

Indice

1	Introduzione all'utilizzo di MATLAB e alla gestione dei dati	19
1.1	L'interfaccia di MATLAB	19
1.2	L'esecuzione del codice	22
1.3	Classi di dati	24
1.4	Gli array	27
1.4.1	Creazione dei diversi tipi di array	27
1.4.2	Estrazione dei dati da un array	31
1.5	Le tabelle	32
1.5.1	Creazione di una tabella	33
1.5.2	Importazione di una tabella	34
1.5.3	Estrazione dei dati da una tabella	36
1.6	Gestione dei dati	39
1.6.1	Salvataggio dei risultati ottenuti	39
1.6.2	Estrazione dei dati in base a dei criteri	41
1.6.3	Intersezione di due table con le stesse variabili	42
1.6.4	Unione e intersezione di due table	43
1.7	Introduzione alla generazione di numeri casuali	48
1.8	Elementi di base di programmazione	49
1.8.1	Le istruzioni if e i cicli	49
1.8.2	L'istruzione assert	53
1.8.3	Alcune costanti speciali	53
2	Algebra lineare di base	57
2.1	Operazioni elementari con le matrici	57
2.2	Le matrici diagonali	60
2.3	Alcune matrici particolari	63
2.4	Le matrici idempotenti	65
2.5	Le matrici ortogonali	65
2.6	Moltiplicazione di matrici trasposte	66
2.7	Moltiplicazione di matrici inverse	67
2.8	La trasposta dell'inversa	68
2.9	Sistemi di equazioni lineari	69
2.10	La traccia	70
2.11	L'espansione implicita	72
2.12	Matrice degli scostamenti standardizzati tramite espressioni matriciali	76

3	La matrice dei dati e le analisi univariate	81
3.1	Analisi univariate di variabili categoriche	83
3.2	Analisi univariate di variabili quantitative	86
3.3	Grafici univariati per dati categorici o quantitativi	89
3.4	Le distribuzioni di frequenze di variabili quantitative	95
3.5	Analisi univariate di variabili quantitative in presenza di sotto- gruppi	98
3.6	Intervalli di confidenza e riepiloghi avanzati	104
4	Variabili casuali: densità e distribuzioni	109
4.1	La variabile Gaussiana o normale	109
4.2	La variabile aleatoria Uniforme	123
4.3	La variabile aleatoria chi quadrato	125
4.4	La variabile aleatoria T di Student	130
4.5	La distribuzione normale bivariata	134
4.6	La distribuzione uniforme discreta	142
5	I trattamenti preliminari dei dati	145
5.1	I dati mancanti e strategie per il loro trattamento	145
5.2	I valori anomali e le strategie per il loro trattamento	148
5.3	Analisi automatica dei missing e dei valori anomali univariati .	154
5.4	Operazioni avanzate con i missing values	157
6	La relazione tra le variabili quantitative: correlazione e cogra- duazione	161
6.1	La matrice di covarianze e la matrice di correlazione	165
6.2	La significatività della correlazione	170
6.3	L'indice di cograduazione di Spearman	177
6.3.1	La cograduazione in presenza di gradi ripetuti	177
6.4	La relazione tra gli indici di correlazione e cograduazione	182
6.5	La correlazione e la cograduazione in presenza di missing values	183
7	L'associazione	187
7.1	Introduzione	187
7.2	Indici di associazione per le tabelle 2×2	187
7.2.1	Misure basate sulla statistica di Pearson	191
7.2.2	Misure basate sul rapporto dei prodotti incrociati	195
7.3	Indici di associazione per le tabelle $I \times J$	198
7.4	Indici di associazione basati sulla riduzione dell'errore	202
7.5	Indici di associazione basati sulla riduzione dell'eterogeneità . .	205

