

INDICE GENERALE

Prima parte ■ INTRODUZIONE ALLA STATISTICA

Capitolo 1

La statistica e i campioni 2

1.1 Che cos'è la statistica? 2

1.2 Campionare le popolazioni 3

- Popolazioni e campioni 4
- Le proprietà dei buoni campioni 4
- Il campionamento casuale 5
- Come ottenere un campione casuale 6
- Il campione di convenienza 7
- La distorsione introdotta dal volontario 7
- I dati reali in biologia 8

1.3 I tipi di dati e le variabili 8

- Variabili qualitative e variabili quantitative 8
- Variabili risposta e variabili esplicative 9

1.4 Distribuzioni di frequenza e distribuzioni di probabilità 9

1.5 I tipi di studi 10

SINTESI 11

PROBLEMI 12

FOCUS 1 *La correlazione non richiede una relazione di causa ed effetto* 16

Capitolo 2

La visualizzazione dei dati 18

2.1 Le regole principali per una visualizzazione efficace dei dati 18

- Come costruire un cattivo diagramma 18
- Come costruire un buon diagramma 19

2.2 Visualizzare i dati di una variabile 21

- Visualizzare una variabile categorica: tabella di frequenza e diagramma a barre 21
- Costruire un buon diagramma a barre 22
- Un diagramma a barre solitamente è migliore di un diagramma a torta 22
- Visualizzare una variabile numerica: tabelle di frequenza e istogrammi 22
- Descrivere la forma di un istogramma 23
- Come costruire un buon istogramma 24
- Altri diagrammi per una variabile numerica 25

2.3 Visualizzare l'associazione tra due variabili e le differenze tra gruppi 25

- Visualizzare l'associazione tra due variabili categoriche 25
- Visualizzare l'associazione tra variabili numeriche: diagramma di dispersione (*scatter plot*) 27
- Visualizzare l'associazione tra una variabile numerica e una variabile categorica 27

2.4 Visualizzare tendenze nel tempo e nello spazio 29

2.5 Come costruire una buona tabella 30

- Le regole per le tabelle di visualizzazione 30

2.6 Come creare file di dati 32

SINTESI 33

PROBLEMI 33

Capitolo 3

Descrivere i dati 42

3.1 La media aritmetica e la deviazione standard 42

- La media campionaria 43
- La varianza e la deviazione standard 43
- L'arrotondamento delle medie, delle deviazioni standard e di altre quantità 44
- Il coefficiente di variazione 45
- Il calcolo della media e della deviazione standard a partire da una tabella di frequenza 45
- L'effetto del cambio di unità di misura 46

3.2 La mediana e la differenza interquartile 46

- La mediana 47
- La differenza interquartile 47
- Il diagramma box plot 48

3.3 Il confronto tra misure di posizione e di dispersione 48

- Il confronto tra media e mediana 49
- Il confronto tra deviazione standard e differenza interquartile 50

3.4 La distribuzione di frequenza cumulativa 50

- Percentili e quantili 50
- Visualizzare le frequenze relative cumulative 50

3.5 Le proporzioni	51
■ Il calcolo di una proporzione	51
■ La proporzione è come una media campionaria	51
SINTESI	52
FORMULARIO	52
PROBLEMI	53

Capitolo 4

La stima con incertezza: le basi dell'inferenza statistica

4.1 La distribuzione campionaria di una stima	60
■ La stima della lunghezza media dei geni con un campione casuale	61
■ La distribuzione campionaria di \bar{Y}	62
4.2 Come misurare l'incertezza di una stima	63
■ L'errore standard	63
■ L'errore standard di \bar{Y}	63
■ L'errore standard di \bar{Y} calcolato a partire dai dati	64
4.3 L'intervallo di confidenza	64
■ La regola pratica dei 2ES	65
4.4 Le barre di errore	65
SINTESI	66
FORMULARIO	66
PROBLEMI	67
FOCUS 2 <i>La pseudoreplicazione</i>	71

