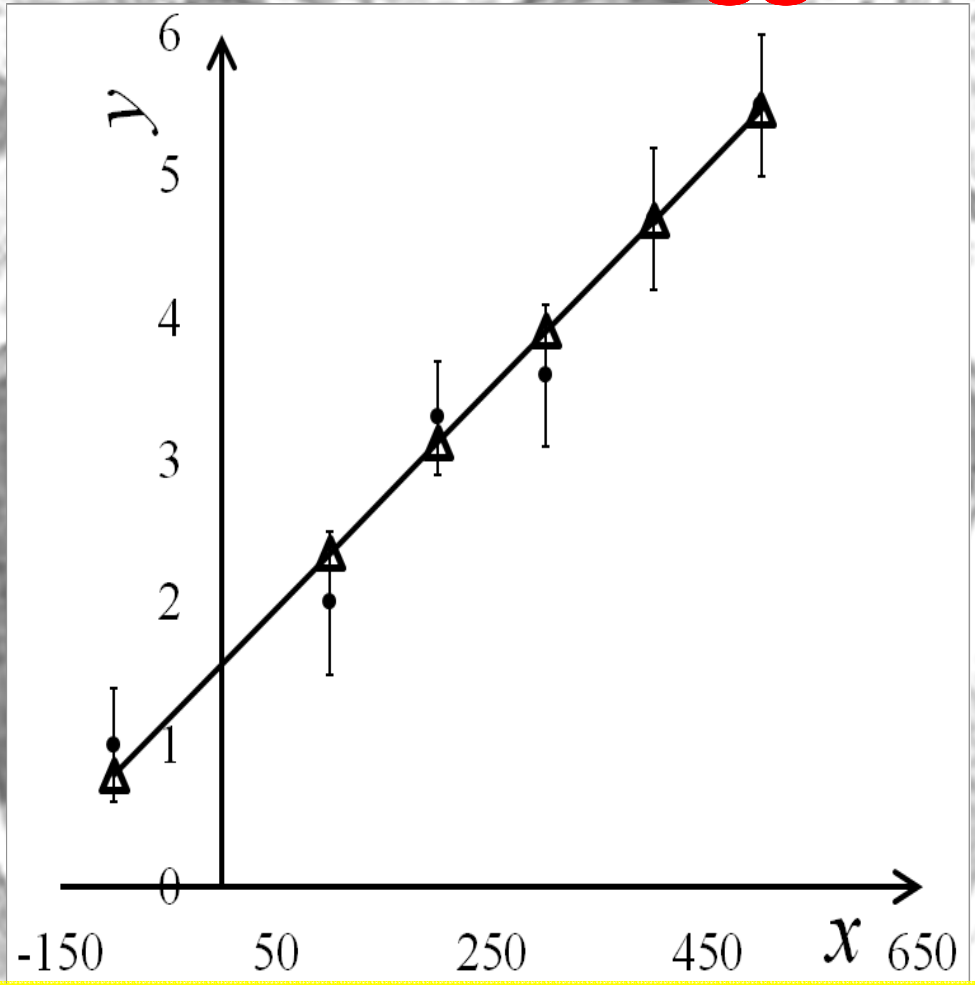
Scappatoie e …rigore.

Indicherò le scappatoie per le scuole superiori,
con l'intestazione scappatoie in sfondo celestino.

Segnalerò con e … rigore a sfondo rosso la trattazione corretta.
E indicherò le parti dove è possibile reperire indicazioni più rigorose dal testo Ciullo G. «Introduzione al Laboratorio di fisica (Springer Verlag, 2014, Milano).

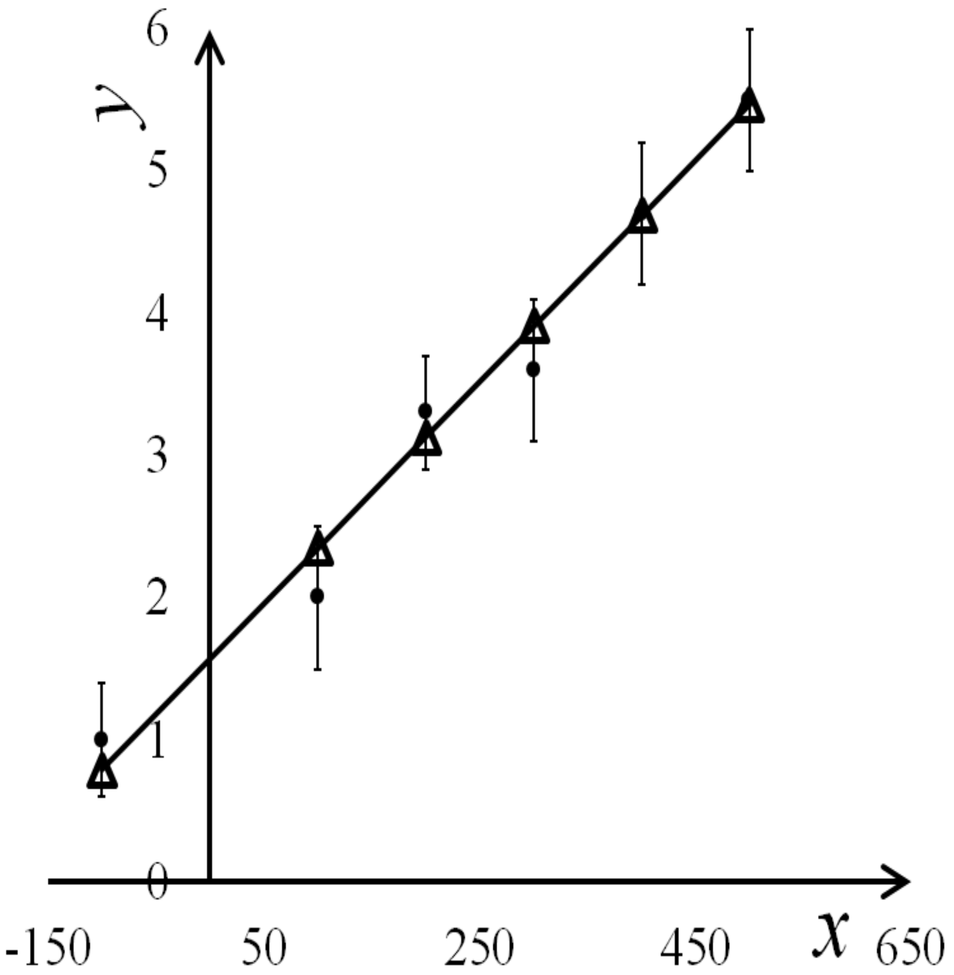
Consideriamo la regressione lineare

La verifica di una legge fisica



La verifica di una legge fisica risulta di maggiore interesse e applicabile nelle scuole. Inoltre è un ottimo strumento didattico, per estendere tale verifica (che prenderà il nome di chi-quadro), anche alle distribuzioni di dati o comunque a istogrammi.

Scappatoie



Quindi $|Y_i - y_i| \leq \delta y_i$

$$\left(\frac{y_i - Y_i}{\delta y_i}\right)^2 \leq 1$$

$$\sum\left(\frac{y_i - Y_i}{\delta y_i}\right)^2 \leq N$$

$$\min\left\{\sum\left(\frac{y_i - Y_i}{\delta y_i}\right)^2\right\}$$

e …rigore.

Se y_i segue la distribuzione gaussiana suddetta, la probabilità di ottenere il valore y_i sarà proporzionale a:

Pg 142

$$P_{Y_i,\sigma_i}(y_i) \propto \frac{1}{\sigma_i} e^{-(y_i-Y_i)^2/2\sigma_i^2} ,$$

dove il valore centrale della gaussiana sarà espresso da una relazione funzionale, nel caso particolare della regressione lineare $Y_i = A + Bx_i$.

Dai dati y_i e x_i si devono determinare dei parametri per descrive la legge, per il caso della regressioni lineare i due parametri A e B .

