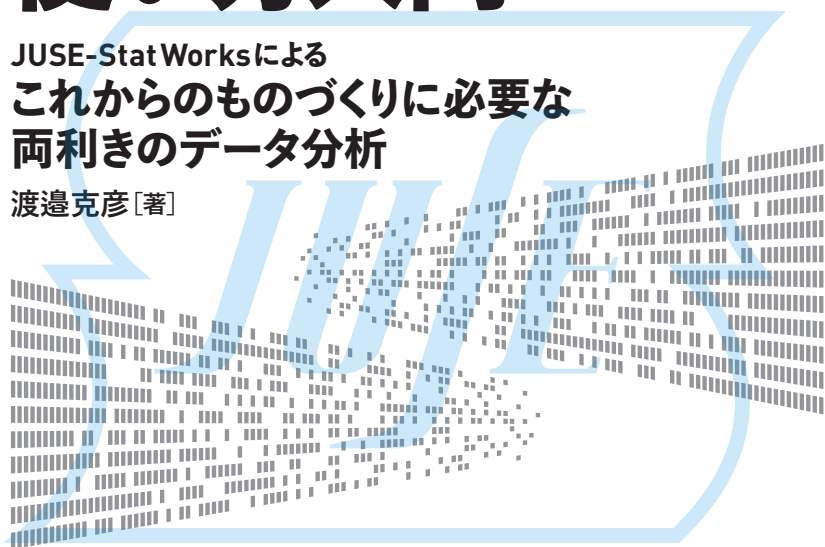


# 上手な機械学習と 統計的品質管理の 使い方入門

JUSE-StatWorksによる  
これからのものづくりに必要な  
両利きのデータ分析

渡邊克彦 [著]



日科技連

## 推薦のことは

日本の品質管理には、顧客満足・トップのリーダーシップ・全員参加を重視するという理念があり、方針管理・機能別管理・日常管理といった品質経営のしくみがある。そして、それらを具体的に支えるために小集団活動が行われる。これは科学的活動でなければならないと強調される。すなわち、データにもとづいた判断が大切である。そのために統計的方法を用いた活動が必要となり、それを統計的品質管理(Statistical Quality Control : SQC)とよぶ。

いま、世の中は混乱している。「品質が悪い」「品質不祥事」といったネガティブな言葉とともに品質管理が話題になり、“良い製品を作るシステム”という品質管理本来の意義が見失われかけている。

SQCは地道に継続していくからこそ力を発揮するのだが、ビッグデータ・AI・機械学習といった言葉が氾濫し、即効性が求められ、大切なものがわからなくなっている。「SQCは古いから最初から機械学習を勉強するのがよい」と言う人がいる。しかし、機械学習はどの範囲の解析手法を意味しているのだろうか？ SQCと機械学習とは何が違うのだろうか？ 巷にあふれる機械学習の書籍や解説文献のなかには、機械学習の手法の一部であるかのように重回帰分析・ロジスティック回帰分析をはじめとする従来のSQC手法が取り扱われている場合がある。「整理が必要だ」と思う。

本書の著者の渡邊克彦氏はトヨタ自動車(株)の技術者である。トヨタはかねてより品質管理のしくみを充実させてきた。比較的最近でも「次工程はお客様」を独自に実施するための「自工程完結活動」を提唱するなど、そのしくみを進化させている。トヨタの品質管理活動は名実ともに世界の模範である。それを支えているのは継続的に実施してきたSQCである。筆者自身もその活動を模範としてきた。渡邊氏は多くの優秀で誠実な先輩諸氏からの薫陶を受けて研鑽されてきた。筆者は、渡邊氏が本書を執筆されていることについて噂で知っていた。「整理が必要だ」と前述したが、渡邊氏ならそれをうまくできるだろう

など感じていた。また、どのように整理されるのを楽しみにしていた。  
そんな折、図らずも推薦のことはを寄稿させていただくことになった。

筆者は、本書を読ませていただいて、次に挙げる著者の視点を強く感じた。

- 従来の SQC の手法と機械学習の手法の特質を明確にする。
- 現代でも変わらない「SQC の手法の効用」を適切に説明する。
- 「機械学習の手法がなぜ必要なのか」を解説する。

トヨタのように SQC を風土として積み上げてきた場合には、上記のように整理した形で教育するのが理想だと思う。多変量解析法の知識をもっている一般の読者に対しても同様である。「渡邊氏なら上手に整理するだろうな」という筆者の予感は的中した。

本書は、「ものづくりにおけるデータ分析」「データ可視化」「層別」「情報の要約」「予測」「分類」「外れ値検出」「相関分析」「総合演習」という章立てになっている。まさに機能別管理(機能で分類)となっている。そして、章末にちりばめられているコラムや補足は秀逸である。渡邊氏の長年のトヨタでの教育と社内コンサルタントの経験がにじみ出ている。