Capitolo 5

La probabilità

5.1 La probabilità di un evento	73
5.2 I diagrammi di Venn	74
5.3 Gli eventi incompatibili	74
5.4 Le distribuzioni di probabilità	75
■ Le distribuzioni di probabilità discrete	75
■ Le distribuzioni di probabilità continue	75
5.5 Va bene sia questo sia quello: sommare le probabilità	76
■ La regola della somma	76
■ La somma delle probabilità di tutti i possibili eventi incompatibili è pari a 1	77
■ La regola della somma generalizzata	77
5.6 L'indipendenza e la regola del prodotto	77
■ La regola del prodotto	78
■ Gli enunciati che implicano «e» o «o»	79
■ L'indipendenza di più di due eventi	79
5.7 Gli alberi di probabilità	79



5.8 Gli eventi dipendenti	80
5.9 La probabilità condizionata e il teorema di Bayes	81
■ La probabilità condizionata	82
■ La regola del prodotto generalizzata	82
■ Il campionamento senza reimmissione	83
■ Il teorema di Bayes	83
SINTESI	84
PROBLEMI	85

Capitolo 6

La verifica delle ipotesi

6.1 La formulazione e l'impiego delle ipotesi statistiche	92
■ L'ipotesi nulla	92
■ L'ipotesi alternativa	92
■ Rifiutare o non rifiutare	92
6.2 La verifica delle ipotesi: un esempio	93
■ Formulare le ipotesi	93
■ La statistica test	93
■ La distribuzione nulla	94
■ La quantificazione dell'incertezza: il <i>P</i> -value	95
■ Trarre le conclusioni	95
■ Comunicare i risultati	96
6.3 Gli errori nella verifica delle ipotesi	96
■ Gli errori di tipo I e di tipo II	96
6.4 Quando l'ipotesi nulla non viene scartata	97
■ Il test	97
■ Come interpretare un risultato non significativo	98
6.5 I test unilaterali	99
6.6 La verifica delle ipotesi e gli intervalli di confidenza	100
SINTESI	100
PROBLEMI	101
FOCUS 3 <i>Perché significatività statistica non è sinonimo di importanza biologica</i>	106

Seconda parte ■ PROPORZIONI E FREQUENZE

Capitolo 7

L'analisi delle proporzioni 108

- 7.1** La distribuzione binomiale 108
 - La formula della distribuzione binomiale 109
 - Il numero di successi in un campione casuale 109
 - La distribuzione campionaria di una proporzione 110
- 7.2** Lo studio di una proporzione: il test binomiale 111
 - Le approssimazioni per il test binomiale 113
- 7.3** La stima delle proporzioni 113
 - La stima dell'errore standard di una proporzione 113
 - L'intervallo di confidenza di una proporzione: il metodo di Agresti e Coull 114
 - L'intervallo di confidenza di una proporzione: il metodo di Wald 114
- 7.4** Come si ricava la distribuzione binomiale 114
- SINTESI** 115
- FORMULARIO** 115
- PROBLEMI** 116
- FOCUS 4** *La biologia e la storia della statistica* 121

Capitolo 8

L'adattamento di modelli probabilistici a dati di frequenza 122

- 8.1** Il test del χ^2 di bontà di adattamento: il modello proporzionale 122
 - L'ipotesi nulla e l'ipotesi alternativa 123
 - Frequenze osservate e frequenze attese 123
 - La statistica test χ^2 124
 - La distribuzione campionaria di χ^2 sotto l'ipotesi nulla 124
 - Il calcolo del P -value 125
 - I valori critici per la distribuzione χ^2 126
- 8.2** Le assunzioni del test χ^2 di bontà di adattamento 127
- 8.3** Il test di bontà di adattamento quando vi sono soltanto due categorie 127



