

Misure di dispersione e della tendenza centrale

Uso e necessità della statistica

- La statistica offre un modello per interpretare i fenomeni psicologici, e ci permette di **studiare le relazioni** fra diverse misurazioni;
- Per esempio: intelligenza e successo scolastico; bisogni fisici (come fame) e motivazione; influenza degli ormoni e orientamento sessuale ecc.
- Ma per capire e osservare le relazioni, dobbiamo inizialmente osservare la distribuzione delle **singole misurazioni**, rilevate secondo la quadruplica classificazione, prima di applicare i modelli esplicativi

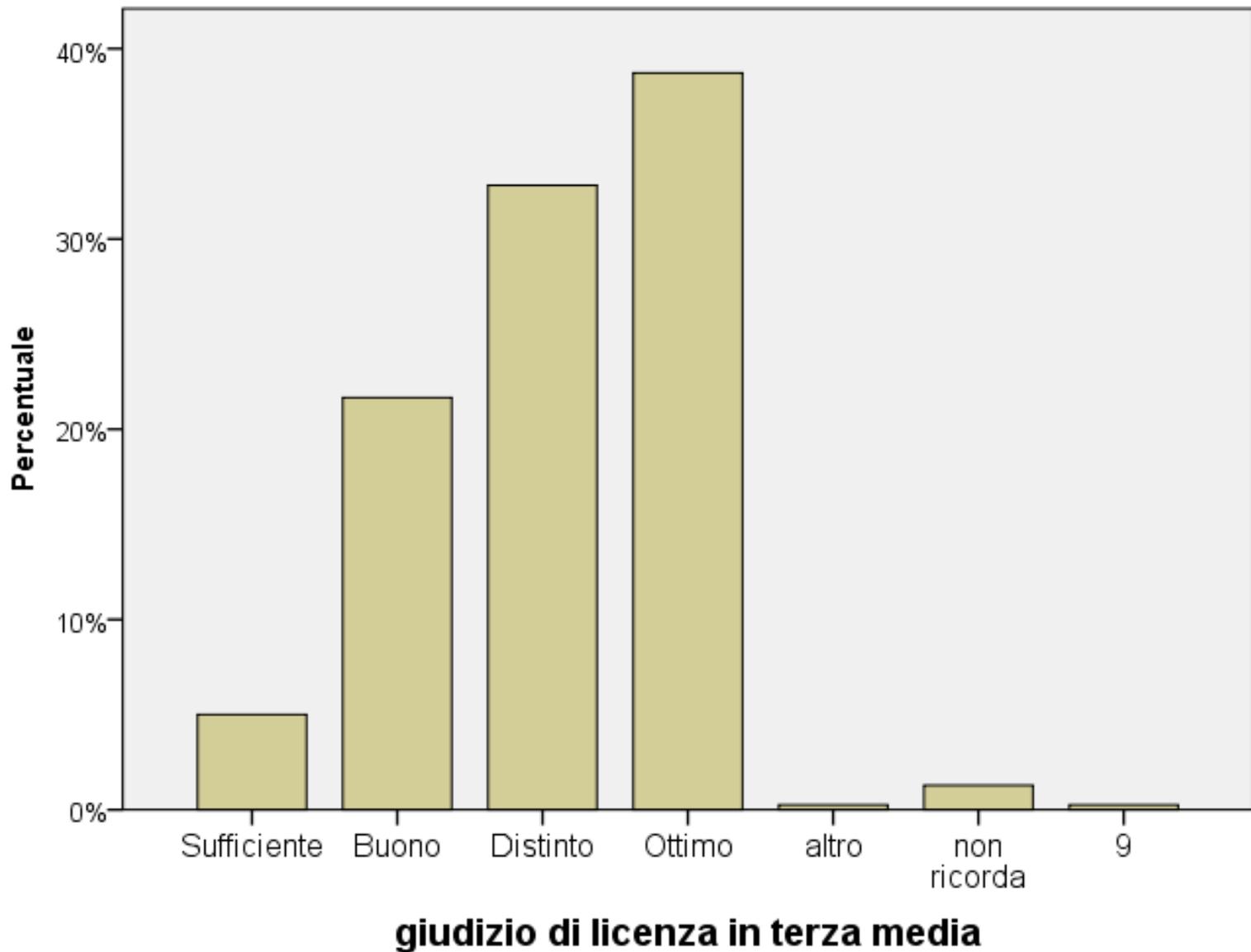
- Dobbiamo imparare quindi alcuni concetti importanti, che si applicano alle **singole** variabili, e permettono di descrivere le loro caratteristiche principali.
- Alcune sono già note, (come la media aritmetica), altre sono poco conosciute e poco usate.
- Cominciamo a rivedere la distribuzione di frequenza.

Un esempio di distribuzione di frequenza

giudizio di licenza in terza media

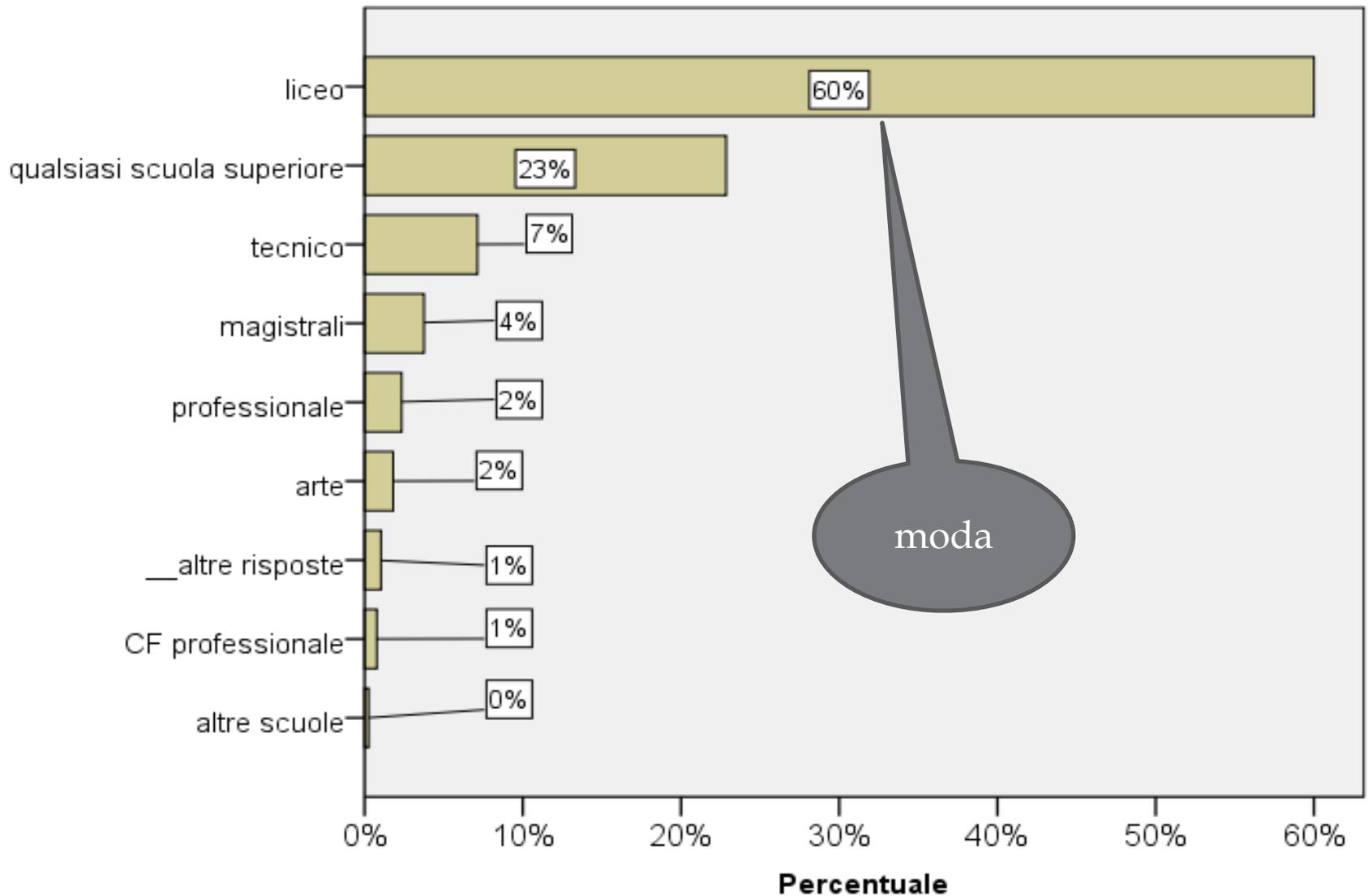
		Frequenza	Percentuale	Percentuale valida	Percentuale cumulativa
Valido	1 Sufficiente	39	5,0	5,0	5,0
	2 Buono	169	21,6	21,7	26,7
	3 Distinto	256	32,7	32,8	59,5
	4 Ottimo	302	38,6	38,7	98,2
	5 altro	2	,3	,3	98,5
	6 non ricorda	10	1,3	1,3	99,7
	9	2	,3	,3	100,0
	Totale	780	99,6	100,0	
Mancante/i	Sistema	3	,4		
Totale		783	100,0		

