Prof.ssa Stefania Bufalino
DISAT- Politecnico di Torino
e-mail: stefania.bufalino@polito.it

Breve Introduzione alla Fisica

[Distribuzione degli errori e Teorema del Limite centrale]

4. Distribuzione degli errori

Nei paragrafi precedenti di questa introduzione alla Fisica, abbiamo detto che le misure sono affette da errori e assumiamo come vero valore di una grandezza la media delle N misurazioni effettuate. In questo paragrafo esaminiamo come questi valori assomigliano e giustificano la scelta del valore medio come valore "vero" della misura. A questo scopo, dimostreremo un teorema di base della teoria delle misure chiamato **"Teorema del limite centrale"**.

4.1 Misure di valori discreti

Iniziamo considerando quantità la cui misura dà solo valori discreti. A titolo di esempio prendiamo in considerazione l'insieme dei punteggi ottenuti da uno studente in tutti gli esami del Corso di laurea + master. Ogni esame è una misura dell'abilità dello studente nell'apprendimento di una materia e il voto è il risultato della misurazione. Ogni segno può assumere i valori discreti 0, 1, 2, ... 18, 29 ..., 30. Al termine degli esami allo/a studente/essa è dato un voto, la media, che è considerato il valore "vero" del/la suo/a abilità . Chiamiamo x_i [i = 0,1, ... n] i voti e: $\bar{x} \equiv \frac{1}{N}\sum_{i=1}^{N} x_i$ il valore "vero". Alcuni degli N valori sono uguali e possiamo definire la frequenza f_k come il numero di volte che il voto di k è stato ottenuto dallo studente. Pertanto, il punteggio medio può essere scritto come:

$$\bar{x} \equiv \frac{1}{N}\sum_{k=1}^{30} f_k \cdot x_k = \frac{1}{N}\sum_{k=k_{\min}}^{k_{MAX}} f_k \cdot x_k = \sum_{k=k_{\min}}^{k_{MAX}} \frac{f_k}{N} \cdot x_k \qquad (4.1)$$

Il numero $\frac{f_k}{N}$ è chiamato "frequenza normalizzata" ed è semplice verificare che

$$\sum_{k=k\min}^{k\max} \frac{f_k}{N} = 1. \qquad (4.2)$$

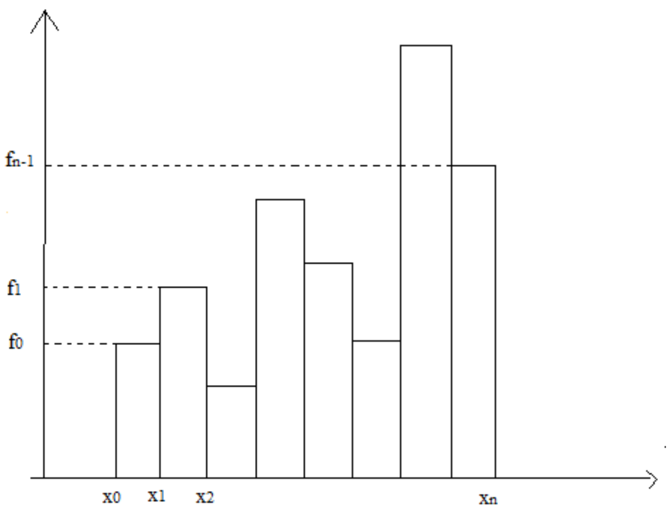
4.2 Misure di valori continui

Consideriamo ora il caso in cui la quantità assume valori continui nel range compreso tra il valore x_{min} e il valore x_{max}. In realtà questo caso non è un caso fisico perché ogni misurazione dà un valore, che è separato dai due valori più vicini (la misura con valore più piccolo e quella con valore maggiore) da una distanza finita: questa distanza è detta "risoluzione" dello strumento. Tuttavia, se la distanza è piccola (strumento ad alta risoluzione) e il numero N di misurazioni è grande, possiamo trattare il caso continuo utilizzando lo stesso concetto di frequenza adottato per il caso discreto.

