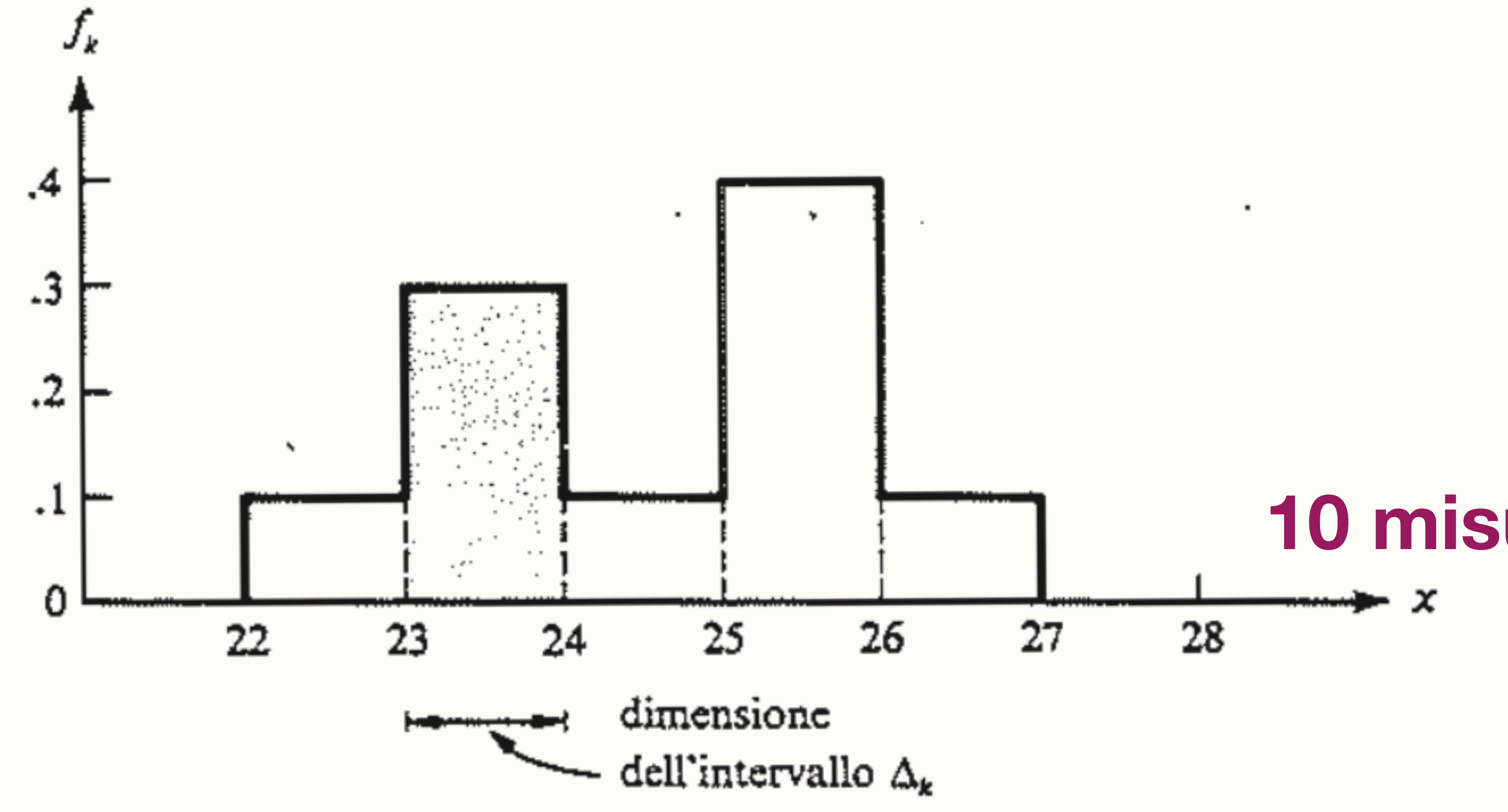


Figura 5.2. *L'istogramma ad intervalli mostra la frazione di misure di x che cadono negli "intervalli" 22-23, 23-24, e così via.*