Grafico degli stessi dati



Le scuole consigliate in terza media agli studenti di un liceo possono essere ordinate secondo la percentuale di occorrenza, per mettere in evidenza la categoria più frequente

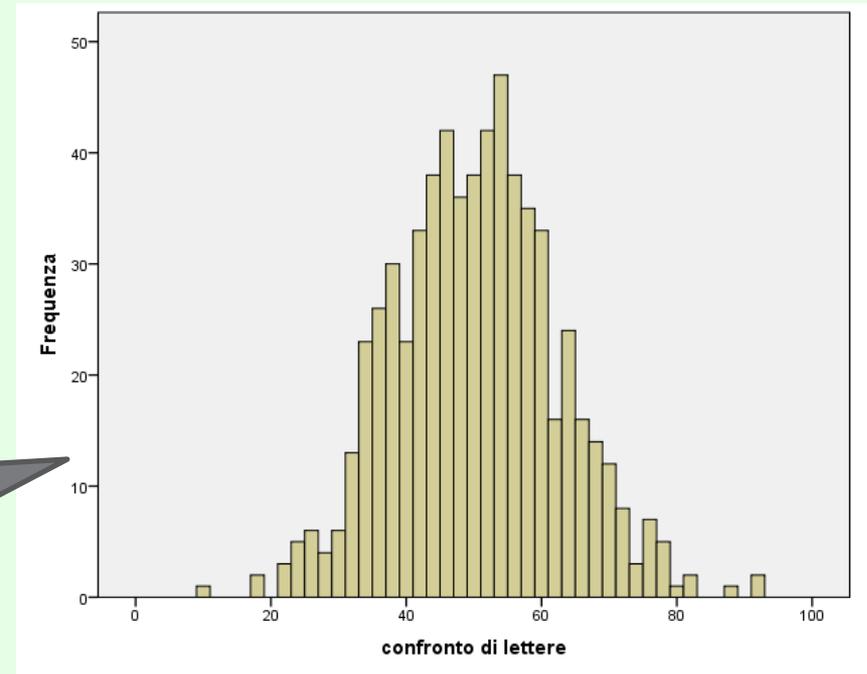
consiglio di orientamento in terza media



Scale a intervalli (Test psicometrici)

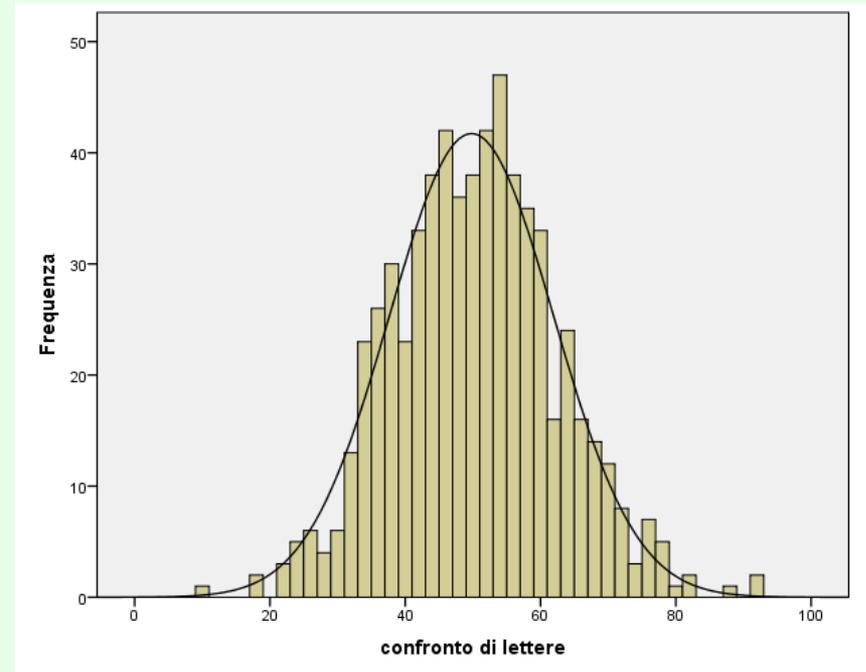
Con i test psicometrici,
la **forma** della
**distribuzione dei
punteggi** diventa
rilevante.

