Università di Ferrara
fondata nel 1391

Elementi di statistica medica

Misure di sintesi numerica

MISURAZIONE E MISURA

- Misurazione: insieme di operazioni che portano alla determinazione del valore del misurando.
- Misura: valore del misurando ottenuto in seguito ad una misurazione (l'unità di misura deve sempre essere espressa).
- Valore vero: è quello che si otterrebbe con una misurazione perfetta (senza errori.)

METODI DI MISURAZIONE

- Metodo di misurazione diretto: la misura è ottenuta mediante l'uso di uno strumento atto alla misurazione della grandezza 'X' del misurando (stessa unità di misura).
- Metodo di misurazione indiretto: il risultato della misura è espresso in termini di valori di altre grandezze (e diverse unità di misura) con una relazione nota tra loro (ad esempio la misura di una grandezza fisica si dice indiretta se viene effettuata attraverso il calcolo di una espressione matematica. v=s/t).

Misurazione Indiretta

Alcuni esempi di misura con il metodo indiretto:

- misura della pressione tramite la misura dell'altezza di una colonna di liquido (es. barometro a mercurio);
- misura della temperatura tramite la misura di una resistenza elettrica (es. termometro a termoresistenza);
- misura del glucosio nel sangue tramite la misura della colorazione di cartine reagenti (es. glucometri).

Questo tipo di misurazione è ormai molto diffusa: è infatti evidente che tutte le misure fatte con l'ausilio di trasduttori e altri sensori sono delle misure indirette; in particolare, quasi tutti trasformano il misurando in una grandezza elettrica (tipicamente una tensione) che poi viene letta e interpretata da uno strumento elettronico.

Tabella di frequenza

Classe	Frequenza assoluta
0 - 2	20
3 - 5	14
6 - 8	15
9 - 11	2
12 - 14	1

Limiti inferiori di classe

Limiti inferiori

Classe	Frequenze
0 - 2	20
3 - 5	14
6 - 8	15
9 - 11	2
12 - 14	1

- Sono i valori più piccoli che possono effettivamente appartenere alla classe.

Limiti superiori di classe

Limiti superiori

Classe	Frequenze
~~0 -~~ 2	20
~~3 -~~ 5	14
~~6 -~~ 8	15
~~9 -~~ 11	2
~~12 -~~ 14	1

- Sono i valori più grandi che possono effettivamente appartenere alla classe.

Valore centrale di classe

Classe			Frequenza
0 –	1	2	20
3 –	4	5	14
6 –	7	8	15
9 –	10	11	2
12 –	13	14	1

Valori Centrali di classe

Ampiezza di classe

- E’ la differenza fra due limiti superiori (inferiori) consecutivi

	Classe	Frequenza
3	0 - 2	20
3	3 - 5	14
3	6 - 8	15
3	9 - 11	2
3	12 - 14	1

MISURAZIONI

Una misura deve essere:

- Attendibile
- Precisa
- Ripetibile
- Controllabile

L’ **ERRORE**

- L’errore è il risultato di una misurazione meno il valore vero del misurando.
- L'errore casuale o statistico è qualsiasi errore di misurazione che può incidere con la stessa probabilità in aumento o in diminuzione sul valore misurato.

- Fatta la misura delle grandezze di un certo fenomeno, si sente la necessità di sintetizzare la distribuzione mediante valori che la caratterizzino e permettano di confrontarla con distribuzioni di fenomeni analoghi osservati in tempi o in luoghi diversi.
- Un primo di questi valori è dato da un valore medio che esprime una tendenza centrale.

INDICI DI POSIZIONE

- Descrivono ed evidenziano alcune caratteristiche della popolazione in esame identificando alcuni elementi della popolazione stessa.

I principali indici di posizione sono:

- Media
- Mediana
- Moda
- Percentili

Concetto e tipo di medie

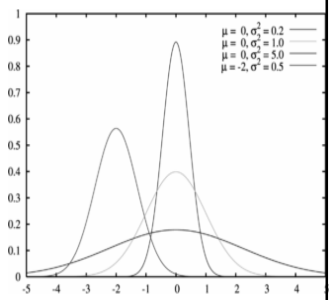
Il valore medio è un valore che esprime una tendenza centrale.

