Comprende versione **ebook**

Gaetano Cannelli

Metodologie sperimentali in Fisica

Introduzione al metodo scientifico

IV Edizione

EdiSES università

Accedi ai contenuti **digitali**

Espandi le tue risorse

un libro che **non pesa**
e si **adatta** alle dimensioni
del **tuo lettore**!

COLLEGATI AL SITO **EDISES.IT**

ACCEDI AL **MATERIALE DIDATTICO**

SEGUI LE **ISTRUZIONI**

Utilizza il codice personale contenuto nel riquadro per registrarti al sito **edises.it** e accedere ai contenuti digitali.

Scopri il tuo **codice personale** grattando delicatamente la superficie

Il volume NON può essere venduto, né restituito, se il codice personale risulta visibile.
L'**accesso ai contenuti digitali** sarà consentito **per 18 mesi**.

Per attivare i **servizi riservati**, collegati al sito **edises.it** e segui queste semplici istruzioni

Se sei registrato al sito

- clicca su *Accedi al materiale didattico*
- inserisci email e password
- inserisci le ultime 4 cifre del codice ISBN, riportato in basso a destra sul retro di copertina
- inserisci il tuo **codice personale** per essere reindirizzato automaticamente all'area riservata

Se non sei già registrato al sito

- clicca su *Accedi al materiale didattico*
- registrati al sito **edises.it**
- attendi l'email di conferma per perfezionare la registrazione
- torna sul sito **edises.it** e segui la procedura già descritta per *utenti registrati*

EdiSES università

I contenuti digitali sono accessibili dalla propria **area riservata** secondo la procedura indicata nel frontespizio.

Dalla sezione **materiali e servizi** della tua area riservata potrai accedere all'**Ebook**, ovvero la versione digitale del testo in formato epub, standard dinamico che organizza il flusso di testo in base al dispositivo sul quale viene visualizzato. Fruibile mediante l'applicazione gratuita BookShelf, consente una visualizzazione ottimale su lettori e-reader, tablet, smartphone, iphone, desktop, Android, Apple e Kindle Fire.

L'accesso ai contenuti digitali sarà consentito per **18 mesi**.

EdiSES università

Gaetano Cannelli

Metodologie sperimentali in Fisica

Introduzione al metodo scientifico

Corso propedeutico alla pratica di laboratorio per gli studenti dei Corsi di Laurea in Scienze, Farmacia ed Ingegneria.
Misure ed incertezze, elaborazione statistica dei dati, formulazione di leggi

IV Edizione

EdiSES università

Metodologie sperimentali in Fisica IV Edizione
Introduzione al metodo scientifico.
Corso propedeutico alla pratica di laboratorio per gli studenti dei
Corsi di Laurea in Scienze, Farmacia ed Ingegneria.
Misure ed incertezze,elaborazione statistica dei dati, formulazione di leggi
GAETANO CANNELLI

9 8 7 6 5 4 3 2 1 0
2028 2027 2026 2025 2024

Le cifre sulla destra indicano il numero e l'anno dell'ultima ristampa effettuata

Stampato presso: Vulcanica S.r.l. – Nola (NA)

per conto della EdiSES Edizioni S.r.l. – Piazza Dante, 89 – Napoli

www.edises.it
assistenza.edises.it

ISBN 978 88 3623 202 4

Dedica dell'Autore

A mia moglie e a mio figlio

Prefazione alla prima edizione

Il libro è rivolto agli studenti delle Facoltà di Scienze, Farmacia ed Ingegneria che frequentano per la prima volta un corso di laboratorio di Fisica e a chiunque debba occuparsi per motivi professionali di scienza sperimentale, includendo tra questi anche i laureati in medicina.

Lo scopo primario che il libro si prefigge è quello di insegnare il "metodo scientifico" che è alla base di ogni scienza moderna. Nella mia esperienza di docente ho avuto modo di constatare che anche gli studenti delle Facoltà scientifiche, con l'esclusione dei laureandi in Fisica quando preparano la loro tesi, arrivano alla laurea senza avere ancora chiaramente inteso con quale metodo procede la Scienza. Sinceramente devo ammettere che di questo stato di cose siamo responsabili anche noi fisici che poco tempo dedichiamo agli aspetti epistemologici dell'attività scientifica e finiamo per impartire un insegnamento dogmatico della Fisica presentando i suoi principi e le sue leggi come verità indiscutibili.

Il contenuto del libro è stato organizzato in tre parti avendo in mente il riordino degli studi universitari, in particolare la laurea di primo livello, e le differenti esigenze di maggiore/minore approfondimento dei diversi corsi di laurea.

*La **I PARTE** e la **III PARTE** sono strettamente connesse: esse corrispondono nel riordino degli studi universitari ad un "modulo" (Unità Didattica, U.D.) di 5 crediti, che richiede da 35 a 40 ore di lezioni includendo alcune ESPERIMENTAZIONI da eseguire in laboratorio. Queste due parti insieme possono anche costituire un corso di insegnamento compiuto che fornisce il minimo bagaglio culturale indispensabile a chi deve occuparsi di scienza sperimentale. In particolare, la **I PARTE** del libro munisce lo studente degli strumenti necessari per quantificare l'osservazione di un fenomeno naturale attraverso il concetto di grandezza fisica e la sua misura. Ciò comporta la definizione di sistemi di unità di misura: tra questi, il Sistema Internazionale, oggi raccomandato dalla comunità scientifica, viene introdotto dopo una breve rassegna dei sistemi di unità che si sono succeduti nel tempo per rendere più comprensibili le motivazioni di certe definizioni operative. Si introducono quindi le idee fondamentali dell'analisi degli errori mettendo in risalto come ogni misura sia affetta da incertezza, che ne determina tra l'altro le cifre significative per esprimerla, e come la sua conoscenza sia indispensabile per estrarre una legge fisica dai dati sperimentali. In particolar modo viene trattato l'errore casuale e le diverse situazioni in cui questo errore si presenti e come possa essere valutato il suo valore massimo. Per quanto concerne l'errore standard di una misura, vengono fornite le fondamentali espressioni per la sua valutazione rinviando la dimostrazione e l'approfondimento del suo notevole significato probabilistico nella **II PARTE**. La **III PARTE** descrive e discute, con lo spirito del "metodo", dodici "ESPERIMENTAZIONI". In questa parte lo studente viene sollecitato ad eseguire l'esperimento non soltanto per verificare quanto appreso dalla lezione teorica, ma anche per ricavare, indipendentemente da essa, attraverso l'acquisizione e l'elaborazione* *dei dati sperimentali la legge fisica del fenomeno in osservazione secondo il procedimento induttivo "del metodo sperimentale galileiano". In alcuni casi, il calcolo dei parametri dell'equazione fisica dedotta rappresenta una diversa maniera per la determinazione indiretta di importanti grandezze fisiche. Per la formulazione della legge fisica vengono proposti due metodi di elaborazione dei dati sperimentali: il metodo grafico, più semplice ed intuitivo (presentato con alcune figure "dal vero" su carta millimetrata) e il metodo analitico che, richiedendo metodi di calcolo più avanzati, è rinviato alla* ***II PARTE****. Per questo modulo è richiesto come strumento di calcolo una comune calcolatrice tascabile programmabile, per l'impiego in laboratorio e per una preliminare elaborazione dei dati.*

