

土木・交通計画のための 多変量解析（改訂版）

博士（工学） 川崎 智也
博士（工学） 稲垣 具志
博士（工学） 寺内 義典 共著
博士（工学） 石坂 哲宏
博士（工学） 兵頭 知

コロナ社

まえがき

本書は、『土木・交通工学のための統計学 基礎と演習』の続刊であり、大学や高等専門学校で土木・交通計画を専攻する学生を対象とした多変量解析手法の入門書として2017年に初版第1刷が発刊されました。今回の改訂版では、交通工学の専門家である兵頭 知氏を共著者に迎え、因子分析(8章)の基本概念やモデル式の説明に重点を置いて再整理し、分析結果解釈とExcel演習のための説明を充実しました。本書の演習問題では、Excelを用いた解法を紹介しています。現在さまざまなソフトウェアが入手可能で、これまでよりも容易にモデル分析が可能となっていますが、多変量解析手法の中身を理解するために、初版に引き続きExcelを用いた演習問題としています。

土木・交通計画が解き明かそうとする対象のデータは、無限大に近い主体の意思や行動の集積であり、出現する現象はさまざまな要素が相互に影響した結果でもあります。たとえば交通事故の解析であれば、交通量などの交通条件、道路線形などの構造条件、天候などの自然条件、そしてドライバーなど交通主体の個人属性など、異なるさまざまな状況が事故発生の要素となります。このような交通現象に対して、多くの要素をまとめて一つの数理モデルに収め、客観的かつ論理的に解き明かすことができる手法が多変量解析手法であり、交通工学分野の研究・実務にとって非常に強力な統計手法です。

今回の改訂版においても、理論や方法の解説に加えて、交通分野における現実の課題に近い例題を多く用意しましたので、学生が理解しやすいだけでなく、そのまま交通工学の実務者や研究者としてのキャリアを始める頃まで、おつきあいいただける内容になっています。

なお、章末の演習問題の略解は、Webページからダウンロードすることができますが、ぜひ解答を見る前にじっくり考察することを勧めます(p.38参照)。

最後に、コロナ社の皆様には、長きにわたりご支援をいただきました。ここに厚く御礼を申し上げます。

2024年1月

著者一同

目 次

1. 土木・交通計画における多変量解析

1.1	土木・交通計画分野における多変量解析の意義	1
1.2	多変量解析手法の選択	3
1.2.1	さまざまな多変量解析手法	3
1.2.2	予 測	3
1.2.3	影 響 力 分 析	6
1.2.4	要 約	7
1.2.5	構 造 分 析	9
1.2.6	手法選択のまとめ	11
1.3	手法の適用における留意点	12
1.3.1	多変量解析の手順	12
1.3.2	多変量解析の実際	13
	コラム：多変量解析で用いるソフトウェア	15

2. 記 述 統 計

2.1	多変量解析における記述統計の役割	17
2.2	データの 種類	18
2.2.1	質 的 デ ー タ	19
2.2.2	量 的 デ ー タ	20
2.3	度数分布表とヒストグラム	22
2.3.1	度数分布表とヒストグラムの作成	22
2.3.2	分布形状の特徴	25
2.3.3	Excel によるヒストグラムの作成	26
2.4	代 表 値	30
	演 習 問 題	38
	コラム：複数選択のアンケート設問における集計	39

3. 2変数の分析（相関分析・分散分析）

3.1	相 関 分 析	41
3.1.1	散 布 図	41

3.1.2	相 関 係 数	41
3.1.3	相関分析の実際	43
3.1.4	Excelによる相関分析	45
3.2	分 散 分 析	49
3.2.1	分散分析の基本的な考え方	49
3.2.2	分散分析の種類	50
3.2.3	データの変動と分離	52
3.2.4	仮 説 検 定	53
3.2.5	Excelによる分散分析の実際 (一元配置)	55
3.2.6	Excelによる分散分析の実際 (二元配置)	56
3.2.7	多 重 比 較	59
3.3	ピボットテーブルを用いたクロス集計表の作成	61
3.3.1	クロス集計表の統計手法	61
3.3.2	例 題	61
	演 習 問 題	66
	コラム：相関係数の目安	69