Secondo ***Cauchy*** *la media di un insieme è un valore compreso tra il minimo e il massimo.*

- Si può chiamare **media di una distribuzione x_1, x,....x_n,** rispetto ad una funzione *x_1, x_2,... x_n),* quella quantità che sostituita alle x_i nella funzione lascia invariato il risultato.
- In statistica si distinguono di solito due tipi di medie.
  - **medie di calcolo*:* sono quelle che si calcolano tenendo conto di tutti i valori della distribuzione;**
  - **medie di posizione; sono quelle che si calcolano tenendo conto solo di alcuni valori della distribuzione**

Misure di sintesi numerica

- La caratteristica più comunemente studiata in una serie di dati è il suo centro, ovvero dove le osservazioni tendono a raccogliersi.
- La **media** di un insieme è un qualsiasi valore compreso tra il minimo e il massimo.



Media

- La misura di tendenza centrale più comunemente utilizzata è la **media aritmetica** o media.
- La media è calcolata sommando tutte le osservazioni in una serie di dati e dividendo per il numero totale delle osservazioni.
- La media ci dice all'incirca l'ordine di grandezza (la posizione sulla scala dei numeri, appunto) dei valori esistenti.

MEDIA

Ad esempio:

- Popolazione 1 5 8 9 6 4 8 5 2 8
- Numero di osservazioni: 10
- Somma: 1+5+8+9+6+4+8+5+2+8 = 56
- Media: m = 56/10 = 5,6
- La media della popolazione in esame è 5,6

MEDIA

Gli obiettivi che ci si prefigge nel calcolo di una media sono sostanzialmente due:

- sostituire a più dati rilevati un solo numero che dia però una efficace rappresentazione del fenomeno dato;
- esprimere l'ordine di grandezza o tendenza centrale dell'insieme dei dati relativi a un fenomeno. Tale ordine di grandezza può a volte sfuggire perché i dati sono spesso differenti fra loro.

Si definisce **media aritmetica** di più numeri quel valore che, sostituito ai dati, lascia invariata la loro somma.

Indicati con x_1, x_2,… , x_n i numeri dati, per la definizione si ha:

$$x_1 + x_2 + \ldots + x_n = M + M + \ldots + M = n \cdot M$$

Da cui si ricava :

$$M = \frac{x_1 + x_2 + \ldots + x_n}{n} = \frac{1}{n}\sum_{i=1}^{i=n} x_i$$

Ovvero, utilizzando il simbolo della sommatoria:

$$\overline{x} = \frac{\Sigma_i{}^N = 1^x{}_i}{N}$$

- La media ci permette così di avere un'idea della quantità complessiva conoscendo soltanto il valore medio e quanti valori ci sono.
- La media può essere utilizzata come misura di sintesi per misurazioni discrete e continue.

Esempio:

- Abbiamo cinque bambini: Nicola, Giovanni, Paola, Tommaso e Chiara.
- Nicola ha 5 cioccolate, Giovanni e Tommaso una sola, mentre Paola e Chiara hanno ciascuna due cioccolate.
- Mediamente, quante cioccolate hanno i cinque bambini? .
- Soluzione:
  - possiamo dire che mediamente i cinque bambini hanno 2,2 cioccolate ciascuno e messi insieme ne hanno 11.
  - È vero che in realtà nessuno dei cinque bambini ha 2,2 cioccolate: o ne hanno di più o ne hanno di meno

Media

Esistono altri modi per calcolare una media:

- media geometrica
- media armonica
- media aritmetica ponderata

Se i valori di x_i hanno frequenze diverse, ossia compaiono più volte nelle osservazioni, per esempio il valore x_1 ha frequenza y_1, x_2 ha frequenza y_2, la condizione d'invarianza della somma diventa:

$$x_1y_1 + x_2y_2 + \ldots + x_ny_n = My_1 + My_2 + \ldots + My_n$$

da cui si ricava:

$$M = \frac{x_1y_1 + x_2y_2 + \ldots + x_ny_n}{y_1 + y_2 + \ldots + y_n} = \frac{\sum_{i=1}^{i=n} x_iy_i}{\sum_{i=1}^{i=n} y_i}$$

che si chiama **media aritmetica ponderata** perché le frequenze sono anche dette ***pesi*** .

