

## LA STATISTICA DESCRITTIVA

Consente di ottenere la descrizione di un **campione** di soggetti o eventi, in termini di **variabili** o **combinazioni di variabili**.

La descrizione statistica si articola in due fasi:

- la **rilevazione** dei dati
- la **elaborazione** dei dati

## LA RILEVAZIONE DEI DATI

Consiste nella **misurazione** (**quantificazione** o **codifica**), attraverso strumenti adeguati (colloqui, osservazioni, questionari, test ecc.) dei valori che la variabile assume per i diversi elementi che costituiscono il campione esaminato.

Il tipo di dati che si ottengono dalla fase di rilevazione dipende dal livello di misura consentito dalla variabile in esame. Tali dati vengono chiamati **grezzi**, in quanto su di essi non è stata ancora effettuata nessuna elaborazione.

I dati grezzi, una volta raccolti, vengono **tabulati** in una **tabella a doppia entrata** (dove ogni riga corrisponde ad un soggetto del campione ed ogni colonna ad una variabile misurata) per facilitare la loro lettura e per consentire la loro immissione in un computer per le successive elaborazioni.

## Esempio di tabulazione dei dati in una tabella a doppia entrata

Sog.	VARIABILI					
	Sesso	Reddito	Scolarità	Motivazione	Punt. Ansi	T.R. Stim. Lu
	(1)	(2)	(3)	(4)	(5)	(6)
1	0	2°	12	ALTA	0	250
2	1	10°	8	BASSA	2	340
3	1	6°	13	MEDIA	4	290
4	0	9°	16	BASSA	3	310
5	1	1°	17	ALTA	4	420
6	0	8°	8	BASSA	7	580
7	1	3°	16	MEDIA	5	380
8	0	5°	12	MEDIA	2	290
9	0	4°	8	MEDIA	8	610
10	1	7°	12	BASSA	6	470

### Legenda

- (1) Sesso: 0 = Maschio; 1 = Femmina.  
Variabile su **scala nominale** (mutabile)
- (2) Graduatoria per il reddito.  
Variabile su **scala ordinale**
- (3) Scolarità, misurata in anni di studio.  
Variabile su **scala a rapporti equivalenti**
- (4) Motivazione a partecipare alla ricerca (da questionario)  
Variabile su **scala ordinale**
- (5) Punteggio di ansia, ottenuto tramite test (da 0 = poco ansioso a 8 = molto ansioso).  
Variabile su **scala ad intervalli equivalenti**
- (6) Tempi di reazione semplici a stimoli luminosi (in millisecondi).  
Variabile su **scala a rapporti equivalenti**

## L'ELABORAZIONE DEI DATI

Nell'ambito della **statistica descrittiva** le principali tecniche di elaborazione sono:

- 1) le rappresentazioni grafiche;
- 2) il calcolo delle misure di tendenza centrale e di variabilità;
- 3) le trasformazioni dei punteggi grezzi.

### Le rappresentazioni grafiche

Le rappresentazioni grafiche tipiche della statistica descrittiva riguardano le **distribuzioni di frequenza** e permettono di rappresentare la frequenza con la quale si presentano i diversi valori della variabile in esame.

### Il grafico a barre e la "torta"

Si utilizzano per rappresentare distribuzioni di frequenza di **variabili quantitative discrete** o di **variabili qualitative**.

### L'istogramma e il poligono di frequenza

Si utilizzano per rappresentare distribuzioni di frequenza di **variabili continue**.

### L'ogiva

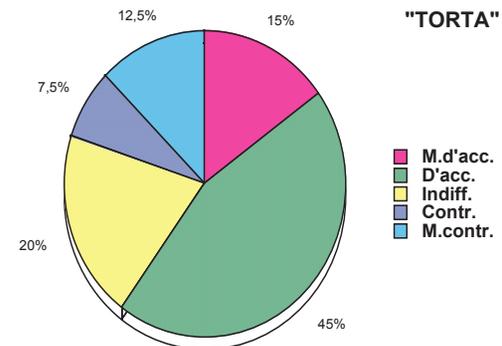
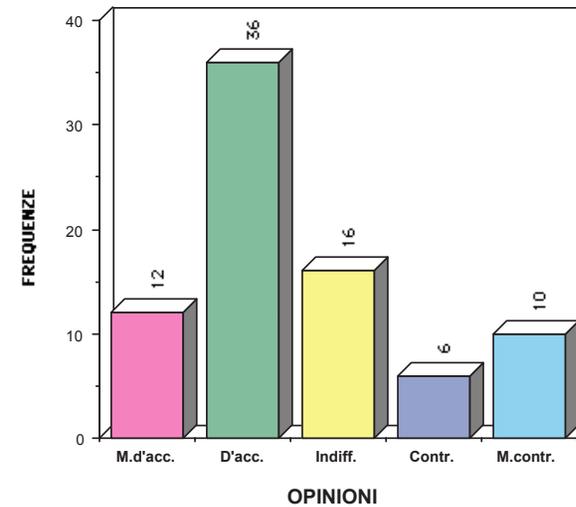
Si utilizza per rappresentare distribuzioni di frequenza **cumulative**.