Esplicitiamo questo al pedice di P, che quindi risulta funzione dei parametri A e B:

$$P_{A,B}(y_i) \propto \frac{1}{\sigma_i} e^{-(y_i-A-Bx_i)^2/2\sigma_i^2} . \tag{9.2}$$

Le y_i sono stocasticamente indipendenti, percui la probabilità di ottenere tutte le misure y_i sarà data dal prodotto delle singole probabilità espresse dalla (9.2):

$$P_{A,B}(y_1, \dots, y_N) = P_{A,B}(y_1) \cdots P_{A,B}(y_N) \propto$$

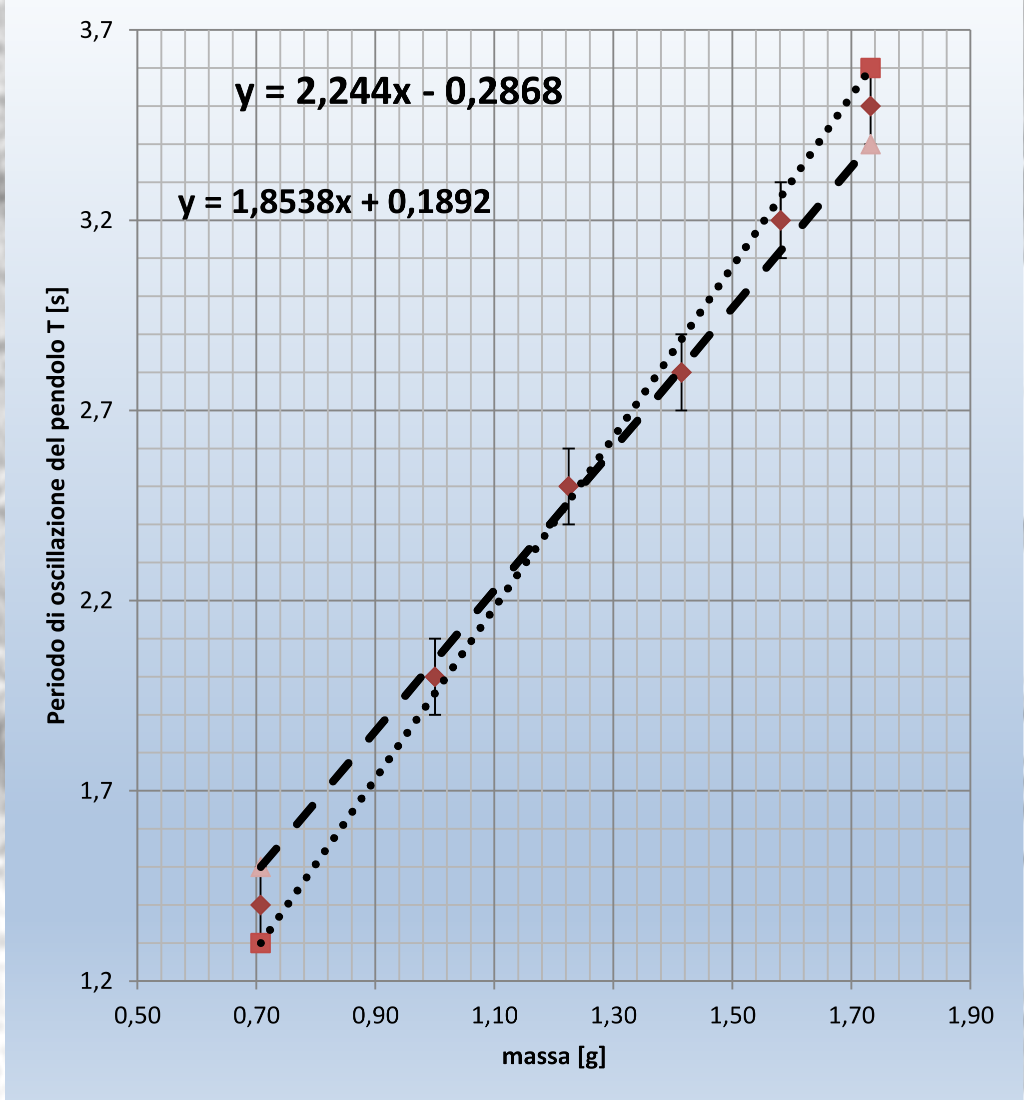
$$\propto \left(\prod_{i=1}^{N} \frac{1}{\sigma_i} \right) e^{-\chi^2/2} ,$$

$$\chi^2 = \sum \left(\frac{y_i - Y_i}{\sigma_i} \right)^2 = \sum (\chi_i)^2$$

$$\min \left\{ \sum \left(\frac{y_i - Y_i}{\sigma_i} \right)^2 \right\}$$

$$\sigma_i^2 = \text{varianza} \equiv (\, \delta y_i \,)^2$$

Scappatoie



Pendenza massima
dai i punti estremi
punteggiata
$y_{max} = A' + B_{max}\, x$
su $y_1 - \delta y_1$ e $y_N + \delta y_N$
e
pendenza minima
Retta tratteggiata
$y_{min} = A'' + B_{min}\, x$
Su $y_1 + \delta y_1$ e $y_N - \delta y_N$

B_{ms} dal valore centrale
δB dalla semidispersione

A_{ms} dal valore centrale
δA dalla semidispersione

e …rigore.

$$\min\left\{\sum\left(\frac{y_i - Y_i}{\sigma_i}\right)^2\right\} \Rightarrow$$

$$A = \frac{\sum x_i^2 \sum y_i - \sum x_i \sum x_i y_i}{N \sum x_i^2 - \left(\sum x_i\right)^2},$$

$$B = \frac{N \sum x_i y_i - \sum x_i \sum y_i}{N \sum x_i^2 - \left(\sum x_i\right)^2}.$$

Dalla minimizzazione del χ^2 si ottengono le migliori stime dei parametro *A* e *B*, avendo assunto che i dati seguano una relazione lineare del tipo *Y=A+Bx.*

Fogli di calcolo forniscono tali valori

e …rigore.

$$A = \frac{\sum x_i^2 \sum y_i - \sum x_i \sum x_i y_i}{N\sum x_i^2 - (\sum x_i)^2},$$

$$B = \frac{N\sum x_i y_i - \sum x_i \sum y_i}{N\sum x_i^2 - (\sum x_i)^2}.$$

Considero *A* e *B*, funzioni delle y_i *(x_i non affette da errore o propagato su y) dalla propagazione delle incertezze*

$$A = \frac{\partial A}{\partial y_1} y_1 + \frac{\partial A}{\partial y_2} y_2 + \cdots + \frac{\partial A}{\partial y_N} y_N,$$

$$\sigma_A^2 \approx \left(\frac{\partial A}{\partial y_1}\right)^2 (\sigma_1)^2 + \left(\frac{\partial A}{\partial y_2}\right)^2 (\sigma_2)^2 + \cdots + \left(\frac{\partial A}{\partial y_N}\right)^2 (\sigma_N)^2.$$

Per il teorema del limite centrale *A è gaussiana con varianza* σ_A

Programmi di analisi dati tali valori:propagazione.

Si ottiene quindi dalla varianza σ_A^2 del parametro A, la corrispondente deviazione standard:

$$\sigma_A = \sigma\sqrt{\frac{\sum x_i^2}{\Delta}} \overset{\sigma\equiv\delta y}{\overbrace{\Longrightarrow}} \left\{ \delta A = \delta y\sqrt{\frac{\sum x_i^2}{\Delta}} \right. . \tag{9.12}$$

Allo stesso modo si ottiene per l'incertezza sul parametro B:

$$\sigma_B = \sigma\sqrt{\frac{N}{\Delta}} \overset{\sigma\equiv\delta y}{\overbrace{\Longrightarrow}} \left\{ \delta B = \delta y\sqrt{\frac{N}{\Delta}} \right. . \tag{9.13}$$

Si osservi che le *incertezze A* e *B dipendono dalle incertezze totali sulle y* etichettate σ per agganciarci al teorema del limite centrale, per il quale A e B sono variabili gaussiane con i valori aspettati dati dalle (9.7) e (9.8) e incertezze rispettivamente date dalle (9.12) e (9.13).

Le incertezze totali nelle equazioni e nella loro deduzione (9.12) ed (9.13) sono etichettate tali per generalità ed uniformità con le variabili gaussiane. Onde evitare confusione abbiamo indicato con la frecce come diventano in fase di applicazione, quindi espresse con simbolo δ.

e ...rigore.

È utile considerare l'enunciato

e …rigore.

Teorema del limite centrale : una combinazione lineare di $g = a_x x + a_y y + a_z z + \cdots$, di *variabili aleatorie indipendenti* $x, y, z, \cdots$, ciascuna avente *legge di distribuzione qualsiasi*, ma con valori aspettati comparabili e varianze finite dello stesso ordine di grandezza, *all'aumentare del numero di variabili* aleatorie (casuali), tende alla distribuzione normale con valore aspettato G e varianza σ_g^2, dati rispettivamente da

$$G = a_x X + a_y Y + a_z Z + \cdots \text{ e}$$

$$\sigma_g^2 = a_x^2 \sigma_x^2 + a_y^2 \sigma_y^2 + a_z^2 \sigma_z^2 + \cdots,$$

dove X, Y, Z, $\cdots$ sono i valori di aspettazione delle rispettive variabili e $\sigma_x^2, \sigma_y^2, \sigma_z^2, \cdots$ le rispettive varianze.

Scappatoie

σ_Y cos'è invece?

$$\sigma_Y = \sqrt{\frac{\sum (y_i - Y_i)^2}{d}}$$

Distanza media (/*N* numero di punti – coppie di dati)

Statistica (/*d* – gradi di libertà = numero di dati – vincoli statistici)

$$\sigma_Y = \sqrt{\frac{\sum_i^N (y_i - Y_i)^2}{d}}$$

vincoli statistici- parametri utilizzati per la stima che si ottengono utilizzando i dati,
osservare la formula per il numero di dati

σ_Y cos'è invece?

e …rigore.