Esempio: Distribuzione dei
punteggi del test di Confronto
di lettere (prima prova del
GATB)



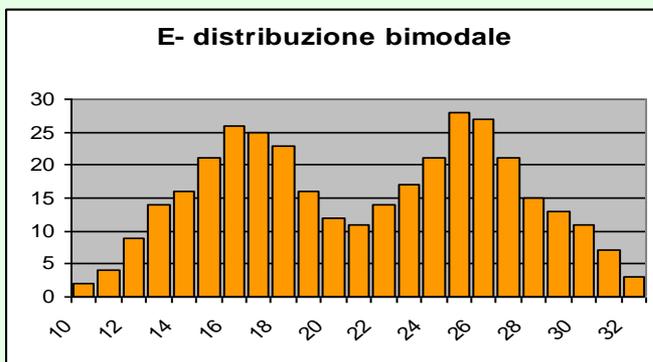
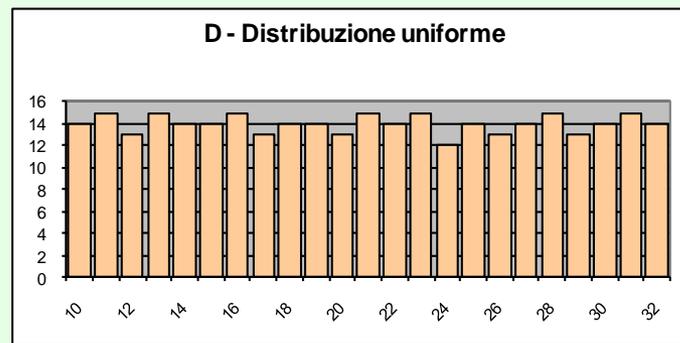
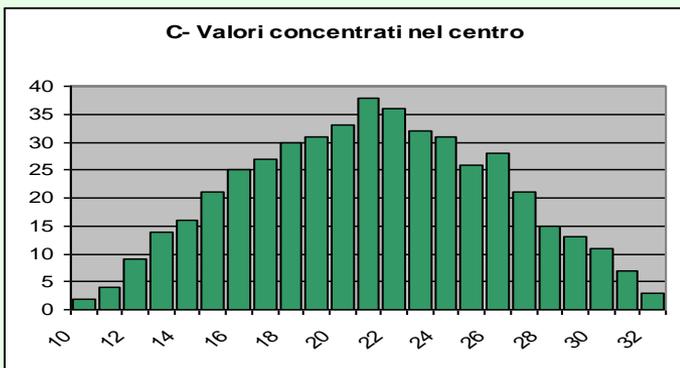
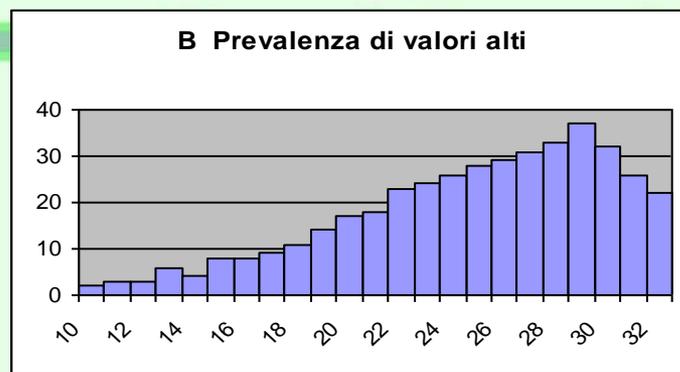
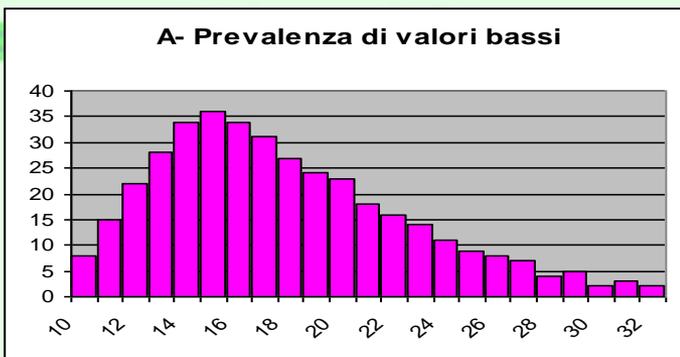
Vorremmo che la distribuzione dei punteggi seguisse la **curva normale (di Gauss)**

Distribuzione dei punteggi del test di Confronto di lettere (prima prova del GATB)



I punteggi non sempre seguono la distribuzione normale

Le misurazioni sono concentrate sul lato sinistro (A), sul lato destro (B), al centro (C), da nessuna parte (D) o in due punti (E).



Quante classi e quali intervalli?

- Il software che elabora i dati pensa a risolvere per noi il problema di quanti intervalli occorre usare per rappresentare graficamente una variabile continua.

L'ogiva

ovvero

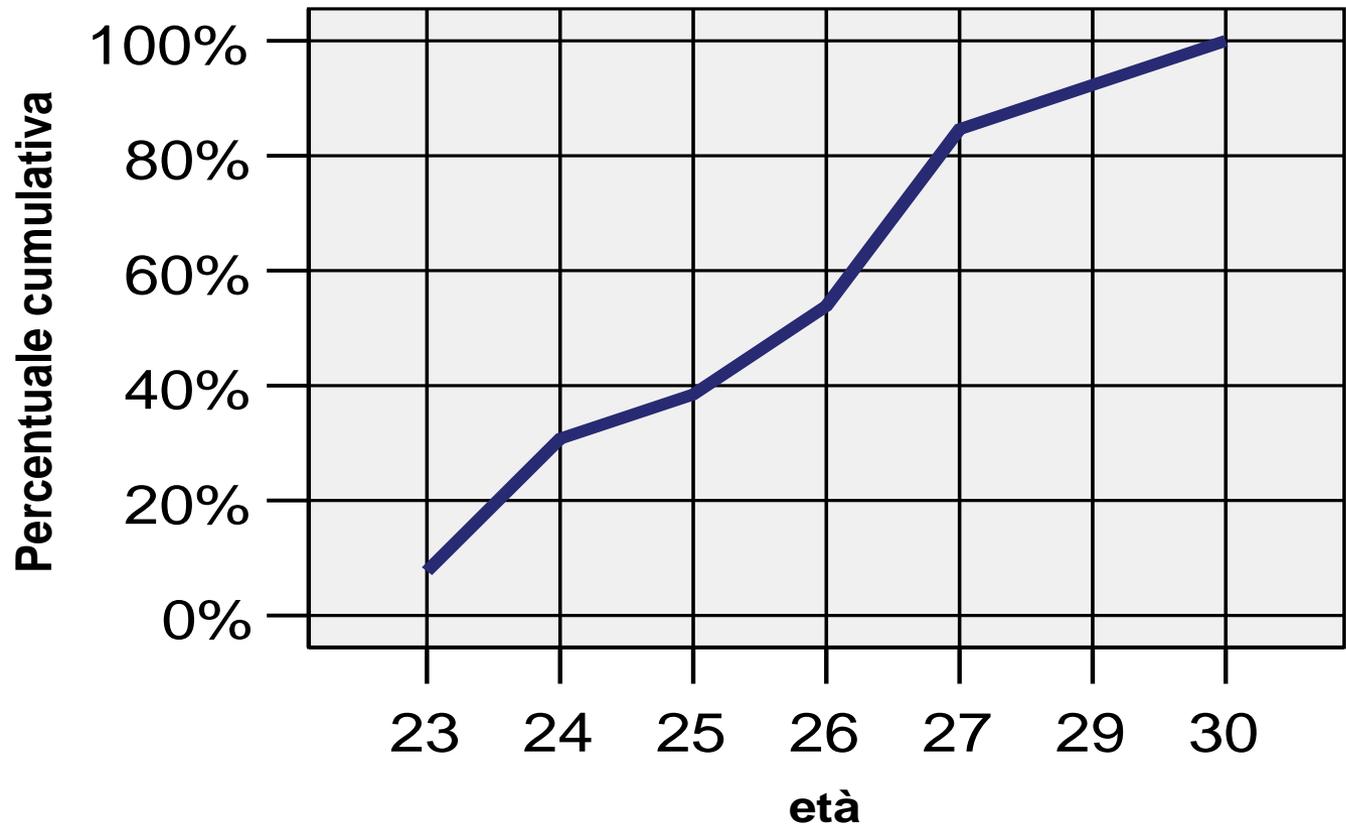
La distribuzione **cumulativa** delle
frequenze