100 misure

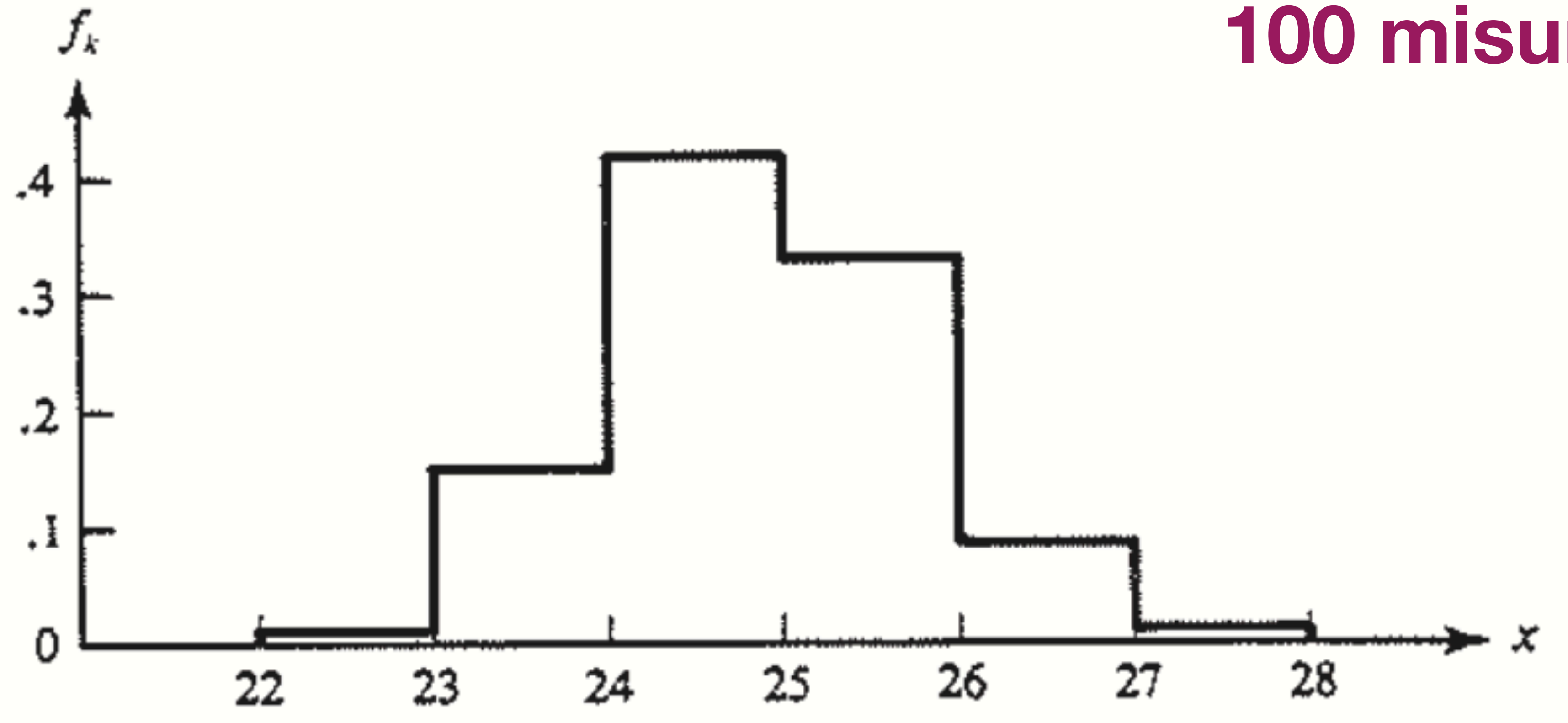


Figura 5.3. *Istogramma per 100 misure della stessa grandezza di Figura 5.2.*

1000 misure

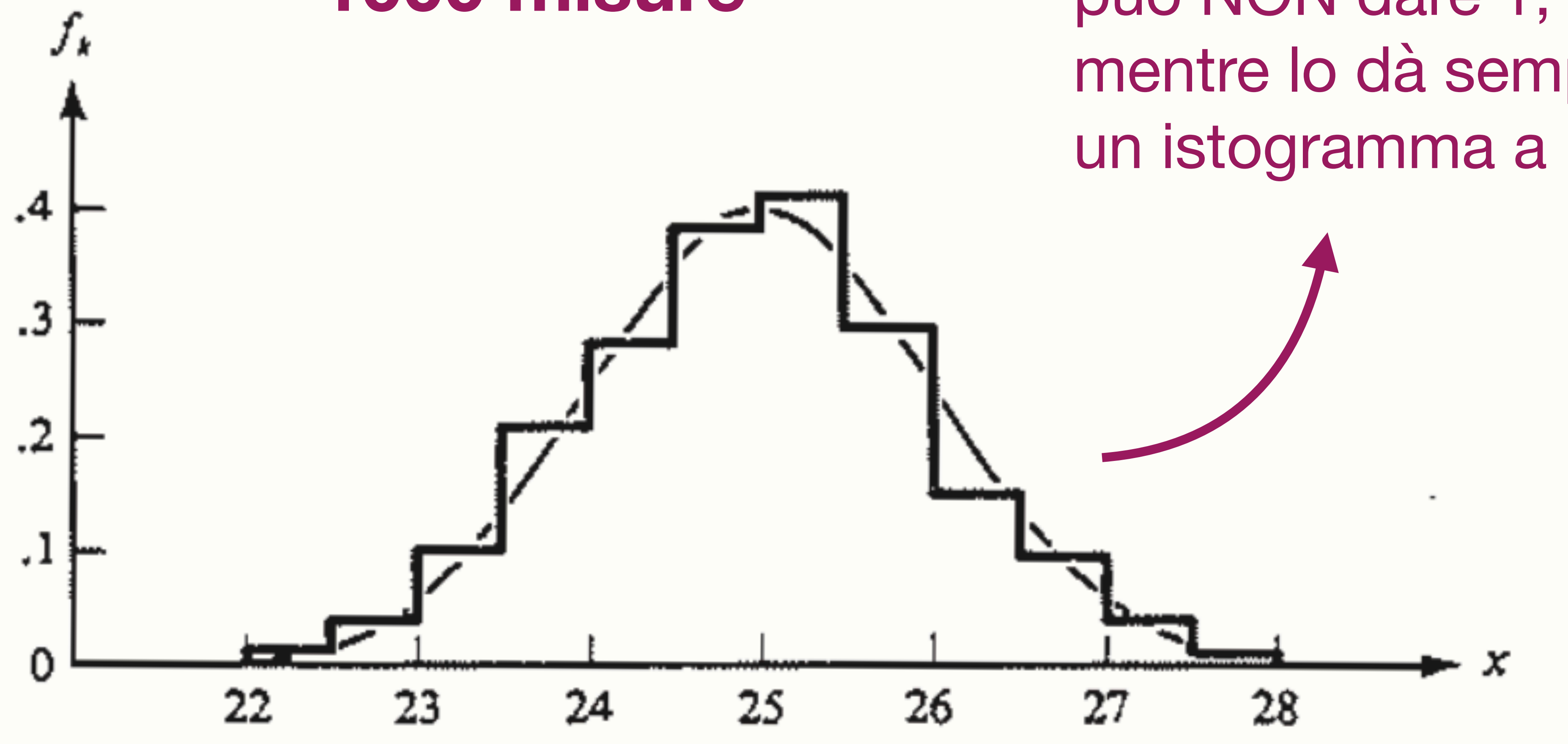


Figura 5.4. *Istogramma per 1000 misure della stessa grandezza di Figura 5.3. La curva tratteggiata è la distribuzione limite.*

NB: La somma delle frequenze relative in un istogramma ad intervalli normalizzato può NON dare 1, mentre lo dà sempre in un istogramma a barre

La distribuzione limite

Quando il numero di misure si avvicina all'infinito, la loro distribuzione si avvicina a qualche curva continua e definita.

Quando ciò accade, la curva continua è chiamata la distribuzione limite.

La **distribuzione limite** è la distribuzione che si otterrebbe se il numero delle misure diventasse infinitamente grande.