Si supponga che tutte le misure y_i siano gaussiane tendenti al valore vero Y_i, espresso da una relazione qualsiasi $Y = Y(x)$ rispetto alle x_i, e aventi tutte la stessa dispersione σ_Y ideale (si noti la Y maiuscola al pedice), pertanto possiamo esprimere la probabilità di ottenere una determinata y_i, che segua questa distribuzione ideale, con centralità $Y_i = Y(x_i)$:

$$P_{A,B}(y_i) \propto \frac{1}{\sigma_Y} e^{-(y_i - Y_i)^2/(2\sigma_Y^2)} , \qquad (9.14)$$

si osservi che si presuppone a priori, che tutte le variabili y_i seguano una gaussiana

P per *N* y_i che seguono una la gaussiana di parametri *Y*, σ_Y)

$$P_{A,B}(y_1, \ldots, y_N) \propto \frac{1}{\sigma_Y^N} e^{-\Sigma(y_i - Y_i)^2/(2\sigma_Y^2)} .$$

P per *n* x_i seguono $G_{X,\sigma}(x)$

$$P(X) \equiv P(x_1, x_2, \ldots, x_{n-1}, x_n) \sim \frac{1}{\sigma^n} e^{-\Sigma_{i=1}^n (x_i - X)^2/(2\sigma^2)} .$$

e ...rigore.

Se applichiamo il principio di massima verosimiglianza a questa ipotesi, rispetto alla migliore stima di σ_Y, (stesso calcolo fatto per la miglior stima del parametro σ_x nel caso della gaussiana con valore centrale X nella (8.10)) si ottiene:

$$\sigma_Y \text{ (ideale) } = \sqrt{\sum (y_i - Y_i)^2 / N} \tag{9.16}$$

La formula riportata nella (9.16), non è altro che la distanza media dei punti sperimentali y_i dai punti (Y_i) della retta Y assunta adatta ai dati. Ma nel quadro della statistica uno stimatore dedotto come valore medio, va diviso per i gradi di libertà pertanto, la (9.16) deve essere invece riscritta come:

$$\sigma_Y \text{ (migliore stima)} = \sqrt{\sum (y_i - Y_i)^2 / d}\,. \tag{9.17}$$

dove d sono i gradi di libertà statistici. Quanto espresso nella (9.17) vale per per qualsiasi legge–relazione $Y = Y(x)$, vediamo il caso particolare per la regressione lineare.

Le stime dei parametri A e B risultano essere due vincoli statistici, percui σ_Y sarà data dalla somma degli scarti quadratici diviso i gradi di libertà $d = N - c$, ovvero $N - 2$, fornendo per σ_Y di una retta:

$$\sigma_Y \overbrace{=}^{Y=A+Bx} \sqrt{\sum_{i=1}^{N} (y_i - A - Bx_i)^2 / (N-2)}\,. \tag{9.18}$$

PAUSA di meditazione

- Gauss: misure ripetute
- X (valore centrale) stimato da x_{media}
- σ (punti di flesso) stimati da σ_x
- se gaussiana per il teorema del limite centrale

$\sigma_{\bar{x}}$

- Regressione: relazioni funzionali
- Y (valore centrale) stimato da A e B se retta
- Se tutte le y_i gaussiane

σ_Y

Conseguenze per gaussiane

- Migliore stima di x, assunta la variabile gaussiana, dalla media aritmetica
- Incertezza statistica, comunque per più di trenta dati, assunta la gaussiana, σ_x.
- Incertezza totale δ somma in quadratura di ε_x e σ_x ed eventuale accuratezza + o – η_x *(vedi calibrazioni).*

- Nocciolo duro per le scuole: se x gaussiana, incertezza statistica, $\sigma_{\bar{x}}$ teorema del limite centrale.

Conseguenze per regressioni

- Migliore stima dei parametri, con formule o «scappatoie».
- Incertezze totale δA *e* δB, sia da formule (propagazione delle incertezze) che da valori centrali e semidispersione.
- Eventuale accuratezza *(vedi calibrazioni) con leggi fisiche (esempi: caduta del grave e calorimetro).*

- Nocciolo duro per le scuole: se la retta è adatta ai dati , incertezza statistica σ_Y, ristima di nuovo.

Come possiamo verificare le ipotesi

- Partiamo dalla regressione: abbiamo ottenuto le stime dei parametri minimizzando il χ^2

$$\min\left\{\sum\left(\frac{y_i - Y_i}{\sigma_i}\right)^2\right\}$$

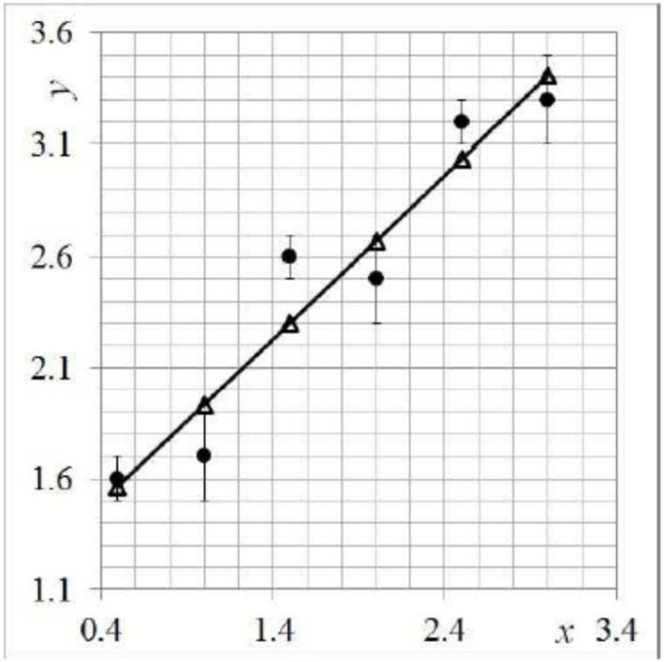
$$\chi^2 = \sum\left(\frac{y_i - Y_i}{\sigma_i}\right)^2 = \sum(\chi_i)^2$$

o meglio

$$\chi^2 = \sum\left(\frac{y_i - Y_i}{\delta y_i}\right)^2 = \sum(\chi_i)^2$$

La verifica del χ^2

La *verifica del* χ^2 *è un controllo*, a posteriori, proprio *sulla minimizzazione* del χ^2 nella (11.1) e fornisce una valutazione quantitativa dell'accettabilità dell'ipotesi, che la relazione funzionale $y = f(x)$, trovata dall'analisi dei dati, o assunta tale a priori, sia il modello teorico adatto ai dati sperimentali.

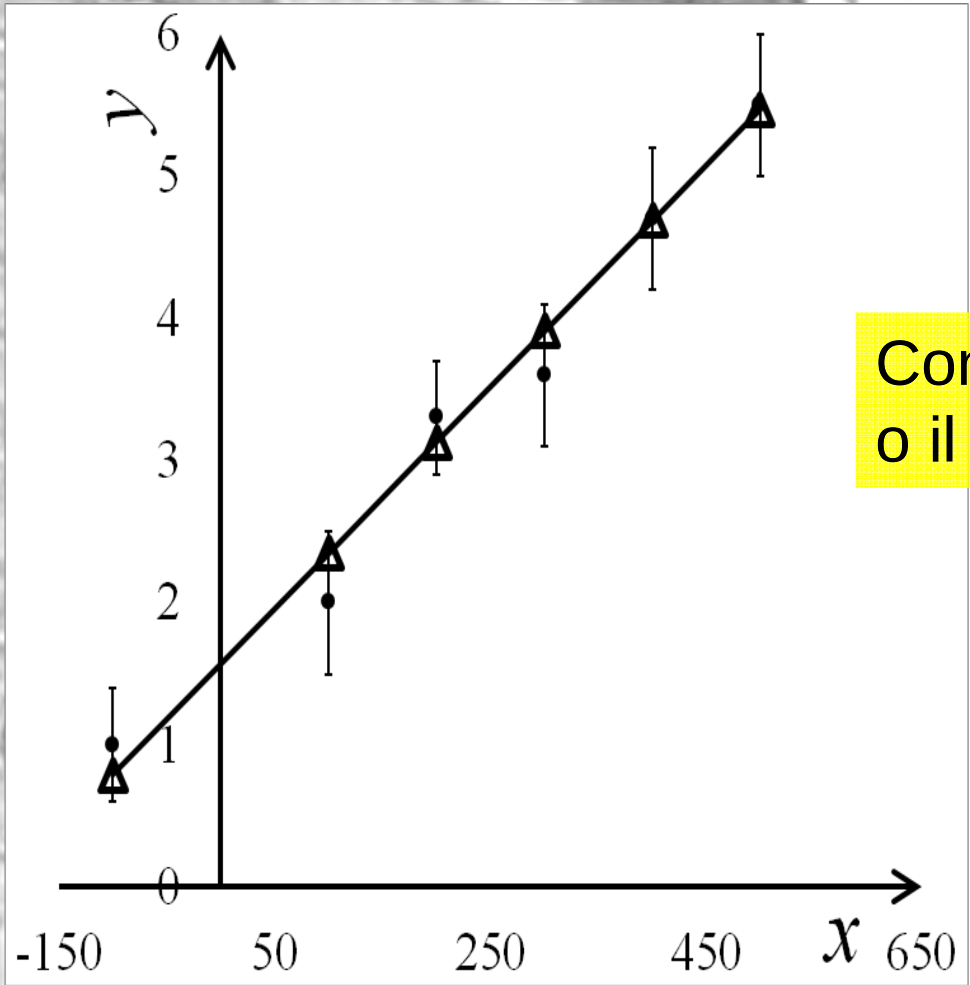


$$\chi^2 = \sum_{i=1}^{N} \left(\frac{y_i - Y_i}{\delta y_i} \right)^2 \leq \sum_{1=1}^{N} 1 = N.$$

