LABORATORIO DI FISICA per Scienze e Tecnologie per l'Ambiente. AA 2012/2013

ESPERIENZA N.1

Determinazione della distribuzione degli errori in una serie di misure ripetute della stessa grandezza fisica

1. Introduzione

Scopo dell' esperienza è verificare che gli errori casuali nella misura di una grandezza fisica ripetuta molte volte nelle stesse condizioni sperimentali seguono la distribuzione normale o di Gauss.
Nell' esperienza proposta, la grandezza da misurare è l' intervallo di tempo impiegato da una slitta che scivola su un piano inclinato con attrito trascurabile, per percorrere una distanza prefissata.

2. Descrizione dell' apparato sperimentale

L'apparato sperimentale è costituito da una "guidovia a cuscino d'aria" schematicamente rap-presentata in **Fig.1**, il cui scopo è di realizzare un moto virtualmente senza attrito della "slitta", da un piccolo elettromagnete (EC) montato alla sommità della guidovia e da un cronometro elettronico collegato a due traguardi fotoelettrici anch' essi fissati sulla guidovia. Questa è così composta:

- un **tubo metallico** lungo all'incirca 1.4 m a **sezione rettangolare**, sulla cui faccia superiore sono praticati numerosi forellini per la fuoriuscita dell'aria compressa; il tubo costituisce il piano sul quale si forma , al passaggio di una slitta opportunamente sagomata, un cuscino d'aria in modo che la slitta stessa possa scorrere con attrito radente trascurabile.

- un basamento metallico sul quale il tubo è montato, che per mezzo di una **vite micrometricaV** può variare la sua inclinazione rispetto al banco di lavoro, approssimativamente orizzontale. **Un giro della vite** corrisponde ad una variazione di inclinazione **$\delta\theta$=5'.**

- **un compressore d'aria**, a portata variabile, collegato ad un estremo della guida tramite un manicotto flessibile.

I **due traguardi fotoelettrici** T_1 e T_2 sono collocati a fianco della guida e possono essere fissati in diverse posizioni lungo una scala millimetrata solidale col basamento. Essi sono collegati ad un cronometro elettronico; l'elemento fotosensibile emette un impulso elettrico al passaggio della slitta, senza che con essa vi sia contatto materiale.

Il **cronometro elettronico**, per il quale i segnali dai due traguardi funzionano rispettivamente da 'start' e 'stop' per la misura del tempo; nella **scala "10K"** il cronometro misura i tempi con sensibilità pari a **0.1 ms** . La partenza del conteggio è automaticamente preceduta dall'azzeramento (che può anche essere ottenuto manualmente premendo il pulsante 'reset').
Si utilizzerà la **slitta con due magnetini sottili** (senza velcro). **L' elettromagnete**, posto all'estremità sinistra della guida, ha il compito di rilasciare la slitta imprimendole un **impulso minimo e riproducibile,** ciò si ottiene tenendo premuto il pulsante collegato all'elettromagnete.

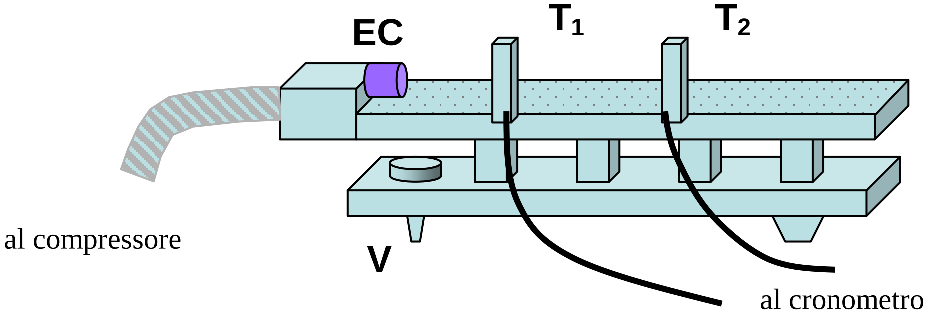

Fig. 1 - Schema dell'apparato sperimentale.