- 8.4** La casualità nello spazio o nel tempo: la distribuzione di Poisson 128
 - La formula della distribuzione di Poisson 129
 - La verifica della casualità con la distribuzione di Poisson 129
 - Il confronto tra media e varianza 131
- SINTESI** 131
- FORMULARIO** 132
- PROBLEMI** 132
- FOCUS 5** *Pianificare l'esperimento* 138

Capitolo 9

L'analisi di contingenza: associazione tra variabili categoriche 139

- 9.1** L'associazione tra due variabili categoriche 139
- 9.2** La stima dell'associazione nelle tabelle 2×2 : il rischio relativo 140
 - Il rischio relativo 140
 - La riduzione del rischio 141
- 9.3** La stima dell'associazione nelle tabelle 2×2 : l'odds ratio 141
 - L'odds 141
 - L'odds ratio 142
 - L'errore standard e l'intervallo di confidenza dell'odds ratio 143
 - Il rischio relativo vs l'odds ratio 143
- 9.4** Il test di contingenza del χ^2 145
 - Le ipotesi 146
 - Le frequenze attese assumendo indipendenza 146
 - La statistica χ^2 146
 - I gradi di libertà 147
 - P -value e conclusione 147
 - Una formula rapida per il calcolo delle frequenze attese 147
 - Il test di contingenza del χ^2 è un caso particolare del test del χ^2 di bontà di adattamento 147
 - Le assunzioni del test di contingenza del χ^2 148
 - La correzione per la continuità 148
- 9.5** Il test esatto di Fisher 148
- SINTESI** 149
- FORMULARIO** 150
- PROBLEMI** 150
- PROBLEMI DI RIEPILOGO** 158

Terza parte ■ CONFRONTO DI VALORI NUMERICI

Capitolo 10

La distribuzione normale 162

- 10.1** Le curve a campana e la distribuzione normale 162
- 10.2** La formula della distribuzione normale 163
- 10.3** Le proprietà della distribuzione normale 164
- 10.4** La distribuzione normale standardizzata e le tavole statistiche 164
 - Come utilizzare le tavole della distribuzione normale standardizzata 164
 - La distribuzione normale standardizzata per descrivere qualsiasi distribuzione normale 166
- 10.5** La distribuzione delle medie campionarie è normale 167
 - Calcolare le probabilità delle medie campionarie 168
- 10.6** Il teorema del limite centrale 169
- 10.7** L'approssimazione normale per la distribuzione binomiale 170

SINTESI 172

FORMULARIO 173

PROBLEMI 173

FOCUS 6 *I controlli negli studi medici* 178

Capitolo 11

L'inferenza in una popolazione con distribuzione normale 179

- 11.1** La distribuzione t per le medie campionarie 179
 - La distribuzione t di Student 179
 - Come trovare i valori critici della distribuzione t 180
- 11.2** L'intervallo di confidenza della media di una distribuzione normale 181
 - L'intervallo di confidenza al 95% della media 182
 - L'intervallo di confidenza al 99% della media 182
- 11.3** Il test t per un campione 183
 - Gli effetti dell'aumento della numerosità campionaria: l'esempio della temperatura corporea rivisitato 185
- 11.4** Le assunzioni del test t per un campione 185
- 11.5** La stima della deviazione standard e della varianza in una popolazione con distribuzione normale 185
 - I limiti di confidenza della varianza 186
 - I limiti di confidenza della deviazione standard 186
 - Le assunzioni 187