La distribuzione cumulativa delle frequenze

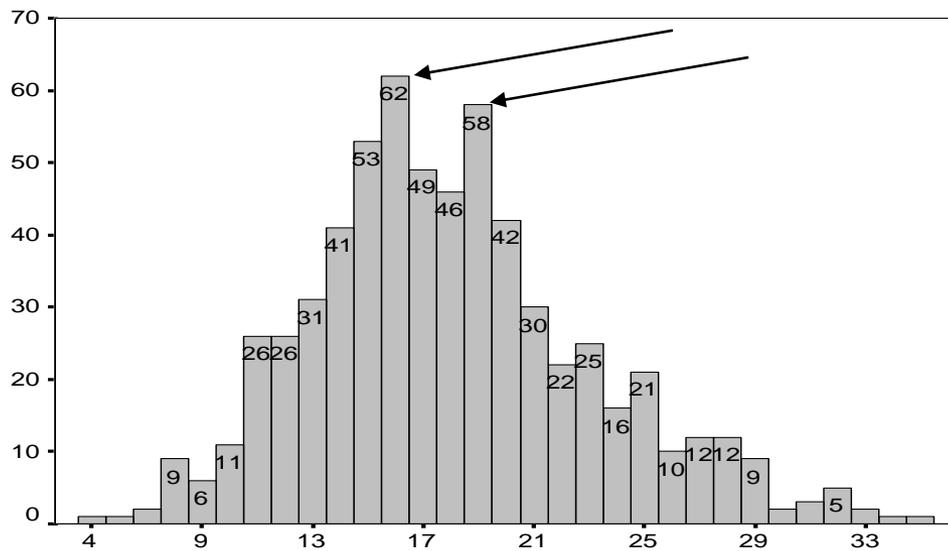
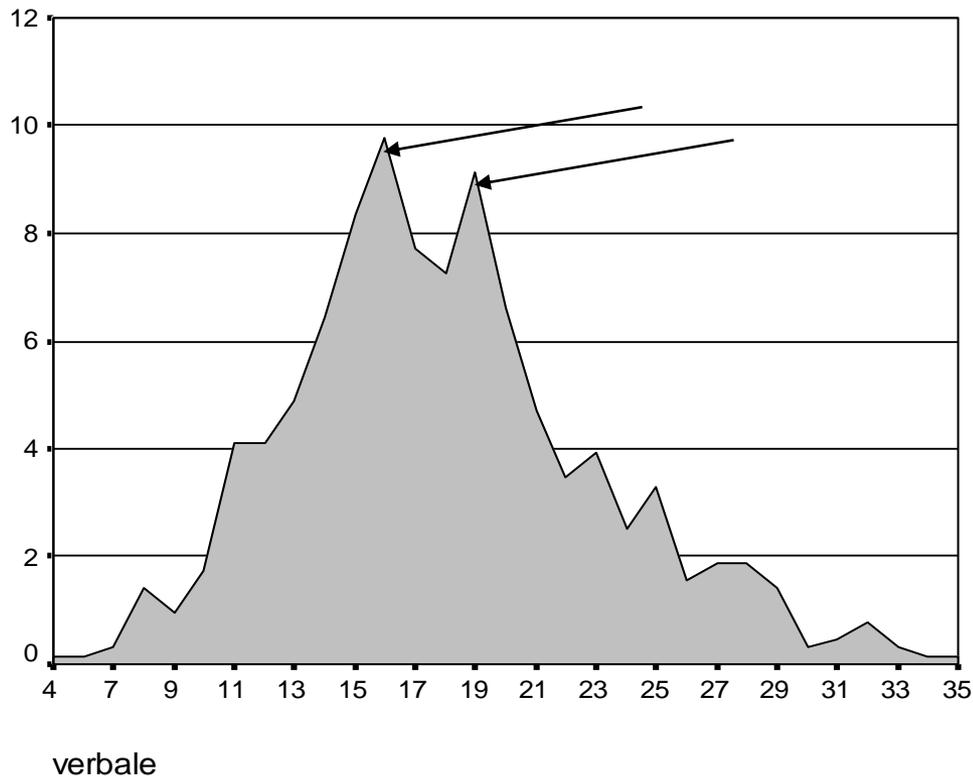
Sommando le frequenze dai valori più bassi ai valori più alti, si ottiene la **distribuzione cumulativa (o cumulata)** delle frequenze:

Età osservata in 13 persone	Frequenza	Percentuale	Frequenza cumulativa	Percentuale cumulativa
23	1	7.7	1	7.7
24	3	23.1	4	30.8
25	1	7.7	5	38.5
26	2	15.4	7	53.8
27	4	30.8	11	84.6
29	1	7.7	12	92.3
30	1	7.7	13	100.0
Totale	13	100.0	13	100.0

Per rappresentare la distribuzione cumulativa di frequenza si usa il **poligono**; Si usano le percentuali piuttosto che le frequenze assolute:



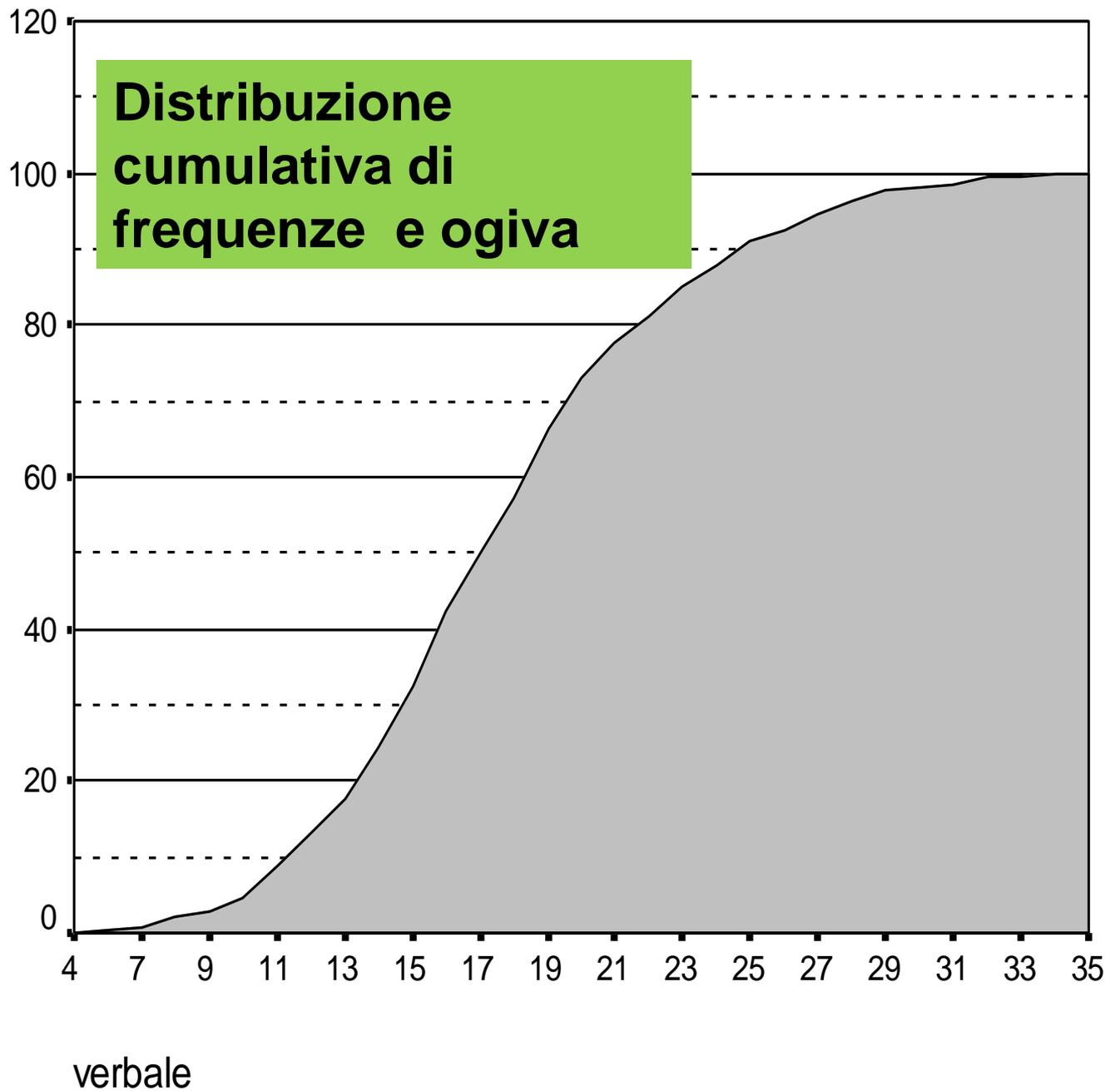
Ogni punto rappresenta un valore della variabile abbinato alla sua percentuale cumulativa. L'unione di tutti i punti dà questo segmento spezzato detto **ogiva**.

A**B**

Istogramma delle frequenze (A) e poligono di frequenza (B) relativi ai punteggi della tabella precedente.

Nella tabella sono mostrati i punteggi relativi a un test di abilità verbale eseguito da 635 studenti di terza media

Punteggio rilevato	Frequenza	Percentuale	Percentuale cumulativa
4	1	0,2	0,2
5	1	0,2	0,3
7	2	0,3	0,6
8	9	1,4	2
9	6	0,9	3
10	11	1,7	4,7
	(abbreviati)		
29	9	1,4	97,8
30	2	0,3	98,1
31	3	0,5	98,6
32	5	0,8	99,4
33	2	0,3	99,7
34	1	0,2	99,8
35	1	0,2	100
Totale	635	100	



[Grafico relativo alla tabella precedente]

I ranghi e i ranghi percentili

I **ranghi** sono dei numeri che indicano la posizione di ciascuna osservazione in rapporto alle altre. Il rango 1 può essere attribuito o al numero più piccolo (disposizione crescente) o al numero più grande (disposizione decrescente). Ogni volta è necessario specificare quale scelta si fa.

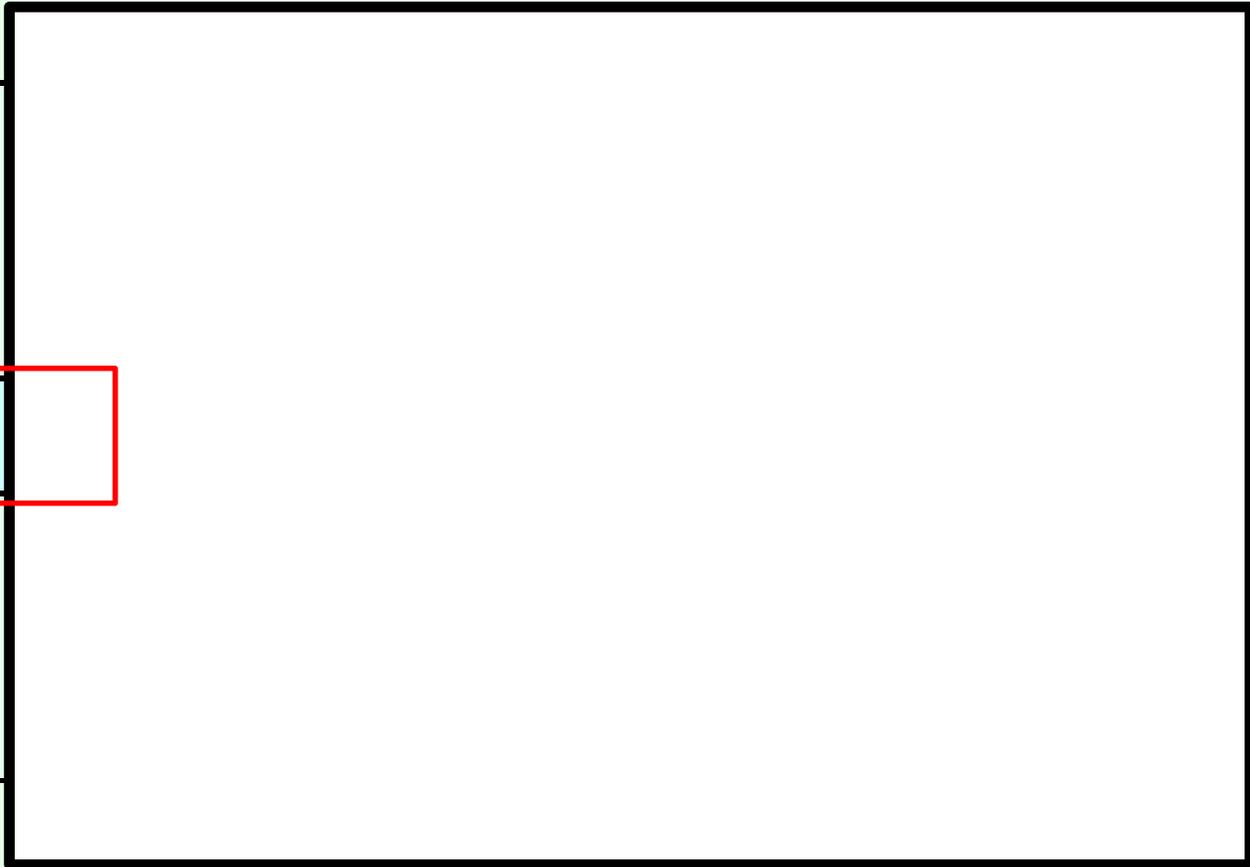
In questa tabella l'osservazione 4 ha **rango a** uguale a **1** perché, ordinando in maniera crescente i punteggi, risulta in prima posizione; **rango b** uguale a **7** perché, ordinando in maniera decrescente, risulta in ultima posizione.