4.2.1 Istogrammi

Supponiamo che un insieme di n valori x_i $[i = 0,1, \ldots n]$ di una quantità x varia tra x_{min} e x_{max} e può assumere tutti i valori continui (approssimazione per un numero grande di misure) all'interno di questo intervallo. Si assume, inoltre, che i valori siano ordinati, in modo che $x_{min} = x_0$ e $x_{max} = x_n$. Per costruire un istogramma si procede in questo modo. Dividiamo l'intervallo $[x_0, x_n]$ in n intervalli di uguale lunghezza $\Delta x \equiv \frac{x_n - x_0}{n}$ e identifichiamo ogni intervallo sulla base del suo limite inferiore x_i: $x_0, x_1, \ldots x_{n-1}$. Nell'intervallo k-esimo (denominato x_k) ci sono valori f_k diversi. Se riportiamo gli n + 1 limiti inferiori degli n intervalli e disegniamo un segmento orizzontale sull'intervallo (x_k, x_{k+1}), ad una altezza f_k per ogni k, si costruisce un grafico chiamato "Istogramma" (vedi Fig. 1)

Fig.1



Teniamo in considerazione il fatto che la somma dei valori contenuti nell'intervallo è uguale al prodotto $\sum_{over\Delta x} x_m = f_k \cdot \frac{\sum_{over\Delta x} x_m}{f_k} = f_k \cdot \bar{x}_k$ dove $\bar{x}_k$ è la media dei valori x_m dentro il k-esimo intevallo Δx.

Se N è molto elevato gli intervalli sono altamente popolati anche se n è un numero grande, e ad esempio, Δx è molto piccolo. Si può osservare che per piccoli Δx il valore di $\boldsymbol{x_m}$ si avvicina alla media $\bar{x}_k$. Quindi possiamo concludere che sotto l'ipotesi di N grane e Δx piccolo il valore medio di x, anche nel caso continuo è data da:

$$\bar{x} = \frac{1}{N}\sum_{k=0}^{n} f_k \cdot \bar{x}_k \cong \sum_{k=0}^{n} \frac{f_k}{N} \cdot x_k \quad (4.3)$$

con: $\sum_{k=0}^{n} \frac{f_k}{N} = 1$ (4.4)

dove la somma è estesa a tutti gli intervalli *k*.
A questo punto possiamo costruire l'istogramma riportando in ascissa i valori x_i, distribuiti negli n intervalli: in ordinata, ad una altezza f_k un segmento orizzontale è disegnato per ogni intervallo k. In questo modo l'istogramma si presenta come una sequenza di rettangoli verticali di base Δx e altezza f_k.

4.2.2 Proprietà degli istogrammi

Dividendo la frequenza normalizzata per Δx si ottiene un nuovo istogramma a colonne di diversa altezza, ma stessa forma. La frequenza normalizzata può essere riscritta come:

$$\frac{f_k}{N} = \frac{f_k}{N \cdot \Delta x} \cdot \Delta x \qquad (4.5)$$

e $\frac{f_k}{N \cdot \Delta x}$ è denominta *densità di frequenza* della variabile *x*. Considerando che f_k è un sottoinsieme del set totale N di valori x, possiamo chiamarlo ΔN_k (come Δx è un sottoinsieme dell'intervallo totale x_{min}- x_{max}) e scrivere le eq. (4.3) e (4.4) come:

$$\bar{x} \cong \sum_{k=0}^{n} \frac{1}{N} \frac{\Delta N_k}{\Delta x} \cdot \Delta x \cdot x_k \qquad (4.6)$$

$$\sum_{k=0}^{n} \frac{1}{N} \frac{\Delta N_k}{\Delta x} \cdot \Delta x = 1 \qquad (4.7)$$

Se l'intervallo Δx diventa sempre più piccolo, anche i numeri ΔN_k diventano sempre più piccoli, ma non il loro rapporto, purché N sia abbastanza grande da riempire sempre gli intervalli. Quando Δx e ΔN diventano infinitesimali il loro rapporto diventa la derivata, il numero di intervalli diventa infinito ed ogni intervallo Δx è indicato con il valore x (nell'intervallo continuo) del suo limite inferiore anziché dall'indice discreto *k*. Pertanto la somma su *k* diventa l'integrale della densità normalizzata tra x_{min} e x_{max} e possiamo scrivere la versione finale delle equazioni (4.6) e (4.7) come:

$$\bar{x} = \int_{x_{\min}}^{x_{MAX}} \frac{1}{N} \frac{dN}{dx} \cdot x \cdot dx \qquad (4.8)$$

$$\int_{x_{\min}}^{x_{MAX}} \frac{1}{N} \frac{dN}{dx} \cdot dx = 1 \qquad (4.9)$$