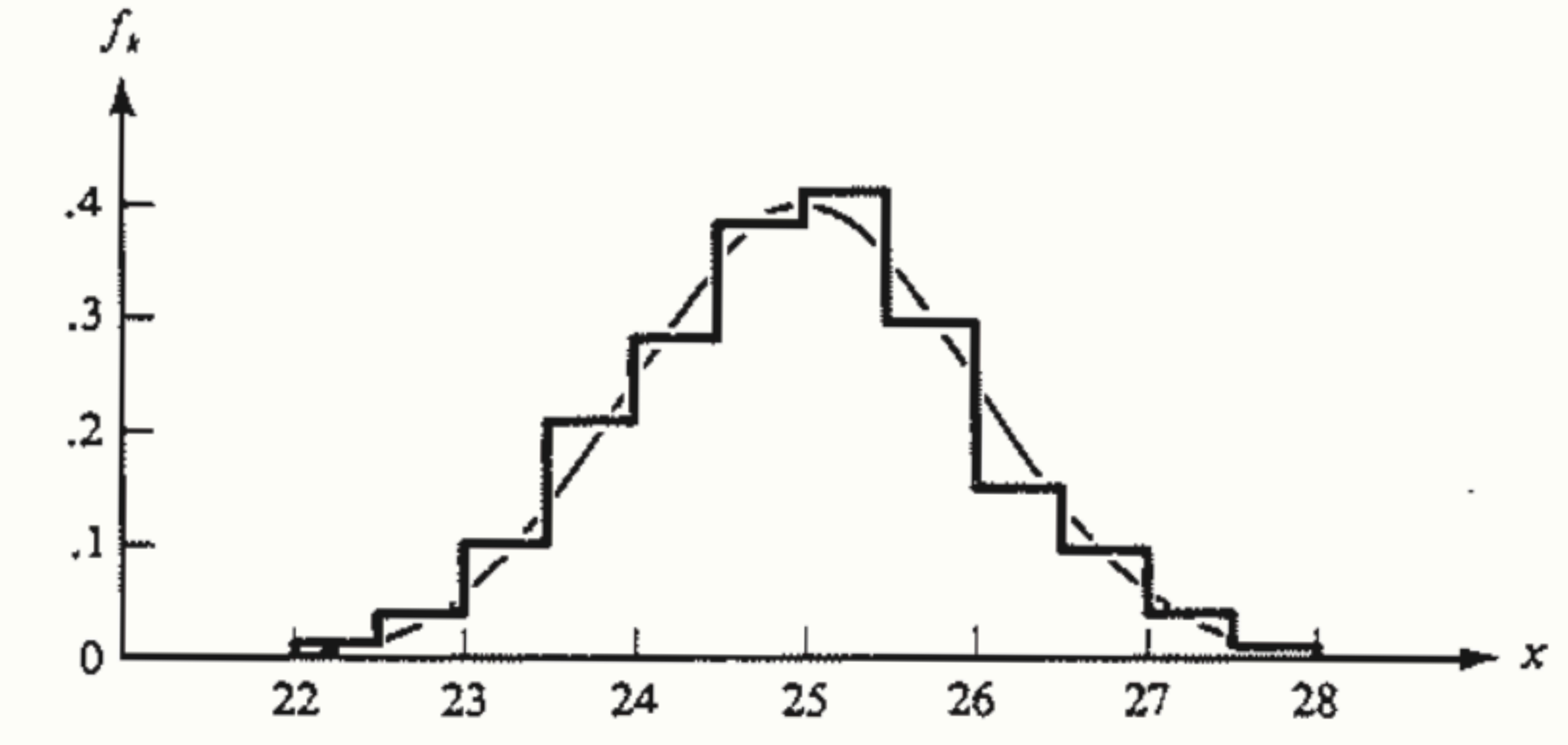


Figura 5.4. *Istogramma per 1000 misure della stessa grandezza di Figura 5.3. La curva tratteggiata è la distribuzione limite.*

La distribuzione limite è una costruzione teorica: non può essere misurata esattamente.
Il nostro istogramma si avvicina asintoticamente tanto più alla distribuzione limite quante più misure si fanno.

La distribuzione limite si otterrebbe nel limite di infinite misure e di intervalli infinitamente stretti.

Una distribuzione limite definisce una **curva continua** che chiamiamo $f(x)$.

La distribuzione limite

Quando il numero di misure si avvicina all'infinito, la loro distribuzione si avvicina a qualche curva continua e definita.

Quando ciò accade, la curva continua è chiamata la distribuzione limite.

La **distribuzione limite** è la distribuzione che si otterrebbe se il numero delle misure diventasse infinitamente grande.