3. Operazioni sperimentali e presa dati.

a. Si accende il compressore, regolandone la portata corrispondente alla posizione 2,5 o 3 della manopola. E' opportuno **non modificare la posizione della manopola** durante l'intera procedura di presa dati.

b. Si definisce approssimativamente la **posizione di orizzontalità** della guidovia ponendo su di essa la slitta in posizione centrale e ruotando la vite micrometrica finché la slitta, lasciata libera di scorrere, rimane in quiete. A questo punto si sblocca la **ghiera** alla base della vite micrometrica (mediante una piccola vite laterale) e la si ruota in modo che lo spigolo dell'asticella corrisponda alla tacca di riferimento sulla vite micrometrica.

c. **Scelta dell'inclinazione**. Si ruoti ora la vite micrometrica di un certo numero intero di giri. Si consiglia un numero compreso fra 4 e 10.

d. Si **posizionano i traguardi** ad una distanza di circa 30 cm, con l'unico accorgimento che la prima fotocellula non deve essere troppo vicina all'elettromagnete. Si verifica che quando la slitta passa in corrispondenza dei traguardi muovendosi verso destra si ottenga l'inizio e la fine del conteggio. Se dovesse accadere il contrario, controllare che il commutatore 'start-stop' sia sulla posizione 'start"

e. Portare la slitta all'estremità sinistra della guida, in modo che il magnetino aderisca all'elettromagnete fisso. Tenendo premuto il pulsante, si otterrà il rilascio della slitta. Ripetere alcune volte la procedura e verificar che i tempi di transito sono circa uguali.

Si inizia quindi la presa dati; si suggerisce di ripetere la misura almeno 200 volte. Allo scopo di evitare errori sistematici, si faccia attenzione a:

- non urtare il tavolo o la guidovia;
- non toccare la vite di livellamento;
- non variare la portata del compressore;
- non premere il pulsante durante il conteggio.

4. Analisi dei dati.

Un esempio di dati (simulati) è riportato in tabella alla fine. I 200 valori di tempo sono elencati nelle colonne gialle. Innanzitutto, si calcola il **valore medio <t>** delle N misure:

$$<t> \equiv \frac{1}{N}\sum_{j=1}^{N} t_j$$

e la **deviazione standard del campione**: $\sigma \equiv \sqrt{\frac{1}{N-1}\sum (t_j - <t>)^2}$.

Le colonne bianche a destra dei dati, in tabella, rappresentano gli **scarti dalla media**: $z_k = t_k - \langle t \rangle$, che per i nostri scopi sono più convenienti da rappresentare.

Una prima valutazione qualitativa della coerenza interna dei dati può essere ottenuta costruendo l'**ideogramma** delle N misure o meglio dei loro scarti, ovvero i valori z_k in funzione del numero di misura k, come in **fig. 2**. Le due linee orizzontali in questa figura corrispondono a $\pm\sigma$.

Questa rappresentazione può mettere in evidenza cambiamenti intervenuti durante la misura (ad es. movimenti meccanici, deriva dell'elettronica ...) che possono influenzare il risultato.

La rappresentazione più importante della distribuzione delle misure è costituita dall'**istogramma degli scarti** (**fig. 3**) ovvero il numero di misure comprese negli intervalli (tutti uguali) in cui si è diviso l'asse Δz. Per costruire l'istogramma bisogna scegliere un valore opportuno dell'intervallo Δz. Per questo si prenda la metà della deviazione standard, eventualmente arrotondata per comodità: $\Delta z \cong \sigma / 2$.