4.2.3 La funzione densità di frequenza

La funzione densità di frequenza $\frac{dN}{dx}$ (chiamata anche distribuzione della variabile *x*) può essere solo approssimata perché è impossibile effettuare una serie di misurazioni della stessa variabile x sufficienti a riempire un numero infinito di intervalli infinitesimi: servirebbe un numero infinito di misurazioni! Pertanto ci si avvicina a questa funzione con una curva mediante una prodecura di fit basata su un processo di costruzione di una curva o di una funzione matematica, che abbia la migliore corrispondenza con una serie di punti assegnati, possibilmente soggetti a limitazioni. Una prima osservazione: la forma dell'istogramma cambia ogni volta che il numero N cambia. Pertanto diverse curve corrispondono a differenti serie di misurazioni: ciò renderebbe insignificante la nostra definizione di densità di frequenza! A seguito di diversi esperimenti è stata trovata una legge che governa il comportamento degli istogrammi di misure e il loro numero. Questa legge è chiamata **"Teorema del limite centrale"**. Ad oggi questa legge è stata sempre verificata nella pratica e possiamo considerarla a tutti gli effetti una legge sperimentale. Di seguito viene riportato un esempio si esperimento di Laboratorio di Fisica I in cui questo teorema è stato verificato.

- Teorema del limite centrale -

"Se viene eseguito un numero N di misurazioni di una determinata quantità, i valori x di tale quantità saranno distribuiti nell'intervallo (-∞, + ∞), avvicinandosi sempre piu' a una funzione gaussiana tanto piu' grande è il numero N di misure effettuate".

In altre parole, la forma degli istogrammi ottenuti con diverse serie di misurazioni non cambia in modo casuale, in modo disordinato, ma converge verso una curva gaussiana all'aumentare di N: pertanto la densità di frequenza sarà espressa come:

$$\frac{dN}{dx} = A \cdot e^{\frac{-(x-a)^2}{b^2}} \qquad (4.10)$$

Come esempio riportiamo di seguito gli istogrammi ottenuti nell'esperienza n.1 del Laboratorio di Fisica I durante l'a.a. 2012/13 dalle squadre A1, A2, A3, A4, A5, A6. Il rateo di decadimento di una sorgente radioattiva di ^{60}Co è stata misurato circa 120 volte da ogni squadra e con tali misure è stato costruito il corrispondente istogramma (uno di essi è mostrato in Fig. 2a). Poi i dati provenienti da tutte le squadre sono stati sommati ed è stato costruito un nuovo istogramma (vedi Fig. 2b). Si può vedere che la forma degli istogrammi tende verso una forma sempre piu' vicina ad una distribuzione gaussiana quando la statistiche aumenta da 120 a circa 700 misurazioni da analizzare.

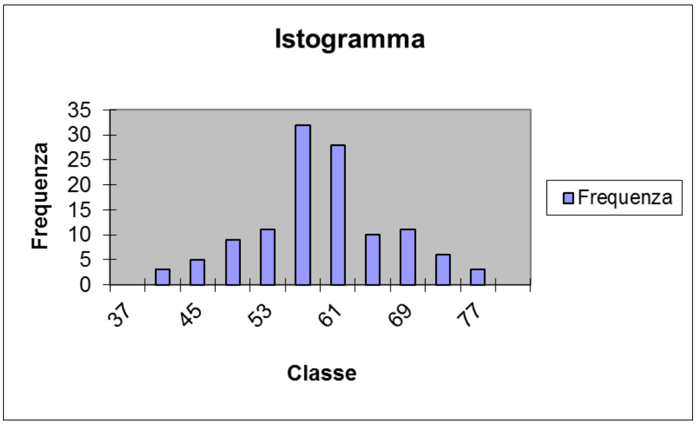


Fig.2a: analisi dei dati analizzati dalla squadra A2 (2013)