SINTESI 187

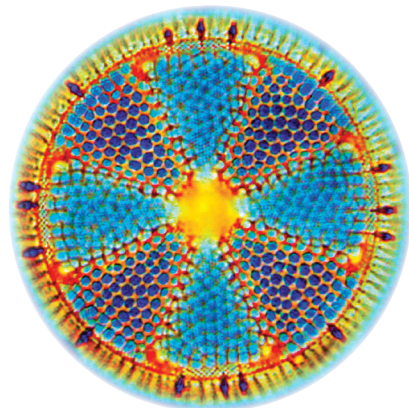
FORMULARIO 187

PROBLEMI 188

Capitolo 12

Il confronto tra due medie 193

- 12.1** Dati appaiati o campioni indipendenti? 193
 - 12.2** Il confronto appaiato di medie 194
 - La stima della differenza media in dati appaiati 195
 - Il test t per dati appaiati 196
 - Le assunzioni 197
 - 12.3** Il confronto tra le medie in due campioni 197
 - L'intervallo di confidenza per la differenza tra due medie 198
 - Il test t per due campioni 199
 - Le assunzioni 200
 - Il test di Welch 200
 - 12.4** Quali sono le unità campionarie corrette? 201
 - 12.5** L'errore del confronto indiretto 203
 - 12.6** Come interpretare la sovrapposizione tra intervalli di confidenza 204
 - 12.7** Confrontare le varianze 204
 - Il test F per l'uguaglianza tra due varianze 204
 - Il test di Levene per l'omogeneità delle varianze 205
- SINTESI** 205
- FORMULARIO** 206
- PROBLEMI** 207
- FOCUS 7** *Quale test devo usare?* 217



Capitolo 13

Che cosa fare quando le assunzioni di un'analisi statistica vengono violate

13.1 Identificare le violazioni della normalità	219
■ I metodi grafici	219
■ I test statistici di normalità	221
13.2 Quando è possibile ignorare le violazioni delle assunzioni	222
■ La violazione della normalità	222
■ La disuguaglianza delle deviazioni standard	223
13.3 Le trasformazioni dei dati	223
■ La trasformazione logaritmica	223
■ Le altre trasformazioni	225
■ Gli intervalli di confidenza nel caso di trasformazioni	225
■ È importante evitare di fare molti test con diverse trasformazioni	226
13.4 Le alternative non parametriche al test t per un campione e al test t per dati appaiati	226
■ Il test dei segni	226
■ Il test dei segni per ranghi di Wilcoxon	228
13.5 Il confronto tra due gruppi: il test U di Mann-Whitney	228
■ I ranghi con lo stesso valore (<i>tied ranks</i>)	230
■ Campioni grandi e approssimazione normale	231
13.6 Le assunzioni dei test non parametrici	231
13.7 Gli errori di tipo I e di tipo II nei test non parametrici	231
13.8 Il test di permutazione	232
■ Le assunzioni dei test di permutazione	234
SINTESI	234
FORMULARIO	235
PROBLEMI	235

Capitolo 14

Il disegno sperimentale

14.1 Gli insegnamenti delle sperimentazioni cliniche	247
■ Le componenti del disegno sperimentale	247
14.2 Come ridurre la distorsione	248
■ Il gruppo di controllo simultaneo	248
■ La randomizzazione	249
■ La cecità	250
14.3 Come ridurre l'influenza dell'errore di campionamento	251
■ La replicazione	251
■ Il bilanciamento	252
■ Il raggruppamento in blocchi	252

■ I trattamenti estremi	254
14.4 Gli esperimenti con più di un fattore	254
14.5 Che cosa fare quando non si possono condurre esperimenti?	256
■ Appaiamento e aggiustamento	256
14.6 La scelta della dimensione campionaria	256
■ Pianificare la precisione	257
■ Pianificare la potenza	258
■ Pianificare per compensare la perdita di dati	259
SINTESI	259
FORMULARIO	260
PROBLEMI	261
FOCUS 8 <i>Data dredging: il problema dei test multipli</i>	265

