

目 次

第 1 章	定性的データの操作と視覚化	1
1.1	定性的データとは	1
1.2	定性的データの形式	1
1.2.1	定性的データの尺度	2
1.2.2	データの記述形式	2
1.3	カテゴリカルデータの集計	3
1.3.1	データの読み込み	3
1.3.2	二元分割表の集計	6
1.3.3	三元分割表の集計	9
1.3.4	相対頻度・割合	10
1.4	分割表の視覚化	11
1.4.1	気球グラフ	12
1.4.2	帯グラフと棒グラフ	12
1.4.3	円グラフ	14
1.4.4	モザイクグラフ	15
1.4.5	網グラフ	16
第 2 章	比率の推測	18
2.1	母集団と標本	18
2.2	推測とは	18
2.3	確率分布	19
2.3.1	二項分布	20

vi 目次

2.3.2	正規分布	23
2.3.3	二項分布の近似	24
2.3.4	t 分布	25
2.4	比率の区間推定	26
2.5	尤度と最尤推定	30
2.5.1	尤度	30
2.5.2	尤度関数と最尤推定法	30
2.5.3	推定量の性質	31
第3章	仮説検定と比率の検定	33
3.1	仮説検定	33
3.1.1	仮説と対立仮説	33
3.1.2	誤りと有意水準	34
3.1.3	検定統計量と P 値	35
3.1.4	検定のプロセス	36
3.2	比率の検定	37
3.3	比率の検定と 2×2 分割表のカイ 2 乗検定	41
3.4	オッズ比の検定	43
3.4.1	オッズとオッズ比	43
3.4.2	オッズ比の区間推定	44
3.4.3	オッズ比の検定	46
3.5	フィッシャーの正確確率検定	47
第4章	$I \times J$ 二元分割表	49
4.1	適合度検定	49
4.2	独立性の検定 (順序なしの場合)	51
4.2.1	カイ 2 乗検定統計量	52
4.2.2	分割表の視覚化 (モザイクグラフ)	56
4.3	連関係数	56
4.4	フィッシャーの正確確率検定	57
4.5	順序尺度の独立性検定	58
4.5.1	片方の変数が順序尺度の場合	59
4.5.2	両変数が共に順序尺度の場合	60
4.5.3	ウィルコクソン・スコア	62

第5章	効果量と検出力	66
5.1	効果量とは	66
5.2	平均差の検定の効果量	67
5.2.1	効果量 d	67
5.2.2	効果量 Hedge's g	67
5.2.3	t 検定統計量の効果量 r	67
5.3	比率差の検定の効果量	68
5.3.1	効果量 Cohen's d	68
5.3.2	効果量 Hedge's g	68
5.4	独立検定の効果量	69
5.5	効果量を求める関数	69
5.5.1	比率検定の効果量を求める関数 <code>prepes</code>	69
5.5.2	分割表検定の効果量を求める関数 <code>chies</code>	71
5.6	検出力とは	73
5.6.1	比率検定の検出力	74
5.6.2	カイ 2 乗検定の検出力	75
5.6.3	必要となる標本サイズの計算	76
第6章	三元分割表	79
6.1	三元分割表	79
6.2	三元表の表記と操作	81
6.2.1	三元表の表記	81
6.2.2	三元表の作成	82
6.3	層別分析とシンプソンのパラドックス	87
6.4	三元分割表の独立性検定	90
6.4.1	独立性の検定	90
第7章	分割表の対数線形モデル	96
7.1	分割表のモデリング	96
7.2	関数 <code>loglm</code> による対数線形モデルの計算	99
7.3	モデルの選択	101
7.4	三元分割表の対数線形モデル	103
7.4.1	モデルの表記	103
7.4.2	モデル間の差異の統計量	106

7.4.3	モデルの選択と変数の選択	107
7.5	モデルのグラフ表現	109
7.6	多元分割表の対数線形モデル	112
第 8 章	回帰分析と数量化 I 類	116
8.1	回帰分析とは	116
8.2	回帰分析	117
8.2.1	回帰係数	118
8.2.2	変数選択	121
8.2.3	モデルの評価指標	123
8.2.4	多重共線性	123
8.2.5	残差分析	124
8.2.6	交互作用項目ありのモデル	125
8.3	カテゴリカルデータの回帰分析	127
8.4	数量化 I 類	130
第 9 章	ロジットモデル	138
9.1	モデリングと確率分布	138
9.1.1	二項分布とロジットモデル	138
9.2	関数 glm によるモデルの推測	140
9.2.1	関数 glm の書式	140
9.2.2	集計データのモデリング	140
9.2.3	ロジットモデルとオッズ比	145
9.3	素データのモデリング	146
9.4	多項ロジットモデル	150
9.5	交差確認	154
9.6	ロジットモデルとその他	155
第 10 章	ポアソン回帰	157
10.1	ポアソン分布	157
10.1.1	ポアソン分布と関連関数	157
10.1.2	ポアソン分布のパラメータの最尤推定値	158
10.2	ポアソン分布によるモデリング	159
10.3	ポアソン回帰モデル	163

