Le basi del metodo sperimentale
– un'introduzione pratica –

G. D'Agostini
Dipartimento di Fisica, Università "La Sapienza", Roma

1 febbraio 2001

Indice

I Introduzione alla metodologia di laboratorio **1**

1 Laboratorio virtuale **3**
1.1 Contatore 3
1.1.1 Conteggi a intervalli di tempo fissati 4
1.1.2 Tempi di attesa per ottenere un numero prestabilito di conteggi 7
1.2 Pallinometro 10
1.2.1 Previsioni 11
1.2.2 Risultati 13
1.3 ↻ Proprietà chimico-fisiche delle acque minerali 13
1.4 ↻ Nascono più femmine che maschi? 13
1.5 ↻ Coincidenze di compleanno 15
1.6 ↻ Numeri ritardatari al lotto 17
1.7 Nota semantica 17

2 Primo sguardo ai dati di laboratorio **19**
2.1 Misura foglio A4 19
2.2 Capacità di interpolazione fra le tacche e incertezza di lettura . 21
2.3 Errore ed incertezza di misura (discussione introduttiva) 23
2.4 ↻ Tempo di reazione e misure di cronometraggio 25
2.5 ↻ Moto uniformemente accelerato 26
2.6 ↻ Allungamento e periodo di oscillazione di una molla 27
2.6.1 Breve richiamo di fisica generale 28
2.6.2 Misure 28
2.6.3 Prime valutazioni di k e di g 29
2.7 ❄ Potere di aspirazione di una pompa da vuoto 31
2.8 Continua 33

3 Logbook e relazione **35**
3.1 Documentazione del lavoro sperimentale 35
3.2 Redazione del quaderno di laboratorio 36
3.3 Stesura della relazione 40
3.4 Cifre significative 41
3.4.1 Dai valori letti ai risultati delle misure 42
3.4.2 Cifre decimali e cifre significative 43
3.4.3 Regole pratiche (da prendere “cum grano salis”) . . . 44
3.4.4 Suggerimenti 46
3.5 Arrotondamenti 46

3.6 Controllo dimensionale e degli ordini di grandezza 47
3.7 Problemi 48

4 Descrizione grafica dei dati sperimentali 49
4.1 Riduzione dei dati 49
4.2 Tabelle, istogrammi e diagrammi a barre 52
4.3 ❄ Box plot 55
4.4 ○ Istogrammi bidimensionali: scatter plot e lego plot 58
4.4.1 Esempi di scatter plot 60

5 Descrizione quantitativa dei dati sperimentali 63
5.1 Statistica descrittiva e statistica inferenziale 63
5.2 Distribuzioni statistiche: notazioni 64
5.3 Misure di posizione 67
5.4 Misure di dispersione 68
5.4.1 Varianza e deviazione standard 70
5.5 Analogia meccanica di media e varianza 71
5.6 Proprietà di media e varianza 72
5.7 Valutazione pratica della deviazione standard 73
5.8 Effetto del raggruppamento in classi 74
5.9 Dispersione relativa e coefficiente di variazione 75
5.10 Misure di dispersione e incertezza della misura - caveat 76
5.11 ❄ Altre misure di forma 77
5.12 ❄ Misure di correlazione 79
5.13 ○ σ_N o σ_{N-1}? Commenti sul fattore correttivo $N/(N-1)$. 83
5.14 Nota sulle cifre significative da utilizzare nei problemi di statistica descrittiva 83
5.15 Problemi 85

6 Analisi grafiche 87
6.1 Studio di andamenti funzionali 87
6.2 Grafici 87
6.3 Grafici lineari: stima grafica dei parametri della retta 90
6.4 Cifre significative dei parametri della retta 93
6.4.1 Rilettura dei punti sperimentali e della retta 94
6.4.2 Scelta e tracciamento della retta 94
6.4.3 Altre incertezze nella stima dei parametri 94
6.4.4 Raccomandazioni 95
6.5 Linearizzazione 96
6.6 ○ Analisi grafica dell'esperienza della molla 96
6.6.1 Dipendenza dal modello 96
6.6.2 Combinazione dei risultati delle tre serie di misure . . 98
6.6.3 Valore di k condizionato dal valore noto di g 98
6.7 Uso di carte logaritmiche 98
6.7.1 Carta semilogaritmica 99
6.7.2 Proprietà delle carte logaritmiche 100
6.7.3 Stima dei parametri 102
6.7.4 Carta doppiologaritmica 105
6.8 ❄ Altre linearizzazioni notevoli 107

6.9 Problemi 109

II Considerazioni probabilistiche sulle esperienze simulate 111

7 Previsioni dei risultati 113
7.1 Introduzione 113
7.2 Pallinometro e distribuzione binomiale 113
7.2.1 Pallinometro “minimale”: calcolo della probabilità degli esiti 113
7.2.2 Pallinometro a molte file di chiodi 114
7.3 Contatore e processo di Poisson 117
7.3.1 Distribuzione di Poisson 117
7.3.2 ⥁ Distribuzione esponenziale 118
7.4 Limite a poissoniana della distribuzione binomiale 121
7.5 ❄ Contatore e distribuzione Gamma 123
7.6 Contatore e distribuzione geometrica 123
7.7 Numeri ritardatari al lotto 124
7.8 Previsioni basate sul teorema del limite centrale 125
7.8.1 Limite a normale della binomiale 125
7.8.2 Limite a normale della poissoniana 128
7.8.3 ❄ Limite a normale della distribuzione Gamma 128
7.8.4 Distribuzione della media aritmetica 128
7.8.5 Numero di teste meno numero di croci 129
7.9 ⥁ Cammino casuale (random walk) 130
7.10 ⥁ Ginnastica riepilogativa 131
7.11 ❄ Le distribuzioni osservate “erano” sempre molto poco probabili! 132
7.12 ❄ Simulazioni 134
7.13 Problemi 139

III Elementi di metrologia 141

8 Misure, strumenti ed errori di misura 143
8.1 Introduzione 143
8.2 Grandezze e unità di misura 143
8.3 Valore vero 147
8.4 Misure: concetti e definizioni 148
8.5 Risultati di misura, errori ed incertezze 149
8.6 Cause delle incertezze di misura 151
8.7 Errori casuali e sistematici 156
8.8 Precisione e accuratezza 156
8.9 Strumenti di misura 158
8.9.1 Introduzione 158
8.9.2 Strumenti a indicazione diretta 159
8.10 Caratteristiche degli strumenti 160
8.10.1 Campo di misura e condizioni di lavoro 160
8.10.2 Dipendenza della risposta dallo stimolo 161

8.10.3 Errori degli strumenti di misura 163
8.11 Correzione di errori sistematici 163
8.12 Esempi 164
8.12.1 Dipendenza delle caratteristiche del termometro a mercurio dai suoi parametri costruttivi 164
Sensibilità 164
Prontezza 164
Capacità del termometro confrontabile con quella del sistema 165
8.12.2 Sensibilità di una misura di capacità termica 165
8.12.3 Sensibilità di una misura di resistenza mediante ponte di Weathstone 165
8.13 Problemi 166

IV Applicazioni dell'inferenza statistica 167

9 Considerazioni generali sulla valutazione dell'incertezza di misura 169
9.1 Breve richiamo dei concetti di probabilità 169
9.2 Valutazione dell'incertezza di misura: schema generale 170
9.3 Imparare dagli esperimenti: il problema dell'induzione 172
9.4 Dalla probabilità degli effetti alla probabilità delle cause . . . 173
9.4.1 Verosimiglianza 173
9.4.2 Probabilità iniziale e probabilità finale 174
9.5 Paura dei "pregiudizi"? Inevitabilità di principio e frequente irrilevanza pratica delle prior 174
9.6 Scorciatoia al ragionamento bayesiano: il cane e il cacciatore . 175
9.7 Imparare dall'esperienza 176
9.8 ❄ Teorema di Bayes e probabilità delle ipotesi 177
9.8.1 Confronto fra due ipotesi 177
9.8.2 Classe continua di ipotesi 179

10 Misure dirette con verosimiglianza gaussiana 185
10.1 Risultati delle misure dirette in assenza di errori sistematici . . 185
10.2 Condizioni di ripetitibità 185
10.3 Singola osservazione con σ_r nota 186
10.4 n osservazioni indipendenti con σ_r nota 187
10.5 Caso di σ_r ignota 189
10.5.1 Misure ripetute della stessa grandezza fisica 189
10.5.2 Singole misure di grandezze fisiche variabili (grafici) . 190
10.5.3 Bisogna sempre ripetere le misure? Rarità delle situazioni in cui σ_r sia completamente ignota 191
10.6 ❄ Uso della t di Student 191
10.7 Presentazione del risultato - cifre significative 191
10.8 Misure di conteggio in approssimazione normale 194
10.8.1 Valutazione parametro della Poissoniana e dell'intensità di un processo di Poisson 194
10.8.2 Valutazione di p di una distribuzione binomiale 197
10.9 Combinazione di più risultati sullo stesso misurando 198

10.10 Problemi 201

11 Misure indirette ed errori sistematici 203

11.1 Propagazione delle incertezze 203

11.1.1 Caso di combinazioni lineari 203

11.1.2 Linearizzazione 204

11.1.3 Incertezze relative 206

11.2 Come tener conto degli errori sistematici 207

11.2.1 Condizioni di riproducibilità 208

11.2.2 Correzione dei risultati per tener conto di errori sistematici noti - calibrazioni 208

11.2.3 Incertezze dovute all'inesatta conoscenza dell'entità di un possibile errore sistematico 208

11.2.4 Imperfetta conoscenza delle costanti di calibrazioni e dei parametri di influenza 209

Errore di zero (offset) 209

Errore di scala 210

Importanza delle misure per differenza 211

11.2.5 Casi di errore di più difficile schematizzazione 213

11.2.6 Incertezza su un fattore di influenza 214

11.2.7 Propagazione senza derivate 215

11.2.8 Calibrazione, intercalibrazione e "randomizzazione" . 215

11.3 Coefficiente di correlazione 216

11.3.1 Valutazione pratica di ρ dovuto ad errori di calibrazione 216

11.4 Propagazione di varianze e covarianze 217

11.4.1 Formula generale per le incertezze relative 219

11.5 Casi notevoli di propagazione di incertezze 219

11.6 Formalismo della matrice di covarianza 220

11.7 Raccomandazioni BIPM/ISO 221

11.8 Valutazione delle incertezze di tipo B 223

11.9 Esempi numerici 225

11.10 Problemi 230

12 Fit 235

12.1 Inferenza sui parametri di una legge 235

12.2 ❄ Come tener conto anche di possibili incertezze sulle X . . . 237

12.3 Formule dei minimi quadrati 238

12.3.1 σ_Y nota e costante 238

12.3.2 σ_{Y_i} ignote e supposte costanti 239

12.3.3 σ_{Y_i} diverse e note a priori 240

12.4 Esempi di applicazione delle formule dei fit 240

12.4.1 Incertezze ignote e presupposte uguali 240

12.4.2 Incertezze note e diverse fra loro 242

12.5 Rette di calibrazione ed estrapolazione 242

12.6 Analisi grafica 244

12.6.1 Stima dei parametri 244

12.6.2 Stima dell'incertezza sui parametri ripetendo le misure 244

12.6.3 Stima dell'incertezza della singola misura dai residui . 246

12.6.4 Valutazione semplificata di σ_r 247

12.6.5 Barre di incertezza 247
12.6.6 Incertezza dei parametri mediante σ_r ricavata dai dati . 247
12.6.7 Analisi nel baricentro 248
12.7 Effetto degli errori sistematici 248
12.7.1 Errori sistematici dipendenti dal valore della grandezza 248
12.7.2 Errore di zero 249
12.7.3 Errore di scala 249
12.7.4 Deviazione dalla linearità 250
12.8 Esempio numerico di un'analisi grafica 250
12.9 Uso e abuso del computer 253
12.10 Problemi 255

V Soluzione dei problemi 257

VI 267

A Appendice critica 269
A.1 Valutazioni usuali delle incertezze 269
A.2 Critica della "teoria degli errori massimi" 270
A.2.1 $\Delta y = \sum_i \left|\frac{\partial y}{\partial x_i}\right| \Delta x_i$ 270
A.2.2 Regola della mezza divisione 274
A.2.3 $\Delta t = 0.2\,\mathrm{s}$? 277
A.2.4 Imperativo categorico di riportare le "barre di errore" . 277
A.2.5 Rette di massima e minima pendenza 277
A.3 Critica degli "errori statistici" 280
A.4 Riassumendo 281

Parte I

Introduzione alla metodologia di laboratorio

Capitolo 1

Laboratorio virtuale

In questo capitolo sono illustrate alcune esperienze basate su dati completamente simulati, o che comunque non richiedono strumentazione di laboratorio. Esse servono a presentare dati "sperimentali" che verranno trattati nel corso dei capitoli seguenti, applicando su di essi le tecniche di analisi che verranno via via presentate. Questi esperimenti, insieme a quelli più reali del prossimo capitolo, costituiscono il filo conduttore dell'intero corso e vanno ben al di là degli usuali esercizi numerici che seguono l'introduzione di nuove formule e metodi.

Inoltre le prime due esperienze possono essere utilizzate come introduzione 'maieutica' al concetto di probabilità, da approfondire in altra sede.

1.1 Contatore

Come prima esperienza simulata consideriamo un rivelatore di radioattività, ad esempio un contatore a scintillazione. Il principio di funzionamento è abbastanza semplice (Fig. 1.1): quando una "particella ionizzante" (non meglio definita per quello che ci interessa al momento) attraversa lo *scintillatore* viene emessa una debolissima luce dalle molecole del materiale. Essa, colpendo un opportuno dispositivo elettronico, chiamato fotomoltiplicatore, è in grado di generare un piccolo segnale elettrico di ampiezza proporzionale al numero dei fotoni prodotti. Questo segnale elettrico ha una durata inferiore al milionesimo di secondo e un'ampiezza che dipende dalla tensione applicata al fotomoltiplicatore. Se l'ampiezza supera un determinato *valore di soglia* il segnale produce un conteggio che viene registrato su un apposito visualizzatore digitale.

Il rivelatore descritto può essere quindi utilizzato, almeno in linea di principio, per stabilire il livello di *radioattività* ambientale, misurata in numero di conteggi al secondo. Per quanto detto sul rivelatore, è chiaro che c'è molta arbitrarietà in questa misura, in quanto il numero di conteggi registrato in un certo intervallo di tempo dipende dalla tensione applicata al fotomoltiplicatore e dalla soglia scelta. Inoltre, contando semplicemente il numero di segnali che superano una certa soglia, si perde l'informazione sulla ionizzazione prodotta da ciascuna delle particelle rivelate. Un altro fattore che complica l'interpretazione dei risultati è che non sempre un conteggio è causato da radioattività, ma può essere dovuto a *rumore* interno allo strumento. Quindi per passare

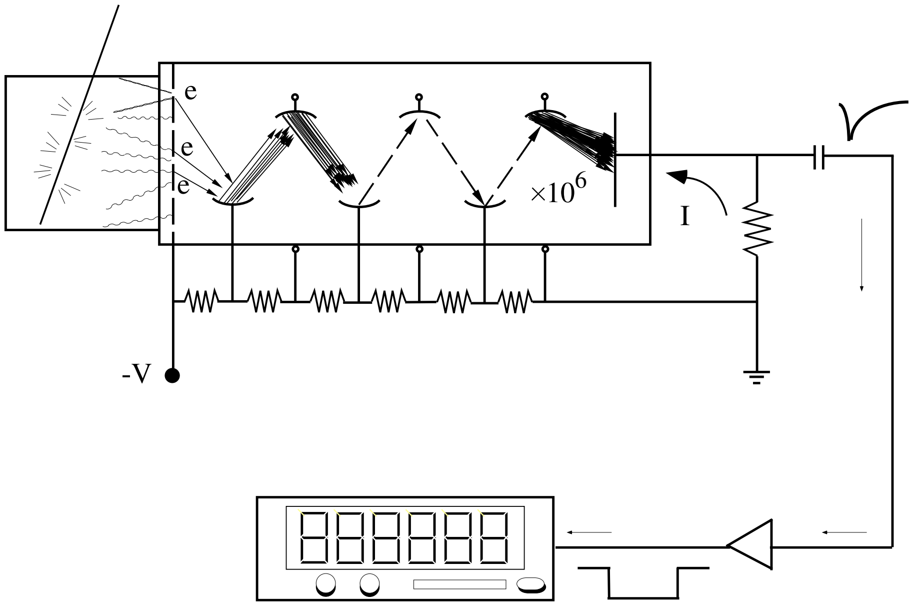

Figura 1.1: Schema del contatore a scintillazione.

dai numeri osservati ad un valore assoluto di radioattività che possa essere di qualche utilità ad un altro ricercatore occorre definire opportune *procedure di misura* e di *calibrazione*.

Per ora non ci preoccupiamo di questi importanti problemi e operiamo come se esistessimo solo noi e il nostro contatore. Assumiamo inoltre che il rumore sia trascurabile. Applichiamo la tensione di alimentazione al rivelatore, aspettiamo un po' che esso si stabilizzi e controlliamo che il contatore effettivamente funzioni, eventualmente agendo sul valore di soglia in modo tale da variare la frequenza di conteggio. Quindi regoliamo la soglia in modo tale da registrare soltanto particelle che producono un'alta ionizzazione, ottenendo così una frequenza di conteggio molto bassa (è il caso più interessante dal punto di vista statistico).

1.1.1 Conteggi a intervalli di tempo fissati

Lo strumento possiede un cronometro interno e un dispositivo per prefissare l'intervallo di tempo durante il quale i conteggi vengono accumulati nel visualizzatore. Trascorso questo tempo si legge il numero di conteggi. Poi si riazzera e si procede con una nuova misura.

Scegliamo un intervallo di **3 secondi** e cominciamo la serie delle misure:

- 1^a misura: 0 conteggi;
- 2^a misura: 0 conteggi;
- 3^a misura: 0 conteggi.

Qualcosa che non va? Il dubbio ovviamente c'è, anche se nelle prove preliminari lo strumento sembrava funzionare. Continuiamo:

- 4^a misura: 0 conteggi;
- 5^a misura: 0 conteggi;
- 6^a misura: 2 conteggi.

Funziona! Prima di procedere poniamoci alcune domande.

1. Quanto vale il livello di radioattività in base alle misure effettuate? Abbiamo ottenuto cinque volte 0 e una volta 0.667 conteggi al secondo. Quale numero prendere?

2. Perché mai all'inizio ci sono stati cinque 0 e improvvisamente un 2? Forse il contatore si era incantato e poi ha ripreso a funzionare?

3. Quale è la *previsione* sul numero di conteggi che saranno registrati nei prossimi 3 secondi? 0, 1, 2, 3, 10, 1000? Dovendo effettuare una *scommessa*[1], su quale di questi numeri puntereste?

4. Se per motivi indipendenti dalla nostra volontà l'esperimento fosse stato interrotto dopo le prime cinque misure, cosa si potrebbe dire sulla radioattività? (Assumiamo che sia di fondamentale importanza stimarne un qualche valore.) Aver misurato degli zeri è come non aver eseguito l'esperimento?

Effettuiamo altre misure, ottenendo ancora: 1, 1, 0, 1, 0,.... Eseguiamo in totale 100 misure e riportiamo a mano a mano i valori in una tabella (vedi tabella 1.1).

A questo punto possiamo riproporci le domande formulate dopo la $6^{\underline{a}}$ misura. A qualcuna di esse sappiamo già rispondere.

- Punto 2: *non sembra* che ci siano motivi per sospettare che il contatore si sia messo a funzionare improvvisamente, nel senso che - "col senno del poi" - cinque zeri consecutivi sembrano rientrare nella norma;

- Punto 3. dovendo scommettere sul risultato della 101-ma misura qualsiasi persona *punterebbe* sullo zero, seguito dall'1 e poi dal 2; più difficile è decidere fra il 3 e il 4, ma se proprio si dovesse scegliere fra i due valori si preferirebbe il 3 al 4. Non tanto perché il 3 si è verificato due volte e il 4 una volta sola, quanto perché *ci sembra di intravedere* che le frequenze con le quali si verificano i diversi numeri di conteggi seguano un andamento decrescente. Purtroppo per rispondere in modo corretto a tutte le altre domande bisognerà acquisire delle conoscenze teoriche su questo tipo di fenomeni *aleatori*;

[1]Si immagini di fare una scommessa con puntate uguali contro una ipotetica persona che accetterebbe. Ci si rende conto che se uno è fortemente propenso a scommettere in favore del verificarsi di un evento, difficilmente si troverà una persona che scommetta la stessa puntata contro, a meno che non sia un "pollo". Tale scommessa non è quindi *equa*. Variando invece il rapporto delle puntate si può invece arrivare al punto in cui si è indecisi se scommettere in favore o contro. Questi argomenti saranno trattati quando si formalizzerà il concetto di probabilità. Per ora, quando si parla di scommesse si pensi sempre di giocare alla pari con l'intento di vincere.

$T = 3\,s$																			
0	0	0	0	0	2	1	1	0	1	0	1	0	0	0	0	0	1	2	0
0	0	0	0	0	1	0	0	0	0	1	0	0	0	0	1	0	1	1	0
0	0	0	1	3	2	0	1	0	3	1	0	0	2	0	0	0	0	4	0
0	1	2	1	0	1	2	1	0	0	0	0	2	1	0	1	1	0	0	1
1	1	1	2	1	1	1	0	0	2	1	1	0	1	0	1	0	1	0	0
$T = 6\,s$																			
0	3	0	1	0	2	1	1	1	2	0	0	0	0	4	0	1	2	1	2
0	1	2	0	1	0	2	0	0	0	0	2	1	2	0	0	3	0	1	1
2	1	0	0	1	3	2	1	1	1	3	0	1	1	2	1	0	1	1	0
1	2	0	1	0	1	1	1	2	0	1	1	1	0	2	1	3	2	0	1
1	2	1	1	2	0	1	1	0	2	0	1	1	0	1	2	2	1	3	0
$T = 12\,s$																			
2	2	2	0	1	0	3	4	2	4	5	4	2	1	0	5	3	2	1	2
0	2	1	2	2	0	4	2	3	1	4	2	2	1	3	5	0	4	3	3
3	2	2	1	2	2	3	4	3	3	2	3	4	2	5	2	5	0	1	1
1	2	1	3	3	1	2	2	2	3	2	5	2	4	2	0	2	3	1	3
2	3	0	5	1	1	3	3	1	1	2	3	2	1	2	1	0	4	0	1
$T = 30\,s$																			
6	4	10	9	7	4	9	6	5	4	6	4	5	7	5	4	3	6	3	5
4	9	4	8	3	7	5	6	5	9	5	7	4	4	4	4	3	8	5	5
1	1	4	7	2	5	2	9	7	5	5	6	7	7	2	2	5	3	4	1
5	5	7	3	6	7	3	6	9	4	2	5	8	6	7	5	7	4	3	6
8	1	4	6	7	3	7	3	6	6	4	4	4	3	6	4	8	3	12	3
$T = 100\,s$																			
14	22	13	20	20	18	16	11	17	12	17	10	19	19	18	25	16	15	29	19
12	21	18	17	14	18	14	17	17	17	22	16	14	19	10	23	14	27	13	17
12	17	22	20	16	13	19	16	22	25	21	16	13	21	27	18	18	13	18	22
19	15	9	29	19	22	19	16	14	17	20	19	20	14	17	12	27	19	24	16
11	18	13	22	24	8	19	14	18	20	14	20	18	18	18	23	18	17	20	22
$T = 300\,s$																			
55	58	59	56	48	62	57	59	43	49	49	60	58	51	51	65	39	52	61	64
53	61	33	40	49	57	39	63	56	45	60	65	58	52	53	63	54	48	46	53
48	63	51	47	53	56	52	57	49	51	44	40	37	49	63	78	61	58	64	52
51	55	54	54	52	54	46	51	53	51	58	71	48	51	45	52	49	54	55	44
57	61	50	52	61	42	67	53	58	61	55	44	50	57	40	53	52	61	61	48

Tabella 1.1: Dati del contatore.

- formuliamo la domanda precedente in altri termini: quale conteggio ha maggiore *probabilità* di verificarsi alla 101-ma misura? (Diamo per ora scontato il concetto di probabilità.) Senza dubbio la risposta che qualsiasi persona darebbe è che il risultato più probabile sarà lo zero, quindi l'1, e così via.

Adesso cambiamo il tempo di misura e portiamolo a **6 secondi**. Cosa ci aspettiamo? La risposta intuitiva è che "il numero di conteggi si raddoppia". Cosa vuol dire esattamente? Che gli 1 diventeranno 2, i 2 diventeranno 4, etc.? Quindi gli 0 rimarranno 0 e non compariranno numeri dispari? Strano. Il risultato (tabella 1.1) è una diminuzione degli zeri, un aumento di 1, di 2 e di 3. Ancora una volta non si è verificato nessun numero di conteggi superiore a 4. C'è qualche ragione o è accaduto solo per caso?

Eseguiamo quindi altre misure, rispettivamente per 12, 30, 100 e 300 secondi (tabella 1.1) e poniamoci altre domande.

1. Quale valore di radioattività si ottiene per ciascuno dei gruppi di misura? Quale gruppo darà la misura più *precisa*? (Per il momento il concetto di "precisione" del linguaggio comune è sufficiente.)

2. È possibile combinare le informazioni di tutti i gruppetti? Se sì, la precisione sulla determinazione della radioattività aumenterà o diminuerà?

3. È possibile *decidere* se la radioattività si è mantenuta costante da una serie di misure all'altra, oppure se essa sta variando nel tempo?

1.1.2 Tempi di attesa per ottenere un numero prestabilito di conteggi

Nei paragrafi precedenti ci siamo interessati al numero di conteggi registrati con un contatore a scintillazione avendo fissato il tempo di lettura. È stata invece persa l'informazione sulla sequenza temporale degli eventi. Prima di passare alla misura, cerchiamo di farci un'idea di cosa ci aspettiamo in base ai dati della tabella 1.1.

- Il valore di radioattività è 'intorno' a 0.18 conteggi al secondo, ovvero abbiamo in media un conteggio ogni 5.6 secondi.

- Non ci dobbiamo attendere che i conteggi arrivino cadenzati al ritmo di uno ogni 5.6 secondi, sia perché è difficile immaginare un meccanismo che controlli la regolarità dell'arrivo di particelle radioattive provenienti da più sorgenti, sia perché i dati sperimentali mostrano grandi fluttuazioni dei tempi di arrivo intorno al *tempo* medio *di attesa*.

- La tabella 1.1 mostra come spesso (11 volte su 100), anche aspettando 12 secondi (oltre il doppio del tempo medio di attesa), non si sia verificato nessun evento. A volte invece (qualche volta ogni cento) in un tempo circa la metà di quello medio di attesa si verificano 3 o 4 eventi.

- Si noti inoltre come su tempi lunghi il processo in qualche modo si regolarizzi: sebbene la variazione fra il minimo e il massimo numero di conteggi aumenta con il tempo di misura, il tempo medio per ogni singola misura varia molto meno, per esempio va da 3.8 a 9.1 secondi (78 e 33 eventi in 300 secondi) nei conteggi registrati in 300 s.

Essendoci convinti che *ci aspettiamo* un evento ogni 5.6 secondi, anche se non ci dobbiamo sorprendere se a volte bisogna attendere molto di più o se altre volte arrivano raffiche di eventi quasi simultanei, poniamoci ancora delle domande.

1. Supponiamo di dividere il tempo di attesa in intervallini di un secondo e di dover fare una scommessa in cui vince chi indovina l'intervallino esatto in cui si verificherà il conteggio, su quale intervallino converrà puntare? Sul sesto, ovvero quello che è quasi centrato sul tempo medio di attesa? Detto in altre parole, è più probabile misurare tempi compresi fra 5 e 6 secondi che misurare tempi compresi fra 0 e 1 secondi, fra 1 e 2, etc.? O, ancora, è più probabile osservare valori compresi fra 0 e 5.6 secondi o valori maggiori di 5.6 secondi?

2. Cosa ci dobbiamo aspettare se ci interessiamo invece al tempo di attesa per registrare k conteggi? È ragionevole pensare che il tempo medio di attesa sia k volte quello necessario per il primo conteggio? Come saranno le fluttuazioni intorno a tale valore?

Anche per queste domande cerchiamo delle risposte intuitive:

1. il tempo di attesa non può essere negativo, ma può essere arbitrariamente grande (anche se le grandissime fluttuazioni sembrano molto poco probabili). Quindi il numero di intervallini è infinito. Ma essendo la media nel sesto intervallino, ci aspettiamo che gli intervallini al si sotto di esso si verifichino con frequenza superiore di quelli al di sopra in modo tale da compensare, nel calcolo della media, i grandi tempi di questi ultimi.

   Anche se da questi argomenti non si può concludere l'andamento decrescente della probabilità in funzione del numero di intervallino (come verificheremo), si può senz'altro affermare che è più probabile misurare tempi inferiori che tempi superiori a quello medio di attesa (anticipando il risultato esatto: 63 % contro 37 %);

2. è abbastanza naturale che il tempo medio di attesa debba essere proporzionale al numero di conteggi. Più complicato è fare affermazioni intuitive sulle fluttuazioni dei tempi intorno alla media e sulle probabilità dei vari intervallini. Ciò nonostante sembra abbastanza naturale attendersi che alla regolarizzazione del processo per grandi tempi di misura se ne debba accompagnare una analoga per un gran numero di conteggi. Per esempio, dalla tabella 1.1 per 300 s si vede come fluttuazioni oltre il $\pm 30\,\%$ del numero tipico (intorno a 50) siano rare. Quindi in media bisogna aspettare circa 300 s per osservare una cinquantina conteggi, ma dovranno essere veramente molto rare le fluttuazioni di oltre il doppio o di meno della metà di tale tempo, cosa invece frequentissima nel caso di un solo conteggio.

1 conteggio									
6.9	6.2	4.8	12.1	17.2	6.6	10.5	3.0	17.2	0.2
1.5	0.6	11.5	1.7	1.7	3.8	2.9	5.5	4.1	8.1
2.5	6.4	0.1	0.3	4.2	2.3	0.2	8.0	5.3	2.4
1.8	2.9	9.6	0.4	9.6	2.6	10.4	5.3	10.2	0.4
17.3	0.7	3.1	11.0	9.0	1.0	5.1	1.8	6.6	2.8
8.1	11.3	0.2	0.0	8.0	1.4	4.1	2.4	6.7	3.8
15.0	2.1	8.9	0.0	0.3	5.9	2.6	1.2	3.8	20.6
1.0	3.4	0.9	5.0	1.8	2.6	10.0	11.4	7.8	4.4
12.6	1.9	0.8	8.0	3.8	18.8	0.6	6.2	0.9	5.7
3.2	4.6	2.9	0.6	2.7	5.7	0.2	14.4	8.8	1.6
2 conteggi									
1.7	17.6	26.1	2.6	9.7	30.5	19.7	14.2	17.1	9.9
5.1	19.2	13.5	9.8	6.6	7.8	7.1	16.3	15.9	25.2
9.9	4.1	12.0	14.8	1.1	28.0	33.8	5.8	34.3	17.1
11.5	1.8	12.4	14.0	2.7	2.2	16.2	7.6	0.6	9.0
11.8	15.2	17.1	18.5	29.5	6.1	7.1	7.2	9.3	15.6
5 conteggi									
42.5	15.3	12.0	40.6	38.9	25.8	18.1	36.8	30.1	30.9
28.2	21.6	35.8	7.6	18.8	20.6	36.8	28.9	17.9	16.1
38.4	19.2	25.5	43.3	16.8	16.8	48.5	21.2	30.2	24.8
36.7	43.1	35.1	40.6	28.4	51.0	34.4	52.1	35.3	34.8
39.0	40.6	24.5	20.5	27.7	33.4	52.2	13.3	39.4	11.6
10 conteggi									
61.0	36.0	55.6	47.6	28.4	66.8	64.5	31.4	29.7	49.9
67.5	42.9	62.3	57.2	70.1	73.1	25.8	46.9	43.4	47.6
55.6	45.1	49.9	47.1	103.9	68.5	61.1	83.7	103.3	63.8
24.5	54.6	67.9	32.9	68.8	60.2	36.9	70.5	54.1	50.4
35.2	85.7	50.7	45.9	89.7	51.4	64.4	109.2	58.7	76.0
20 conteggi									
115.7	136.7	86.7	145.2	160.9	121.0	110.9	98.8	94.2	87.3
117.9	110.2	124.1	86.6	146.2	120.2	114.4	124.8	64.7	107.1
129.7	108.4	115.7	120.8	141.8	157.7	168.7	96.0	101.1	118.4
82.1	80.0	98.8	140.5	78.2	128.6	150.1	114.4	83.5	120.7
118.5	122.4	113.3	108.7	99.6	141.8	149.8	131.0	136.7	146.8
50 conteggi									
346	228	327	228	209	297	285	291	245	280
284	272	325	201	286	301	298	335	233	346
294	296	267	254	240	251	297	244	344	267
190	265	267	249	225	238	266	286	285	235
227	305	246	246	296	253	255	286	208	285
100 conteggi									
590	669	580	586	498	513	629	537	579	568
627	542	493	527	459	518	629	518	586	619
516	617	548	586	594	524	564	551	515	558
549	627	570	575	490	626	533	491	489	529
488	663	608	545	550	652	574	560	623	623

Tabella 1.2: Tempi di attesa per osservare un numero prefissato di eventi. Questi dati simulano quelli ottenibili con contatore che lavori nelle stesse condizioni di quello con cui sono stati ottenuti i dati di tabella 1.1.

I dati sperimentali (simulati) sono riportati in tabella 1.2. Si notino le enormi fluttuazioni per attendere il primo conteggio e le minori fluttuazioni relative per registrare 100 conteggi (per un conteggio sono state riportati più valori "misurati" per meglio capirne il comportamento).

Per comprendere le ragioni della migliore stabilità per grandi conteggi è importante la seguente osservazione: essendo il fenomeno che determina un conteggio del tutto casuale, si ha la stessa distribuzione di tempi di attesa sia che si faccia partire il cronometro manualmente a tempi arbitrari, sia che si prenda per "start" del cronometro l'istante di arrivo dell'ultimo evento. Quindi il tempo di attesa di n eventi è pari alla somma di n tempi di attesa di un evento, ognuno indipendente dall'altro. Per questo motivo le enormi fluttuazioni intorno al tempo di attesa medio tenderanno a compensarsi, "in media", ma non necessariamente!

1.2 Pallinometro

Consideriamo come secondo esperimento il quinconce[2] di Galton, familiarmente chiamato "pallinometro" dagli studenti e indicato con tale nome in questo testo. Il "vero" pallinometro è costituito da una tavola con dei chiodi infissi "a quinconce". La tavola è disposta verticalmente, il numero di chiodi per ciascuna riga cresce dall'alto verso il basso. Sopra ai chiodi, parallella alla tavola, c'è una lastra di plexiglass. In alcuni punti la lastra di plexiglass ha dei fori che permettono l'inserimento di palline, le quali scendono poi per gravità. La dimensioni delle palline e la spaziatura verticale e orizzontale fra chiodi sono scelte in modo tale che ogni pallina, scendendo da una fila all'altra, urti un chiodo circa centralmente. La lastra di plexiglass serve a costringere il moto delle palline su un piano verticale. La figura 1.2 mostra la schematizzazione di possibili traiettorie delle palline (le traiettorie reali da un chiodo all'altro sono più propriamente dei tratti di parabola). Le palline sono raccolte in fondo al pallinometro in opportune feritoie.

Un pallinometro ha il grosso "inconveniente" di essere un oggetto "reale" e quindi poco adatto allo scopo di effettuare degli esperimenti in condizioni quasi ideali per il quale era stato concepito. Infatti:

- il posizionamanto dei chiodi non è "perfetto" e si modifica con l'uso;
- il comportamento delle palline dipende moltissimo dal materiale di cui esse sono costituite e non è semplicissimo trovare quelle appropriate per mostrare l'effetto voluto;
- le esperienze richiedono un grande numero di prove, e non è affatto interessante perdere tempo a lanciare palline senza imparare nessuna manualità utile per il laboratorio, mentre la parte più interessante dell'esperienza consiste nell'analisi dei dati.

Si preferisce quindi usare un pallinometro simulato su personal computer (PC), con tanto di effetti grafici e di suono (naturalmente da escludere quando si fan-

[2]Con quinconce non si intende nient'altro che il modo di disporre degli oggetti a righe sfalsate, come mostrato in figura 1.2.

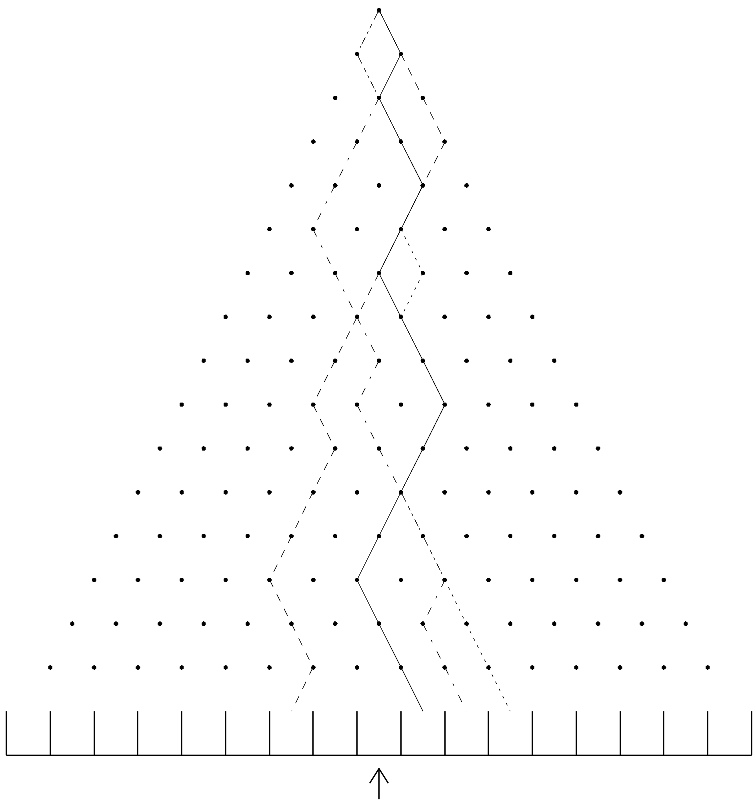

Figura 1.2: SChema del quinconce di Galton, familiarmente "pallinometro" per generazioni di studenti.

no molti lanci). Nei ragionamenti che seguono fingeremo di utilizzare un "vero" pallinometro "ideale" sul quale faremo delle considerazioni di meccanica come in un "gedankenen Experiment"[3].