4. 回 帰 分 析

4.1	基本的な概念と位置づけ	70
4.2	回帰モデル式と最小二乗法	72
4.2.1	回帰モデル式	72
4.2.2	最小二乗法によるモデルパラメータの導出	73
4.2.3	Excelによる最小二乗法の適用	77
4.3	決 定 係 数	80
4.4	回帰係数と決定係数の検定	83
4.4.1	回帰係数の検定	84
4.4.2	決定係数の検定	85
4.5	重回帰分析への拡張	86
4.5.1	標準偏回帰係数	86
4.5.2	自由度修正済み決定係数	87
4.5.3	多 重 共 線 性	88
4.5.4	変数選択と偏回帰係数の検定	88
4.6	Excelによる重回帰分析の実際	90
4.6.1	重回帰分析の手順	91
4.6.2	重回帰モデルによる予測と影響度評価	96

演習問題	98
コラム：見せかけの回帰	99

5. ロジスティック回帰分析

5.1 基本的な概念と位置づけ	100
5.2 回帰式の当てはめ	101
5.2.1 ロジスティック回帰曲線	101
5.2.2 オ ッ ズ	103
5.2.3 最尤法によるモデルパラメータの算出	104
5.3 Excelによるロジスティック回帰分析の実践	106
演習問題	115

6. 判別分析

6.1 基本的な概念と位置づけ	116
6.2 線形判別式	118
6.2.1 判別の考え方	118
6.2.2 判別モデルパラメータの導出	119
6.2.3 Excelによる線形判別式の判別	123
6.3 マハラノビスの距離	128
6.3.1 判別の考え方	128
6.3.2 多変量への拡張	130
6.3.3 Excelによるマハラノビスの距離の判別	132
演習問題	134

7. 主成分分析

7.1 基本的な概念と位置づけ	135
7.1.1 基本的な概念	135
7.1.2 主成分分析の結果のイメージ	139
7.2 主成分分析の解法と手順	140
7.2.1 主成分の求め方(2変数でのイメージ)	140
7.2.2 主成分の一般化	142
7.2.3 主成分の求め方(ラグランジュ未定乗数法)	142
7.2.4 主成分の求め方(分散共分散行列, 相関行列)	145
7.2.5 主成分分析の結果の解釈方法	147
7.3 Excelを用いた主成分分析の演習	149

7.3.1	第1主成分の算出	149
7.3.2	第2主成分の算出	151
7.3.3	残りの主成分の算出	153
7.3.4	寄与率, 累積寄与率, スクリーンプロット	153
7.3.5	主成分負荷量の算出および主成分の解釈	154
7.3.6	主成分得点の算出とそれぞれの地域の解釈	155
演習問題		156

8. 因子分析

8.1	基本的な概念と位置づけ	157
8.2	因子分析の解法と手順	159
8.2.1	基本モデル	159
8.2.2	因子負荷量の推定方法	161
8.2.3	因子数の決定	161
8.2.4	因子軸の回転	162
8.2.5	結果の解釈	163
8.3	Excelによる因子分析の実践	166
8.3.1	準備(因子数の決定, データの前準備)	166
8.3.2	因子分析用データの準備	167
8.3.3	因子負荷量の推定	169
8.3.4	分析結果の解釈	170
演習問題		174

9. クラスタ分析

9.1	基本的な概念と位置づけ	175
9.1.1	クラスタ分析の基本的な考え方	175
9.1.2	階層的手法と非階層的手法	177
9.1.3	クラスタ分析の手順	177
9.2	個体間の非類似度	178
9.3	クラスタ間距離の算出方法	179
9.3.1	階層的クラスタリング	179
9.3.2	クラスタリングの基本アルゴリズム	180
9.3.3	非階層的クラスタリング	188
9.4	デンドログラム	189
9.5	Excelによるクラスタ分析	190
演習問題		193