La ***II PARTE****, che equivale ad un secondo modulo di 5 crediti, illustra attraverso le distribuzioni di probabilità (binomiale, Gauss, Poisson) il significato probabilistico dell'errore e introduce metodi più avanzati per l'elaborazione statistica dei dati sperimentali (regressione di curve). I metodi statistici appresi in questa parte consentono allo studente di elaborare con maggiore approfondimento i dati sperimentali delle ESPERIMENTAZIONI già acquisiti e preliminarmente trattati con il più semplice metodo grafico. Alcuni argomenti come la propagazione dell'errore standard sono stati esposti in maniera intuitiva sacrificando un'impostazione rigorosa che può essere demandata a testi più specialistici. Questa parte del corso, con l'esclusione delle distribuzioni di Poisson e di Gauss, potrebbe anche non essere considerata essenziale dal punto di vista della comprensione e della applicazione del "metodo" poiché, per l'elaborazione dei dati sperimentali, lo studente potrebbe ricorrere agli opportuni programmi di cui sono muniti i calcolatori. In ogni caso, questa parte pur essendo raccomandata ai fisici, è anche rivolta a matematici, chimici e biologi che volessero rendersi conto di come vengono impiegati i metodi statistici per il trattamento dei dati. Gli argomenti svolti sono stati resi più comprensibili mediante numerosi esempi; quelli ritenuti meno importanti sono segnalati da un asterisco. Tutta la trattazione è alla portata di studenti che posseggono una cultura matematica corrispondente al programma di due o tre moduli di Calcolo.*

Infine qualche precisazione su altre caratteristiche del libro.

Gli esperimenti descritti fanno volutamente riferimento ad una strumentazione il più possibile semplice ed essenziale, qualcuno potrebbe definire "rozza". La ragione di ciò è dovuta al fatto che non si vuole distrarre lo studente dall'aspetto prettamente fisico dell'esperimento con apparati muniti di automatismi sofisticati e inutilmente complicati. Nel primo contatto con il laboratorio è didatticamente raccomandabile che lo studente principiante acquisisca i dati sperimentali ed esegua i grafici manualmente. La gran parte degli esperimenti descritti vertono su argomenti di meccanica e termodinamica; un capitolo è dedicato alla determinazione di resistenze di conduttori metallici e semiconduttori con il metodo dei quattro contatti, un altro capitolo alle distribuzioni di probabilità di Poisson e di Gauss. La scelta degli argomenti è determinata unicamente dalla circostanza che il primo approccio con il laboratorio avviene quando lo studente conosce soltanto tali argomenti. Comunque, dal punto di vista dell'apprendimento e dell'applicazione del "metodo", un esperimento di meccanica o di elettromagnetismo o di fisica moderna sono perfettamente equivalenti.

Tutte le ESPERIMENTAZIONI sono presentate seguendo uno schema costituito dai seguenti paragrafi: 1. Introduzione, 2. Strumentazione, 3. Procedimento, 4. Valutazione dell'errore a priori, 5. Risultati sperimentali, 6. Elaborazione dei dati, 7. Conclusioni. Tale schema è anche riportato in fondo alla ***III PARTE*** *dove sono stati riservati a ciascun paragrafo opportuni spazi del foglio per la stesura della relazione di laboratorio. Tramite questo schema, lo studente è sollecitato a descrivere l'esperimento nella sua essenzialità, ma con la massima chiarezza e completezza. Inoltre, esso facilita considerevolmente la correzione degli elaborati da parte del docente evitandogli la lettura di relazioni prolisse e disarticolate, qualora lo studente fosse lasciato senza una traccia che lo guidi.*

Nel testo non viene data alcuna nozione di programmazione al calcolatore poiché ormai, con il *riordino degli studi universitari, ogni corso di laurea prevede come obbligatori alcuni moduli dedicati all'informatica.*

Oltre alla descrizione di esperimenti da eseguire nel laboratorio, vengono riportati numerosi esempi che propongono la formulazione di equazioni fisiche da serie di dati sperimentali. In alcuni di questi esempi viene fatto osservare, come spesso avviene effettivamente nella ricerca, che la scelta dell' equazione fisica, considerata l'incertezza delle misure, non è sempre univoca ma va fatta in base a valutazioni probabilistiche che possono convalidare o "falsificare" modelli proposti dalla formulazione teorica.

Desidero ringraziare il Prof. R. CANTELLI del Dipartimento di Fisica dell'Università di Roma "La Sapienza" e i Proff. E. COLAVITA, E. LAMANNA del Dipartimento di Fisica dell'Università della Calabria per l'incoraggiamento, gli utili commenti e i suggerimenti avuti nel corso della stesura del testo.

Un particolare riconoscimento va ai dottori R. BARTUCCI e A. PAPA del Dipartimento di Fisica dell'Università della Calabria e al Dr. G.B. CANNELLI dell'Istituto "O.M. Corbino" del CNR che hanno avuto la costanza di leggere l'intero manoscritto segnalando errori, omissioni e stimolandomi a migliorare quelle parti che potevano risultare oscure per uno studente al primo approccio con il laboratorio di Fisica.

Infine sono grato al Prof. G. RUSSO del Dipartimento di Fisica dell'Università della Calabria, al Dr. G. RAELE, e al Dr. F. TREQUATTRINI del Dipartimento di Fisica dell'Università di Roma "La Sapienza" per il loro prezioso aiuto nell'elaborazione di figure e grafici al calcolatore.

GAETANO CANNELLI
Febbraio 2000

Prefazione alla seconda edizione

Il libro come nella prima edizione si propone di insegnare il "metodo scientifico" agli studenti delle Facoltà di Scienze, Farmacia ed Ingegneria enfatizzando ancor più alcuni suoi aspetti. Introdotto il concetto di misura di una variabile, che in questa seconda edizione non è necessariamente sempre una grandezza fisica, viene mostrato come dalla elaborazione di una serie di dati sperimentali si giunga alla formulazione di leggi, ipotesi che possono convalidare o invalidare modelli, teorie, assunzioni preesistenti o anche stimolare la proposizione di nuove idee.

Il contenuto del libro, equivalente nel nuovo ordinamento degli studi universitari a due moduli didattici (di circa 40 ore ciascuno, incluso il laboratorio), è strutturato in maniera che possa essere utilizzato da un ampio numero di studenti con interessi su discipline diverse ma che tutte necessitano dell'uso del "metodo". Il primo modulo, fondamentale per tutti i corsi di laurea, introduce il concetto di misura e i sistemi di unità di misura, tratta l'errore probabilistico attraverso le più importanti distribuzioni di probabilità (binomiale, Poisson, Gauss) e può essere completato con l'esecuzione di alcuni esperimenti, scelti tra quelli della terza parte del libro, i cui dati possono essere elaborati con il semplice metodo grafico. Il secondo modulo, dedicato agli studenti dei corsi di laurea di Fisica, Matematica, Chimica ed Ingegneria, può essere rappresentato dalla seconda parte del libro, più formalistica, dove vengono adottati metodi più avanzati (minimi quadrati, test del "chi quadrato", correlazioni, interpolazioni) per l'elaborazione dei dati degli stessi esperimenti della terza parte.