Media aritmetica ponderata

- Il ***peso*** di ciascun valore è in genere rappresentato dal numero di volte in cui i valori figurano (frequenza), ma può significare anche l'importanza (oggettiva o soggettiva) che il singolo valore riveste nella distribuzione.
- La divisione di conseguenza non viene fatta con il numero di valori, ma con la somma dei pesi.
- Pertanto, si può affermare che la media aritmetica ponderata è la somma dei prodotti dei valori x_i per le corrispondenti frequenze relative.

Esempio

Esempio: determinare l'età media di 50 giovani presenti in una pizzeria un sabato sera; indichiamo con x_i l'età e con y_i il numero di giovani per ogni classe di età.

età (x_i)	N° giovani (y_i)	
	maschi	femmine
15-20	5	4
21-30	13	11
31-35	9	8
Totale	27	23

ESERCIZIO 1

Quale delle due seguenti distribuzioni ha media maggiore?

- Serie a: 1 5 9 4 5 6
- Serie b: 2 4 8 5 1 10

- ❑ Serie a
- ❑ Serie b
- ❑ Sono uguali

MEDIANA

- La **mediana** di una distribuzione ordinata di valori è il valore dell'unità che si trova a metà della distribuzione, in modo che il 50% dei valori della serie sia uguale o inferiore a esso e il restante 50% sia superiore.
- Per calcolare la mediana è necessario che la variabile sia quantitativa o qualitativa ordinata.

MEDIANA

Se n è dispari, la mediana è il valore centrale dell'insieme ordinato

Esempio:

- La serie: 2 6 6 4 8 9 5 1 2 dispari
- Li ordino: 1 2 2 4 5 6 6 8 9
- La mediana: 1 2 2 4 5 6 6 8 9

Se n è pari, la mediana è la media dei 2 valori centrali dell'insieme ordinato

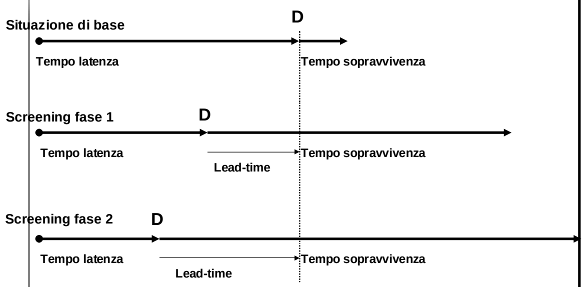
Esempio:

- La serie: 2 6 7 4 8 9 5 1 2 10 pari
- Li ordino: 1 2 2 4 5 6 7 8 9 10
- Ottengo: 1 2 2 4 5 6 7 8 9 10
- La mediana: (5+6)/2 = 5,5

CONFRONTO TRA MEDIA E MEDIANA

- Molto spesso la media e la mediana presentano valori simili.
- Ciò accade in particolare quando la distribuzione della variabile è simmetrica.
- Se però la distribuzione presenta forti asimmetrie, le due misure possono divergere notevolmente.
- Le due misure di tendenza centrale non si escludono a vicenda.
- Entrambe possono essere utilizzate per descrivere un fenomeno perché forniscono informazioni diverse.

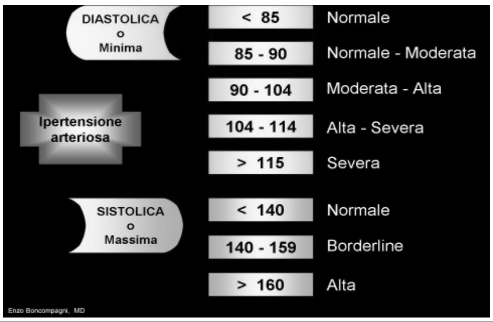
Effetto anticipazione diagnostica (lead-time) negli studi di screening



Moda o valore normale

- La *moda* o *valore normale* (detta anche *valore modale* o *norma*) è un valore caratteristico di una distribuzione di frequenze la cui determinazione non richiede alcun calcolo.
- In altre parole, si dice **moda** o **valore normale** di una distribuzione di frequenze la modalità o il valore della variabile al quale corrisponde la massima frequenza.

