

Indice

Parte I

Capitolo 1 Statistica e campioni

■ 1.1 Che cos'è la statistica?	2
■ 1.2 Campionare le popolazioni	3
• Popolazioni e campioni	4
• 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
• Dati reali in biologia	8
■ 1.3 Tipi di dati e 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 Tipi di studi	10
■ 1.6 Riepilogo	10
Problemi di approfondimento	11
Ulteriori problemi	12
■ Scheda 1 <i>La biologia e la storia della statistica</i>	14

Capitolo 2 La visualizzazione dei dati

■ 2.1 Rappresentare le distribuzioni di frequenza	16
• Tabelle di frequenza e diagrammi a barre per dati categorici	16
• Semplici regole per costruire un diagramma a barre	16
• Tabelle di frequenza e istogrammi per dati numerici	17
• Descrivere la forma di un istogramma	18
• L'ampiezza degli intervalli può influenzare la forma di un istogramma	19
• Semplici regole per costruire un istogramma	20
■ 2.2 Quantili di una distribuzione di frequenza	20
• Percentili e quantili	20
• La distribuzione di frequenza cumulativa	20
■ 2.3 Associazione tra variabili categoriche	21
• Tabelle di contingenza	21
• Il diagramma a barre raggruppate	22
• Il grafico a mosaico (mosaic plot)	22

■ 2.4 Confrontare variabili numeriche tra gruppi	23
• Confronto di istogrammi tra gruppi	23
• Confronto delle frequenze cumulative	23
■ 2.5 Visualizzare la relazione tra una coppia di variabili numeriche	24
• Il diagramma di dispersione (o scatter plot)	24
• Il diagramma a linee	24
• Le mappe	25
■ 2.6 Le regole principali per una visualizzazione efficace dei dati	25
• Le regole per la visualizzazione grafica	26
• Le regole per le tabelle sono simili	27
■ 2.6 Riepilogo	28
Problemi di approfondimento	29
Ulteriori problemi	32

Capitolo 3 Descrivere i dati

■ 3.1 La media aritmetica e la deviazione standard	37
• La media campionaria	37
• La varianza e la deviazione standard	37
• Arrotondamento delle medie, delle deviazioni standard e di altre quantità	39
• Il coefficiente di variazione	39
• Calcolo della media e della deviazione standard a partire da una tabella di frequenza	39
■ 3.2 La mediana e la differenza interquartile	40
• La mediana	40
• La differenza interquartile	41
• Il diagramma box plot	41
■ 3.3 Confronto tra misure di posizione e di dispersione	42
• Confronto tra media e mediana	42
• Confronto tra deviazione standard e differenza interquartile	43
■ 3.4 Le proporzioni	44
• Calcolo di una proporzione	44
• La proporzione si può vedere come una media campionaria	44
■ 3.5 Riepilogo	44
■ 3.6 Riepilogo delle formule	44
Problemi di approfondimento	46
Ulteriori problemi	48

Capitolo 4
La stima con incertezza:
le basi dell'inferenza statistica

■ 4.1 La distribuzione campionaria di una stima	52
<ul style="list-style-type: none"> • Stima della lunghezza media dei geni con un campione casuale • La distribuzione campionaria di \bar{Y} 	52 53
■ 4.2 Come misurare l'incertezza di una stima	54
<ul style="list-style-type: none"> • L'errore standard • L'errore standard di \bar{Y} • L'errore standard di \bar{Y} calcolato a partire dai dati 	54 54 54
■ 4.3 L'intervallo di confidenza	55
<ul style="list-style-type: none"> • La regola pratica dei 2ES 	56
■ 4.4 Riepilogo	56
■ 4.5 Riepilogo delle formule	56
<ul style="list-style-type: none"> Problemi di approfondimento Ulteriori problemi 	57 57
■ Scheda 2 Pseudoreplicazione	59