Scappatoia

$$\sigma_Y = \sqrt{\frac{\sum (y_i - Y_i)^2}{d}}$$

$$\sigma_Y \approx \sqrt{\frac{\sum (y_i - Y_i)^2}{N}}$$



$$\chi^2 = \sum \left(\frac{y_i - Y_i}{\delta y_i} \right)^2 = \sum (\chi_i)^2 \leq N$$

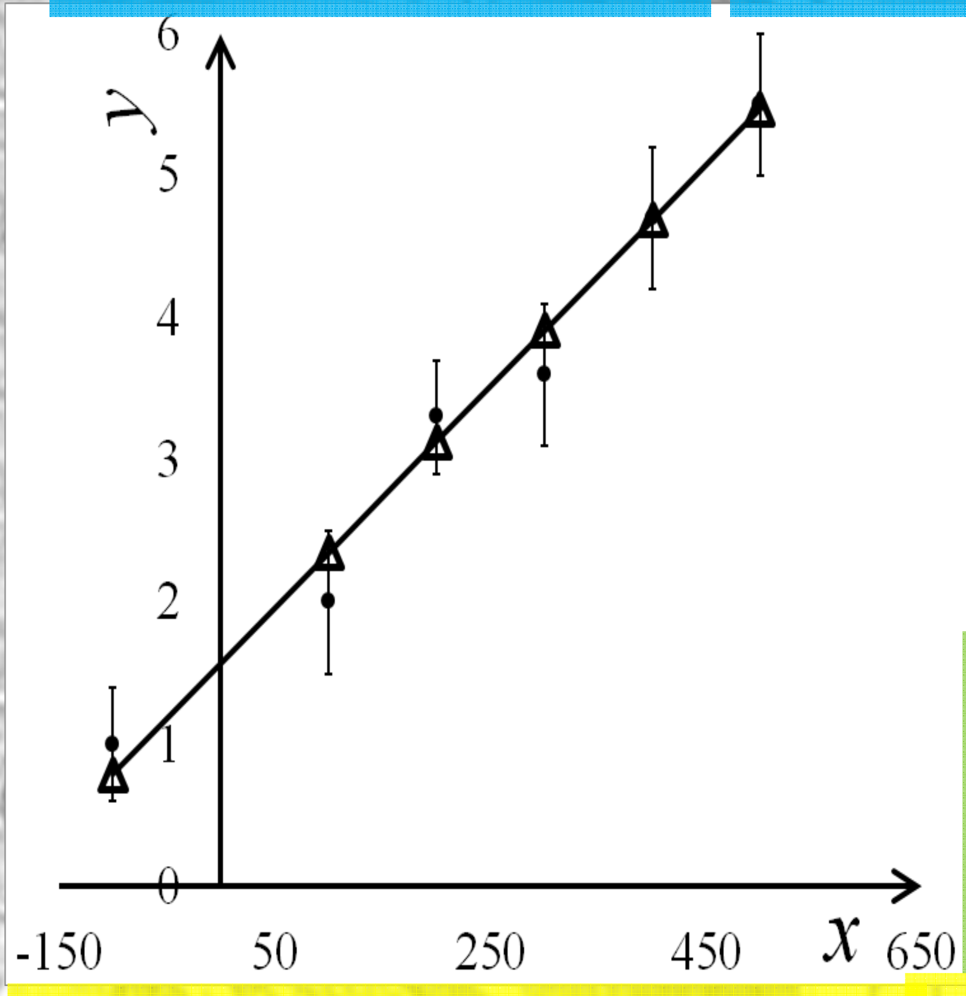
Considero δy_i uguali per tutte le i o il valore medio se cambiano.

$$\chi^2 = \frac{1}{(\delta y)^2} \sum \left(\frac{y_i - Y_i}{1} \right)^2 \leq N$$

$$\sigma_Y^2 N = \sum (y_i - Y_i)^2$$

$$\chi^2 = \frac{1}{(\delta y)^2} \left(\sigma_Y^2 N \right) \leq N \equiv \sigma_Y^2 \leq (\delta y)^2$$

Scappatoia per la verifica del χ^2



Se $\sigma_Y \leq \delta y$

La legge è appropriata per i dati:

una buona stima dell'incertezza statistica è σ_Y

Invece di $(\delta y_i)^2 = \varepsilon^2_{y_i} + \sigma^2_{y_i}$

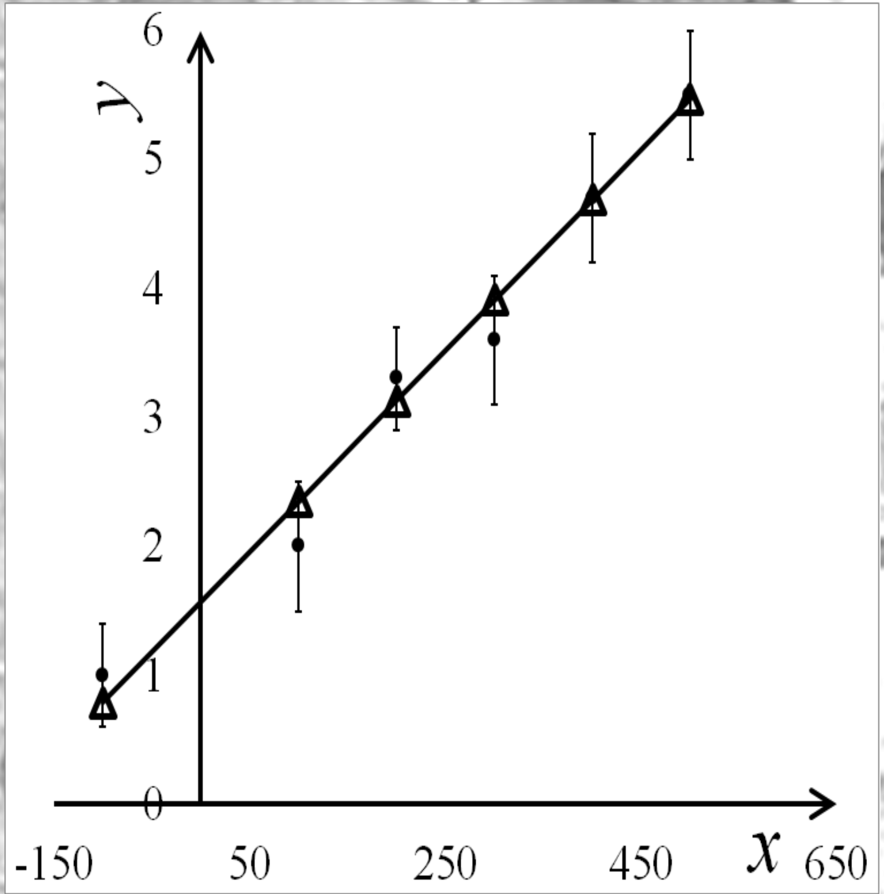
$$(\delta y_i)^2 = \varepsilon^2_{y_i} + \sigma^2_{Y}$$

Scorciatoia $(\delta Y)^2 = \varepsilon^2_{y} + \sigma^2_{y}$

$$(\delta Y)^2 = \varepsilon^2_{y} + \sigma^2_{Y}$$

Scappatoie

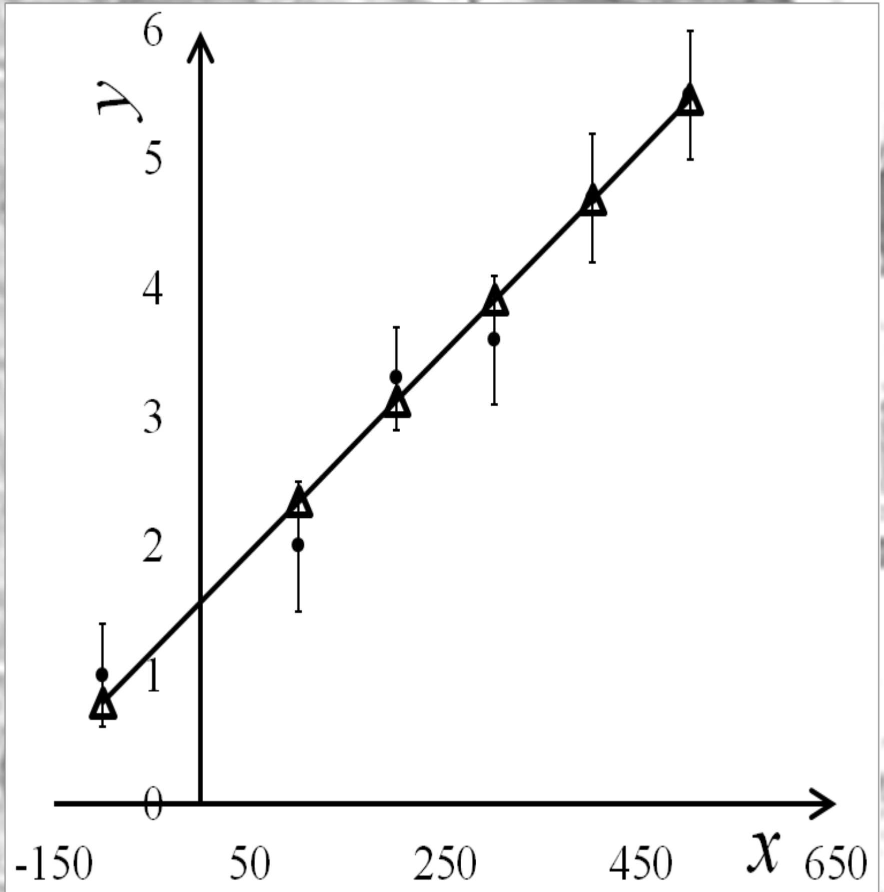
$$\text{Se } \sigma_Y \le \delta y$$



Possiamo abbattere l'incertezza statitisca e quindi ricalcolare le incertezze su *A* e *B*