Valori normali della pressione arteriosa



MODA

1° Esempio:

- Serie: 1 2 4 1 5 5 8 9 5
- Moda: 5 (unico valore presente 3 volte)

2° Esempio

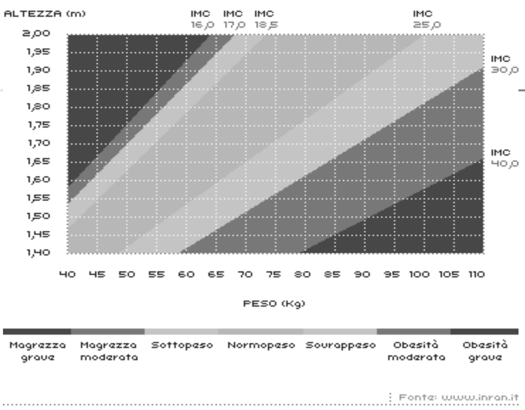
- Serie: 1 2 4 1 5 5 8 9 5 1
- Moda: 1 e 5 (valori presente 3 volte)

MODA

- La *moda* è il valore (o i valori) che si presenta con maggiore frequenza nella distribuzione.
- Se nessun valore si presenta più frequentemente degli altri la distribuzione è senza moda.
- Sono anche possibili distribuzioni con due o più valori modali.
- Può essere calcolata per qualunque tipo di variabili, anche qualitative sconnesse.

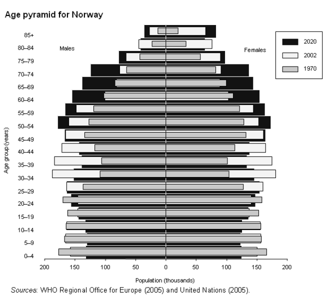
Moda o valore normale

- Se i dati sono raggruppati in classi, il calcolo della moda presenta maggiori difficoltà.
- Se l'ampiezza della classe è costante si dirà **classe modale** quella che ha la frequenza maggiore.
- Se le classi hanno ampiezza diversa, si divide ogni frequenza per l'ampiezza della rispettiva classe e la classe modale è la classe alla quale corrisponde il rapporto maggiore.



Moda o valore normale

- Talvolta per individuare la classe modale è sufficiente esaminare l'istogramma della distribuzione: la classe modale è quella che è base del rettangolo di altezza massima.
- Il valore modale è, fra tutti i valori medi, il più significativo in quanto è un dato che esprime il valore di una concreta osservazione sul fenomeno, mentre le medie di calcolo possono o meno coincidere con un valore della distribuzione.



Mediana

- La **mediana** è una media di posizione e rappresenta il valore centrale della distribuzione quando i dati sono ordinati.

Precisamente:

- Siano x_1, x_2, ..., x_n i valori ordinati in senso non decrescente, si dice **mediana Me** il valore che bipartisce la successione, ossia il valore non inferiore a metà dei valori e non superiore all'altra metà.

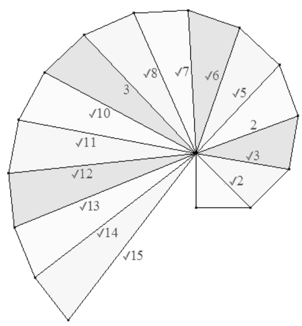
Mediana

- Ordinati i valori, se il numero *n* dei termini è *dispari*, la mediana è proprio il valore centrale.
- Se *n* è *pari*, si assume come mediana la semi somma dei due valori centrali.
- Tale procedimento si applica per le serie.
- Per le distribuzioni di frequenze con valori discreti, i dati sono generalmente già ordinati; occorre allora calcolare le frequenze assolute cumulate, che si ottengono associando a ogni valore la somma della rispettiva frequenza con tutte quelle che la precedono, e determinare *quale valore* corrisponde.

