

Capitolo 10

Test del χ^2

I metodi di stima dei parametri esposti nei capitoli precedenti permettono di ottenere i valori dei parametri incogniti di una funzione in modo da approssimare nel modo migliore i dati sperimentali, tuttavia questi metodi non danno alcuna indicazione su come la funzione ottenuta si adatti ai dati ovvero non danno alcuna indicazione sulla *qualità* del *fit*.

Per valutare la qualità di un *fit* è estesamente utilizzato il cosiddetto *test del χ^2* . Questo test, se sono soddisfatte alcune condizioni, dà una valutazione probabilistica della “bontà del fit¹” attraverso la distribuzione di densità di probabilità del χ^2 , già descritta nel paragrafo 5.9.7.

Il test del χ^2 ha due forme, una che si applica quando si prende in considerazione una dipendenza funzionale tra due grandezze e un'altra che si applica a risultati esprimibili con istogrammi.

10.1 Test del χ^2 per relazioni funzionali

Il test del χ^2 si utilizza per la valutazione della qualità del *fit* ottenuto da dati sperimentali con l'uso di metodi di stima puntuale oppure per verificare se una teoria, anche senza parametri da determinare, è supportata o meno dai dati sperimentali. Si consideri una grandezza fisica Y misurata in funzione di un'altra grandezza X e sia $Y = g(X|\hat{\Lambda})$ la funzione adattata ai dati (il *fit*); $\hat{\Lambda}$ rappresenta l'insieme dei k parametri i cui valori sono stati ottenuti con uno dei metodi di stima puntuale oppure sono determinati con altri metodi ad esempio con ipotesi teoriche. Siano $y_i \pm u_i$ le misure di Y effettuate in corrispondenza a N valori x_i , ($i = 1, \dots, N$) noti con incertezza trascurabile. Supponiamo inoltre che le misure y_i abbiano una distribuzione *normale* di valore medio $g(x_i|\hat{\Lambda})$, valore atteso per la variabile y_i secondo il *fit* ottenuto, e deviazione standard u_i . La figura 10.1 esemplifica quanto esposto: le curve a campana disegnate attorno ai valori attesi delle y_i rappresentano le distribuzioni normali delle y_i riportate nell'asse delle ordinate. Con le ipotesi fatte la variabile $z_i = (y_i - g(x_i|\hat{\Lambda}))/u_i$ è una variabile normale standardizzata con distribuzione $\mathcal{N}(0, 1)$ e quindi la variabile:

$$\chi^2 = \sum_{i=1}^N z_i^2 = \sum_{i=1}^N \frac{(y_i - g(x_i|\hat{\Lambda}))^2}{u_i^2} \quad (10.1)$$

segue la distribuzione di probabilità del χ^2 con un numero di gradi di libertà dato dal numero N dei punti sperimentali da cui si deve eventualmente sottrarre il numero dei k parametri

¹In inglese *Goodness-of-fit* alle volte abbreviato con l'acronimo GOF

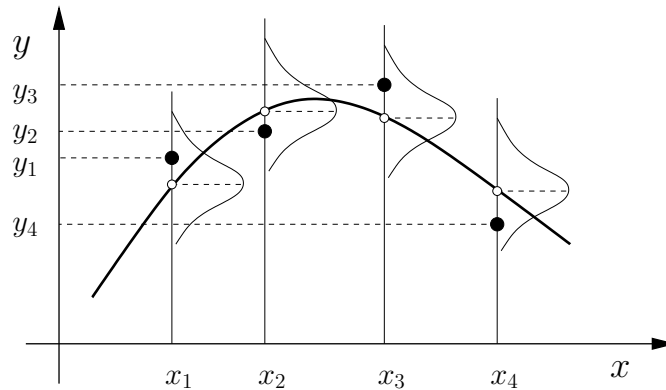


Figura 10.1: Test del χ^2 per relazioni funzionali. I cerchi pieni rappresentano le misure sperimentali, i cerchi vuoti i valori attesi delle y_i e le curve a campana sono le gaussiane che descrivono come si distribuiscono le y_i attorno al valore atteso. Le barre di incertezza sulle y_i non sono state riportate per evitare la sovrapposizione di simboli grafici che avrebbero reso più difficoltosa la comprensione del grafico.