Capitolo 5
La probabilità

■ 5.1 La probabilità di un evento	61
■ 5.2 I diagrammi di Venn	62
■ 5.3 Eventi incompatibili	62
■ 5.4 Le distribuzioni di probabilità	63
<ul style="list-style-type: none"> • Distribuzioni di probabilità discrete • Distribuzioni di probabilità continue 	63 63
■ 5.5 Va bene sia questo che quello: sommare le probabilità	64
<ul style="list-style-type: none"> • La regola della somma • La somma delle probabilità di tutti i possibili eventi incompatibili è pari a 1 • La regola della somma generalizzata (teorema della somma delle probabilità) 	64 65 65
■ 5.6 L'indipendenza e la regola del prodotto	65
<ul style="list-style-type: none"> • La regola del prodotto • Enunciati che implicano «e» o «oppure» • Indipendenza di più di due eventi 	66 66 66
■ 5.7 Alberi di probabilità	67
■ 5.8 Eventi dipendenti	68
■ 5.9 La probabilità condizionata e il teorema di Bayes	70
<ul style="list-style-type: none"> • La probabilità condizionata • La regola del prodotto generalizzata (teorema della probabilità composta) • Il teorema di Bayes 	70 70 71

■ 5.10 Riepilogo	72
<ul style="list-style-type: none"> Problemi di approfondimento Ulteriori problemi 	73 75

Capitolo 6
La verifica delle ipotesi

■ 6.1 La formulazione e l'impiego delle ipotesi	78
<ul style="list-style-type: none"> • L'ipotesi nulla • L'ipotesi alternativa • Rifiutare o non rifiutare 	78 78 78
■ 6.2 La verifica delle ipotesi: un esempio	79
<ul style="list-style-type: none"> • Formuliamo le ipotesi • La statistica test • La distribuzione nulla • La quantificazione dell'incertezza: il <i>P</i>-value • La significatività statistica • Comunicare i risultati 	79 79 79 80 81 82
■ 6.3 Gli errori nella verifica delle ipotesi	82
<ul style="list-style-type: none"> • Errori di tipo I e di tipo II 	82
■ 6.4 Quando l'ipotesi nulla non viene scartata	83
<ul style="list-style-type: none"> • Il test • Come interpretare un risultato non significativo 	83 84
■ 6.5 I test unilaterali	84
■ 6.6 Verifica delle ipotesi e intervalli di confidenza	85
■ 6.7 Riepilogo	86
<ul style="list-style-type: none"> Problemi di approfondimento Ulteriori problemi 	87 88
■ Scheda 3 Perché significatività statistica non è sinonimo di importanza biologica	89

Parte II

Capitolo 7
L'analisi delle proporzioni

■ 7.1 La distribuzione binomiale	92
<ul style="list-style-type: none"> • La formula della distribuzione binomiale • Il numero di successi in un campione casuale • La distribuzione campionaria di una proporzione 	93 93 94
■ 7.2 Lo studio di una proporzione: il test binomiale	95
<ul style="list-style-type: none"> • Approssimazioni per il test binomiale 	97
■ 7.3 La stima delle proporzioni	97
<ul style="list-style-type: none"> • La stima dell'errore standard di una proporzione • L'intervallo di confidenza di una proporzione: il metodo di Agresti e Coull 	97 98