10. 数量化理論

10.1	基本的な概念と位置づけ	194
10.2	数量化理論 I 類	195
10.2.1	質的データの数量化	195
10.2.2	カテゴリーウェイト	196
10.2.3	Excel による数量化理論 I 類の分析手順	197
10.3	数量化理論 II 類	201
10.3.1	質的データの数量化	201
10.3.2	カテゴリーウェイト	201
10.3.3	Excel による数量化理論 II 類の分析手順	205
10.4	数量化理論 III 類	208
10.4.1	基本的な概念と位置づけ	208
10.4.2	解 法	210
10.4.3	推定結果の解釈	212
10.4.4	Excel による数量化理論 III 類の分析手順	213
10.4.5	結 果 の 解 釈	216
	演 習 問 題	218
	付 録	220
	引用・参考文献	228
	索 引	230

執筆者一覧 (執筆順)

寺内 義典 (国士舘大学)	1 章, 3.1, 3.3, 3 章コラム
川崎 智也 (東京大学)	2 章, 3.2, 5 章, 9 章, 10.1 ~ 10.3 1 章コラム, 4 章コラム
稲垣 具志 (東京都市大学)	4 章, 6 章, 2 章コラム
石坂 哲宏 (日本大学)	7 章, 10.4
兵頭 知 (徳島大学)	8 章

(所属は 2024 年 1 月現在)

1

土木・交通計画における多変量解析

本章では、事例を交えながら土木・交通分野における多変量解析手法の有用性について解説する。また、分析の目的やデータの種別にあわせた手法選択の概要や多変量解析を実施するための手順や実際の適用について概説する。

1.1 土木・交通計画分野における多変量解析の意義

本書は、土木・交通計画分野を学ぶ皆さんに**多変量解析** (multivariate analysis) を知り理解してもらうことを目的としている。この多変量解析は、その名のとおりに二つ以上の変数の関係を処理する統計解析手法の総称である。この多変量解析という統計手法の特徴を、事例を交えて説明しよう。

まず、バス路線の利用者数の実態を調査したデータがあるとする。このデータは、ある1日の路線ごとに便名と利用者数が並んでいる。これを用いて、路線ごとの日利用者数や、それを1日の便数で割った便あたり平均利用者数を求めることができる。記述統計手法により、利用者数が最多となる便や、利用者数の四分位数やヒストグラムを見ることで混雑状況なども把握できる。つまり統計とは、データを理解するための道具であると言えるだろう。

路線の日利用者数と沿線人口との関係を知りたいければ相関係数を求めることもできる。ここで、沿線人口は利用者数に影響を及ぼす関係にあることから、この沿線人口を**要因** (factor) と呼ぶ。この要因にデータという実態をいれるための箱を**変数** (variable) と呼ぶ。多変量解析は、この要因を表す変数が二つ以上あり、その変数間の関係や特徴を示す統計手法である。

実際に、利用者数は沿線人口だけでなく、複数の要因の影響を受けている。そこで、多変量解析の応用例として、新たなバス路線の利用者数を予測するケースを考えてみよう (図 1.1)。まず、沿線の人口や施設・土地利用が利用



小型車両で循環ルート運行

路線沿線には公営住宅や病院がある

図 1.1 バス路線の例 (M市I交通が運行するループバス)

者数を左右する要因となる。そのほか、ルート（直行型・両周り循環型・片周り循環型）、ダイヤ（間隔・パターンか否か・始発・終発など）、停留所（上屋・イス・近接性など）、車両（大きさ・席数・ステップの有無など）などのサービス水準や料金も要因となりうる。複数の要因が影響する現象を、人が理解できるかたち、人が使えるかたちにする統計手法が多変量解析なのである。