$\hat{\Lambda} : \{\hat{\Lambda}_1, \dots, \hat{\Lambda}_k\}$ stimati nel calcolo del fit *utilizzando gli stessi N punti sperimentali*. In conclusione il numero dei gradi di libertà della variabile χ^2 è $\nu = N - k$. Ad esempio supponiamo che la curva continua nella figura 10.1 sia una parabola: $g(x|\hat{\Lambda}) = \hat{\Lambda}_1 + \hat{\Lambda}_2 x + \hat{\Lambda}_3 x^2$ e che tramite il fit dei 4 punti sperimentali avessimo determinato il valore dei tre parametri $\hat{\Lambda}_1, \hat{\Lambda}_2, \hat{\Lambda}_3$, la variabile χ^2 ottenuta tramite la (10.1) ha $4 - 3 = 1$ grado di libertà.

Una volta calcolato il valore χ_o^2 dell'espressione (10.1) e il numero di gradi di libertà, *se la funzione g ipotizzata descrive correttamente i dati*, possiamo calcolare la probabilità di ottenere casualmente un valore di χ^2 maggiore di quello ottenuto χ_o^2 con il seguente espressione:

$$P(\chi^2 > \chi_o^2) = \int_{\chi_o^2}^{\infty} f_{\chi^2}(\chi^{2'}) d\chi^{2'} \quad (10.2)$$

I valori dell'integrale (10.2) si ottengono tramite tabelle² o con programmi di calcolo³. Parlando qualitativamente ci aspettiamo che se la funzione $g(x|\hat{\Lambda})$ descrive correttamente i dati, $P(\chi^2 > \chi_o^2)$ sia abbastanza grande⁴, e dell'ordine del 50%, mentre valori piccoli di tale probabilità ci dicono che, se la funzione $g(x|\hat{\Lambda})$ descrivesse correttamente i dati, il valore trovato (χ_o^2) sarebbe estremamente improbabile. Quindi se il valore di $P(\chi^2 > \chi_o^2)$ è "piccolo" (convenzionalmente si prende come valore minimo il 5%, ma si parla sempre di probabilità e non di certezza) probabilmente la teoria, rappresentata dalla funzione ($g(x|\hat{\Lambda})$), non si adatta ai dati e quindi è da rigettare.

10.1.1 Il χ^2 ridotto

Notiamo che se la funzione $g(X|\hat{\Lambda})$ descrive correttamente i dati, il valore atteso della variabile χ^2 (10.1) è pari al numero di gradi di libertà ν cioè $\mathbb{E}[\chi^2] = \nu$. Per rendere il valore

²Vedi l'appendice ?? nella quale è tabulato il χ^2 ridotto, ovvero $\tilde{\chi}^2 = \chi^2/\nu$. Per la definizione del χ^2 ridotto vedi il prossimo paragrafo.

³ Tra i programmi per il calcolo dell'integrale (10.2) si può utilizzare EXCEL o il suo equivalente *open source*

⁴Estremizzando, se i punti sperimentali coincidessero esattamente con i valori previsti avremo $\chi_o^2 = 0$ e quindi una probabilità pari a 1 di ottenere in un esperimento successivo un valore maggiore di χ^2 . Tuttavia vedremo che anche valori troppo piccoli di χ^2 sono indice di un problema nella valutazione della qualità del fit. Vedi il successivo Esempio 3 a pagina 130.

atteso del χ^2 indipendente dal numero dei gradi di libertà è stato introdotto il χ^2 ridotto definito dalla relazione

$$\tilde{\chi}^2 = \frac{\chi^2}{\nu} \quad (10.3)$$

E' facile vedere che, qualsiasi sia il valore dei gradi di libertà, il valore atteso di $\tilde{\chi}^2$ è $\mathbb{E}[\tilde{\chi}^2] = 1$. Il χ^2 ridotto è utilizzato per semplificare la stesura di alcuni tipi di tabelle come ad esempio quelle citate nel paragrafo precedente.