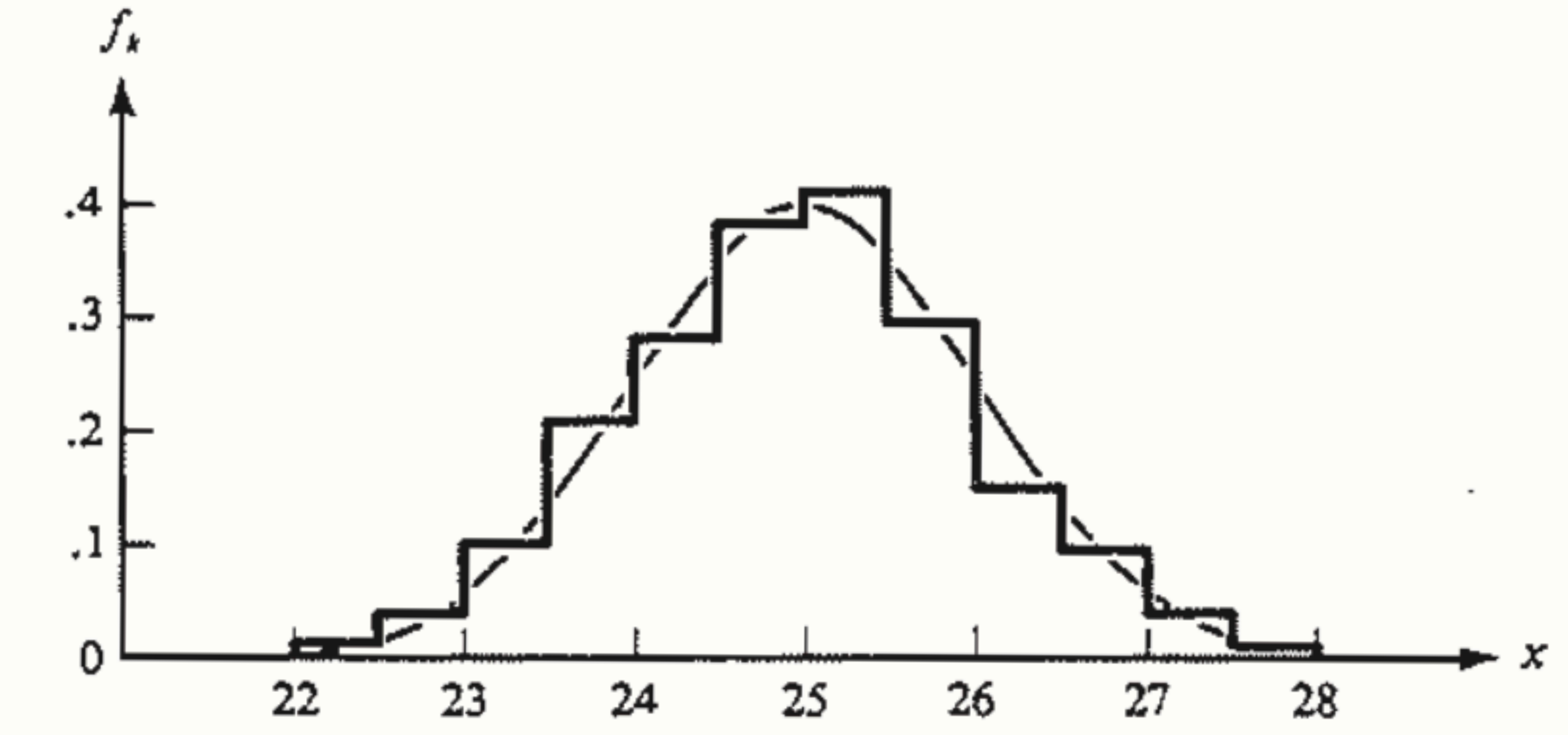


Figura 5.4. *Istogramma per 1000 misure della stessa grandezza di Figura 5.3. La curva tratteggiata è la distribuzione limite.*

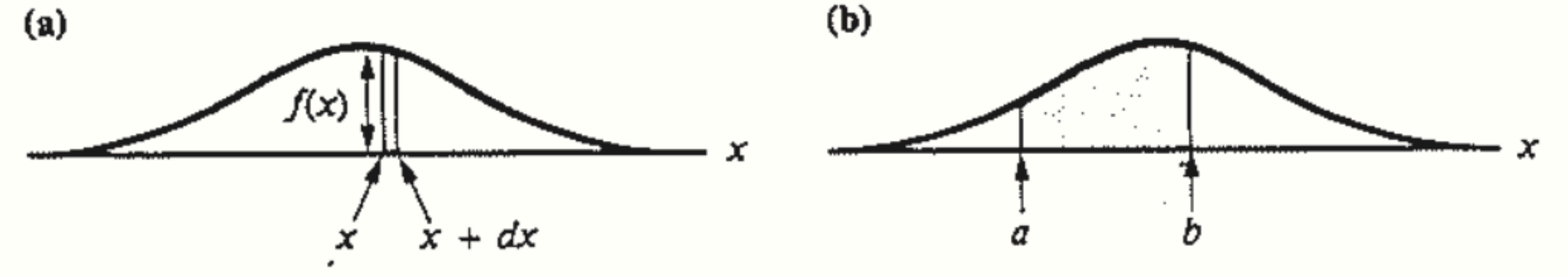


Figura 5.5. *Una distribuzione limite* f(x). *(a) Dopo molte misure la frazione che cade fra* x *e* x + dx *è l'area* f(x)dx *della stretta striscia. (b) La frazione che cade fra* x = a *e* x = b *è l'area ombreggiata.*

Significato della distribuzione limite: all'aumentare del numero di misure della grandezza x, l'istogramma diventerà indistinguibile da $f(x)$. Allora **la frazione di misure che cadono tra x e x+dx è uguale all'area** $f(x)dx$

La distribuzione limite

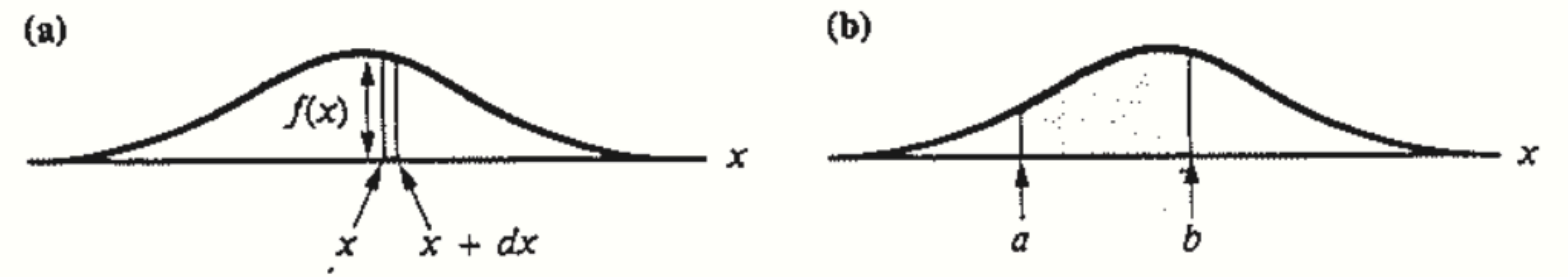


Figura 5.5. *Una distribuzione limite* f(x). *(a) Dopo molte misure la frazione che cade fra* x *e* x + dx *è l'area* f(x)dx *della stretta striscia. (b) La frazione che cade fra* x = a *e* x = b *è l'area ombreggiata.*