La convenienza di questa scelta si può apprezzare dal fatto che, per una distribuzione gaussiana, la teoria prevede che il 99 % dei valori degli scarti dalla media siano compresi in un intervallo di

ampiezza $\pm 3\sigma$ intorno allo zero. Si ottiene così una distribuzione con qualche decina di conteggi negli intervalli centrali.
Si passa quindi a verificare se la distribuzione ottenuta approssimativamente la legge normale degli errori casuali, sia in modo grafico che in modo quantitativa. Graficamente, si sovrappone all'isto-gramma sperimentale la curva gaussiana. La densità di probabilità gaussiana è esprimibile come

$$f(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(x-\bar{x})^2}{2\sigma^2}}$$

dove $\bar{x}$ è il valore medio e σ la deviazione standard. Se si usano gli scarti z la distribuzione è centrata intorno allo zero.

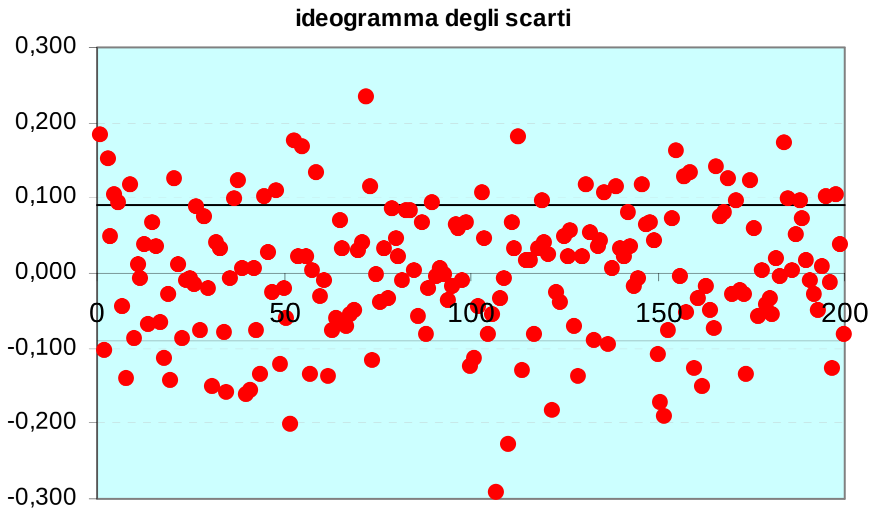

Fig. 2 Ideogramma relativo ad una serie di 200 misure dei tempi di percorrenza di un tratto prefissato della guidovia da parte di una slitta .

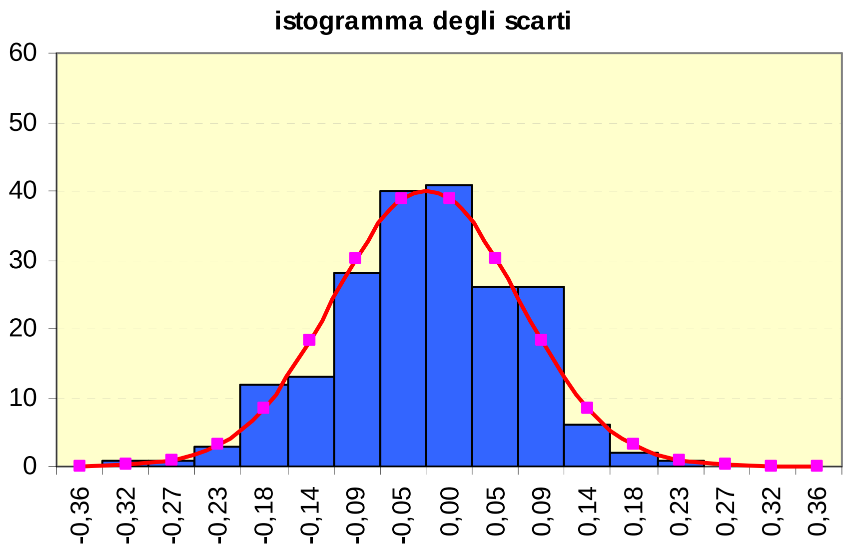

Fig. 3 Istogramma degli scarti dalla media per i dati relativi alla Fig. 2.