1.2.1 Previsioni

Con il pallinometro si possono effettuare diverse esperienze. In questa esperienza simulata lo utilizzeremo nella sua forma pressoché minimale. Supponiamo di lasciar cadere una pallina sopra un chiodo della seconda fila dal basso (vedi figura 1.3), facendo in modo che l'urto sia il più centrale possibile.

1. Dove andrà la pallina? La meccanica insegna che, sotto tali condizioni, la pallina rimbalzerà in alto e, se l'urto è perfettamente elastico, segui-

[3] Ad essere franchi il quinconce meccanico è una specie di assurdità didattica. Molto spesso infatti la teoria del suo funzionamento viene presentata come se fosse effettivamente la migliore descrizione del movimento delle palline. Invece è lo strumento che viene costruito allo scopo di rappresentare un modello di moto casuale che sia ha già in mente e che nessun quinconce reale riuscirà mai a riprodurre. Per questo è preferibile la "sua simulazione" al computer (il termine simulazione è improprio in quanto - ripetiamo - con il computer non si prova a simulare nessuna realtà, ma si ottiene semplicemente una migliore realizzazione pratica del modello di cammino casuale ideale).

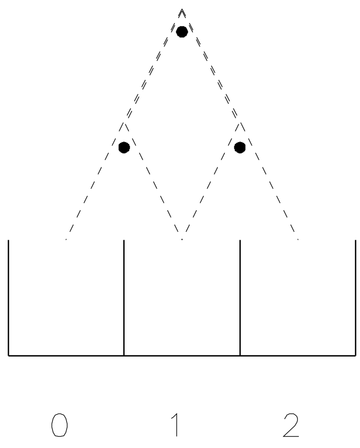

Figura 1.3: Pallinometro a due file di chiodi. Sono indicate schematicamente le quattro possibili traiettorie della pallina.

terà infinitamente in tale moto periodico. L'esperienza insegna invece che la pallina, rimbalzando sul chiodo, scende nella fila sottostante deviando a destra o a sinistra del chiodo che ha urtato. Anche se ci si mette tanta buona volontà è molto difficile far rimbalzare più di due volte una pallina sullo stesso chiodo. Quindi, per semplificare il problema assumiamo che non lo faccia mai (che lo possa fare è inessenziale per quello che ci interessa).

2. Dopo un urto la pallina andrà a sinistra o a destra del chiodo? Non c'è nessun motivo che ci porti a preferire una direzione. Diciamo quindi che le due possibilità hanno la stessa *probabilità*.

3. In quali delle tre feritoie (o celletta, *"bin"*) di raccolta andrà la pallina? Considerando le possibili traiettorie che la pallina può percorrere, la celletta centrale sembra la favorita, in quanto ci sono due *possibilità* su quattro che la pallina ci finisca. Quindi sembra ragionevole dire che la sua probabilità sia doppia rispetto alle altre due.

Per inciso, notiamo che, anche se non abbiamo ancora definito esattamente cosa intendiamo con il termine "probabilità", abbiamo già fatto uso di questa parola, come già fatto nell'esperimento precedente, contando sull'intuizione e sull'uso che se ne fa nella vita quotidiana. È da notare comunque come lo stesso concetto sia stato utilizzato in due diversi modi di ragionare:

- nel caso del contatore avevamo detto che è più probabile il risultato che si era presentato più frequentemente nel passato;

- nel caso del pallinometro abbiamo detto che è più probabile il risultato che ha "più possibilità" di verificarsi, senza dover effettuare alcun esperimento.

Poniamoci la seguente domanda: supponiamo di mostrare i risultati dell'esperimento della caduta di molte palline ad una persona razionale che non conosca il meccanismo interno del pallinometro. Converrà sul valore di probabilità che stimiamo noi? Ad esempio, affermerà, analizzando i dati, che la probabilità della celletta centrale è doppia delle altre due e quindi la probabilità che la pallina finisca al centro è pari alla probabilità che cada altrove?

1.2.2 Risultati

Eseguiamo un certo numero di lanci di palline. Per esempio lanciamo 10 volte 10 palline e riportiamo i risultati in una tabella (tabella 1.3). Praticamente ogni volta si ottengono risultati diversi. Considerando soltanto la prima sequenza si direbbe che i bin 1 e 2 hanno probabilità doppia del bin 0. Considerando solo la sesta si direbbe invece che è da favorire il bin 2. In totale le palline sono andate 53 volte nel bin centrale, 21 volte in quello a sinistra e 26 in quello a destra. Cosa se ne potrebbe concludere sulla probabilità - ripetiamo - se non si conoscesse il meccanismo interno dell'apparato?

Eseguiamo quindi altri lanci di palline, in numero maggiore, e riportiamo il risultato nella stessa tabella 1.3. Osservando i dati si può dire che effettivamente, se si dovessero dare previsioni sul prossimo lancio, la probabilità del bin centrale è circa pari a quella degli altri due messi insieme e che i due laterali hanno probabilità circa uguale. È da notare come ciò non possa essere affermato esattamente senza conoscere (o "ipotizzare") il meccanismo interno di funzionamento.

1.3 ↻ Proprietà chimico-fisiche delle acque minerali

È possibile raccogliere dati interessanti su cui applicare le metodologie di analisi che incontreremo semplicemente leggendo le etichette delle bottiglie di acqua minerale. La tabella 1.4 mostra le proprietà chimico-fisiche riportate su alcune acque in. Lo scopo dell'esperienza è quello di analizzare tali dati anche senza le dovute conoscenze di chimica fisica, considerando l'acqua una "scatola nera" dalla quale sono state estratte tali informazioni, e cercando di capire se esiste qualche legge che le leghi fra loro.

1.4 ↻ Nascono più femmine che maschi?

Tutti sanno che nella nostra società ci sono più donne che uomini. Ce lo ripetono alla vigilia di ogni elezione o referendum e, vista la frequenza con cui tali eventi si verificano in Italia, è impossibile non esserne al corrente. L'idea è talmente comune che la maggior parte delle persone è convinta che effettivamente nascano più femmine che maschi. La tabella 1.5 offre il quadro della situazione italiana fotografata nel 1981. Si vede come effettivamente nelle fasce giovanili la percentuale dei maschi supera il 51 %. Questo è un fatto noto da secoli agli statistici e dipende molto poco dal tempo e dalle varie popolazioni. Qualcuno addirittura aveva utilizzato questo dato come prova della Divina Provvidenza!

Nr di palline	Sequenza	Bin 0	Bin 1	Bin 2
10	1	2	4	4
	2	2	7	1
	3	3	5	2
	4	2	5	3
	5	2	5	3
	6	2	3	5
	7	1	6	3
	8	0	9	1
	9	4	4	2
	10	3	5	2
30	1	7	17	6
	2	6	17	7
	3	6	16	8
	4	3	15	12
	5	7	13	10
	6	10	12	8
	7	11	13	6
	8	9	12	9
	9	8	17	5
	10	10	15	5
100	1	26	50	24
	2	22	54	24
	3	24	52	24
	4	26	49	25
	5	24	49	27
	6	19	52	29
	7	22	52	26
	8	23	47	30
	9	23	54	23
	10	19	55	26
300	1	59	165	76
	2	78	145	77
	3	64	148	88
	4	83	138	79
	5	78	151	71
	6	80	132	88
	7	78	154	68
	8	69	156	75
	9	73	153	74
	10	85	142	73
1000	1	254	487	259
	2	247	514	239
	3	259	499	242
	4	229	522	249
	5	234	497	269
	6	258	493	249
	7	251	508	241
	8	246	505	249
	9	260	512	228
	10	247	508	245

Tabella 1.3: Dati simulati del relativi al pallinometro da due file di chiodi.

Acqua	T (°C)	pH	Conducibilità (μS/cm)	Res. fisso (mg/l)	CO_2 libera (mg/l)
Levissima	5.8	7.8	107.5	73.5	-
Ferrarelle	-	6.1	1800	1283	2150
Acqua di Nepi	-	5.68	741	567.6	1930
San Benedetto	16	7.72	354	233	-
Appia	19.2	-	1000	681.3	373
Egeria	18.1	5.87	980	617.9	683
Cottorella	11	7.4	420	286	1.9
Capannelle	18.05	6.24	1002	685	532
Claudia	22.7	5.75	965	763	?
Cinzia	24.2	-	465	307.5	?
Rocchetta	-	7.41	260.8	173.32	10.76
Uliveto	-	6.0	1410	1004	820
Fabia	14.7	7.02	643	436.7	58
Fiuggi	-	6.2	156	106.5	-
Lete	12	6.1	1160	873.5	898.8
Fonte Serena	12.2	7.6	248	156.1	0.5
Sangemini	18	6.33	1333	955	928
Gaia	-	7.45	465	315	22
Cutolo	15	6.52	668	496	1130
Natta	-	6.5	410	318	147
Santa Croce	6.6	7.61	$2.97 \cdot 10^2$	169.5	12.0
L'Acqua	15.7	7.4	490	274	14.8

Tabella 1.4: Grandezze fisico-chimiche di acque minerali: temperatura alla sorgente, pH (a 18 o 20 °C), conducibilità elettrica specifica (a 18 o 20 °C) e residuo fisso a 180 °C.

Oltre che per valutazioni sociologiche e biologiche (oltre che teologiche...) questi dati offrono spunti interessanti per questioni probabilistiche. Si provi a rispondere alle seguenti domande (o almeno a porsi il problema):

- Quanto vale la probabilità che il primo nato dell'anno 2000 in Italia sia maschio?
- Quanto vale la probabilità che un italiano scelto a caso nel 2023 abbia un'età superiore a 70 anni?
- Quanto vale la probabilità che, sempre nel 2023, un italiano scelto a caso nella fascia fra 40 e 44 anni sia maschio?

1.5 ↻ Coincidenze di compleanno

Succede di conoscere persone che hanno il compleanno lo stesso giorno, oppure che tale coincidenza si verifichi in un gruppo poco numeroso di persone (relativo al numero di giorni di un anno) come ad esempio una classe di liceo o un'associazione sportiva. Utilizzando la propria esperienza in materia, quanto

	età	femmine	maschi	totale
Nord	0 - 5	616892	646485	1263377
	5 - 9	832085	873859	1705944
	10 - 14	929932	970990	1900922
	15 - 19	974798	1013410	1988208
	20 - 24	880068	913861	1793929
	25 - 29	851161	867405	1718566
	30 - 34	927736	936186	1863922
	35 - 39	851873	855159	1707032
	40 - 44	937079	933549	1870628
	45 - 49	865947	848897	1714844
	50 - 54	878616	837733	1716349
	55 - 59	872694	770533	1643227
	60 - 64	613678	504937	1118615
	65 - 69	717096	553601	1270697
	70 - 74	642787	444824	1087611
	≥ 75	887527	449488	1337015
	totale nord	13279969	12420917	25700886
Centro	0 - 5	277706	293329	571035
	5 - 9	360455	377193	737648
	10 - 14	390073	407175	797248
	15 - 19	410161	425315	835476
	20 - 24	373701	383883	757584
	25 - 29	358627	357313	715940
	30 - 34	386649	378719	765368
	35 - 39	351628	338179	689807
	40 - 44	387684	369514	757198
	45 - 49	362934	345973	708907
	50 - 54	372289	354167	726456
	55 - 59	374274	340965	715239
	60 - 64	267313	232756	500069
	65 - 69	290970	241285	532255
	70 - 74	251889	192795	444684
	≥ 75	346760	201017	547777
	Totale centro	5563113	5239578	10802691
Sud e Isole	0 - 5	740855	787195	1528050
	5 - 9	859726	911866	1771592
	10 - 14	903724	948074	1851798
	15 - 19	917793	947362	1865155
	20 - 24	792216	800113	1592329
	25 - 29	696269	687834	1384103
	30 - 34	689458	681001	1370459
	35 - 39	574053	566134	1140187
	40 - 44	595946	571924	1167870
	45 - 49	588462	553263	1141725
	50 - 54	596173	552901	1149074
	55 - 59	569231	509721	1078952
	60 - 64	404716	342237	746953
	65 - 69	434155	370333	804488
	70 - 74	363471	298187	661658
	≥ 75	481227	317714	798941
	totale sud e isole	10207475	9845859	20053334
	Totale Italia	29050557	27506354	56556911

Tabella 1.5: Popolazione italiana secondo il censimento del 1981.

Top 15 dei numeri ritardatari (12/3/94 $< t <$ 19/3/94)																
BA	**num.**	**38**	**63**	**28**	**85**	**27**	**23**	**39**	**21**	**68**	**57**	**15**	**45**	**73**	**54**	**48**
	rit.	149	82	75	55	52	43	37	37	35	31	31	30	29	29	28
CA	**num.**	**8**	**16**	**55**	**14**	**20**	**48**	**70**	**39**	**47**	**35**	**32**	**24**	**60**	**45**	**41**
	rit.	93	75	72	71	67	60	50	49	47	45	44	43	40	39	38
FI	**num.**	**3**	**51**	**42**	**49**	**41**	**82**	**57**	**47**	**26**	**60**	**38**	**79**	**23**	**71**	**37**
	rit.	103	58	50	45	44	44	41	41	37	34	33	31	29	28	28
GE	**num.**	**66**	**4**	**3**	**67**	**80**	**41**	**51**	**48**	**36**	**44**	**11**	**13**	**24**	**83**	**75**
	rit.	76	75	58	44	38	37	36	36	35	34	34	31	31	31	30
MI	**num.**	**85**	**23**	**24**	**7**	**62**	**64**	**57**	**5**	**2**	**38**	**6**	**42**	**75**	**26**	**33**
	rit.	62	62	61	54	47	43	42	42	38	38	37	37	36	34	34
NA	**num.**	**30**	**54**	**50**	**80**	**37**	**13**	**36**	**53**	**67**	**90**	**33**	**73**	**2**	**63**	**71**
	rit.	78	57	53	52	51	49	45	43	37	34	33	32	32	28	25
PA	**num.**	**56**	**2**	**50**	**10**	**49**	**31**	**17**	**33**	**90**	**44**	**11**	**15**	**39**	**62**	**36**
	rit.	62	53	52	48	45	44	43	42	40	37	37	36	33	33	31
Roma	**num.**	**49**	**44**	**52**	**55**	**54**	**61**	**72**	**42**	**12**	**87**	**7**	**22**	**10**	**29**	**32**
	rit.	79	64	58	53	52	51	50	50	50	49	46	44	43	40	39
TO	**num.**	**18**	**11**	**35**	**38**	**66**	**53**	**87**	**49**	**41**	**7**	**61**	**69**	**45**	**72**	**59**
	rit.	81	79	76	61	56	55	54	51	49	48	44	37	36	33	31
VE	**num.**	**3**	**33**	**70**	**78**	**43**	**37**	**50**	**7**	**81**	**65**	**34**	**36**	**63**	**30**	**61**
	rit.	132	76	72	61	48	48	47	46	42	40	39	37	37	34	34
Tutte	**num.**	**21**	**3**	**6**	**59**	**13**	**5**	**17**	**30**	**33**	**42**	**50**	**63**	**71**	**2**	**8**
	rit.	6	5	5	5	4	3	3	3	3	3	3	3	3	2	2

Tabella 1.6: Tabella dei numeri maggiormente ritardatari al lotto italiano dopo l'estrazione di sabato 12 marzo 1994.

vale la probabilità che fra trenta persone ce ne siano almeno due che hanno il compleannno lo stesso giorno? Non c'è bisogno di dare un numero preciso. Basta provare a darne una stima qualitativa ("nulla", "bassa", . . ., "altissima").

1.6 ↻ Numeri ritardatari al lotto

La tabella 1.6 mostra i maggiori ritardi registrati al gioco del lotto intorno alla metà di marzo 1994. Quanto è da riternersi sorprendente che dei numeri ritardino per oltre 50, 100 o 150 settimane? Ritieni che i numeri ritardatari debbano più facilmente degli altri, in quanto "si debbono mettere in regola con le leggi della statistica", oppure debbano uscire con la stessa frequenza degli altri? Oppure, sono forse da considerare più probabili i numeri usciti più facilmente nel passato ("numeri caldi")? Si immagini di scommettere sui numeri che al momento in cui stai leggendo questo testo ci sia almeno un numero che ritardi almeno 100 estrazioni. È conveniente scommettere pro o contro? (ovviamente, si assume il lettore non sia informato della situazione ritardi).

1.7 Nota semantica

Nei primi paragrafi sono state utilizzate delle parole o delle espressioni delle quali il lettore si sarebbe atteso una qualche definizione. Ad esempio, si è parlato di "misura", "procedure di misura", "valore vero", "evento", "probabilità", "aleatorio", "calibrazione", "incertezza", "precisione", "verifica" e "media". Per ora il senso di queste parole è quello che attribuisce loro, in base al contesto, una persona di media cultura. Nel seguito molti di questi concetti verranno meglio definiti per evitare le eventuali ambiguità che essi hanno nel linguaggio comune.

Sono state inoltre incontrate altre espressioni inusuali per un libro di testo, specie di una disciplina scientifica, come "ci sembra che ...", "punterebbe su ...", "assumiamo che ...", "è ragionevole che ...", "praticamente" e "ci attendiamo che ...". Chi è abituato soltanto ad espressioni del tipo *"dato ... segue ..."*, *"Principio di Tizio"* e *"Legge di Caio"* sarà rimasto quanto meno sorpreso. Si sarà chiesto se è questo il modo di ragionare del mondo scientifico. Senza entrare in sottili discussioni epistemologiche che ci porterebbero lontano, possiamo fin d'ora affermare che la risposta è positiva. Si lavora costantemente fra dati empirici da un lato e supposizioni, intuizioni, assunzioni, ipotesi di ragionevolezza e addirittura semplici principi "estetici" ed "economici" dall'altro. Questi principi valgono in particolare nella formulazione di ipotesi scientifiche. Ciò nonostante, quando la massa di dati sperimentali è grande ed è alto numero di controlli positivi, si è "praticamente certi" che il mondo funziona in un certo modo o che una grandezza fisica abbia un certo valore.

Capitolo 2

Primo sguardo ai dati di laboratorio

Vengono ora descritte semplici esperienze effettivamente realizzate in laboratorio. Esse saranno analizzate in dettaglio - nel testo o negli esercizi - nei prossimi capitoli. Esse sono state scelte con il criterio di offrire spunti di analisi su tutti gli argomenti che saranno trattati nel seguito, pur non richiedendo conoscenze profonde di Fisica Generale. In particolare, esse offrono fin dall'inizio spunti di riflessioni sui temi di errori e incertezze di misure e mostrano come le soluzioni intuitive a tali problemi non sempre sono corrette. Inoltre le semplicissime esperienze dell'interpolazione fra le tacche e delle misure manuali di tempo risolvono in modo pragmatico tante discussioni accademiche su tali argomenti. Esse sono pertanto vivamente raccomandate ad ogni studente.

2.1 Misura foglio A4

Questa esperienza è molto semplice e si presta a schematizzare alcune delle problematiche che intervengono nella misura. Lo scopo è di determinare le dimensioni di un foglio di carta formato DIN A4 (quello delle normali fotocopie). Ogni studente riceve un foglio ed uno strumento di misura. I fogli sono stati presi dallo stesso blocco ed i lati sono stati in qualche modo segnati in modo tale da garantire al massimo che tutti stiano misurando la stessa grandezza fisica. Gli strumenti di misura consistono in fotocopie di un righello disegnato su un calendarietto omaggio. Il righello è lungo appena 9 cm ed inoltre è stato riprodotto con rapporti di ingrandimento variabili, così da avere differenze dell'ordine di qualche parte per mille tra uno strumento e l'altro (non è banale con le comuni fotocopiatrici...). Nelle misure c'è un problema dovuto al fatto che il righello è molto più corto dei lati da misurare. Esso deve essere quindi spostato più volte, facendo corrispondere la nuova posizione dello zero con la precedente posizione del fondo scala. Gli studenti sono stati invitati, nell'eseguire questa operazione, a segnare la fine del righello su un foglio diverso da quello da misurare, in modo tale che ogni misura sia il più possibile indipendente dalle altre.

	N.C.			V.P.		
i	a (cm)	b (cm)	c (cm)	a (cm)	b (cm)	c (cm)
1	29.40	20.85	36.10	29.53	20.90	36.11
2	29.43	20.87	36.08	29.50	20.92	36.19
3	29.49	20.90	36.00	29.58	20.91	36.14
4	29.23	20.65	36.20	29.55	20.93	36.16
5	29.21	20.66	36.10	29.60	20.88	36.14
6	29.39	20.89	36.05	29.54	20.89	36.15
7	29.41	20.86	35.90	29.52	20.91	36.18
8	29.40	20.82	35.85	29.54	20.94	36.22
9	29.42	20.81	35.87	29.50	20.89	36.16
10	29.48	20.89	35.93	29.57	20.90	36.13
11	29.43	20.81	36.00	29.53	20.91	36.12
12	29.47	20.87	36.05	29.54	20.93	36.15
13	29.43	20.86	36.15	29.50	20.92	36.19
14	29.48	20.79	36.10	29.49	20.94	36.12
15	29.40	20.85	36.12	29.47	20.92	36.10
16	29.41	20.88	36.01	29.51	20.88	36.23
17	29.45	20.69	35.99	29.45	20.87	36.12
18	29.43	20.83	36.05	29.48	20.90	36.18
19				29.51	20.91	36.19
20				29.50	20.89	36.21
	A.P.			A.S.		
i	a (cm)	b (cm)	c (cm)	a (cm)	b (cm)	c (cm)
1	29.78	21.05	36.35	29.42	20.80	36.08
2	29.60	20.95	36.40	29.43	20.85	36.10
3	29.59	21.00	36.25	29.43	20.85	36.08
4	29.55	20.95	36.40	29.49	20.90	36.06
5	29.61	21.00	36.39	29.40	20.84	36.10
6	29.70	20.95	36.45	29.39	20.90	36.06
7	29.65	20.95	36.30	29.42	20.89	36.02
8	29.67	21.00	36.32	29.40	20.89	36.10
9	29.59	21.01	36.40	29.42	20.88	36.10
10	29.65	21.00	36.25	29.39	20.88	36.08
11	29.61	20.95	36.35	29.41	20.86	36.08
12	29.58	21.01	36.35	29.40	20.90	36.05
13	29.73	20.95	36.32	29.43	20.86	36.10
14	29.61	21.00	36.30	29.48	20.85	36.06
15	29.55	21.00	36.40	29.42	20.86	36.08
16	29.65	20.95	36.25	29.46	20.89	36.08
17	29.65	21.05	36.30	29.44	20.88	36.06
18	29.59	21.00	36.32	29.46	20.88	36.10
19	29.70	20.95	36.36	29.42	20.85	36.09
20	29.55	21.00	36.40	29.42	20.88	36.10

Tabella 2.1: Dati della misura del foglio DIN A4. a, b e c indicano i due lati e la diagonale del foglio.

I dati ottenuti da quattro studenti sono riportati in tabella 2.1. Essi saranno analizzati in dettaglio nei prossimi capitoli. Limitiamoci, per ora, a porci alcune domande.

1. Perché gli studenti, nonostante tutta la buona volontà, ottengono ad ogni misura dei numeri diversi?

2. Cosa potrà affermare ogni studente sul valore dei lati del foglio di carta, se assume che il suo righello sia perfettamente calibrato e non conosce le misure degli altri studenti?

3. Con le ipotesi del punto precedente, all'aumentare delle misure ogni studente migliorerà o peggiorerà la sua conoscenza delle grandezze fisiche a cui è interessato? (Si direbbe che più misure si fanno e più facilmente capitano valori lontani da quelli tipici: è quindi preferibile effettuare poche misure?)

4. Confrontando le misure dei singoli studenti, l'incertezza sui valori delle grandezze fisiche diminuirà o aumenterà? (Si provi a dare una risposta intuitiva osservando attentamente i dati sperimentali, ad esempio limitandosi al lato a.)

2.2 Capacità di interpolazione fra le tacche e incertezza di lettura

Mostriamo ora una semplicissima misura legata alla questione della liceità di leggere il valore indicato da uno strumento stimando un valore fra le tacche di una scala analogica. Infatti nell'esperienza precedente i valori sono stati riportati fino al decimo di millimetro e qualcuno potrebbe avere il dubbio che tutto quello che è al di sotto dell'intervallo di scala (la divisione da 1 mm) sia soltanto frutto della fantasia e irrilevante ai fini della determinazione delle grandezze di interesse. La migliore soluzione del problema è quella di effettuare un esperimento in laboratorio affinché ciascuno possa valutare le proprie capacità di lettura. Uno strumento che si presta bene per tale scopo è il *calibro* ventesimale (vedi figura 2.1). La scala è graduata in millimetri e per facilitare la lettura delle frazioni di millimetro il calibro è dotato di un *nonio*. La posizione dello zero del nonio nella scala principale indica la lettura da eseguire. Il nonio è costituito da 20 divisioni che hanno una lunghezza totale pari a 19 divisioni della scala principale (ovvero 19 mm). Quindi se la tacca di zero del nonio coincide esattamente con una tacca della scala principale anche l'ultima tacca del nonio (la 21-ma) coinciderà con una tacca della scala principale. Ad ogni spostamento successivo di 1/20 mm del cursore, tacche successive del nonio verranno a coincidere una alla volta con le tacche superiori. Quindi la tacca del nonio che meglio coincide con una delle tacche superiori indica i ventesimi di millimetro.

L'esperienza consiste nel posizionare a caso il cursore e nello stimare al meglio in valore misurato coprendo il nonio (eccetto la prima tacca, ovviamente). Questo valore *stimato* viene riportato in una tabella. Servendosi poi dell'informazione del nonio si scrive il valore *letto* sulla seconda colonna (è

i	L.T. l_S	L.T. l_N	L.T. Δ	P.P. l_S	P.P. l_N	P.P. Δ
1	18.35	18.45	-0.10	21.90	21.80	$+0.10$
2	46.00	46.10	-0.10	34.15	34.20	-0.05
3	37.15	37.10	$+0.05$	72.20	72.25	-0.05
4	28.20	28.25	-0.05	15.25	15.30	-0.05
5	5.60	5.65	-0.05	27.30	27.40	-0.10
6	34.20	34.30	-0.10	56.50	56.55	-0.05
7	48.90	48.75	$+0.15$	34.40	34.45	-0.05
8	16.75	16.85	-0.10	9.40	9.55	-0.15
9	23.45	23.30	$+0.15$	47.25	47.40	-0.15
10	72.20	72.30	-0.10	18.90	18.80	$+0.10$
i	C.P. l_S	C.P. l_N	C.P. Δ	S.M. l_S	S.M. l_N	S.M. Δ
1	40.85	40.85	0.00	30.75	30.70	$+0.05$
2	62.15	62.10	$+0.05$	44.95	44.90	$+0.05$
3	23.20	23.15	$+0.05$	65.60	65.50	$+0.10$
4	12.25	12.35	-0.10	33.70	33.80	-0.10
5	59.20	59.30	-0.10	26.90	26.95	-0.05
6	43.70	43.65	$+0,05$	74.00	74.10	-0.10
7	23.20	23.30	-0.10	7.90	8.00	-0.10
8	37.00	37.10	-0.10	22.75	22.65	$+0.10$
9	47.35	47.20	$+0.15$	66.70	66.60	$+0.10$
10	29.75	29.80	-0.05	43.25	43.35	-0.10
i	C.U. l_S	C.U. l_N	C.U. Δ	L.P. l_S	L.P. l_N	L.P. Δ
1	49.15	49.15	0.00	18.95	18.75	$+0.20$
2	20.35	20.50	-0.15	15.50	15.55	-0.05
3	22.00	22.00	0.00	27.10	27.20	-0.10
4	29.55	29.80	-0.05	30.50	30.55	-0.05
5	38.15	38.20	-0.05	**18.50**	19.00	$\mathbf{-0.50}$
6	24.45	24.45	0.00	13.00	13.00	0.00
7	25.00	25.00	0.00	19.20	19.30	-0.10
8	28.85	28.75	$+0.10$	26.55	26.55	0.00
9	18.50	18.60	-0.10	17.95	17.95	0.00
10	34.25	34.20	$+0.05$	24.00	24.05	-0.05
i	F.N. l_S	F.N. l_N	F.N. Δ	A.M. l_S	A.M. l_N	A.M. Δ
1	23.75	23.65	$+0.10$	35.15	35.20	-0.05
2	28.50	28.55	-0.05	41.55	41.45	$+0.10$
3	33.95	33.95	0.00	39.95	39.90	$+0.05$
4	13.80	13.70	$+0.10$	47.35	47.40	-0.05
5	14.85	14.80	$+0.05$	51.05	51.15	-0.10
6	21.15	21.20	-0.05	63.85	63.85	0.00
7	20.15	20.25	-0.10	35.00	35.05	-0.05
8	7.85	7.75	$+0.10$	40.95	40.95	0.00
9	19.45	19.45	0.00	87.00	87.10	-0.10
10	18.05	18.00	$+0.05$	42.35	42.40	-0.05

Tabella 2.2: Dati sperimentali relativi alla capacità di interpolazione fra le tacche di alcuni studenti (identificati dalle iniziali del nome) alla prima esercitazione di laboratorio. l_S è il valore stimato interpolando a occhio e l_N il valore letto sul nonio. Tutti i valori sono in millimetri. Il valore in grassetto è sospetto: sembra più un errore di scrittura che di valutazione.

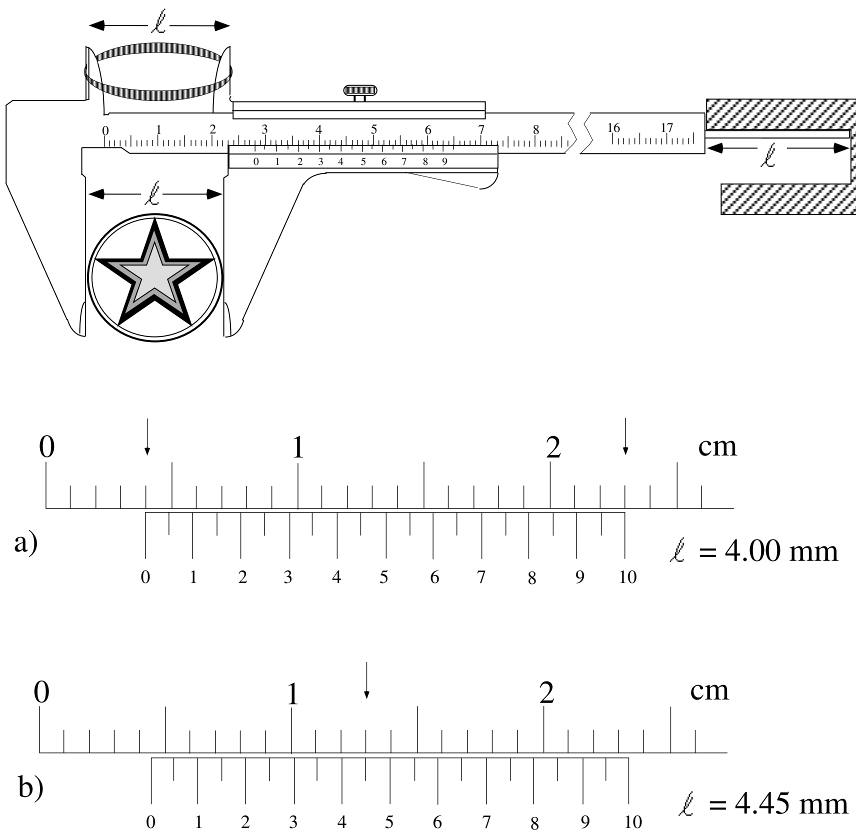

Figura 2.1: Calibro ventesimale e lettura del nonio.

preferibile che la lettura del nonio sia effettuata da un'altra persona). Le differenze fra i due numeri, opportunamente elaborate statisticamente a tempo debito, forniranno una stima quantitativa dell'incertezza di interpolazione fra le tacche. La tabella 2.2 riporta i dati sperimentali ottenuti da studenti alla loro *prima* esercitazione (in qualche modo i risultati daranno una stima pessimistica della capacità di interpolazione).

La figura 2.2 permette di valutare rapidamente la propria capacità di interpolazione senza dover eseguire l'esperienza con un calibro.

2.3 Errore ed incertezza di misura (discussione introduttiva)

L'esperienza appena descritta illustra molto bene i concetti di errore e di incertezza di misura, già introdotti a livello intuitivo e sui quali si tornerà in termini più formali nel seguito.

La differenza fra i valori stimati e quelli letti al nonio rappresenta l'errore di stima che si commette di volta in volta, in quanto il valore letto al nonio è in buona approssimazione quello "vero". Dalla conoscenza di questi *errori tipici*

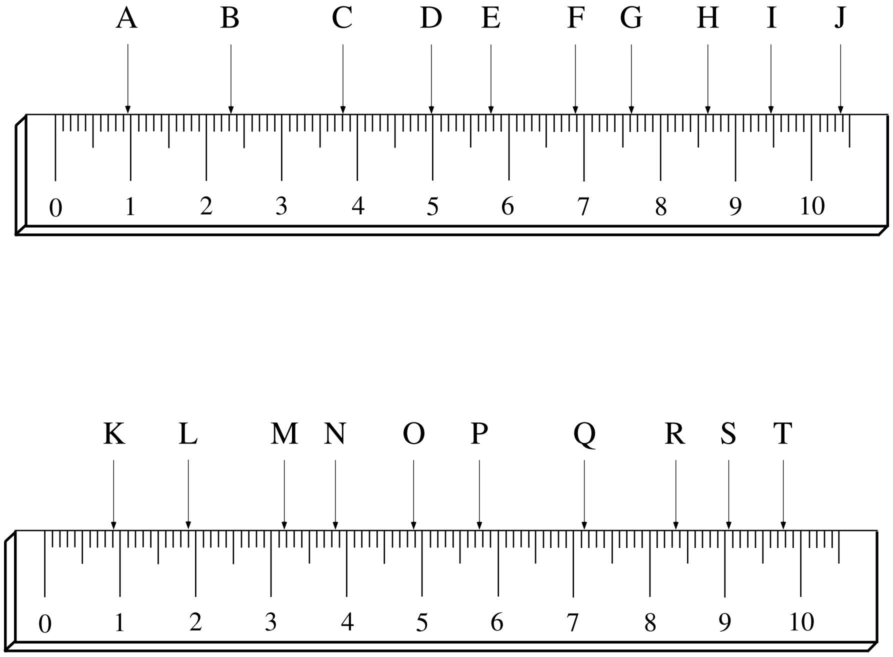

Figura 2.2: Esercizio di lettura dei decimi di divisione (vedere le soluzioni a pagina ...).

(da meglio definire) si può risalire all'incertezza della misura derivante dalla sola stima. Consideriamo alcuni casi particolarmente istruttivi.

1. Supponiamo di sapere che un'ipotetica persona A (nessuna di quelle della tabella 2.2) si sbaglia "in media" di 0.02 cm sia in più che in meno. Se in una misura succesiva effettuata in condizioni analoghe lui stimerà 23.54 cm, quanto si crederà a questa a affermazione? Ragionevolmente si tenderà a credergli "entro 0.02 cm".

2. Immaginiamo invece che un'altrettanto ipotetica persona B tenda *in media* a sovrastimare le lunghezze di 0.02 cm e che la dispersione degli errori intorno a questo errore sistematico medio sia soltanto di 0.01 cm. Se costui stimerà 15.67 cm chiaramente si tenderà a credere che il valore vero sia "entro 0.01 cm" intorno a 15.65 cm, avendo corretto la sua stima per l'errore medio.

3. Supponiamo che sei persone che si comportino come A stimino 18.33, 18.36, 18.34, 18.31, 18.34, 18.37 cm. Chiaramente, non essendoci nessun motivo per ritenere qualcuna di queste stime più o meno giusta delle altre si tenderà a pensare che "in medio stat virtus". E fin qui va bene. Però se si prova a stimare l'ampiezza dell'intervallo entro cui si può credere che il valore vero sia compreso, l'intuizione potrebbe dare una risposta ("entro 0.03 cm") ben lontana da quella corretta ("entro 0.008 cm"). Per arrivare a queste conclusioni bisognerà prima aver imparato - anticipiamo il risultato - che le stime medie effettuate indipendentemen-

Tempi di riflessi (ms): G.D.									
199	186	189	190	199	193	156	229	196	263
201	176	184	192	176	186	261	196	167	174
181	202	202	181	179	187	209	195	154	173
Tempi di riflessi (ms): D.P.									
188	233	245	195	165	241	201	185	197	206
243	194	194	187	178	187	299	208	243	199
213	199	192	200	196	219	187	176	189	193
Periodo del pendolo (s): G.D.									
1.45	1.46	1.40	1.51	1.44	1.44	1.49	1.47	1.49	1.46
1.45	1.45	1.46	1.46	1.44	1.46	1.50	1.39	1.50	1.48
1.48	1.42	1.46	1.46	1.50	1.50	1.44	1.47	1.47	1.46
Periodo del pendolo (s): D.P.									
1.57	1.44	1.45	1.45	1.41	1.41	1.46	1.47	1.42	1.43
1.40	1.40	1.48	1.46	1.40	1.48	1.46	1.43	1.47	1.43
1.48	1.48	1.45	1.47	1.47	1.46	1.41	1.44	1.41	1.45

Tabella 2.3: Due situazioni di misure manuali di tempo: tempo di riflesso e cronometraggio applicato a misure del periodo del pendolo. In queste ultime la costanza del periodo durante le due serie di misure è stata monitorata con un cronometro elettronico, il quale ha sempre indicato 1.454 s.

te da 6 persone equivalenti sbagliano tipicamente di un fattore $1/\sqrt{6}$ rispetto a quanto sbaglia la singola persona.

4. Supponiamo di porci nelle condizioni del punto 1, ma sapendo che lo strumento usato non è calibrato perfettamente. Più precisamente, lo sperimentatore sa che esso può essere scalibrato al massimo dello 0.5 % in più o in meno. Ciò significa che, anche se la lettura fosse perfetta, si crederà al risultato entro 0.12 cm. Quindi l'incertezza dovuta all'errore di lettura di 0.02 mm diventa trascurabile. È anche chiaro, se si è interessati ad una sola grandezza, non ha alcun senso fare medie fra più letture o sforzarsi di leggere al meglio, essendo consapevoli di una "inevitabile"incertezza di questo tipo. In particolare, se questa incertezza non è tollerabile ai fini che l'esperimento si propone, bisogna ricalibrare lo strumento o procurarsene un altro migliore.