Sconto per le superiori buone le stime e buona la legge niente ricalcolo

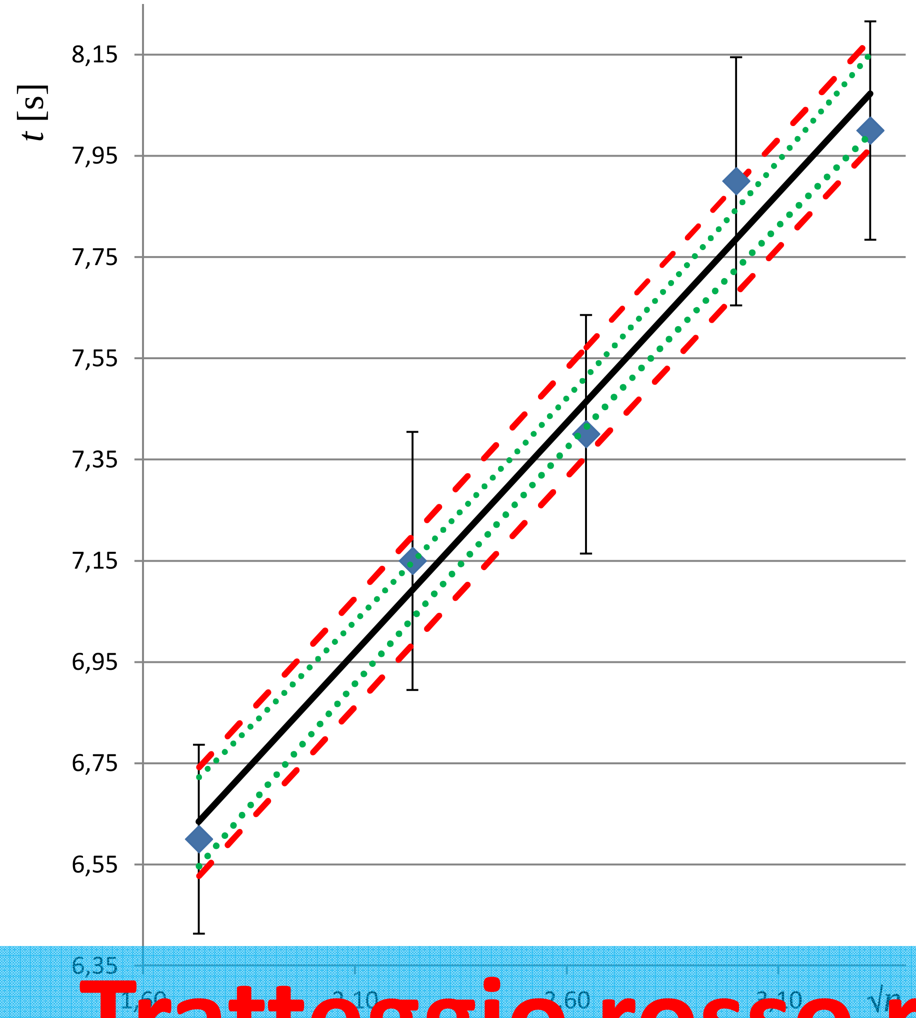
Stima sull'interpolazione



Ci permette di capire cosa vogliamo dire

Incertezza sulla y dedotta dalla legge (interpolazione). Per esempio calibriamo un Sistema, ed usiamo il valore y dedotto da y=A+Bx

Se possiamo accettare la legge?



$$\text{Se } \sigma_Y \leq \delta y$$

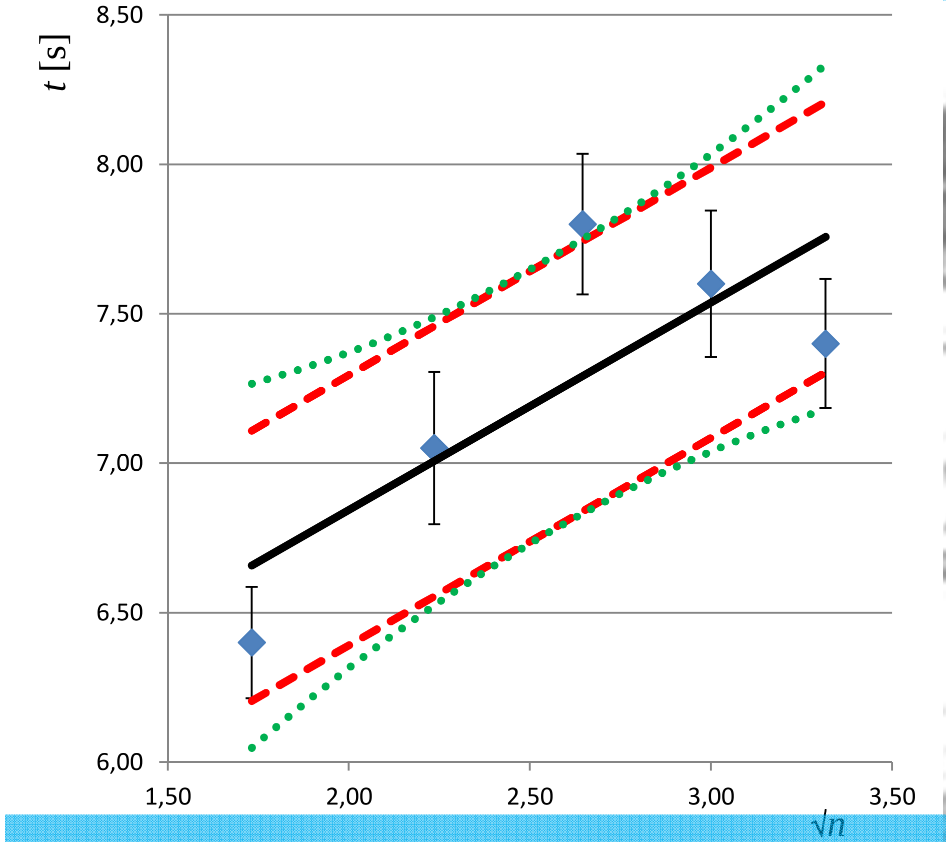
$$\delta Y = \sqrt{\sigma_Y^2 + \varepsilon_y^2}$$

Figura 1: dati sperimentali (rombi celesti) con barre d'errore rispettive δy_i secondo la tabella 1, retta di regressione (linea continua nera), pallini rossi e retta tratteggiata in rosso $Y \pm \delta Y$, curve punteggiante in verde $Y \pm \sigma_{y\text{-}interp}$.

Tratteggio rosso rette $Y+\delta Y$ e $Y-\delta Y$

previsione al 68 % incluse le incertezza

Se non possiamo accettare la legge?



$$\text{Se } \sigma_Y > \delta y$$

$$\delta Y = \sqrt{\sigma_Y^2 + (\delta y)^2}$$

Figura 1: dati sperimentali (rombi celesti) con barre d'errore rispettive δy_i secondo la tabella 1, retta di regressione (linea continua nera), pallini rossi e retta tratteggiata in rosso $Y \pm \delta Y$, curve punteggiante in verde $Y \pm \sigma_{y\text{-}interp}$.

Non accettiamo la legge

Forniamo comunque una stima per ***Y+δY e*** *Y-δY* al 68 % incluse le incertezze

e …rigore.

$$\bar{\chi}^2 = \sum_{i=1}^{N} \left(\frac{y_i - Y_i}{\delta y_i} \right)^2 \Big/ N \; ..$$

La media dei $\chi_i{}^2$ *espressa in forma statistica è detta chi-quadro-ridotto*, indicato di solito con la sovrapposizione di una tilde $\tilde{\chi}^2$, e si ottiene, *dividendo* χ^2 per i *gradi di libertà* $d = (N$ (numero dei dati utilizzati) - c (vincoli statici)).

$$\tilde{\chi}^2 = \frac{\chi^2}{d} = \sum_{i=1}^{N} \left(\frac{y_i - Y_i}{\delta y_i} \right)^2 \Big/ d \; . \qquad (11.5)$$

e …rigore.

Quanto descritto finora ha basi statistiche, che è opportuno accennare, perché il χ^2 è necessario per effettuare la verifiche di significatività per un modello (legge) teorico atteso e si estende anche a distribuzioni di probabilità attese.