Esempio 1. In un esperimento si misurano due grandezze tra le quali si ipotizza che esista una relazione lineare⁵: $y = a + bx$. Le misure ottenute nell'esperimento sono riportate nelle prime due colonne della tabella 10.1. Eseguito il *fit* con il metodo dei minimi quadrati si

Tabella 10.1

x (u.a.)	y (u.a.)	y_{th}	residuo	(residuo) ²
10	11 ±1	12.0	0.969	0.938
20	17 ±1	16.6	-0.353	0.125
30	21 ±1	21.3	0.326	0.106
40	28 ±1	26.0	-1.996	3.984
50	30 ±1	30.7	-0.683	0.466
60	36 ±1	35.4	-0.639	0.408
70	39 ±1	40.0	1.040	1.081
				$\chi^2 = 7.11$

ottengono i valori dei parametri a e b ed è possibile confrontare la funzione trovata con i punti sperimentali come è mostrato in figura 10.2. Il valore del χ^2 si ottiene sommando in quadratura i residui $R_i = (y_i - a - bx_i)/u_i$ (vedi l'ultima colonna della tabella 10.1). Il calcolo del χ^2 per questi dati è 9.11. Poiché il numero di gradi di libertà è $\nu = 7 - 2 = 5$, dalle tabelle si ricava $P(\chi^2 > 9.11) = 10\%$ valore che rende statisticamente accettabile la descrizione dei dati fornita dal *fit*.

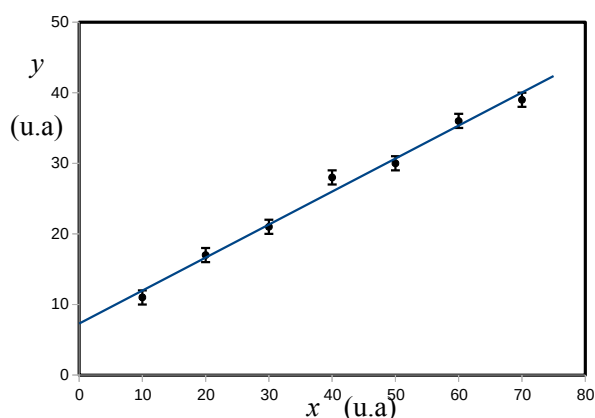


Figura 10.2: Grafico dei dati in Tabella 10.1. Il *fit* dei dati con il metodo dei minimi quadrati è: $0.460x + 6.70$.

⁵Per concretezza, la grandezza indicata con x potrebbe essere il valore della massa applicata ad una molla e quella indicata con y il relativo l'allungamento.

Esempio 2. Consideriamo un'altra serie di coppie di dati e anche tra queste si ipotizzi che esista una relazione lineare. Analogamente all'esempio precedente compiliamo la tabella 10.2 nella quale le prime due colonne contengano i valori misurati delle due grandezze. Nella

Tabella 10.2

x (u.a.)	y (u.a.)	y_{th}	residuo	(residuo) ²
10	6.3 ±1	4.1	-2.204	4.859
20	9.4 ±1	8.9	-0.500	0.250
30	11.7±1	13.7	1.972	3.888
40	16.7±1	18.5	1.810	3.275
50	22.5±1	23.3	0.815	0.663
60	27.3 ±1	28.1	0.786	0.618
70	35.6 ±1	32.9	-2.676	7.162
				$\chi^2 = 20.71$

figura 10.3 sono riportati i punti sperimentali e il fit lineare ottenuto con il metodo dei minimi quadrati. Il calcolo del χ^2 per questi dati è 20.71. Poiché il numero di gradi di libertà anche in questo caso è $\nu = 7 - 2 = 5$, dalle tabelle si ricava $P(\chi^2 > 20.71) = 0.1\%$ valore molto minore del minimo 5% usualmente accettato come limite inferiore. Si deduce che la retta non descrive correttamente i dati. Il rigetto della descrizione lineare dei dati è ulteriormente confermato dall'analisi dei residui come sarà mostrato nel prossimo paragrafo.