本書は(株)日本科学技術研修所の統計解析ソフトウェア JUSE-StatWorks のガイドブックという性質をもっているため、このソフトウェアを操作できる読者が主な対象である。また、本書では数式をほとんど用いずに要点を言葉で解説しているので、手法の数理をより深く理解したい読者は、中上級の書籍を参照する必要がある。しかし、本書には「まず SQC との違いを把握したうえで機械学習の手法の特徴を掴んでほしい」「そしてとりあえずは使えるようになってほしい」「そのうえで次のステップに進んでほしい」という思いが込められていると感じる。

本書で学ばれることをお勧めしたい。

2021 年 2 月

早稲田大学創造理工学部経営システム工学科

教授 永田 靖

## まえがき

トヨタ自動車では約70年間、統計的品質管理(SQC)を問題解決の有効なツールとして位置づけ、ものづくりにおける品質確保や技術力向上に活用してきた。近年、IoTの発展により機械学習が注目され、SQCでは対応が困難であった問題の解決が可能となってきている。このように機械学習の登場で、さらなる問題解決力の向上が期待される一方で、両者は同じデータ分析手法でありながら生まれが違うなどの理由から、活用にあたっていくつかの課題が存在している。そのなかの1つに「SQCで既に活用されている多変量解析法と機械学習との役割分担」があり、これらの整理が求められている。

筆者は、この「多変量解析法と機械学習との役割分担」というテーマで、2019年12月に開催された「第29回 JUSEパッケージ活用事例シンポジウム」(主催：㈱日本科学技術研修所)の特別報告として「品質技術力向上に繋げるSQCと機械学習のより良い使い方について」を講演した<sup>1)</sup>。本書はその内容をもとに書き下ろしたものであり、ものづくりにおいて国内で広く使われている統計解析ソフトウェア JUSE-StatWorks/V5 を用いて、目的や対象データに沿ってSQCと機械学習のより良い使い方、いわゆる「両利きのデータ分析」を解説している。詳細な分析手順まで示しているのだから、読者の皆様に実践で活用してもらい、出す成果を少しでも高いものにできれば幸いである。

本書の読者対象は、主にSQCに触れている「ものづくりにかかわる実務者」のなかで機械学習をこれから勉強してみたい方々である。とはいえ、「機械学習を活用している実務者」のなかでSQCにも興味があるという方々にも参考になるだろう。いずれにしても実務者に役立つことを第一に考えている。

筆者は本書をまとめる際、以下の3つを意識した。

- ① SQCおよび機械学習の優劣を論じるよりは、それぞれの優れた点を目的に応じて使い分けできるような内容とすること
- ② 読者が両者の使い分けをより理解しやすくするため、具体的なデータ

を使ってデータ分析を体感できること。そのため、数式などの理論の記述は極力少なくした。

- ③ 理論の記述を少なくする分、本書の内容を実務で活用する際に抑えておくべきポイントを示すこと(章末の「コラム」でも実践活用のコツや陥りやすい罠をまとめている)

上記で「統計解析ソフトウェア JUSE-StatWorks/V5 を用いた」と書いた。これを読んだ読者のなかには「機械学習を活用するのに、どうして R や Python を使わないのか?」と思った方がいるかもしれない。もちろん、筆者には明確な理由がある。それは、JUSE-StatWorks が「プログラムを組む必要なく(ノンプログラミングで)手軽に Excel 感覚で活用できるソフトウェアだから」である。

R や Python は、確かに有名かつ便利なオープンソースのソフトウェアである。しかし、これらは読者に一定以上のデータ分析とプログラミング力を要求する。ものづくりの世界は「ポストコロナ」の影響もあり、ただでさえ実務者にますます負担がかかる状況下にある。こうした点を考慮し、「手軽に両利きのデータ分析を知り、実践での活用に繋げてもらえる」と考えたツールが、JUSE-StatWorks だったのである。

JUSE-StatWorks を使ったことのない方は、(株)日本科学技術研修所のウェブサイトから本書の内容に対応した無料トライアル版(扱えるデータは最大 30 変数 × 200 サンプル)を利用できる。また、上記のウェブサイトには JUSE-StatWorks の製品概要や活用事例なども掲載されている。本書を活用するにはぜひ「目次」の直後(p.xii 参照)にある「JUSE-StatWorks/V5 総合編 & 機械学習編 R2 のご案内」を見たらうで、ダウンロードしてほしい。

本書で扱う SQC 手法は「多変量解析法」を中心した手法群が多く、機械学習の手法は多変量解析法の拡張版ともいえる「正則化回帰分析」「カーネル主成分分析」をはじめとした個々の手法を解説している(詳細は 1.2 節参照)。