Mediana

- La mediana non è influenzata dai valori estremi della distribuzione, quindi anche se le classi estreme sono aperte, non occorre chiuderle.
- Inoltre, se la distribuzione è molto asimmetrica, il valore mediano è più appropriato della media aritmetica per esprimere un valore sintetico della distribuzione.

- Nel caso di una variabile continua con valori raggruppati in classi, per il calcolo approssimato della mediana e dei percentili è molto utile servirsi del grafico della poligonale delle frequenze relative cumulative. È infatti sufficiente trovare le ascisse dei punti corrispondenti alle ordinate ½, ¼, ¾.



PERCENTILI

Si dice percentile x (sinonimi: centile, frattile e quantile) di una distribuzione ordinata quel valore che ha alla sua sinistra il x% dei valori delle frequenze cumulate.

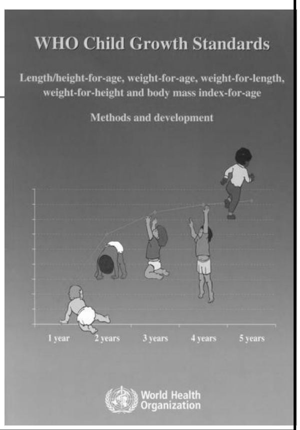
- Dal punto di vista pratico, determinare un percentile significa determinare il valore di un insieme ordinato tale per cui lui e tutti gli altri valori inferiori siano in quantità pari, rispetto al numero di elementi dell'insieme stesso, al valore di percentile visto come percentuale.
- Ad esempio, il 10° percentile di un insieme di valori è l'elemento dell'insieme per cui il 10% dei valori dell'insieme ha valore inferiore o uguale ad esso.

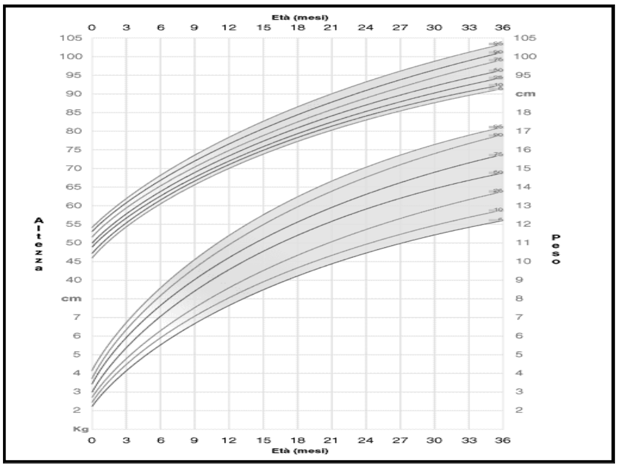
PERCENTILI
Un esempio: la crescita corporea

- Dal momento che la crescita presenta una grande variabilità individuale nei tempi e nelle misure con cui avviene, sono stati fissati dei **limiti di normalità** dividendo il range dei dati raccolti in 100 parti, chiamate **percentili**.
- Tale divisione è stata eseguita in modo che una proporzione definita dei bambini di un campione si trovasse sopra e sotto misure particolari in età particolari.

PERCENTILI
Un esempio: la crescita corporea

- Il 50° percentile rappresenta i valori di crescita medi, mentre un'altra curva, corrispondente ad esempio al 30° percentile, ci informa che una determinata percentuale di bambini presenta valori inferiori (in questo caso il 30%), che diventano superiori in un'altrettanto determinata popolazione (in questo caso il 70%). Potremmo quindi paragonare i percentili di crescita a dei binari di normalità.





25°-75° percentile:	perfetto stato di normalità
75°-90° percentile:	livello "borderline" o normo-superiori o normo-inferiori
90°-97° e 10°-3° percentile:	patologia lieve o comunque ad alto rischio
97° e 3° percentile:	patologia conclamata; in riferimento alla statura, un'altezza inferiore al 3° percentile (nanismo) non è necessariamente patologica; per questo motivo si richiede una visita specialistica approfondita (valutazione dell'età ossea, della velocità di crescita, esami ematochimici ecc.); analogo discorso per stature superiori al 97° percentile.