

# Indice degli argomenti

<b>Prefazione</b>	<b>15</b>
<b>Software e materiale di corredo</b>	<b>23</b>
<b>1 Introduzione all'utilizzo di MATLAB e alla gestione dei dati</b>	<b>27</b>
1.1 L'interfaccia di MATLAB . . . . .	27
1.2 L'esecuzione del codice . . . . .	32
1.3 Introduzione agli array . . . . .	35
1.3.1 Esempi di creazione di array . . . . .	35
1.3.2 Estrazione dei dati da un array . . . . .	36
1.4 Tipologia di dati in MATLAB . . . . .	38
1.5 Le cell . . . . .	41
1.5.1 Esempi di creazione di cell . . . . .	41
1.5.2 Estrazione di dati dalla cell . . . . .	43
1.6 Le struct . . . . .	43
1.6.1 Esempi di creazione di struct . . . . .	43
1.6.2 Estrazione di dati dalla struct . . . . .	44
1.7 Le table . . . . .	44
1.7.1 Esempi di creazione di table . . . . .	45
1.7.2 Importazione di una table da file esterni . . . . .	46
1.7.3 Importazione di una table tramite task . . . . .	49
1.7.4 Estrazione dei dati da una table . . . . .	51
1.7.5 Estrazione dei dati da una table in base a criteri . .	54
1.7.6 La somma di due table e informazioni aggiuntive pre- senti nelle table . . . . .	57
1.8 Salvataggio dei risultati . . . . .	60
1.9 Introduzione alla generazione di numeri casuali . . . . .	62
1.10 Elementi di base di programmazione . . . . .	63
1.10.1 Le istruzioni <code>if</code> e i cicli . . . . .	63
1.10.2 L'istruzione <code>assert</code> . . . . .	67

1.10.3	Alcune costanti e funzioni speciali . . . . .	68
1.11	Il dictionary . . . . .	69
1.11.1	Esempi di creazione di un dictionary . . . . .	70
1.11.2	Estrazione di dati dal dictionary . . . . .	70
1.12	Il sistema di help di MATLAB . . . . .	74
	Esercizi di riepilogo . . . . .	76
<b>2</b>	<b>Algebra lineare</b>	<b>79</b>
2.1	Operazioni elementari con le matrici . . . . .	79
2.2	Le matrici diagonali . . . . .	82
2.3	Alcune matrici particolari . . . . .	85
2.4	Ripasso di algebra lineare . . . . .	86
2.5	Sistemi di equazioni lineari . . . . .	88
2.6	Le operazioni matriciali elemento per elemento . . . . .	89
2.7	L'espansione implicita . . . . .	91
2.7.1	L'espansione implicita nelle table . . . . .	95
2.8	La norma di un vettore . . . . .	98
2.9	Il prodotto scalare . . . . .	100
2.10	Le forme quadratiche . . . . .	101
2.10.1	Estrazione degli elementi tramite forme quadratiche . . . . .	103
2.11	Spazi vettoriali . . . . .	103
2.11.1	Dipendenza, indipendenza lineare e base di uno spazio vettoriale . . . . .	104
2.12	Il rango . . . . .	107
2.13	Autovalori e autovettori . . . . .	108
2.14	Polinomio caratteristico . . . . .	109
2.15	Routine per il calcolo degli autovalori e degli autovettori . . . . .	111
2.16	Scomposizione spettrale . . . . .	112
2.16.1	La scomposizione spettrale attraverso il calcolo simbolico . . . . .	113
2.17	Proiezioni ortogonali . . . . .	115
	Esercizi di riepilogo . . . . .	120
<b>3</b>	<b>Analisi esplorative dei dati e tabelle pivot</b>	<b>129</b>
3.1	Creazione e gestione dei categorical arrays . . . . .	130
3.2	Analisi univariate di variabili categoriche . . . . .	132
3.3	Analisi univariate di variabili quantitative . . . . .	133
3.4	Grafici univariati per dati categorici o quantitativi . . . . .	137
3.5	Le distribuzioni di frequenze di variabili quantitative . . . . .	143