Abbiamo visto che, nel caso di una distribuzione gaussiana, possiamo verificare, se un determinato singolo valore può, oppure no, essere accettato, confrontandoci con la densità di probabilità gaussiana. Dalla probabilità che un valore x sia compreso nell'intervallo $[-x_O, x_O]$:

$$P(|x| \leq x_O) = 2\int_0^{x_O} G_{X,\sigma}(x)\mathrm{d}x \ ,$$

si può calcolare la probabilità complementare di ottenere un valore $|x| \geq x_O$: $2\int_{x_O}^{\infty} G_{X,\sigma}(x)\mathrm{d}x$.

Sono stati definiti dei livelli di significatività, per rigettare l'ipotesi, che non ci sia differenza tra il nostro campione ed il valore x_O.

Osserviamo che per l'adattamento di una retta, dovremmo controllare per ogni singolo valore sperimentale y_i, se il valore atteso Y_i, sia accettabile.

Quindi per ogni i, avremmo che la probabilità di ottenere un determinato valore compreso tra zero e y_{iO} secondo la gaussiana è:

$$P(|y_i| \leq y_{iO}) = 2\int_0^{y_{iO}} G_{Y_i,\sigma_i}(y_i)\mathrm{d}y_i = 2\frac{1}{\sqrt{2\pi}}\int_0^{y_{iO}} \mathrm{e}^{-(y_i - Y_i)^2/(2\sigma_i^2)}\mathrm{d}y_i \ .$$

Se le variabili y_i sono stocasticamente indipendenti tra di loro, possiamo ottenere la probabilità totale come prodotto delle probabilità (gli estremi degli integrali non vengono esplicitati per semplicità di scrittura):

$$\left(\frac{1}{\sqrt{2\pi}}\right)^N \int\int \ldots \int \mathrm{e}^{-(\Sigma z_i^2/2)}\mathrm{d}z_1 \cdots \mathrm{d}z_N \ , \qquad (11.7)$$

dove abbiamo utilizzato le note variabili standardizzate $z_i = (y_i - Y_i)/\sigma_i$, che in questo capitolo abbiamo rietichettato χ_i.

e …rigore.

L'integrale multiplo in (11.7) si riduce ad un integrale semplice, che ha come integrando:

$$f(\chi^2) = \frac{1}{2^{d/2}\Gamma(d/2)}(\chi^2)^{d/2-1}e^{-\chi^2/2}. \qquad (11.8)$$

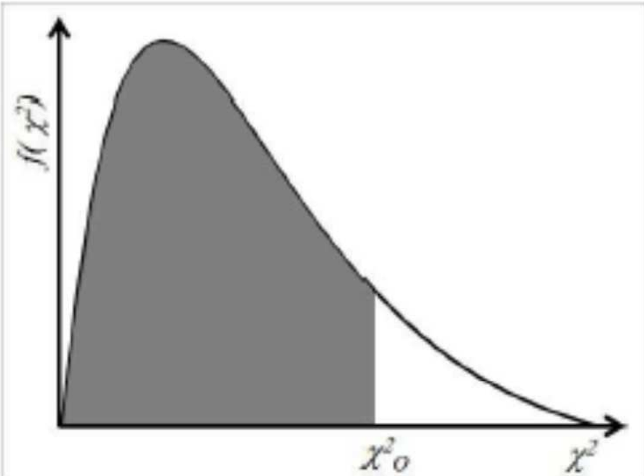


Figura 11.2 Densità di probabilità del χ^2. L'area ombreggiata fornisce l'integrale $\int_0^{\chi_0^2} f(\chi^2)d(\chi^2)$.

L'integrando $f(\chi^2)$ è la densità di probabilità, che segue la variabile χ^2, ottenuta dall'avere assunto, che le variabili y_i siano tutte gaussiane.

Per d > 100

$\chi^2 = d$

Deviazione Standard =

= ($2d$)$^{1/2}$

Altrimenti tabella della probabilità di ottenere
Un chi-quadro ridotto (χ^2/d) $\geq$ *di quello osservato*

e …rigore.

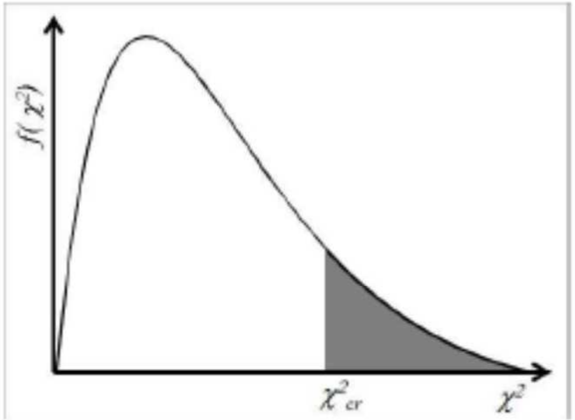


Grafico dell'area corrispondente al valore di probabilità per $\chi^2 \geq \chi^2_{cr}$: $P(\chi^2 \geq \chi^2_{cr}) = \int_{\chi^2_{cr}}^{\infty} f(\chi^2)\mathrm{d}\chi^2$. Nella Tabella C.1 sono riportati invece $\tilde{\chi}^2_{cr} = \chi^2_{cr}/d$, per ricordare che la probabilità dipende anche da d.

Coda destra

Tabella C.2 Tabella dei valori $\tilde{\chi}^2_{cr}$ per i quali seguendo la riga corrispondente ai gradi di libertà d (1^a colonna) si ottiene la probabilità (2^a riga) di ottenere $\tilde{\chi}^2$ maggiore del $\tilde{\chi}^2_{cr}$ tabulato. Critico in quanto oltre tale valore, se si pone un dato livello di significatività, si può rigettare l'ipotesi. In enfasi le probabilità *0.01* (1 %) e *0.05* (5 %), che individuano i livelli di significatività più in uso.

	Probabilità $P(\tilde{\chi}^2_d \geq \tilde{\chi}^2_O)$								
d	0.005	*0.010*	0.025	*0.050*	0.100	0.250	0.500	0.750	0.900
1	7.88	6.63	5.02	3.84	2.71	1.32	0.455	0.102	0.0158
2	5.30	4.61	3.69	3.00	2.30	1.39	0.69	0.29	0.11
3	4.28	3.78	3.12	2.61	2.08	1.37	0.79	0.40	0.19
4	3.71	3.32	2.79	2.37	1.95	1.35	0.84	0.48	0.27
5	3.35	3.02	2.57	2.21	1.85	1.33	0.87	0.53	0.32
6	3.09	2.80	2.41	2.10	1.77	1.31	0.89	0.58	0.37
7	2.90	2.64	2.29	2.01	1.72	1.29	0.91	0.61	0.40
8	2.74	2.51	2.19	1.94	1.67	1.28	0.92	0.63	0.44
9	2.62	2.41	2.11	1.88	1.63	1.27	0.93	0.66	0.46
10	2.52	2.32	2.05	1.83	1.60	1.25	0.93	0.67	0.49

Verifica di significatività 1 %

Ipotesi da rigettare

Legge non appropriata

O anche tabella della probabilità di ottenere
Un chi-quadro ridotto (χ^2/d) $\leq$ *di quello osservato*

e …rigore.

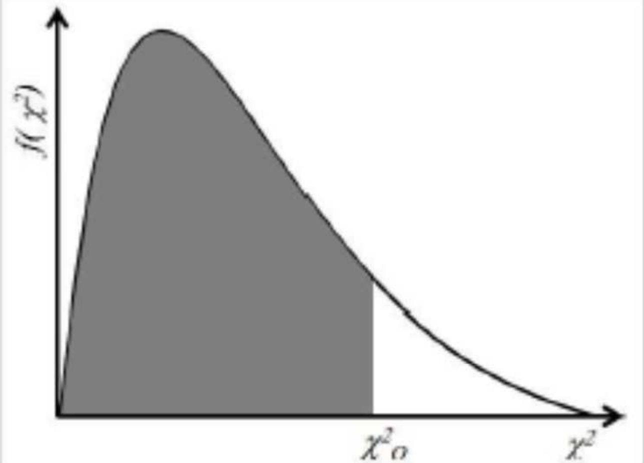


Grafico dell'area corrispondente al valore di probabilità per χ^2 tra zero e χ^2_O:
$P(\chi^2 \leq \chi^2_O) = \int_0^{\chi^2_O} f(\chi^2)\mathrm{d}\chi^2$.
Tale valore dipende dai gradi di libertà come riportato nella Tabella C.1.