Altri tipi di grafici (diagrammi di dispersione, grafici di funzione, grafici a serie temporale) vengono utilizzati soprattutto nell'ambito della verifica delle ipotesi (**statistica inferenziale**) e saranno esaminati in seguito.

Variabile: Opinione su una proposta di legge, valutata tramite sondaggio

Valori:	M. d'acc.	D'acc.	Indiff.	Contr.	M. contr.
Freq.:	12	36	16	6	10
Perc. :	15.0%	45.0%	20.0%	7.5%	12.5%

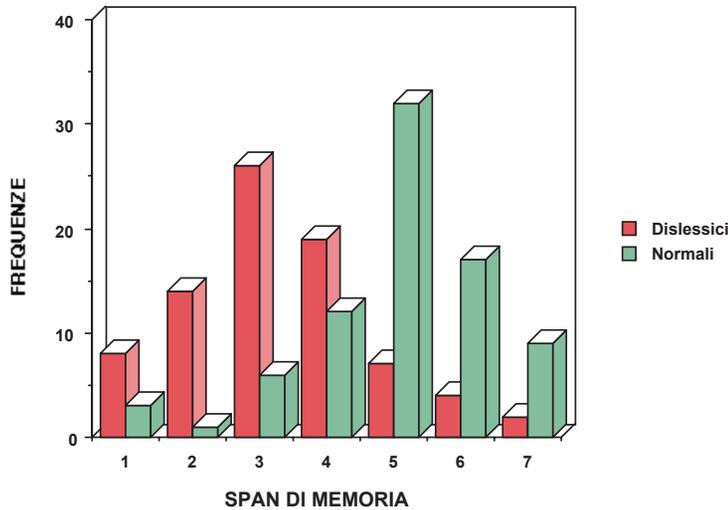
GRAFICO A BARRE



VARIABILE: Ampiezza dello span di memoria di un gruppo di dislessici (N=80) e di un gruppo di bambini con normali abilità di lettura (N=80).

Valori:	1	2	3	4	5	6	7
Freq. Dislessici:	8	14	26	19	7	4	2
Freq. Normali:	3	1	6	12	32	17	9

GRAFICO A BARRE



## ISTOGRAMMA

Consente di rappresentare distribuzioni di frequenza di **variabili continue**

### Differenze con il grafico a barre:

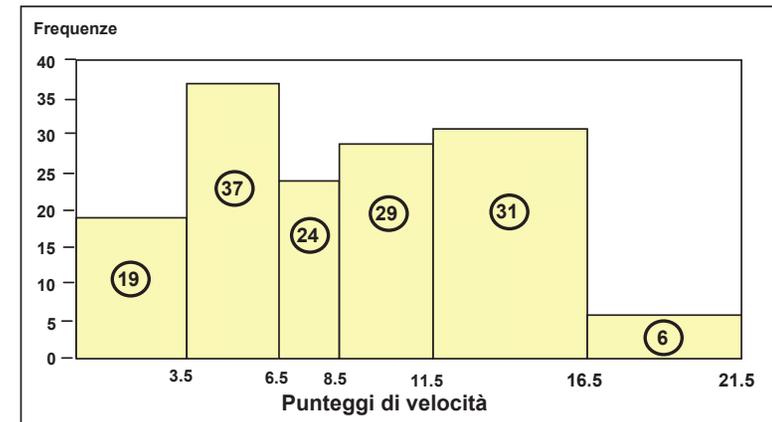
- 1) Sulla scala dei valori (ascissa) si indicano i **limiti reali delle classi**
- 2) I rettangoli devono essere **contigui**
- 3) La **frequenza** di ogni classe è rappresentata dall'**area del rettangolo** e non dalla sua altezza (l'altezza si ricava dividendo la frequenza per l'ampiezza della classe).

Dire se il seguente istogramma è costruito correttamente:

VARIABILE: Punteggio di velocità in una prova di denominazione di oggetti (n° oggetti denominati / tempo in sec. x 60) relativo a bambini di prima elementare (N = 146).