5. Continuando sul punto precedente, è altrettanto ovvio che, se misuriamo due grandezze con lo stesso strumento e siamo interessati al loro rapporto, l'errore ignoto di scala dello strumento non avrà alcuna influenza sul risultato. Queste riflessioni insegnano che gli "errori sistematici" andranno trattati con cautela, per non sovrastimare o sottostimare sostanzialmente l'incertezza sul risultato finale.

2.4 ↻ Tempo di reazione e misure di cronometraggio

Semplici esperimenti complementari a quelli descritti nel paragrafo precedente permettono di testare la propria capacità dieffettuare misure di tempo. Queste

	serie 1	serie 2	serie 3	serie 4	serie 5	serie 6
z	t	t	t	t	t	t
(cm)	(s)	(s)	(s)	(s)	(s)	(s)
0	0.	0	0	0.	0	0
510	1.556013	1.630131	1.611060	1.621914	1.593325	1.632937
15	1.976810	2.053967	2.035277	2.046783	2.016730	2.057339
20	2.329293	2.408044	2.389096	2.400443	2.369708	2.411299
25	2.644171	2.724131	2.704712	2.716117	2.684988	2.726711
30	2.927985	3.008754	2.989123	3.000911	2.969174	3.011155
35	3.194652	3.276158	3.256609	3.268559	3.236645	3.279085
40	3.437351	3.519370	3.499438	3.511345	3.479116	3.521723
45	3.665107	3.747497	3.727759	3.739757	3.707301	3.750176
50	3.883512	3.966447	3.946807	3.958853	3.926480	3.969745
55	4.091030	4.174128	4.154152	4.166161	4.133710	4.177082
60	4.289556	4.373025	4.353041	4.365218	4.332379	4.375825
65	4.482196	4.565777	4.545679	4.557905	4.525146	4.568780
70	4.663909	4.747730	4.727502	4.739650	4.706941	4.750554
75	4.842332	4.926470	4.906120	4.918255	4.885343	4.928937
80	5.011102	5.095287	5.074691	5.086945	5.054073	5.097858
85	5.181138	5.265103	5.244506	5.256917	5.223947	5.268059
90	5.344091	5.427422	5.406430	5.418815	5.385940	5.430183
95	5.507233	5.588757	5.567767	5.580020	5.547225	5.591583

Tabella 2.4: Dati sperimentali della caduta di un grave connesso ad un volano. I valori di tempo sono letti mediante fotocellule.

misure permettono di valutare le incertezze su misure eseguite in condizioni analoghe.

Il primo esperimento consiste nel misurare il tempo di reazione fra la comparsa di un segnale luminoso sul monitor di un computer e l'istante in cui viene premuto un tasto. Il secondo consiste nel misurare molte volte il periodo di un pendolo che effettua piccole oscillazioni. La costanza del periodo viene controllata con un sistema elettronico. I dati sono riportati in tabelle 2.3 e si riferiscono a due persone.

Chiaramente le situazioni di tempo di reazione è diversa da quella di cronometraggio. Mentre nella prima c'e' un ritardo medio rispetto allo stimolo con fluttuazioni che a volte possono diventare anche grandi a causa di piccole distrazioni, nel secondo caso si hanno fluttuazioni positive e negative rispetto al tempo "vero". Inoltre nel secondo caso si combina una fluttuazione sullo *start* e una sullo *stop*.

2.5 ⟳ Moto uniformemente accelerato

Un metodo per ottenere un moto uniformemente accelerato con accelerazione minore di g, e alternativo al classico piano inclinato, consiste nel sospendere un oggetto ad un filo avvolto in un rocchetto coassiale e solidale ad un disco (volano). La maggiore inerzia del sistema riduce l'accelerazione con cui l'oggetto scende e facilita lo studio del moto. Mentre il corpo scende, un computer registra i tempi di passaggio su traguardi costituiti da fotocellule. I traguardi

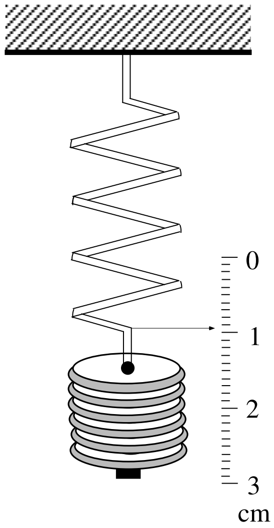

Figura 2.3: Molla.

sono equidistanti 5 cm "nominali" (diciamo al decimo di millimetro). Il peso viene rilasciato, a mano, da una posizione leggermente sopra la prima fotocellula e quindi la posizione di rilascio può differire leggermente da una serie di misure all'altra.

La tabella 2.4 riporta i dati sperimentali così come essi sono stampati dal computer. Essi verranno analizzati nel seguito. Poniamoci, come al solito, alcune domande:

- Ha senso misurare il milionesimo di secondo (1 μs) in questa esperienza?
- Come si fa a stabilire il numero di cifre da prendere "sul serio" nel corso dell'analisi?
- È possibile "dimostrare" che il moto è uniformemente accelerato?
- Come combinare le informazioni delle sei serie di misure per migliorare l'informazione sull'accelerazione? Può essere una buona idea mediare i valori dei tempi a ciascun traguardo e utilizzare questi valori nell'analisi?

2.6 ↻ Allungamento e periodo di oscillazione di una molla

Si tratta di un'altra semplice esperienza che non richiede conoscenze di fisica di livello universitario. Scopo dell'esperienza è la verifica dell'allungamento lineare di una molla in funzione della forza applicata e della dipendenza dalla massa del periodo di oscillazione del sistema. Come risultato dell'esperienza si otterrà una misura dell'accelerazione di gravità.

		Prima serie		Seconda serie		Terza serie	
n	M (g)	l (mm)	$T \times 10$ (s)	l (mm)	$T \times 10$ (s)	l (mm)	$T \times 10$ (s)
0	63	0	-	0	-	0	-
1	142	0	-	0	-	0	-
2	221	0	-	0	-	0	-
3	300	14	5.01	16	5.09	16	5.19
4	379	32	5.57	33	5.66	33	5.68
5	458	49	6.24	51	6.27	51	6.34
6	536	66	6.78	68	6.82	69	6.94
7	615	85	7.28	86	7.33	87	7.28
8	694	103	7.79	103	7.81	103	7.86
9	773	119	8.13	121	8.31	121	8.24
10	852	137	8.63	139	8.77	139	8.70

Tabella 2.5: Dati dell'esperienza della molla. n è il numero dei dischetti sospesi, l è l'allungamento ($l_\circ = 0$) e T il periodo di oscillazione. Il peso di 10 dischetti è pari a 789 g e la massa della molla stessa (parte oscillante) è pari a circa 63 g.

2.6.1 Breve richiamo di fisica generale

Il caso di una molla di massa trascurabile posta con l'asse disposto verticalmente e bloccata in alto, a cui viene applicata una massa, è un classico problema di fisica generale: la coordinata l dell'estremo della molla, in condizione di equilibrio, è data dalla relazione

$$l = l_\circ + \frac{gM}{k} = l_\circ + \frac{g}{k}M\,, \tag{2.1}$$

dove $l_\circ$ è la posizione in assenza di massa applicata, k è la costante della molla, M è la massa dell'oggetto e g l'accelerazione di gravità. Se l'estremo della molla è spostato dalla posizione di equilibrio esso oscilla con periodo

$$T = 2\pi\sqrt{\frac{M}{k}} = \frac{2\pi}{\sqrt{k}}\sqrt{M}\,. \tag{2.2}$$

Effettuando quindi una serie di misure del periodo in funzione della massa applicata si può ricavare la costante della molla. Questa può essere sostituita nella (2.1) per ottenere, da misure dell'allungamento in funzione della massa, il valore di g.

2.6.2 Misure

La tabella 2.5 contiene i risultati delle misure di allungamento e periodo in funzione della massa applicata effettuate da un gruppo di studenti.

Per comodità nei dati della tabella è stato sottratto il valore $l_\circ$, il quale è stato misurato essere lo stesso nelle tre serie.

I diversi valori di massa sospesa sono ottenuti mediante un numero n di dischetti metallici *uguali* fra di loro. I 10 dischetti a disposizione sono pesati tutti insieme ed hanno una massa totale di 789 g. Anche la massa della molla, pari a 63 g, è stata considerata nella valutazione di M. Dal punto di vista fisico sicuramente questo non è corretto in quanto i 63 g sono distribuiti lungo la molla e non concentrati all'estremità. Questa osservazione è rilevante per il periodo di oscillazione. Può essere istruttivo fingersi nell'incapacità di stimare il contributo esatto della massa della molla all'inerzia del sistema nelle oscillazioni e di procedere in modo empirico. Quindi, come detto, per ora assumiamo che la massa della molla influenzi l'inerzia del sistema come se essa fosse tutta concentrata nell'estremo[1].

Poiché il periodo di oscillazione è inferiore al secondo si è preferito eseguire la misura su 10 oscillazioni. Questo "trucco", simile a quello adottato nel misurare tutti insieme i 10 dischetti, permette di ridurre l'incertezza di misura. Infatti, innanzitutto questo equivale a mediare su 10 oscillazioni e quindi piccoli effetti spuri si possono compensare. Secondo, se si compie un piccolo errore di cronometraggio, questo è importante su tempi brevi ma trascurabile su tempi lunghi.

È da notare che le misure di lunghezza sono effettuate leggendo soltanto fino al millimetro. Secondo gli studenti che hanno eseguito le misure, la strumentazione non era tale da consentire agevolmente la lettura dei decimi. Anche rispettando la loro opinione, derivata dall'esperienza diretta, si potrebbe obiettare che in alcuni casi, a parità di massa applicata, alcuni valori di allungamento si ripetono esattamente. Sorge quindi la curiosità di conoscere la cifra successiva. Vedremo nel seguito come stabilire in modo quantitativo se in questo caso i decimi di millimetro sarebbero stati significativi o no. Per ora possiamo tranquillizzare il lettore affermando che, ai fini del risultato finale, aver arrotondato il risultato al millimetro non produce un grosso effetto.

Osservando le prime misure della tabella 2.5, effettuate per 1 o 2 dischetti ci si rende conto che qualcosa non va con le formule (2.1) e (2.2). Infatti fino a oltre 200 g la molla non si sposta dalla posizione iniziale: a differenza delle (2.1) e (2.2), ricavate da una molla ideale, l'esperimento è stato effettuato, necessariamente, con una *molla reale*. Si può pensare che tali relazioni siano valide soltanto al di sopra di una certa *massa critica* oltre la quale l'allungamento della molla in funzione della forza applicata è *lineare* (vedi (2.1)) e il periodo va effettivamente come la radice quadrata della massa (vedi (2.2)). Supponiamo che questo sia vero per masse intermedie, ad esempio per un numero di dischetti intorno a 6 (per masse più elevate ci sono da temere possibili effetti di saturazione; si ragioni "per limite" e si immagini di applicare alla molla una tonnellata ...).

2.6.3 Prime valutazioni di k e di g

Cerchiamo di valutare k e g dalla differenza di allungamento dovuta all'aggiunta del sesto dischetto. Prendendo i dati della prima serie di misure, dalla

[1] La massa equivalente ai fini dell'inerzia del sistema e localizzata all'estremità della molla è pari a un terzo della massa della molla.

$n_\circ \to n_1$	Δl (mm)	Prima serie k (N/m)	g (m/s^2)	Seconda serie k (N/m)	g (m/s^2)	Terza serie k (N/m)	g (m/s^2)
$5 \to 6$	17	46.0	9.9	45.5	9.8	43.9	9.5
$6 \to 7$	19	45.8	11.0	45.2	10.9	45.2	10.9
$7 \to 8$	18	45.1	10.3	44.9	10.3	44.3	10.1

Tabella 2.6: Valori della costante elastica della molla (k) e dell'accelerazione di gravità ottenuti da un'analisi parziale della tabella 2.5.

(2.2) si ottiene[2]

$$\begin{aligned} k = \frac{4\pi^2 M}{T^2} &= \frac{39.5 \times 536\,\mathrm{g}}{0.678^2\ \mathrm{s}^2} \\ &= 4.60 \cdot 10^4\ \mathrm{g/s}^2 \\ &= 46.0\,\mathrm{N/m}\,. \end{aligned} \tag{2.3}$$

Poiché l'allungamento della molla causato dall'aggiunta del sesto dischetto ($\Delta M = 78.9\,\mathrm{g}$) è stato di $\Delta l = 17$ mm = 0.017 m si ottiene

$$\begin{aligned} g = \frac{k\Delta l}{\Delta M} &= \frac{46.0 \times 0.017\,\mathrm{Nm}^{-1}\mathrm{m}}{0.0789\,\mathrm{kg}} \\ &= 9.9\,\mathrm{m/s}^2\,. \end{aligned} \tag{2.4}$$

È da notare come il valore di 9.9 m/s^2 potrebbe indurre a credere che g sia stata determinata con un incertezza di qualche parte per cento. In realtà ai fini dell'esatta valutazione di g è determinante il contributo di Δl. Infatti non ci sarebbe da meravigliarsi se esso venisse in altre misure 0.016 o 0.018 cm, o anche se deviasse un po' di più da 0.017 cm. Quindi Δl è determinato a circa 1 parte su 17, ovvero al 6 %. Questo si rifletterebbe su g con una incertezza dell'ordine di 0.6 m/s^2. Pertanto l'ottimo accordo con il "valore vero" di 9.8 m/s^2 è da ritenersi soltanto un caso fortunato.

Possiamo ripetere lo stesso esercizio per le tre serie e, considerando anche le variazioni dal sesto al settimo e dal settimo all'ottavo dischetto, riportare i risultati in tabella 2.6. I risultati mostrano valori di g confrontabili con il valore atteso di 9.80 m/s^2. La differenza fra ciascuno dei valori misurati e il valore "vero" di g ci dà un'idea dell'*errore* commesso nella misura. Esso varia fra -0.3 e $+1.2$ m/s^2. Abbiamo ottenuto variazioni comprese fra -3 e $+12$ % rispetto al valore vero, ovvero *errori percentuali* (in valore assoluto) *dell'ordine* del $\approx$ 5-10 %.

Purtroppo non è usuale nella ricerca sapere già il risultato della misura, in quanto non è interessante perdere tempo a misurare una grandezza nota con esattezza, se non per ragioni didattiche o per per *calibrare* l'apparato sperimentale e il *metodo di misura*. In genere prima della misura il risultato di interesse è sempre ignoto o incerto (si crede che sia più in un certo intervallo che altrove). È quindi importante essere in grado di stimare l'ordine di grandezza dell'errore (ignoto!) che può essere stato commesso senza conoscere il

[2] Si noti come il valore di $4\pi^2 = 39.5$ è riportato con 3 cifre in quanto ulteriori cifre sarebbero ininfluenti alla fine del risultato finale, come sarà chiarito nel paragrafo 3.4.

valore vero della grandezza da misurare. Questo è uno degli scopi di questo corso e l'argomento verrà trattato ampiamente nel seguito.

Dai dati di tabella 2.5 e dai risultati ottenuti in tabella 2.6 sorgono spontanee alcune domande

1. I risultati sono stati ottenuti dall'ipotesi di linearità dell'allungamento dalla massa applicata e dalla formula 2.2. La relativa stabilità dei valori ottenuti e l'accordo qualitativo con il valore vero di g indicano che tali ipotesi sono ragionevoli. È possibile stabilire dei criteri per verificare tali ipotesi? È possibile analizzare soltanto i dati che meglio soddisfano tali ipotesi al fine di diminuire gli errori?

2. Come si possono combinare insieme tutte le informazioni di una serie di misure per ottenere un valore di k e di g? Si sarebbe tentati di *mediare* in qualche modo i valori di g ottenuti. Bisogna però prestare attenzione al fatto che, con il metodo utilizzato sopra, ogni misura di allungamento influenza due valori di g. Per esempio, se al posto di $l = 66\,\text{mm}$ per $n = 6$ si fosse misurato 67 mm (non improbabile) i due valori di g corrispondenti della tabella 2.6 sarebbero diventati rispettivamente 10.5 e 10.4 m/s^2.

Come al solito, prima di poter dare una risposta a tale domande occorrerà aver acquisito delle conoscenze teoriche di probabilità e di statistica.

2.7 ❄ Potere di aspirazione di una pompa da vuoto

Questa esperienza è molto istruttiva ai fini dell'apprendimento dell'analisi dei dati. Viene introdotta in questo capitolo e l'analisi grafica viene lasciata come esercizio sulle linearizzazioni. Successivamente verrà ripresa in dettaglio quando si parlerà dei *fit*.

Consideriamo una pompa da vuoto connessa ad un recipiente e messa in funzione a partire da un istante $t = 0$. Essa produrrà una depressione nella *linea del vuoto* (definita come l'insieme della pompa, recipienti, tubi, raccordi, rubinetti e manometro opportunamente connessi fra di loro). Dall'istante in cui il circuito è chiuso *a tenuta* l'andamento della depressione D in funzione del tempo è di tipo esponenziale

$$D = D_F(1 - e^{-t/\tau}), \tag{2.5}$$

caratterizzato da una *costante di tempo* τ che dipende dal volume totale della linea di vuoto e dalla potenza della pompa, quantificata nella cosiddetta *velocità di aspirazione*, definita da

$$S_p = \frac{V_L}{\tau}\,. \tag{2.6}$$

È possibile ottenere S_p dalla misura della costante di tempo per un dato valore di V_L. Ma il volume totale della linea contiene un contributo costante - ma di

	$V = 137.3\,\mathrm{ml}$		$V = 212.5\,\mathrm{ml}$		$V = 509.3\,\mathrm{ml}$	
i	t (s)	ΔP (cm Hg)	t (s)	ΔP (cm Hg)	t (s)	ΔP (cm Hg)
1	2.0	45.0	4.0	45	6.0	50.8
2	4.0	34.5	8.0	31	12.0	39.3
3	6.0	26.5	12.0	22	18.0	31.8
4	8.0	22.5	16.0	16	24.0	25.3
5	10.0	17.0	20.0	13	30.0	20.8

Tabella 2.7: Dati sperimentali dell'andamento della pressione all'interno di una linea da vuoto in funzione del tempo e del volume del recipiente V. Il volume totale della linea contiene anche un contributo costante $V_\circ$ dovuto alla pompa stessa e ai raccordi. Per semplicità le misure dirette di P e P_F sono presentate come differenza $\Delta P = P - P_F$. Si noti come nelle misure per $V = 212.5\,\mathrm{ml}$ non è stata eseguita la lettura del decimo e quindi tale informazione non è recuperabile. Ciò può essere dovuto al fatto che queste misure sono state effettuate per prime e soltanto successivamente gli studenti hanno deciso di poter apprezzare i decimi. Sarà interessante vedere al momento opportuno, mediante l'analisi dei residui, se i decimi sono significativi.

difficile misura - dovuto alla pompa stessa, raccordi e manometro, che indicheremo con $V_\circ$. È quindi opportuno effettuare misure di τ per diversi volumi V del recipiente e dalla relazione

$$S_p = \frac{1}{\tau}(V_\circ + V)\,, \tag{2.7}$$

ovvero

$$\tau = \frac{1}{S_p}V + \frac{V_\circ}{S_p}\,, \tag{2.8}$$

ottenere S_p e $V_\circ$.

Ai fini delle misure, effettuate con un normale manometro, e dell'analisi è preferibile esprimere la (2.5) come:

$$P - P_F = (P_\circ - P_F)e^{-t/\tau}, \tag{2.9}$$

o

$$\Delta P = \Delta P_\circ e^{-t/\tau}\,. \tag{2.10}$$

dove P è la pressione misurata all'istante t, $P_\circ$ è la pressione all'istante $t = 0$ e P_F è la pressione raggiunta asintoticamente.

I dati sperimentali degli andamenti di P in funzione di t per tre volumi sono riportati in tabella 2.7. Essi sono dati reali raccolti da un gruppo di studenti.

2.8 Continua ...

Le esperienze presentate in questo capitolo costituiscono il filo conduttore del corso e su di esse verranno applicate le tecniche di analisi che incontreremo nei prossimi capitoli. Per comodità riportiamo i paragrafi in cui compare il proseguimento delle analisi (per completezza sono state elencate anche altre esperienze che verranno presentate nel seguito ... o per niente in questa versione del testo).

Contatore: 4.2; 4.4; 5.3; 5.6; 5.9; 6.7.2;

Pallinometro: 4.4;

Foglio A4:

Moto grave: 3.4

Molla: 6.6; 6.7.4

Pompa da vuoto:

Interpolazione fra le tacche:

Riflessi:

Scarica condensatore: 6.7.2;

Capitolo 3

Logbook e relazione

3.1 Documentazione del lavoro sperimentale

Lo studente che affronta per la prima volta il tema della sperimentazione è stato precedentemente esposto a questioni scientifiche attraverso la lettura di libri di testo, di voci di enciclopedia o di articoli e libri divulgativi. Queste fonti di documentazione rappresentano gli ultimi passi del processo di riduzione di dati, spesso effettuata su scala planetaria, mettendo insieme informazioni dei laboratori di tutto il mondo.

Passiamo in rassegna i diversi livelli di documentazione di un lavoro scientifico:

Quaderno di laboratorio: rappresenta il diario di bordo dello sperimentatore (in inglese *logbook*). In esso sono riportati i dati originali ed eventuali informazioni che potranno servire alla loro elaborazione. È concepito ad uso dello sperimentatore o del gruppo che esegue le misure. Nel seguito daremo alcuni consigli sul modo di redigerlo, al fine di trarre il massimo profitto, in fase di analisi, delle informazioni in esso contenute.

Nota tecnica: è indirizzata ai colleghi di un programma di ricerca. Essa è comunque utile anche quando si lavori in piccoli gruppi e addirittura alla stessa persona che ha eseguito le misure. Infatti, scrivendo ci si rende conto di salti logici, controlli non effettuati ed altre incongruenze. A questo punto un logbook ben redatto è di vitale importanza per non dover ritornare in laboratorio a ripetere misure, o addirittura a dover ricominciare l'intero lavoro in caso di forti dubbi su alcune condizioni di lavoro non considerate o delle quali si è persa memoria. Nella nota tecnica si cerca di descrivere il più fedelmente possibile l'apparato sperimentale e le procedure eseguite. In pratica tale documentazione deve essere in grado di fornire a noi stessi e ai colleghi le informazioni sufficienti per poter ripetere le misure nelle stesse condizioni.

Articolo scientifico: rappresenta il punto conclusivo di un esperiemento o di una serie di esperimenti. In gergo è chiamato anche "il lavoro" ed indica appunto che esso rappresenta il risultato di un lavoro di ricerca ed è infatti utilizzato come uno dei criteri di valutazione delle capacità di un

ricercatore ("quello che non è pubblicato non è stato fatto", usa ripetere un collega tedesco; "publish or perish" dicono i ricercatori americani facendo riferimento alla politica di finanziamento scientifico).
L'articolo viene pubblicato su una rivista scientifica (tipicamente a diffusione internazionale e quindi in inglese) in base al consenso di un "referee" (arbitro) anonimo che giudica la validità del contenuto al fine dell'aumento di conoscenza della comunità scientifica su quel determinato argomento. I risultati sono presentati in forma sintetica, inquadrati nella linea di ricerca del settore e comprensibili agli specialisti del settore, indipendentemente dalla loro conoscenza del dettaglio della strumentazione e delle procedure utilizzate. Deve contenere invece una descrizione accurata dell'*oggetto della misura* (a meno che esso non sia ben noto alla comunità scientifica) in modo tale da permettere ad altri sperimentatori di riprodurre il risultato. (La differenza fra *ripetitività* e *riproducibilità* verrà discussa nel paragrafo 11.2.1).

Articolo di rassegna: presenta in un quadro unitario risultati sperimentali e previsioni teoriche di un certo settore, con l'intento di metterne in risalto l'eventuale accordo o disaccordo e di trarre delle conclusioni su quello che la comunità scientifica ha imparato dopo anni di ricerca, citando i contributi dei vari ricercatori che hanno contribuito.

Monografia: a questo punto la distinzione fra cosa sia un articolo di rassegna, cosa una monografia e cosa un libro specialistico diventa sottile.

Libro di testo: è pensato per essere usato da studenti e quindi dovrebbe essere in grado di coprire sufficientemente gli argomenti svolti in un intero corso o in una parte di esso (in realtà poi nessun docente è completamente soddisfatto di un unico testo, e gli studenti si lamentano che i libri costano . . .).
In un libro di testo, anche di argomenti specialistici, il riconoscimento del contributo dei vari ricercatori diventa inessenziale (con eccezioni di casi notevoli) e si dà la priorità all'inquadramento generale della materia e alle esigenze didattiche. Spesso la coerenza logica del libro di testo è anteposta al reale processo storico - a volte con risultati al limite della mistificazione - e induce gli studenti a farsi delle idee "strane" su cosa succede nell'ambiente della ricerca. Non c'è migliore antidoto contro tale pericolo che la frequenza del laboratorio.

3.2 Redazione del quaderno di laboratorio

Durante lo svolgimento di un esperimento è utile documentare il lavoro svolto. Per mantenere un buon quaderno di laboratorio è importante partire da alcune considerazioni.

- Esso rappresenta la sorgente primaria dei dati sperimentali. Va compilato durante le misure e deve poter essere riletto a distanza di tempo dalle stesse persone che partecipano all'esperimento; le annotazioni devono essere quindi chiare ed essenziali, cioè eliminando il superfluo (ma all'inizio non è facile distingere cosa lo sia e cosa non).

- Va compilato prima, durante e dopo la raccolta di dati. L'idea di "prima lavorare e poi scrivere" in genere non funziona. Tantomeno è impensabile di compilarlo la sera (nella vita professionale in genere non ci si porta il logbook a casa) o il giorno dopo.

- Nel seguito daremo dei consigli su come gestire il logbook, assumendo che esso possa contenere tutte le informazioni concernenti l'esperimento. Questo è abbastanza vero per le esperienze didattiche. Nel mondo della ricerca invece gli esperimenti tendono ad essere complessi e i calcolatori svolgono un ruolo essenziale nell'acquisizione, archiviazione e trattamento dei dati. Il logbook esiste ancora, ma esso rappresenta soltanto la parte cartacea di una enorme massa di informazione dislocata su dischi di computer ed altri supporti di dati. Esso contiene allora soltanto informazioni molto generali, un numero progressivo che identifica l'insieme di misure effettuate e qualche annotazione dello sperimentatore.

Veniamo ora ad alcune raccomandazioni per una buona conduzione del logbook.

- Scegliere un **quaderno** tradizionale, **a quadretti** e possibilmente a fogli uniti, di **formato DIN A4** (quello delle normali fotocopie che circolano in Europa) per poterne fotocopiare, all'occorrenza, delle parti nel modo più efficiente ed economico.

- **Numerare le pagine**, in modo tale da poter fare dei rimandi.

- Prima di cominciare le misure annotare immediatamente **data**, ora e nome dei **partecipanti** all'esperienza; riportare di nuovo l'ora prima di ogni nuova serie di misure.

- Indicare il **titolo** dell'esperienza (o, in generale, descrivere brevemente cosa si vuol misurare).

- Elencare il **materiale** a disposizione, eventualmente cercando di identificare, attraverso il numero di matricola o in qualche altro modo, gli strumenti utilizzati. Questa accortezza può essere utile sia per evitare di confondere due strumenti dello stesso tipo (ma che possono differire leggerement in calibrazione) che nell'eventualità che si debba tornare in laboratorio per completare una esperienza.

- Aggiungere eventualmente uno **schizzo** dell'apparato con sopra i vari simboli utilizzati. Pur senza essere abili come Leonardo, spesso anche un disegno non necessariamente artistico, ma con dei simboli inequivocabili, può essere più utile di tante parole.

- Riportare i **dati originali direttamente** sul quaderno a penna (ovvero **in modo indelebile**) e mai su foglietti volanti. Naturalmente all'inizio non è facile convincere gli studenti, abituati al concetto di "brutta" e "bella", seguire questa regola. I motivi di quella che alla prima esercitazione può sembrare una tortura sono:

Grandezza misurata	Modello			
	A	B	C	D
Condizioni della prova				
Massa vettura provata (kg)	1337	1139	1087	-
Temperatura (°C)	26	23	17	19
Pressione atmosferica (mb)	1022	1030	1033	1017
Umidità relativa (%)	52	38	78	48
Vento (m/s)	0.7-1.8	0-0.3	≈ 0	1.2-1.7
Densità carburante (g/l)	750	746	742	740
Prestazioni				
Velocità massima (km/h)	186.211	187.325	166.363	198.269
(dichiarata)	(187)	(188)	(170)	(198)
Consumo a 90 km/h (km/l)	15.4	20.1	17.8	15.1
(dichiarato)	(16.4)	(20.8)	(20.8)	(15.8)
Consumo a 120 km/h (km/l)	12.3	13.7	13.4	11.6
(dichiarato)	(12.2)	(16.1)	(14.7)	(12.7)

Tabella 3.1: Dati di prove di autovetture (da “Quattroruote”, Novembre 1995).

  - trascrivere dalle minute alla bella copia è una perdita di **tempo**; è preferibile dedicare il tempo a disposizione per altre misure, per verifiche o per cominciare ad elaborare i dati;
  - si possono commettere **errori** di trascrizione;
  - si rischia di essere **selettivi**, ovvero di eliminare dei valori che sembrano inutili o “sbagliati”.

- L’unico uso consentito - e anche raccomandato - dei **foglietti** volanti è per eseguire dei conti ripetitivi (ad esempio sui dati di una tabella). È però importante che sia riportata sul quaderno la formula utilizzata, eventualmente accompagnata da un esempio numerico, per essere sicuri dell’algoritmo utilizzato.

- Anche i **grafici** vanno incollati sul quaderno. È raccomandabile aggiungere delle didascalie per ricordare da quali dati ogni grafico è stato ricavato. È anche opportuno mettere vicino alla tabella un rimando alla pagina dove è riportato il grafico, al fine di facilitare la rilettura del logbook.

- Registrare i valori delle grandezze che possono influenzare i risultati delle misure (ad esempio temperatura, umidità, etc.). La scelta di quali tali **fattori di influenza** registrare dipende dalla conoscenza che lo sperimentatore ha delle grandezze da misurare. Ad esempio la tabella 3.1 mostra le prestazioni di alcune vetture secondo i test effettuati da un mensile. Confidando nell’esperienza dei tecnici si può evincere che i dati meteorologici riportati abbiano un qualche effetto sulle prestazioni delle vetture (come esercizio si provi a pensare perché essi possono in-

t (s)	x (cm)	x^2 (cm^2)	. . . spazio per altre elaborazioni . . .
...	...	...	

Tabella 3.2: Esempio di preparazione di una tabella del quaderno di laboratorio

fluenzare velocità massima e consumi). *** Dire da qualche parte che sarebbe meglio riscalare i valori ***

- Preparare subito delle **tabelle** su cui riportare i valori letti e le successive elaborazioni previste. Lasciare dello spazio per ulteriori colonne che potessero aggiungersi, come mostrato nell'esempio di tabella 3.2.

- Se in alcune colonne i valori sono tutti uguali è preferibile non riportarli tutti in tabella. In questo modo ci si rende conto subito che un valore rimarrà costante durante le misure (ad esempio la pressione atmosferica in un esperimento di breve durata) esso può essere indicato a parte. Anche se era già stata preparata una colonna si può indicare mediante opportuni simboli che i valori che seguono sono tutti uguali (altrimenti per arrivare alla stessa conclusione si dovrà scorrere, in fase di rilettura, l'intera colonna).

- Tutti i valori letti vanno **registrati direttamente** senza elaborazioni intermedie (cambiamenti di scala, sottrazioni di tara, etc.). Qualora sia possibile, è buona norma leggere il valore, riportarlo sul quaderno e ricontrollarlo sullo strumento.

- Se ci si accorge di aver commesso degli errori, sia nello scrivere un numero che un'intera serie di misure, è preferibile **non cancellare** o bianchettare i valori "sbagliati"; fare semplicemente delle barre su di essi in modo che essi risultino ancora leggibili. Non è raro infatti il caso di accorgersi in seguito che quello che si riteneva sbagliato era in verità giusto o che, comunque, questa potrebbe fornire delle indicazioni su un malfunzionamento dell'apparato o su un cattivo modo di procedere dello sperimentatore.

- Se proprio si rendesse necessario eseguire **lavori in parallelo** fra diversi membri del gruppo è conveniente utilizzare fogli che poi andranno incollati sul quaderno; questi devono essere considerati parte integrante del logbook e non semplici foglietti volanti.

- Nel caso di esercitazioni didattiche il quaderno di laboratorio contiene anche le successive elaborazioni. Esse devono ugualmente contenere informazioni sufficienti a facilitare la loro **rilettura** e la rielaborazione (in caso di dubbi): vi devono essere le formule utilizzate (ma non tutti i conti) e così via.

- Non è raccomandabile dilungarsi in **introduzioni teoriche** sull'esperimento, a meno che questo non sia esplicitamente richiesto dal docente. Infatti esse, oltre che non aiutano a capire l'esperimento, sono un lavoro inutile e "rischioso" (in quanto si possono commettere errori concettuali che poi il docente non può far finta di ignorare ...).

- Alcune elaborazioni possono protrarsi per qualche settimana, accavallandosi alle successive esperienze di laboratorio. A volte può anche succedere che sia richiesto che una esperienza debba essere rielaborata, o ulteriormente analizzata, alla luce di nuove conoscenze teoriche, anche a distanza di mesi. È quindi preferibile riportare esperienze ed elaborazioni in **ordine cronologico**, senza lasciare pagine bianche fra una esercitazione e l'altra. Avendo numerato le pagine è facile fare dei rimandi in avanti e indietro.

3.3 Stesura della relazione

La relazione dell'esperienza di laboratorio è un elaborato con scopo prettamente didattico e per questo non era stata considerata fra la documentazione del lavoro sperimentale. Essa rappresenta, in principio, il passo di elaborazione dati successivo quaderno al di laboratorio. Ma rispetto a questo è più difficile dare consigli generali sulla sua redazione in quanto dipende dalle esigenze del docente. Per esempio qualcuno preferisce non separare il quaderno di laboratorio dalla relazione e chiamare relazione un qualcosa di intermedio fra i due.

Volendo dare delle indicazioni si può dire, in linea di massima, che:

- in genere la relazione comincia con un'introduzione concisa sullo scopo della misura e illustra qualche aspetto teorico utile per la comprensione del lavoro;

- si cerca di descrivere brevemente il metodo di misura utilizzato e la successione delle misure e delle procedure di analisi utilizzate;

- qualora si ritenga necessario viene anche spiegata qualche tecnica di analisi non banale che ha giocato un ruolo importante nella comprensione dei dati;

- le misure sono presentate in ordine logico anziché cronologico; ad esempio se le misure del raggio di un volano o della massa di palline è stata effettuata con calma alla fine delle misure (o addirittura dopo giorni se si era dimenticato di farlo) non per questo va presentata alla fine della relazione;

- non c'è bisogno di riportare tutte le informazioni e i dati che non sono stati realmente utilizzati per conseguire i risultati;

- i risultati, sia finali che intermedi, vanno ben evidenziati;

- è raccomandata una discussione dei risultati raggiunti, confrontati con quanto ci si attendeva da previsioni teoriche che possono essere "assolute" (ad esempio la misura dell'accelerazione di gravità g o il calore specifico di un metallo) oppure dipendenti dalla strumentazione utilizzata (ad esempio il momento di inerzia di un volano o la frequenza di risonanza di un circuito RLC).

La relazione va redatta subito dopo che sono terminate le elaborazioni dei dati, ma è di grande importanza didattica provare a scriverla dopo alcune settimane (o mesi). Ci si accorgerà di quanto non sono ovvie molte delle "ovvietà" non riportate sul logbook e se ne ricaverà un insegnamento per le esperienze successive.

3.4 Cifre significative

Un tipico problema legato alle prime esperienze è quello del numero di cifre da considerare per il valore letto sullo strumento e per le successive elaborazioni. L'argomento sarà trattato più a fondo nel seguito, ma si può avere un'idea del problema dando uno sguardo ai dati della tabella 3.1. Si provi ad esempio a rispondere - all'occorrenza in modo intuitivo - alle seguenti domande:

- Supponiamo che la velocità massima della vettura A sia stata misurata dal tempo necessario affinché la vettura passi fra due traguardi posti alla distanza di un metro e che il posizionamento dei traguardi sia "perfetto". Quali valore di velocità si sarebbe ottenuto se i traguardi fossero stati posizionati per errore ad una distanza di 0.5 mm maggiore (o minore) di quella nominale?

- Ammesso che la misura di velocità sia "perfetta", è importante per il potenziale acquirente sapere che la vettura A raggiunge 186.211 km/h, invece di 186.210 o di 186.212?

- Se i tecnici del mensile avessero avessero avuto a disposizione dieci vetture dello stesso modello, è ragionevole pensare che essi avrebbero misurato sempre una velocità massima di 186.211 km/h?

- In un caso (C) la velocità massima accertata è quasi 4 km/h inferiore a quella dichiarata dalla casa: è indice di scarsa serietà del costruttore o del fatto che sia stata provata una vettura "poco riuscita"? Oppure dipenderà dal fatto che essa è stata provata in una giornata più fredda, più umida e con più alta pressione delle altre?

Anche se è chiaro che per rispondere esattamente a queste domande è richiesta una buona conoscenza del funzionamento dei motori, di aereodinamica e delle norme che regolano la dichiarazione delle prestazioni da parte delle case automobilistiche, è altresì chiaro che le cifre dopo la virgola non danno nessuna informazione rilevante sulla effettiva velocità massima che potrà essere raggiunta dalla vettura acquistata dal lettore.

Si provi anche a dare uno sguardo ai dati chimico-fisici delle acque minerali riportati in tabella 1.4. Anche se non si raggiungono i livelli assurdi delle velocità pubblicate da Quattroruote si cerchi di capire se hanno senso tutte le cifre riportate (per esempio che la temperatura alla sorgente sia 18.05°C).

3.4.1 Dai valori letti ai risultati delle misure

Il problema del numero di cifre con cui presentare un risultato verrà trattato per gradi. In questo capitolo ci occupiamo soltanto del numero di cifre con cui scrivere sul logbook il valore misurato *direttamente* (ossia letto sullo strumento) e quante cifre portarsi dietro nei conti nell'effettuare misure *indirette*. A tempo debito tratteremo più a fondo il problema con criteri basati su argomenti probabilistici.

Fino agli inizi degli anni '70 era importante stabilire a priori il numero di cifre con le quali lavorare, poiché ogni cifra in più voleva dire molti conti da fare a mano o con l'ausilio di tabelle e di regoli calcolatori. L'uso dei calcolatorini tascabili ha praticamente eliminato tale problema dal punto di vista tecnico ma, ciò nonostante, è inopportuno portarsi dietro più cifre di quelle necessarie ad esprimere il risultato con la precisione con cui è stato ottenuto. Infatti, se esse non danno *informazioni significative* sulle grandezze misurate sono soltanto ingombranti, facilitano la possibilità di errore di trascrizione e riducono l'*intelleggibilità* di tabelle e risultati.