• L'intervallo di confidenza di una proporzione: il metodo di Wald	98		
■ 7.4 Come si ricava la distribuzione binomiale	98		
■ 7.5 Riepilogo	99		
■ 7.6 Riepilogo delle formule	99		
Problemi di approfondimento	100		
Ulteriori problemi	102		
■ Scheda 4 La correlazione non richiede una relazione di causa ed effetto	105		
<hr/>			
Capitolo 8			
Adattamento di modelli probabilistici a dati di frequenza			
<hr/>			
■ 8.1 Un esempio di modello casuale: il modello proporzionale	107		
■ 8.2 Il test del χ^2 di bontà di adattamento	108		
• L'ipotesi nulla e l'ipotesi alternativa	108		
• Frequenze osservate e frequenze attese	108		
• La statistica test χ^2	109		
• La distribuzione campionaria di χ^2 sotto l'ipotesi nulla	110		
• Il calcolo del P -value	110		
• I valori critici per la distribuzione χ^2	111		
■ 8.3 Le assunzioni del test χ^2 di bontà di adattamento	112		
■ 8.4 Il test di bontà di adattamento quando vi sono soltanto due categorie	112		
■ 8.5 L'adattamento della distribuzione binomiale	113		
■ 8.6 La casualità nello spazio o nel tempo: la distribuzione di Poisson	115		
• La formula della distribuzione di Poisson	116		
• La verifica della casualità con la distribuzione di Poisson	116		
• Il confronto tra media e varianza	118		
■ 8.7 Riepilogo	118		
■ 8.8 Riepilogo delle formule	118		
Problemi di approfondimento	119		
Ulteriori problemi	121		
■ Scheda 5 Pianificare l'esperimento	124		
<hr/>			
Capitolo 9			
Le tabelle di contingenza per l'analisi dell'associazione tra variabili categoriche			
<hr/>			
■ 9.1 L'associazione tra due variabili categoriche	125		
■ 9.2 La stima dell'associazione in tabelle 2×2: l'odds ratio	126		
• L'odds	126		
• L'odds ratio	127		
• L'errore standard e l'intervallo di confidenza per l'odds ratio	127		
■ 9.3 Il test χ^2 per l'analisi delle tabelle di contingenza	128		
• Le ipotesi	129		
• Le frequenze attese assumendo vera l'ipotesi nulla di indipendenza	129		
• La statistica χ^2	130		
• I gradi di libertà	130		
• P -value e conclusione	130		
• Una formula rapida per il calcolo delle frequenze attese	131		
• Il test χ^2 per l'analisi delle tabelle di contingenza è un caso particolare del test χ^2 di bontà di adattamento	131		
• Le assunzioni del test χ^2 per l'analisi delle tabelle di contingenza	131		
• La correzione di Yates per la continuità	131		
■ 9.4 Il test esatto di Fisher	132		
■ 9.5 Il test G	133		
■ 9.6 Riepilogo	133		
■ 9.7 Riepilogo delle formule	133		
Problemi di approfondimento	135		
Ulteriori problemi	137		
<hr/>			
Parte III			
<hr/>			
Capitolo 10			
La distribuzione normale			
<hr/>			
■ 10.1 Le curve a campana e la distribuzione normale	142		
■ 10.2 La formula della distribuzione normale	144		
■ 10.3 Le proprietà della distribuzione normale	144		
■ 10.4 La distribuzione normale standardizzata e le tavole statistiche	144		
• Come utilizzare le tavole della distribuzione normale standardizzata	145		
• La distribuzione normale standardizzata per descrivere qualsiasi distribuzione normale	146		
■ 10.5 La distribuzione delle medie campionarie è normale	147		
■ 10.6 Il teorema del limite centrale	149		
■ 10.7 Approssimazione normale per la distribuzione binomiale	151		

■ 10.8 Riepilogo	152
■ 10.9 Riepilogo delle formule	152
Problemi di approfondimento	153
Ulteriori problemi	155
■ Scheda 6 I controlli negli studi medici	157

Capitolo 11
L'inferenza in una popolazione
con distribuzione normale

■ 11.1 La distribuzione t per le medie campionarie	159
• La distribuzione t di Student	160
• Come trovare i valori critici della distribuzione t	160
■ 11.2 L'intervallo di confidenza della media di una distribuzione normale	161
• L'intervallo di confidenza al 95% della media	161
• L'intervallo di confidenza al 99% della media	162
■ 11.3 Il test t per un campione	162
• Gli effetti dell'aumento della numerosità campionaria: l'esempio della temperatura corporea rivisitato	165
■ 11.4 Le assunzioni del test t per un campione	165
■ 11.5 La stima della deviazione standard e della varianza in una popolazione con distribuzione normale	165
• I limiti di confidenza della varianza	165
• I limiti di confidenza della deviazione standard	166
• Le assunzioni	166
■ 11.6 Riepilogo	167
■ 11.7 Riepilogo delle formule	167
Problemi di approfondimento	168
Ulteriori problemi	169