もう一つ、土木・交通分野における多変量解析の有用性について、ある失敗事例から説明しよう。その分析者は、生活道路を走行する自動車について、道路幅員と区間速度との関係を分析したいと考えた。そこで、幅員の広い道路区間 A と狭い道路区間 B で、十分な台数の自動車の速度を計測した。しかし、A と B の平均速度の差はわずかで、期待した有意な差がみられなかった。失敗の原因は、A が駅に近いことから、通行する自転車・歩行者交通の影響を受けたことであった。とはいえ、一般道の A と B の交通をすべて制御して計測することは非現実的である^{†1}。

この問題を解決する一つの方法として多変量解析がある^{†2}。多くの道路区間を対象に、速度に影響を及ぼすあらゆる要因のデータを収集し多変量解析を用

†1 試験路を用いる方法もあるが、生活道路の沿道条件や交通環境の再現性が課題として残る。そもそも試験路のある実験環境が、非常に貴重である。

†2 一般に他の要因を消す方法としてはランダム化が広く知られている。この場合、幅員以外の要因が同じように混在する程度に多数の道路区間での計測が必要となる。道路の場合、幅員そのものが沿道条件や交通環境と関連して整備されているため、厳密なランダム化より多変量解析による分析が有利と考える。

いることで、それらの要因のなかから幅員の影響をとり出すことができる。土木・交通工学の研究では、実験室のような厳密に比較可能な対象を用意できる環境を整えることが、実現性、再現性、費用といった面から困難であることも多い。多変量解析は、土木・交通工学において非常に有用である。

1.2 多変量解析手法の選択

1.2.1 さまざまな多変量解析手法

多変量解析は、ある現象を多数の要因どうしの関係の形（構造）として仮定し、その構造を数理的にモデル化し、統計的に検証しようとする手法である。沿線人口などの複数の要因とバス利用者数との間の関係が構造である。この構造にデータをあてはめ、バス利用者数の予測をするモデルを得ることが、分析目的である。異なる要因間の構造を仮定することで新たな手法が開発される。手法と分析目的は鶏とたまごの関係であり、分析目的からの要請に応じて進化してきた。さらに、量的データだけでなく質的データを取り扱いたいケースなど、変数の尺度や分布の違いに対応することも求められてきた。

こうして、あらゆる分野の人々が分析ニーズのなかで統計学を進歩させ、さまざまな分析目的、さまざまなデータに対応した多変量解析の手法が開発されてきた。現在の私たちは、先人たちによって用意されたこの多様な多変量解析の手法の特徴を理解し、みずからの分析目的とデータの種類に応じて、適切に分析手法を選択することが重要である。

多変量解析手法を選択するにあたり、本書ではその分析目的を「予測」、「影響力分析」、「要約」、「構造分析」の四つに大きく分類している。この4分類をもとに多変量解析手法の選択について解説する。

1.2.2 予 測

多変量解析では因果関係の結果に相当する一つの変数を一般に**目的変数** (response variable) と呼ぶ。また、因果関係の原因に相当する一つ以上の変数を一般に**説明変数** (explanatory variable) と呼ぶ。

〔1〕 目的変数も説明変数もおもに量的データの場合

バス路線の利用者数を予測する事例から、具体的に考えてみよう。ここで説明変数となるデータは、サービス水準を示す「運賃」、「運行頻度」、沿線の「人口」、「施設数」である。変数が量的データの場合、**重回帰分析**（multiple regression analysis）が選択される。これは、回帰分析^{†1}の説明変数を増やしたものである。

重回帰分析を用いると、既存のバス路線における「運賃」、「運行頻度」、「人口」、「施設数」のデータから、もっとも確からしい「利用者数」を統計的に推計する予測モデルを得ることができる。この予測モデルのあてはまりに問題がなければ、新規のバス路線の利用者数や、既存バス路線のサービス改善効果を予測することができる。このイメージを図1.2、図1.3に示す。