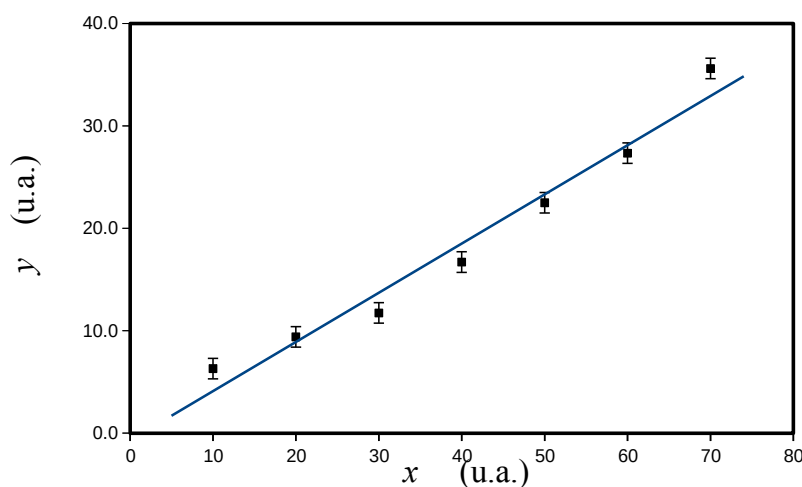


Figura 10.3: Grafici dei dati in Tabella 10.2. Il fit dei dati con una retta ha come risultato: $0.480x - 0.709$, da cui si ricava $\chi^2 = 20.71$, $\nu = 5$, $P(\chi^2 > 7.11) = 0.1\%$. Poiché $P(\chi^2 > 7.11) < 5\%$, si deduce che il fit non descrive adeguatamente i dati.

Esempio 3. riprendiamo in esame i dati dell'esempio 1 supponendo ora che in modo erroneo le incertezze siano state valutate il triplo di quelle corrette (3 u.a. invece di 1 u.a.). E' facile calcolare il nuovo valore di χ^2 come $7.11/9 = 0.79$. Dalle tabelle ricaviamo che $P(\chi^2 > 0.79) = 98\%$. Questo valore, per quanto rispetti in pieno il criterio dato nel paragrafo 10.1 è indice di un problema. Infatti *sopravvalutare le incertezze è il principale*

motivo per cui il χ^2 prende valori piccoli. Se la funzione g è quella che descrive correttamente i dati e se le incertezze sono state valutate correttamente il valore atteso della variabile aleatoria χ^2 è dato dal numero dei gradi di libertà ν . Valori che si discostano in modo marcato sia in positivo ma anche in negativo dal valore atteso sono indice che l'esperimento, i dati e l'analisi vanno riesaminati criticamente.

10.1.2 Analisi dei residui

Nell'analisi dei dati può essere utile osservare il grafico dei residui, dati dall'espressione

$$\frac{y_i - g(x_i)}{u_i}$$

in funzione di x_i . Se la funzione con la quale descriviamo i dati è quella "vera", l'andamento dei residui deve essere del tutto casuale. Se al contrario, l'andamento dei residui mostra una *struttura*, ovvero un andamento che abbia una certa regolarità, allora possiamo dedurre che la funzione con cui confrontiamo i dati è inadeguata e deve essere cambiata. Ad esempio consideriamo gli esempi 1 e 2 del paragrafo precedente. Nella figura 10.4 sono mostrati entrambi. In quello relativo all'esempio 1 i valori dei residui variano in modo casuale attorno allo zero senza evidenziare un andamento che possa essere approssimato da una funzione e quindi questo grafico è un'ulteriore conferma che il *fit* ottenuto descrive correttamente i dati. Al contrario nel grafico relativo all'esempio 2, in cui i dati sono stati approssimati da una retta, si riconosce facilmente un andamento dei residui prima crescente e poi decrescente. Questa osservazione suggerisce che l'aggiunta di un termine quadratico alla funzione lineare potrebbe migliorare la qualità del *fit*.