### Variabile continua su scala a rapporti equivalenti

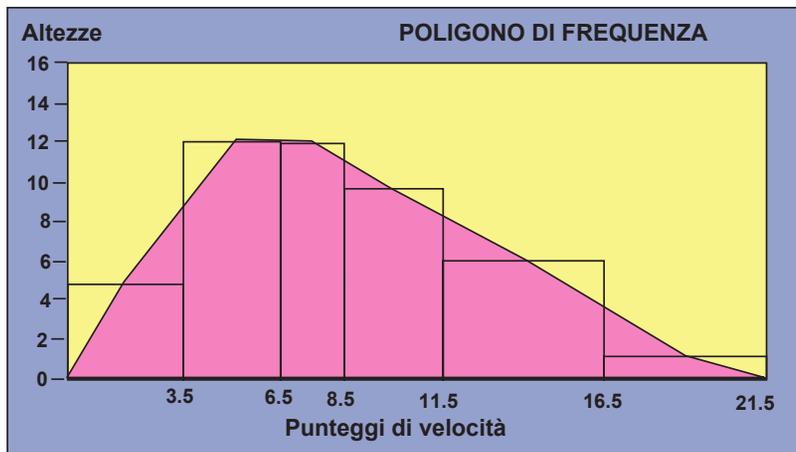
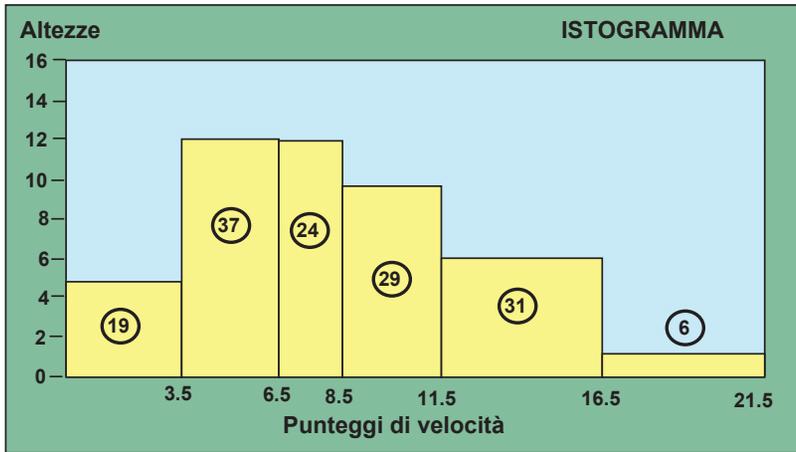
Classi di valori:	0-3	4-6	7-8	9-11	12-16	17-21
Frequenze:	19	37	24	29	31	6



Il grafico è costruito in modo errato perchè le aree dei rettangoli non sono proporzionali alle frequenze. La 4<sup>a</sup> e la 5<sup>a</sup> classe hanno frequenza simile (29 e 31), eppure la 5<sup>a</sup> classe sembra avere frequenza quasi doppia rispetto alla 4<sup>a</sup>. Quando le classi sono di ampiezza diversa, ogni frequenza va divisa per l'ampiezza della classe, altrimenti il grafico risulterà distorto.

**Sempre sugli stessi dati (procedura corretta):**

Classi di valori:	0-3	4-6	7-8	9-11	12-16	17-21
<b>Ampiezza (i):</b>	<b>4</b>	<b>3</b>	<b>2</b>	<b>3</b>	<b>5</b>	<b>5</b>
Frequenze:	19	37	24	29	31	6
<b>Altezze (Freq/i):</b>	<b>4.75</b>	<b>12.3</b>	<b>12.0</b>	<b>9.7</b>	<b>6.2</b>	<b>1.2</b>



POLIGONO DI FREQUENZA: Si ottiene unendo con una linea continua i punti centrali dei lati superiori dei rettangoli.

Quando si suddividono i valori in classi della stessa ampiezza è consentito rappresentare le frequenze con le altezze dei rettangoli, in quanto queste sono comunque proporzionali alle aree. Il seguente istogramma è costruito sugli stessi dati del precedente esempio, ma considerando classi di ampiezza 4.

L'**ogiva** si costruisce segnando un punto all'incrocio tra il limite reale superiore della classe e la frequenza cumulata corrispondente, ed unendo i vari punti.

Classi di valori:	0-3	4-7	8-11	12-15	16-19	20-23
Frequenze:	19	47	43	25	10	2
Freq. cumulate:	19	66	109	134	144	146