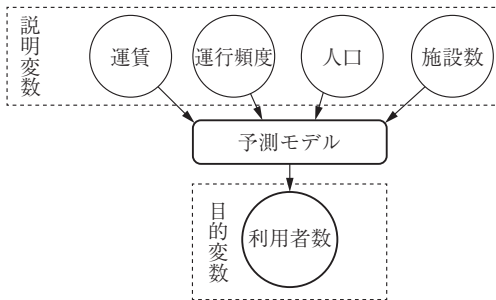


図1.2 重回帰分析

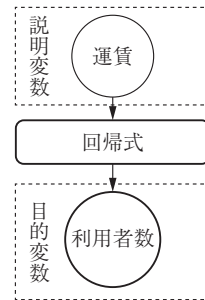


図1.3 回帰分析

〔2〕 目的変数が質的データの場合

目的変数を少し変えて、別の問題を考えてみよう。ある個人が移動をしようとするとき、バスを利用する／しない、のどちらかを推測したい。この目的変数は「利用する」、「利用しない」の2値の質的データとなる（この問題の説明変数には、「運賃」、「頻度」などのサービス水準と、個人属性や移動属性^{†2}を

†1 轟 朝幸ほか著：土木交通工学のための統計学，6章回帰分析，コロナ社（2015）を参照のこと。

†2 これらは質的データとなることが一般的だが，ダミー変数とすることで判別分析やロジスティック回帰分析を用いることができる。

用いる)。この場合、多変量解析では**判別分析** (discriminant analysis) や**ロジスティック回帰分析** (logistic regression analysis) を用いる。二つの分析手法は数理モデルが異なる。判別分析とロジスティック回帰分析のイメージを図 1.4、図 1.5 に示す。

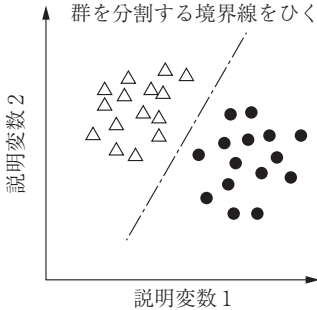


図 1.4 判別分析のイメージ

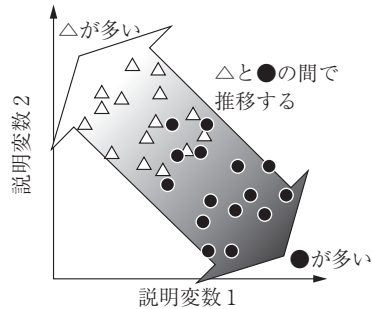


図 1.5 ロジスティック回帰分析のイメージ

判別分析は、各個体が属する群間の境界として適切な線を求める手法である。一方で、ロジスティック回帰は、二つの状態の間で、各個体が選択する確率が推移するイメージである。

〔3〕 扱う変数がおもに質的データの場合

扱う変数がおもに質的データの場合は、**数量化理論 I 類, II 類** (quantification methods of first type, second type) を用いる。目的変数が量的データで説明変数が質的データであれば、数量化理論 I 類を選択する。目的変数も説明変数も質的データであれば、数量化理論 II 類を選択する。なお、重回帰分析、判別分析、ロジスティック回帰分析において、説明変数の一部に質的データの変数を用いたい場合は、ダミー変数を用いることで適用可能である[†]。