データ分析の目的として、「データ可視化(第 2 章)」「層別(第 3 章)」「情報の要約(第 4 章)」「予測(回帰)(第 5 章)」「分類(第 6 章)」「外れ値検出(第 7

章)「相関分析(第8章)」を挙げ、第2～8章で、それぞれに対応するSQC手法および機械学習手法を並べて解説している。これらの理解度を図る集大成として実践に即した事例を用意し、「総合演習(第9章)」とした。

また、各章末には筆者が長年トヨタおよびトヨタグループにおける困り事解決の実践支援に携わった経験から得たデータ分析の「コツ」や「陥りやすい罠」をコラムとしてまとめたので、これらも参考にさせていただきたい。

本書をまとめるにあたり、実に多くの方々より、有益なご意見や励ましをいただいた。これらは筆者が執筆を進めるうえで大きな力となった。

特に、早稲田大学の永田靖教授には、お忙しいなか、「推薦のこぼし」を賜っただけでなく、原稿の隅々まで精査いただき、筆者が気づかなかった用語の使い方をはじめ、本書をよりよく仕上げるご意見を数多く頂戴した。

さらに、(株)日本科学技術研修所の片山清志取締役、犬伏秀生部長、(株)日科技連出版社の戸羽節文社長、鈴木兄宏部長、田中延志係長、またトヨタ自動車(株)の北村晋部長、四手井肇前室長には、本テーマの重要性をご理解いただき、本書を執筆する機会を与えていただいた。

また、トヨタグループデータ分析分科会メンバーである愛知製鋼(株)の知念広秋氏、(株)ジェイテクトの内藤甲矢雄氏、トヨタ車体(株)の松本浩司氏、豊田通商(株)の伊藤勇介氏、アイシン精機(株)の丸谷守氏、(株)デンソーの鈴木則之氏、トヨタ紡織(株)の森部総一氏、(株)豊田中央研究所の水野雅彦氏、トヨタ自動車東日本(株)の花石昇氏、豊田合成(株)の井上純矢氏、日野自動車(株)の田中誠氏、ダイハツ工業(株)の本多匠氏、トヨタホーム(株)の土屋清明氏、トヨタ自動車(株)の近藤一雄氏、田中宏一氏には、毎月の会合で、SQCと機械学習のより良い使い方のヒントを数多くいただいた。そのなかでも特に、(株)豊田自動織機の久保田享氏、トヨタ自動車九州(株)の佐々木康博氏、則尾新一氏、トヨタ自動車(株)の阿部誠氏には、草稿の段階から丹念に査読をいただき、有益なコメントを多数賜ることができた。

最後に、トヨタ自動車(株)の小杉敬彦氏には、筆者が2006年にTQM推進部

に異動して以来，SQCの基礎から応用までをご指導いただけただけでなく，講師スキルや社内に推進体制をつくるノウハウ，学会発表や社外有識者との交流機会の創出など，TQM活動についても，幅広くご指導いただいた。

筆者にとって2作目となる本書でも相当の苦勞をしたが，小杉氏から非常に有益なアドバイスがあったおかげで，何とか道筋をつけられた。筆の進まない筆者が原稿を完成させ，無事に刊行できたのは，相当量の時間を原稿の査読・検討・助言に費やしていただいた小杉氏のご尽力によるところが大きい。

皆様方にはこの場を借りて厚く感謝申し上げるとともに，今後も問題解決力の向上に繋がる議論，研鑽のご指導・ご鞭撻を賜れば幸いである。

2021年2月

渡邊克彦



JUSE

## 目 次

推薦のことば *iii*

まえがき *v*

<b>第1章</b>	<b>ものづくりにおけるデータ分析</b>	<b>1</b>
1.1	トヨタグループにおける問題解決とデータ分析	1
1.2	本書でのSQCと機械学習の枠組み	2
1.3	データ分析の全体像	5
1.4	データサイズが大きくなる場合のSQCでの問題点	9
<b>第2章</b>	<b>データ可視化</b>	<b>15</b>
2.1	SQC—多変量関連図	15
2.2	機械学習—濃淡散布図, 密度プロット, 等高線図	20
2.3	本章のまとめ	22
2.4	補足—データクリーニング(前処理)	23
<b>第3章</b>	<b>層別</b>	<b>33</b>
3.1	クラスター分析と混合ガウス分布の使い分け	33
3.2	SQC—クラスター分析	34
3.3	機械学習—混合ガウス分布	39
3.4	本章のまとめ	42
3.5	補足—階層的クラスター分析	43