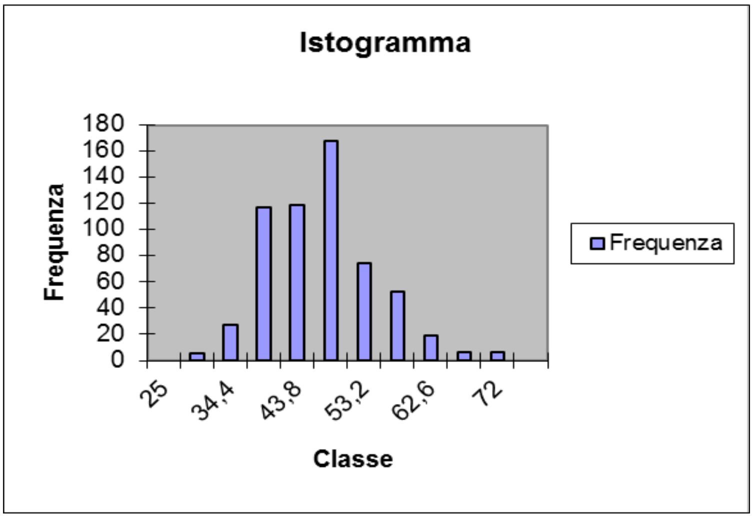


Fig.2b Analisi dei dati analizzati delle squadre A1+A2+A3+A4+A5+A6

4.2.4 Proprietà della densità di frequenza gaussiana

Un gaussiana è una funzione simmetrica a campana, sempre positiva, centrata attorno al massimo di coordinate (x = a, y = A). Ulteriori caratteristiche possono essere verificate a partire dal seguente integrale definito, che viene dimostrato nei corsi di Analisi Matematica:

$$\int_{-\infty}^{\infty} e^{-x^2} dx = \sqrt{\pi} \qquad (4.11)$$

Si deve aggiungere che l'integrale indefinito non ha soluzione analitica, in quanto può essere verificata provando e riprovando.

- **Proprietà 1 – Normalizzazione di una densità gaussiana di frequenza**:

Una distribuzione di densità $\frac{dN}{dx}$ si dice "normalizzata", se $\int_{x_{\min}}^{x_{MAX}} \frac{dN}{dx} \cdot dx = 1$

Se la distribuzione di una variabile x è gaussiana nell'intervallo ($x_{min} = -\infty$, $x_{max} = +\infty$) (vedi eq. 4.10.) e se è normalizzata, il parametro A è correlato al parametro b come segue:

Dimostrazione proprietà 1 definendo $z \equiv \frac{(x-a)}{b}$ possiamo scrivere

$dz \equiv \frac{dx}{b}, x = bz + a$ e dalla definizione (4.9) si ottiene

$$1 = \int_{-\infty}^{\infty} A \cdot e^{\frac{-(x-a)^2}{b^2}} \cdot dx = Ab \int_{-\infty}^{\infty} \cdot e^{-z^2} \cdot dz = Ab\sqrt{\pi} \Rightarrow A = \frac{1}{b\sqrt{\pi}} \qquad (4.12)$$

- Proprietà 2 – Valore medio di una variabile x

Se la distribuzione di una variabile x è una gaussiana normalizzata nell'intervallo (x_{min} = -∞, x_{max} = + ∞), cioè è data dall'eq. (4.10) e l'eq. (4.12) è soddisfatta, il parametro a è il valore medio di x.

- Dimostrazione proprietà 2.

Definendo $z \equiv \frac{(x-a)}{b}$ possiamo scrivere $dz \equiv \frac{dx}{b}$, $x = bz + a$: dalla definizione (4.8) e dall'eq. (4.12) si ottiene

$$\bar{x} = \int_{x_{min}}^{x_{MAX}} \frac{1}{N}\frac{dN}{dx} \cdot x \cdot dx = \int_{-\infty}^{\infty} A \cdot e^{\frac{-(x-a)^2}{b^2}} \cdot x \cdot dx = Ab \int_{x=-\infty}^{x=\infty} \cdot e^{-z^2} \cdot (bz+a) \cdot dz =$$
$$= Ab^2 \int_{x=-\infty}^{x=\infty} \cdot e^{-z^2} z \cdot dz + Aba \int_{x=-\infty}^{x=\infty} \cdot e^{-z^2} \cdot dz = \frac{Ab^2}{2} \int_{x=-\infty}^{x=\infty} \cdot e^{-z^2} \cdot dz^2 + Aba\sqrt{\pi} = 0 + Aba\sqrt{\pi} = a \qquad (4.13)$$

- Proprietà 3 - Varianza di una gaussiana

Se la distribuzione di una variabile x è una gaussiana normalizzata nell'intervallo (x_{min} = -∞, x_{max} = + ∞), cioè è data dall'eq. (4.10) ed è soddisfatta l'eq. (4.12), il parametro b è la varianza della distribuzione gaussiana.