予測に用いる個別の分析手法ごとに説明をしてきたが、実際には目的変数と説明変数のデータの尺度に応じて手法を選択するとよい。その選択フローを図 1.6 に示す。

† 数量化理論 I 類とは、すべての変数をダミー変数とした重回帰分析と同じである。数量化理論 II 類は、実質的にダミー変数による判別分析と考えてよい。

索 引

<p>【あ行】</p> <p>異常値 26</p> <p>一元配置分散分析 50</p> <p>因果関係 (因果) 14</p> <p>因子 157</p> <p>因子寄与 164</p> <p>因子得点 158</p> <p>因子負荷量 158, 163</p> <p>因子分析 9, 157</p> <p>ウォード法 187</p> <p>オッズ 103</p> <p>【か行】</p> <p>回帰係数 77, 84</p> <p>回帰式 70</p> <p>回帰直線 73</p> <p>回帰分析 70</p> <p>回帰平方和 82</p> <p>回帰モデル式 72</p> <p>階 級 22</p> <p>階級値 22</p> <p>カイザー基準 149, 161</p> <p>外 挿 97</p> <p>階層的方法 177</p> <p>外的基準 194</p> <p>カテゴリーウェイト 196</p> <p>カテゴリーデータ 19</p> <p>間隔尺度 20</p> <p>幾何平均 30</p> <p>疑似相関 13, 44</p> <p>記述統計 17</p> <p>共通因子 158</p> <p>共通性 165</p> <p>共分散 43</p> <p>共分散構造分析 10</p> <p>距離行列 179</p>	<p>寄与率 147, 164, 212</p> <p>鎖効果 182</p> <p>クラスター 8</p> <p>クラスター分析 7, 175</p> <p>クロス集計 13, 61</p> <p>群間変動 52, 120</p> <p>群内変動 52, 120</p> <p>群平均法 184</p> <p>決定係数 80, 85, 90</p> <p>検証の因子分析 10</p> <p>検 定 83, 90</p> <p>合成変数 135</p> <p>構 造 3</p> <p>誤判別率 127</p> <p>固有値 145</p> <p>固有値問題 146</p> <p>固有ベクトル 146</p> <p>【さ行】</p> <p>最小値 34</p> <p>最小二乗法 73, 75, 161</p> <p>最大値 34</p> <p>最短距離法 180</p> <p>最長距離法 183</p> <p>最頻値 33</p> <p>最尤法 104, 161</p> <p>残 差 73, 96</p> <p>残差平方和 75, 82</p> <p>算術平均 30</p> <p>散布図 13</p> <p>サンプルスコア 196</p> <p>シグモイド曲線 103</p> <p>時系列データ 45</p> <p>実績値 73</p> <p>質のデータ 19</p> <p>四分位数 1, 34</p> <p>斜交回転 162</p>	<p>主因子法 161</p> <p>重回帰分析 4, 71, 86</p> <p>重心法 186</p> <p>重相関係数 93</p> <p>従属変数 70</p> <p>自由度修正済み決定係数 88</p> <p>縮約化 135</p> <p>樹形図 177</p> <p>主成分 135</p> <p>主成分得点 7, 147</p> <p>主成分負荷量 148</p> <p>主成分分析 7</p> <p>順序尺度 20</p> <p>情報量 138</p> <p>数量化理論 I 類 5</p> <p>数量化理論 II 類 5</p> <p>数量化理論 III 類 7</p> <p>スクリープロット 149</p> <p>スクリープロット基準 161</p> <p>スタージェスの公式 24</p> <p>正規分布 32, 36</p> <p>正規方程式 76</p> <p>成分 (軸) の数 213</p> <p>説明変数 70</p> <p>線形式 73</p> <p>潜在因子 9</p> <p>尖 度 36</p> <p>層 化 25</p> <p>相関係数 212</p> <p>相関比 122</p> <p>総合化 135</p> <p>相対度数 23</p> <p>総変動 52</p> <p>双峰性 25</p> <p>ソルバー 77, 123</p>
--	--	--