Capitolo 15

Il confronto tra le medie di più di due gruppi

15.1 L'analisi della varianza	267
■ Le ipotesi	267
■ L'ANOVA in breve	268
■ La tabella dell'ANOVA	269
■ Scomporre la somma dei quadrati	270
■ Calcolare la media dei quadrati	271
■ Il rapporto tra varianze, F	271
■ La variabilità spiegata: R^2	272
■ L'ANOVA per due gruppi	273
15.2 Assunzioni e alternative	273
■ La robustezza dell'ANOVA	273
■ Le trasformazioni dei dati	273
■ Le alternative non parametriche all'ANOVA	273
15.3 I confronti pianificati	274
■ Il confronto pianificato tra due medie	274
15.4 I confronti non pianificati	275
■ Confrontare tutte le coppie di medie con il metodo di Tukey-Kramer	275
■ Le assunzioni	276
15.5 Effetti fissi ed effetti casuali	276
15.6 L'ANOVA per gruppi scelti casualmente	277
■ I calcoli dell'ANOVA	278
■ Le componenti della varianza	278
■ La ripetibilità	279
■ Le assunzioni	279
SINTESI	279
FORMULARIO	280
PROBLEMI	281
FOCUS 9 <i>Gli errori durante l'esperimento e durante l'analisi statistica</i>	290
PROBLEMI DI RIEPILOGO	291

Quarta parte ■ REGRESSIONE E CORRELAZIONE

Capitolo 16

La correlazione tra variabili numeriche

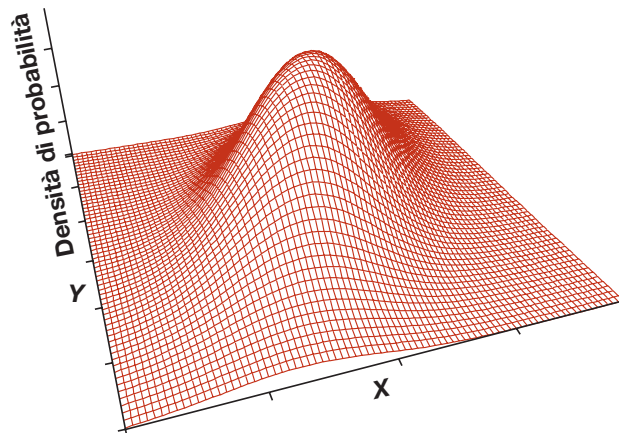
16.1 La stima del coefficiente di correlazione lineare	296
<ul style="list-style-type: none"> ■ Il coefficiente di correlazione 296 ■ L'errore standard 298 ■ L'intervallo di confidenza approssimato 299 	
16.2 Verificare l'ipotesi nulla di assenza di correlazione	299
16.3 Le assunzioni	301
16.4 Il coefficiente di correlazione dipende dall'intervallo di valori considerato	302
16.5 La correlazione per ranghi di Spearman	302
<ul style="list-style-type: none"> ■ La procedura per n grande 304 ■ Le assunzioni della correlazione per ranghi di Spearman 304 	
16.6 Gli effetti dell'errore di misura sulla correlazione	304
SINTESI	305
FORMULARIO	305
PROBLEMI	306
FOCUS 10 <i>Il bias di pubblicazione</i>	315

Capitolo 17

La regressione

17.1 La regressione lineare	318
<ul style="list-style-type: none"> ■ Il metodo dei minimi quadrati 319 ■ L'equazione della retta di regressione 319 ■ Calcolo di pendenza e intercetta della retta di regressione 320 	

<ul style="list-style-type: none"> ■ Popolazioni e campioni 321 ■ I valori previsti 321 ■ I residui 322 ■ L'errore standard della pendenza 322 ■ L'intervallo di confidenza della pendenza 322 	
17.2 La qualità delle previsioni	322
<ul style="list-style-type: none"> ■ Gli intervalli di confidenza delle previsioni 323 ■ L'estrapolazione 324 	
17.3 La verifica delle ipotesi sulla pendenza	324
<ul style="list-style-type: none"> ■ Il test t per la pendenza della retta di regressione 325 ■ L'approccio ANOVA applicato alla regressione 325 ■ L'uso di R^2 per misurare l'adattamento della retta di regressione ai dati 326 	
17.4 La regressione verso la media	326
17.5 Le assunzioni della regressione	327
<ul style="list-style-type: none"> ■ Gli outlier 328 ■ L'identificazione della non linearità 328 ■ L'identificazione della non normalità e di varianze disuguali 328 	
17.6 Le trasformazioni	329
17.7 Gli effetti dell'errore di misura sulla regressione	330
17.8 La regressione non lineare	330
<ul style="list-style-type: none"> ■ Le curve con asintoto 331 ■ Le curve quadratiche 332 ■ L'adattamento di una curva senza formula 332 	
17.9 Regressione logistica: funzioni per una variabile risposta di tipo binario	333
SINTESI	335
FORMULARIO	335
PROBLEMI	337
FOCUS 11 <i>La meta-analisi</i>	348
PROBLEMI DI RIEPILOGO	350