Volendo stabilire dei criteri da seguire, diciamo innanzitutto che le letture vanno effettuate al meglio, sforzandosi di interpolare ad occhio fra le tacche di strumenti a lettura analogica. Per esempio, nella lettura di una lunghezza con un normale righello da disegno bisogna tentare di apprezzare il decimo di millimetro. Torneremo a tempo debito sull'incertezza da attribuire a tale stima personale e a come essa si rifletta sull'incertezza totale della misura. È comunque chiaro dall'esperienza dell'interpolazione fra le tacche del paragrafo 2.2 che la lettura dei decimi di divisione è ragionevole, anche se, a volte, può essere vanificata poi da altri errori in gioco. Ma questo è un altro problema che andrà trattato con la giusta attenzione al momento dovuto.

Per quanto riguarda le cifre sulle elaborazioni dei dati originali, prendiamo l'esempio di misure effettuate con un righello su un tondino di acciaio (un cilindro molto alto). Chiamiamo d il diametro e assumiamo, facendo del nostro meglio, di stimare che esso sia pari a 8.0 mm. È da notare subito l'importanza dello zero dopo la virgola, sebbene si dica comunemente che esso non conta: nell'ambito delle misure esso conta moltissimo! Sta infatti ad indicare che, dovendo scegliere fra 7.9, 8.0 e 8.1 (ovvero[1] fra 7.5, 8.0 e 8.5) chi ha effettuato la lettura dello strumento ha deciso che 8.0 fosse il valore più ragionevole (o più probabile). Se - caso assurdo per un righello, ma possibilissimo con altri tipi di strumenti - lo sperimentatore fosse stato in grado di scegliere fra 7.99, 8.00 e 8.01, egli avrebbe dovuto registrare 8.00, in quanto entrambi gli zeri erano stati accertati.

Calcoliamo ora l'area della sezione del cilindro mediante la formula $A = \pi(d/2)^2$. La calcolatrice fornisce un valore di 50.26548246. Ma se noi prendessimo sul serio questo numero è come se veramente stessimo decidendo fra 50.26548245, 50.26548246 e 50.26548247. Non sembra molto sensato.

[1] Per ora consideriamo che l'incertezza della misura sia dell'*ordine di grandezza* dell'ultima cifra registrata.

Se nel calcolo avessimo utilizzato gli altri due valori fra i quali eravamo indecisi (7.9 e 8.1, tanto per fissare le idee) avremmo ottenuto rispettivamente 49.01669938 e 51.52997350. Quindi essenzialmente stiamo decidendo fra 49, 50 e 51. Tutte le altre cifre che seguono non sono affatto *significative*. Per coerenza scegliamo 50, ovvero $50\,\mathrm{cm}^2$.

È da notare che il valore numerico dipende dall'unità di misura. Si sarebbe potuto riportare il risultato in centimetri o in metri. Il valore numerico del diametro sarebbe stato 0.80 e 0.080 nei due casi. Ripetendo il ragionamento precedente si sarebbe trovato che il valore dell'area sarebbe stato $0.50\,\mathrm{cm}^2$ e $0.0050\,\mathrm{mm}^2$ nei due casi.

3.4.2 Cifre decimali e cifre significative

È evidente che gli "zeri a sinistra", ovvero la posizione della virgola, dipendono dall'unità di misura utilizzata. Ne segue che, per considerare il numero di cifre da portarsi dietro, non ci si può basare semplicemente sul numero di cifre dopo la virgola. L'informazione sulla "qualità" della misura è contenuta nelle due cifre "80" (o "50"), le quali sono traslate rispetto alla virgola a seconda che che il risultato sia dato in metri, centimetri, millimetri, micron o chilometri. Esse rappresentano le *cifre significative*. Si dice allora che, nel nostro esempio, il diametro e l'area della sezione sono stati misurati con 2 cifre significative.

Un modo per rendere evidente l'indipendenza del numero delle cifre significative dall'unità di misura è di scrivere il numero che esprime il risultato in notazione "scientifica" (secondo le istruzioni dei calcolatorini) o "esponenziale" (secondo alcuni linguaggi di programmazione) nella forma

$$\{\mathit{mantissa}\} \cdot \{\mathit{potenza\ di\ 10}\}\,,$$

ad esempio $8.0 \cdot 10^0$ mm, $8.0 \cdot 10^{-1}$ cm, etc (chiaramente si può omettere 10^0).

Supponiamo ora di misurare l'altezza del cilidro con il righello e che essa risulti $h = 57.83$ cm, ovvero $h = 578.3$ mm. In questo caso si è riusciti ad accertare ben quattro cifre, ovvero abbiamo 4 cifre significative.

Interessiamoci ora al volume del tondino. Moltiplicando l'area di base per l'altezza otteniamo dalla calcolatrice i valori 28915.00000 mm^3 o 28.91500000 cm^3, etc. Quante di queste cifre sono veramente significative? In realta, se ci ricordiamo che l'area poteva essere compresa fra 49 e 51 mm^2 e l'altezza fra 578.2 e 578.4 mm, possiamo costruire una tabella dei valori del volume per le 9 combinazioni che potevamo scegliere (tabella 3.3).
I valori oscillano intorno a 29000 con oscillazioni di circa 500 in alto o in basso. Abbiamo riottenuto un risultato *sostanzialmente*[2] a due cifre significative. In questo caso i tre zeri che seguono "29" non sono significativi ma servono "di appoggio" e quindi possono trarre in inganno se dal contesto non è chiaro

[2]In questo caso in effetti due cifre significative sembrano poche, ma tre sembrano troppe. È chiaro che il numero di cifre significative dipende dal sistema di numerazione adottato. Per esempio le tre cifre significatice di "125" diventano sette nel sistema binario ("1111101") e si riducono a due nel sistema esadecimale ("7D"). Vedremo al momento opportuno come presentare il risultato con il numero di cifre significative che dipende dalla *qualità* della misura effettuata e non dall'uso del sistema decimale.

Un esempio estremo che mostra molto bene le grandi discontinuità prodotte da un uso ferreo

h (mm)	A (mm^2)		
	49	50	51
578.2	28331.80000	28910.00000	29488.20000
578.3	28336.70000	28915.00000	29293.30000
578.4	28341.60000	28920.00000	29498.40000

Tabella 3.3: Possibili valori (in mm^3) per il volume di un cilindro di cui sia la base che l'altezza hanno un'incertezza di ± 1 sulla cifra meno significativa

quali cifre sono veramente significative. In questo caso è allora preferibile la scrittura in notazione esponenziale: $2.9 \cdot 10^4$ mm^3 o $29 \cdot 10^3$ mm^3. Meglio ancora, 29 cm^3.

3.4.3 Regole pratiche (da prendere "cum grano salis")

Se osserviamo attentamente la tabella 3.3 scopriamo che le variazioni sono molto più ampie lungo le righe che lungo le colonne in quanto il maggior responsabile delle variazioni del risultato è il fattore con il minor numero di cifre significative. Infatti, essendo la misura dell'area incerta di circa una parte su cinquanta, anche se l'altro fattore fosse un valore "esatto" il prodotto sarebbe ancora affetto da una incertezza di una parte su cinquanta. Si dice allora che "l'incertezza di un fattore si è *propagata* sull'incertezza sul prodotto". Da questo ragionamento si evince come l'incertezza sul prodotto sia maggiormente influenzata dal fattore che ha la maggiore *incertezza relativa*, ove per incertezza relativa si indica l'incertezza in unità del valore della grandezza.

Otteniemo quindi la seguente regola pratica:

> *in una moltiplicazione fra due grandezze il numero di cifre del* **prodotto** *è determinato dal fattore avente il minor numero di cifre significative.*

Per vedere cosa succede nel caso di una somma, consideriamo una lastra metallica del quale misuriamo lo spessore con un micrometro. Ammettiamo di essere riusciti a leggere 0.234 mm. Calcoliamo la somma di tale spessore con il diametro e con l'altezza del tondino (ricordiamo che valevano 8.0 e 578.3 mm rispettivamente). Senza ripetere in dettaglio i ragionamenti fatti sopra, si può vedere che nei due casi i risultati sono 8.2 mm e 578.5 mm. Le informazioni sui centesimi e millesimi di millimetro dello spessore della lastra sono irrilevanti alla fine della somma in quanto le altre due grandezze sono già incerte al decimo di millimetro. Quindi nella somma non è il numero di cifre

delle regole sulle cifre è il seguente:

$$\begin{aligned} 3.6 \times 2.8 &= 10\,, \\ \text{ma}\;\; 3.6 \times 2.7 &= 9.7\,. \end{aligned}$$

Nel primo caso la cifra meno significativa è un decimo del valore della grandezza, nel secondo un centesimo. Chi legge soltanto i due risultati è indotto a pensare, erroneamente, che i due risultati derivino da esperimenti di diverso livello di qualità.

significative ad essere importante, quanto l'addendo che ha l'*incertezza (assoluta)* maggiore (l'aggettivo "assoluto" è ridondante e serve solo a contrapporsi a "relativo" incontrato poc'anzi; in particolare, non ha niente a che vedere con il concetto di valore assoluto nel senso algebrico). Quindi ne segue che

> *in una addizione fra due grandezze l'ordine di grandezza della cifra meno significativa della* **somma** *è pari al maggiore degli ordini di grandezza delle cifre meno significative degli addendi.*[3]

Si può verificare facilmente che quanto detto per il prodotto vale anche per il rapporto, e quanto detto per la somma vale anche per la differenza.

Nel seguito troveremo delle regole generali per *propagare* l'incertezza e stabilire il numero adeguato di cifre significative per qualsiasi operazione. Per ora vale la pena di soffermarci sui logaritmi, mentre il caso delle potenze può essere assimilato in prima approssimazione a quello dei prodotti.

Provando con un po' di esempi, del tipo di quelli mostrati per somme e prodotti, ci si rende conto di una regola empirica che troverà giustificazione nella regola generale ma che per ora può sembrare un po' sorprendente:

> *il* **logaritmo** *ha tante cifre decimali significative quante sono le cifre significative dell'argomento.*

Ad esempio, se la grandezza y è legata ad x dalla relazione $y = \ln x$, con $x = 8.51$, 85100, 0.851 e 0.0000851, y varrà rispettivamente[4] rispettivamente 2.141, 11.352, -0.161 e -9.372. Questo comportamento "anomalo" del logaritmo può essere giusticato dal punto di vista fisico ricordandosi che l'*argomento del logaritmo deve essere adimensionale*. Quindi il valore di una grandezza fisica argomento del logaritmo deve essere sempre rapportata ad un'altra grandezza ad essa omogenea che funge da *scala*. Molto spesso questa è semplicemente l'unità di misura di tale grandezza. Ma l'unità di misura è in principio arbitraria, con il semplice effetto di moltiplicare per un fattore (privo di incertezza) il valore numerico della grandezza. Il logaritmo sarà quindi pari al logaritmo della mantissa più il logaritmo di tale fattore. Quindi le cifre del logaritmo potrebbero crescere arbitrariamente, mentre la cifra meno significativa della mantissa rimane sempre la stessa, con la conseguente crescita del numero di cifre significative del risultato.

Questo apparente paradosso non deve spaventare, in quanto nei casi normali di laboratorio il calcolo dei logaritmi compare sempre in combinazioni del tipo

$$y = \ln x_2 - \ln x_1 = \ln \frac{x_2}{x_1} \qquad (3.1)$$

che chiaramente non dipendono dall'unità di misura (comune!) di x_1 e x_2.

Gli *esponenziali* si comportano in modo opposto ai logaritmi.

[3] Per la serie "ci vuole più a dirlo che a farlo"...

[4] L'uso del logaritmo decimale mostra meglio l'effetto delle cifre significative del logaritmo semplicemente dovute alla posizione della virgola. In questo caso si otterrebbe rispettivamente 0.930, 4.930, -0.070 e -4.070 .

3.4.4 Suggerimenti

A conclusione di questo paragrafo diamo due suggerimenti sul numero di cifre cifre da utilizzare nelle prime esperienze di laboratorio.

- Riportare i dati originali con tutte le cifre che si ritiene essere in grado di leggere e di stimare.
- Ne caso non sia ovvio dal contesto, annotare chiaramente il tipo di strumento con cui è stata eseguita la lettura. Ad esempio, una lettura di 1.35 mm è qualitativamente diversa a seconda che sia stata eseguita con un calibro ventesimale (1.35 è in alternativa a 1.30 o 1.40) o con un calibro centesimale (1.35 è in alternativa a 1.34 o 1.36).
- Nell'effettuare i conti ricordarsi delle regole sulle cifre significative di somme/differenze, prodotti/rapporti e logaritmi/esponenziali, cercando di interpretarre tali regole con un po' di giudizio; eventualmente ci si può portare dietro una cifra in più per evitare problemi di *arrotondamento*; nel caso di dubbi è meglio abbondare un po' in cifre e arrotondare alla fine (ma questo si può fare solo quando si conoscono le regole di propagazione delle incertezze) piuttosto che essere costretti a rifare tutti i conti (le cifre mancanti non possono essere inventate!).

L'esperienza insegnerà a poco a poco come ci si deve regolare. Ad esempio:

- nella lettura di uno strumento digitale è inutile riportare le eventuali cifre che fluttuano in continuazione: chiaramente esse non sono significative;
- se è nota a priori con quale precisione è sufficiente ottenere un certo risultato è inutile misurare le grandezze da cui questo risultato dipende con una precisione più elevata di quella necessaria (anche per queste valutazioni si presuppone di conoscere le regole di propagazione dell'incertezza).

A questo stadio non siamo ancora in grado di fornire il risultato accompagnato da una valutazione quantitativa di un'incertezza di misura che abbia un ben preciso significato probabilistico. È quindi sufficiente seguire le regole delle cifre significative, al fine di non dare valori ridicoli (ad esempio $g = 9.9231453\ \mathrm{m/s^2}$). Se è possibile, è raccomandabile di ripetere più volte le misure. La variabilità dei risultati danno un'idea di quali cifre sono veramente significative. Ad esempio, fra 40.65, 42.65 e 43.38, un risultato combinato di 42 è più che ragionevole.

3.5 Arrotondamenti

A questo punto è d'obbligo un commento sugli arrotondamenti. Non ci sono dubbi sul fatto che le cifre da 1 a 4 si arrotondano per difetto e quelle dal 6 al 9 per eccesso. Il 5 crea in genere un po' di imbarazzo. Secondo una tradizione esso dovrebbe essere arrotondato una volta in difetto e una volta in eccesso per evitare effetti sistematici sui risultati (ad essere pignoli questa operazione andrebbe fatta "a caso" - assolutamente non banale! - e non alternativamente).

Non riteniamo il problema di importanza cruciale e lo lasciamo al gusto personale. Comunque bisogna tener conto che:

- le calcolatrici tascabili nell'opzione *"fix n"* che fissa il numero di cifre decimali arrotondano il 5 per eccesso;[5]
- quando si saprà valutare quantitativamente l'incertezza di misura in genere si sarà sempre nelle condizioni tali per cui queste finezze di arrotondamento saranno inessenziali.
- se proprio l'"eliminazione" del 5 fa nascere dei conflitti di coscienza è meglio riportarlo che rischiare di non dormirci sopra ...

3.6 Controllo dimensionale e degli ordini di grandezza

Come per i problemi di Fisica Generale anche per l'elaborazione dei dati raccolti in laboratorio il controllo dimensionale delle grandezze ottenute riduce la possibilità di sbagli. Però se si controllano soltanto le dimensioni del risultato con quanto "dovrebbe venire" non è semplice capire dove si è sbagliato. Si raccomanda pertanto di riportare ad ogni passo dell'analisi il valore numerico della grandezza insieme alla sua unità di misura, cercando di utilizzare le unità che sono più naturali per l'esperimento in corso. Quando si lavora simultaneamente con grandezze meccaniche, elettriche e termiche è opportuno esprimere tutti i valori nel sistema internazionale (SI) ed esprimere l'energia in Joule (anziché calorie), la potenza in Watt, il tempo in secondi, e così via.

Oltre che a controlli dimensionali è opportuno effettuare controlli sugli ordini di grandezza dei valori ottenuti. Ma per questo bisogna avere una certa familiarità con gli ordini di grandezza tipici, ovvero delle conoscenze di Fisica Generale. Quindi 100 m/s^2 deve far pensare a una accelerazione tremenda, 1 μs a un tempo non accessibile al cronometraggio manuale, 0.1 F a una capacità mostruosa, 10^{-3} Joule a una energia piccolissima, e così via. Quindi il risultato "raffinato" ottenuto con i metodi di analisi più o meno "sofisticati" che incontreremo deve essere confrontato con quello che si può dare da una stima rozza e veloce. Quest'ultima infatti può sbagliare sì del 20, 30 o 50 %, ma difficilmente sbaglierà di un fattore 1000, come può succedere quando si commettono errori di distrazione in calcoli complicati.

A questo riguardo è altamente raccomandabile effettuare la stima rozza prima di intraprendere il calcolo complicato. Ad esempio, nell'esperimento della molla di paragrafo 2.6 si vede subito come una variazione di circa 0.5 kg ($\approx$ 5 N) abbia prodotto un allungamento di una decina di centimetri (0.1 m) e quindi k è dell'ordine di grandezza di 50 N/m. Se dall'analisi grafica o dal fit risulta un valore di 0.5 N/m c'e' sicuramente qualcosa che non va (significa che 50 g produrrebbero un allungamento di 1 m!).

[5]Le operazioni automatiche tendono a tranquillizzare gli indecisi (per la serie "lo ha detto il computer").

3.7 Problemi

1. Leggere le posizioni indicate dalle freccie di figura 2.2, cercando di stimare i decimi di millimetro.

2. Indicare per ciascuno di questi valori il numero di cifre significative: a) 3.14; b) 1234.0; c) 123; d) 0.777; e) 0.0003; f) $6.022 \cdot 10^{23}$; g) 3.200; h) 20000; i) 20001; j) 0.0020; k)0.2000000; l) 13400; m) $0.7 \cdot 10^{-12}$; n) 300.001; o) $0.0030 \cdot 10^{-100}$.

3. Eseguire le seguenti operazioni elementari fra risultati di misure fornendo un ragionevole numero di cifre significative:

   (a) $3.14 + 6.77 + 21.1$;

   (b) $6 - 0.1178$;

   (c) $3.14 + 0.01246$;

   (d) $\pi + 0.01246$;

   (e) $C = 2\pi R$, con $R = 7.895\,\mathrm{cm}$;

   (f) 45.9×0.0023;

   (g) $(34.334 + 1.1) \times 0.02435$;

   (h) $V = 4/3\pi R^3$, con $R = 4.45$;

   (i) $2.68 + 1.23 \cdot 10^{-2} + 9.88 \cdot 10^{-8}$;

   (j) $75.78 + 3.91 \cdot 10^3 + 1.122 \cdot 10^4$;

   (k) $23456.239 + 1.334 \cdot 10^6 + 5.5 \cdot 10^5$;

   (l) $2.3001 \cdot 10^6 \times 0.052 \cdot 10^5$;

   (m) $2.6 \times 1.2345 \cdot 10^{120} \times 8.8800 \cdot 10^{144}$;

   (n) $2.543 \cdot 10^{105} + 0.000077 \cdot 10^{108} + 0.00122 \cdot 10^{106}$;

   (o) $7.41 \cdot 10^{-130} + 4.5823 \cdot 10^{-134} \times 32$;

   (p) $\ln 2.7$;

   (q) $\ln 27$;

   (r) $\ln 6.02 \cdot 10^{23}$;

   (s) $m = \frac{\ln y_2 - \ln y_1}{x_2 - x_1}$, con $x_1 = 10.0\,\mathrm{s}$, $x_2 = 213.2\,\mathrm{s}$, $y_1 = 1.5\,°\mathrm{C}$ e $y_2 = 33.6\,°\mathrm{C}$;

   (t) $m = \frac{\ln y_2 - \ln y_1}{x_2 - x_1}$, con $x_1 = 15.2\,\mathrm{ms}$, $x_2 = 148.0\,\mathrm{ms}$, $y_1 = 12.6\,\mathrm{V}$ e $y_2 = 15.2\,\mathrm{mV}$;

4. Dai dati del diagramma ad albero 4.1 stimare il numero di abitanti di età inferiore a 10 anni. (mettere come domanda nellla didascalia della figura).

Capitolo 4

Descrizione grafica dei dati sperimentali

4.1 Riduzione dei dati

Riprendiamo l'esperimento simulato del contatore del paragrafo 1.1. La tabella 1.1 contiene quasi tutte le informazioni che riguardano l'esperimento. Il "quasi tutte" si riferisce al fatto che mancano tanti dati importanti: luogo, data e ora delle misure; nome dello sperimentatore; condizioni ambientali (temperatura, etc.); specifiche e condizioni di lavoro dello strumento (tensione di alimentazione, valore di soglia, etc.); e così via. Eppure, se si vuol capire quello che "sta succedendo", le informazioni offerte dalla tabella sono, in un certo senso, "troppe". È chiaro che al fine del risultato della misura siamo interessati a rispondere alle domande formulate sopra e quindi a *condensare* le informazioni in pochi numeri. È quindi necessaria una *riduzione dei dati* (espressione che è più o meno equivalente ad *analisi dei dati*) generalmente effettuata attraverso passi successivi.

Per capire meglio il problema, consideriamo il caso dello studio della popolazione di un certo paese. Interessiamoci ad esempio alle età degli abitanti di una nazione. Si capisce subito che quello che si desidera non è la lista di nome, cognome, data di nascita e altre informazioni per ciascun cittadino. Ci vorrebbe una vita a scorrere tutti i numeri e alla fine non si sarà imparato niente. È preferibile raggruppare le informazioni in *classi omogenee*, per esempio maschi e femmine di ogni età. Quindi si possono comporre due *tabelle*, una per ciascun sesso, in cui viene riportato il numero di persone per ogni classe di età. Un ulteriore passo per descrivere le informazioni di interesse nella forma più sintetica consiste nel darne una *rappresentazione grafica*. Nel caso specifico la rappresentazione più conveniente consiste nel familiare *diagramma ad albero* della popolazione. Quello relativo alla Germania del 1992 è mostrato in figura 4.1. Come i cerchi di un tronco d'albero, tale figura mostra a colpo d'occhio la storia del paese.

Ovviamente riducendo i dati si *perdono* informazioni ma se ne *acquistano* altre di natura globale:

- dai dati originali alla tabella: si perde l'informazione sull' individuo, guadagnando quella relativa al numero di persone di ogni classe di età;

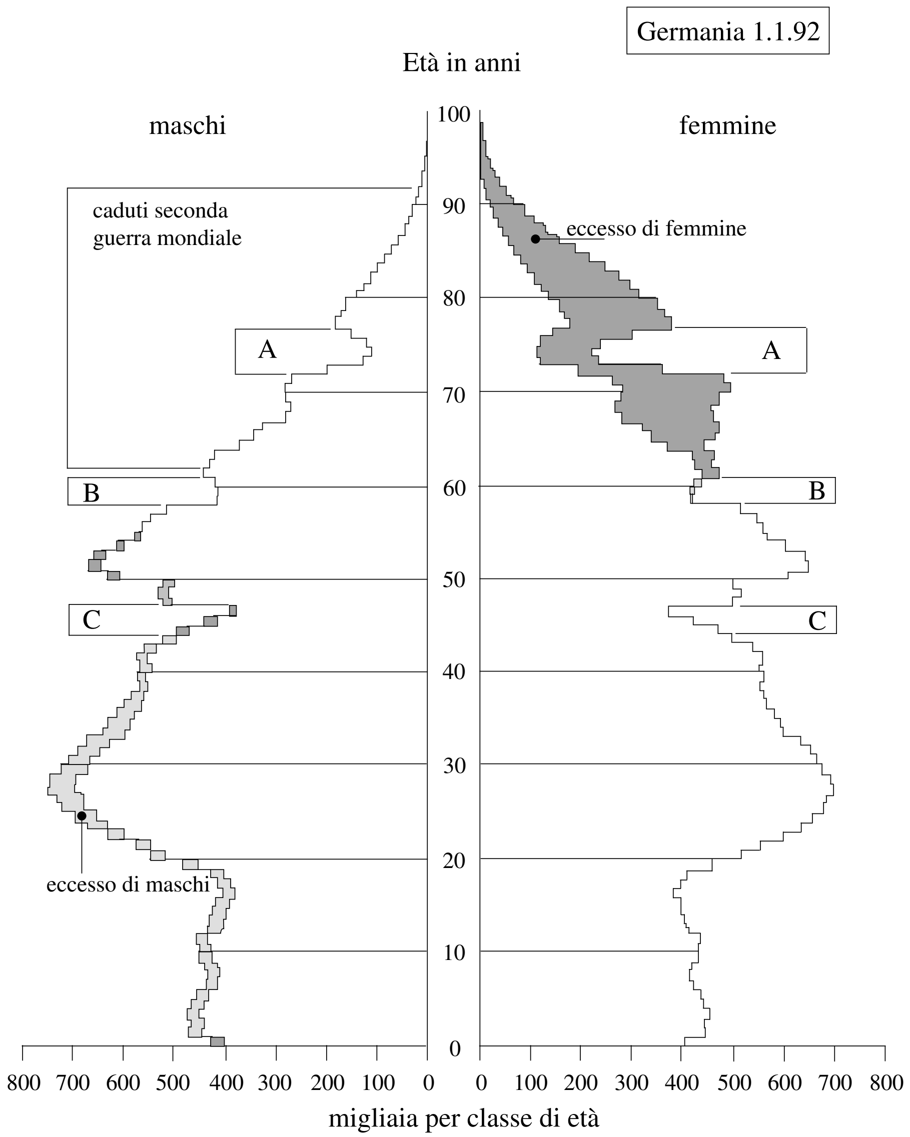

Figura 4.1: Diagramma ad albero della popolazione tedesca. Si notino: la maggiore natalità di maschi dovuto alle misteriose leggi della natura; l'eccesso di donne anziane dovute alla loro più elevata vita media (l'effetto maggiore è dovuto ai caduti della 2^a guerra mondiale); la caduta delle nascite in corrispondenza alle guerre mondiali (A e C) e alla crisi economica del '32 (B); il boom economico con aumento delle nascite dell'inizio degli anni '60 e il controllo delle nascite dagli anni '70 in poi.

conteggi	$T = 3\,s$	$T = 6\,s$	$T = 12\,s$	$T = 30\,s$	$T = 100\,s$
0	56	33	11		
1	32	40	20	4	
2	9	20	32	5	
3	2	6	20	13	
4	1	1	10	20	
5			7	17	
6				14	
7				14	
8				5	1
9				6	1
10				1	2
11					2
12				1	4
13					6
14					9
15					2
16					8
17					11
18					13
19					11
20					8
21					3
22					8
23					2
24					2
25					2
26					
27					3
28					
29					2
30					

Tabella 4.1: Esperienza del contatore: frequenze di conteggio delle misure da 3 a 100 secondi.

$T = 300\,s$										
Nr di conteggi	31	32	33	34	35	36	37	38	39	40
Nr di eventi			1				1		2	3
Nr di conteggi	41	42	43	44	45	46	47	48	49	50
Nr di eventi		1	1	3	2	2	1	5	6	2
Nr di conteggi	51	52	53	54	55	56	57	58	59	60
Nr di eventi	8	8	7	5	4	3	5	6	2	2
Nr di conteggi	61	62	63	64	65	66	67	68	69	70
Nr di eventi	8	1	4	2	2		1			
Nr di conteggi	71	72	73	74	75	76	77	78	79	80
Nr di eventi	1							1		

Tabella 4.2: Esperienza del contatore: frequenze di conteggio delle misure di 5 minuti.

- dalla tabella alla rappresentazione grafica: si perde l'informazione dell'esatto numero di persone per ogni età (che in ogni caso cambia il giorno dopo quello preso in considerazione) ottenendo in compenso un quadro d'insieme a volte difficilmente quantificabile e che offre allo sperimentatore lo spunto per ulteriori considerazioni ed analisi[1].

I dati possono essere ulteriormente ridotti attraverso successivi stadi di analisi, fino ad arrivare ad una sola informazione, ad esempio il numero di abitanti o l'età media. Ovviamente a questo punto il numero di informazioni perse è enorme. Ad esempio, si noti come la consueta riduzione di informazioni a due - numero totale di abitanti per sesso - induca molti a credere che nascano più femmine che maschi, mentre è noto da secoli il contrario.

4.2 Tabelle, istogrammi e diagrammi a barre

Tornando alla nostra misura di radioattività con il contatore a scintillazione, un modo per presentare le informazioni in forma più compatta è quello di raccoglierle in tabelle in cui le classi omogenee sono rappresentate dai conteggi registrati (tabelle 4.1 e 4.2). Il numero di volte in cui si è verificato un certo conteggio è chiamato *frequenza*. L'informazione persa in questa riduzione dei dati è quella della sequenza temporale con cui si sono verificati i conteggi[2]. Essa è comunque inessenziale se *si presuppone* che la radioattività *praticamente* non cambia nel tempo o, meglio ancora, se *è controllato* che tale ipotesi è *ragionevole*.[3]

Il modo più semplice di rappresentare graficamente le frequenze di conteggio consiste nel costruire un *istogramma* per ogni serie di misure (figura

[1]Gli istogrammi del tipo di quelli mostrati in figura 4.2 vanno visti come funzionali alla costruzione delle tabelle

[2]Ricordiamo che formando segnali *logici* ("sì" o "no") dagli impulsi elettrici in uscita dal fotomoltiplicatore si era già persa l'informazione esatta sulla ionizzazione prodotta.

[3]Qualcuno sarebbe tentato di parlare di ipotesi corretta ("vera"). Non ci vuole molto a convincersi che quando ha a che vedere con fenomeni aleatori, è raro che si possa arrivare a definire certa, o vera, una data ipotesi.

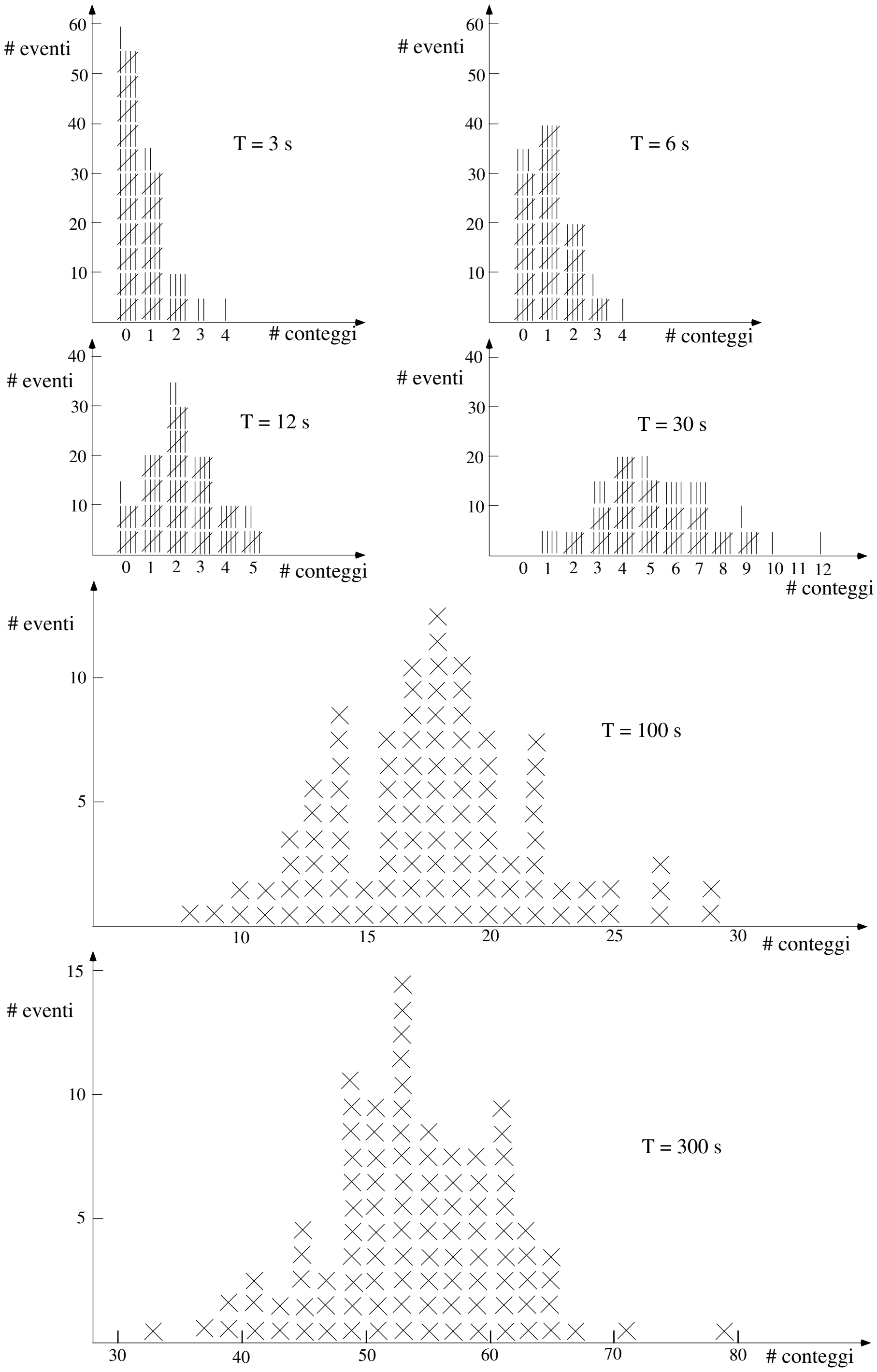

Figura 4.2: Istogrammi dei risultati dell'esperienza del contatore. Il simbolo "#" sta per "numero". Il termine "evento" è qui usato - come avviene usualmente - nel senso di "occorrenza" o "numero di volte".

4.2). L'uso di rappresentazioni grafiche di questo tipo è talmente comune sia nei libri che nei mass media che la loro lettura non necessita di alcuna spiegazione. Qualche parola va invece spesa sul modo di realizzarli. Gli istogrammi mostrati in figura 4.2 sono fra quelli che più frequentemente si incontrano in laboratorio, specialmente per dare un'occhiata veloce ad una manciata di dati e per contare le occorrenze di ciascuna classe[4]. In genere per operazioni più complesse entrano immediatamente in gioco i calcolatori. Osservando la figura si notano tre tipi di istogrammi.

- In quello relativo alle misure da 100 s c'è una crocetta per ogni occorrenza di un certo numero di conteggi (il simbolo "#" sta per "numero di"). Le *occorrenze* sono usualmente chiamate anche *"eventi"* dai fisici, anche se vedremo che nello studio della probabilità tale nome verrà utilizzato con altro significato (è opportuno abituarsi subito a qualche inevitabile ambiguità di linguaggio, ricordandosi che, come scrisse Wittgenstein "un significato di una parola è un suo mode del suo impiego").

- Nei quattro istogrammi in alto il numero di eventi è talmente elevato che la tecnica precedente non può funzionare. È preferibile allora mettere delle barrettine verticali per ogni evento e raggruppare gli eventi con un tratto obliquo ogni cinque eventi.

- Nelle misure da 300 s si ha invece un problema di dispersione di dati sull'asse delle ascisse. Anche se si fossero tracciate delle crocette più piccole, la dispersione sarebbe tale da far perdere la visione d'insieme. Si è allora preferito formare nuove classi raggruppando conteggi contigui. Il numero tracciato sull'asse delle ascisse rappresenta il limite inferiore della classe. Ovvero vengono riportati all'interno di un bin gli eventi per i quali è valida la relazione

  $$\text{limite inferiore} \le X < \text{limite superiore}\,. \tag{4.1}$$

  L'istogramma indica, ad esempio, che si sono verificati 10 eventi per i quali il numero di conteggi è maggiore o uguale di 60 e minore di 62. A differenza degli altri due tipi di istogrammi, qui si è persa l'informazione della quantità di eventi per numero di conteggi all'interno di ciascuna classe.

  Quest'ultimo tipo di istogramma è quello comunemente utilizzato se un risultato si può presentare in un grande numero di classi. Tale numero può diventare virtualmente infinito per grandezze i cui valori possono assumere valori con continuità - dal punto di vista pratico, non dal punto di vista puramente matematico - in un certo intervallo.

[4]Sottolineamo che questi istogrammi sono utili non soltanto per avere una rappresentazione grafica delle informazioni, ma anche per contare le occorrenze di ciascuna classe. Quindi, anche se per ragioni didattiche le tabelle 4.1 e 4.2 sono state introdotte prima degli istogrammi di figura 4.2, in realtà questi ultimi sono stati fatti per primi. Per convincersi che tali istogrammi rappresentano il modo più rapido e sicuro per contare occorrenze multiple si suggerisce di provare, come esercizio, a ricostruire le frequenze di conteggio per $T = 3$ s a partire dai dati di tabella 1.1.

Un esempio di questa applicazione è mostrato in figura 4.3, dove sono istogrammati i tempi di attesa per ottenere un numero prefissato di conteggi. Per meglio mostrare le distribuzioni dei tempi questi istogrammi sono ottenuti utilizzando 10˙000 eventi ottenuti nelle stesse condizioni di quelli della tabella 1.2.

È da notare come i quattro istogrammi in alto, della figura 4.2, realizzati con le stesse scale in ascissa e ordinata, si prestino più facilmente a confronti degli altri due. Questo insegna che, quando è possibile, è preferibile procedere in questo modo (ancor meglio se gli istogrammi possono essere disposti in colonna).

A volte nemmeno l'informazione della frequenza di ciascuna classe è importante. Ad esempio, il fatto che per 56 volte si siano verificati zero conteggi nelle misure da 3 secondi dipende dal numero totale di misure effettuate. Se invece di 100 misure ne avessimo eseguite 1000 avremmo *ragionevolmente* ottenuto un numero di zeri *circa* 10 volte maggiore. Ben diverso è invece il significato del rapporto fra il numero con cui si verificano zero conteggi e quello con cui se ne verifica uno. Ci *attendiamo* infatti che tale rapporto resti *circa costante* all'aumentare del numero di misure. Invece del rapporto fra le frequenze di due classi è preferibile allora calcolare la *frequenza relativa*, ovvero la frequenza diviso il numero di prove (le misure effettuate, nel nostro caso).