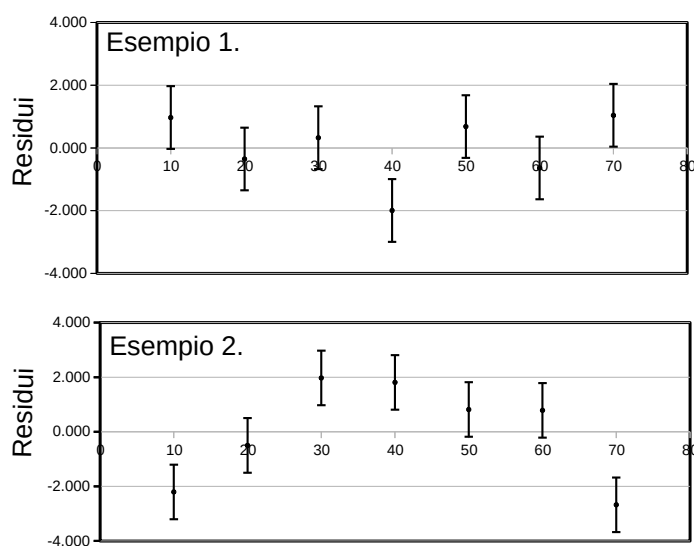


Figura 10.4: Andamento dei residui per i due fit del paragrafo precedente Esempio 1 e Esempio 2 in funzione di x_i .

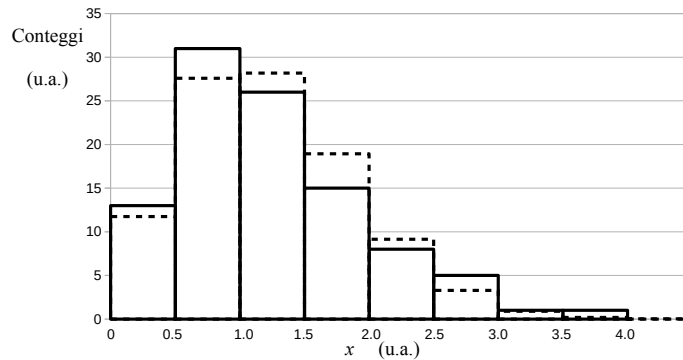


Figura 10.5: Istogramma di 100 realizzazioni di una variabile aleatoria continua raggruppate in *bin* (linea continua). In ciascuno dei *bin* sono contenuti O_i eventi osservati. La linea tratteggiata rappresenta invece il valore teorico atteso dei conteggi E_i . Per questo esempio la distribuzione di riferimento scelta è $f(x) = x \exp(-x^2/2)$. I *bin* sono equispaziati, ma per ottenere un valore significativo della probabilità del χ^2 nel confronto tra dati osservati e dai attesi, i *bin* meno popolati (quelli per $x > 2$) vanno raggruppati come indicato nel testo.

10.2 Test del χ^2 di Pearson per istogrammi

L'analisi dei fenomeni fisici, come quelli statistici, si basa molto spesso sull'uso degli istogrammi nei quali i valori delle variabili aleatorie sono raggruppati in classi o *bin*. Consideriamo ad esempio la misurazione di una determinata grandezza X ripetuta N volte (x_1, x_2, \dots, x_N) . Per costruire l'istogramma di queste misurazioni dividiamo l'asse reale in *bin* tipicamente di eguale dimensione⁶ in modo tale che in ognuno dei *bin* compaia un numero *congruo*⁷ di misurazioni. L'operazione di raggruppamento dei dati in *bin* fa apparire la distribuzione della variabile X come discreta anche se X è continua. Supponiamo di volere confrontare una distribuzione di dati sperimentali di una variabile continua raggruppati in n *bin* con la distribuzione di riferimento $f(X)$ che si ritiene descriva correttamente i dati. Nel *bin* i -esimo, nel quale sono stati raggruppati O_i eventi, gli eventi attesi, E_i sono dati da:

$$E_i = Np_i \quad \text{con} \quad p_i = \int_{x_i}^{x_{i+1}} f(X)dX \quad (10.4)$$

dove x_i e x_{i+1} sono gli estremi dell' i -esimo *bin* e N è il numero totale degli eventi osservati. Nella figura 10.5 è esemplificato quanto detto: L'istogramma a linea continua rappresenta le osservazioni sperimentali mentre quello a linea tratteggiata indica la previsione teorica con cui si voglio confrontare i dati sperimentali. Si noti che le E_i sono descritte dalla distribuzione multinomiale⁸ (che è una generalizzazione della distribuzione binomiale). Se consideriamo un particolare *bin* la distribuzione del numero di eventi attesi in quel *bin* è una binomiale il cui valore medio è Np_i . Per N sufficientemente alto questa binomiale tende alla distribuzione di Poisson con media e varianza pari Np_i . Come è noto al crescere di N la distribuzione di Poisson tende ad una distribuzione normale di valore medio Np_i e deviazione standard

⁶ La uguale dimensione di bin non è obbligatoria anzi, come vedremo, in certe circostanze è necessario utilizzare *bin* di dimensione differente.

⁷ La scelta della dimensione del *bin* è un aspetto delicato di questa procedura e torneremo su questo punto nel seguito

⁸vedi il paragrafo 5.11.4

$\sqrt{Np_i}$). Sotto le condizioni indicate, Pearson ha dimostrato che la grandezza

$$\chi^2 = \sum_{i=1}^n \frac{(Np_i - O_i)^2}{Np_i}$$

asintoticamente (ovvero al crescere di N) segue la distribuzione del χ^2 con $\nu = n - 1$ gradi di libertà. Il numero dei gradi di libertà è pari al numero dei *bin* in cui sono stati raggruppati i dati sperimentali meno uno. La diminuzione di una unità dei gradi di libertà è giustificata dalla osservazione che il numero totale degli eventi (N) è fissato e quindi il contenuto dell'ultimo *bin* è determinato dal contenuto di tutti gli altri. Ovviamente se la distribuzione con la quale confrontiamo i dati avesse k parametri stimati da una procedura di *fit* allora il numero dei gradi di libertà della variabile χ^2 sarebbe $\nu = n - k - 1$. Come regola pratica per un'applicazione corretta del test del χ^2 di Pearson il contenuto di ogni *bin* deve essere maggiore o uguale a 6.

Esempio. La misurazione del periodo T di un pendolo fisico è eseguita con un apparato provvisto di un traguardo ottico collegato ad un cronometro la cui risoluzione è $10 \mu\text{s}$. Sono stati acquisiti 198 valori del periodo delle piccole oscillazioni e l'istogramma dei valori è mostrato nella figura 10.6. Media e deviazione standard dei valori acquisiti sono $\bar{T} = 1.51196 \text{ s}$ e $s = 0.00012 \text{ s}$. Ci chiediamo se la distribuzione dei valori ottenuti è compatibile con una distribuzione normale con valore medio e deviazione standard ottenuti dai valori calcolati. Come prima operazione trasformiamo le grandezze T_i misurate, nelle corrispondenti variabili normalizzate z_i con la relazione:

$$z_i = \frac{T_i - \bar{T}}{s} \quad (10.5)$$

Il nuovo asse è rappresentato al di sotto di quello originale nella figura 10.6. Il passo successivo è calcolare per ogni *bin* il contenuto di eventi atteso, che nell'ipotesi fatta della distribuzione normale, è:

$$E_i = N \int_{z_i}^{z_{i+1}} \frac{1}{\sqrt{2\pi}} e^{-t^2/2} dt = \frac{N}{2} \left[\text{erfc} \left(\frac{z_{i_1}}{\sqrt{2}} \right) - \text{erfc} \left(\frac{z_{i_{1+1}}}{\sqrt{2}} \right) \right] \quad (10.6)$$

dove $\text{erfc}(x)$ è la funzione complementare degli errori (vedi pagina 73).