Quando l'istogramma diventa indistinguibile da $f(x)$, **la frazione di misure che cadono tra x e x+dx è uguale all'area** $f(x)dx$

In generale, il numero di misure che cadono tra **a** e **b** è l'area totale sotto il grafico tra x=a e x=b
Quest'area è l'integrale definito di $f(x)$:

$$\int_a^b f(x)\,dx$$ = frazione di misure che cadono tra x=a e x=b

$f(x)\,dx$ **è la probabilità che una singola misura dia un risultato compreso tra x e x+dx**

Se si conosce la distribuzione limite $f(x)$ per la misura di una grandezza x, allora si conosce la probabilità di ottenere un risultato in qualsiasi intervallo a $\leq$ x $\leq$ b.

La distribuzione limite

Dal momento che la probabilità totale di ottenere un risultato in qualunque punto tra $-\infty$ e $+\infty$ deve essere 1, la funzione limite $f(x)$ deve soddisfare:

$$\int_{-\infty}^{+\infty} f(x)\,dx = 1 \qquad \Longrightarrow f(x) \text{ è } \textbf{normalizzata}$$

$f(x)$ è essenzialmente zero al di fuori dell'intervallo in cui le misure sono più probabili, per cui non ci si deve aspettare di ottenere valori della grandezza misurata che siano prossimi a $\pm\infty$

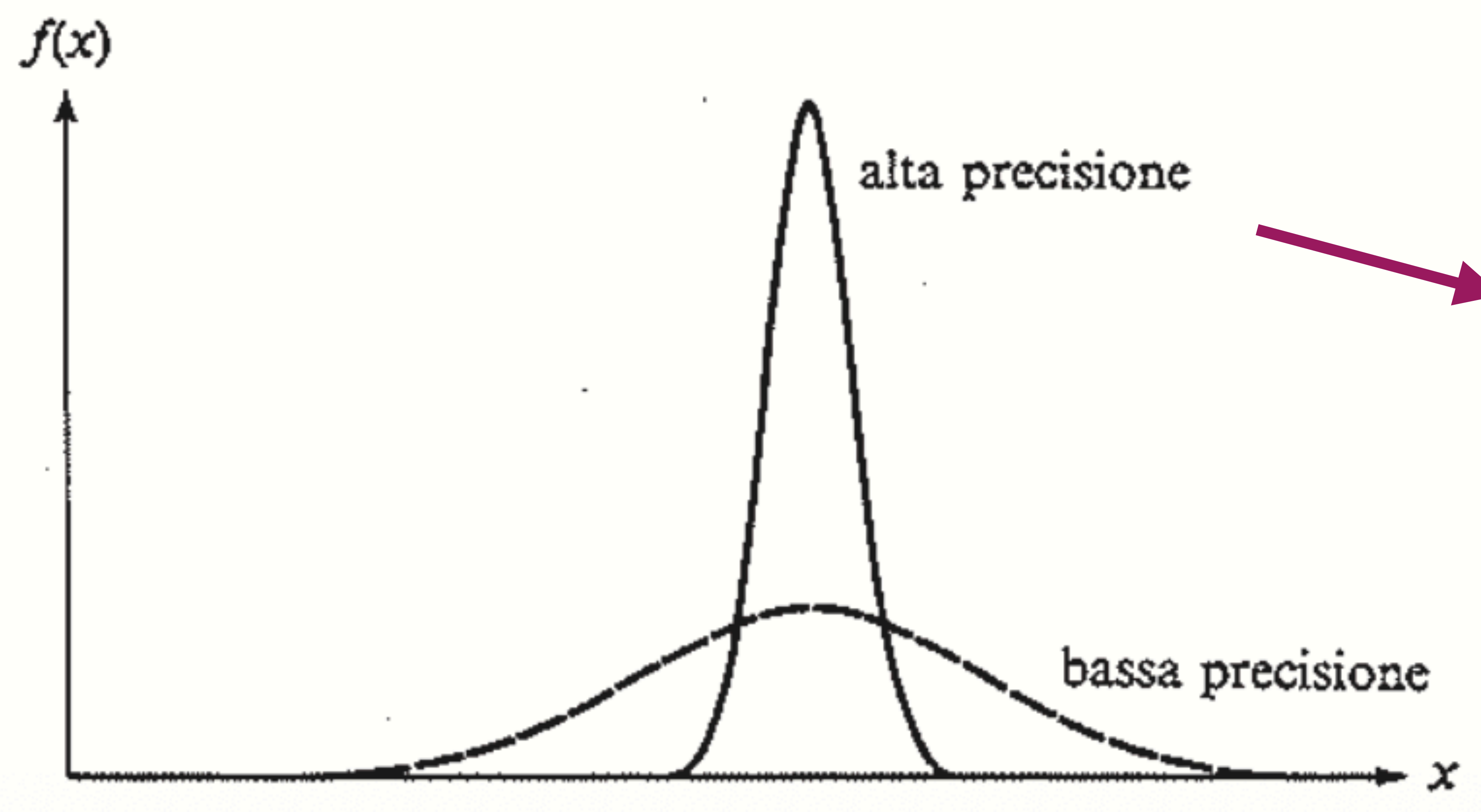


Figura 5.6. *Due distribuzioni limite, una per una misura ad alta precisione, l'altra per una misura a bassa precisione.*