Spesso per comodità si preferisce evitare l'uso persistente di numeri minori di uno e si passa alle frequenze percentuali. Questa operazione è nota a tutti, ma bisogna fare attenzione a scrivere esplicitamente il simbolo %, e a leggere correttamente i valori graficati, per non incorrere nel rischio di sbagliare di un fattore 100. I 6 "istogrammi" di figura 4.4 mostrano le frequenze relative dei conteggi espresse in percentuali. È da notare come in questo caso non ha senso in genere mettere delle crocette, poiché le frequenze relative sono tipicamente numeri razionali. Per questo motivo questo tipo di rappresentazione grafica è più propriamente chiamata *diagramma a barre*.

4.3 ❄ Box plot

Spesso si è nella necessità di mostrare molti istogrammi contemporaneamente, in genere al variare di una certa grandezza. Il metodo suggerito nel paragrafo 4.2 era quello di riportare gli istogrammi incolonnati e con la stessa scala.

Un'alternativa che cominciano a diffondersi in certi settori della letteratura scientifica è quella dei *box plot* (o *box-and-whisker plot*), di cui è riportato un esempio in figura 4.5. Il simbolo condensa le seguenti informazioni:

- l'intervallo entro cui sono compresi i dati sperimentali (barre esterne, i "baffi", Whisker in inglese);

- l'intervallo entro cui è compreso il 50 % dei dati (scatola);

- il punto mediano dei dati nel senso definito nel paragrafo 5.3, ovvero il valore rispetto al quale si verificano tanti valori minori quanti maggiori (barretta bella scatola).

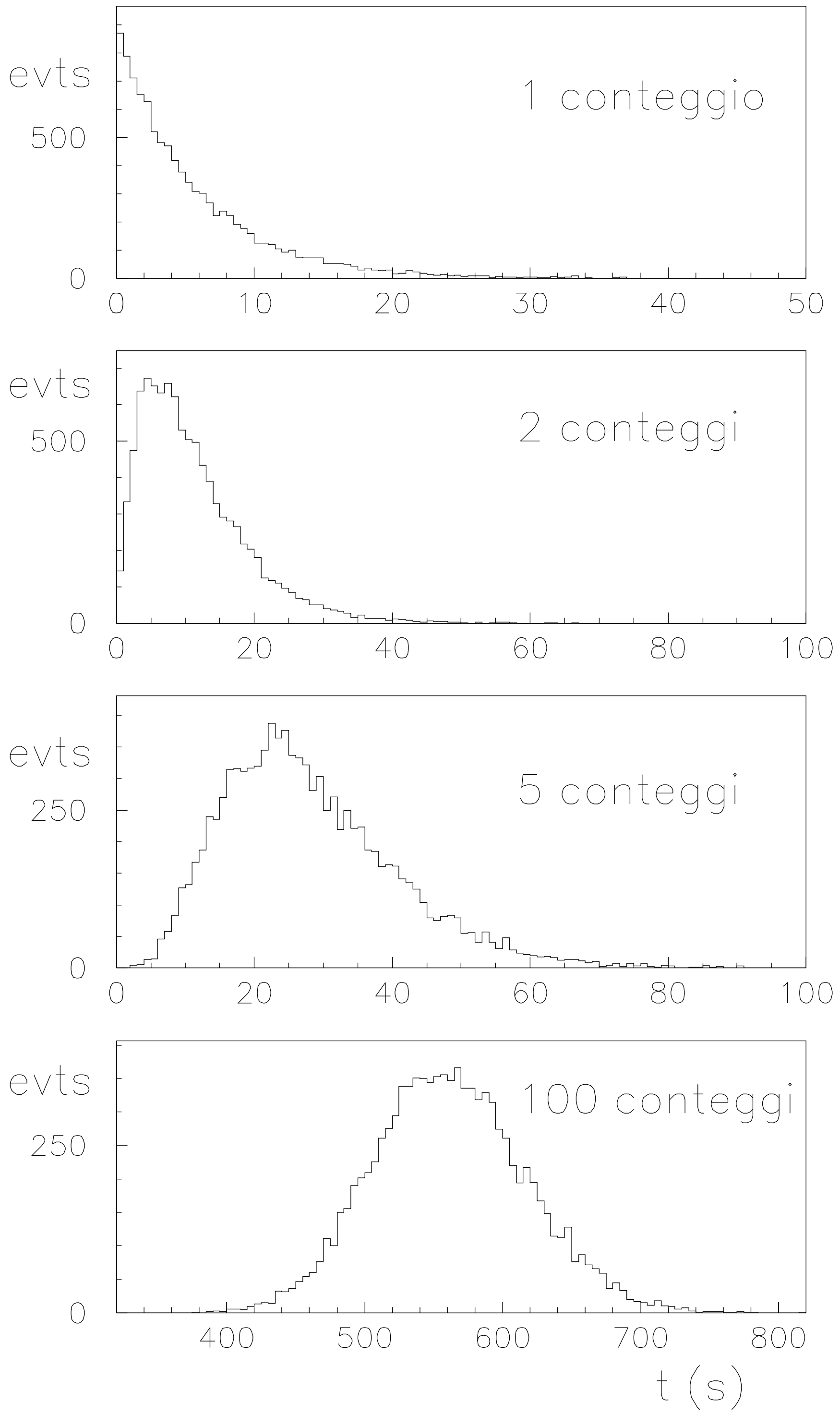

Figura 4.3: Istogrammi dei tempi di attesa per 1, 2, 5 e 100 conteggi, basati ciascuno su 10000 eventi simulati (i dati simulati della tabella 1.2 ne rappresentano un piccolo sotto-campione). "evts" è un altro modo di scrivere "numero di eventi". Un altro modo di rappresentare gli stessi eventi è mostrato in figura 6.7.

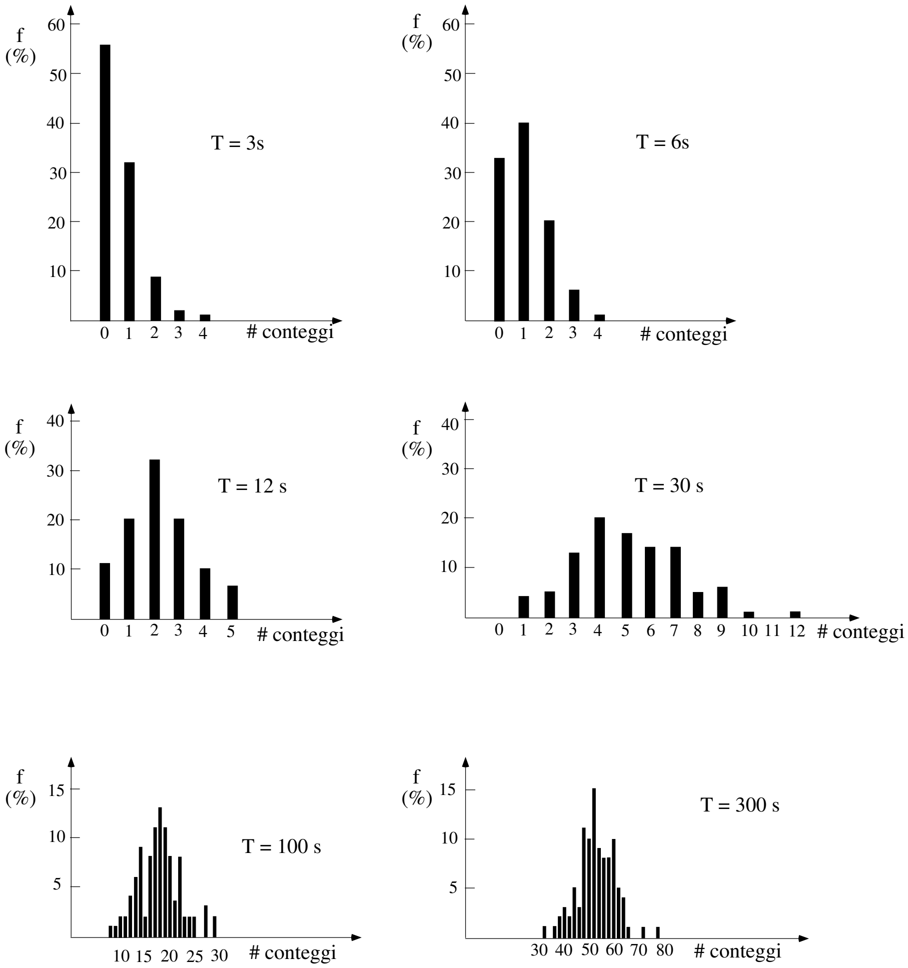

Figura 4.4: Dati dell'esperienza del contatore a scintillazione. Diagrammi a barre delle frequenze relative di conteggio.

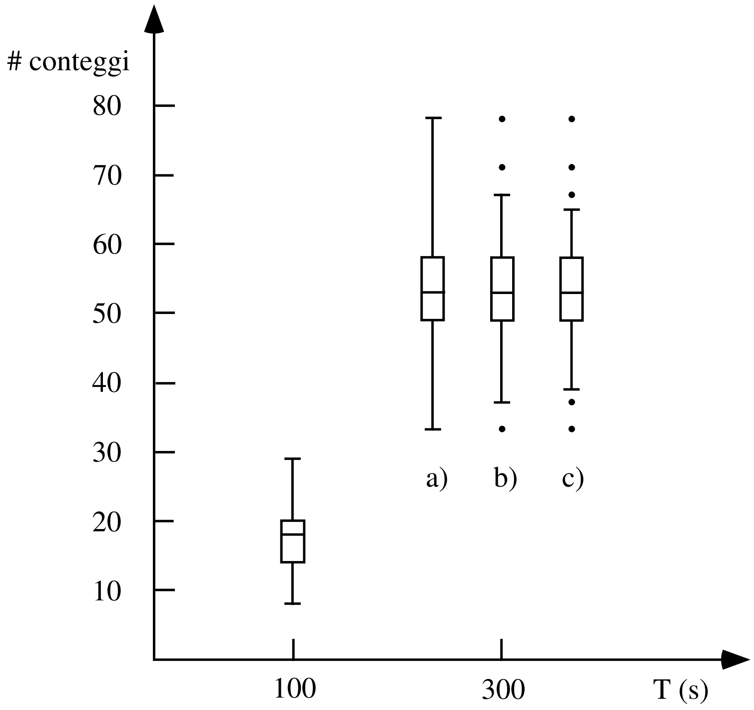

Figura 4.5: Rappresentazione a plot box-whisker ("scatola e baffi") delle distribuzioni dei conteggi per 100 e 300 s (tabelle 4.1 e 4.2). Per il secondo caso sono mostrati tre modi di rappresentare le code della distribuzione.

A volte, se ci sono dei dati sperimentali lontani dal raggruppamento principali, essi vengono considerati alla stregua di fluttuazioni anomale e indicati individualmente. La figura 4.5 mostra due possibili alternative di disegnare il box plot a seconda di come si considerano i punti più isolati[5].

4.4 ⥀ Istogrammi bidimensionali: scatter plot e lego plot

A volte si è interessati a quante volte si verifica ciascuna delle possibili coppie con le quali si possono presentare due variabili casuali. Esse possono essere tabulate in tabelle doppie, purché il numero di righe e di colonne lo permetta.

Per mostrare graficamente la distribuzione delle grandezze con istogrammi o diagrammi a barre del tipo di quelli visti sopra le crocette o le barre andrebbero collocate sul piano cartesiano definito dalle due grandezze e ... dovrebbero uscire fuori dal piano. Questo crea ovvi inconvenienti pratici.

Ci sono diverse tecniche di grafica per risolvere il problema, ciascuna con i suoi pro e contro. Ne indichiamo soltanto un paio. I programmi di grafica al computer sono ricchi di tante varianti.

[5]Si noti come la pratica di considerare fuori norma dei punti isolati può creare delle strane idee sugli eventi aleatori. Ad esempio, gli eventi della figura 4.5 non hanno assolutamente niente di "anomalo" rispetto a quelli che sono nel mezzo della distribuzione.

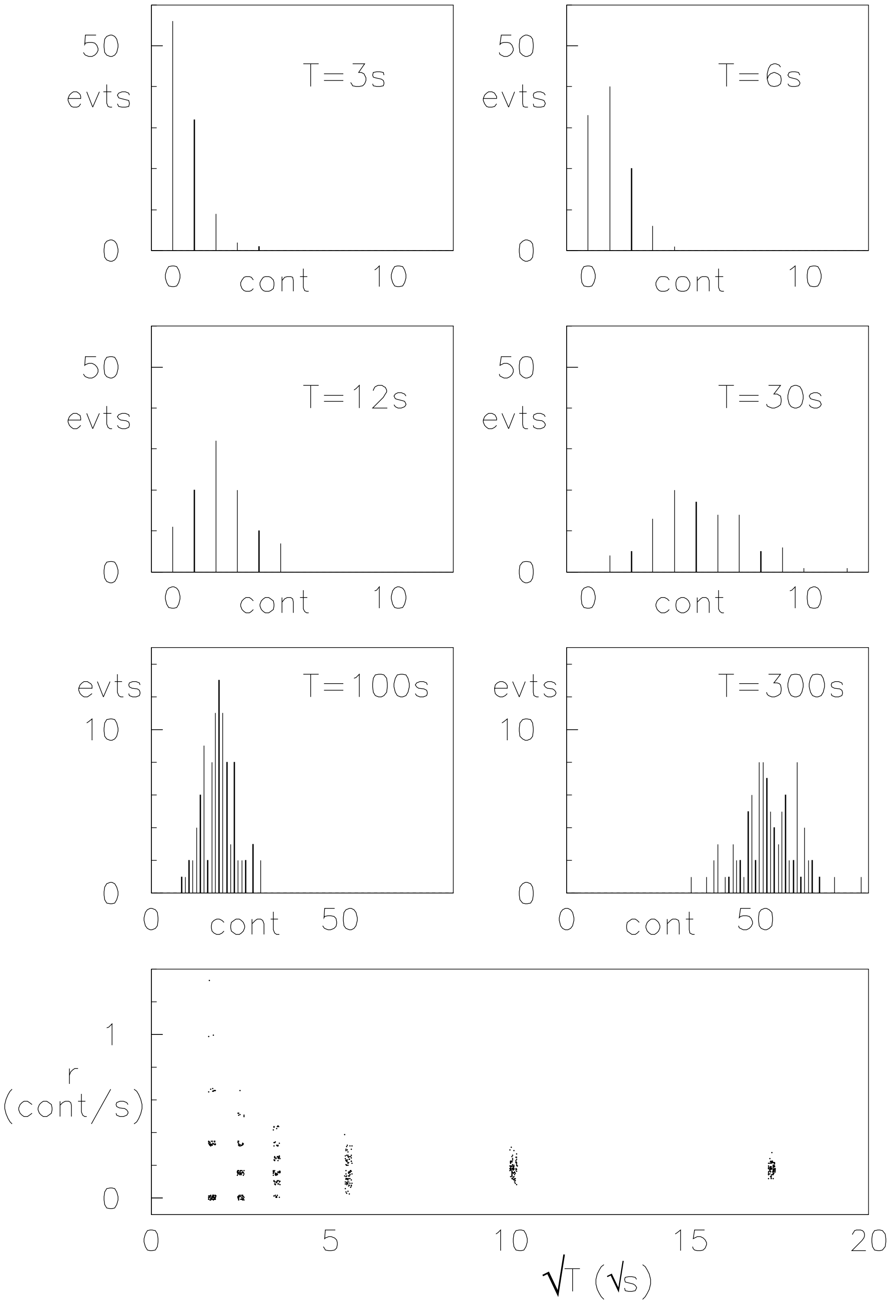

Figura 4.6: Dati dell'esperienza del contatore a scintillazione. Scatter plot della radioattività r in funzione della radice quadrata del tempo. (In quest'ultimo i punti sperimentali sono stati volutamente distribuiti intorno al valore esatto, al fine di renderli visibili).

- La tecnica che somiglia di più all'istogramma consiste nel mostrare il grafico tridimensionale (2 dimensioni per le grandezze e 1 per il numero di eventi) in visione prospettica, sia con barre che con istogrammi a tratto continuo. Questi ultimi si chiamano *lego-plot* perché somigliano alle note costruzioni per bambini. Hanno l'inconveniente che le barre alte possono mascherare quello che c'è dietro.

- Se le grandezze sono continue, o comunque tali da rendere minima la probabilità che una coppia con esattamente gli stessi valori si verifichi più di una volta, si possono disegnare dei puntini per ciascuna delle occorrenze. Questa tecnica può essere eseguita anche a mano. Questo tipo di rappresentazione si chiama scatter plot ed alcuni esempi sono mostrati nel prossimo paragrafo. Contrariamente ai lego plot questa tecnica può essere praticata anche a mano.

4.4.1 Esempi di scatter plot

Come nuovo passo nell'analisi dei dati dell'esperienza del contatore, calcoliamo il valore della radioattività in funzione della durata della misura e riportiamolo su un grafico (figura 4.6). Allo scopo di condensare la scala orizzontale (in realtà vedremo in seguito che c'è dietro anche una ragione più profonda) riportiamo sull'asse delle ascisse la radice quadrata del tempo. Ogni puntino rappresenta una singola misura. Si vede chiaramente come per ogni classe di tempo di misura i valori sono distribuiti intorno ad un valore di radioattività circa costante. Cambia invece la dispersione dei valori. Intuitivamente si può dire che una singola misura da 300 s determina il *valore vero* della radioattività con minore *incertezza* di una singola misura da 3 s. Per essere più precisi, nell'ipotesi che la radioattività si mantenga costante una misura da 300 s deve fornire un'informazione di qualità confrontabile a 100 misure da 3 s ciascuna.

Come prima analisi dei dati del pallinometro grafichiamo (figura 4.7) le frequenze relative del numero di volte con cui le palline terminano nel bin centrale e in quello laterale in funzione del numero di lanci (in realtà si è preferita la sua radice quadrata, per gli stessi motivi dell'esperimento precedente). Ogni puntino della figura rappresenta quindi il risultato di una sequenza. Le frequenze relative tendono ad addensarsi intorno a 0.5 per il bin centrale e intorno a 0.25 per quello laterale. Se si prova ad estrapolare il comportamento della frequenza per un *grandissimo numero di lanci* si può ragionevolmente affermare che *sembra poco probabile* che possano verificarsi valori di frequenza *troppo lontani* rispettivamente da 0.5 o da 0.25.

Un'altra analisi interessante è quella della differenza fra il numero di palline che termina nel bin centrale e la somma del numero di palline che va negli altri due in funzione del numero di lanci. Ingenuamente si potrebbe pensare che, poiché le frequenze tendono ad essere circa uguali, tale differenza tenderà a ridursi. In realtà si può verificare facilmente che da un lancio all'altro la differenza oscilla sempre intorno a zero, ma l'entità delle fluttuazioni cresce in entrambe le direzioni all'aumentare del numero di lanci (figura 4.8).

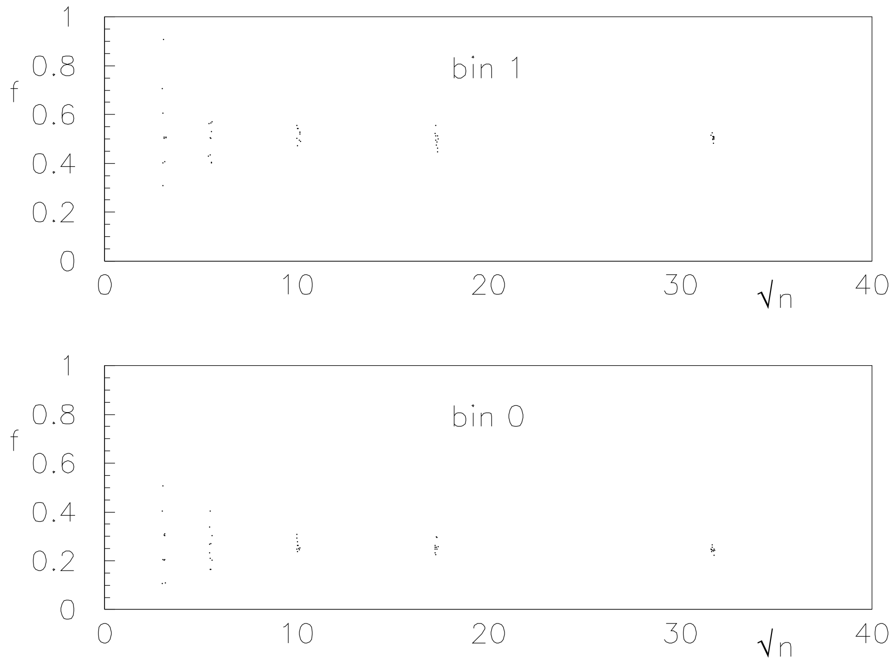

Figura 4.7: Dati dell'esperienza del pallinometro. Frequenza relativa con cui le palline cadono nella casella centrale e in quella laterale in funzione della radice quadrata del numero di palline lanciate.

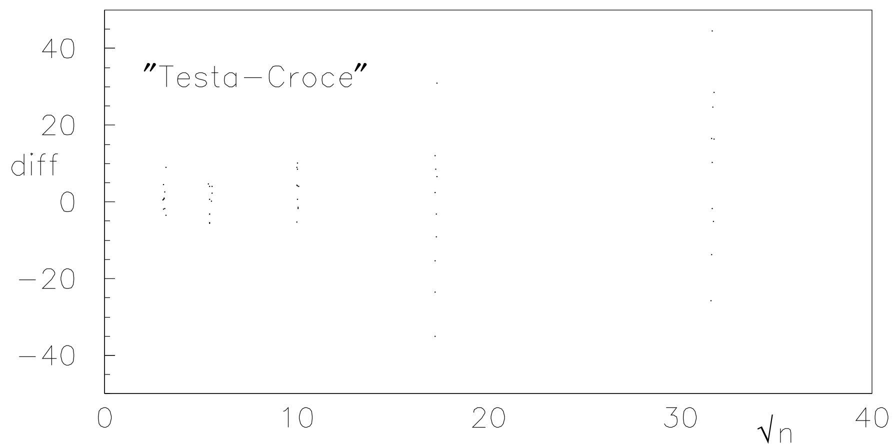

Figura 4.8: Esperienza del pallinometro. Differenza fra il numero di palline cadute nel bin centrale e quelle cadute altrove. È lo stesso andamento che si avrebbe se lanciando n volte una moneta si calcolasse la differenza fra il numero di teste e il numero di croci.

Capitolo 5

Descrizione quantitativa dei dati sperimentali

Finora abbiamo visto come ridurre in tabelle e grafici i dati sperimentali delle misure di conteggio. Occupiamoci ora di come sintetizzare in pochi numeri significativi, chiamati in gergo *riassunti statistici*, le informazioni raccolte. Tipicamente questi numeri sono legati a posizione e forma - soprattutto larghezza - dell'istogramma. Ad esempio, dalla figura 4.2 si vede come l'istogramma si sta "spostando" e "allargando" all'aumentare del tempo di misura.

5.1 Statistica descrittiva e statistica inferenziale

Prima di procedere cerchiamo di chiarire il termine *statistica* che viene usato in vari contesti e a volte anche come sinonimo di probabilità.

Senza voler entrare nei dettagli diciamo che bisogna distinguere le *statistiche*, di cui si parla in continuazione, come sinonimo di dati su aspetti sociali, economici, politici, geografici: "le statistiche dicono" che il ... % delle popolazione è ultrasessantenne, che questa è l'estate più calda degli ultimi n anni, che ... % di coppie divorziano nei primi 5 anni di matrimonio, e così via.

Con *statistica* si intende invece la disciplina che, in senso lato, si interessa della raccolta e dell'analisi dei dati e dell'interpretazione dei risultati. In particolare, la *statistica descrittiva* si occupa di descrivere la massa dei dati sperimentali con pochi numeri o grafici significativi. Quindi, per così dire si occupa di fotografare una data situazione e di sintetizzarne le caratteristiche salienti.[1] La *statistica inferenziale* utilizza i dati statistici, anche opportunamente sintetizzati dalla statistica descrittiva, per fare previsioni di tipo probabilistico su situazioni future o comunque incerte. Ad esempio esaminando un piccolo campione estratto da una grande popolazione cerca di valutare la frazione della popolazione che possiede una certa caratteristica, ha un certo reddito o voterà per un certo candidato.

[1]Per una interessante analisi sulla nascita della statistica, si veda I. Hacking, "Il caso domato", 1990 (il Saggiatore, 1994). Per esempio, a pag. 25 vi si legge la seguente il seguente passo scritto nel 1798 da J. Sinclair: "Per statistica si intende in Germania una ricerca fatta ai fini di accertare la forza politica di un paese, o i problemi relativi agli affari si stato."

Per quello che riguarda la teoria e la pratica delle misure, indubbiamente la parte di maggiore interesse è la statistica inferenziale in quanto lo scopo delle misure è quello di fare affermazioni sul valore di una grandezza a partire da un numero limitato di misure, effettuate con strumenti non ideali e con parametri e disturbi ambientali non noti con esattezza.

La parte di statistica descrittiva è importante perché nella maggior parte dei casi non è necessario conoscere in dettaglio tutti i dati sperimentali per inferire qualcosa, ma sono sufficienti pochi numeri nei quali i dati sono stati precedentemente sintetizzati.

5.2 Distribuzioni statistiche: notazioni

Introduciamo innanzitutto il concetto di *distribuzione statistica*. Le tabelle 4.1 e 4.2 e le figure 4.1 e 4.2 mostrano come sono *distribuiti* dei dati sperimentali di interesse, rispettivamente l'età dei cittadini tedeschi e i conteggi registrati in un certo intervallo di tempo. Abbiamo detto prima come le tabelle e gli istogrammi si ottengono dal conteggio delle occorrenze di ciascuna delle classi in cui sono stati raggruppate le informazioni.

Nei due casi illustrati per ogni *unità statistica* (il singolo cittadino tedesco o la singola misura) si era interessati ad un solo *carattere* (età e numero di conteggi, rispettivamente). In genere la scelta del carattere a cui ci si interessa può essere arbitraria e dipende dall'applicazione particolare. Ad esempio, trattandosi di studenti universitari si può essere interessati al tipo di maturità conseguita, alla città di provenienza, al sesso o al colore dei capelli. A volte i caratteri possono essere delle informazioni quantitative, come il voto riportato alla maturità, il numero di esami sostenuti, l'altezza o il peso. Questo *dato statistico* (quantitativo) può essere *discreto* (come voto ed esami sostenuti) o - almeno in principio - *continuo* (come altezza e peso). Altra ovvia considerazione è che, per ciascun carattere preso in considerazione, ogni *unità statistica* (l'individuo o il singolo risultato di una misura) appartenente ad una *popolazione* (la totalità degli individui o dei dati sperimentali) appartenga ad una ed una sola classe (lo studente non può aver conseguito 42 <u>e</u> 60 alla maturità), mentre non è vero il contrario (molti studenti possono avere avuto lo stesso voto). Non sempre i dati sono relativi all'intera popolazione: a volte si hanno dati relativi soltanto ad un *campione* della popolazione; altre popolazioni possono essere virtualmente infinite - come quella legata a tutte le misure possibili di una certa grandezza fisica.

Si dice che i dati statistici (quantitativi) costituiscono una *distribuzione statistica* quando a ciascuno di essi è associato il numero di volte con il quale si è verificato.

Introduciamo dei simboli per operare sulle distribuzioni statistiche.

- X è il nome generico della variabile di interesse associata alle unità statistiche.

- N è il numero totale di dati sperimentali (unità statistiche), ovvero la *dimensione*, o *numerosità* del campione o della popolazione.

- i è l'indice del generico dato sperimentale (unità statistica); quindi $i = 1, 2, \ldots, N$.
- N_c sta per il numero totale di classi.
- x_i il valore numerico associato all'i-ma unità statistica.
- k è l'indice della generica classe; quindi $k = 1, 2, \ldots, N_c$.
- x_k è il valore numerico associato alla classe k.
- n_k è il numero di *occorrenze* della classe k, ovvero la *frequenza* con cui si verifica il valore x_k; ne segue che

$$\sum_{k=1}^{N_c} n_k = N . \tag{5.1}$$

- f_k, definito da

$$f_k = \frac{n_k}{N} = \frac{n_k}{\sum_{k=1}^{N_c} n_k} , \tag{5.2}$$

  è la *frequenza relativa* della classe k; essa fornisce il *peso* (*relativo*) di tale classe rispetto alle altre ed è quindi chiamata anche *peso*, o *peso statistico* ed è indicata anche con w_k (come *weight*); vale ovviamente la condizione

$$\sum_{k=1}^{N_c} f_k = 1 \quad (\text{ovvero} \sum_{k=1}^{N_c} w_k = 1) . \tag{5.3}$$

  Il peso statistico del singolo dato sperimentale vale $w_i = 1/N$.

Al fine di alleggerire le formule eviteremo d'ora innanzi di scrivere esplicitamente gli estremi delle sommatorie, a meno che non ci siano ambiguità oppure, saltuariamente, per rinfrescare le convenzioni. Quindi saranno generalmente da intendersi le seguenti abbreviazioni:

$$\sum_{i=1}^{N} \longrightarrow \sum_{i} \tag{5.4}$$

$$\sum_{k=1}^{N_c} \longrightarrow \sum_{k} . \tag{5.5}$$

Nell'effettuare la classificazione dei dati sperimentali possono capitare due situazioni.

1. I valori numerici (x_k) delle classi sono tutti i valori numerici assunti dalle unità statistiche (x_i), come ad esempio le classificazioni delle tabelle 4.1 e 4.2.
2. I valori numerici delle classi sono inferiori a quelli delle unità statistiche. Questo succede se il numero di possibili classi è talmente elevato che è

Raggruppamento 1				
k	estremi	x_k	n_k	w_k
1	$0 \le t < 2$	1	30	0.30
2	$2 \le t < 4$	3	20	0.20
3	$4 \le t < 6$	5	15	0.15
4	$6 \le t < 8$	7	8	0.08
5	$8 \le t < 10$	9	10	0.10
6	$10 \le t < 12$	11	8	0.08
7	$12 \le t < 14$	13	2	0.02
8	$14 \le t < 16$	15	2	0.02
9	$16 \le t < 18$	17	3	0.03
10	$18 \le t < 20$	19	1	0.01
11	$20 \le t < 22$	21	1	0.01
Raggruppamento 2				
k	estremi	x_k	n_k	w_k
1	$0 \le t < 1$	0.5	19	0.19
2	$1 \le t < 2$	1.5	11	0.11
3	$2 \le t < 3$	2.5	13	0.13
4	$3 \le t < 4$	3.5	7	0.07
5	$4 \le t < 6$	5	15	0.15
6	$6 \le t < 8$	7	8	0.08
7	$8 \le t < 10$	9	10	0.10
8	$10 \le t < 14$	12	10	0.10
9	$14 \le t < 18$	16	5	0.05
10	$18 \le t < 22$	20	2	0.02

Tabella 5.1: Due modi di raggruppamento in classe dei valori dei tempi di attesa per un conteggio (tabella 1.2).

preferibile raggruppare più classi elementari[2] contigue. Viene così persa l'informazione degli esatti valori x_i acquistati dai dati sperimentali originali e per valore numerico della classe si prende semplicemente il punto medio di ciascuno degli intervalli che definiscono la classe. Si ha sempre questo secondo caso quando i valori numerici delle unità statistiche sono continui.

La seconda situazione avrà delle conseguenze sul valore numerico dei riassunti statistici, i quali differiranno da quelli che si ottengono dai dati non raggruppati. Comunque la differenza è in genere trascurabile, specialmente se si hanno molte classi e la distribuzione dei dati all'interno di esse è abbastanza "regolare" (vedi esempio nel paragrafo 5.8).

Come esempio eseguiamo una suddivisione in classi dei tempi di attesa per 1 conteggio (tabella 1.2). La tabella 5.1 mostra due modi per riunire in classi questi dati continui. Nel secondo raggruppamento le classi sono di diversa ampiezza. Sarà interessante vedere le variazioni dei risultati dovuti ai raggruppamenti in classi.

[2]"Elementare" sta per una classe all'interno della quale è impossibile - o irrilevante - differenziare l'unità statistica rispetto al carattere preso in esame.

5.3 Misure di posizione

È abbastanza comune la consuetudine che, volendo riassumere in un solo numero dei dati statistici, se ne dia il valore medio: media dei voti di uno studente, età media di un gruppo di persone, reddito medio, e così via. In effetti, la media aritmetica è la più semplice e conosciuta misura di posizione di una distribuzione statistica. Ricordiamo che essa è definita da

$$\overline{x} = \frac{1}{N}\sum_i x_i \,. \tag{5.6}$$

Nel caso di dati suddivisi in classi la sommatoria $\sum_{i=1}^{N} x_i$ si può riscrivere come somma dei valori numerici associati alle classi, ciascuno moltiplicato per il numero di volte che esso compare, ottenendo quindi

$$\overline{x} = \frac{1}{N}\sum_k n_k x_k = \frac{\sum_k n_k x_k}{\sum_k n_k} \tag{5.7}$$

$$= \sum_k f_k x_k \quad \left(\text{ovvero } \sum_k w_k x_k\right), \tag{5.8}$$

ovvero è una *media pesata* con la frequenza, o con la frequenza relativa, di ciascuna delle classi.

Se i dati sperimentali sono stati raggruppati senza unire classi elementari adiacenti è facile dimostrare che i due valori della media ottenuti dalle (5.6) e (5.7-5.8) sono esattamente uguali (è una semplice applicazione della proprietà commutativa e associativa della somma). Nel caso contrario essi sono soltanto approssimativamente uguali. L'approssimazione è tanto migliore quanto più grande è il numero di classi e quanto più regolarmente sono distribuiti i valori dei dati originali all'interno delle classi.

Per come è definita la media, può capitare che il suo valore numerico non coincida con nessuno dei valori dei dati sperimentali. Quindi "medio" non va inteso come quello che si verifica di più, ma semplicemente ... come medio[3].

Ci sono casi in cui si è effettivamente interessati ai valori che capitano più frequentemente. Si introduce allora il concetto di *moda* come il valore che si verifica più spesso, ovvero quello che ha peso statistico maggiore o, per dirlo in altri termini, "quello che va più di moda". Per come è definita si capisce che la moda può essere non unica. Si parla allora di *distribuzione multimodale*. Nei dati di conteggio per 3 e 30 secondi la moda vale rispettivamente 0 e 4.

Un altro modo per indicare la posizione di una distribuzione è di considerare il *valore centrale* - *mediano* - nel senso del valore rispetto al quale ci siano tanti valori più grandi quanti valori più piccoli. Tale valore è chiamato *mediana*. Per trovarlo si ordinano prima i dati e poi:

a) se il numero di dati N è dispari si prende come mediana il valore centrale

$$\text{mediana} = x_{\frac{N+1}{2}}\,; \tag{5.9}$$

[3] Molto istruttiva e divertente, a tale proposito, è la poesia di Trilussa in cui il poeta ironizza sul fatto che se uno mangia due polli e l'altro niente, "secondo la statistica" essi mangiano un pollo a testa.

b) se il numero di dati N è pari si prende come mediana la media aritmetica fra i due valori centrali:

$$\text{mediana} = \frac{x_{\frac{N}{2}} + x_{\frac{N}{2}+1}}{2} \qquad (5.10)$$

Nelle misure di conteggio per 3 e 30 secondi la mediana vale rispettivamente 0 e 5.

Le tre misure di posizioni introdotte non sono utilizzate una in alternativa dell'altra, anzi molto spesso la conoscenza simultanea delle tre fornisce delle utili informazioni sulla distribuzione. Per esempio si immagini tre comunità il cui reddito medio sia di 1000$ l'anno, ma i cui valori di moda e mediana siano:

1. moda = mediana = 1000$;
2. moda = 100$, mediana = 100$;
3. moda = 100$, mediana = 800$.

Anche se da questi tre soli valori è impossibile risalire alle esatte distribuzioni di reddito, essi offrono tre quadri della popolazione completamente diversi. La prima mostra una più equa ripartizione del reddito, mentre la seconda mostra il maggiore squilibrio fra ricchi e poveri.

Anche se sono state introdotte per completezza, la moda e la mediana, utili per molte applicazione di statistica descrittiva, in questo testo faremo uso soltanto della media.

5.4 Misure di dispersione

Ci sono molti modi possibili per quantificare la "larghezza" di una distribuzione statistica. Il più semplice è quello di fornire il *campo di variabilità* dei dati sperimentali, ovvero il valore minimo e quello massimo assunti dalle unità statistiche[4]. Questa misura può però dipendere da occasionali valori eccezionali che non danno l'idea dell'effettiva dispersione dei dati sperimentali intorno al loro centro. Ad esempio, nei dati del contatore per 300 s il campo di variabilità è compreso fra 33 e 78, da cui risulterebbe che la distribuzione è larga 45. Però tali valori estremi sono alquanto isolati ed è facile immaginare che in un'altra misura effettuata nelle stesse condizioni si possano ottenere grandi variazioni da 45. In altre parole è interessante trovare delle misure di dispersione del "tipo" di distribuzione e non soltanto di quella effettivamente osservata[5].

Un altro problema del campo di variabilità è che esso dipende dalla quantità dei dati osservati. Se prendiamo in considerazione i primi 20 valori dei dati

[4] Un derivato del campo di variabilità è la *semidispersione massima*, definita come metà del campo di variabilità.