Coda sinistra

Tabella C.1 Tabella dei χ^2_O: in corrispondenza di d gradi di libertà nella 1ª colonna, si hanno le probabilità $P(\chi^2 \leq \chi^2_O)$ nella 2ª riga, per un dato χ^2_O

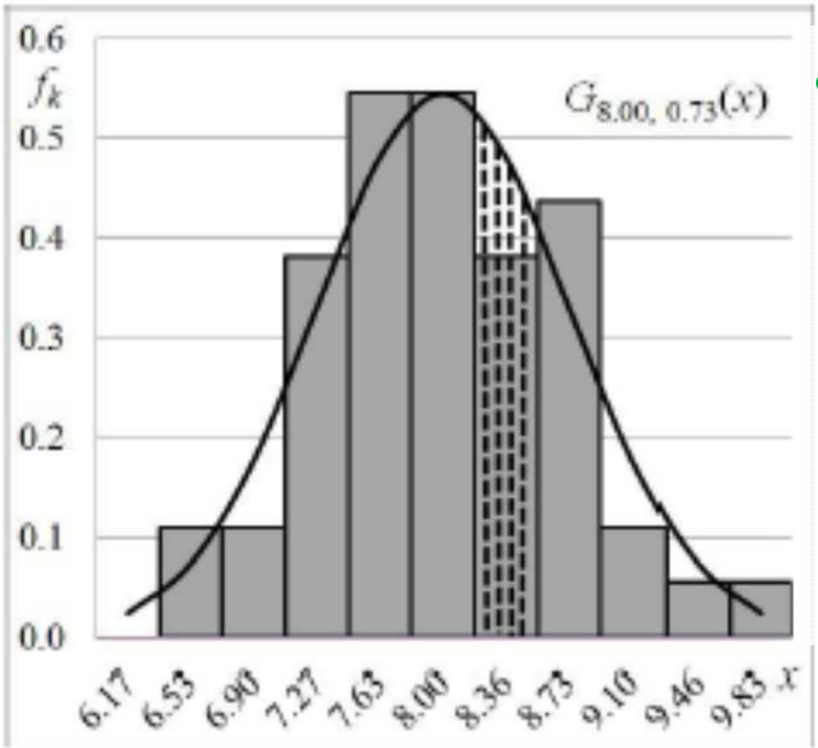
	Probabilità $P(\chi^2 \leq \chi^2_O)$											
d	0.005	*0.01*	*0.05*	0.100	0.250	0.500	0.750	0.900	0.950	0.975	0.990	0. 995
1	.000039	.00016	.0393	0.0158	0.102	0.455	1.32	2.71	3.84	5.02	6.63	7.88
2	.0100	.0201	.103	0.211	0.575	1.39	2.77	4.61	5.99	7.38	9.21	10.6
3	.0717	.115	.352	0.584	1.21	2.37	4.11	6.25	7.81	9.35	11.3	12.8
4	.207	.297	.711	1.06	1.92	3.36	5.39	7.78	9.49	11.1	13.3	14.9
5	.412	.554	.1.15	1.61	2.67	4.35	6.63	9.24	11.1	12.8	15.1	16.7
6	.676	.872	1.64	2.20	3.45	5.35	7.84	10.6	12.6	14.4	16.8	18.5
7	.989	1.24	2.17	2.83	4.25	6.35	9.04	12.0	14.1	16.0	18.5	20.3
8	1.34	1.65	2.73	3.49	5.07	7.34	10.02	13.4	15.5	17.5	20.1	22.0
9	1.73	2.09	3.33	4.17	5.90	8.34	11.4	14.7	16.9	19.0	21.7	23.6
10	2.16	2.56	3.94	4.87	6.74	9.34	12.5	16.0	18.3	20.5	23.2	25.

Verifica di significatività 1 %

Incertezze elevate

Estensione della verifica del χ^2

- Per le distribuzioni di tipo istogramma:



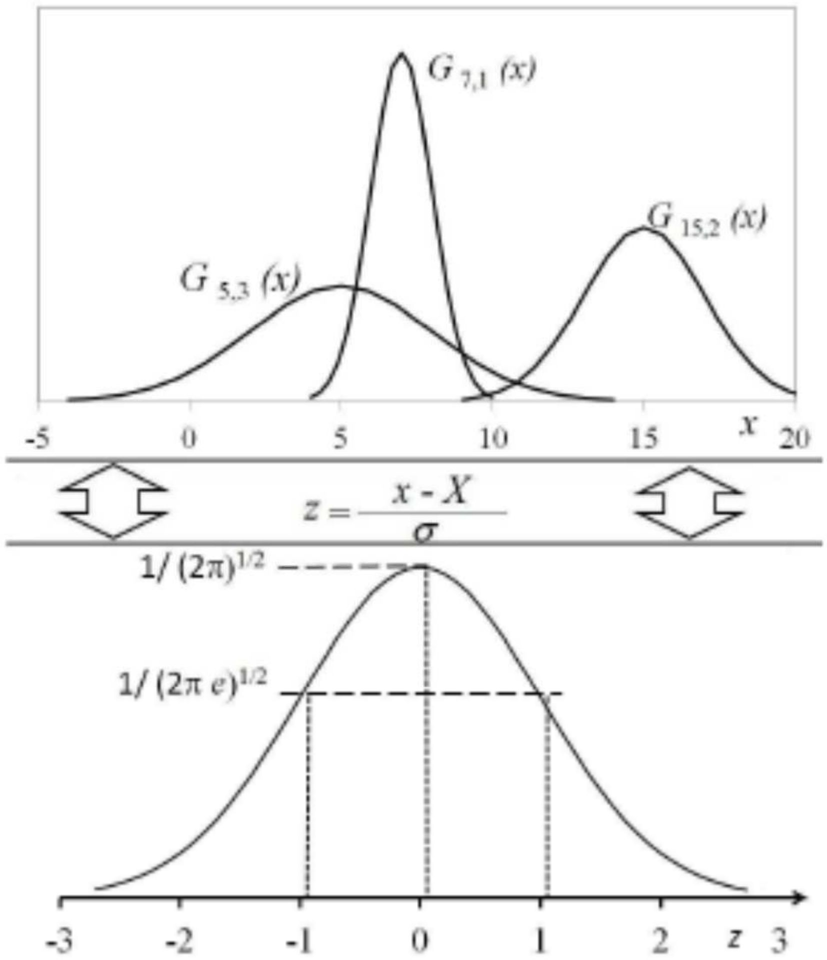
- Organizzo per classi e confronto quante
  - occorrenze ho ottenuto in una classe

  con
  - le aspettative calcolate per una gaussiana

Ogni gaussiana riconducile a $G_{0,1}(z)$

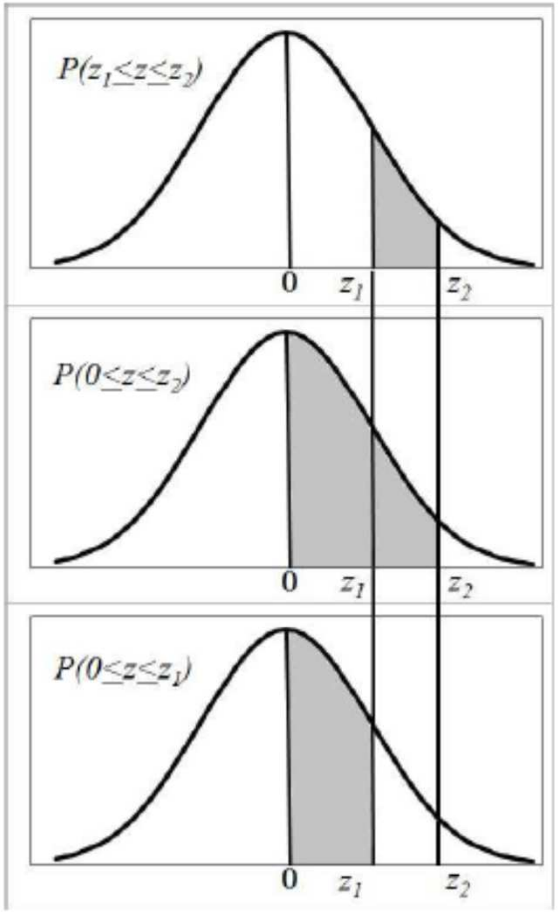
Figura 8.4 Nella figura superiore sono riportate varie gaussiane $G_{X,\sigma}(x)$ con centralità X e larghezza σ differenti. Grazie alla trasformazione delle gaussiane mediante la variabile standardizzata $z=(x-X)/\sigma$, tutte si riconducono a $G_{0,1}(z)$, gaussiana di centralià Z=0 e deviazione standard σ=1.

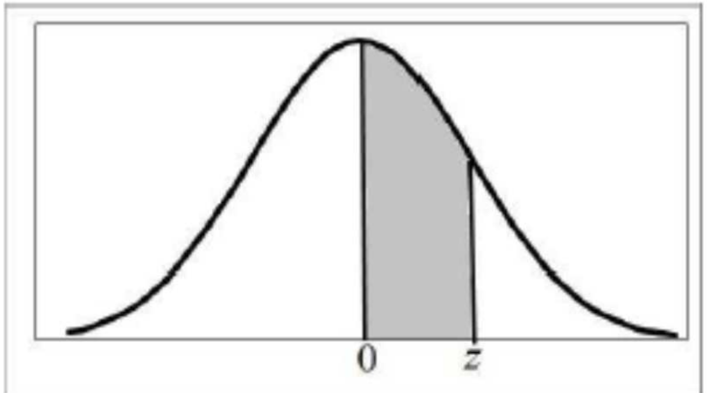
$$z=\frac{x-X}{\sigma}$$



Integrale normale *tra zero e z (Tab)*

Figura A.2 Rappresentazione grafica del calcolo dell'area per un intervallo compreso tra z_1 e z_2 con l'ausilio della Tabella A.1. L'integrale tra z_1 e z_2, riportato nella parte superiore si ottiene sottraendo all'integrale tra zero e z_2 grafico nella parte centrale, l'integrale tra zero e z_1, grafico nella parte inferiore.