distribuzione limite stretta e piccata

distribuzione limite più ampia

La media di una distribuzione

Se conosciamo $f(x)$ e dobbiamo calcolare il **valor medio** $\bar{x}$:

$$\bar{x} = \int_{-\infty}^{+\infty} x f(x)\, dx$$

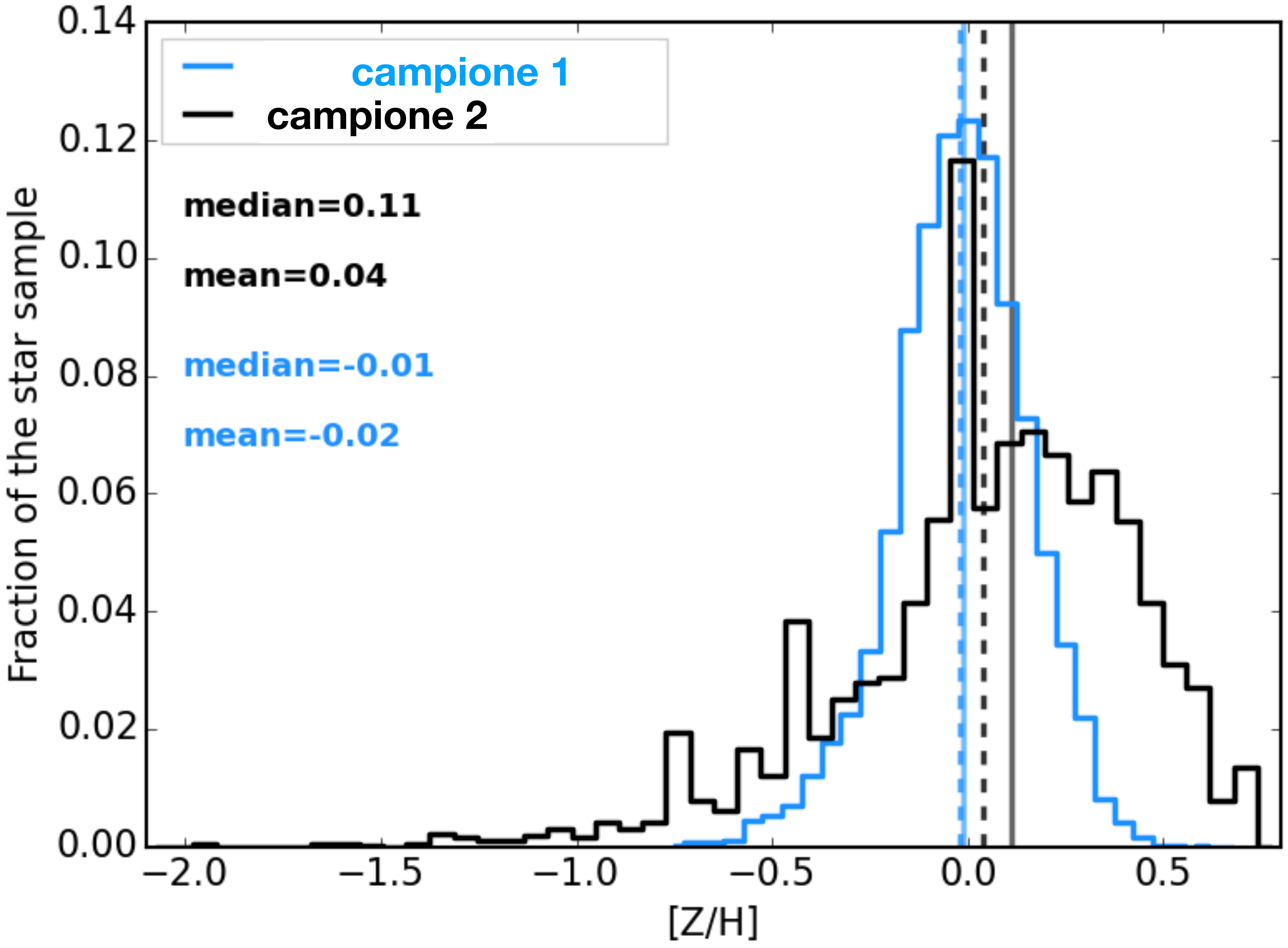
dove $f(x)\, dx$ è la frazione di valori in ciascun intervallo dx, in analogia con $\bar{x} = \sum_k x_k F_k$

Con ragionamento simile, la **varianza** può essere calcolata come:

$$\sigma_x^{\ 2} = \int_{-\infty}^{+\infty} (x - \bar{x})^2 f(x)\, dx$$

scritta come la media della deviazione standard al quadrato.