7.6	Indici di associazione per variabili ordinali	209
8	Le rappresentazioni grafiche	215
8.1	Rappresentazioni grafiche per serie storiche univariate	215
8.2	I grafici a cascata	217
8.3	Rappresentazioni grafiche per serie storiche bivariate	219
8.4	I grafici ad imbuto	221
8.5	Grafici per la stima della densità univariata	221
8.6	Grafici a dispersione personalizzati	227
8.7	Grafici a dispersione con istogrammi o boxplot ai margini	229
8.8	Grafici con istogrammi bivariati	233
8.9	Grafici esplorativi per l'analisi di regressione	234
8.10	Grafici a tre dimensioni	239
8.11	Il balloonplot	242
8.12	La matrice dei diagrammi di dispersione	244
8.12.1	La matrice dei diagrammi di dispersione con variabile di raggruppamento	248
8.13	Le heatmap	251
8.14	I grafici in coordinate parallele	255
8.14.1	I grafici in coordinate parallele con variabile di raggruppamento	257
8.15	Le rappresentazioni iconiche	260
8.15.1	Le stelle	260
8.15.2	Le facce di Chernoff	264
9	Algebra lineare avanzata	267
9.1	La norma di un vettore	267
9.2	Il prodotto scalare	270
9.3	Le forme quadratiche	274
9.3.1	Estrazione degli elementi tramite forme quadratiche	281
9.4	Lo spazio vettoriale	282
9.4.1	Dipendenza, indipendenza lineare e base di uno spazio vettoriale	282
9.5	Il rango	287
9.6	Autovalori e autovettori	289
9.7	Polinomio caratteristico	290
9.8	Routines per il calcolo degli autovalori e degli autovettori	292
9.9	Scomposizione spettrale	292
9.9.1	La scomposizione spettrale attraverso il calcolo simbolico	293

9.10	Introduzione ai poligoni	295
9.11	Proiezioni ortogonali	301
10	Le distanze e gli indici di similarità	307
10.1	Definizione di distanze	307
10.2	Alcuni tipi di distanza	308
10.3	Gli indici di distanza e gli indici di dissimilarità	316
10.4	Lo spazio euclideo ponderato	317
10.5	La distanza di Mahalanobis	319
10.5.1	Proprietà della distanza di Mahalanobis	325
10.6	La scala di misura delle distanze	326
10.7	Gli indici di similarità	327
10.7.1	Indici di similarità per fenomeni dicotomici	328
10.7.2	Indici di similarità in presenza di fenomeni misti	334
11	La riduzione delle dimensioni	339
11.1	Analisi in componenti principali (PC): introduzione	339
11.2	La prima PC come combinazione lineare delle variabili origina- rie	340
11.3	Le prime k PC come combinazioni lineari delle variabili origi- narie	342
11.3.1	Relazione tra autovalori traccia e determinante	343
11.4	La scomposizione in valori singolari (svd)	345
11.5	Le prime k PC come migliore rappresentazione di rango k delle variabili originarie	347
11.6	PC come proiezione ortogonale dei punti in un sottospazio di dimensione ridotta	350
11.6.1	Retta di regressione e retta associata alla prima compo- nente principale	351
11.6.2	Ricostruzione della matrice originaria con una matrice di rango ridotto	356
11.6.3	Componenti principali come rotazione degli assi carte- siani	358
11.7	L'analisi in componenti principali in pratica	366
11.8	Il biplot	377
11.8.1	Prima rappresentazione dei punti riga e dei punti colonna	381
11.8.2	Seconda rappresentazione dei punti riga e dei punti co- lonna	382
11.9	La funzione pcaFS	386
11.10	Componenti principali su \tilde{X} oppure su Z	394

12 L'analisi delle corrispondenze	399
12.1 Notazione	402
12.2 Giudizi sulla bontà dell'analisi e punteggi	414
12.3 Contributi all'inerzia del punto o all'inerzia della dimensione latente	419
12.4 La funzione CorAna	421
13 Cluster analysis	429
13.1 Metodi di cluster gerarchico	431
13.2 Il dendrogramma	436
13.2.1 Concetti alla base della strategia agglomerativa	436
13.2.2 Definizione di distanza tra due gruppi e metodo di rag- gruppamento	438
13.3 Caratteristiche dei metodi gerarchici	440
13.4 Le scelte nell'analisi dei gruppi	441
13.5 Metodi di clustering non gerarchici	444
13.5.1 Il metodo delle k medie	447
13.6 La scelta del numero ottimo di gruppi	452
13.7 Il confronto tra partizioni	454
13.8 Classificazioni basati su modelli di misture finite	459