Questa seconda edizione, rispetto alla prima, si differenzia per le seguenti integrazioni, enfatizzazioni, modifiche.

i) *Tenendo conto delle esigenze degli studenti dei corsi di laurea in Biologia, Scienze Naturali, Farmacia è stato dato maggiore risalto ad elaborazioni di variabili che non siano esclusivamente grandezze fisiche mettendone in risalto aspetti in un contesto più ampio di quello delle applicazioni in Fisica. A tale scopo sono stati presentati alcuni esempi di distribuzioni di variabili in campo biomedico; è stato ampliato il Cap. 10 per mostrare in tale campo l'impiego della distribuzione di Student; inoltre è stata introdotta la distribuzione di Fisher.*

ii) *Gli esperimenti non richiedono una conoscenza approfondita della Fisica Generale che sarà insegnata in moduli successivi. Questa circostanza, contrariamente a quanto si possa credere, rende più favorevole la comprensione del "metodo" da parte dello studente. Infatti le leggi della Fisica, che tradizionalmente sono presentate agli studenti come "verità" indiscutibili, nel testo emergono solo dopo la elaborazione e la discussione dei dati sperimentali. Ad esempio, nella caduta libera di un grave, nel primo esperimento sul periodo del pendolo semplice, nell'andamento della resistenza di un semiconduttore con la temperatura si richiede di formulare le relative leggi empiriche che successivamente, si constaterà, possono essere descritte, entro le incertezze, attraverso i principi e i modelli che la Fisica elabora.*

iii) *È stato spiegato con maggiore dettaglio il metodo grafico per la determinazione di una relazione funzionale che descriva una serie di dati sperimentali. Ciò poiché per alcuni corsi di laurea,* *dove le conoscenze dell'analisi matematica sono limitate, il metodo grafico è l'unica maniera possibile per spiegare l'adattamento di una relazione funzionale ai dati sperimentali ("best fit").*

iv) *Sono stati introdotti sei nuovi esperimenti: caduta libera, bilancia idrostatica, misura di densità di liquidi, misura con l'ohmetro di una resistenza metallica e di un semiconduttore in funzione della temperatura, distanza focale e ingrandimento di una lente. Con l'aggiunta di questi esperimenti, la seconda edizione offre una varietà completa di prove di laboratorio che può essere considerata esaustiva per tutti i corsi di laurea ad esclusione di quello di Fisica.*

v) *È stato dato maggiore risalto (Cap. 11) al significato della variabile "chi quadrato" e alle sue applicazioni; un nuovo paragrafo (12.11) è stato dedicato alle interpolazioni ed estrapolazioni di grandezze fisiche. Nel paragrafo 2.10 "Unità di misura fuori sistema" sono state illustrate ampiamente le definizioni di unità di massa atomica e di massa molecolare (grammomolecola) la cui conoscenza da parte di molti studenti è solitamente piuttosto confusa.*

vi) *Un nuovo capitolo è stato dedicato agli esercizi, svolti ed ampiamente discussi, parte dei quali assegnati alle prove di esame. I principali argomenti trattati sono: le equazioni dimensionali, le distribuzioni binomiale, di Poisson, di Gauss, il test del "chi quadrato", la regressione di curve e la formulazione di leggi.*

vii) *Infine, sono stati corretti gli errori della prima edizione segnalati dai miei studenti nel corso delle lezioni.*

L'autore desidera esprimere la sua gratitudine ai laureati in Scienza dei Materiali S. Abate e G. Desiderio per l'allestimento dei nuovi esperimenti.

GAETANO CANNELLI
Roma, Settembre 2003

Prefazione alla terza edizione

La terza edizione del libro lascia sostanzialmente invariate la I e II Parte, mentre la III Parte è stata arricchita con altre quattro Esperimentazioni.

Nel Capitolo 14. ***Dinamica del corpo rigido. Elasticità di scorrimento*** *sono state inserite due nuove Esperimentazioni: "Formulazione della legge del periodo del pendolo di torsione" ; "Determinazione del modulo elastico di scorrimento di alcuni metalli". Queste due Esperimentazioni sono strettamente legate poiché la costante elastica di torsione di un filo, determinata tramite il periodo del pendolo, è legata da una semplice relazione al modulo elastico di scorrimento del materiale di cui è fatto il filo. Tutte e due le Esperimentazioni sono introdotte da paragrafi che spiegano la teoria.*

Il Capitolo 20. ***Distribuzioni di probabilità*** *è stato rivisto introducendo due nuove Esperimentazioni: "Determinazione della legge di distribuzione di un campione di misure di una stessa grandezza" ; "Determinazione della legge di distribuzione delle misure di diametri di cilindretti metallici". Con questi due esperimentazioni si è voluto enfatizzare che, mentre per le misure ripetute di una stessa grandezza fisica ci si aspetta una distribuzione gaussiana, per le misure di una grandezza riferita a oggetti distinti, ma per quanto possibile eguali, la distribuzione non è necessariamente gaussiana.*

Infine, sono stati corretti errori ed imprecisioni della II edizione.

Un ringraziamento è dovuto alla Professoressa Rosa Bartucci del Dipartimento di Fisica dell'Università della Calabria per la sua disponibilità e la cura posta nel revisionare i nuovi paragrafi di questa III edizione.

GAETANO CANNELLI
Roma, Giugno 2010

Prefazione alla quarta edizione

Nella IV edizione del libro è stato revisionato il paragrafo 2.8 con la descrizione del nuovo ***Sistema Internazionale SI delle unità di misura*** *entrato in vigore il 20 maggio 2019, approvato dai rappresentanti di 62 Paesi riuniti a Versailles nella 26ª Conferenza generale dei pesi e delle misure* ***(CGPM)****, novembre 2018.*

Nel nuovo Sistema SI è avvenuta una vera e propria rivoluzione, le 7 unità di misura fondamentali, metro, kilogrammo, secondo, kelvin, ampere, candela, mole non sono più legate a campioni fisici o a proprietà della materia ma tutte definite in termini di costanti fisiche fondamentali: velocità della luce nel vuoto, costante di Planck, frequenza della transizione iperfine dell'atomo Cesio-133, costante di Boltzmann, carica elementare dell'elettrone, intensità luminosa di una radiazione monocromatica di una data frequenza, costante di Avogadro; rimangono invariate le definizioni del metro m, del secondo s e della candela cd. Le nuove definizioni hanno il vantaggio che i campioni di unità di misura ora sono sempre riproducibili, senza errori, in qualsiasi laboratorio.

L'autore desidera esprimere un ringraziamento al Prof. Alessandro Papa, docente di Fisica Teorica del Dipartimento di Fisica dell'Università della Calabria, per la collaborazione nella stesura delle definizioni delle unità di misura del nuovo ***Sistema Internazionale SI.***

Indice generale

Distribuzioni di probabilità di piccoli campioni: distribuzione di Student, distribuzione di Fisher, distribuzione della variabile χ^2. Test per la verifica di ipotesi

10

10.1 – La distribuzione della variabile *t* di Student

Quando un campione di N valori della variabile x, estratto da una popolazione distribuita normalmente, è piccolo, le stime del valore aspettato μ e della deviazione standard sono molto incerte e la distribuzione di Gauss non è più adeguata. In tal caso è più corretto adottare la distribuzione proposta da Gosset con il suo pseudonimo Student. Questa distribuzione nota come "distribuzione di Student" considera la variabile

$$t = \frac{\bar{x} - \mu}{\hat{s}/\sqrt{N}}, \tag{10.1}$$

dove $\bar{x}$ è la media campionaria e $\hat{s} = \sqrt{\dfrac{\sum_{i=1}^{N}(x_i - \bar{x})^2}{N-1}}$ è lo stimatore della deviazione standard tramite il campione; nel caso che nella definizione della variabile di Student si faccia intervenire la deviazione standard empirica $s = \sqrt{\dfrac{\sum_{i=1}^{N}(x_i - \bar{x})^2}{N}}$, la (10.1) assume la forma $t = \dfrac{\bar{x} - \mu}{s/\sqrt{N-1}}$ osservando che $\dfrac{\hat{s}}{s} = \sqrt{\dfrac{N}{N-1}}$.