[5] È chiaro che la scelta della grandezza opportuna non è semplicemente parte della statistica descrittiva, ma in qualche modo è legato alla statistica inferenziale, cioè al problema di astrarre informazioni di carattere generale con cui fare eventuali predizioni per le osservazioni future a partire dai dati osservati. Questo argomento sarà trattato al momento opportuno. Nel paragrafo 5.13 è discusso un noto caso di interferenza fra statistica descrittiva e statistica inferenziale

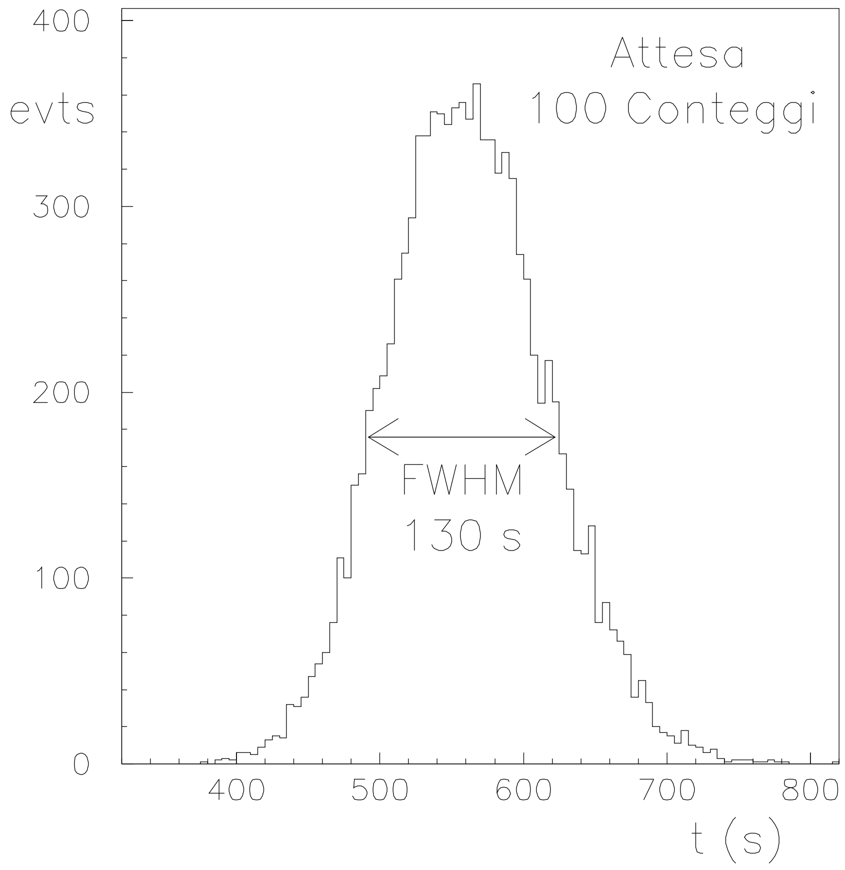

Figura 5.1: Esempio di larghezza a metà altezza (full width half maximum).

da 300 s l'intervallo è quasi la metà di quello ottenuto per 100 valori. Si può immaginare ragionevolmente che più dati raccogliamo e più è facile che prima o poi compaiano valori sempre più lontani dalla media. È opportuno quindi scegliere altre misure di larghezza più legate alla forma e meno dipendenti dalla *scala* della distribuzione, ovvero dal numero totale di dati raccolti. Si può utilizzare ad esempio l'intervallo centrato intorno alla media che contiene una certa percentuale dei dati sperimentali (50 o 90 %), oppure la larghezza della distribuzione misurata ove i pesi statistici sono una certa frazione (5, 10, 50 %) della frequenza massima.

Gli statistici usano molto i *quantili* e *percentili* che indicano la percentuale di dati sperimentali compresi in un certo intervallo, ma queste tecniche vanno al di là del nostro interesse e non presentano nessuna difficoltà di comprensione.

Delle varie misure grafiche di dispersione vale la pena di accennare alla *larghezza a metà altezza*, in inglese *full width half maximum* (FWHM), che permette di misurare la larghezza di una distribuzione da un istogramma senza dover fare conti. Per come è definità essa è anche poco sensibile a eventuali code estreme della distribuzione e soprattutto ad eventi di fondo casuale (cioè non derivanti dal segnale fisico di interesse) che si verifica a volte in modo circa uniforme estendendosi anche alcuni ordini di grandezza intorno ai valori tipici del segnale, inficiando le altre misure di dispersione (in questo caso anche le usuali misure di posizione perdono valore e bisogna procedere un po' a occhio e un po' cercando di modellizzare segnale e fondo, ma per ora non ci occupiamo di questi problemi).

La figura 5.1 mostra un esempio di valutazione di FWHM sulla distribuzione dei tempi di attesa per avere 100 conteggi, già riportata in figura 4.3. Si noti come a prescrizione di "larghezza a metà altezza" in pratica non viene presa alla lettera, ma smussando a occhio la distribuzione.

La metà della FWHM è chiamata *"semilarghezza a metà altezza"* ed è indicata come HWHM (*half width half maximum*).

La misura di dispersione più utile in assoluto e di uso universale è comunque quella legata agli scarti dei dati sperimentali rispetto alla loro media e sarà trattata nel prossimo paragrafo

5.4.1 Varianza e deviazione standard

Più la distribuzione è "larga" e più è frequente trovare grandi scarti $d_i = x_i - \overline{x}$, sia positivi che negativi. Quindi, come la media può essere utilizzata per valutare la posizione tipica dei dati sperimentali, così la media degli scarti potrebbe essere utilizzata come misura della dispersione dei dati. Ma, poiché gli scarti positivi sono compensati da scarti negativi, la media degli scarti, presi con il loro segno, è nulla. Si potrebbe ovviare a questo inconveniente considerando la media del modulo degli scarti, ma per comodità si preferisce calcolare la *media dei quadrati degli scarti* e prendere poi la radice quadrata (positiva). Queste due grandezze sono chiamate rispettivamente *varianza* e *deviazione standard* (o *scarto quadratico medio*). I simboli utilizzati in questo testo sono Var e σ^2 per la varianza, σ per la deviazione standard[6]. Dalla definizione abbiamo quindi

$$\sigma^2(X) = \text{Var}(X) = \frac{1}{N}\sum_i d_i^2 = \frac{1}{N}\sum_i (x_i - \overline{x})^2 \tag{5.11}$$

[6]La deviazione standard come è qui definita è spesso indicata con σ_N o s_N o nelle calcolatrici tascabili. In molti testi il simbolo σ sarà riservato alla varianza delle distribuzioni di probabilità e s al suo stimatore, definito tipicamente come $s^2 = \sum_i (x_i - \overline{x})^2/(N-1)$. Come sarà commentato nel seguito (vedi paragrafi i bisogna fare attenzione a non confondere problemi di statistica descrittiva e problemi di statistica inferenziale. Inoltre bisogna anche abituarsi ad una certa flessibilità. Per questo motivo indicheremo con σ le deviazioni standard di distribuzioni sia statistiche che di probabilità, a meno che non sorgano ambiguità. Lo stesso discorso vale per il seimbolo 'Var'.

ovvero, se i dati sono riuniti in classi,

$$\sigma^2(X) = \mathrm{Var}(X) = \frac{1}{N}\sum_k n_k \cdot d_k^2 \quad = \quad \frac{1}{N}\sum_k n_k \cdot (x_k - \overline{x})^2 \qquad (5.12)$$

$$= \frac{\sum_k n_k \cdot (x_k - \overline{x})^2}{\sum_k n_k} \qquad (5.13)$$

$$= \sum_k f_k \cdot (x_k - \overline{x})^2 \qquad (5.14)$$

$$= \sum_k w_k \cdot (x_k - \overline{x})^2 \qquad (5.15)$$

e

$$\sigma = \sqrt{\sigma^2}\,. \qquad (5.16)$$

Vedremo nel paragrafo 5.6 il modo più opportuno per calcolare la varianza.

5.5 Analogia meccanica di media e varianza

Si noti come la formula della media pesata (5.7) ricordi la formula del centro di massa di un corpo esteso, se si pensa che ogni classe posta in x_k abbia una massa n_k, ovvero un peso (relativo) w_k. Anche la formula (5.6) per dati non raggruppati ha lo stesso significato per il caso di N punti di massa unitaria. Per questa ragione la media aritmetica è spesso chiamata *baricentro della distribuzione*. Questa osservazione è anche importante per stimare ad occhio la media di dati istogrammati: essa è localizzata nel punto dell'asse delle ascisse rispetto al quale l'istogramma, visto come un corpo rigido bidimensionale, si manterrebbe in equilibrio se sottoposto ad una forza gravitazionale.

Anche la varianza ha una semplice interpretazione meccanica, non essendo altro che il *momento di inerzia* del corpo rispetto al baricentro, come si vede dalla (5.15).

Da queste analogie meccaniche è possibile ottenere interessanti proprietà della media e della deviazione standard.

1. La media è il valore rispetto al quale è minima la somma dei quadrati degli scarti.

2. Se tutti i punti sono traslati in egual modo anche la media è traslata della stessa quantità.

3. La varianza - e quindi la deviazione standard - è invariante per traslazione.

4. Se tutti i punti vengono moltiplicati per lo stesso *fattore di scala*, media e deviazione standard variano dello stesso fattore.

5. Il momento di inerzia rispetto al punto $x = 0$ è pari al momento di inerzia rispetto al baricentro più il quadrato della posizione del baricentro moltiplicata per la massa totale (unitaria) dei punti sperimentali. Ne

segue la seguente propietà:

$$\sum_k w_k x_k^2 = \sum_k w_k (x_k - \overline{x})^2 + \overline{x}^2 \sum_k w_k = \sum_k w_k (x_k - \overline{x})^2 + \overline{x}^2 \,. \tag{5.17}$$

Se si considerano dati sciolti, ciascuno di massa unitaria, si ottiene

$$\sum_i x_i^2 = \sum_i (x_i - \overline{x})^2 + N\overline{x}^2 \,. \tag{5.18}$$

Dalle (5.17) e (5.18) (dividendo quest'ultima per N) si ottiene che

$$\overline{x^2} = \text{Var}(X) + \overline{x}^2 \,, \tag{5.19}$$

dove con $\overline{x^2}$ si è indicata la media dei quadrati. Ne segue che

$$\text{Var}(X) = \overline{x^2} - \overline{x}^2 \,. \tag{5.20}$$

Queste proprietà possono molto utili per semplificare i calcoli e saranno viste formalmente nel prossimo paragrafo.

5.6 Proprietà di media e varianza

Innanzitutto verifichiamo le proprietà di media e varianza trovate dalle analogie meccaniche, seguendo lo stesso ordine dei punti del paragrafo precedente.

1. La somma dei quadrati degli scarti rispetto al generico punto x è pari a $\sum_i (x_i - x)^2$. Dalla condizione di minimo

$$\frac{d}{dx} \sum_i (x_i - x)^2 = 0 \tag{5.21}$$

$$2 \sum_i x_i - 2Nx = 0 \,, \tag{5.22}$$

   da cui segue che $x = \sum_i x_i / N = \overline{x}$.

2. Se $x_i' = x_i + c$, allora la nuova media sarà

$$\overline{x'} = \frac{1}{N} \sum_i x_i' = \frac{1}{N} \sum_i (x_i + c) \tag{5.23}$$

$$= \overline{x} + c \,. \tag{5.24}$$

3. Analogalmente per la varianza otteniamo

$$\sigma^2(X') = \frac{1}{N} \sum_i (x_i + c - (\overline{x} + c))^2 = \sigma^2(X) \,. \tag{5.25}$$

4. In modo analogo, è semplice dimostrare le proprietà di media e deviazione standard per un cambiamento di scala e viene lasciato come esercizio;

5. Infine, la proprietà $\text{Var}(X) = \overline{x^2} - \overline{x}^2$ sarà ripresa nel prossimo paragrafo.

In generale si ha quindi:

$$\overline{ax+c} = a\,\overline{x} + c \tag{5.26}$$
$$\sigma^2(aX+c) = a^2\sigma^2(X) \tag{5.27}$$
$$\sigma(aX+c) = |a|\,\sigma(X)\,. \tag{5.28}$$

Queste proprietà formali possono essere utili nel fare i conti. Infatti:

- se si deve calcolare media e deviazione standard di valori in cui variano soltanto le ultime cifre (ad esempio: 102.3, 104.8, 111.3, 105.5) è molto più rapido e sicuro considerare solo le cifre fluttuanti (nell'esempio riportato sottraendo 100) e aggiungere la parte costante alla fine sulla sola media;

- se una grandezza è ottenuta da un'altra mediante una trasformazione lineare (ad esempio la conversione da tensione a temperatura di una termocoppia) e sono note media e deviazione standard di una delle grandezze si ottengono i valori relativi all'altra senza dover rianalizzare i singoli valori di temperatura.

5.7 Valutazione pratica della deviazione standard

Per quanto riguarda il calcolo della deviazione standard, la sua definizione potrebbe far supporre che bisognerebbe trovare prima la media, poi calcolare gli N scarti d_i, farne il quadrato, e così via. Questa operazione è chiaramente lunga e macchinosa e, anche se effettuata con un calcolatore, richiederebbe di archiviare in memoria tutti i valori di x_i per poterli utilizzare nel calcolo degli scarti. In realtà è possibile utilizzare una procedura più snella utilizzando la proprietà del punto 5 del paragrafo 5.5, che ridimostriamo come esercizio:

$$\sigma^2(X) = \frac{1}{N}\sum_i (x_i - \overline{x})^2 \tag{5.29}$$
$$= \frac{1}{N}\sum_i \left(x_i^2 - 2\overline{x}x_i + \overline{x}^2\right) \tag{5.30}$$
$$= \frac{1}{N}\left(\sum_i x_i^2 - 2\overline{x}\sum_i x_i + N\overline{x}^2\right) \tag{5.31}$$
$$= \frac{\sum_i x_i^2}{N} - 2\overline{x}^2 + \overline{x}^2 \tag{5.32}$$
$$= \overline{x^2} - \overline{x}^2\,, \tag{5.33}$$

dove è stata indicata con $\overline{x^2} = \sum_i x_i^2/N$ la *media dei quadrati*. Quindi si è trovato che *"la varianza è pari alla media dei quadrati meno il quadrato della media"*. Ricordandosi che il peso statistico dell'i-mo evento è pari a $w_i = 1/N$ si vede come la quantità $\frac{\sum_i x_i^2}{N} = \sum_i w_i x_i^2$ è pari al momento di

i	N	$\sum x$	$\sum x^2$	$\overline{x}$	$\overline{x^2}$	σ
1-50	50	24	42	0.48	0.84	0.85
51-100	50	36	60	0.72	1.20	0.83
1-100	100	60	102	0.60	1.02	0.81

Tabella 5.2: Calcolo dettagliato di medie e varianze dai dati di conteggio per 3 secondi (tabella 4.1).

inerzia della distribuzione rispetto all'origine delle ascisse, come era già stato discusso nel paragrafo 5.5

È quindi preferibile calcolare, contemporaneamente a $\sum_i x_i$ anche $\sum_i x_i^2$ e dalle (5.33) ottenere media e varianza. Questa è la tecnica utilizzata anche nei programmi al calcolatore e nelle calcolatrici tascabili con funzioni statistiche. In queste ultime appositi tasti, generalmente contrassegnati da $\boxed{\sum x}$ e $\boxed{\sum x^2}$, permettono di leggere il valore raggiunto dopo che è stato inserito l'i-mo dato. Queste funzioni possono tornare utili per calcolare valori parziali e totali di media e deviazione standard di una lunga serie di misure senza dover inserire di nuovo tutti i numeri. Infatti le sommatorie parziali possono essere annotate e sommate fra di loro per combinare vari gruppi di dati. Questo è particolarmente importante quando si ha una lunga serie di valori: è preferibile annotare di tanto in tanto i risultati parziali in modo da non dover ricominciare da capo se si commette un errore.

Come esempio consideriamo i dati del contatore per 3 secondi e valutiamo media e deviazione standard della distribuzione e delle due distribuzioni formate dividendo i dati in due campioni. I risultati sono in tabella 5.2. Per l'intero campione svolgiamo in dettaglio i conti

$$\begin{aligned}
\sum x &\equiv \sum_k n_k x_k = 56\cdot 0 + 32\cdot 1 + 9\cdot 2 + 2\cdot 3 + 1\cdot 4 = 60 \\
\sum x^2 &\equiv \sum_k n_k x_k^2 = 56\cdot 0 + 32\cdot 1 + 9\cdot 4 + 2\cdot 9 + 1\cdot 16 = 102 \\
\overline{x} &= \frac{\sum x}{N} = 0.60 \\
\overline{x^2} &= \frac{\sum x^2}{N} = 1.02 \\
\sigma^2 &= 1.02 - 0.60^2 = 0.66\,.
\end{aligned} \tag{5.34}$$

5.8 Effetto del raggruppamento in classi

Come accennato precedentemente, i valori dei riassunti statistici per dati raggruppati possono differire numericamente da quelli calcolati dai dati individuali, in quanto vengono usati i centri degli intervalli (x_k) e non i valori originali (x_i). In genere l'effetto è trascurabile se le classi sono ben popolate e le distribuzioni "abbastanza regolari". Ovviamente gli effetti più importanti possono

dati statistici	$\overline{x}$	σ
dati sciolti	5.28	4.72
Raggruppamento 1	5.46	4.72
Raggruppamento 2	5.44	4.83

Tabella 5.3: Dipendenza della media e varianza dal raggruppamento dei dati.

venire dai punti isolati sulle code della distribuzione. Questi andrebbero trattati quindi con cautela e possibilmente individualmente. Comunque, quando è richiesta un'alta precisione nei risultati è conveniente effettuare i conti a partire dai dati originali. Come esempio riportiamo in tabella 5.3 media e deviazione standard dei tempi di attesa per 1 conteggio (tabella 1.2) utilizzando i dati sciolti o quelli dei due raggruppamenti della tabella 5.1. In questo caso l'asimmetria della distribuzione produce un risultato apprezzabile. Questo è dovuto al fatto che all'interno di ciascuna classe la media dei dati è sempre inferiore al centro dell'intervallo in quanto i numeri piccoli sono più frequenti di quelli grandi.

5.9 Dispersione relativa e coefficiente di variazione

Gli istogrammi di figura 4.2 mostrano un chiaro allargamento della distribuzioni in funzione del tempo. Questo può essere quantificato con le misure di dispersione viste nel paragrafo 5.4. Però, come è noto dalla vita quotidiana, la dispersione assoluta non è sempre il miglior criterio per stabilire se ci sono fluttuazioni rispetto ai valori tipici. Immaginiamo che una indagine mostri che il prezzo di una scatola di pelati sia intorno a 800 lire, ma con un campo di variazione compreso fra 400 e 1500 lire, mentre lo champagne varia fra 32˙000 e 45˙000 lire con un prezzo medio di 40˙000. Si concorderà nel dire che le fluttuazioni sul prezzo dei pelati è superiore di quello dello champagne sebbene il campo di variabilità di quest'ultimo sia di un ordine di grandezza superiore a quello dei pelati (se i prezzi di una moto variano al più di 10˙000 lire da un concessionario all' altro, diciamo, giustamente, che sono "esattamente uguali"). Infatti, spesso siamo interessati alle *fluttuazioni relative*. Esse sono usualmente quantificate in termini della dispersione rapportata ai valori tipici ed è immaginabile pensare alla grande varietà di modi con cui può essere espressa la *dispersione relativa*. Accenniamo soltanto al *coefficiente di variazione*, definito a partire da media e deviazione standard:

$$v = \frac{\sigma}{|\overline{x}|}\,, \tag{5.35}$$

espressa a volte anche in percentuale. Il modulo serve a rendere tale misura di larghezza relativa indipendente dal segno della media (come lo è già la deviazione standard). Ad esempio, la distribuzione dei conteggi per $T = 3\,\mathrm{s}$ ha un coefficiente di variazione pari 1.35, o del 135 %, mentre quella per $T = 300\,\mathrm{s}$ lo ha del 14 %. Si dice quindi che la seconda distribuzione è relativamente più *stretta* della prima.

5.10 Misure di dispersione e incertezza della misura - caveat

Anche se l'incertezza della misura è un problema di statistica inferenziale che verrà trattato nel seguito, è opportuno premettere un invito alla cautela rivolto a coloro che pensano di essere già in grado di esprimere l'incertezza di misura dalla conoscenza delle misure di dispersione. Infatti si potrebbe credere che, siccome posizione e dispersione possono condensare le informazioni dei dati, esse possono essere utilizzate tout court per fornire il risultato dell'esperimento

Soffermiamoci sulle misure di radioattività con un tempo di 300 s. Dalla media e deviazione standard del numero di conteggi (tabella 5.4) e utilizzando le proprietà sotto una trasformazione lineare (ricordiamo che la nostra definizione di radioattività è "$r = \#$conteggi/T") si trova $\overline{r} = 0.178$ cont/s e $\sigma(r) = 0.025$ cont/s. Ovvero si potrebbe inferire frettolosamente che il risultato dell'esperimento sia che r è uguale a 0.178 "piu o meno" 0.025 conteggi al secondo (si è scritto "più o meno" secondo l'uso del linguaggio colloquiale e non "$\pm$" per indicare che l'espressione andrebbe meglio precisata, ma per ora questa può essere considerata una finezza). Ripetendo l'esercizio per le misure da 3 s si ottiene 0.20 "più o meno" 0.27. Cosa significa? Compatibilità con radioattività negativa?

Riprendiamo i risultati per 300 s. Questi mostrano una distribuzione abbastanza simmetrica e quindi si prestano a ragionamenti di simmetria. Cerchiamo di capire quello che si può ragionevolmente inferire dalla conoscenza di $\overline{x}$ e σ e quello che è invece una illazione sconsiderata. Le informazioni descrittive dicono semplicemente che nelle 100 misure i valori di radioattività sono distribuiti intorno a 0.178 cont/s (o intorno a 0.177 cont/s se si fosse considerata la mediana) con deviazioni (quadratiche) medie di 0.025 cont/s. Quando invece si presenta il risultato della misura si vuole affermare qualcosa sul "valore vero" della radioattività.

L'intuizione secondo cui "è più facile che il valore vero si trovi intorno alla media che altrove" è ragionevole, almeno in questo caso.

Per quanto concerne la deviazione standard, essa ci quantifica la dispersione delle singole misure intorno alla media e quindi, in prima approssimazione, intorno al valore vero. Quindi invertendo il ragionamento si potrà dire, al più, che il valore vero è lontano "mediamente" (media quadratica) di 0.025 cont/s dal risultato ottenuto da ogni singola misura da 300 s. Quindi da questa sola informazione (senza conoscere il modello statistico del contatore) ed avendo effettuato una sola misura possiamo affermare che, prendendo ad esempio la prima misura di tabella 1.1 per 300 s, r è uguale a 0.183 "più o meno" 0.025 cont/s. In effetti questo risultato è compatibile con la media delle 100 misure. Quindi questo caso semplice è simile a quello relativo alla lettura di una scala analogica considerato nel paragrafo 2.2. Ma anche allora si era fatto cenno alla non banale estensione ad un insieme di misure.[7]

Per concludere, si può dire che

[7] Vedremo come il caso del contatore sia particolarmente istruttivo in quanto mostra che in certi casi non è necessaria alcuna valutazione di dispersione - preliminare o dai dati spessi - per la stima dell'incertezza. Una sola misura, accompagnata da un modello probabilistico del sistema, è sufficiente a fornire informazioni esaurienti ai fini del risultato.

- è senz'altro lecito affermare che

  i valori di radioattività ottenuti nelle 100 misure da 300 s sono distribuiti intorno a 0.178 "più o meno" 0.025 cont/s,

- ma è fuorviante sostenere che

  i risultati delle 100 misure da 300 s indicano un valore vero di radioattività pari a 0.178 "più o meno" 0.025 cont/s.

5.11 ❄ Altre misure di forma

Come il coefficiente di variazione offre un modo per riassumere in un numero la larghezza relativa (alla media) della distribuzione, così a volte si può essere interessati a quantificare altre caratteristiche di forma della distribuzione. Quelle di maggiore interesse e uso sono la asimmetria e la dispersione rapportata al campo di variabilità (invece che alla media).

Per mettere in evidenza eventuali asimmetrie si può utilizzare la media delle potenze dispari degli scarti, con esclusione del primo ordine che sappiamo essere identicamente nulla. Per semplicità si sceglie il cubo degli scarti. Poiché la potenza è maggiore di 1 vuol dire che scarti lontani hanno importanza maggiore ai fini della media. Ne segue che se la media dei cubi degli scarti è positiva significa che la distribuzione ha una coda verso "destra" (per valori maggiori della media), se è negativa ha una coda verso sinistra. Per ottenere una misura di asimmetria con la quale poter confrontare distribuzioni che si estendono su ordini di grandezza diversi e che non dipenda dall'unità di misura usata è conveniente esprimere la media dei cubi degli scarti in una unità che sia "naturale" per la distribuzione di interesse. Questa *scala naturale* è scelta essere il cubo della deviazioni standard.

Il cubo degli scarti in unità del cubo della deviazione standard è chiamato *skewness* (skew in inglese significa infatti sbilenco, ma "sbilenchità" suona decisamente male[8]...):

$$\begin{aligned} \text{skewness} &= \frac{1}{\sigma^3}\frac{\sum_i (x_i-\overline{x})^3}{N} && (5.36)\\ &= \frac{1}{\sigma^3}\frac{\sum_k n_k(x_k-\overline{x})^3}{N} && (5.37)\\ &= \frac{\sum_k w_k(x_k-\overline{x})^3}{\sigma^3} && (5.38)\\ &= \frac{m_3}{\sigma^3}\,, && (5.39)\end{aligned}$$

dove con m_3 è stato indicato il *momento terzo* rispetto alla media.[9]

[8] Ad esempio, la versione italiana di EXCEL della Microsoft traduce skewness con "asimmetria", ma a quanto mi risulta gli statistici italiani preferiscono decisamente skewness.

[9] In generale, il momento k-mo rispetto a c è definito come $\sum_i (x_i - c)^k/N$.

Svolgiamo i conti in dettaglio per il caso del contatore per $T = 3\,\mathrm{s}$:

$$\begin{aligned}
\sum_k n_k(x_k-\overline{x})^3 &= 56\cdot(-0.226)+32\cdot 0.064+9\cdot 2.74+2\cdot 13.82 \\
&\quad +1\cdot 39.304 \\
&= -12.7+2.0+24.7+27.6+39.3 \;=\; 80.9 \\
m_3 &= 0.809 \\
\frac{m_3}{\sigma^3} &= 1.52\,.
\end{aligned}$$

Il valore positivo è in accordo con l'ispezione visiva dell'istogramma, che mostra la coda verso destra più pronunciata di quella a sinistra.

L'ultima misura di forma che incontriamo è la *curtosi* (in inglese *kurtosis*). Essa indica l'acutezza della distribuzione, ovvero se la forma ricorda più un "picco aguzzo" o una specie di "altopiano" (a parte le ovvie "irregolarità"). Essa è costruita da una media della quarta potenza degli scarti, opportunamente scalata alla quarta potenza della deviazione standard:

$$\begin{aligned}
\text{curtosi} &= \frac{1}{\sigma^4}\frac{\sum_i (x_i-\overline{x})^4}{N} && (5.40)\\
&= \frac{1}{\sigma^4}\frac{\sum_k n_k(x_k-\overline{x})^4}{N} && (5.41)\\
&= \frac{\sum_k w_k(x_k-\overline{x})^4}{\sigma^4} && (5.42)\\
&= \frac{m_4}{\sigma^4}\,. && (5.43)
\end{aligned}$$

Anche qui m_4 sta per il *momento quarto* rispetto alla media. La media della quarta potenza da molto più peso alle grandi deviazioni della varianza. Quindi, qualitativamente, se il valore della kurtosis è "grande" significa che si verificano molti valori in un intorno della media dell'ordine della deviazione standard, più code molto distanti dalla media[10] che danno più contributo a m_4 che a σ^4. Ad esempio le distribuzioni costituite dai dati sperimentali $\{1, 2, 3, 4, 5, 6, 7, 8, 9\}$ e $\{0.65, 4, 5, 5, 6, 9.35\}$ hanno stessa media e deviazione standard ($\overline{x} = 5$, $\sigma = 2.58$) e sono simmetriche (skewness $= 0$), ma differiscono per la curtosi (1.8 e 2.8 rispettivamente).

Per capire l'ordine di grandezza dei valori della curtosi si pensi che se la forma è "a campana" regolare (quella della la famosa distribuzione di probabilità gaussiana) la curtosi vale 3; se è piu aguzza si hanno valori maggiori, se più appiattita minori. Si noti invece come il valore della curtosi non dipende dalla larghezza della campana, in quanto il momento quarto è già rapportato alla larghezza misurata dalla deviazione standard.

Un'osservazione più generale è che dalle definizioni si verifica facilmente che sia skewness che curtosi sono invarianti per traslazione e per cambiamenti di scala.

[10]Nel calcolo della deviazione standard i dati sulle code devono compensare con l'"ampio braccio di leva rispetto al baricentro" l'alta frequenza di dati intorno al centro della distribuzione.

conteggi (cfr. figura 4.2)				
T (s)	$\overline{x}$ (cont.)	σ (cont.)	skewness	curtosi
3	0.60	0.81	1.52	5.57
6	1.02	0.93	0.71	3.05
12	2.19	1.35	0.31	2.51
30	5.16	2.15	0.38	3.05
100	17.81	4.25	0.28	3.14
300	53.48	7.52	0.36	3.56
tempi di attesa (cfr. figura 4.3)				
Nr. di conteggi	$\overline{x}$ (s)	σ (s)	skewness	curtosi
1	5.57	5.59	2.01	8.94
2	11.13	8.83	1.43	6.06
5	28.14	12.65	0.92	4.26
100	561.54	55.53	0.20	3.05

Tabella 5.4: Misure di posizione, dispersione, asimmetria e di acutezza delle distribuzioni di conteggi e di tempi di attesa (figure 4.2 e 4.3).

Per il calcolo pratico di m_3 e m_4 si utilizzano, in analogia al caso della varianza, le seguenti proprietà:

$$m_3 = \overline{x^3} - 3\overline{x^2}\,\overline{x} + 2\overline{x}^3 \tag{5.44}$$

$$m_4 = \overline{x^4} - 4\overline{x^3}\,\overline{x} + 6\overline{x^2}\,\overline{x}^2 - 3\overline{x}^4. \tag{5.45}$$

Si utilizzano quindi le sommatorie $\sum x$, $\sum x^2$, $\sum x^3$ e $\sum x^4$.

La tabella 5.4 riporta la skewness e la curtosi delle distribuzioni del contatore mostrate nelle figure 4.2 e 4.3. Per completezza sono state riportate anche media e deviazione standard. (L'abbondanza delle cifre significative è giustificata soltanto da esigenze di formattazione dei valori.) Si vede come l'ultima distribuzione di figura 4.3 ha raggiunto una forma di campana regolare (curtosi ≈ 3).

5.12 ❄ Misure di correlazione

A volte i dati statistici si presentano a coppie (o a gruppi logici di dimensione maggiore) in quanto possono essere stati ottenuti contemporaneante nella stessa misura o sono semplicemente due caratteri della stessa unità statistica (coppie tempo e posizione di un corpo; temperatura e allungamento di una barretta; voto alla maturità e media dei voti all'Università di uno studente; altezza e peso di una persona; e così via). Non è necessario che le coppie abbiano questo legame "forte" attraverso un processo pressoché simultaneo di misura o si riferiscano alla stessa persona, ma possono essere anche più generali (temperatura registrata in due città diverse; temperatura massima estiva e temperatura minima invernale in una certa città; altezza del padre e quella del figlio; indice

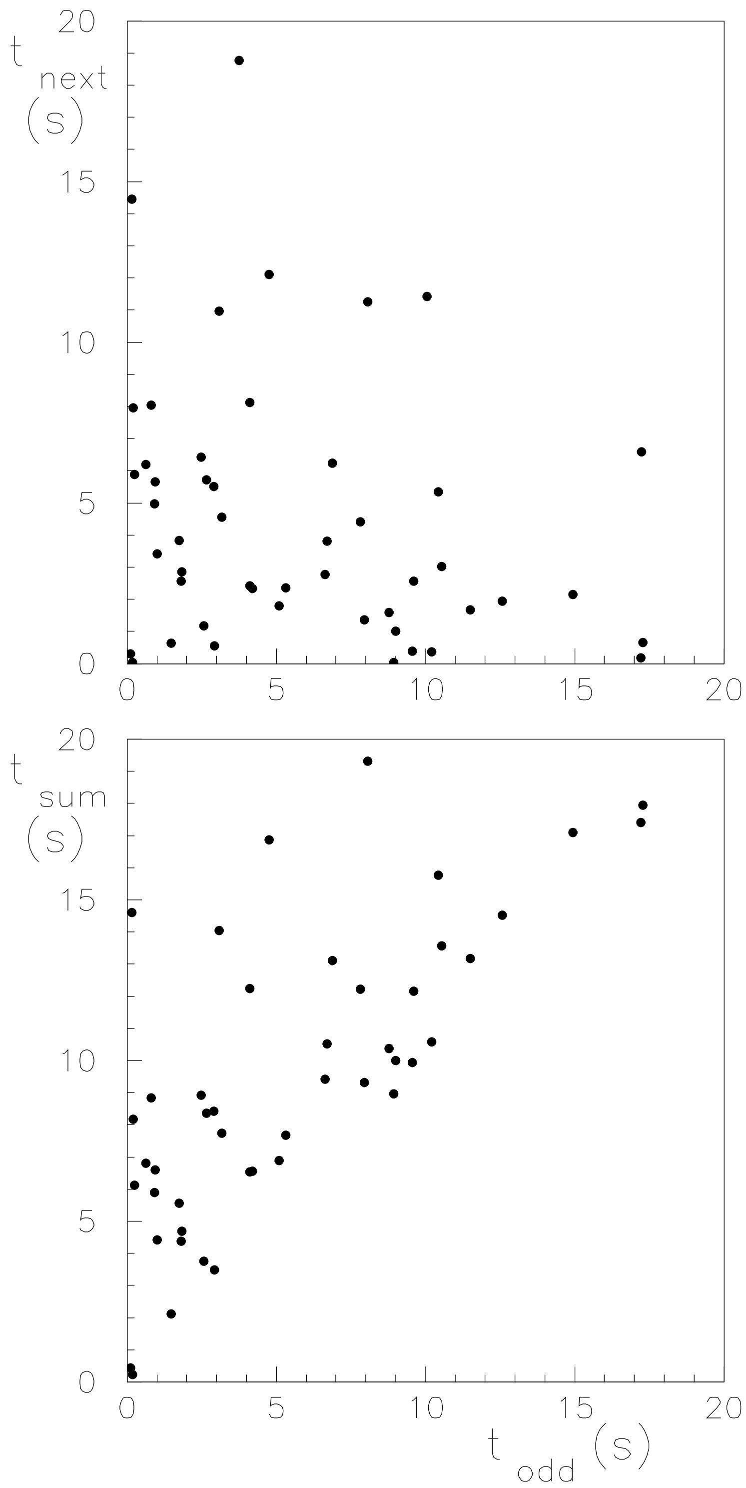

Figura 5.2: Tempi di attesa per registrare un conteggio (tabella 1.2. Lo scatter plot in alto mostra la distribuzione congiunta del risultato di una misura dispari (t_{odd}) e della misura immediatamente seguente (t_{next}). In quello in basso sull'asse delle ordinate si ha invece la somma del risultato della misura dispari e della pari successiva (t_{sum}).

della borsa di Milano e livello dell'acqua a Venezia). La scelta è basata sul sospetto che esse possano essere in qualche modo legate fra di loro.

L'estensione a gruppi di n caratteri quantitativi (grandezze fisiche nelle nostre applicazioni di laboratorio), chiamati *"n-tuple"*, è abbastanza automatica e quindi ci limitiamo a considerare il caso *bidimensionale* e chiamiamo le due grandezze x e y.

Per ciascun carattere si possono calcolare tutte le misure di posizione, dispersione e forma che abbiamo incontrato, semplicemente considerandone uno alla volta. Questa operazione corrisponde quindi a proiettare la distribuzione bidimensionale in due distribuzioni unidimensionali.

Ne segue che le suddette misure di forma unidimensionali non sono sensibili a caratteristiche legate alla struttura bidimensionale dei dati. In particolare si perdono le correlazioni fra le due grandezze, ovvero la preferenza di y ad assumere certi particolari valori per ciascuno dei valori della x. Ad esempio se le due città in cui si misura la temperatura sono vicine, o almeno sullo stesso emisfero, quando è caldo in una farà mediamente caldo anche nell'altra, e analogalmente per il freddo.

Il modo migliore per studiare le eventuali correlazioni fra i dati è di osservarli su uno scatter plot o con altra rappresentazione grafica opportuna (un normale grafico se c'è un solo punto per ogni coppia di possibili valori $[x, y]$) e considerare se è *ragionevole* che ci sia una dipendenza funzionale fra una grandezza e l'altra. Questo sarà argomento del capitolo sui cosiddetti fit.

Per ora presentiamo una variabile statistica atta a quantificare il grado di *correlazione lineare* fra le due grandezze, anche se ne sconsigliamo l'uso a questo livello, specie se non accompagnato da una ispezione grafica della distribuzione bidimensionale.

In analogia alla varianza di una variabile, si definisce la *covarianza* come media dei prodotti degli scarti delle due grandezze rispetto alla media:

$$\begin{aligned}
\mathrm{Cov}(X,Y) &= \frac{\sum_i (x_i-\overline{x})(y_i-\overline{y})}{N} && (5.46)\\
&= \frac{\sum_k n_k (x_k-\overline{x})(y_k-\overline{y})}{N} && (5.47)\\
&= \sum_k f_k (x_k-\overline{x})(y_k-\overline{y}) && (5.48)\\
&= \sum_k w_k (x_k-\overline{x})(y_k-\overline{y})\,. && (5.49)
\end{aligned}$$

L'analogia formale consiste nel fatto che la covarianza di una variabile con se stessa è pari alla varianza

$$\mathrm{Cov}(X,X) = \mathrm{Var}(X)\,, \tag{5.50}$$

come risulta ovvio dalla definizione.

Se $\mathrm{Cov}(X,Y)$ è positiva vuol dire che quando una delle due grandezze presenta scarti positivi (rispetto alla media) anche l'altra grandezza ha mediamente scarti positivi; se una ha scarti negativi anche l'altra ha mediamente scarti negativi. Se invece la covarianza è negativa significa che gli scarti si presentano preferibilmente con il segno opposto. Si dice che nei due casi le grandezze sono (linearmente) correlate positivamente o negativamente. Se essa è nulla non c'è correlazione (lineare) fra le due grandezze (ma ci può essere una correlazione più complicata).

Se si prova a calcolare la covarianza su coppie di valori anche scelti a caso raramente si avrà esattamente zero. Va quindi precisato meglio cosa si intende per covarianza "piccola" (prossima a zero). Per questo motivo si preferisce una variabile adimensionale, ottenuta dividendo la varianza per le deviazioni standard di ciascuna delle grandezze, prese come unità di scala. Si ottiene così il *coefficiente di correlazione*:

$$\rho(x,y) = \frac{\mathrm{Var}(x,y)}{\sigma(x)\sigma(y)}\,, \tag{5.51}$$

che per costruzione assume valori compresi fra -1 e 1. Gli estremi corrispondono ai casi estremi di correlazione in cui tutti i punti sono perfettamente allineati su una retta con pendenza positiva ($\rho = +1$) o negativa ($\rho = -1$).