Esaminando la tabella 10.3 si nota che alcuni *bin* hanno un contenuto inferiore a 5 per cui, per applicare in modo corretto il test del χ^2 di Pearson, devono essere accorpati fino a raggiungere una numerosità maggiore di 5. Allo scopo i valori attesi dei primi 5 e degli ultimi 3 *bin* sono stati sommati tra loro e confrontati con la somma dei corrispondenti valori osservati (vedi la tabella 10.3). In queste condizioni si ottiene $\chi^2 = 2.08$. Il numero dei gradi di libertà è $\nu = 2$ (5 *bin*, 2 parametri stimati dai dati) con la probabilità $P(\chi^2 > 2.08) = 0.36$. Questo valore conferma l'ipotesi di normalità della distribuzione dei periodi del pendolo su cui sono state effettuate le misurazioni, almeno per misure di T che distano meno di due deviazioni standard dalla media.

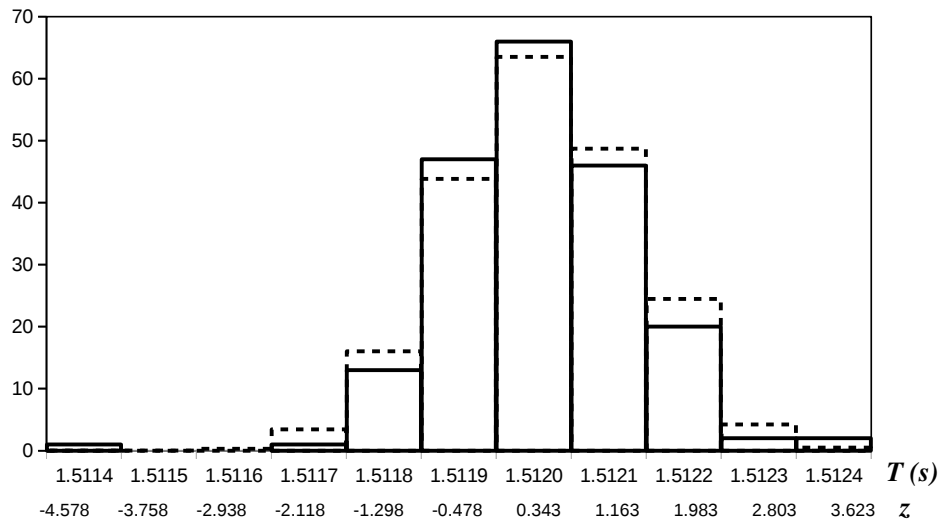


Figura 10.6: Istogramma delle misure del periodo di un pendolo. Sono rappresentati due assi: in alto i valori del periodo in secondi, in basso la corrispondente variabile $z = (T - \bar{T})/s$. I numeri indicano l'estremo inferiore del *bin*.

Tabella 10.3: Tabella per l'esecuzione del test di χ^2 di Pearson sui dati del periodo di oscillazione di un pendolo fisico. Nella colonna dei valori attesi (E_i) i numeri in corsivo si riferiscono ai dati accorpati come spiegato nel testo

T_i (s)	z_i	O_i	E_i	$(O_i - E_i)^2/E_i$
1.5114	-4.578	1	—	
1.5115	-3.758	0	—	
1.5116	-2.938	0	—	
1.5117	-2.118	1	—	
1.5118	-1.298	13	19.2	0.936
1.5119	-0.478	47	43.4	0.296
1.5120	0.343	66	62.9	0.156
1.5121	1.163	46	48.2	0.101
1.5122	1.983	20	24.2	0.002
1.5123	2.803	2	—	
1.5124	3.623	2	—	
				$\chi^2 = 2.06$