La variabile t può essere intesa come la variabile standardizzata della media campionaria rispetto alla stima della deviazione standard della media. Gosset ha mostrato che la funzione densità di distribuzione della variabile t è

$$f(t) = C \cdot \left(1 + \frac{t^2}{\upsilon}\right)^{-(\upsilon+1)/2}, \tag{10.2}$$

dove la costante C è determinata dalla condizione di normalizzazione e $\nu = N - 1$, unico parametro della funzione di Student, è il numero di gradi di libertà.

La funzione di distribuzione (10.2) è simmetrica, con forma a campana, con valore aspettato, $t = 0$, ove presenta il massimo (Fig. 10.1); si può vedere che la varianza della variabile è $N/(N-2)$

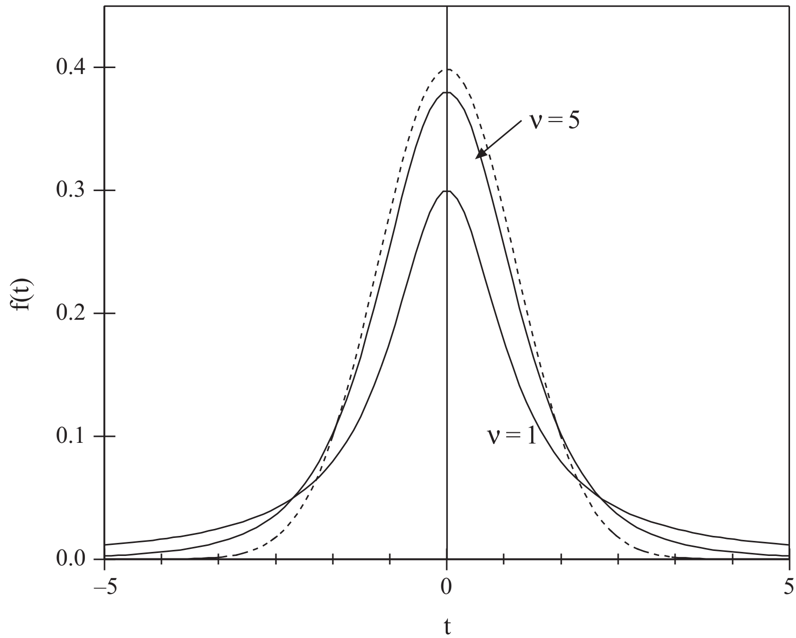

Fig. 10.1 – Funzione densità di probabilità di Student per due valori del grado di libertà $\nu = N - 1$. La curva tratteggiata rappresenta la funzione di Gauss.

se $N > 2$. All'aumentare del numero di gradi di libertà ν (o di N), la distribuzione di Student tende alla distribuzione normale standardizzata $f(t) = 1/\sqrt{2\pi} \cdot e^{-\frac{1}{2}t^2}$. Già per valori di $N \geq 30$ le due distribuzioni non sono più distinguibili, mentre per $N < 30$ la distribuzione $f(t)$ presenta due code che decadono più lentamente di quelle della distribuzione gaussiana. La distribuzione di Student, dunque, può essere applicata indifferentemente per piccoli e grandi campioni.

10.2 – Livelli di confidenza ed intervalli di fiducia nella distribuzione di Student

Vediamo ora alcuni livelli di confidenza e i corrispondenti intervalli di fiducia per la variabile della distribuzione di Student. Da questi possiamo stimare gli intervalli di fiducia nei quali presumiamo sia contenuto il valore aspettato μ della popolazione. Illustriamo ciò con un esempio.

Esempio 10.1

Un campione di $N = 10$ misure del valore di uno spessore, eseguite con un palmer, presenta un valore della media campionaria $\bar{x} = 4.745$ mm e una stima della deviazione standard $\hat{s} = 0.053$ mm. Trovare, al livello di confidenza del 95% e del 99%, i limiti dell'intervallo di fiducia nel quale presumiamo cada il valore vero dello spessore.

I gradi di libertà sono $\nu = N - 1 = 9$, la deviazione standard della media campionaria è $\sigma_{\bar{x}} = \hat{s}/\sqrt{N} = 0.053/\sqrt{10} = 0.0168\ mm$.

Un livello di confidenza del 95% significa probabilità del 95% di osservare un valore della variabile t tra i due limiti $-t_{0.975}$ e $+t_{0.975}$ dell'intervallo di fiducia, cioè $P(-t_{0.975} \leq t \leq +t_{0.975}) = 95\%$. Con tale notazione si vuole indicare che sopra $+t_{0.975}$ e sotto $-t_{0.975}$, la curva simmetrica di Student delimita due aree (code) eguali che rappresentano ciascuna la probabilità del 2.5% (Fig. 10.2).

Il valore di $t_{0.975}$ resta determinato dalle equazioni integrali

$$\int_{-t_{0.975}}^{+t_{0.975}} f(t) = 95\%, \qquad \int_{-\infty}^{-t_{0.975}} f(t) = \int_{+t_{0.975}}^{+\infty} f(t) = 2.5\%, \qquad \int_{-\infty}^{+t_{0.975}} f(t) = 97.5\% \tag{10.3}$$

che esprimono le aree racchiuse dalla curva di Student: tra $-t_{0.975}$ e $+t_{0.975}$; al di sotto di $-t_{0.975}$; al di sopra di $t_{0.975}$; tra $-\infty$ e $+t_{0.975}$, rispettivamente.

Nell'uso corrente, per un certo livello di confidenza, il valore del limite superiore dell'intervallo di fiducia, detto **valore critico $\mathbf{t_c}$**, è espresso nella forma dell'ultimo integrale di (10.3). Il suo valore, per alcuni livelli di confidenza al variare dei gradi di libertà ν, è riportato nella Tabella A II.

Nel nostro caso, per il livello di confidenza del 95%, con $\nu = N - 1 = 9$, $t_{0.975} = 2.26$. Quindi esiste il 95% di probabilità che la variabile t cada nell'intervallo

$$-2.26 \leq \frac{\bar{x} - \mu}{\hat{s}/\sqrt{N}} \leq +2.26 \ . \tag{10.4}$$

Pertanto confidiamo, con il 95% di probabilità, che il valore vero dello spessore sia compreso nell'intervallo

$$\bar{x} - 2.26\,\frac{\hat{s}}{\sqrt{N}} \leq \mu \leq \bar{x} + 2.26\,\frac{\hat{s}}{\sqrt{N}}, \tag{10.5}$$

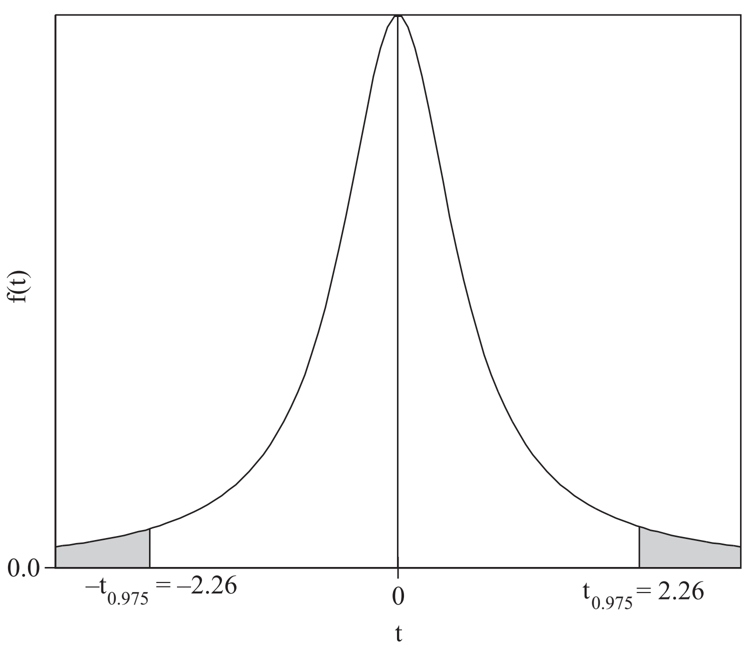

Fig. 10.2 – Intervallo di fiducia $(-t_{0.975} \leq t \leq t_{0.975})$ per la variabile t di Student al livello di confidenza del 95%, per il grado di libertà $\nu = 9$.

e sostituendo i valori:

$$4.707 \text{ mm} \leq \mu \leq 4.783 \text{ mm}.$$

Per il livello di confidenza del 99%, i limiti dell'intervallo di fiducia, con analoghe considerazioni, sono indicati come $-t_{0.995}$, $+t_{0.995}$. Per $\nu = 9$, dalla Tabella A II il valore critico risulta $t_{0.995} = 3.25$.