Come per la varianza, skewness e curtosi per il calcolo della covarianza non vengono utilizzate le formule che la definiscono, bensì si fa uso della proprietà

$$\mathrm{Cov}(x,y) = \overline{xy} - \overline{x}\cdot\overline{y}\,, \tag{5.52}$$

dove con $\overline{xy}$ è stato indicata la media del prodotto

$$\overline{x\cdot y} = \frac{\sum xy}{N} = \frac{\sum_i x_i y_i}{N} = \frac{\sum_k n_k x_k y_k}{N} = \sum_k w_k x_k y_k \tag{5.53}$$

Come esempio di ricerca di correlazioni mostriamo in alto della figura 5.2 lo scatter plot di due tempi di attesa consecutivi registrati al contatore. Non si notano correlazioni al di fuori di quelle che possono essere simulate da fluttuazioni casuali. La figura in basso mostra invece la somma di due tempi di attesa in funzione del primo di essi. In questo caso chiaramente la correlazione è stata forzata ed è infatti visibile nello scatter plot. Il coefficiente di correlazione vale nei due casi rispettivamente -0.24 e +0.63. Si noti come la figura indichi la presenza o assenza di correlazione molto più chiaramente di quanto non si possa evincere dalla differenza (dei moduli) dei due coefficienti di correlazione. In particolare, la correlazione di -0.24 è chiaramente prodotta dai grandi valori dei tempi, i quali sono seguiti o preceduti più frequentemente da numeri piccoli semplicemente perché le combinazioni di due tempi grandi sono più rare e quindi non sono apparse nelle 50 coppie della figura. Ma anche eliminando i punti in cui una delle due coordinate è maggiore di 14 s il coefficiente di correlazione resta ancora abbastanza diverso da zero (-0.14).

Quindi, in conclusione, il coefficiente di correlazione misura il grado di correlazione lineare fra due grandezze, <u>ma</u>

- il suo valore non consente sempre di farsi una'idea effettiva della correlazione fra i dati;
- esso è insensibile a correlazioni più complicate di quelle lineari;
- è molto più istruttivo osservare un istogramma bidimensionale dei dati.

L'argomento delle correlazioni verrà trattato più esaurientemente in relazione alle variabili casuali e ai fit. Quindi è preferibile non utilizzare tale variabile per il momento.

5.13 ↻ σ_N o σ_{N-1}? Commenti sul fattore correttivo $N/(N-1)$

Qualcuno saprà già, o si sarà accorto utilizzando la calcolatrice, che si incontra un'altra definizione di varianza (e quindi di deviazione standard), ottenuta dividendo la somma dei quadrati degli scarti per $N-1$ anziché per N. Ovvero le formule della varianza sono state riscalate per un fattore $N/(N-1)$ (e quelle della deviazione standard per $\sqrt{N/(N-1)}$). Lo stesso fattore correttivo viene applicato ovunque ci sono formule con medie di potenze di scarti (momento 3° e 4° nel calcolo della skewness e della curtosi) o loro combinazioni (covarianza).

Sulle calcolatrici le deviazioni standard ottenute con $N-1$ sono chiamate σ_{N-1}.

La ragione di questa correzione è legata ad una interferenza fra statistica descrittiva e statistica inferenziale (una scuola di pensiero di questa, ad essere più precisi). È chiaro che se vogliamo mantenere una varianza che conservi l'analogia con il momento di inerzia si può solo prendere σ_N. Se invece da un piccolo campione si vuole inferire la deviazione standard di una grande popolazione, allora si tratta di un problema ben diverso, che andrà affrontato con la dovuta cautela.

Quindi la deviazione σ_N, indicata più semplicemente con σ, è adeguata all'uso che si fa per ora di questa quantità.

5.14 Nota sulle cifre significative da utilizzare nei problemi di statistica descrittiva

Parlando del quaderno di laboratorio è stato introdotto il problema delle cifre significative e sono state suggerite delle raccomandazioni, da usare con una certa flessibilità. Il problema si ripropone per il calcolo delle grandezze statistiche.

È importante capire come in un ambito di pura statistica descrittiva il problema non ha molto senso in quanto i riassunti statistici servirebbero soltanto a descrivere quantitativamente la distribuzione statistica osservata. Ma, come è già stato fatto notare precedentemente e nel paragrafo 5.4, è difficile svincolare la statistica puramente descrittiva da quella inferenziale. In particolare, ogni volta che si presenta un risultato di una misura di una grandezza fisica si sta compiendo un'operazione di statistica inferenziale e allora ha senso parlare di incertezza del risultato e di cifre significative.

Purtroppo è impossibile astenersi da fare qualsiasi conto prima di aver chiaro come presentare il risultato, cosa che si saprà fare correttamente e a ragione soltanto quando saranno chiari diversi concetti di probabilità e di statistica inferenziale. Converrà quindi procedere per gradi tenendo conto che di quanto segue.

- Se i conti servono come semplice esercizio sui metodi statistici introdotti è sufficiente calcolare le grandezze di interesse a 2-3 cifre significative, in modo tale da controllare la bontà delle procedure a qualche parte su cento o su mille.

- Gli argomenti sulle cifre "dopo la virgola" sono in genere fallaci.
- Nel caso che i risultati dei conti svolti a questo stadio iniziale possano essere riutilizzati per successive elaborazioni che conducano ad un risultato finale con stima quantitativa dell'incertezza converrà abbondare con le cifre ed effettuare gli arrotondamenti alla fine.
- Il dilemma fra la pesantezza delle troppe cifre inutili e il rischio di dover rifare tutti i conti è momentaneo: alla fine del corso si saprà perfettamente come comportarsi; quindi eventuali "disavventure" iniziali vanno considerate fisiologiche e pertanto formative.

Per quanto concerne invece i conti è importante utilizzare tutte le cifre della calcolatrice (10, generalmente) e arrotondare solo alla fine. Altrimenti possono sorgere problemi numerici che falsano drammaticamente i risultati. Anche se questo può sembrare in contraddizione con quanto si va ripetendo sulle cifre significative, si rifletta sul fatto che la varianza (e anche covarianza, skewness e curtosi) sono calcolate in pratica da (somme e) differenze di numeri che a volte possono essere molto simili fra di loro. Quindi le ultime cifre (quelle più a destra) giocano un ruolo decisivo ai fini del risultato.

Come esempio convincente prendiamo la distribuzione statistica costituita dai seguenti valori (fra parentesi la loro frequenza): 100.21 (3); 100.22 (8); 100.23 (10); 100.24 (7); 100.25 (2). I valori della media aritmetica e delle media dei quadrati sono (con tutte le cifre della calcolatrice):

$$\begin{aligned} \overline{x} &= 100.2296552 \\ \overline{x^2} &= 10045.98378\,. \end{aligned}$$

da cui ne segue che la varianza è calcolata come

$$\begin{aligned} \sigma 2 &= 10045.98388 - 10045.98378 \\ &= 0.00010\,, \end{aligned} \tag{5.54}$$

ovvero il risultato è deciso dalla 9^a e 10^a cifra!

Il modo per ovviare a questo problema consiste nell'utilizzare le proprietà di trasformazione di media e varianza e considerare nei conti solo la parte eccedente 100. In alcuni casi, quando la dispersione è dell'ordine di una parte su 10000 o meno questo è l'unico modo di operare (a meno di non disporre di computer a "doppia precisione") in quanto anche le 10 cifre delle calcolatrici diventano insufficienti.

5.15 Problemi

1. Dimostrare che la media degli scarti è nulla.
2. Valutare moda, mediana e media e deviazione standard dei seguenti dati statistici: 6, 7, 10, 1, 8, 12, 7, 5, 7, 9, 6, 8.
3. Valutare moda, mediana e media e deviazione standard dei seguenti dati statistici: 0, 6, 7, 10, 1, 8, 0, 12, 7, 5, 7, 9, 6, 8.
4. Valutare moda, mediana e media e deviazione standard dei seguenti dati statistici: 3, 4, 0, 6, 7, 10, 1, 8, 0, 12, 7, 5, 7, 9, 6, 8.
5. Valutare moda, mediana e media e deviazione standard dei seguenti dati statistici: 6, 3, 4, 0, 6, 7, 10, 1, 8, 0, 12, 7, 5, 7, 9, 6, 8.
6. Calcolare moda, mediana, media, deviazione standard e coefficiente di variazione delle distribuzioni di conteggi per 6, 12 e 30 della tabella 4.1, riportando anche i valori di $\sum x$ e di $\sum x^2$.
7. Trovare moda, mediana, media, deviazione standard e coefficiente di variazione della distribuzione di conteggi per 100 secondi della tabella 4.2.
8. Trovare moda, mediana, media, deviazione standard e coefficiente di variazione della distribuzione di conteggi per 300 secondi della tabella 4.2.
9. Calcolare media e deviazione standard dei seguenti valori: 1000000.2, 1000000.3, 1000000.1, 1000000.2, 1000000.4
10. Calcolare media e deviazione standard degli scarti fra i valori stimati e quelli letti al nonio dello studente L.T. della tabella 2.2. Se quantifichiamo l'errore tipico con la deviazione standard degli scarti e in una misura successiva effettuata nelle stesse condizioni egli affermerà di leggere 44.76 cm quanto si crederà a tale valore?
11. Da 100 valori si ottiene $\overline{x} = 5.50$ e $\sigma = 0.25$. Altri 50 valori danno $overlinex = 5.46$ e $\sigma = 0.29$. Quanto valgono media e deviazione standard raggruppando i 150 valori?

Capitolo 6

Analisi grafiche

6.1 Studio di andamenti funzionali

Abbiamo già visto negli esempi dei paragrafi precedenti l'importanza delle rappresentazioni grafiche al fine di sintetizzare informazioni e di fornire una visione d'insieme delle misure.

Ci interessiamo adesso al caso speciale in cui si voglia mettere in risalto un certo andamento funzionale. Per esempio, avendo misurato due grandezze x e y si vuole trovare il tipo di andamento funzionale $y = f(x)$. Molto spesso la grandezza associata alla variabile indipendente è il tempo.

Prima di passare allo studio dei diversi andamenti diamo alcune raccomandazioni sull'importanza dei grafici e sul modo di disegnarli.

6.2 Grafici

Innanzitutto è importante convincersi che i grafici vanno fatti tempestivamente poiché essi offrono un quadro d'insieme superiore a quello delle tabelle. Da un andamento "sbagliato", non percettibile esaminando i numeri, si può decidere che le misure sono fatte in modo errato e si può reagire in tempo ripetendo parzialmente o completamente le misure.

Come detto a proposito del logbook, è altresì importante che un grafico sia ben leggibile anche a distanza di tempo. La ricostruzione dell'esatto valore numerico a partire da un punto graficato può essere non interessante e all'occorrenza si deve ricorrere alle informazioni contenute nelle tabelle. Deve essere invece ottenibile con rapidità una stima che sia una buona approssimazione del valore esatto.

Da queste regole generali segue che

- I grafici vanno disegnati su apposita *carta millimetrata* e, in mancanza di questa, possibilmente su carta a quadretti.

- Su ciascun asse deve essere riportato il nome della grandezza (o un simbolo per esso) e la sua unità di misura. Quest'ultimo è generalmente posto fra parentesi. Ad esempio

$$\textit{velocità}\ (\mathrm{m/s}) \qquad\qquad v\ (\mathrm{m/s})\,.$$

Alcuni preferiscono separare il nome e l'unità di misura dal simbolo "/", ad indicare che il valore numerico riportato sul grafico è pari al valore della grandezza *diviso* l'unità di misura, ad esempio

$$altezza/\mathrm{m} \qquad\qquad h/\mathrm{m}\,.$$

Anche se questo secondo modo è abbastanza diffuso ed accettato bisogna fare un po' di attenzione quando le grandezze o le unità di misura sono indicate con simboli complicati. Ad esempio, già

$$velocità/\mathrm{m/s} \qquad\qquad v/\mathrm{m/s}$$

non è gradevole.

- Eventualmente le potenze di 10 dei valori graficati possono essere riportate insieme all'unità di misura. Per esempio se tutti i valori di massa sono compresi fra 10000 e 50000 kg si può scrivere

$$\mathrm{M}\ (10^4\mathrm{kg})$$

e riportare i valori delle decine di migliaia (valori compresi fra 1 e 5). È però preferibile scegliere

$$\mathrm{M}\ (10^3\mathrm{kg}) \tag{6.1}$$

e riportare le migliaia (valori compresi fra 10 e 50) in quanto le potenze di 10 multiple di 3 (millesimi, migliaia, milioni, etc.) sono generalmente meglio percepite e memorizzate. Approfittiamo per ricordare nella tabella 6.1 i prefissi che indicano le potenze di 10. È da notare che a volte si incontra la potenza di 10 affianco al simbolo della grandezza. Essa sta ad indicare che il valore è stato moltiplicato per tale potenza. Nel nostro caso si avrebbe

$$\mathrm{M} \times 10^{-3}(\mathrm{kg})\,. \tag{6.2}$$

Bisogna prestare attenzione a riportare e rileggere i valori in modo consistente con la notazione usata. Un minimo di confusione fra le notazioni (6.1) e (6.2) causa uno sbaglio di un fattore 10^6!

Multipli			Sottomultipli		
	simbolo	nome		simbolo	nome
10^1	da	deca-	10^{-1}	d	deci-
10^2	h	etto-	10^{-2}	c	centi-
10^3	k	chilo-	10^{-3}	m	milli-
10^6	M	mega-	10^{-6}	μ	micro-
10^9	G	giga-	10^{-9}	n	nano-
10^{12}	T	tera-	10^{-12}	p	pico-
10^{15}	P	peta-	10^{-15}	f	femto-
10^{18}	E	exa-	10^{-18}	a	atto-

Tabella 6.1: Simboli e prefissi delle potenze di dieci più comunemente usate.

- Sulla carta millimetrata gli assi sono graduati con delle tacche. Soltanto su alcune di esse, poste equidistanti e più marcate delle altre, viene indicata la scala.

- Si ricorda che, affinché i valori sugli assi siano definiti in modo univoco, occorre che ci siano almeno due tacche esplicitamente numerate (fa eccezione la carta logaritmica che incontreremo fra breve, per la quale ne è sufficiente una).

- Non occorre che un asse comprenda necessariamente lo zero, anche se a volte può essere comodo per apprezzare la variazione relativa delle grandezze. In genere si preferisce segnare gli assi in modo da avere la massima leggibilità sulla regione effettivamente misurata, o su quella nella quale si sia interessati a fare delle *estrapolazioni*. Ne segue che non è corretto indicare un solo punto, "assumendo" che l'asse cominci da zero.

- Se i dati sono concentrati in due gruppi molto lontani fra di loro può essere conveniente *spezzare* l'asse con opportuni simboli.

- Al fine di agevolare sia la scrittura che la rilettura del valore dei punti è opportuno che l'intervallo fra le tacche sia (a meno di una potenza di dieci) di $1, 2$ e 5. Provare ad esempio a rileggere il valore degli stessi sei punti riportati con diverse scale in figura 6.1.

- I valori vanno riportati con dei puntini o altri simboli ($\circ$, $\bullet$, $\times$, $\triangle$, $\square$ etc.). Nel caso essi facciano riferimento a diverse serie o tipi di misure è d'obbligo utilizzare simboli diversi e riportare le corrispondenze nell'apposita didascalia. È preferibile invece evitare l'uso di crocette ($+$) in quanto si potrebbero confondere con le *barre di incertezza*.

- Non vanno riportati sugli assi i valori sperimentali. Come già detto, scopo del grafico non è quello di riportare gli esatti valori numerici. Numerelli scritti lungo gli assi (o vicino ai punti) generano solo disordine e diminuiscono la percezione d'insieme del grafico.

- Se il valore numerico dell'incertezza è noto[1] esso va riportato con una *barra di incertezza*.

La figura 6.2, estratta da una delle prime pubblicazioni della Collaborazione UA1 al CERN di Ginevra sulla scoperta del bosone $W^{\pm}$ (il capo della Collaborazione era Carlo Rubbia), mostra come le regole sopra elencate per disegnare i grafici siano sostanzialmente le stesse seguite nella ricerca avanzata.

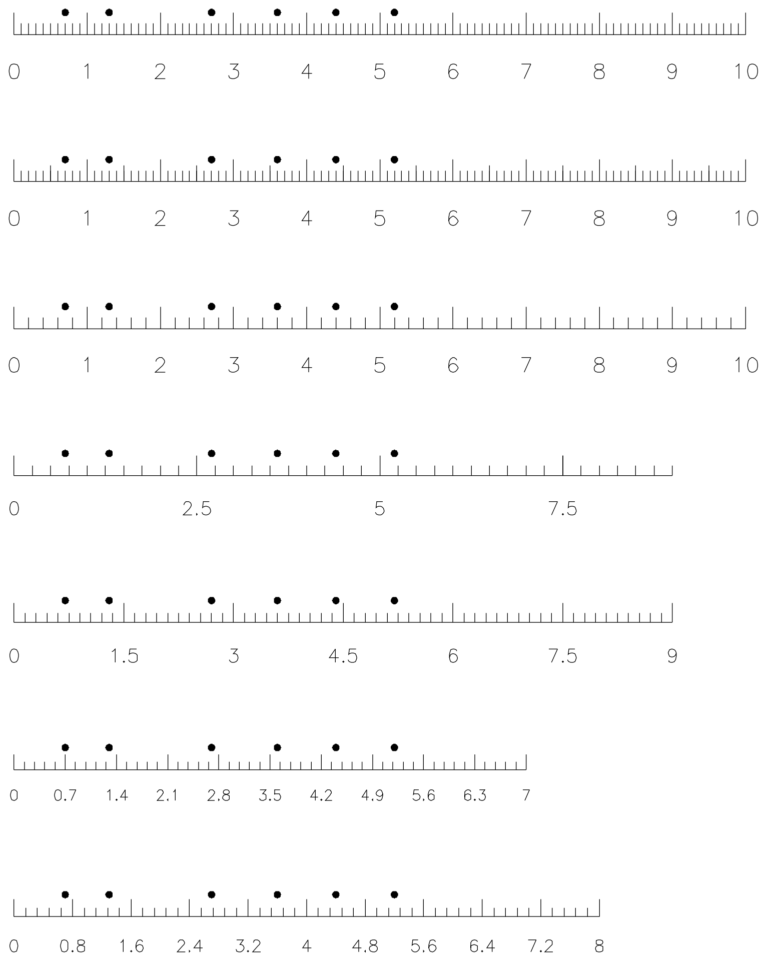

Figura 6.1: Quale scelta degli assi è la più conveniente?

6.3 Grafici lineari: stima grafica dei parametri della retta

La figura 6.3 mostra i dati della prima serie di misure di tabella 2.5 riportati su opportuni grafici. Il primo grafico mostra l'allungamento, espresso in centimentri, in funzione della massa applicata espressa in chilogrammi. Il secondo mostra invece il periodo di oscillazione, espresso in secondi, in funzione della radice quadrata della massa applicata espressa in unità di radice quadrata di chilogrammo.

[1] In molti casi invece l'incertezza è ricavata dalle fluttuazioni dei punti sperimentali intorno all'*andamento medio* della curva che li descrive. Allora le barre possono essere apportate soltanto in una seconda fase di elaborazione dei dati sperimentali. Questa nota serve a sottolineare che non c'è un imperativo categorico di riportare i punti sperimentali sempre accompagnati dalle barre di incertezza.

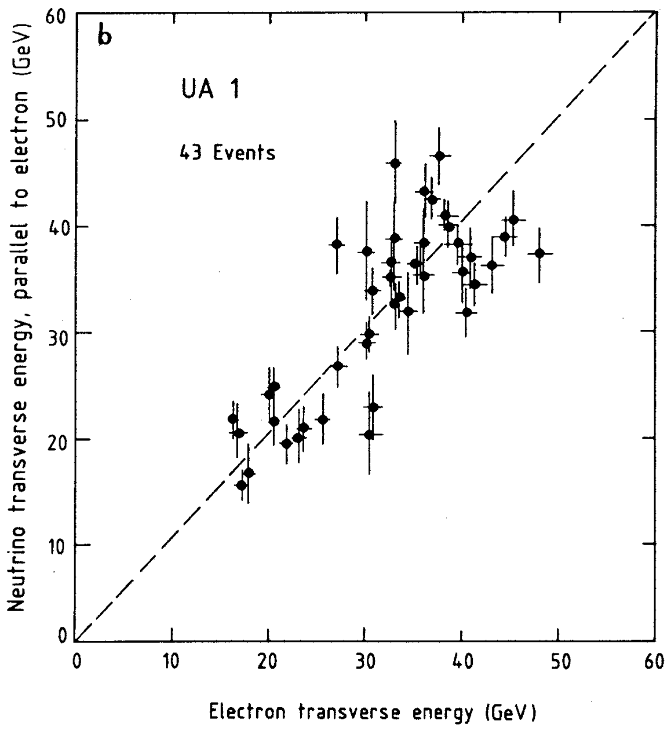

Figura 6.2: Dati dell'esperimento UA1 al protosincrotrone SPS del CERN (Ginevra) relativi all'evidenza sperimentale del bosone intermedio $W^{\pm}$.

Entrambi i grafici mostrano, al di sopra di una massa critica, un andamente lineare dei punti. Questo indica che le relazioni (2.1) e (2.2) sono *ragionevolmente* soddisfatte. È quindi possibile tracciare la retta che *meglio passa* per i punti, determinando per via grafica i parametri di tale andamento e associarli, attraverso le (2.1-2.2) alle grandezze fisiche di interesse (k e g).

Definiamo la procedura in via generale pensando a due quantità x e y legate dalla relazione

$$y = mx + c\,. \tag{6.3}$$

m e c sono rispettivamente il *coefficiente angolare* (o *pendenza*, in inglese *"slope"*) e l'*intercetta* della retta. Per analizzare i dati ci si basa sul seguente modello:

- l'equazione (6.3) è quella ipotizzata dai *valori veri* delle grandezze fisiche.

- i punti sperimentali non giacciono tutti sulla retta - nel *senso geometrico* - a causa degli inevitabili errori di misura;

- la migliore stima della retta *vera* è quella che passa meglio per i punti. Per ora la sua valutazione sarà effettuata ad occhio cercando, mediante un righello trasparente, di trovare la retta che meglio approssima globalmente i punti.

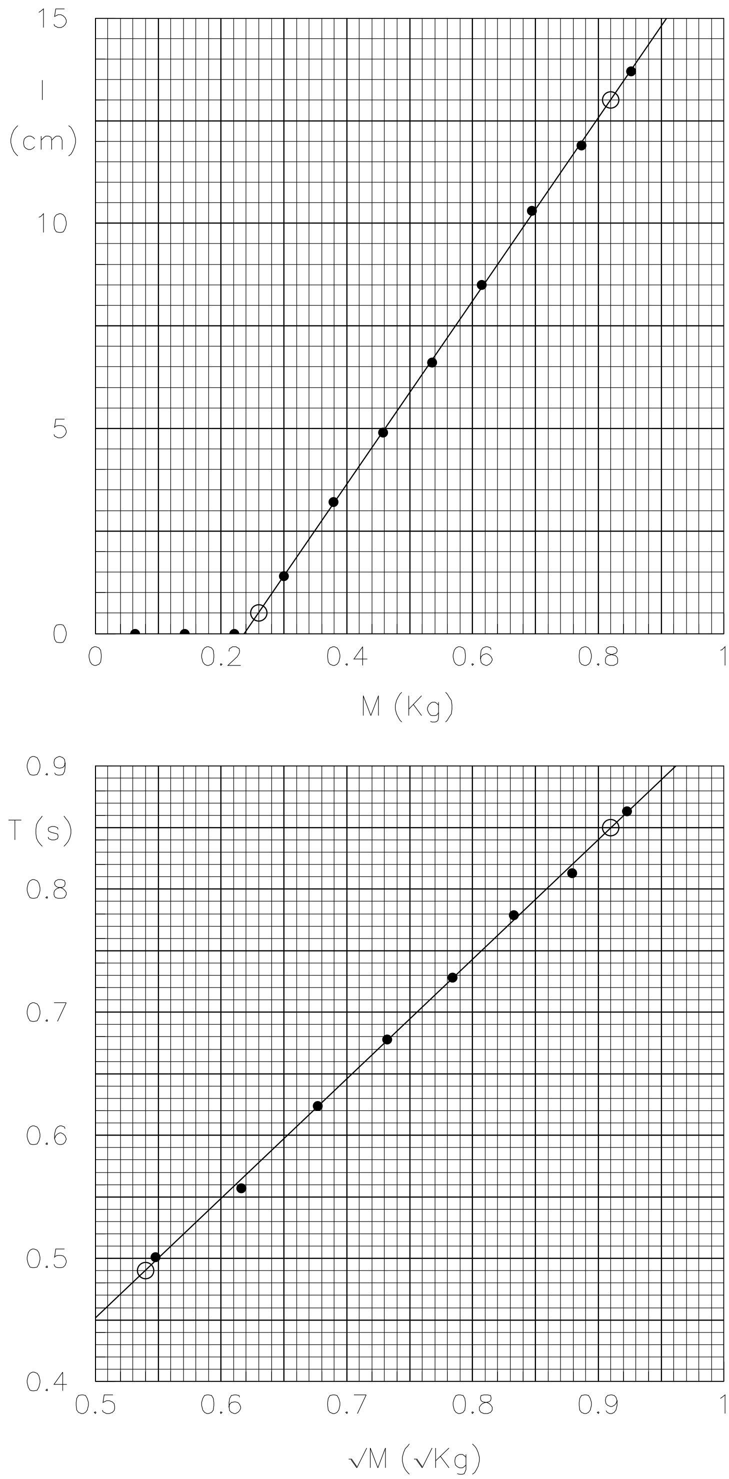

Figura 6.3: Grafici dell'allungamento in funzione della massa applicata alla molla e del suo periodo di oscillazione in funzione della radice quadrata della massa. I cerchietti indicano i punti utilizzati per il calcolo dei parametri della retta.

Trovata la retta che meglio *si adatta* ai dati i parametri m e c sono determinati considerando due punti, in principio "arbitrari", appartenenti alla retta. Chiamando i due punti $P_1 = (x_1, y_1)$ e $P_2 = (x_2, y_2)$ si può scrivere l'equazione parametrica della retta

$$\frac{y - y_1}{y_2 - y_1} = \frac{x - x_1}{x_2 - x_1} \Longrightarrow y = \frac{y_2 - y_1}{x_2 - x_1} x + y_1 - \frac{y_2 - y_1}{x_2 - x_1} x_1 \,, \tag{6.4}$$

ovvero

$$m = \frac{y_2 - y_1}{x_2 - x_1} \tag{6.5}$$

$$c = y_1 - \frac{y_2 - y_1}{x_2 - x_1} x_1 = y_1 - m x_1 \,. \tag{6.6}$$

È da notare che, mentre la (6.5) è effettivamente usata per la stima di m, per valutare l'intercetta si preferisce determinare direttamente, per via grafica, il valore per il quale la retta interseca l'asse delle ordinate (a meno che lo zero dell'asse delle ascisse sia fuori scala):

$$c = y(x = 0) \,. \tag{6.7}$$

Per quanto riguarda la scelta dei punti per valutare m (ed eventualmente c) si seguono questi criteri:

- i punti vanno scelti sulla retta e non in corrispondenza dei valori sperimentali (a meno che per un puro caso un punto sperimentale non giaccia esattamente sulla retta). Infatti, nonostante intuitivamente si possa pensare che un punto misurato sia "più vero" di un generico punto sulla retta, i punti della retta tengono conto simultanemente di tutti i punti misurati e quindi sono meno sensibili alle possibili fluttuazioni legate all'errore della singola misura;

- i punti vanno scelti in modo che siano ben leggibili e quindi preferibilmente in prossimità dell'incrocio di due linee della carta millimetrata;

- al fine di ridurre l'incertezza dovuta alla lettura dei punti è opportuno scegliere questi molto lontani fra di loro. Infatti, facendo riferimento alla (6.5), ci si può facilmente convincere che il numero di cifre significative su m ottenibili con il metodo grafico è legata al numero di cifre significative del numeratore e del denominatore. Queste ultime dipendono a loro volta dal valore assoluto di $y_2 - y_1$ e di $x_2 - x_1$ (si provi a considerare il limite di due punti molto vicini);

6.4 Cifre significative dei parametri della retta

Il numero di cifre significative da utilizzare per i parametri della retta merita una piccola digressione. Diciamo subito che, come al solito, non ha senso dilungarsi troppo su questi problemi senza avere una conoscenza approfondita della teoria dell'incertezza. Cerchiamo qui soltanto di fornire delle indicazioni di massima, mostrando quali sono le cause di incertezza.

6.4.1 Rilettura dei punti sperimentali e della retta

Cominciamo col considerare il numero di cifre con cui è possibile leggere i punti della retta che passa per i punti. Intuitivamente si può pensare a due criteri estremi, di cui uno è chiaramente errato e l'altro è un po' pessimista, anche se a prima vista "potrebbe" risultare in accordo con quanto detto precedentemente.

- Il primo consiste nel riportare tutte le cifre con le quali si riescono a rileggere i punti sul grafico; per capire come questo metodo applicato indiscriminatamente porta a delle assurdità, si pensi a una decina di punti riportati su una scala talmente dilatata che la variazione dell'ultima cifra (la meno significatica) del valore misurato produca una variazione di un centimetro sul grafico. Rileggendo i decimi di millimetri si ottengono i punti ricostruiti con due cifre significative in più dei risultati sperimentali (un fattore 1/100 nell'incertezza relativa!).

- In alternativa si potrebbe essere tentati a rileggere i punti sperimentali con un numero di cifre pari a quelle dei punti sperimentali. In questo secondo caso si sbaglia per difetto. Infatti la retta che meglio si adatta ai punti tiene conto simultaneamente di tutte le informazioni e quindi i suoi punti sono più stabili (più credibili) dei singoli punti misurati.

Per arrivare ad un criterio quantitativo applicabile in tutte le circostanze servono conoscenze teoriche che verranno introdotte soltanto nel capitolo 12. Per adesso possiamo affermare che, per le esperienze di laboratorio in cui vengono graficati una decina di punti, i punti sulla retta possono essere riletti - tipicamente - con una cifra in più dei dati sperimentali.

Come esempio di applicazione di queste tecniche grafiche vedremo fra un po' l'analisi dei dati della molla.

6.4.2 Scelta e tracciamento della retta

Oltre all'incertezza sulla rilettura dei punti ce ne sono altre legate alla rozzezza della tecnica grafica.

- La scelta "ad occhio" della retta ha indubbiamente un carattere soggettivo e persone diverse possono scegliere rette diverse, ma in genere il margine di incertezza sui parametri dovuto a questo effetto è piuttosto piccolo.

- Dalla scelta della retta al suo tracciamento con la matita sul foglio di carta millimetrata ci possono essere delle piccole differenze: serve un po' di pratica per stimare, ad esempio, la distanza fra la linea seguita dalla matita e il bordo del righello.

- La rilettura dei punti può produrre un ulteriore errore.

6.4.3 Altre incertezze nella stima dei parametri

Le incertezze di cui si è parlato possono essere eliminate passando da metodi grafici a metodi più sofisticati. Rimangono invece altre sorgenti di incertezza legate al processo di misura e alla strumentazione usata:

- Ripetendo le misure in condizioni apparentemente uguali si ottengono valori diversi che produrranno rette diverse indipendentemente dalla scelta del metodo di analisi.

- A questa incertezza se ne sommano altre dovute alla non perfetta calibrazione degli strumenti, a fattori esterni di influenza, e così via.

Tutte queste verranno trattate al momento opportuno con metodi probabilistici (capitolo 12). Per il momento l'alternativa è di ripetere più volte le misure e l'analisi dei dati e di confrontare i risultati per farsi un'idea dell'intervallo in cui possono essere ragionevolmente compresi i valori delle grandezze fisiche di interesse.

6.4.4 Raccomandazioni

Tenendo conto delle varie sorgenti di incertezza che intervengono sulla ricostruzione dei parametri, suggerire delle semplici prescrizioni sul numero di cifre da considerare è, a questo livello, ancora più problematico di quanto visto nel paragrafo 3.4. Sempre ricordando di prendere con spirito critico le raccomandazioni e tenendo conto della natura transitoria - per scopi didattici - del problema, si possono dare le seguenti raccomandazioni.:

- Se i punti sono bene allineati e con fluttuazioni intorno alla retta dell'ordine di grandezza della cifra meno significativa di essi allora è possibile leggere i punti della retta con una cifra in più dei punti sperimentali.

- Si tenga conto che, nel caso precedente, la qualità sui parametri della retta non può essere peggiore (e quindi il numero di cifre non può essere minore) di quella che si avrebbe considerando soltanto i due punti sperimentali estremi.

- Se le fluttuazioni intorno alla retta sono molto maggiori dell'ultima cifra allora la lettura dei punti della retta può essere fatta con un numero di cifre inferiore a quelle dei punti sperimentali.

- Non è possibile basarsi soltanto sulle fluttuazioni dei punti intorno alla retta in quanto l'incertezza dipende in modo non semplice anche - a parità dispersione dei punti - dal numero dei punti e dalla loro configurazione.

- Nel seguito (paragrafo 12.6) verranno date delle regole empiriche per stabilire in modo approssimativo le incertezze sui parametri senza dover fare conti complicati, ma riteniamo che queste regolette debbano venire dopo la trattazione rigorosa del problema.

- se è possibile, effettuare più serie di misure e valutare l'ordine di grandezza dell'incertezza dalla dispersione dei risultati ottenuti per ciascuna serie;

- anche gli effetti sistematici verranno trattati nel seguito e per alcuni casi semplici di essi verranno date regolette pratiche.

6.5 Linearizzazione

Si è visto nel paragrafo precedente che per graficare l'andamento del periodo di oscillazione in funzione della massa si è preferito cambiare variabile e riportare sull'asse delle ascisse $\sqrt{M}$. Per comprendere meglio la ragione di tale scelta riportiamo in figura 6.4 anche il grafico di T in funzione di M. Nella figura sono anche tracciate due curve. Quella continua corrisponde esattamente all'andamento rettilineo di figura 6.3. Quella tratteggiata mostra una diversa parametrizzazione della relazione $T = f(M)$. È ovvio che è molto più semplice tracciare la retta che meglio si adatta a dei punti piuttosto che la funzione $\sqrt{\cdot}$. Per questo motivo si preferisce a volte *linearizzare* un andamento mediante opportuno cambiamento di variabili. Questa pratica è particolarmente utile non soltanto in laboratorio per eseguire delle rapide stime di parametri o per delle rapide verifiche, ma anche nelle pubblicazioni scientifiche per convincere il lettore che una legge segue un certo andamento.

Una classe importante di linearizzazioni è quella legata ad andamenti esponenziali (o logaritmici) e a leggi di potenza. Queste verranno trattate nel paragrafo 6.7. Si accennerà successivamente ad altre linearizzazioni notevoli.

6.6 ⟳ Analisi grafica dell'esperienza della molla

Come esempio dell'uso di grafici lineari analizziamo i dati della prima serie di misure della tabella 2.5, riportati in figura 6.3. I punti sulla retta utilizzati per i calcoli sono indicati con un cerchietto.

Dal grafico di T in funzione di $\sqrt{M}$ valutiamo il coefficiente angolare e quindi la costante elastica della molla:

$$m = \frac{(0.850 - 0.491)\,\mathrm{s}}{(0.910 - 0.540)\,\sqrt{\mathrm{kg}}} = 0.970\,\frac{s}{\sqrt{\mathrm{kg}}} \tag{6.8}$$

$$k = \frac{4\pi^2}{m^2} = 41.9\,\frac{\mathrm{kg}}{\mathrm{s}^2} = 41.9\,\frac{\mathrm{N}}{\mathrm{m}}\,. \tag{6.9}$$

Dal grafico di l in funzione di M troviamo che il coefficiente angolare m' vale

$$m'\left(=\frac{g}{k}\right) = \frac{(13.00 - 0.51)\,10^{-2}\,\mathrm{m}}{(0.820 - 0.260)\,\mathrm{kg}} = 0.223\,\frac{\mathrm{m}}{\mathrm{kg}}\,. \tag{6.10}$$

Ne segue

$$g = km' = 41.9\,\frac{\mathrm{N}}{\mathrm{m}} \times 0.223\,\frac{\mathrm{m}}{\mathrm{kg}} = 9.34\,\frac{\mathrm{m}}{\mathrm{s}^2}\,. \tag{6.11}$$

È da notare che il numero di cifre significative segue coerentemente dalla risoluzione con cui è possibile leggere i punti della retta sui grafici e non tiene conto di altre eventuali incertezze.

6.6.1 Dipendenza dal modello

C'è un ultimo punto che va precisato. Nello stimare la costante elastica della molla dal grafico di T in funzione di $\sqrt{M}$ la retta è stata scelta senza imporre

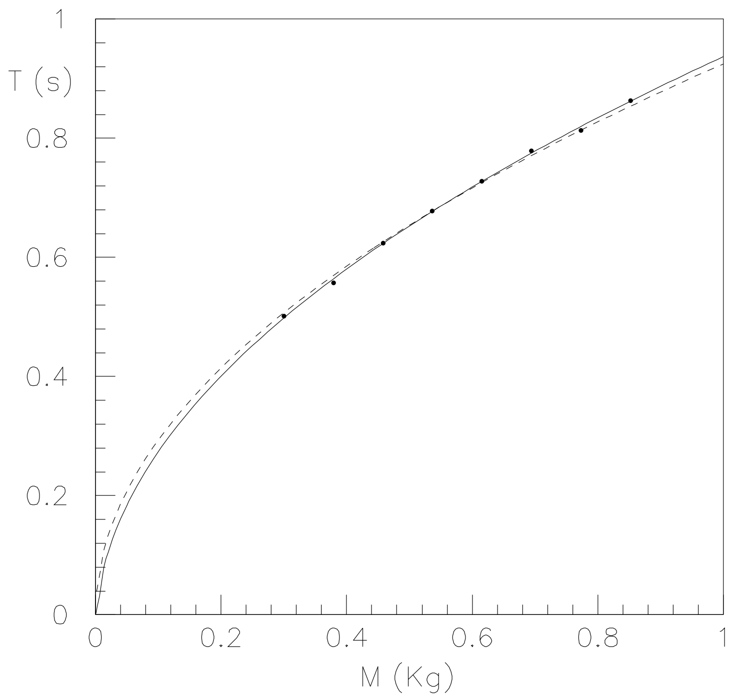

Figura 6.4: Periodo di oscillazione della molla in funzione della massa applicata. La linea continua corrisponde all'andamento ottenuto con due parametri ricavati dal grafico del periodo in funzione della radice quadrata della massa. La linea tratteggiata mostra la curva che meglio si adatta ai dati se si impone il passaggio per l'origine degli assi.

il passaggio per l'origine degli assi. Questo significa che implicitamente si è assunto che la legge (2.2) potesse essere non perfettamente valida e che ci potesse essere un termine correttivo costante al periodo:

$$T = \frac{2\pi}{\sqrt{k}}\sqrt{M} + T_\circ\,. \tag{6.12}$$

Utilizzando la (6.6) si trova

$$T_\circ (= c) = 0.491 - 0.970 \times 0.540\,\mathrm{s} = -0.033\,\mathrm{s}\,. \tag{6.13}$$

Questo valore di $T_\circ$ "diverso da zero" fa sorgere alcune domande:

- innanzitutto, $T_\circ$ è veramente *significativamente* diverso da zero?
- in tal caso significa che la nostra legge è sbagliata?
- quale valore di g si ottiene se si impone $T_\circ = 0$, ovvero che la legge (2.2) sia quella giusta?