Teorema della somma di Pearson

- Quante classi dobbiamo scegliere?

Teorema della somma di Pearson:
In un istogramma di un numero n_{classi} di classi, costruito da un campione casuale di n dati e avente probabilità vera P_k in ogni classe, la quantità:

$$Q = \sum_{k=1}^{n_{classi}} \left(\frac{O_k - nP_k}{\sqrt{nP_k}} \right)^2 ,$$

dove la $\sum_{k=1}^{n_{classi}} O_k = n$ e O_k sono il numero di eventi osservati per la classe corrispondente k, tende ad avere, per $n \to \infty$, la densità di probabilità χ^2 espressa nella (11.8) con "$n_{classi} - 1$" gradi di libertà.

- Q tende alla funzione χ^2, se in ogni classe si aspettano $E_k = nP_k > 10$

$$\overline{\chi}^2 = \sum_{i=1}^{n_{classi}} \chi_k^2 \Big/ n_{classi} = \sum_{i=1}^{n_{classi}} \frac{(O_k - E_k)^2}{E_k} \Big/ n_{classi}$$

Per una distribuzione gaussiana il *chi–quadro–ridotto* — $\tilde{\chi}^2$ — risulta:

$$\tilde{\chi}^2 = \frac{\chi^2}{d} = \sum_{k=1}^{n_{classi}} \left(\frac{O_k - nP_k(\overline{x}, \sigma_x)}{\sqrt{nP_k(\overline{x}, \sigma_x)}} \right)^2 \Big/ (n_{classi} - 3) \ ,$$

dove si è esplicitato, che P_k è ottenuta dai parametri $\overline{x}$ e σ_x.

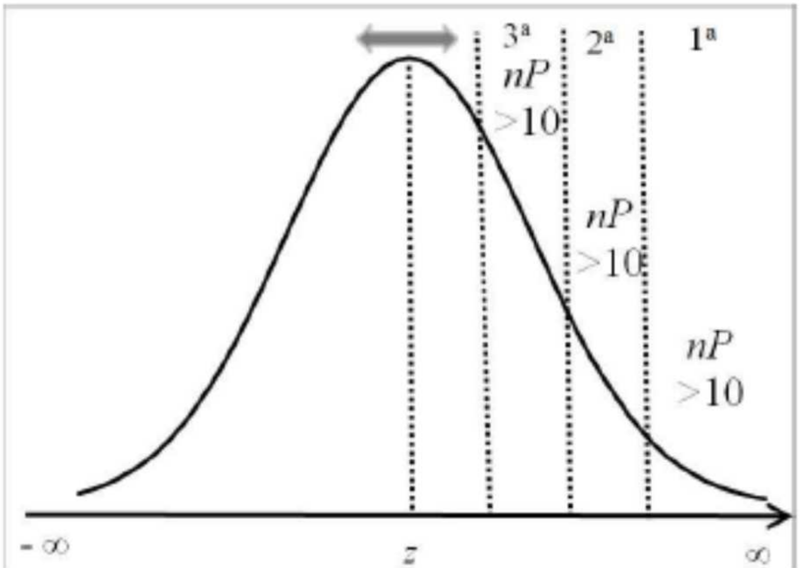


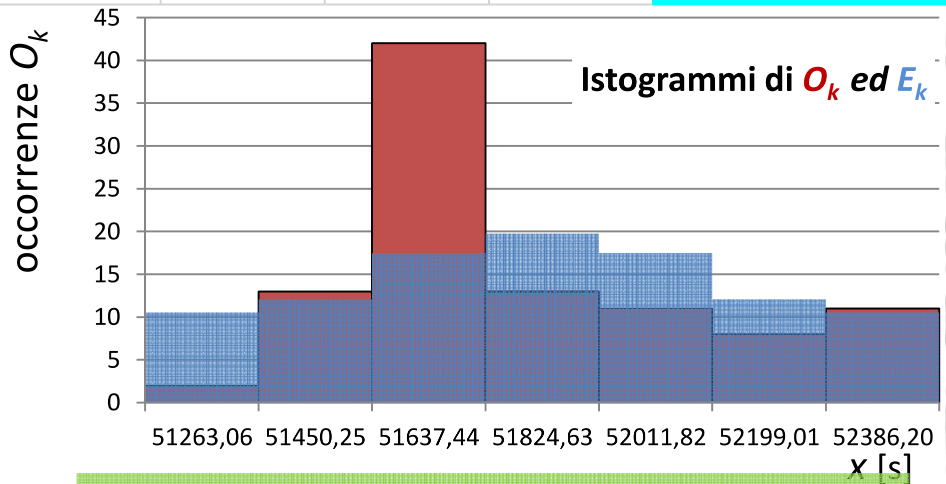
Figura 11.4 Sequenze dell'organizzazione delle classi per la verifica del χ^2. La prima linea tratteggia a destra delimita la 1^a operazione da effettuare sulla coda, partendo da $+\infty$, per trovare z_{min} per avere almeno dieci E_k, poi a scalare trovare le variabili z, che soddisfino sempre le condizione del teorema della somma di Pearson. La doppia freccia indica, che per simmetria, quanto calcolato da $z = 0$ a ∞, vale da $z = 0$ a $-\infty$.

Verifica sui conteggi

$z_{k\,min}$	$z_{k\,max}$	$P(0<z<z_{k\,min})$	$(0<z<z_{k\,ma}$	P_k	E_K		E_K	O_K	χ_K^2
					4.01				
					6.55		10.56	2	7
					12.1		12.1	13	0
					17.47		17.47	42	34
-0.250	0.250	0.0987	0.0987	0.1974	19.74		19.74	13	2
0.250	0.750	0.0987	0.2734	0.1747	17.47		17.47	11	2
0.750	1.250	0.2734	0.3944	0.121	12.1		12.1	8	1
1.250	1.750	0.3944	0.4599	0.0655	6.55		10.56	11	0
1.750	infto	0.4599	0.5	0.0401	4.01				
			somma	100.00			100.00	100.00	
								χ^2 /d	11.89
								χ^2 /n_{classi}	6.79

Sconto per le superiori

$$\chi^2/n_{classi} \le 1 \text{ gaussiana}$$
$$\chi^2/n_{classi} > 1 \text{ non gaussiana}$$



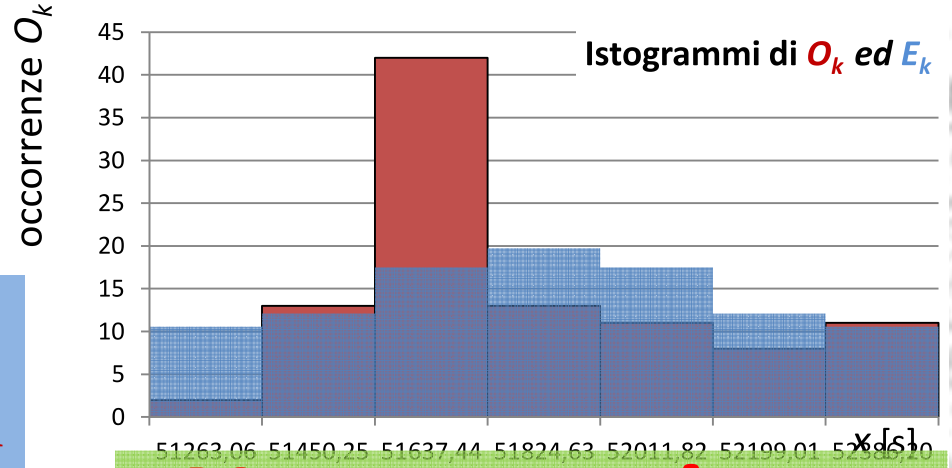
Non gaussiana

Verifica sul pendolo

$z_{k\,min}$	$z_{k\,max}$	$P(0<z<z_{k\,min})$	$P(0<z<z_{k\,max})$	P_k	E_K		E_K	O_K	χ_K^2
					3.61				
					5.90		9.50	10	0.03
					10.89		10.89	6	2.20
					15.72		15.72	9	2.87
-0.250	0.250	0.0987	0.0987	0.1974	17.77		17.77	27	4.80
0.250	0.750	0.0987	0.2734	0.1747	15.72		15.72	22	2.51
0.750	1.250	0.2734	0.3944	0.121	10.89		10.89	10	0.07
1.250	1.750	0.3944	0.4599	0.0655	5.90		9.50	6	1.29
1.750	infto	0.4599	0.5	0.0401	3.61				
			somme		90.00		90.00	90.00	
							χ^2/d	3.44	
							χ^2/n_{classi}	**1.97**	

Sconto per le superiori

$$\chi^2/n_{classi} \leq 1 \text{ gaussiana}$$
$$\chi^2/n_{classi} > 1 \text{ non gaussiana}$$



Non gaussiana

Conclusioni

- L'obiettivo di un'esperienza del laboratorio, è fornire una misura, in ogni caso.
- La verifica di una legge fisica sicuramente risulta più diretta ed interessante.
- La verifica di una variabile gaussiana, un po' artificiosa. Ma da quanto visto usare la deviazione standard del campione è conservativo.
- Per dati maggiori di trenta comunque se non si può verificare che sia gaussiana, comunque l'incertezza statistica è stimabile con la deviazione Standarda del campione.