In tal caso possiamo essere confidenti al 99% che il valore vero cada nell'intervallo di fiducia

$$4.690 \text{ mm} \leq \mu \leq 4.800 \text{ mm}.$$

Nel caso si fosse usata la distribuzione di Gauss, la variabile standardizzata z, al livello di confidenza del 95%, avrebbe assunto i valori $-z_{0.975} = -1.96$, $z_{0.975} = 1.96$. Il corrispondente intervallo di fiducia sarebbe risultato di minore ampiezza con i seguenti valori dei limiti per il valore dello spessore

$$4.745 - 1.96 \times 0.053/\sqrt{10} = 4.712\ mm \qquad e \qquad 4.745 + 1.96 \times 0.053/\sqrt{10} = 4.778\ mm.$$

10.3 – Test di ipotesi mediante la distribuzione di Student. Alcuni esempi di applicazione anche in campo medico e farmaceutico

Nei capitoli precedenti abbiamo visto come descrivere una popolazione di variabili aleatorie (che non necessariamente debbano essere grandezze fisiche) mediante alcuni parametri, la media (valore aspettato), la deviazione standard etc. e come questi parametri possano essere stimati da campioni tratti dalla popolazione. Di seguito faremo uso di questi parametri per test di verifica di ipotesi (test di ipotesi o di significatività) che possono riguardare non solo la Fisica, ma anche la Bio-Fisica e altre discipline. Ad esempio ci si potrebbe chiedere se due stime del valore di una grandezza fisica, ottenute con diverse procedure siano consistenti cioè appartengano alla stessa popolazione (Esempio 10.3), oppure se la somministrazione di un certo farmaco abbia un effetto su un qualche parametro clinico di una popolazione di individui (Esempi 10.4, 10.5, 10.6). La procedura del test è basata sull'assunzione, detta **ipotesi nulla o ipotesi $\mathbf{H_0}$**, che non ci sia differenza tra i parametri (medie campionarie) di due campioni a livelli di significatività del 5 o dell'1%. La validità dei test richiede che i campioni analizzati siano indipendenti e tratti a caso da popolazioni distribuite normalmente; poiché tali popolazioni possono essere descritte esaurientemente mediante i due parametri media e deviazione standard, le procedure sono definite "metodi statistici parametrici".

Vediamo ora alcuni tipici test di ipotesi nei quali viene impiegata la distribuzione di Student.

10.3.1 – *Consistenza della media campionaria di un piccolo campione con il valore aspettato*

Può capitare di dover stabilire se il valore di una grandezza fisica, determinato tramite un piccolo campione di misure, sia compatibile, ad un certo livello di significatività (5% o 1%), con un valore noto a priori. Illustriamo ciò con un esempio.

Esempio 10.2

Il valore aspettato della durata media di scarica, nelle stesse condizioni d'uso, di pile prodotte per diversi anni da una grossa industria è $\mu = 120$ min. Un campione di $N = 10$ pile tratto dalla produzione di una nuova filiale dell'industria mostra una durata media più piccola, $\bar{x} = 111$ min, con una stima della deviazione standard $\hat{s} = 13$ min.

Si vuole stabilire, al livello significativo del 5%, se il valore della durata media del prodotto della filiale sia consistente con quello della casa madre.

Nel nostro caso il valore osservato della variabile di Student risulta

$$t_0 = \frac{\bar{x} - \mu}{\hat{s}/\sqrt{N}} = \frac{111 - 120}{13/\sqrt{10}} = -2.19.$$

Dobbiamo valutare la probabilità $P(-\infty \le t \le -2.19) = \int_{-\infty}^{-2.19} f(t)\, d\, t$: se questa risulta superiore al 5% accettiamo l'ipotesi H_0 cioè il valore della durata media osservato è compatibile con il valore aspettato poiché la differenza è solo dovuta alle fluttuazione statistiche del campionamento casuale; se la probabilità è inferiore al 5% rigettiamo l'ipotesi. In questo caso eseguiamo il test ad una coda comparando la probabilità $P(-\infty \le t \le -2.19)$ con quella corrispondente all'intervallo tra $-\infty$ e $-t_{0.95}$ per il quale l'area racchiusa dalla curva di Student vale il 5% (vedi Esempio 10.1)

$$\int_{-\infty}^{-t_{0.95}} f(t)\, d\, t = 5\% \tag{10.6}$$

L'equazione integrale (10.6) determina il limite di integrazione $-t_{0.95}$, che è il 5° percentile (paragrafo 9.2) ed è tabulato nella Tabella AII: per $\nu = N - 1 = 9$ gradi di libertà, $-t_{0.95} = -1.83$ (si tenga conto della simmetria della curva $f(t)$). Poiché $t_0 = -2.19 < -t_{0.95} = -1.83$, la coda di probabilità $P(-\infty \le t \le -2.19)$ è più piccola del 5%, dunque rigettiamo l'ipotesi H_0. Il valore della durata di scarica tratto dal campione non è consistente con il valore aspettato poiché la differenza non è spiegabile con le fluttuazioni statistiche. Il prodotto della filiale appare di qualità inferiore a quello della casa madre; però se si considerasse il livello di significatività dell'1%, il percentile sarebbe $-t_{0.99} = -2.82$ e il valore osservato potrebbe essere ritenuto consistente.

10.3.2 – *Consistenza di due valori medi*

Può capitare di disporre di due piccoli campioni di misure della stessa grandezza che diano luogo a due diverse medie campionarie. Si vuole stabilire se la diversità delle medie campionarie sia unicamente dovuta a fluttuazioni statistiche; in tal caso esse appartengono a popolazioni aventi lo stesso valore aspettato.

Indichiamo con N_1 e N_2 due piccoli campioni aventi medie campionarie $\bar{x}_1$ e $\bar{x}_2$ e rispettive stime, s_1 e s_2, delle deviazioni standard. Siano inoltre σ_1 e σ_2 le deviazioni standard delle popolazioni da cui sono stati tratti i due campioni.