Istogrammi



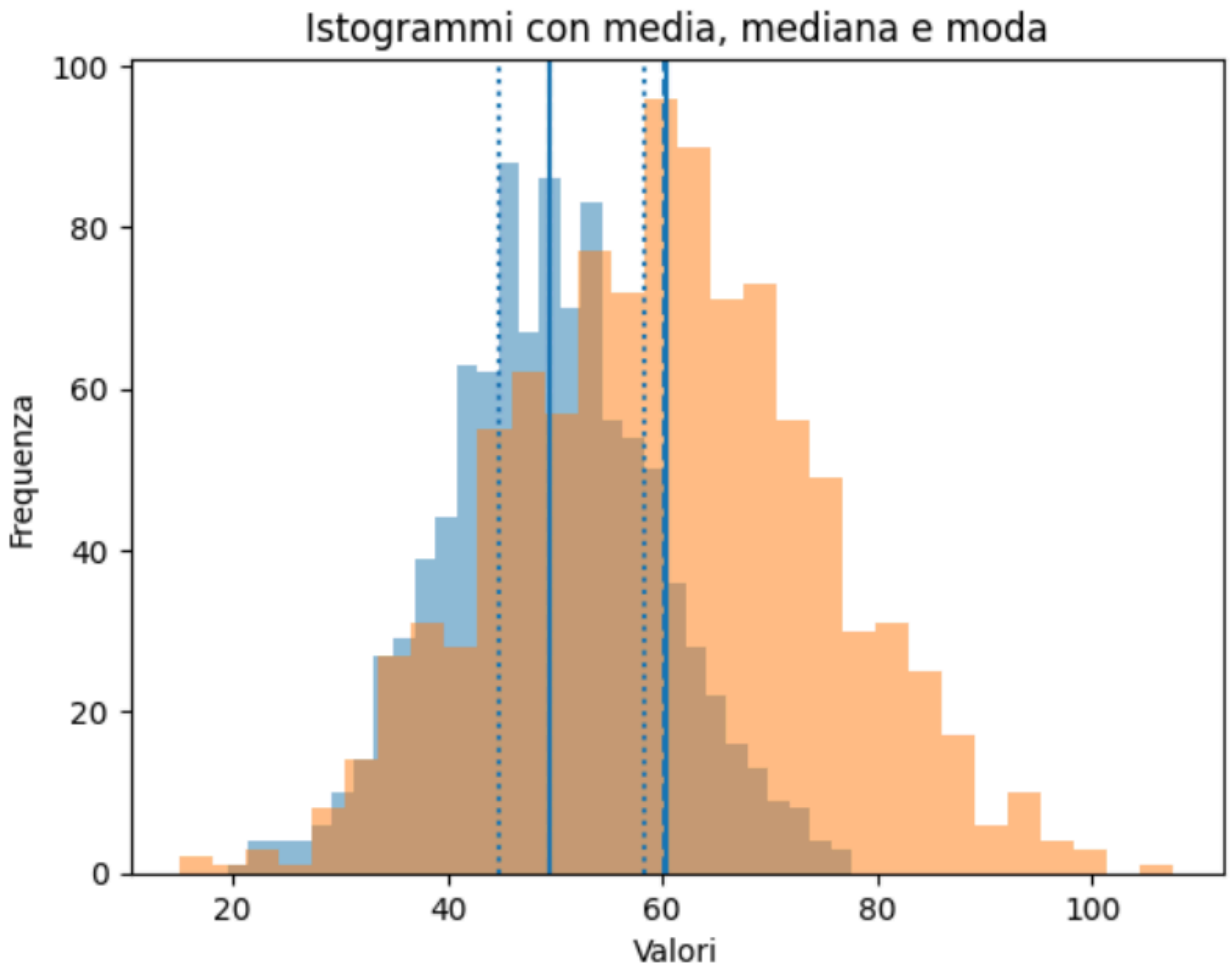
campione 1
campione 2
median=0.11
mean=0.04
median=-0.01
mean=-0.02
Fraction of the star sample
[Z/H]
0.14
0.12
0.10
0.08
0.06
0.04
0.02
0.00
−2.0
−1.5
−1.0
−0.5
0.0
0.5

Istogrammi

Linee verticali: media - - -
mediana —
moda . . .

Istogramma 1:
media = 49.55
mediana = 49.42
moda = 44.70

Istogramma 1:
media = 60.20
mediana = 60.39
moda = 58.24



Istogrammi

Istogramma
non normalizzato!