

目次

第I部 データ分析

第1章 データ	3
1.1 変数と変量とデータ	5
1.2 関係形式	8
1.2.1 湿度吸収実験データ	10
1.2.2 複式簿記データ	13
1.3 データの代表値	16
1.3.1 最小値, 最大値	18
1.3.2 平均値	18
1.3.3 標準偏差	21
1.4 偏差値	24
第2章 データ分布	27
2.1 1次元散布図	27
2.2 度数分布表とヒストグラム	29
2.3 度数分布多角形	32
2.4 ひとつこと	33
第3章 データ分布の代表値	35
3.1 データ分布の中心	35
3.1.1 平均値	35
3.1.2 中央値	37
3.2 データ分布の広がり	41
3.2.1 標準偏差	41
3.2.2 平均絶対偏差	42
3.2.3 四分位数	43
3.3 データ分布の要約値	44
3.4 データ例	45
3.4.1 春の訪れ	45

3.4.2	音楽アルバム	48
第4章	箱ひげ図	51
4.1	ネットワークの応答速度	51
4.2	箱ひげ図とロウソク足	54
4.3	ひとこと	59
第5章	2変量データ	61
5.1	変量の相関関係	61
5.2	2元度数分布表と2次元散布図	63
5.3	2変量同時分布の代表値	64
5.3.1	共分散と相関係数	64
第II部	データサイエンス	
第6章	データサイエンス入門	69
6.1	データに語る	70
6.2	ストラテジー	70
6.3	モデル	72
6.4	データサイエンス	75
6.4.1	データの上流から下流まで	79
6.4.2	データリテラシー	89
6.5	データサイエンティスト	95
6.6	データファイル	97
6.6.1	フラットファイル	98
6.6.2	マークアップファイル	99
6.6.3	バイナリーファイル	100
6.7	関係形式データベース	101
6.7.1	テーブルに対する演算	101
6.7.2	SQL	102
6.7.3	キー	103
6.7.4	ドメイン	107
6.8	関係形式データベースを超えて	108
6.8.1	2次データ	108
6.8.2	テーブル間の関係	112
6.8.3	さまざまなレベルでの属性	114
6.8.4	DandD	115

6.9	ソフトウェア	119
6.9.1	表計算ソフトウェア	120
6.9.2	統計解析ソフトウェア	120
6.9.3	汎用なデータ解析環境 S と R	121
6.10	データ行列と線形代数	122
6.10.1	データ行列	123
6.10.2	個体空間と変量空間	123
6.10.3	中心化	124
6.10.4	尺度規準化	125
第7章	個体の雲の探索	129
7.1	クラスタリング	129
7.1.1	ウイルス RNA 変異データ	132
7.2	主成分分析	142
7.2.1	新しい座標軸の求め方	143
7.2.2	都道府県之力	148
7.2.3	特異値分解	153
7.3	高次元個体空間の可視化 - TextilePlot	156
第8章	変量間の関係	167
8.1	回帰モデル	167
8.1.1	放射性物質拡散データ	173
8.1.2	湿度吸収実験データ	179
8.1.3	脊柱後弯症データ	183
8.1.4	真鯛放流捕獲データ	186
8.2	非線形回帰モデル	187
8.3	正準相関分析とコレスポンデンス分析	188
8.3.1	正準相関分析	188
8.3.2	コレスポンデンス分析	193
第9章	変量間の相関	199
9.1	相関係数と偏相関係数	199
9.2	相関と独立性	205
9.3	偏相関と条件付き独立性	210

第 10 章 確率モデル	215
10.1 地震データ	215
10.1.1 マグニチュードの分布	216
10.1.2 地震の発生間隔の分布	229
10.2 真鯛放流捕獲データ	232
10.3 株価収益率データ	236
10.4 心拍データ	239
10.5 ひとつ、ふたこと	241
10.5.1 連続収益率と複利	241
10.5.2 正規分布	242
10.5.3 ブラウン運動	243
10.5.4 地震の予測	244
10.5.5 多変量正規分布	246
10.5.6 ガンマ分布	247
参考文献	249
索引	252