Se i campioni appartengono a popolazioni distribuite normalmente aventi lo stesso valore aspettato, la variabile $\bar{x}_1 - \bar{x}_2$ deve avere come valore aspettato

$$E[\bar{x}_1 - \bar{x}_2] = 0 \quad , \tag{10.7}$$

e varianza

$$\sigma^2_{\bar{x}_1 - \bar{x}_2} = \sigma^2_{\bar{x}_1} + \sigma^2_{\bar{x}_2} = \frac{\sigma_1^2}{N_1} + \frac{\sigma_2^2}{N_2} \ . \tag{10.8}$$

Nel caso si possa ritenere che i piccoli campioni provengano da popolazioni con la stessa varianza, $\sigma_1 = \sigma_2 = \sigma$, possiamo dare una stima di tale varianza mediante l'espressione

$$\sigma^2 \cong s^2 = \frac{\sum_i (x_{1i} - \bar{x}_1)^2 + \sum_i (x_{2i} - \bar{x}_2)^2}{(N_1 - 1) + (N_2 - 1)} = \frac{(N_1 - 1)\hat{s}_1^2 + (N_2 - 1)\hat{s}_2^2}{N_1 + N_2 - 2} \ , \tag{10.9}$$

dove l'ultima quantità può essere vista come la media delle stime delle varianze pesate con i gradi di libertà dei due campioni. La (10.8) può allora essere scritta come

$$\sigma^2_{\bar{x}_1 - \bar{x}_2} = \frac{s^2}{N_1} + \frac{s^2}{N_2} \ . \tag{10.10}$$

Per valutare la consistenza delle due medie campionarie dobbiamo considerare la variabile

$$t = \frac{(\bar{x}_1 - \bar{x}_2) - 0}{\sigma_{\bar{x}_1 - \bar{x}_2}} = \frac{\bar{x}_1 - \bar{x}_2}{\sqrt{\dfrac{s^2}{N_1} + \dfrac{s^2}{N_2}}} \tag{10.11}$$

la cui distribuzione è data dalla distribuzione di Student con un numero di gradi di libertà $\nu = N_1 + N_2 - 2$.

Esempio 10.3

Due piccoli campioni di misure di una grandezza, eseguiti con due diversi strumenti, presentano le seguenti medie campionarie e rispettive stime delle deviazioni standard

$$N_1 = 8 \qquad \bar{x}_1 = 4.81 \ ua \qquad \hat{s}_1 = 0.40 \ ua$$

$$N_2 = 10 \qquad \bar{x}_2 = 5.07 \ ua \qquad \hat{s}_2 = 0.34 \ ua.$$

Si stabilisca, al livello significativo del 5%, se i due valori osservati siano consistenti.

Calcoliamo la stima della deviazione standard comune

$$s = \sqrt{\frac{(N_1 - 1)\hat{s}_1^2 + (N_2 - 1)\hat{s}_2^2}{N_1 + N_2 - 2}} = \sqrt{\frac{7 \times 0.40^2 + 9 \times 0.34^2}{18 - 2}} = 0.3674\ldots \cong 0.37 \ ua.$$

I valori della variabile di Student (10.11) da considerare sono due corrispondenti alle due possibili differenze, $\bar{x}_1 - \bar{x}_2$ e $\bar{x}_2 - \bar{x}_1$,

$$+t_0 = \frac{5.07 - 4.81}{0.37\sqrt{\dfrac{1}{8} + \dfrac{1}{10}}} = 1.48 \qquad , \qquad -t_0 = \frac{4.81 - 5.07}{0.37\sqrt{\dfrac{1}{8} + \dfrac{1}{10}}} = -1.48;$$

ora il test va fatto sulle due code della distribuzione. Al livello significativo del 5% con $\nu = N_1 + N_2 - 2 = 16$ (Tabella A II), il percentile che separa il 97.5% inferiore di probabilità dal 2.5% superiore (coda superiore) è $t_{0.975} = 2.12$ e il percentile che separa il 2.5% inferiore dal 97.5% superiore (coda inferiore) è $-t_{0.975} = -2.12$. Poiché i valori osservati della variabile $-t_o$ e $+t_o$ sono entro l'intervallo da –2.12 a 2.12, essi determinano una probabilità superiore al 5%. La differenza tra i valori delle due medie non è significativa ma dovuta unicamente a fluttuazioni statistiche, dunque i valori osservati sono consistenti.

10.3.3 – *Alcuni esempi di applicazioni in campo medico e farmaceutico*

Vediamo ora come il test di ipotesi di Student trovi applicazione per la verifica dell'efficacia di farmaci.

Esempio 10.4 Test di Student su un farmaco adottato nella riduzione del tasso di colesterolo totale

Un medico vuole verificare, in una comunità (popolazione) di 300 suoi pazienti, l'efficacia di un farmaco (statina) largamente impiegato per la riduzione del tasso di colesterolo totale. A tale scopo sceglie a caso nella comunità, due gruppi di N = 15 individui adulti di età superiore a 60 anni nelle stesse condizioni cliniche. Ad un gruppo di controllo somministra un placebo e all'altro gruppo il farmaco. L'elaborazione dei dati grezzi (che non mostriamo) dà luogo ai seguenti valori delle medie campionarie e delle stime delle deviazioni standard

gruppo trattato con il placebo $\quad c_{pla} = 263 \text{ mg/100 ml} \;,\; s_{pla} = 23 \text{ mg/100 ml} \;;$

gruppo trattato con il farmaco $\quad c_{far} = 236 \text{ mg/100 ml} \;,\; s_{far} = 29 \text{ mg/100 ml} \;;$

abbiamo omesso, per economia di scrittura, il simbolo di stimatore $\hat{s}$ nelle stime delle deviazioni standard.

Il medico vuole stabilire se la differenza osservata tra le medie campionarie della concentrazione di colesterolo totale possa essere attribuita unicamente alle fluttuazioni statistiche del campionamento casuale o all'efficacia del farmaco.

Facciamo l'ipotesi che il farmaco non abbia effetto sulla concentrazione del colesterolo totale (ipotesi nulla o ipotesi H_0) e che la differenza tra le medie campionarie sia dovuta alle fluttuazioni tra i possibili campioni tratti dalla popolazione. Si osservi che il numero delle diverse coppie di campioni che possono essere tratti dalla popolazione è elevatissimo. Nella comunità costituita da 300 individui il numero delle coppie di possibili campioni è ${}_{300}C_{30} \cong 10^{41}$! I due gruppi, nell'ipotesi H_0, si possono considerare come campioni della stessa popolazione. Le due deviazioni standard empiriche sono entrambe stime della deviazione standard σ della popolazione e poiché le dimensioni dei due campioni sono eguali, la stima della varianza σ^2 data dalla (10.9) è la media delle due stime

$$\sigma^2 \;\cong\; s^2 \;=\; \frac{1}{2}\,(s^2_{pla} \;+\; s^2_{far}) \;=\; \frac{1}{2}\,(23^2 \;+\; 29^2) \;=\; 685\ (\text{mg/100 ml})^2 \;.$$

Il valore delle variabile di Student è pertanto

$$t_o \;=\; \frac{\overline{c}_{pla} - \overline{c}_{far}}{\sqrt{\dfrac{s^2}{N} + \dfrac{s^2}{N}}} \;=\; \frac{263 \;-\; 236}{\sqrt{2\,\dfrac{685}{15}}} \;=\; \frac{27}{9.55} \;=\; 2.83 \;.$$

Dobbiamo considerare anche la possibilità di un valore negativo della differenza delle medie. Infatti per l'ipotesi nulla, le due medie campionarie, derivate da campioni scelti a caso dalla stessa popolazione, possono dar luogo, fluttuando, a valori positivi o negativi della loro differenza. Questo comporta l'esecuzione del test a due code che, a parità di livello di significatività, fa aumentare il valore critico della variabile con conseguente maggiore affidabilità del test.