10.3.1	回帰モデルの計算	164
10.3.2	比率データのポアソン回帰 (オフセットによる調整)	168
10.4	過分散の問題と負の二項回帰モデル	170
10.4.1	過分散の問題	170
10.4.2	負の二項回帰モデル	172
10.5	ゼロ過剰の回帰モデル	176
10.5.1	データ bioChemists のポアソン回帰モデル	176
10.5.2	データ bioChemists の Hurdle 回帰モデル	178
10.5.3	データ bioChemists の Zero-inflated 回帰モデル	180
10.5.4	データ bioChemists の 4 つのモデルの比較	182
第 11 章	一般化線形モデル	184
11.1	指数分布族	184
11.1.1	指数分布族と関連のパラメータ	184
11.1.2	指数分布族の最大対数尤度	186
11.2	一般化線形モデルとリンク関数	187
11.2.1	一般化線形モデル	187
11.2.2	関数 glm と主なりリンク関数	188
11.3	最大対数尤度推定と加重最小 2 乗解	189
11.3.1	最大対数尤度の推定	190
11.3.2	加重最小 2 乗法	193
11.3.3	反復加重最小 2 乗法	194
11.4	一般化線形モデルの残差	197
11.4.1	残差	197
11.4.2	逸脱度と関連の検定	198
第 12 章	一般化線形混合効果モデル	203
12.1	線形混合効果モデル	203
12.1.1	線形混合効果モデルとは	203
12.1.2	線形混合効果モデル定式	204
12.1.3	分散成分の推定	205
12.1.4	線形混合効果モデルの関数 lmer と例	206
12.2	一般化線形混合効果モデル	213
12.2.1	一般化線形混合効果モデルと関数 glm	213

第 13 章	ツリーモデル	217
13.1	ツリーモデルとは	217
13.2	分割基準	220
13.3	R 上でのツリーモデルの操作	222
13.3.1	関数 <code>rpart</code> による木の生成	223
13.3.2	木の図示	225
13.3.3	木の生長のコントロールと剪定	230
13.3.4	ツリーモデルによる予測	234
13.4	回帰木	236
13.5	その他	239
第 14 章	アンサンブル学習モデル	242
14.1	アンサンブル学習とは	242
14.2	バギング	243
14.2.1	ブートストラップ	244
14.2.2	バギングのアルゴリズム	244
14.2.3	R 関数と事例分析	246
14.3	ランダムフォレスト	253
14.3.1	ランダムフォレストのアルゴリズム	253
14.3.2	R 関数と事例分析	254
14.4	ブースティング	258
14.4.1	ブースティングアルゴリズム	259
14.4.2	R 関数と事例分析	260
14.5	その他	263
第 15 章	対応分析	264
15.1	対応分析とは	264
15.2	対応分析のアルゴリズム	265
15.3	対応分析関数とデータ解析	266
15.3.1	関数 <code>corresp</code> による対応分析	266
15.3.2	関数 <code>CA</code> を用いた場合	269
15.4	多重対応分析	272
15.4.1	ダミー変数について	273
15.4.2	カテゴリラベルで記録したデータ	275

15.5	多元分割表の対応分析	277
15.5.1	多元分割表の度数を重みとして扱う	277
15.5.2	多元分割表を直接対応分析	279
15.6	その他	281
第 16 章	類似度によるデータ分析	283
16.1	多変量データ解析と類似度	283
16.2	定性的データの関連係数	284
16.2.1	名義尺度	285
16.3	主成分分析	295
16.3.1	主成分分析とは	295
16.3.2	関数 <code>principal</code> による主成分分析	297
16.3.3	平行分析	301
16.3.4	主成分負荷量の回転	302
16.4	因子分析	305
16.4.1	因子分析のモデル	305
第 17 章	距離データの解析	310
17.1	距離とは	310
17.2	階層的クラスター分析	313
17.2.1	階層的クラスター分析のプロセス	314
17.2.2	階層的クラスター分析の方法	316
17.2.3	階層的クラスター分析の諸方法	317
17.2.4	階層的クラスターリングの関数	317
17.2.5	樹形図の切断	321
17.2.6	デンドログラムの属性で樹形図の操作	321
17.2.7	クラスターのヒートマップ	324
17.2.8	距離行列とコーフェン行列の相関	325
17.3	非階層的クラスター分析	326
17.3.1	k 平均法	326
17.3.2	関数 <code>kmeans</code> と例	327
17.4	その他	328

第 18 章 ネットワーク分析	330
18.1 ネットワーク分析とは	330
18.2 ネットワークの作成	331
18.2.1 隣接行列のネットワークグラフ	331
18.2.2 データフレーム形式のネットワークデータ	333
18.2.3 レイアウト	334
18.2.4 ノードとエッジ装飾	335
18.3 ネットワークの統計量	337
18.3.1 基本要素の数と次数	337
18.3.2 密度	337
18.3.3 中心性	339
18.3.4 クラスター係数	339
18.3.5 次数の相関係数	340
18.3.6 パスの長さ	341
18.3.7 グラフの比較	342
18.4 コミュニティ分析	344
18.4.1 コミュニティの分割	344
18.4.2 コミュニティとオーバーラッピング	348
18.5 テキスト分析におけるネットワーク分析	353
18.5.1 テキストと語句間の関連性	353
18.5.2 語句間の関連性	354
18.6 その他	358
第 19 章 アソシエーション分析	360
19.1 アソシエーション分析	360
19.2 相関ルール	361
19.2.1 相関ルールとは	361
19.2.2 相関ルールの評価指標	361
19.2.3 相関ルール抽出のアルゴリズム Apriori	362
19.2.4 データの操作とルールの抽出	363
19.3 頻出アイテムの抽出	373
19.3.1 頻出アイテム抽出アルゴリズム Eclat	373
19.3.2 関数 eclat と例	374

19.4 抽出結果の補助分析	376
19.4.1 関数 dissimilarity	376
19.4.2 ルールのクラスター分析	376
19.4.3 その他	378
索 引	379