Capitolo 12
Il confronto tra due medie

■ 12.1 Dati appaiati o campioni indipendenti?	171
■ 12.2 Confronto appaiato di medie	172
• Stima della differenza media in dati appaiati	172
• Il test t per dati appaiati	174
• Le assunzioni	175
■ 12.3 Confronto tra le medie in due campioni	175
• Intervallo di confidenza della differenza tra due medie	176
• Il test t per due campioni	177
• Le assunzioni	178

• Un test per due campioni quando le deviazioni standard sono diverse	178
■ 12.4 Quali sono le unità campionarie corrette?	178
■ 12.5 L'errore del confronto indiretto	180
■ 12.6 Come interpretare la sovrapposizione tra intervalli di confidenza	181
■ 12.7 Confrontare le varianze	182
• Il test F per l'uguaglianza tra due varianze	182
• Il test di Levene per l'omogeneità delle varianze	182
■ 12.8 Riepilogo	183
■ 12.9 Riepilogo delle formule	183
Problemi di approfondimento	185
Ulteriori problemi	189
■ Scheda 7 Quale test devo usare?	193

Capitolo 13
Cosa fare quando le assunzioni
di una analisi statistica vengono violate

■ 13.1 Identificare le violazioni dalla normalità	196
• I metodi grafici	196
• Test statistici di normalità	197
■ 13.2 Quando è possibile ignorare le violazioni delle assunzioni	198
• La violazione della normalità	198
• Deviazioni standard disuguali	199
■ 13.3 Le trasformazioni dei dati	199
• La trasformazione logaritmica	199
• La trasformazione arcoseno	201
• La trasformazione radice quadrata	201
• Altre trasformazioni	202
• Gli intervalli di confidenza nel caso di trasformazioni	202
• Un avvertimento: evitate le trasformazioni multiple per ricercare risultati statisticamente significativi	202
■ 13.4 Alternative non parametriche al test t per un campione e al test t per dati appaiati	203
• Il test dei segni	203
• Il test dei segni per ranghi di Wilcoxon	205
■ 13.5 Il confronto tra due gruppi: il test U di Mann-Whitney	205
• Ranghi con lo stesso valore (tied ranks)	207
• Campioni grandi e approssimazione normale	207
■ 13.6 Le assunzioni dei test non parametrici	208