Dalla Tabella A II, per un numero di gradi di libertà $\nu = 2(N - 1) = 28$, osserviamo che al livello del 5% (o al livello dell'1%), i valori della variabile (percentili) che separano le due code inferiore e superiore sono

$-t_{0.975} = -2.05$ e $t_{0.975} = 2.05$ ($-t_{0.995} = -2.76$ e $t_{0.995} = 2.76$, al livello dell'1%). Il valore da noi osservato, $t_o = 2.83$, è fuori da ambedue gli intervalli e quindi ad esso è associata una probabilità inferiore all'1%. Questa probabilità è troppo bassa, dunque l'ipotesi che le due medie campionarie appartengano alla stessa popolazione (ipotesi nulla) è da rifiutare anche al livello significativo dell'1%. Pertanto si può ritenere che il farmaco sia efficace nell'abbassare in media il tasso di colesterolo nei pazienti che lo hanno assunto.

Esempio 10.5 Test di Student per la verifica dell'efficacia di un diuretico

Un medico vuole verificare l'efficacia di un diuretico, somministrato ad un certo dosaggio, sulla diuresi di una comunità numerosa (popolazione) di suoi pazienti. A tale scopo sceglie a caso nella comunità, due gruppi di $N = 14$ individui adulti di età superiore ai 60 anni nelle stesse condizioni cliniche. Ad un gruppo di controllo viene somministrato un placebo e all'altro gruppo il diuretico. La diuresi viene misurata rilevando il volume di urina prodotta da un individuo nelle 24 ore; questa variabile viene espressa in l/giorno (litri/giorno). L'elaborazione (che non riportiamo) dei dati grezzi (volumi giornalieri di urina di ciascun individuo) mostra i seguenti valori delle medie campionarie e delle stime delle deviazioni standard

gruppo trattato con il placebo $\quad \nu_{pla} = 1.65$ l/giorno , $s_{pla} = 0.21$ l/giorno ;

gruppo trattato con il farmaco $\quad \nu_{diu} = 1.73$ l/giorno , $s_{diu} = 0.19$ l/giorno ;

abbiamo omesso, per economia di scrittura, il simbolo di stimatore $\hat{s}$ nelle stime delle deviazioni standard.

Il medico vuole stabilire se la differenza osservata tra le medie campionarie della quantità di urina giornaliera possa essere attribuita unicamente alle fluttuazioni statistiche o se invece è dovuta all'efficacia del farmaco.

Si osservi che il numero delle diverse coppie di campioni che possono essere tratti dalla popolazione è elevatissimo. Se la comunità fosse costituita da 300 individui il numero delle coppie di campioni possibili sarebbe ${}_{300}C_{28} \cong 10^{39}$!

Fatta l'ipotesi che il diuretico non abbia effetto sulla diuresi (ipotesi H_0 o ipotesi nulla), i due gruppi si possono allora considerare come campioni estratti a caso dalla stessa popolazione. Le due deviazioni standard sono entrambe stime della deviazione standard σ della popolazione e poiché le dimensioni dei due campioni sono eguali, la stima della varianza σ^2 può essere ottenuta dalla media delle stime delle due varianze

$$\sigma^2 \cong s^2 = \frac{1}{2}(s^2_{pla} + s^2_{far}) = \frac{1}{2}(0.21^2 + 0.19^2) = 0.040\ (\text{l/giorno})^2 \ .$$

Il valore delle variabile di Student osservato è

$$t_o = \frac{\overline{\nu}_{pla} - \overline{\nu}_{diu}}{\sqrt{\frac{s^2}{N} + \frac{s^2}{N}}} = \frac{1.65 - 1.73}{\sqrt{2\,\frac{0.040}{14}}} = \frac{0.080}{0.076} = -1.06 \ .$$

Dobbiamo considerare anche la possibilità di un valore positivo della differenza delle medie. Infatti per l'ipotesi H_0 fatta, le due medie campionarie, derivate da campioni scelti a caso dalla stessa popolazione, possono dar luogo, fluttuando, a valori positivi o negativi della loro differenza. Questo comporta l'esecuzione del test a due code che, a parità di livello di significatività, fa aumentare il valore critico della variabile con conseguente maggiore affidabilità del test.

Dalla Tabella A II, per un numero di gradi di libertà $\nu = 2(N-1) = 26$, osserviamo che al livello del 5% (o al livello dell'1%), i valori della variabile (percentili) che separano le due code inferiore e superiore sono $-t_{0.975} = -2.06$ e $t_{0.975} = 2.06$ ($-t_{0.995} = -2.78$ e $t_{0.995} = 2.78$, al livello dell'1%). Il valore osservato della variabile,

$t_o = 1.06$, è ben dentro questi intervalli, quindi ad esso è associata una probabilità che è molto superiore al 5%. L'ipotesi che le due medie campionarie appartengano alla stessa popolazione (ipotesi nulla) non è da rigettare: la differenza tra i due valori di diuresi osservati non è significativa e pertanto si può ritenere che il farmaco assunto a quel dosaggio sia inefficace.

10.3.4 – *Test di Student per dati appaiati*

Quando si vuole esaminare l'efficacia di un farmaco su un unico campione di individui si prende in considerazione la variazione di un parametro clinico prima e dopo la somministrazione del farmaco. La variabile aleatoria è rappresentata in questo caso dalla variazione del parametro in ciascun individuo. Ora è più facile discernere l'eventuale efficacia del farmaco poiché non si verifica la maggiore variabilità esistente tra individui appartenenti a due campioni indipendenti: uno trattato con il farmaco e l'altro con il placebo.

Il test t che si esegue è detto **test per dati appaiati** e si basa sempre sull'ipotesi che mediamente il farmaco non alteri il parametro clinico in osservazione negli individui del campione (ipotesi nulla).

Per determinare la variabile di Student calcoliamo la media campionaria delle variazioni d_i del valore del parametro clinico osservate su ciascun individuo del campione e la stima della deviazione standard

$$\bar{d} = \frac{1}{N}\sum_{i=1}^{N} d_i \quad ; \quad s_d = \sqrt{\frac{\sum_{i=1}^{N}(d_i - \bar{d})^2}{N-1}} \; .$$

La variabile di Student assume ora l'espressione

$$t = \frac{\bar{d} - \mu_d}{s_{\bar{d}}} = \frac{\bar{d}}{s_{\bar{d}}} \; , \tag{10.12}$$

dove $s_{\bar{d}} = \dfrac{s_d}{\sqrt{N}}$ è la deviazione standard della media e il valore aspettato μ_d della variabile d è zero nell'ipotesi che il farmaco non abbia alcun effetto sui pazienti.

Si osservi che la stima della deviazione standard s_d è stata valutata direttamente dalle differenze d_i del parametro clinico rilevate sui singoli individui. Se i dati non fossero stati appaiati, la deviazione standard della differenza sarebbe stata calcolata tramite la somma in quadratura delle varianze (10.8) nel campione trattato con placebo e nel campione trattato con il farmaco. Avremmo ottenuto così un valore della deviazione standard più grande di s_d giustificato dalla circostanza che i dati, provenienti da due campioni indipendenti, mostrerebbero una maggiore variabilità.

Esempio 10.6 Test di Student per la verifica dell'efficacia di un antipiretico su un unico campione di pazienti

Un farmaco antipiretico viene somministrato ad un campione di $N = 10$ bambini influenzati. Le temperature di ciascun paziente prima e dopo due ore dalla somministrazione del farmaco sono riportate nella Tabella 10.1. Si vuole accertare se l'antipiretico produca una significativa riduzione della temperatura corporea.

Tabella 10.1 – *Temperature corporee prima e dopo la somministrazione di un antipiretico*

Paziente	T_p (°C) **prima**	T_d (°C) **dopo 2 ore**	$d = T_p - T_d$ (°C)
1	38.8	38.3	0.5
2	38.6	38.0	0.6
3	38.4	38.7	–0.3
4	37.9	37.2	0.7
5	38.7	37.9	0.8
6	39.2	38.3	0.9
7	39.3	38.3	1.0
8	39.7	38.8	0.9
9	38.3	37.1	1.2
10	38.5	37.2	1.3

La variabile aleatoria da considerare è la differenza d, in ciascun paziente, tra le temperatura T_p prima e la temperatura T_d dopo l'assunzione del farmaco.