La gaussiana così scritta è normalizzata all'unità (l'area sottesa vale 1) e non è direttamente confrontabile con l'istogramma delle misure. L'area di un canale dell'istogramma in cui cadono *n* misure è $n\Delta z$; perciò, essendo Δz costante, l'area totale è:

$$\sum_i n_i \Delta z = \Delta z \sum_i n_i = N \cdot \Delta z$$

si dovrà moltiplicare la distribuzione gaussiana per questo fattore per poterla confrontare, ottenendo

$$f(z) = A \exp\left(-\frac{z^2}{2\sigma^2}\right) \quad \text{con} \quad A = \frac{N \cdot \Delta z}{\sigma\sqrt{2\pi}}.$$

Si tratta allora di calcolare questa funzione per valori corrispondenti al centro di ogni singolo canale dell'istogramma.
La curva gaussiana così ottenuta è mostrata in **Fig. 3** sovrapposta all'istogramma. E' opportuno che i punti della curva siano calcolati al centro dei vari intervalli, altrimenti nel grafico ottenuto con EXCEL curva e istogramma risultano "sfasati".

Per un confronto numerico della distribuzione sperimentale con la distribuzione normale, che uno si aspetta, si procederà al calcolo del "**chi-quadro**" (χ^2), dato dalla formula:

$$\chi^2 = \sum_K \frac{(N_K - G_K)^2}{N_K}$$

dove N_K è il conteggio sperimentale nell'intervallo k-esimo e G_K il valore della gaussiana al centro dell'intervallo. Per questa somma si considerino 10-11 intervalli centrali e, se in uno di questi intervalli non ci sono conteggi, si consideri il denominatore pari a 1. Conviene dividere il Chi-quadro per il numero di punti (10 o 11): orientativamente, se il valore ottenuto risulta minore di 2 possiamo dire che la nostra distribuzione è in buon accordo con una distribuzione gaussiana, entro le incertezze statistiche; se il valore è maggiore di 10 la distribuzione non è compatibile con una gaussiana.

La relazione dovrà contenere:

- titolo, nomi degli autori e breve introduzione (scopo della misura e strumenti a disposizione)
- il numero del tavolo;
- l'inclinazione scelta per la guidovia e la posizione dei sensori di start e stop
- la tabella delle N misure, e i loro scarti
- il tempo medio e la sua deviazione standard
- l'ideogramma delle misure
- l'istogramma degli scarti con sovrapposta la distribuzione gaussian di confronto
- il calcolo del chi-quadro
- breve commento finale se opportuno (ad es. la distribuzione ottenuta è effettivamente normale? L'ideogramma evidenza problemi nel corso della misura? ...)

si riporteranno anche, esplicitamente
- il valore Δz dell'intervallo scelto per l'istogramma
- il valore della costante di normalizzazione della gaussiana

La relazione dovrebbe essere **un documento Word**, o equivalente, comprensivo di tabelle e figure, e può essere inviata per posta elettronica.

Appendice.

Consigli per costruire l'istogramma e la gaussiana in EXCEL.

Supponiamo di aver effettuato N misure di tempo (t_K , con k=1...N) , di cui si è calcolata la media $\langle t \rangle$, la deviazione standard σ gli scarti dalla media $z_K = t_K - \langle t \rangle$.

- Riportare le N misure (tipicamente 200) in una colonna EXCEL. Supponiamo, per fissare le idee, che siano nella colonna C, dalla casella C4 alla C203 (cioè nell'intervallo C4:C203). Nella colonna A si è riportato il numero progressivo corrispondente.
- calcolare la media e la deviazione standard delle misure. Ciò ottiene in EXCEL con le funzioni MEDIA(C4:C203) e DEV.ST(C4:C203). Cioè scrivendo, ad es. nella casella G2 "= MEDIA(C4:C203)" e analogamente per la deviazione standard (nell'esempio in casella G3). Non è necessario scrivere queste formule a mano: basta scrivere nella casella selezionata il segno "=", quindi selezionare la funzione desiderata nel menu funzioni (premere " v " a fianco di "Σ" sulla barra degli strumenti) e selezionare con il mouse le caselle da considerare;
- calcolare gli scarti, ad es. nella colonna D, da D4 a D203. Il modo più semplice è scrivere nella casella D4 "=B4 - G2" e trascinando la casella fino alla D203. Nel trascinamento il riferimento B4 viene aggiornato a B5, B6 ... mentre il $ garantisce che il riferimento alla casella G2 (media) non viene modificato.