3.6	Analisi univariate di variabili quantitative in presenza di sottogruppi . . . . .	146
3.7	Intervalli di confidenza e riepiloghi avanzati: funzione <code>grpstats</code> . . . . .	150
3.7.1	La funzione <code>grpstatsFS</code> . . . . .	154
3.8	Tabelle pivot tramite la creazione guidata . . . . .	156
3.9	Tabelle pivot avanzate . . . . .	158
3.10	Boxplot in base a più variabili di raggruppamento . . . . .	165
3.11	Grafici a violino in base a più variabili di raggruppamento . . . . .	167
	Esercizi di riepilogo . . . . .	168
<b>4</b>	<b>Importazione dei dati dal mondo web (in tempo reale)</b>	<b>175</b>
4.1	Importazione dei dati da una pagina web . . . . .	175
4.2	Importazione dei dati da una pagina di GitHub . . . . .	176
4.2.1	Gestione dei file scaricati tramite GitHub . . . . .	179
4.3	Importazione di dataset di grandi dimensioni . . . . .	182
4.4	Introduzione alle timetable . . . . .	184
4.4.1	Estrarre record dalle timetable . . . . .	184
4.4.2	Creare sequenze di date . . . . .	188
4.5	Importazione serie storiche ISTAT . . . . .	189
4.6	Importazione personalizzata . . . . .	192
4.6.1	Duration e <code>calendarDuration</code> . . . . .	194
4.6.2	Cambiamento della periodicità della serie storica . . . . .	195
4.7	Importazione in tempo reale di serie storiche finanziarie o dai social media . . . . .	196
	Esercizi di riepilogo . . . . .	198
<b>5</b>	<b>Variabili casuali: densità, distribuzioni e stime dei parametri</b>	<b>203</b>
5.1	La variabile Gaussiana o normale . . . . .	203
5.2	La variabile aleatoria Uniforme . . . . .	213
5.3	La variabile aleatoria chi quadrato . . . . .	219
5.4	La variabile aleatoria $T$ di Student . . . . .	222
5.5	Stima dei parametri delle distribuzioni . . . . .	225
5.6	La distribuzione normale bivariata . . . . .	226
5.7	La distribuzione normale multivariata . . . . .	232
5.8	La distribuzione uniforme discreta . . . . .	235
	Esercizi di riepilogo . . . . .	236
<b>6</b>	<b>I trattamenti preliminari dei dati</b>	<b>245</b>
6.1	I dati mancanti e strategie per il loro trattamento . . . . .	245

6.2	I valori anomali e le strategie per il loro trattamento . . . . .	249
6.3	Analisi automatica dei missing e dei valori anomali univariati	252
6.4	Operazioni avanzate con i missing values . . . . .	254
6.5	Verifica dell'ipotesi di normalità . . . . .	257
	Esercizi di riepilogo . . . . .	259
<b>7</b>	<b>La relazione tra le variabili quantitative: correlazione e cograduazione</b>	<b>261</b>
7.1	La matrice di covarianze e la matrice di correlazione . . . . .	264
7.2	La distribuzione del coefficiente di correlazione quando $\rho = 0$	268
7.3	La distribuzione del coefficiente di correlazione quando $\rho \neq 0$	275
7.4	L'indice di cograduazione di Spearman . . . . .	280
7.4.1	La cograduazione in presenza di gradi ripetuti . . . . .	280
7.5	La relazione tra gli indici di correlazione e cograduazione . . . . .	284
7.6	La correlazione e la cograduazione in presenza di missing values	285
	Esercizi di riepilogo . . . . .	287
<b>8</b>	<b>Le rappresentazioni grafiche</b>	<b>293</b>
8.1	Rappresentazioni grafiche per serie univariate . . . . .	293
8.2	I grafici a cascata . . . . .	294
8.3	Rappresentazioni grafiche per serie bivariate . . . . .	296
8.4	I grafici ad imbuto . . . . .	298
8.5	Grafici per la stima della densità univariata . . . . .	299
8.6	Grafici a dispersione personalizzati . . . . .	302
8.7	Grafici a dispersione con istogrammi o boxplot ai margini . . . . .	304
8.8	Grafici con istogrammi bivariati . . . . .	307
8.9	Grafici esplorativi per l'analisi di regressione . . . . .	309
8.10	Grafici a tre dimensioni . . . . .	311
8.11	Il balloonplot . . . . .	315
8.12	I diagrammi a bolle (bubblecharts) . . . . .	316
8.12.1	I grafici a mappa: bubblechart geografico . . . . .	319
8.13	I geoplots . . . . .	322
8.13.1	L'app <code>geoplotsFS</code> . . . . .	326
8.13.2	I geoiconcharts . . . . .	326
8.14	I grafici a sciame (swarmchart) . . . . .	328
8.15	La matrice dei diagrammi a dispersione . . . . .	330
8.15.1	Matrice dei diagrammi di dispersione con contorni bivariati e rette di regressione . . . . .	331
8.15.2	Matrice dei diagrammi di dispersione con numeri, visualizzazione simbolica e colore di sfondo . . . . .	333