Le prime due domande sono quelle a quelle è più difficile rispondere a questo livello. Per quanto riguarda la terza, basta estendere il grafico di T in funzione di $\sqrt{M}$ in modo da includere anche l'origine degli assi, tracciare la nuova retta e rifare i conti. I dettagli vengono lasciati per esercizio. Il risultato è

$$g' = 10.28\,\frac{\mathrm{m}}{\mathrm{s}^2}\,. \tag{6.14}$$

In figura 6.4 sono riportate, insieme ai dati sperimentali, le funzioni relative ai due modelli: la curva continua per la legge (6.12) e quella tratteggiata per la legge (2.2). Uno sguardo al grafico e ai risultati ottenuti ci porta ad ulteriori questioni aperte.

- A occhio il modello con $T_\circ \neq 0$ si comporta meglio dell'altro. Ma è chiaro che, avendo un *parametro libero* in più, la curva ha più *libertà* di adattarsi ai dati sperimentali.

- Sempre ad occhio, sembra che addirittura una relazione lineare fra il periodo e la massa possa descrivere decentemente i dati. Come si può stabilire qual' è il modello migliore?

- Come abbiamo visto, per ogni modello abbiamo un risultato di g. È possibile combinare in qualche modo i diversi valori e presentare un solo risultato? Per esempio, facendo la media fra i due valori si otterrebbe $g = 9.81\,\mathrm{m/s}^2$, molto prossimo a quello vero. È questo un puro caso oppure la prova che la procedura è corretta?

Per ora lasciamo queste domande in sospeso e riprendiamo la trattazione dei grafici.

6.6.2 Combinazione dei risultati delle tre serie di misure

6.6.3 Valore di k condizionato dal valore noto di g

6.7 Uso di carte logaritmiche

Fra le linearizzazioni che si incontrano più di frequente ci sono quelle che trasformano una o entrambe le variabili attraverso la funzione logaritmo. La frequenza con la quale tale tecnica è usata, combinata con la difficoltà (fino a qualche decennio fa) del calcolo dei logaritmi e antilogaritmi (esponenziali), ha fatto sì che si diffondesse l'uso di speciale carta millimetrata per agevolare tale compito. Anche se oggi molti grafici si fanno direttamente al calcolatore e opportuni programmi permettono di trovare in modo rapido e automatico i parametri della curva che meglio si adatta ai dati è tuttora valida la rappresentazione su *scala logaritmica* per mostrare a colpo d'occhio se i dati seguono o no un certo andamento. Quindi, anche i programmi commerciali di rappresentazioni hanno l'opzione di mostrare una o entrambe le coordinate in scala logaritmica.

Un altro uso della carta logaritmica è di espandere la scala in un modo tale da permettere di apprezzare valori su diversi ordini di grandezza.

t (ms)	V (mV)	t (ms)	V (mV)
2.0	411	12.0	15
4.0	211	14.0	8
6.0	108	16.0	4
8.0	56	18.0	2
10.0	29	20.0	1

Tabella 6.2: Misure di tensione durante la scarica di un condensatore.

6.7.1 Carta semilogaritmica

Supponiamo che due variabili siano legate da una relazione esponenziale

$$y = ae^{bx} \,. \tag{6.15}$$

Esempi di fenomeni che seguono tale legge sono: decadimento radioattivo; accrescimento di colonie di batteri; scambio di calore di un corpo con l'ambiente circostante; carica e scarica di un condensatore (un esempio di quest'ultimo è mostrato nella tabella 6.2 e nella figura 6.5, dove sono riportati i valori di tensione ai capi del condensatore in funzione del tempo.) Essi hanno tutti in comune la caratteristica che *il tasso istantaneo di variazione temporale di una grandezza è proporzionale al valore della grandezza in quel determinato istante*:

$$\frac{dx}{dt} = bx \,. \tag{6.16}$$

Se dividiamo per a entrambi i membri della (6.15) e ne prendiamo il logaritmo naturale otteniamo

$$\ln \frac{y}{a} = bx \,, \tag{6.17}$$

la quale può essere riscritta come

$$\ln y = \ln a + bx \,. \tag{6.18}$$

se y e a, grandezze omogenee, sono prese con le stesse unità di misura. Si ricorda a tale proposito che gli argomenti di logaritmi ed esponenziale devono essere adimensionali e quindi la 6.18 andrebbe più correttamente scritta come

$$\ln \frac{y}{U_y} = \ln \frac{a}{U_a} + bx \tag{6.19}$$

dove con U è stata indicata l'unità di misura. Quindi riportando sulle ascisse i valori della variabile x e sulle ordinate i valori di $\ln y$ si ottiene un andamento lineare.

La carta *semilogaritmica* (o semplicemente *"semilog"*), con l'asse delle ordinate proporzionale al logaritmo del valore, agevola tale compito in quanto risparmia il calcolo dei logaritmi.

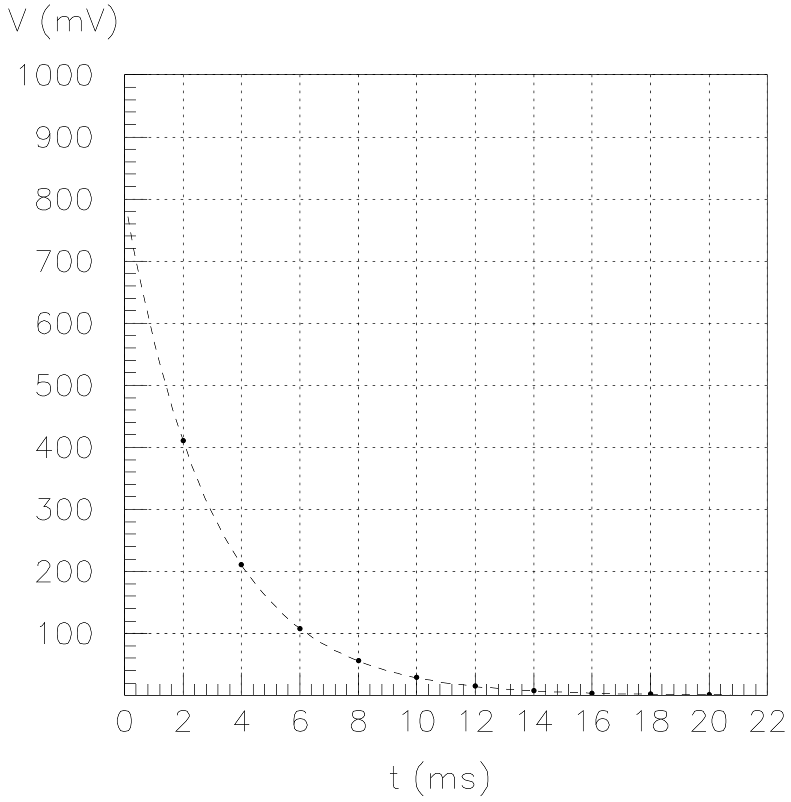

Figura 6.5: Misure di scarica di un condensatore riportate su carta lineare.

6.7.2 Proprietà delle carte logaritmiche

La figura 6.6 mostra un esempio di carta logaritmica. Possiamo fare delle osservazioni generali che derivano dalle proprietà dei logaritmi:

- la scala delle ordinate si infittisce in prossimità di 10^n, con n intero e successivamente si dirada. Questo è dovuto al fatto che, passando all'ordine di grandezza successivo, la spaziatura della scala viene eseguita in unità del nuovo ordine di grandezza. Poiché questo brusco cambiamento di spaziatura porrebbe problemi di lettura dei valori, la scala viene successivamente risuddivisa (questa nuova suddivisione, presente nelle carte logaritmiche commerciali, non è riportata in figura 6.6);

- la distanza fra due punti che differiscono di un certo numeri di ordini di grandezza è costante. Questo deriva dalle proprietà dei logaritmi. Infatti, se ad esempio due numeri y_1 e y_2 differiscono di n ordini di grandezza essi possono essere scritti come $y_2 = 10y_1$, da cui segue che

$$\ln y_2 - \ln y_1 = \ln \frac{y_2}{y_1} = n \ln 10 \,. \tag{6.20}$$

  La distanza misurata in centimetri fra due punti che differiscono di un fattore 10 è pari quindi alla distanza che passa fra 1 e 10, che è pari alla distanza fra 10 e 100 e così via;

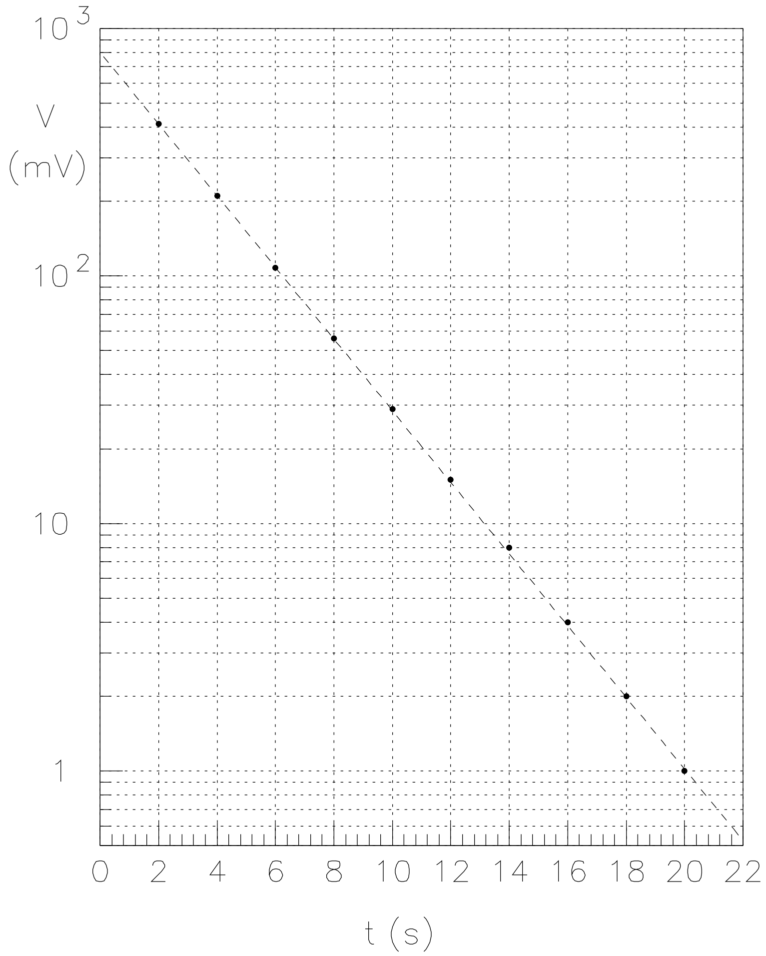

Figura 6.6: Misure di scarica di un condensatore riportate su carta semilogaritmica.

- gli intervalli fra potenze di 10 successive si chiamano *decadi* e, come detto, sono di lunghezza costante su scala logaritmica. La carta del nostro di figura 6.6 ha 3 decadi e mezza;
- il valore 0 non può comparire su una scala logaritmica in quanto $\lim_{x\to 0} \ln x = -\infty$.;
- due rette che risultano parallele su carta logaritmica differiscono di una costante per quanto riguarda i valori delle ascisse, mentre differiscono per un fattore moltiplicativo per quanto riguarda le ordinate. Infatti se abbiamo una legge del tipo $y = a\,e^{bx}$ ed effettuiamo una traslazione delle ascisse di una costante

$$x \longrightarrow x' = x + \delta$$

si ottiene

$$\begin{aligned} y \longrightarrow y' &= ae^{b(x+\delta)} = ae^{b\delta}e^{bx} \\ &= e^{b\delta}y \end{aligned} \qquad (6.21)$$

- nell'interpolare ad occhio fra le tacche contigue bisogna prestare attenzione al fatto che la scala non è lineare. È comunque vero che fra due punti molto vicini l'approssimazione lineare è abbastanza buona.

È da notare inoltre che, mentre la carta semilogaritmica è stata introdotta per risolvere problemi legati ad esponenziali (e quindi a *logaritmi naturali*), per quanto riguarda l'uso di carta logaritmica si è parlato poi di potenze di 10. Questo è dovuto a due ragioni:

- innanzitutto, da un punto di vista grafico, ciò che è lineare nel logaritmo naturale è anche lineare nel logaritmo decimale (si ricordi infatti che $\ln x = \ln 10 \log x$);

- in secondo luogo bisogna notare che l'affermazione fatta precedentemente secondo cui "l'ordinata è proporzionale al logaritmo del valore" non è perfettamente corretta e lascerebbe presupporre un assortimento infinito di carte millimetrate a seconda degli ordini di grandezza di interesse. Ma poiché su scala logaritmica passare da un ordine di grandezza all'altro corrisponde ad una traslazione, è preferibile disegnare le decadi, lasciando all'utilizzatore il compito di segnare la potenza di 10 di interesse.

Le carte logaritmiche commerciali offrono un certo numero di decadi (tipicamente da 2 a 4) e agli estremi di ogni decade è indicato sempre "10". È lo sperimentatore a scrivere la successione di potenze adatta a riportare le misure. Per esempio, nel caso della figura 6.6, se la tensione fosse stata misurata in Volt si sarebbe riportato sulla scale, da basso verso l'alto, 10^{-3}, 10^{-2}, 10^{-1} e 10^{0}.

Un'altra applicazione della carta semilog, oltre a quello di evidenziare andamenti esponenziali e di valutarne i parametri, è quello di presentare dati che variano di molti ordini di grandezza. Questo può essere mostrato dal confronto del grafico di figura 6.5 con quello di figura 6.6. Mentre su carta lineare i valori di tensione corrispondenti a tempi superiori a 14 ms si confondono con lo zero, su carta semilog tutti i valori hanno lo stesso grado di leggibilità, nel senso che l'*incertezza di lettura* dei valori è *in percentuale* circa *costante* in tutto l'intervallo.

Questa caratteristica della scala è utile non soltanto per grafici di una grandezza in funzione di un'altra, ma anche per rappresentare gli istogrammi. In questo caso vengono evidenziati piccoli effetti sulle code delle distribuzioni che è impossibile apprezzare su scala lineare. Si confrontino per esempio le figure 4.3 e 6.7 dei tempi di attesa per ottenere un certo numero di conteggi con un contatore a scintillazione. Si notino in particolare le piccole asimmetrie delle code ancora presenti per 100 conteggi.

6.7.3 Stima dei parametri

Se i punti hanno un andamento lineare su carta semilog si può tracciare la retta che meglio si adatta ad essi e quindi stimare i parametri dell'andamento esponenziale. Con la notazione di formula 6.15, si può verificare quanto segue.

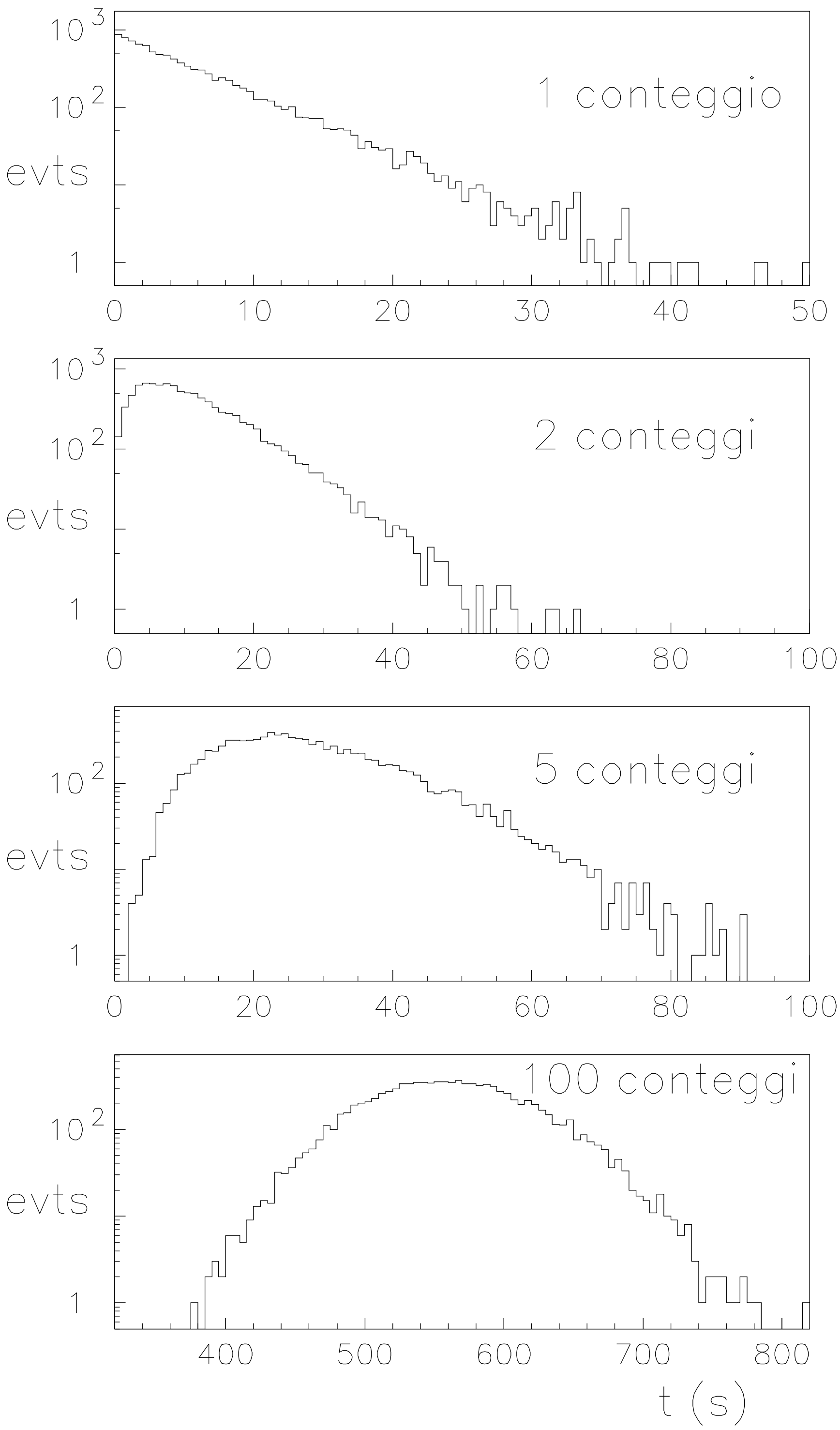

Figura 6.7: Istogrammi dei tempi di attesa per 1, 2, 5 e 100 conteggi in scala semilog. Si confronti con gli stessi istogrammi in scala lineare di figura 4.3.

- Il parametro a è dato semplicemente dal valore della Y che corrisponde all'intersezione della retta con l'asse delle ordinate
$$a = y(x = 0)\,. \tag{6.22}$$
- Il parametro b è dato dal coefficiente angolare della retta. Scegliendo due punti - si ricordi: lontani e ben leggibili - e utilizzando la 6.18 si ottiene, mediante l'equazione parametrica della retta
$$\frac{\ln y - \ln y_1}{\ln y_2 - \ln y_1} = \frac{x - x_1}{x_2 - x_1}\,, \tag{6.23}$$
$$b \;=\; \frac{\ln y_2 - \ln y_1}{x_2 - x_1} = \frac{\ln y_2/y_1}{x_2 - x_1}\,. \tag{6.24}$$
- Nel caso che l'asse delle ordinate ($x = 0$) per motivi di convenienza non sia riportato nel grafico può essere più comodo ricavarsi il parametro a dai due punti della retta:
$$\ln a \;=\; \ln y_1 - \frac{\ln y_2 - \ln y_1}{x_2 - x_1}x_1 = \ln y_1 - bx_1 \tag{6.25}$$
$$a \;=\; y_1 e^{-bx_1}\,. \tag{6.26}$$
(L'alternativa sarebbe di continuare la retta su altri fogli di carta millimetrata.)

Come esempio numerico stimiamo i parametri dell'andamento dei dati della tabella 6.2 e delle figure 6.6 e 6.5:

- dal valore della retta per $t = 0$ si trova $a \approx 800$ mV. L'incertezza della lettura della scala è dell'ordine di 10 mV. Per evitare che, scrivendo 800 mV, si possa pensare di aver apprezzato anche il secondo zero si può riportare il risultato come $a \approx 8.0 \cdot 10^2$ mV, ovvero
$$a \approx 0.80\ \text{V}\,.$$
- Scegliendo come punti della retta $P_1 = (2.0,\ 4.1{\cdot}10^2)$ e $P_2 = (20.0,\ 1.0)$:
$$b \approx \frac{\ln 1.0 - \ln 410}{(20.0 - 2.0)\,\text{ms}} = -0.33\,\text{ms}^{-1}\,.$$
A volte questo tipo di esponenziali negativi in funzione del tempo vengono descritti mettendo all'esponente un parametro definito positivo che abbia le stesse dimensioni del tempo e indicato usualmente τ:
$$V = V_\circ e^{-t/\tau}\,.$$
Nel nostro caso abbiamo $\tau \approx 3.0$ ms.
- Utilizziamo anche il secondo modo di ricavarci a:
$$a = 0.41 e^{2\times 0.33} = 0.79\,\text{V}\,,$$
in *buon accordo* con il valore di 0.80 V ottenuto con il metodo precedente.

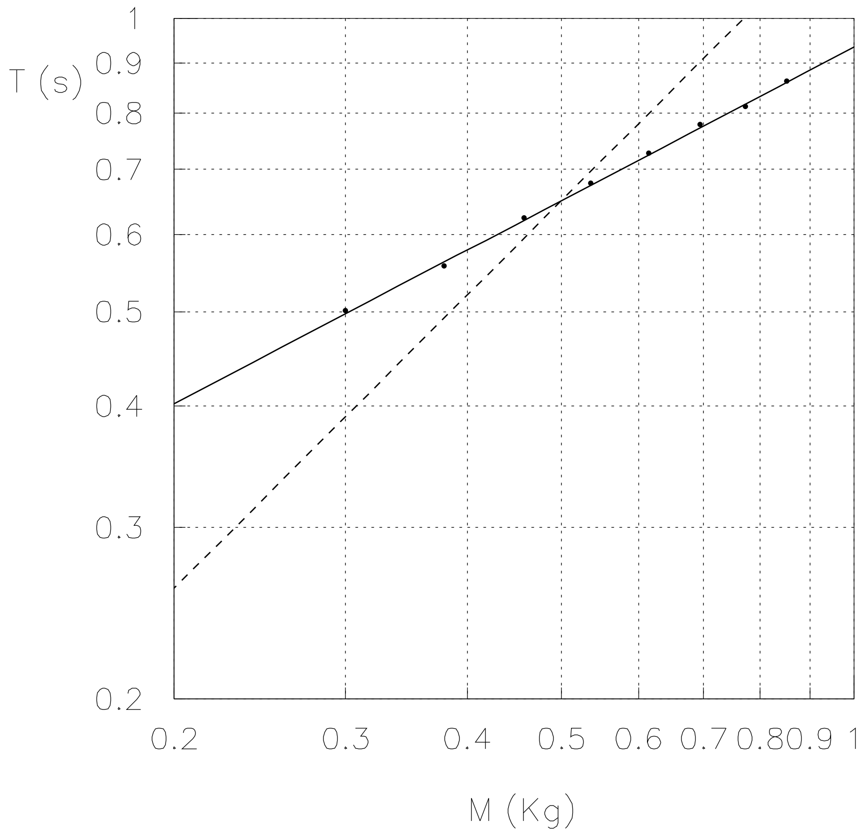

Figura 6.8: Dati della molla. Grafico su carta doppiologaritmica del periodo di oscillazione in funzione della massa applicata.

6.7.4 Carta doppiologaritmica

Un'altra importante classe di leggi fisiche comprende quelle descritte da *leggi di potenza*

$$y = \beta x^{\alpha} \,. \tag{6.27}$$

Esempi di queste sono la dipendenza della forza di gravità dalla distanza ($\alpha = -2$), la pressione di un gas perfetto in funzione del volume a temperatura costante ($\alpha = -1$), l'energia cinetica di un punto materiale in funzione della velocità ($\alpha = 2$), e così via.

Una legge di potenza può essere linearizzata prendendo i logaritmi di entrambi i membri della (6.27):

$$\log y = \log \beta + \alpha \log x \,. \tag{6.28}$$

In questo caso la relazione è formalmente la stessa qualsiasi sia la base dei logaritmi. C'è però da prestare un po' più di attenzione alle dimensioni delle grandezze. Se chiamiamo U_x e U_y le dimensioni di x e di y, possiamo riscrivere la (6.27)

$$\frac{y}{U_y} = \frac{\beta U_x^\alpha}{U_y}\frac{x^\alpha}{U_x^\alpha} \tag{6.29}$$

$$\log\frac{y}{U_y} = \log\frac{\beta U_x^\alpha}{U_y} + \log\frac{x^\alpha}{U_x^\alpha}\,. \tag{6.30}$$

Anche se quest'ultima relazione è quella che tiene conto correttamente delle dimensioni - e da cui segue che le dimensioni di β sono U_y/U_x^α - in genere si utilizza la (6.28), sottintendendo che le dimensioni siano aggiustate correttamente.

In analogia alla carta semilog di cui abbiamo parlato esiste la carta millimetrata *doppiologaritmica* (o *"log-log"*).

Una volta tracciata la retta che meglio approssima i dati sperimentali i parametri α e β si determinano in modo simile a quanto visto precedentemente:

- la potenza α è data dal coefficiente angolare della retta;

$$\alpha = \frac{\log y_2 - \log y_1}{\log x_2 - \log x_1} = \frac{\log y_2/y_1}{\log x_2/x_1} \tag{6.31}$$

- il <u>valore numerico</u> del coefficiente β è dato dall'intercetta, ovvero quello ottenuto in corrispondenza a $\log x = 0$ (o per essere precisi $\log x/U_x = 0$), ovvero per $x = 1$. Delle dimensioni di β si è parlato precedentemente. Notiamo per inciso che non ha invece senso il valore 0 della variabile su scala logaritmica: dove andrebbe collocato?

- anche in questo caso è utile ricavarsi una formula che dia il coefficiente β quando il valore $\log x = 0$ è fuori scala:

$$\log\beta = \log y_1 - \frac{\log y_2 - \log y_1}{\log x_2 - \log x_1}\log x_1 \tag{6.32}$$

$$\beta = y_1 x_1^{-\alpha}\,. \tag{6.33}$$

Come primo esempio di applicazione della carta doppio logaritmica torniamo all'esperienza della molla e verifichiamo che l'andamento del periodo di oscillazione in funzione dellla massa segua una legge di radice quadrata, ovvero $\alpha = 1/2$. Questa verifica è particolarmente interessante in quanto dalla figura 6.4 potrebbe sembrare che anche una legge lineare descriva altrettanto bene i dati sperimentali.

Il grafico di figura 6.8 mostra che i punti sperimentali sono ben allineati e i parametri della legge di potenza sono:

$$\alpha = \frac{\log 0.93 - \log 0.405}{\log 1.00 - \log 0.200} \approx 0.52 \tag{6.34}$$

$$\beta/U_\beta \approx 0.93\,. \tag{6.35}$$

Quindi la potenza dell'andamento $T = \beta M^\alpha$ è prossimo a 0.5 e quindi la (2.2) è da "ritenersi" verificata Per mostrare come un andamento lineare non possa descrivere i dati sperimenatli nella figura 6.8 è mostrata (linea tratteggiata) anche una retta avente $\alpha = 1$ e passante fra i punti.

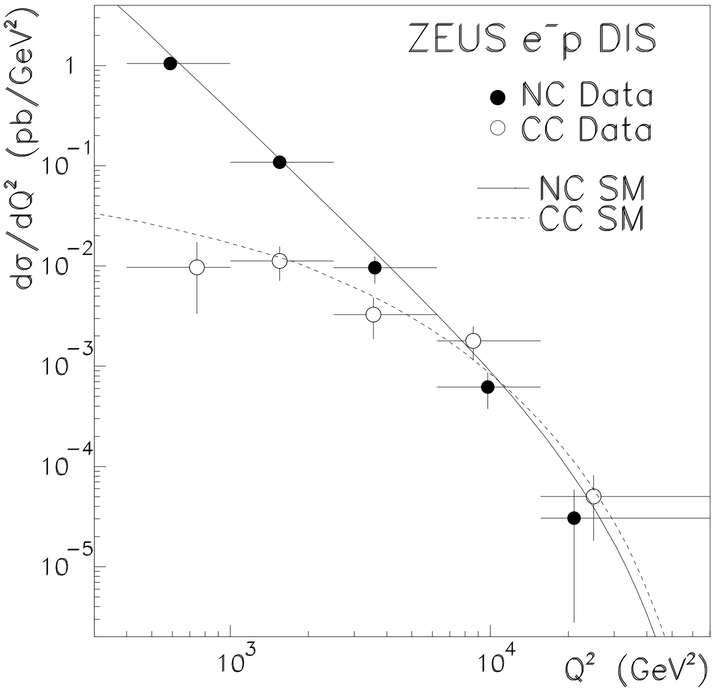

Figura 6.9: Esempio di recenti dati risultati di un esperimento di Fisica delle Particelle Elementari. Senza entrare nel significato fisico delle misure, si notino: l'uso di scale logaritmiche per riportare valori che si estendono su più ordini di grandezza; la convenzione standard di indicare le grandezze fisiche e le unità di di misura; i diversi simboli per indicare differenti misure e predizioni teoriche; le barre di incertezza sulle quantità riportata sulle ordinate. Le barre orizzontali in questo caso non rappresentano invece incertezza, ma l'intervallo dei valori delle ascisse nel quale è stato misurato il valore riportato sulle ordinate.

6.8 ❄ Altre linearizzazioni notevoli

Terminiamo questa introduzione all'analisi grafica dei dati sperimentali illustrando altri trucchi per linearizzare alcuni degli andamenti che capitano più di frequente in laboratorio. Per comodità includiamo nella lista anche gli andamenti esponenziali e di potenza trattati in dettaglio precedentemente. La notazione y Vs x va letta "y versus x", dove "*versus*" sta per "contro", "in funzione di".

funzione	linearizzazione	Note
$y = ae^{bx}$	$\log y$ Vs x	
$y = ax^b$	$\log y$ Vs $\log x$	
$y = y_F\left(1 - e^{-t/\tau}\right)$	$\log\left(y_F - y\right)$ Vs t	
$z = \frac{1}{2}at^2 + v_\circ t$	$\frac{z}{t}$ Vs t	
$y = a + bx + cx^2$	$\frac{y-y_1}{x-x1}$ Vs x	si ricavano b e c
$y = \frac{x}{a+bx}$	$\frac{1}{y}$ Vs $\frac{1}{x}$	
$y = \frac{x}{a+bx} + c$	$\frac{x-x_1}{y-y_1}$ Vs x	si ricavano a e b
$y = ae^{bx+cx^2}$	$\log\left[\left(\frac{y}{y_1}\right)^{1/(x-x_1)}\right]$ Vs x	si ricavano b e c

Tabella 6.3: Linearizzazioni notevoli. “Vs” - leggasi “versus” - sta per “in funzione di”.

6.9 Problemi

1. Trovate gli andamenti funzionali che meglio descrivono le seguenti coppie di punti sperimentali ($[x, y]$);
   (a) [4.00, 11.0], [4.40, 10.40], [4.80, 9.78], [5.30, 9.00], [5.85, 8.20];
   (b) [0.17, 80], [0.21, 46], [0.30, 20], [0.46, 7.0], [0.70, 2.5];
   (c) [1.0, 420], [3.1, 95], [5.3, 20], [8.0, 3.0], [10.5, 0.50];
   (d) [1.3, 10.0], [3.0, 7.0], [7.0, 3.0], [11.0, 1.30], [15.5, 0.50];
   (e) [2.0, 86.0], [4.2, 92.0], [6.0, 97.0], [7.2, 102.0], [9.1, 105.6];
   (f) [0.15, 0.25], [0.34, 4.0], [0.95, 7.0], [4.0, 16], [19.0, 38.0];
   (g) [0.8, 0.48], [2.5, 0.90], [5.0, 2.30], [8.0, 7.0], [10.8, 20];
   (h) [1.0, 10.0], [3.0, 3.33], [5.0, 2.00], [7.0, 1.43], [9.0, 11.1];
   (i) [10, $3.14 \cdot 10^{-2}$], [20, $7.85 \cdot 10^{-3}$], [30, $3.49 \cdot 10^{-3}$], [40, $1.96 \cdot 10^{-3}$], [50, $1.26 \cdot 10^{-3}$];
   (j) [2.0, 6.93], [4.0, 13.9], [6.0, 17.9], [8.0, 20.8], [10.0, 23.0], [12.0, 24.8];
2. Usando i risultati dei problemi precedenti trovare le leggi empiriche che legano media, deviazione standard e coefficiente di variazione delle distribuzioni di conteggi (tabelle 4.1 e 4.2) in funzione del tempo di misura.
3. Dalla prima serie di misure della tabella 2.4 ricavare accelerazione (a) e velocità iniziale ($v_\circ$) mediante un grafico del rapporto spazio/tempo in funzione del tempo. Valutare graficamente la velocità finale del corpo.
4. Dai dati della tabella 2.7 valutare le costanti di tempo di svuotamento per ciascuno dei tre volumi del recipiente. Valutare quindi, facendo uso della (2.8) la velocità di aspirazione della pompa e il volume di pompa e raccordi.

Soluzioni dei problemi

Capitolo 3

1. Letture in cm: A) 0.960; B) 2.320; C) 3.805; D) 4.980; E) 5.770; F) 6.880; G) 7.620; H) 8.630; I) 9.460; J) 10.370; K) 0.915; L) 1.900; M)3.170; N) 3.850; O) 4.885; P) 5.755; Q) 7.140; R) 8.350; S) 9.050; T) 9.775. Effetti dovuti a quantizzazioni di stampa potrebbero causare piccoli scarti (± 0.005 cm) da questi valori che però non alterano la sostanza dell'esercizio.
2. a) 3; b) 5; c) 3; d) 3; e) 1; f) 4; g) 4; h) 5 oppure 1; i) 5; j) 2; k) 7; l) 5 oppure 3; m) 1; n) 6; o) 2.
3. a) 31.0; b) 6; c) 3.15; d) 3.15405; e) $C = 49.61$ cm (il 2 della formula è, come π, esatto e non il risultato di una misura); f) 0.11; g) 0.862; h) 369; i) 2.69; j) $1.521 \cdot 10^4$; k) $1.91 \cdot 10^6$; l) $1.2 \cdot 10^{10}$; m) $2.9 \cdot 10^{265}$; n) $2.632 \cdot 10^{105}$; o) $7.42 \cdot 10^{-130}$; p) 0.99; q) 3.30; r) 54.755; s) 0.0153 s^{-1}; t) 0.05060 ms^{-1};
4. Circa 3˙530˙000 maschi e 3˙320˙000 femmine (diverse stime che differiscano da queste di qualche decina di migliaia di abitanti sono ugualmente compatibili con il diagramma).

Capitolo 5

1. $$\frac{\sum_i (x_i - \overline{x})}{N} = \frac{\sum_i x_i}{N} - \frac{N \cdot \overline{x}}{N} = 0\,.$$
2. 7; 7; 7.2; 2.6 .
3. 7; 7; 6.1; 3.5 .
4. 7; 6.5; 5.8; 3.4 .
5. 6 e 7 (bimodale); 6, 5.8, 3.3 .
6. i valori di moda, mediana, $\sum x$, $\sum x^2$, $\overline{x}$, σ_N e v sono:

   6 s: 1; 1; 102; 190; 1.02; 0.93; 0.91 (91 %);

   12 s: 2; 2; 219; 663; 2.19; 1.35; 0.62 (62 %);

   30 s: 4; 5; 516; 3126; 5.16; 2.15; 0.42 (42 %);
7. 18; 18; 17.8; 4.2; 0.24 (24 %) .
8. 51, 52 e 61 (multimodale); 53; 53.5; 7.5; 0.14 (14 %) .
9. $\overline{x} = 1000000.24$, $\sigma_N = 0.10$. Perché con la calcolatrice (in genere) non funziona? Utilizzare le proprietà di trasformazione di media e deviazione standard.
10. $\overline{x} = -0.025$ mm; $\sigma_N = 0.098$ mm; il valore 44.76 mm potrebbe avere quindi un errore tipico di 0.1 mm. L'errore medio di -0.025 mm è ben al di sotto delle fluttuazioni della singola lettura (oltre al fatto che esso non è significativamente diverso da 0, come si vedrà).

Capitolo 6

1. (a) $y = -1.513 \cdot x + 17.05$;
   (b) $y = 1.04 \cdot x^{-2.2}$;
   (c) $y = 8.7\mathrm{e}^{-0.71 \cdot x}$;
   (d) $y = 13.2\mathrm{e}^{-0.211 \cdot x}$;
   (e) $y = 2.76 \cdot x + 80.5$;
   (f) $y = 7.3 \cdot x^{0.56}$;
   (g) $y = 0.34\mathrm{e}^{0.37 \cdot x}$;
   (h) $y = 10.0/x$;
   (i) $y = 3.14/x^2$;
   (j) $y = 10 \ln x$.
2. Media: $\propto$ lineare; $\sigma_N \propto \sqrt{T}$; $v \propto 1/\sqrt{T}$.
3. $a \approx 4.7$ cm/s^2; $v_\circ 1.0$ cm/s; $v_f \approx 14.15$ cm/s.
4. $\tau = 8.2$, 12.5 e 26.7 s rispettivamente; $S_p = 20.5$ ml/s, $V_\circ = 40$ ml (scarti dell'ordine di grandezza della cifra meno significativa sono nella norma).

Capitolo 7

1. ....

Capitolo 8

1. ....
2. Dall'uguaglianza dei momenti delle forze: $M_x = SqrtM_1 \times M_2$ (media geometrica).

Capitolo 10

1. Non corretti: b), g), h) e m).
2. $\leq 7.4\,\%$. (Che distribuzione di probabilità seguono le misure?)
3. 143 ± 25;
   $n = 5$: 143 ± 49.
4. $68.3\,\%$ C.L.: $t = (12.9 \pm 0.9)\,s$, ovvero $12.1 \leq t \leq 13.8\,s$;
   $95\,\%$ C.L.: $t = (12.9 \pm 2.3)\,s$, ovvero $10.6 \leq t \leq 15.2\,s$;