Ora ci sono tutti gli elementi per disegnare l'**ideogramma** come in fig.2. Conviene usare la modalità Dispers(X,Y), quindi selezionare Serie, Aggiungi, scegliendo come valori X il numero progressivo di misura e come Y gli scarti. Per cambiare il formato dei dati sul grafico (ad es. eliminare la linea che li congiunge, cambiare forma e colore ...) "cliccare" su un punto qualsiasi con il pulsante destro del mouse e selezionare "Formato serie di dati".
Per inserire le linee corrispondenti a +/- 1 deviazione standard, creare una tabella come la seguente, dove la 2ª colonna corrisponde a +σ, la 3ª a –σ:

0	0,096	-0,096
200	0,096	-0,096

Aggiungere questi dati in tabella selezionando Aggiungi come in precedenza.

Per costruire l'**istogramma** è necessario definire una serie di intervalli di larghezza opportuna Δz, e trovare il numero di misure che cade in ognuno di questi intervalli.

- Scegliere un valore opportuno Δz di larghezza degli intervalli, per esempio la metà della deviazione standard (nell'esempio di fig. 4 il valore Δz (nella casella H5) è definito proprio come "=H4/2"
- costruire una colonna degli intervalli, ad es. da -10Δz a +10Δz, ciò si può fare preparando dapprima una colonna di interi da -10 a +10, nell'esempio H9:H29, e a fianco (I9:I29) la colonna degli intervalli [Basta scrivere in I9 " =H9*G$4" e trascinare la casella fino a I29.
- Ora abbiamo la "matrice dei dati" (D4:D203) e la "matrice delle classi" (I9:I29), con le quali calcolare il numero di eventi nei diversi intervalli (classi). Il modo più semplice è utilizzare la funzione FREQUENZA di EXCEL.
- Nella casella J8 selezionare la funzione FREQUENZA(matrice_dati; matrice_classi), ovvero scrivere "=FREQUENZA(D4:D203;I9:I29)". Per usare la funzione in modalità "matrice" bisogna a questo punto
  - con il mouse, selezionare le caselle da J9 a J29. **Nota**: qui non si deve "trascinare" la funzione ma solo selezionare le caselle)
  - premere il tasto F2
  - premere contemporaneamente Ctr - Shift - Enter.
- Ora abbiamo i dati per l'istogramma, ma conviene calcolare i valori medi fra gli estremi dei diversi intervalli: in K9 si scriva "=(I9+I10)/2" e si trascini fino a L29. Questo perché nella casella J9 la funzione frequenza conta il numero di valori che cadono fra I9 e I10 ecc.

- Come "tipo di grafico" si sceglie "istogramma" e come dati l'intervallo J9:J29 (le "etichette" in ascissa saranno K9:K29.

Bisogna ora calcolare la distribuzione normale o gaussiana da confrontare con i dati. Nell'esempio essa è calcolata nei punti L9:L29.

- Definire in casella K5 "=DISTRIB.NORM.N(L21; media; dev.std; 0)*N*Δz". Dove "media"=0 se stiamo considerando gli scarti, "dev.std" può essere copiata manualmente o inserire il riferimento alla casella in cui si trova il valore (nel nostro esempio G3, col $ per i motivi spiegati sopra). N è il numero di misure (200 nel nostro esempio), e Δz sarà G4. Si moltiplica per N*Δz per far sì che la distribuzione normale (per sè di area unitaria) abbia la stessa area dell'istogramma.