8.15.3	La matrice dei diagrammi a dispersione con variabile di raggruppamento . . . . .	337
8.16	Le heatmap . . . . .	339
8.17	I grafici polari (polar plots) . . . . .	343
8.17.1	Il compassplot . . . . .	343
8.18	Polar histogram . . . . .	346
8.18.1	La funzione <code>polarhistogramFS</code> . . . . .	348
8.19	I grafici in coordinate parallele . . . . .	351
8.19.1	I grafici in coordinate parallele con variabile di raggruppamento . . . . .	352
8.20	Le rappresentazioni iconiche . . . . .	355
8.20.1	Le stelle . . . . .	355
8.20.2	Le facce di Chernoff . . . . .	357
	Esercizi di riepilogo . . . . .	357
<b>9</b>	<b>L'associazione</b>	<b>365</b>
9.1	Introduzione . . . . .	365
9.2	Indici di associazione per le tabelle $2 \times 2$ . . . . .	365
9.2.1	Misure basate sulla statistica di Pearson . . . . .	369
9.2.2	Misure basate sul rapporto dei prodotti incrociati . . . . .	372
9.3	Indici di associazione per le tabelle $I \times J$ . . . . .	374
9.4	Indici di associazione basati sulla riduzione dell'errore . . . . .	379
9.5	Indici di associazione basati sulla riduzione dell'eterogeneità . . . . .	382
9.6	La funzione <code>corrNominal</code> . . . . .	383
9.7	Indici di associazione per variabili ordinali . . . . .	386
9.8	La funzione <code>corrOrdinal</code> . . . . .	390
	Esercizi di riepilogo . . . . .	392
<b>10</b>	<b>Le distanze e gli indici di similarità</b>	<b>397</b>
10.1	Definizione di distanze . . . . .	397
10.2	Alcuni tipi di distanza . . . . .	398
10.3	Gli indici di distanza e gli indici di dissimilarità . . . . .	405
10.4	Lo spazio euclideo ponderato . . . . .	406
10.5	Poligoni e distanze . . . . .	408
10.6	La distanza di Mahalanobis . . . . .	414
10.6.1	Proprietà della distanza di Mahalanobis . . . . .	419
10.7	La scala di misura delle distanze . . . . .	421
10.8	Gli indici di similarità . . . . .	422
10.8.1	Indici di similarità per fenomeni dicotomici . . . . .	422
10.8.2	Indici di similarità in presenza di fenomeni misti . . . . .	428

Esercizi di riepilogo . . . . .	430
<b>11 La riduzione delle dimensioni</b>	<b>433</b>
11.1 Analisi in componenti principali (PC): introduzione . . . . .	433
11.2 La prima PC come combinazione lineare delle variabili originarie	434
11.3 Le prime $k$ PC come combinazioni lineari delle variabili originarie . . . . .	436
11.3.1 Relazione tra autovalori traccia e determinante . . . . .	437
11.4 La scomposizione in valori singolari (svd) . . . . .	439
11.5 Le prime $k$ PC come migliore rappresentazione di rango $k$ .	441
11.6 PC come proiezione ortogonale dei punti in un sottospazio .	443
11.6.1 Retta di regressione e retta associata alla prima componente principale . . . . .	444
11.6.2 Ricostruzione della matrice originaria con una matrice di rango ridotto . . . . .	448
11.6.3 Retta principale e retta di regressione . . . . .	450
11.6.4 Componenti principali come rotazione degli assi cartesiani . . . . .	451
11.7 L'analisi in componenti principali in pratica . . . . .	460
11.8 Il biplot . . . . .	471
11.8.1 Prima rappresentazione dei punti riga e dei punti colonna . . . . .	473
11.8.2 Seconda rappresentazione dei punti riga e dei punti colonna . . . . .	475
11.9 Qualità della proiezione per ogni punto . . . . .	479
11.10 La funzione <code>pcaFS</code> . . . . .	484
11.11 Componenti principali su $\tilde{X}$ oppure su $Z$ . . . . .	492
11.12 Componenti principali con dati geografici . . . . .	492
Esercizi di riepilogo . . . . .	497
<b>12 L'analisi delle corrispondenze</b>	<b>505</b>
12.1 Notazione . . . . .	508
12.2 Giudizi sulla bontà dell'analisi e punteggi . . . . .	519
12.3 Contributi all'inerzia del punto o all'inerzia della dimensione latente . . . . .	523
12.4 La funzione <code>CorAna</code> . . . . .	525
12.5 Strumenti avanzati di analisi delle corrispondenze . . . . .	529
12.5.1 Il moonplot . . . . .	529
12.5.2 Personalizzazione dei grafici di analisi delle corrispondenze . . . . .	531