OSSERVAZIONI	RANGO A	RANGO B
4	1	7
18	2	6
19	3	5
20	4	4
22	5	3
50	6	2
120	7	1

Come si procede con due osservazioni uguali ?

Conversione in ranghi dei

Osserv.	rango	Osserv. con due valori
4	1	4
18	2	18
19	3	19
		19
20	4	20
22	5	22
50	6	50
120	7	120



Nel caso di due o più osservazioni uguali si procede calcolando la **media** dei ranghi iniziali, attribuiti inizialmente allo stesso punteggio.

Conversione in ranghi dei punteggi ordinati

Osserv.	rango	Osserv. con due valori	ranghi iniziali	equivalenti	ranghi finali
4	1	4	1		1
18	2	18	2		2
19	3	19	3	3,5	3,5
		19	4	3,5	3,5
20	4	20	5		5
22	5	22	6		6
50	6	50	7		7
120	7	120	8		8

Come si procede con osservazioni uguali?

Un punteggio che appare più volte riceve la media dei ranghi iniziali

Conversione in ranghi dei punteggi ordinati

Osservazioni con tre valori	ranghi iniziali	equivalenti	ranghi finali
4	1		1
18	2		2
19	3	4	4
19	4	4	4
19	5	4	4
20	6		6
22	7		7
50	8		8
120	9		9

Come si procede con osservazioni uguali?

Un punteggio che appare più volte riceve la media dei ranghi iniziali

Conversione in ranghi dei punteggi ordinati

Osserv.	rango	Osserv. con due valori	ranghi equivalenti		ranghi finali	Osserv. con tre valori	ranghi equivalenti		ranghi finali
			iniziali				iniziali		
4	1	4	1		1	4		1	
18	2	18	2		2	18		2	
19	3	19	3	3,5	3,5	19	3	4	4
			4	3,5	3,5		4	4	4
					5		4	4	
20	4	20	5		5	20		6	
22	5	22	6		6	22		7	
50	6	50	7		7	50		8	
120	7	120	8		8	120		9	

Il **rango percentile** indica la posizione di un'osservazione all'interno di un insieme prescindendo dalla numerosità del campione (come le percentuali).

Il percentile 70 è, per esempio, il valore superiore al 70 % dei casi, e inferiore al 30% dei casi.

Ranghi percentili		
Punto grezzo	rango	Rango centile o percentile
2	12	4,17
4	11	12,50
18	10	20,83
19	8	37,50
19	8	37,50
19	8	37,50
20	6	54,17
22	5	62,50
28	4	70,83
38	3	79,17
50	2	87,50
120	1	95,83

Ranghi percentili

Si calcola con questa formula :

Formula di calcolo

N = Numero delle osservazioni

G = Rango

RC = Rango percentile (o centile)

$$RC = 100 \cdot \frac{(N - (G - 0,5))}{N}$$

Punto grezzo	rango	Rango centile o percentile
2	12	4,17
4	11	12,50
18	10	20,83
19	8	37,50
19	8	37,50
19	8	37,50
20	6	54,17
22	5	62,50
28	4	70,83
38	3	79,17
50	2	87,50
120	1	95,83

Questa colonna mostra che le percentuali cumulative sono quasi uguali ai ranghi percentili

osservazione	rango ascendente	(quasi) percentile	rango percentile
1002	1	10	0,5
1004	2	20	10,5
1018	3	30	20,5
1019	4	40	30,5
1020	5	50	40,5
1021	6	60	50,5
1022	7	70	60,5
1026	8	80	70,5
1028	9	90	80,5
1036	10	100	90,5

I ranghi percentili non hanno mai un valore pari a 100

Gli stessi valori della tabella precedente, ma ordinati in senso inverso

osservazione	rango discendente	rango percentile
1002	10	90,5
1004	9	80,5
1018	8	70,5
1019	7	60,5
1020	6	50,5
1021	5	40,5
1022	4	30,5
1026	3	20,5
1028	2	10,5
1036	1	0,5

Come interpretare i dati usando l'ogiva e i percentili

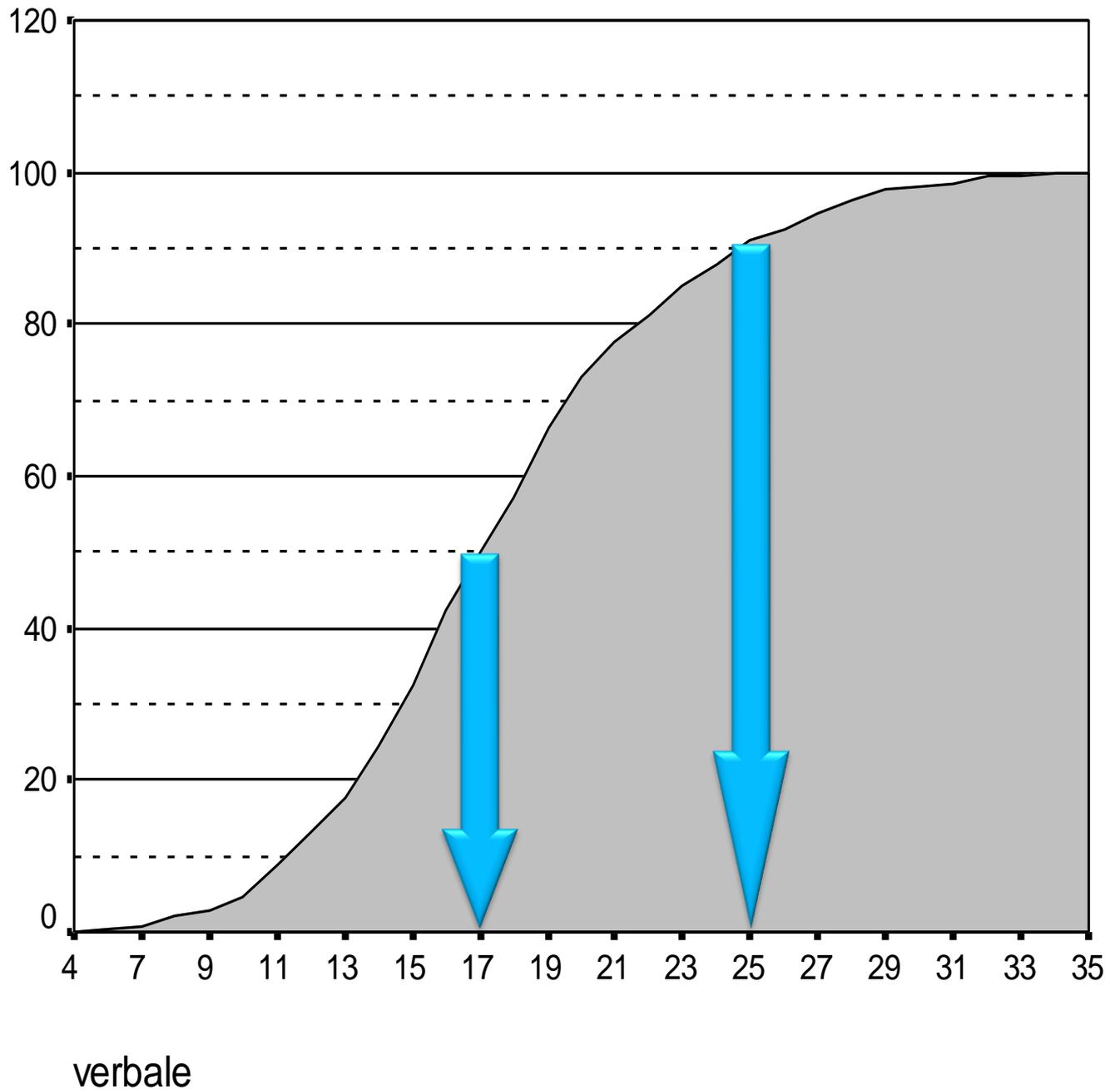
Distribuzione cumulativa di frequenze e ogiva