【た行】	非階層的方法	177	【や行】		
	ヒストグラム	1, 22	尤 度		104
対称性	25		尤度比検定		111
対数尤度	105		ユークリッド距離		129
多元配置分散分析	52		予 測		96
多重共線性 (マルチコ)	13, 88				
多重比較	50		【ら行】		
多重ロジスティック			量的データ		20
回帰分析	101		理論値		73
多峰性	25		類似度		175
ダミー変数	5, 19		累積寄与率		147
単回帰分析	71		累積相対度数		23
探索的因子分析	10		レンジ		34
単峰性	25		ロジスティック回帰式		103
中位数	33		ロジスティック回帰分析		5, 100
中央値	33				
柱状図	22		【わ行】		
直交回転	162		歪 度		35
デンドログラム	177				
独自因子	158		【英字】		
独自性	165		ANOVA		50
独立性の検定	65		CFA		10
独立変数	70		CV		32
度 数	22		EFA		10
度数分布表	22		F 分布		223
			k 平均法		177, 188
【な行】			q パーセンタイル		34
二元配置分散分析	51		q 分布		227
			Turkey-Kramer の方法		59
【は行】			t 分布		222
箱ヒゲ図	34		χ^2 分布		221
外れ値	26, 44				
判別直線	118		【ま行】		
判別得点	119		マハラノビスの距離		129
判別分析	5, 116		無次元量		32
			名義尺度		19
			メジアン法		187
			目的変数		70
			モデル		3
			モデルパラメータ		73

— 著者略歴 —

川崎 智也 (かわさき ともや)
2011年 東京工業大学大学院理工学研究科博士後
期課程単位取得退学 (国際開発工学専攻)
2012年 博士 (工学) (東京工業大学)
現在, 東京大学講師

稲垣 具志 (いながき ともゆき)
2008年 大阪市立大学大学院工学研究科後期博士
課程修了 (都市系専攻), 博士 (工学)
現在, 東京都市大学准教授

寺内 義典 (てらうち よしのり)
2000年 福井大学大学院工学研究科博士後期課程
修了 (システム設計工学専攻), 博士 (工学)
現在, 国士館大学教授

石坂 哲宏 (いしざか てつひろ)
2007年 日本大学大学院理工学研究科博士後期課
程修了 (社会交通工学専攻), 博士 (工学)
現在, 日本大学准教授

兵頭 知 (ひょうどう さとし)
2016年 愛媛大学大学院理工学研究科博士後期課
程単位取得退学 (生産環境工学専攻)
2016年 博士 (工学) (愛媛大学)
現在, 徳島大学准教授

土木・交通計画のための多変量解析 (改訂版)

Multivariate Analysis for Infrastructure and Transportation Planning (Revised Edition)

©Kawasaki, Inagaki, Terauchi, Ishizaka, Hyodo 2017, 2024

2017年7月31日 初版第1刷発行

2024年3月20日 初版第3刷発行 (改訂版)



検印省略

著者	川崎 智也
	稲垣 具志
	寺内 義典
	石坂 哲宏
	兵頭 知
発行者	株式会社 コロナ社
	代表者 牛来 真也
印刷所	萩原印刷株式会社
製本所	有限会社 愛千製本所

112-0011 東京都文京区千石 4-46-10

発行所 株式会社 コロナ社

CORONA PUBLISHING CO., LTD.

Tokyo Japan

振替 00140-8-14844・電話 (03)3941-3131 (代)

ホームページ <https://www.coronasha.co.jp>

ISBN 978-4-339-05282-4 C3051 Printed in Japan

(西村)



JCOPY

< 出版者著作権管理機構 委託出版物 >

本書の無断複製は著作権法上での例外を除き禁じられています。複製される場合は、そのつど事前に、出版者著作権管理機構 (電話 03-5244-5088, FAX 03-5244-5089, e-mail: info@jcopy.or.jp) の許諾を得てください。

本書のコピー、スキャン、デジタル化等の無断複製・転載は著作権法上での例外を除き禁じられています。購入者以外の第三者による本書の電子データ化及び電子書籍化は、いかなる場合も認めていません。落丁・乱丁はお取替えいたします。