12.6	L'app CorAnaAPP e i punti supplementari . . . . .	534
	Esercizi di riepilogo . . . . .	534
<b>13</b>	<b>Cluster analysis</b>	<b>539</b>
13.1	Metodi di cluster gerarchico . . . . .	542
13.2	Il dendrogramma . . . . .	543
13.2.1	Concetti alla base della strategia agglomerativa . . .	543
13.2.2	Definizione di distanza tra due gruppi e metodo di raggruppamento . . . . .	545
13.3	Caratteristiche dei metodi gerarchici . . . . .	548
13.4	Le scelte nell'analisi dei gruppi . . . . .	548
13.5	Metodi di clustering non gerarchici . . . . .	551
13.5.1	Il metodo delle $k$ -medie . . . . .	554
13.6	La scelta del numero ottimo di gruppi . . . . .	558
13.7	Il confronto tra partizioni . . . . .	560
13.8	Classificazioni basate su modelli di misture finite . . . . .	564
	Esercizi di riepilogo . . . . .	570
<b>14</b>	<b>Analisi delle serie storiche</b>	<b>577</b>
14.1	Introduzione . . . . .	577
14.2	Le trasformazioni della serie e le loro caratteristiche descrittive	578
14.2.1	Confronto nel tempo e misura delle variazioni . . . .	579
14.2.2	Aggregazione contemporanea e comparazione di ag- gregati nel tempo . . . . .	583
14.2.3	Dipendenza seriale: autocovarianza e autocorrelazione	586
14.2.4	Identificazione dei cicli: il periodogramma . . . . .	591
14.3	Modelli e metodi per serie temporali . . . . .	593
14.4	Il modello classico (deterministico) di scomposizione di una serie temporale . . . . .	595
14.4.1	Trend . . . . .	596
14.4.2	Ciclo . . . . .	596
14.4.3	Stagionalità . . . . .	596
14.4.4	Irregolarità del calendario, valori anomali e cambia- menti strutturali . . . . .	604
14.4.5	Stima e previsione . . . . .	606
14.5	Regressione polinomiale locale e medie mobili . . . . .	613
14.5.1	Polinomi locali e stima kernel . . . . .	614
14.5.2	Il trattamento delle estremità delle serie. La stima in tempo reale . . . . .	619
14.5.3	Inferenza . . . . .	621

14.5.4	Stima di $h$ per convalida incrociata . . . . .	623
14.5.5	Stima di $\sigma^2$ e stima intervallare di $\mu_t$ . . . . .	624
14.5.6	Le medie mobili . . . . .	624
14.5.7	Dipendenza seriale spuria, effetto ampiezza e fase . . . . .	626
14.6	Trend locali e livellamento esponenziale . . . . .	629
14.6.1	Previsione mediante livellamento esponenziale . . . . .	629
14.6.2	Il livellamento esponenziale applicato alla misura del rischio di mercato . . . . .	632
14.6.3	Previsione mediante il metodo di Holt-Winters . . . . .	633
14.6.4	Metodo di Holt-Winters stagionale . . . . .	635
14.7	La modellazione del breve periodo . . . . .	636
14.7.1	Stazionarietà . . . . .	637
14.7.2	Previsione lineare ottimale . . . . .	639
14.7.3	Funzione di autocorrelazione parziale . . . . .	640
14.7.4	Stima della media e della funzione di autocovarianza . . . . .	641
14.7.5	La verifica dell'ipotesi di incorrelazione . . . . .	641
14.7.6	Il teorema di Wold . . . . .	642
14.7.7	Processi Autoregressivi . . . . .	643
14.7.8	Processi Media Mobile . . . . .	648
14.7.9	Processi misti . . . . .	650
14.7.10	Inferenza . . . . .	653
14.7.11	Verifica del modello stimato . . . . .	656
14.7.12	Nonstazionarietà: trend deterministici o stocastici? . . . . .	656
14.7.13	Modelli ARIMA . . . . .	658
14.7.14	Modelli ARIMA stagionali . . . . .	659
14.7.15	Modelli ARIMA con variabili esogene e modelli regA- RIMA . . . . .	661
14.7.16	Previsione . . . . .	663
	Appendice: Stima del modello di regressione multipla . . . . .	672
	Esercizi di riepilogo . . . . .	677

## Bibliografia

681