Per fare la figura selezionare la modalità "Istogramma" quindi Serie, Aggiungi. **Attenzione**: in questa modalità non si inseriscono i valori X,Y, ma solo i valori Y (Valori: nell'esempio M5:M25) mentre in ascissa il programma assume numeri successivi a partire da 1. Per mettere le "etichette" giuste inserire i valori L5:L25 [Etichette asse categorie (X)]. Il risultato sono due istogrammi; per tracciare una linea "cliccare" sul secondo istogramma (gaussiana) col pulsante destro del mouse, selezionare "Tipo di grafico" e quindi Linee. Per questioni di stile (colore e spessore della linea, eventuale eliminazione dei punti ecc.) cliccare su uno dei punti (o sulla linea) e selezionare "Formato serie di dati" ecc.

	A	B	C	D	E	F	G
1							
2						media	0,8905
3		**k**	**t_k**	**z_k**		dev. st.	0,0026
4		1	0,8906	0,00007		Δz	0,0013
5		2	0,8905	-0,00003			
6		3	0,8874	-0,00313			
			. . . . .				
200		197	0,8894	-0,00113			
201		198	0,8894	-0,00113			
202		199	0,8882	-0,00233			
203		200	0,8884	-0,00213			

Fig. 4. Tabella delle misure, corrispondente per ipotesi alle colonne B, C, D.

	G	H	I	J	K	L	M	N
8				0				
9		-10	-0,0131	0	-0,0124	0,001		
10		-9	-0,0118	0	-0,0111	0,005		
11		-8	-0,0105	0	-0,0098	0,035		
12		-7	-0,0092	0	-0,0085	0,203		
13		-6	-0,0079	2	-0,0072	0,909		

14	-5	-0,0065	4	-0,0059	3,174	-0,826	0,21498
15	-4	-0,0052	7	-0,0046	8,628	1,628	0,30709
16	-3	-0,0039	18	-0,0033	18,265	0,265	0,00384
17	-2	-0,0026	30	-0,002	30,114	0,114	0,00043
18	-1	-0,0013	35	-0,0007	38,667	3,667	0,34773
19	0	0	45	0,00065	38,667	-6,333	1,0373
20	1	0,00131	32	0,00196	30,114	-1,886	0,11815
21	2	0,00262	15	0,00327	18,265	3,265	0,58361
22	3	0,00393	8	0,00458	8,628	0,628	0,04567
23	4	0,00524	3	0,00589	3,174	0,174	0,00954
24	5	0,00655	0	0,0072	0,909		
25	6	0,00786	1	0,00851	0,203		
26	7	0,00917	0	0,00982	0,035		
27	8	0,01048	0	0,01113	0,005		
28	9	0,01179	0	0,01244	0,001		
29	10	0,0131	0	0,01375	0,000		
30	11	0,01441					
31							
32						χ^2	**2,67**
33						P(2.67,8)	**0,95**

Fig. 5. Tabella utilizzata per costruire l'istogramma, con la curva della distribuzione normale. Il numero di riga si legge a sinistra, quello della colonna in alto. Le ultime due colonne corrispondono al calcolo del "chi-quadro".