- Dimostrazione proprietà 3:
Dimostriamo come prima cosa un lemma preliminare: **la densità di frequenza di una funzione *y(x)* è uguale alla densità di frequenza delle x**.
Infatti si deve osservare che nell'istogramma, costruito per la variabile x, ogni intervallo Δx contiene un numero f_k di valori x_i: se consideriamo una funzione *y(x)*, nell'intervallo Δx è contenuto lo stesso numero f_k di valori $y_i = y(xi)$. Pertanto un istogramma avente x in ascissa e il numero di conteggi di y nell'ordinata sarà uguale all'istogramma di x. **In altre parole, la densità di frequenza di *y(x)* è la stessa di x**.

Un altro modo equivalente per dimostrare che la densità di frequenza delle *y(x)* è uguale a quella di x è il seguente. Consideriamo l'intervallo ***Δy→dy*** corrispondente a ***Δx→dx***: per definizione, chiamando dN_y la frequenza dei valori y in ***Δy***, la densità di frequenza è $\frac{dN_y}{dy}$. Poche righe sopra abbiamo ricordato che il numero dN_y dei valori di *y* nell'intervallo ***Δy*** è lo stesso dei valori *x* in ***Δx***, cioè dN. Utilizzando una proprietà delle derivate, possiamo scrivere:

$$\frac{dN_y}{dy}dy = \frac{dN}{dy}dy = \frac{dN}{dy}\frac{dy}{dx}dx = \frac{dN}{dx}dx \qquad (4.14)$$

Possiamo quidi determinare il valore medio di y (partendo dalla definizione) come:

$$\bar{y} = \int_{y_{\min}}^{y_{MAX}} y \cdot \frac{dN_y}{dy} \cdot dy = \int_{x_{\min}}^{x_{MAX}} y(x) \cdot \frac{dN}{dx} \cdot dx \qquad (4.15)$$

Possiamo quindi concludere che il valore medio di una funzione *y(x)* è ottenuto dalla densità di frequenza di *x.*

Consideriamo nuovamente la distribuzione Gaussiana e valutiamo la media σ^2 (varianza) della quantità $(x-\bar{x})^2$. Definendo $z \equiv \frac{(x-\bar{x})}{b}$, possiamo scrivere:

$$\sigma^2 \equiv \frac{1}{b\sqrt{\pi}} \int_{-\infty}^{\infty} e^{\frac{-(x-\bar{x})^2}{b^2}} \cdot (x-\bar{x})^2 \cdot dx = \frac{1}{b\sqrt{\pi}} \int_{-\infty}^{\infty} e^{\frac{-(x-\bar{x})^2}{b^2}} \cdot \frac{(x-\bar{x})^2}{b^2} \cdot b^3 \cdot d(\frac{x-\bar{x}}{b}) =$$
$$= \frac{b^2}{\sqrt{\pi}} \int_{-\infty}^{\infty} e^{-z^2} \cdot z^2 \cdot dz = \frac{b^2}{2\sqrt{\pi}} \int_{-\infty}^{\infty} z \cdot e^{-z^2} \cdot dz^2 = \frac{-b^2}{2\sqrt{\pi}} \int_{-\infty}^{\infty} z \cdot d(e^{-z^2}) = \qquad (4.16)$$
$$\frac{-b^2}{2\sqrt{\pi}} \left[z \cdot e^{-z^2} \Big|_{-\infty}^{\infty} - \int_{-\infty}^{\infty} e^{-z^2} \cdot dz \right] = \frac{-b^2}{2\sqrt{\pi}} \left[0 - \sqrt{\pi} \right] = \frac{b^2}{2} \Rightarrow b^2 = 2\sigma^2$$

Infine otteniamo la distribuzone Gaussiana normalizzata riportata di seguito:

$$\frac{dN}{dx} = \frac{1}{\sqrt{2\pi\sigma^2}} e^{\frac{-(x-\bar{x})^2}{2\sigma^2}} \qquad (4.17)$$

Istrogramma e distribuzone Gaussiana

Proviamo a determinare i valori di x in corrispondenza dei quali una distribuzone gaussiana (non normalizzata) vale metà del suo valore massimo, cioè le coordinate x del punto di intersezione della curva gaussiana con una retta parallela all'asse x ad un'altezza *A/2* . Dobbiamo affermare la seguente condizione:

$$\frac{1}{2} A = A \cdot e^{\frac{-(x-\bar{x})^2}{2\sigma^2}} \Rightarrow x_{\pm} = \bar{x} \pm \sqrt{2\ln 2} \cdot \sigma \qquad (4.18)$$

La differenza:

$$x_+ - x_- = 2\sqrt{2\ln 2} \cdot \sigma \qquad (4.19)$$

è chiamata Full Width Half Maximum (FWHM) della Gaussiana.