Determiniamo la media campionaria della variabile d e la stima della deviazione standard

$$\bar{d} = \frac{1}{N}\sum_{i=1}^{N} d_i = 0.76°\mathrm{C} \quad ; \quad s_d = \sqrt{\frac{\sum_{i=1}^{N}(d_i - \bar{d})^2}{N-1}} = \sqrt{\frac{1.804}{9}} = 0.45°\mathrm{C} \ .$$

L'errore standard della media è

$$s_{\bar{d}} = \frac{s_d}{\sqrt{N}} = \frac{0.45}{\sqrt{10}} = 0.14°\mathrm{C} \ ,$$

da cui otteniamo per la variabile di Student

$$t_o = \frac{\bar{d}}{s_{\bar{d}}} = \frac{0.76}{0.14} = 5.43 \ .$$

In questo caso consideriamo il test ad una coda poiché vogliamo considerare i valori della variabile d positivi. Dalla Tabella A II, per $\nu = N - 1 = 9$ gradi di libertà, osserviamo che: il valore critico della variabile di Student al livello significativo del 5% (ovvero il percentile che separa il 95% di probabilità inferiore dal restante 5% superiore) è $t_{0.95} = 1.83$; il valore critico al livello significativo dell'1% è $t_{0.99} = 2.82$. Il valore osservato della variabile, $t_o = 5.43$, molto più grande dei valori critici, presenta una probabilità di verificarsi nettamente inferiore all'1%; dunque l'ipotesi nulla che mediamente il farmaco non abbia effetto sulla temperatura degli individui del campione va rigettata. Possiamo ritenere pertanto che l'antipiretico sia efficace nell'abbassare la temperatura corporea.

10.4 – Analisi della varianza nel confronto di diversi campioni. Distribuzione di Fisher. Test di ipotesi mediante la distribuzione di Fisher

Nel paragrafo10.3 abbiamo usato la distribuzione di Student per test di ipotesi mediante il confronto di due campioni di dati. Ora consideriamo una nuova **variabile F** che considera il confronto tra le varianze associate ad un numero qualsiasi di gruppi (campioni). Il metodo di analisi statistica che impiega questa variabile è denominato **analisi della varianza** o **test di Fisher**. Un caso particolare di questa distribuzione è quello di due soli campioni con stessa numerosità per il quale si mostra, nel paragrafo 10.5, che $F = t^2$ ove t è la variabile di Student. Anche il test di Fisher è basato sull'**ipotesi nulla**: trattamenti diversi applicati a diversi campioni non ne alterano i parametri a livelli significativi del 5 o 1%, ciò è come dire che tutti i campioni, anche se hanno subito trattamenti diversi, appartengono alla stessa popolazione. Questo tipo di analisi nel caso si debba rifiutare l'ipotesi nulla non indica quale o quali campioni si differenzino dagli altri. La validità del test richiede che i campioni analizzati siano indipendenti e tratti a caso da popolazioni distribuite normalmente. Poiché tali popolazioni possono essere descritte esaurientemente mediante la media e la deviazione standard, le procedure adottate sono definite "metodi statistici parametrici".

Per introdurre la variabile F facciamo riferimento ad un esempio.

Esempio 10.7 Effetto di diverse diete sul tasso di colesterolo HDL di individui

In una vasta comunità (popolazione) di adulti sani si sceglie in modo casuale $n_g = 4$ gruppi (campioni) di $N = 20$ individui nutriti per 2 mesi con: 1) dieta mediterranea che privilegia alimenti vegetali; 2) dieta di solo pesce; 3) dieta con abbondanza di carne; 4) dieta varia. Dopo i due mesi di dieta si misura il valore della concentrazione di colesterolo HDL di ciascun individuo.

La medie campionarie della concentrazione del colesterolo HDL (in milligrammi/decilitro, mg/dl) e le rispettive stime della varianza, ottenute dai dati grezzi per ciascun gruppo, risultano essere

1) *dieta mediterranea*	$\bar{c}_1 = 58.3$ mg/dl	$s_1^2 = 17.3^2 = 299.3$ (mg/dl)^2 ;
2) *dieta di pesce*	$\bar{c}_2 = 64.8$ mg/dl	$s_1^2 = 14.4^2 = 207.4$ (mg/dl)^2 ;
3) *dieta di carne*	$\bar{c}_3 = 45.2$ mg/dl	$s_1^2 = 14.1^2 = 198.8$ (mg/dl)^2 ;
4) *dieta varia*	$\bar{c}_4 = 40.8$ mg/dl	$s_1^2 = 15.0^2 = 225.0$ (mg/dl)^2 .

Si vuole stabilire se le diversità delle concentrazioni del colesterolo HDL nei 4 gruppi possa essere attribuita alla diversità delle diete oppure se essa sia semplicemente connessa al fatto che il tasso di colesterolo varia in modo casuale tra individui diversi.

Partiamo dall'ipotesi che tutte le diete siano equivalenti (ipotesi nulla); nessuna dieta influenza più delle altre il tasso di colesterolo HDL di un adulto sano. Ciò significa che i quattro gruppi appartengono alla stessa popolazione, quindi la stima delle varianza in ciascun campione è una stima della varianza dell'intera popolazione. Naturalmente ogni campione tratto a caso dalla popolazione presenta diversa media campionaria e diversa stima della deviazione standard, ma se l'ipotesi dell'equivalenza delle diete è vera, i loro valori dovrebbero essere comparabili e le differenze osservate sono unicamente dovute alle fluttuazioni statistiche del campionamento. Si osservi che il numero delle possibili quaterne di campioni che possono essere tratti dalla popolazione è elevatissimo: se la comunità fosse costituita da 300 individui il numero sarebbe ${}_{300}C_{80}$.

Determiniamo la varianza della popolazione mediante due diverse stime dette: **varianza entro i gruppi** e **varianza tra i gruppi**.

La **varianza entro i gruppi** viene calcolata eseguendo la media delle 4 stime della varianza

$$s^2_{entro} = \frac{1}{4}\left(s_1^2 + s_2^2 + s_3^2 + s_4^2\right) \ ; \tag{10.13}$$

la denominazione "entro" mette in evidenza che queste stime sono ottenute all'interno di ogni gruppo.

Per ottenere la **"varianza tra i gruppi"** stimiamo prima la varianza delle medie campionarie dei gruppi

$$s_{\bar{c}}^2 = \frac{1}{n_g - 1}\left[(\bar{c}_1 - \bar{\bar{c}})^2 + (\bar{c}_2 - \bar{\bar{c}})^2 + (\bar{c}_3 - \bar{\bar{c}})^2 + (\bar{c}_4 - \bar{\bar{c}})^2\right] \ , \tag{10.14}$$

Gaetano Cannelli

Metodologie sperimentali in Fisica

Introduzione al metodo scientifico

Accedi ai **contenuti digitali** › **Espandi** le tue risorse › con un libro che **non pesa** e si **adatta** alle dimensioni del tuo **lettore**

All'interno del volume il **codice personale** e le istruzioni per accedere ai **contenuti digitali**.
L'accesso alle risorse digitali è **gratuito** ma limitato a **18 mesi dalla attivazione del servizio**.

EdiSES università

ISBN 978-88-3623-202-4

9 788836 232024