Quinta parte ■ MODERNI METODI STATISTICI

Capitolo 18

L'analisi con più di una variabile esplicativa

18.1 L'ANOVA e la regressione lineare sono modelli lineari	356
■ Il modello di regressione lineare	357
■ La generalizzazione della regressione lineare	357
■ I modelli lineari	359
18.2 L'analisi di esperimenti con raggruppamento in blocchi (blocking)	359
■ L'analisi dei dati ottenuti con un disegno a blocchi randomizzati	359
■ La formula del modello	360
■ L'adattamento del modello ai dati	360
18.3 L'analisi di disegni fattoriali	361
■ La formulazione del modello	362
■ I test sui fattori	363
■ L'importanza di distinguere tra fattori fissi e fattori casuali	364
18.4 Come considerare gli effetti di una covariata	364
■ Testare l'interazione	365
■ Adattare il modello senza il termine di interazione	366
18.5 Le assunzioni dei modelli lineari	367
SINTESI	368
PROBLEMI	368
FOCUS 12 <i>Le specie diverse come singole osservazioni nelle analisi statistiche</i>	375

Capitolo 19

I metodi computazionali intensivi

19.1 La verifica delle ipotesi con la simulazione	377
19.2 Il bootstrap per calcolare errori standard e intervalli di confidenza	379
■ L'errore standard di bootstrap	381
■ Gli intervalli di confidenza di bootstrap	381
■ Il bootstrap nel caso di gruppi multipli	382
■ Le assunzioni e i limiti del bootstrap	383
SINTESI	383
PROBLEMI	383

Capitolo 20

La verosimiglianza

20.1 Che cos'è la verosimiglianza?	388
20.2 Due applicazioni della verosimiglianza in biologia	389

■ La stima di una filogenesi	389
■ La mappatura genica	389
20.3 La stima di massima verosimiglianza	390
■ Il modello probabilistico	390
■ La formula per la verosimiglianza	391
■ La stima di massima verosimiglianza	391
■ Gli intervalli di confidenza basati sulla verosimiglianza	392
20.4 La versatilità della stima di massima verosimiglianza	393
■ Il modello probabilistico	393
■ La formula per la verosimiglianza	394
■ La stima di massima verosimiglianza	394
■ La distorsione	394
20.5 Il test del rapporto di verosimiglianza	395
■ La statistica per il test del rapporto di verosimiglianza	395
■ Testare una proporzione	395
SINTESI	396
FORMULARIO	396
PROBLEMI	397

Capitolo 21

L'analisi di sopravvivenza

21.1 Le curve di sopravvivenza	403
■ Riepilogo dei calcoli	404
■ Gli intervalli di confidenza	405
■ La mediana del tempo di sopravvivenza	405
■ Le assunzioni	405
21.2 Paragonare le curve di sopravvivenza	406
■ Rapporto di rischio	406
■ Il calcolo del rapporto di rischio	407
■ Il test di Logrank	408
■ Le assunzioni	409
SINTESI	409
FORMULARIO	410
PROBLEMI	410
RISPOSTE AI PROBLEMI	415
RISPOSTE AI PROBLEMI DI RIEPILOGO	441
BIBLIOGRAFIA	447
TAVOLE STATISTICHE	460
INDICE ANALITICO	475
FONTI DELLE ILLUSTRAZIONI	482