Il punteggio 17 o inferiore è ottenuto all'incirca da 50 % dei soggetti; un punteggio uguale o inferiore a 25 è ottenuto dal 90% dei soggetti.

Percentili	5	10	25	50	75	90	95
Punteggio di abilità verbale	11	12	15	17	21	25	28

Distribuzione cumulativa di frequenze e ogiva

[Grafico relativo alla tabella precedente]



LA MEDIANA

dove N è il numero di osservazioni

Mediana (Mdn): in un insieme di dati ordinati, è quel valore che occupa la posizione centrale nella distribuzione dei dati stessi.

Divide in due metà (50% inferiore e 50% superiore) la distribuzione dei dati:

Mdn=

$$\frac{N + 1}{2}$$

Il modo di calcolare la mediana dipende dal numero pari o dispari del campione

	punto grezzo ordinato	rango	mediana	
con N dispari	2	1		
	5	2		
	6	3		
	9	4	mediana = 9	
	10	5		
	11	6		
	15	7		
con N pari	2	1		
	5	2		
	6	3	mediana = 7,5	
	9	4		
	10	5		
	11	6		

Esempio:

- N dispari = 7
{2,5,6,9,10,11,15}

$$\text{Mdn} = \frac{N + 1}{2} = \frac{7 + 1}{2} = 4$$

$N_4 = 9$ (è nella quarta posizione)

- N pari = 6
{2,5,6,9,10,11}

$$\text{Mdn} = \frac{N + 1}{2} = \frac{6 + 1}{2} = 3,5$$

La posizione 3,5 si colloca tra i valori 6 e 9 e dunque è necessario calcolare la media aritmetica tra i due valori $6+9=7,5$; la mediana in questo caso non esiste e si pone come convenzione pari al valore medio fra le due posizioni.

I QUANTILI

Se dividendo una distribuzione a metà otteniamo la mediana, dividendola in quattro parti otterremo tre valori corrispondenti ai **quartili**. Allo stesso modo, dividendo in tre parti si ottengono i **terzili**, in cinque i **quintili**, in cento i **centili**, e così via.



Equivalenze

Q1 (primo quartile) = 25° percentile

Q2 (secondo quartile) = 50° percentile = Mediana

Q3 (terzo quartile) = 75° percentile

Media, varianza e deviazione standard

MEDIA: valore centrale di un insieme di dati, dato dalla somma di tutti i valori della distribuzione diviso per il numero totale dei casi che la compongono. La media è un concetto astratto, che però è entrato bene nel senso comune, e non presenta difficoltà di comprensione

$$M = \frac{\sum x_i}{N}$$

Media, varianza e deviazione standard

VARIANZA: Media (al quadrato) delle distanza delle osservazioni dalla media

somma degli scarti dalla media elevati al quadrato e divisi per il numero totale dei casi.

$$s^2_x = \frac{\sum (x_i - M)^2}{N} \quad \text{oppure, semplificando} \quad s^2_x = \frac{\sum (x_i)^2}{N} - M^2$$

Media, varianza e deviazione standard

DEVIAZIONE STANDARD: : **Media delle distanza delle osservazioni dalla media.** ottenibile tramite la radice quadrata della varianza.

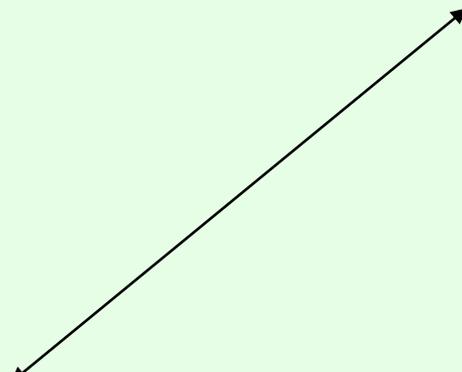
Poiché per calcolare la varianza abbiamo elevato al quadrato le distanze dalla media, allora riconvertiamo tali distanze calcolando la radice quadrata della media delle distanze

$$s = \sqrt{s^2} = \sqrt{\frac{\sum (x_i - M)^2}{N}}$$

Calcolo di media, varianza e deviazione standard

Soggetto	Punteggio	Scarto dalla media	Quadrato dello scarto	Quadrato del punteggio
1 Alessio	2	-2	4	4
2 Bruno	3	-1	1	9
3 Carlo	4	0	0	16
4 Davide	2	-2	4	4
5 Enrico	9	5	25	81
Totale	20	0	34	114
Media	4	0	6,8	22,8

Media	4
Deviazione Standard	2,61
Varianza	6,8



Media dei quadrati degli scarti

Calcolo della varianza e deviazione standard- primo modo

Soggetto	Punteggio	Scarto dalla media	Quadrato dello scarto
1 Alessio	2	-2	4
2 Bruno	3	-1	1
3 Carlo	4	0	0
4 Davide	2	-2	4
5 Enrico	9	5	25
Totale	20	0	34
Media	4	0	6,8

Media	4
Deviazione Standard	2,61
Varianza	6,8

➔ Media dei quadrati degli scarti

$$s^2_x = \frac{\sum (x_i - M)^2}{N} = \frac{34}{5} = 6,8$$

Calcolo della varianza e deviazione standard- secondo modo

Soggetto	Punteggi o	Quadrato del punteggio
1 Alessio	2	4
2 Bruno	3	9
3 Carlo	4	16
4 Davide	2	4
5 Enrico	9	81
Totale	20	114
Media	4	22,8



media dei **quadrati** meno **quadrato** della media.