La FWHM, frequentemente utilizzata in analisi dei dati, può essere ottenuta anche in un istogramma tracciando una linea retta parallela all'asse x ad una altezza pari alla metà della colonna più alta (canale) dell'istogramma e misurando la distanza

orizzontale tra il 2 colonne intersecate. L'istogramma dei valori delle misurazioni ripetute avvicina una gaussiana per numero di misurazioni crescente. Pertanto, la FWHM dell'istogramma dovrebbe soddisfare sempre meglio l'equazione (4.18) all'aumentare del numero di misure. Utilizzando la media dei dati $\bar{x}$ e la σ ottenuta dalla FWHM dell'istogramma (eq. 4.19), si può tracciare una curva gaussiana sull'istogramma: la curva dovrebbe adattarsi tanto meglio all'istogramma dei dati quanto alto è il numero di misurazioni effettuate. Ad esempio, in Fig. 3a e 3b sono riportate le curve gaussiane corrispondenti agli istogrammi di Fig. 2a e 2b.
La media $\bar{x}$ e la σ delle 2 funzioni gaussiane disegnate in blu chiaro sono stati ottenuti seguendo la procedura sopra descritta. È evidente che la curva corrispondente al set più abbondante di dati (3b) si adatta meglio ai canali dell'istogramma rispetto all'istogramma di Fig. 3a contenente un numero inferiore di misurazioni.

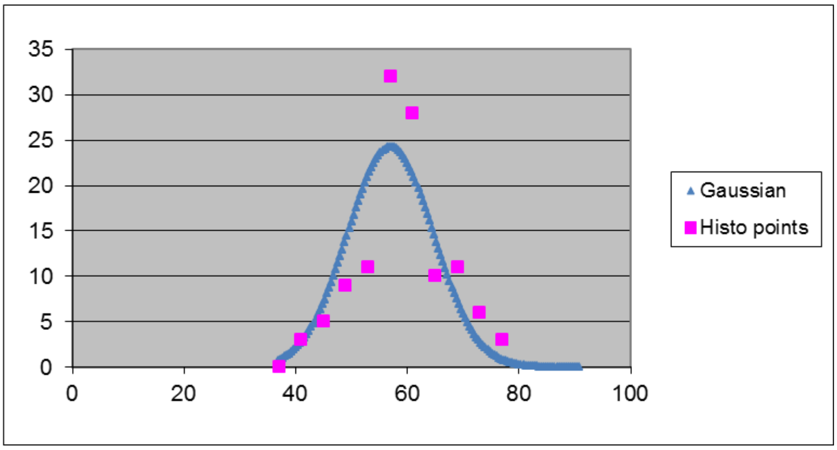


Fig. 3a dati analizzati dal gruppo A2

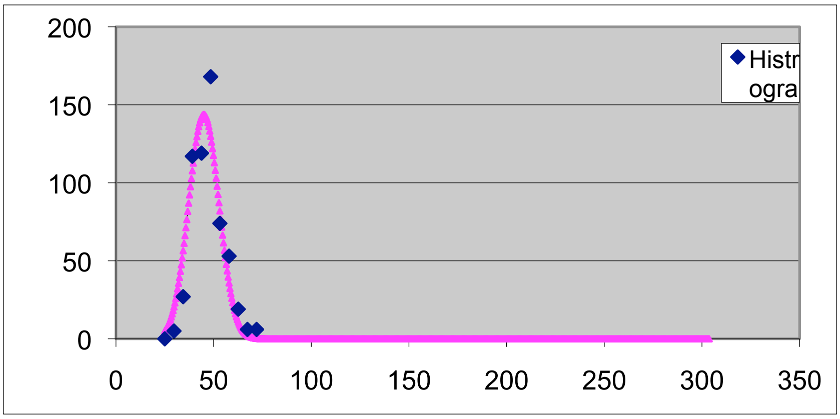


Fig. 3b dati analizzati dai gruppi A1+A2+A3+A4+A5+A6