■ 13.7 Errori di tipo I e di tipo II nei test non parametrici	208	• L'ANOVA in breve	240
■ 13.8 Riepilogo	208	• Il calcolo delle medie dei quadrati	241
■ 13.9 Riepilogo delle formule	209	• Il rapporto tra varianze, F	242
Problemi di approfondimento	210	• La tabella dell'ANOVA	243
Ulteriori problemi	214	• La variabilità spiegata: R^2	244
		• ANOVA per due gruppi	244
<hr/>			
Capitolo 14			
Il disegno sperimentale			
<hr/>			
■ 14.1 Perché si fanno gli esperimenti?	218	■ 15.2 Assunzioni e alternative	244
• Le variabili di confondimento	218	• La robustezza dell'ANOVA	244
• Gli artefatti creati dall'esperimento	219	• Le trasformazioni dei dati	244
■ 14.2 Gli insegnamenti delle sperimentazioni cliniche	219	• Le alternative non parametriche all'ANOVA	245
• Le componenti del disegno sperimentale	220	■ 15.3 I confronti pianificati	245
■ 14.3 Come ridurre la distorsione	220	• Il confronto pianificato tra due medie	245
• Il gruppo di controllo	221	■ 15.4 I confronti non pianificati	246
• La randomizzazione	221	• Confrontare tutte le coppie di medie con il metodo di Tukey-Kramer	247
• La cecità	222	• Le assunzioni	248
■ 14.4 Come ridurre l'influenza dell'errore di campionamento	223	■ 15.5 Effetti fissi ed effetti casuali	248
• La replicazione	223	■ 15.6 ANOVA per gruppi scelti casualmente	249
• Il bilanciamento	224	• La tabella dell'ANOVA	249
• Il raggruppamento in blocchi	225	• Le componenti della varianza	250
• I trattamenti estremi	226	• La ripetibilità	250
■ 14.5 Esperimenti con più di un fattore	226	• Le assunzioni	251
■ 14.6 Cosa fare quando non si possono condurre esperimenti?	227	■ 15.7 Riepilogo	251
• Appaiamento e aggiustamento	228	■ 15.8 Riepilogo delle Formule	251
■ 14.7 La scelta della dimensione campionaria	228	Problemi di approfondimento	253
• Pianificare la precisione	228	Ulteriori problemi	256
• Pianificare la potenza	230	■ Scheda 9 Errori durante l'esperimento e durante l'analisi statistica	260
• Pianificare per compensare la perdita di dati	230		
■ 14.8 Riepilogo	231	<hr/>	
■ 14.9 Riepilogo delle formule	231	Parte IV	
Problemi di approfondimento	233	Capitolo 16	
Ulteriori problemi	235	La correlazione tra variabili numeriche	
■ Scheda 8 // data dredging,	237	<hr/>	
<hr/>			
Capitolo 15			
Il confronto tra le medie di più di due gruppi			
<hr/>			
■ 15.1 L'analisi della varianza	239	■ 16.1 La stima del coefficiente di correlazione lineare	262
• Le ipotesi	240	• Il coefficiente di correlazione	263
		• L'errore standard	265
		• L'intervallo di confidenza approssimato	265
		■ 16.2 Verificare l'ipotesi nulla di assenza di correlazione	266
		■ 16.3 Le assunzioni	267
		■ 16.4 Il coefficiente di correlazione dipende dall'intervallo di valori considerato	268
		■ 16.5 La correlazione per ranghi di Spearman	268
		• La procedura per n grande	270
		• Le assunzioni della correlazione per ranghi di Spearman	271

■ 16.6 Gli effetti dell'errore di misura sulla correlazione	271
■ 16.7 Riepilogo	271
■ 16.8 Riepilogo delle formule	272
Problemi di approfondimento	274
Ulteriori problemi	276
■ Scheda 10 Il bias di pubblicazione	280

Capitolo 17 La regressione

■ 17.1 La regressione lineare	284
• Il metodo dei minimi quadrati	285
• L'equazione della retta di regressione	285
• Calcolo di pendenza e intercetta della retta di regressione	285
• Popolazioni e campioni	286
• I valori previsti	286
• I residui	287
• L'errore standard della pendenza	287
• L'intervallo di confidenza della pendenza	288
■ 17.2 La qualità delle previsioni	288
• Gli intervalli di confidenza delle previsioni	288
• L'estrapolazione	289
■ 17.3 La verifica delle ipotesi sulla pendenza	289
• Il test t sulla pendenza della regressione lineare	290
• L'approccio ANOVA applicato alla regressione	291
• L'uso di R^2 per misurare l'adattamento della retta di regressione ai dati	291
■ 17.4 La regressione verso la media	292
■ 17.5 Le assunzioni della regressione	293
• Gli outlier	293
• Identificazione della non linearità	294
• Identificazione della non normalità e di varianze disuguali	294
■ 17.6 Le trasformazioni	295
■ 17.7 Gli effetti dell'errore di misura sulla regressione	296
■ 17.8 La regressione non lineare	296
• Una curva con un asintoto	297
• Le curve quadratiche	297
• Adattamento di una curva (curve fitting) senza la formula	298
• Funzioni per una variabile risposta di tipo binario	298
■ 17.8 Riepilogo	299
■ 17.10 Riepilogo delle formule	300