Media	4
Deviazione Standard	2,61
Varianza	6,8

$$s^2_x = \frac{\sum(x_i)^2}{N} - M^2$$

$$6,8 = \frac{114}{5} - 4^2$$

Altri indici descrittivi

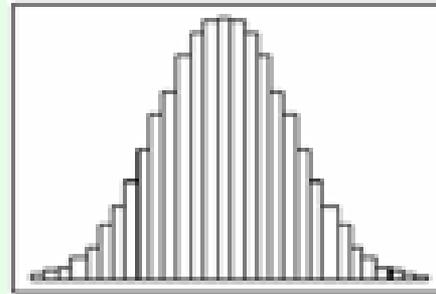
L'indice di **asimmetria** (English: Skewness)
l'indice di **curtosi** (English: Kurtosis),

Quando si situano attorno allo zero indicano
asimmetria e piattezza simile a quella della curva
normale

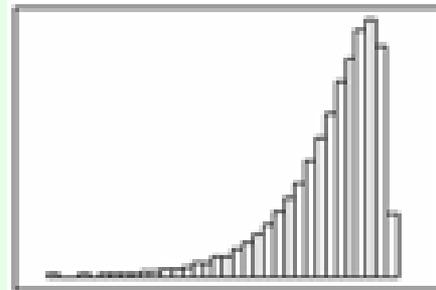
Il loro uso è ancora controverso, ma si può
affermare che valori oltre 2 in senso assoluto
sono indicatori di un andamento non gaussiano
dei valori osservati

Indice di asimmetria

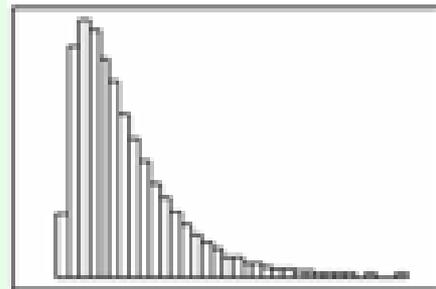
- Se diverso da zero...
- indica che la distribuzione dei punteggi ha una coda a sinistra (asimmetria negativa)...
- ...oppure a destra (asimmetria positiva)



Symmetric
Bell shaped



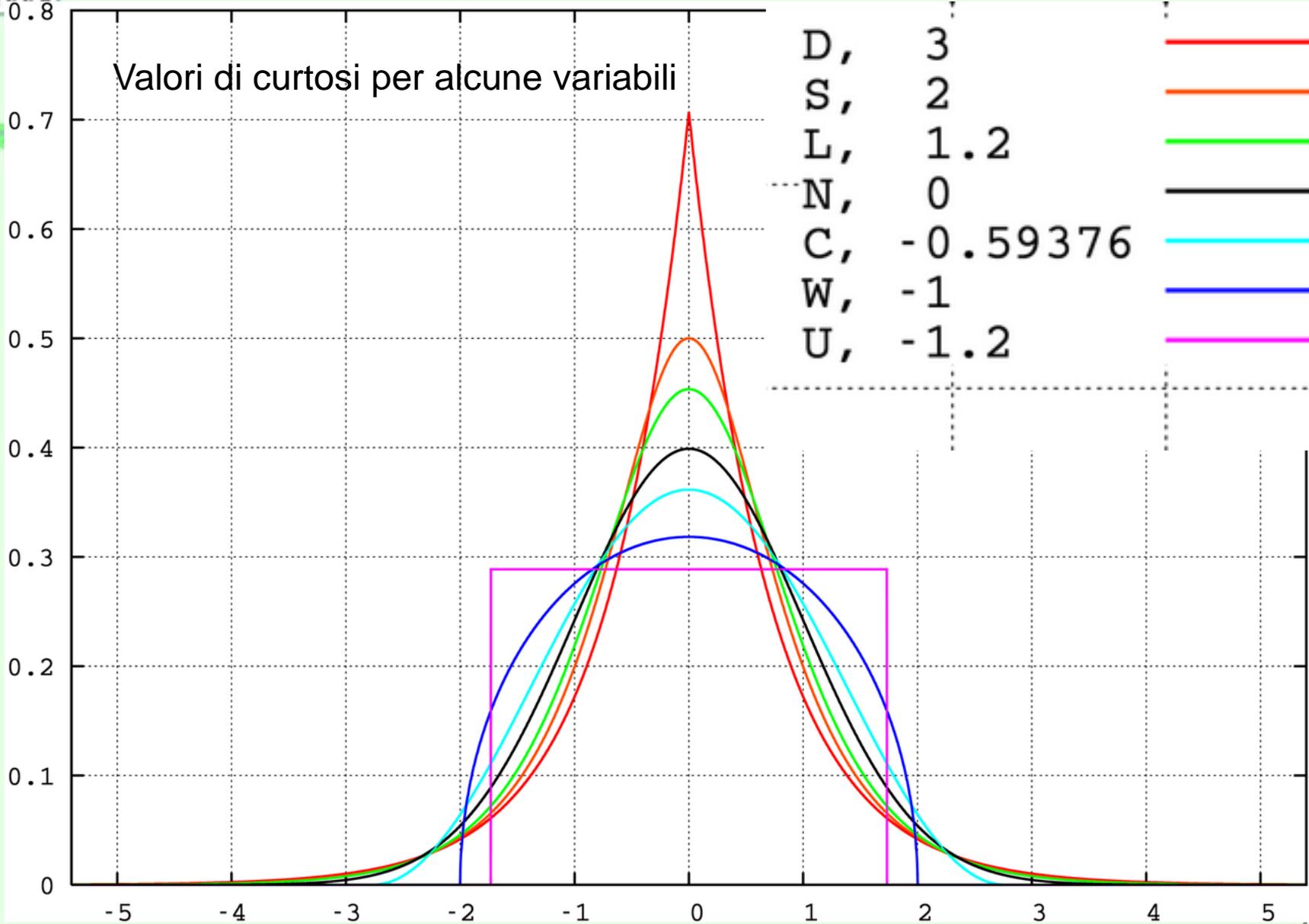
Skewed to
the Left



Skewed to
the Right

Valori di curtosi per alcune variabili

D,	3	
S,	2	
L,	1.2	
N,	0	
C,	-0.59376	
W,	-1	
U,	-1.2	



Quattro variabili del GATB 1

Statistiche descrittive

	N	Intervallo	Minimo	Massimo	Media	Dev std.	Varia nza
	Statistica	Statistica	Statistica	Statistica	Statistica	Statistica	Statistica
confronto di lettere	635	82	10	92	49,79	12,140	147,37
calcoli	635	27	3	30	19,75	4,175	17,429
visualizzazione spaziale	635	32	3	35	17,16	4,586	21,033
vocabolario	635	31	4	35	18,03	5,157	26,594
N. valido (a livello di elenco)	635						

Quattro variabili del GATB 2

Statistiche descrittive

	Media	Deviazione std.	Varianza	Asimmetria	Curtosi
confronto di lettere	49,79	12,140	147,37	,150	,191
calcoli	19,75	4,175	17,429	-,615	,815
visualizzazione spaziale	17,16	4,586	21,033	,328	,355
vocabolario	18,03	5,157	26,594	,490	,212

Indici di asimmetria e curtosi inferiori a $|1|$: possono essere considerati indici di normalità

Dopo l'esame delle caratteristiche della curva normale, torneremo sugli indici di asimmetria e curtosi