Resta da calcolare il **chi-quadro** (χ^2) che ci serve come stima di quanto bene (o male) la distribuzione sperimentale approssima la distribuzione normale. Utilizziamo soltanto gli intervalli centrali, nell'esempio da -5Δz a +5Δz (11 valori). La colonna M rappresenta le differenze fra i valori dell'istogramma e quelli della gaussiana: l'elemento M14 corrisponde a "L14-J14" e analogamente gli altri. La casella N14 corrisponde invece a "=M14^2/J14" e la casella N32 corrisponde alla somma di questi quadrati "=somma(N14:N23)".
Il valore così ottenuto ci dirà se il nostro istogramma è compatibile con una gaussiana, nonostante le differenze, oppure differisce in modo significativo. Per far ciò si deve prendere in considerazione il numero di "gradi di libertà", che nel nostro caso è il numero di punti presi in considerazione (11 nell'esempio) meno il numero di parametri aggiustati (3). Nell'esempio, i gradi di libertà sono 8 e χ^2=2,67.
Si deve valutare la probabilità che il chi-quadro, con 8 gradi di libertà, sia superiore al valore trovato, cosa che possiamo fare utilizzando la funzione di EXCEL come in casella N33: "=1-DISTRIB.CHI.QUAD(valore misurato; gradi di libertà; 1)" (Nell'esempio il valore misurato è 2,67 e i gr. lib. sono 8). Non entriamo nei dettagli di questa formula, diciamo solo che rappresenta la probabilità che il chi-quadro con 8 gradi di libertà superi il valore misurato.
Se questa probabilità è maggiore di 0,05 (5%) diremo che l'istogramma è compatibile con una gaussiana; se è minore di 0,05 diremo che differisce significativamente, mentre se fosse inferiore a 0,01 (1%) diremo che differisce in modo molto significativo.
In alternativa al calcolo con EXCEL si aggiunge una tabella dei valori critici, in funzione dei gradi di libertà.

	10%	5%	2%	1%	0,5%	0,2%	0,1%
n=**1**	2,71	3,84	5,41	6,63	7,88	9,55	10,83
2	4,61	5,99	7,82	9,21	10,60	12,43	13,82
3	6,25	7,82	9,84	11,34	12,84	14,80	16,27
4	7,78	9,49	11,67	13,28	14,86	16,92	18,47
5	9,24	11,07	13,39	15,09	16,75	18,91	20,52
6	10,64	12,59	15,03	16,81	18,55	20,79	22,46
7	12,02	14,07	16,62	18,47	20,28	22,60	24,32
8	13,36	15,51	18,17	20,09	21,96	24,35	26,12
9	14,08	16,92	19,68	21,67	23,59	26,06	27,88
10	15,99	18,31	21,16	23,21	25,19	27,72	29,59
11	17,27	19,68	22,62	24,72	26,76	29,35	31,27
12	18,55	21,03	24,05	26,22	28,30	30,96	32,91
13	19,81	22,36	25,47	27,69	29,82	32,54	32,53
14	21,06	23,68	26,87	29,14	31,32	34,09	36,12
15	22,31	25,00	28,26	30,58	32,80	35,63	37,70

Tabella 1. Valori critici del chi-quadro corrispondenti alle probabilità critiche riportate nella riga superiore, in funzione dei gradi di libertà (colonna a sinistra).

Nell'esempio (tratto da una misura reale), la probabilità vale 0,95 che è ben maggiore del valore critico di 0,05. Diremo pertanto che i risultati sono in buon accordo con l'ipotesi di una distribuzione gaussiana.
Utilizzando la tabella avremmo osservato che, con 8 gr. lib., il chi-quadro trovato di 2,67 è minore del valore critico di 15,51 e anzi è minore anche del valore 13,36 corrispondente ad una probabilità del 10%. Quindi possiamo parlare di buon accordo.
In generale, se il chi-quadro calcolato è minore del valore riportato nella colonna gialla l'ipotesi di distribuzione gaussiana è confermata; se è compreso fra il valore della colonna gialla e quella rosa la differenza è significativa; se è maggiore del valore della colonna rosa diremo che la differenza è molto significativa.

Osservazione: non è affatto raro che il risultato trovato in questa esperienza sia incompatibile con una distribuzione gaussiana: lo studente non sarà penalizzato per questo ma solo per gli eventuali errori nella trattazione dei dati.

Nota finale. Si ricorda che il "valore critico" di 0,05 (o 5,0%) come pure quello dell' 1% sono solo delle convenzioni, quelle che si utilizzano più spesso, ma che si possono usare convenzioni diverse.