Problemi di approfondimento	302
Ulteriori problemi	306

■ Scheda 11 Le specie diverse come singole osservazioni nelle analisi statistiche	311
--	-----

Parte V

Capitolo 18

Analisi con più di una variabile esplicativa

■ 18.1 Dalla regressione lineare ai modelli lineari generali	314
• Il modello di regressione lineare	315
• La generalizzazione della regressione lineare	315
• L'analisi di una variabile trattamento di tipo categorico	316
■ 18.2 L'analisi di esperimenti con raggruppamento (blocking)	317
• L'analisi dei dati ottenuti con un disegno a blocchi randomizzati	318
• La formula del modello	318
• L'adattamento del modello ai dati	318
■ 18.3 L'analisi di disegni fattoriali	319
• L'analisi di due fattori fissi	319
• La formulazione del modello	320
• L'adattamento del modello ai dati	320
• L'importanza di distinguere fattori fissi e fattori casuali	321
■ 18.4 Come tenere in considerazione gli effetti di una covariata	322
• Omettere il termine corrispondente all'interazione	323
■ 18.5 Le assunzioni dei modelli lineari generali	324
■ 18.6 Riepilogo	325
Problemi di approfondimento	325
Ulteriori problemi	327

Capitolo 19

Metodi computazionali intensivi

■ 19.1 La verifica dell'ipotesi con la simulazione	330
■ 19.2 Il test di randomizzazione	332
• Le assunzioni dei test di randomizzazione	336
■ 19.3 Il bootstrap per calcolare errori standard e intervalli di confidenza	336
• L'errore standard di bootstrap	337
• Gli intervalli di confidenza di bootstrap	338
• Il bootstrap nel caso di campioni multipli	339

• Assunzioni e limitazioni del bootstrap	339
■ 19.3 Riepilogo	339
Problemi di approfondimento	340
Ulteriori problemi	343

Capitolo 20 La verosimiglianza

■ 20.1 Che cos'è la verosimiglianza?	346
■ 20.2 Due applicazioni della verosimiglianza in biologia	347
• La stima di una filogenesi	347
• La mappatura genica	347
■ 20.3 La massima verosimiglianza	348
• Il modello probabilistico	348
• La formula per la verosimiglianza	349
• La stima di massima verosimiglianza	349
• Intervalli di confidenza basati sulla verosimiglianza	351
■ 20.4 La versatilità della stima di massima verosimiglianza	351
• Il modello probabilistico	352
• La formula per la verosimiglianza	352
• La distorsione	353
■ 20.5 Il test del rapporto di verosimiglianza	353
• La statistica per il test del rapporto di verosimiglianza	353
• Testare una proporzione	353
■ 20.6 Riepilogo	354
■ 20.7 Riepilogo delle formule	355
Problemi di approfondimento	355
Ulteriori problemi	358

Capitolo 21 La metanalisi: combinare i risultati ottenuti in più studi

■ 21.1 Che cos'è una metanalisi?	363
• Perché ripetere uno studio?	363
■ 21.2 La potenza della metanalisi	363
■ 21.3 La metanalisi può fornire una visione equilibrata	364
■ 21.4 I passaggi da seguire in una metanalisi	365
• Definire la domanda	365
• Leggere attentamente le pubblicazioni scientifiche	366
• Calcolare la dimensione dell'effetto	367
• Determinare la dimensione media dell'effetto	367
• Calcolare intervalli di confidenza e testare le ipotesi	368
• Studiare l'impatto della qualità degli studi	369
• Cercare associazioni	369
■ 21.5 Il problema dello "studio nel cassetto"	369
■ 21.6 Come rendere accessibile un articolo alle metanalisi	370
■ 21.7 Riepilogo	370
■ 21.8 Riepilogo delle formule	371
Problemi di approfondimento	371
Ulteriori problemi	372