<b>第4章 情報の要約</b>	55
4.1 SQC —主成分分析	55
4.2 機械学習—カーネル主成分分析	60
4.3 本章のまとめ	63
<b>第5章 予測</b>	73
5.1 SQC —単回帰分析	73
5.2 SQC —重回帰分析	76
5.3 機械学習—正則化回帰(リッジ回帰, lasso 回帰, Elastic Net)	83
5.4 本章のまとめ	88
5.5 補足—クロスバリデーション	89
5.6 補足—ダブルクロスバリデーション	92
5.7 補足—モデル評価について	93
5.8 補足—SQC 重回帰分析の解析におけるチェック項目	97
<b>第6章 分類</b>	105
6.1 「SQC —判別分析」と「機械学習—サポートベクターマシン(SVM)」	105
6.2 「SQC —AID, CAID」と「機械学習—ランダムフォレスト」	115
<b>第7章 外れ値検出</b>	137
7.1 SQC —多変量管理図(マハラノビス距離による外れ値検出)	137
7.2 機械学習—1クラスSVM(カーネル法による外れ値検出)	141
7.3 本章のまとめ	145
<b>第8章 相関分析</b>	151
8.1 SQC —グラフィカルモデリング	151
8.2 機械学習—glasso	159
8.3 本章のまとめ	163

8.4 補足— glasso での外れ値の確認 ..... 164  
 8.5 補足— glasso での新たなサンプルデータにおける異常度の予測 ..... 165

第9章 総合演習 ..... 175

あとがき ..... 193

参考文献 ..... 195

索引 ..... 199

コラム一覧

コラム 1	異常と不良の違い	28
コラム 2	管理図で $\bar{X}$ を使う理由	31
コラム 3	ばらつきを考慮する重要性	46
コラム 4	検定と推定の使い方	49
コラム 5	○×評価に必要なサンプルサイズ	64
コラム 6	実験計画の種類と選択方法について	66
コラム 7	ばらつきに効く因子	70
コラム 8	回帰式の寄与率と予測精度	99
コラム 9	回帰式への質的変数の取り込み	101
コラム 10	実践における信頼性データ解析(その1)	130
コラム 11	実践における信頼性データ解析(その2)	132
コラム 12	調査に必要なサンプルサイズ	145
コラム 13	ものづくりにおける感性評価	167
コラム 14	数量化Ⅲ類における解析の工夫	171
コラム 15	DRBFM の講義で伝えること	173

## JUSE-StatWorks/V5 総合編 & 機械学習編 R2 のご案内

### ■「JUSE-StatWorks/V5 総合編 & 機械学習編 R2 トライアル版」[本書のサンプルデータ] ダウンロードのご案内

本書で使用しているパッケージ版のトライアル版(インストール後、30日間利用可能)、およびそのなかで使用しているサンプルデータを下記の(株)日本科学技術研修所の JUSE-StatWorks ウェブサイト(書籍用体験版・サンプルデータダウンロード)からダウンロードできます。

<https://www.i-juse.co.jp/statistics/support/pm/download.html>

実際に JUSE-StatWorks を動かしながら、本書の解説や解析手法の出力結果をお読みいただくとさらなる学習効果が期待できます。また、ウェブサイトからは、JUSE-StatWorks の製品概要や活用事例、簡易手順、パッケージの購入方法、典型的な研修カリキュラム、研修内容なども入手できます。

### ■注意事項

#### (1) JUSE-StatWorks/V5 総合編 & 機械学習編 R2 の使用制限

- ① 体験版で扱えるデータは最大 30 変数× 200 サンプルです。

制限を超える大きなファイルを読み込んだ場合は、左から 30 変数、200 サンプルまでが読み込まれ、制限を超えた部分のデータは読み込まれませんのでご注意ください(なお、製品版は 1,000 変数× 100,000 サンプルとなっています)。

- ② データの保存はできません。

#### (2) その他

- ① 上記の方法でうまくいかない場合は、下記の(株)日本科学技術研修所の JUSE-StatWorks ウェブサイト(お問い合わせ窓口)からご連絡ください。

<https://www.i-juse.co.jp/st/contact>

- ② 著者および(株)日本科学技術研修所、(株)日科技連出版社のいずれも、ダウンロードデータを利用した際に生じた損害についての責任、サポート義務を負うものではありません。
- ③ 上記のサンプルデータは、著者および(株)日本科学技術研修所に著作権があります。利用に当たっては、本書の購入者または購入者の所属する組織内でのみ使用を許諾します。許可のない外部公開や営利目的での使用は禁止します。

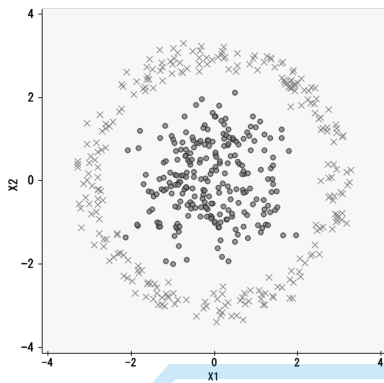
図表 1.1 本書で扱う SQC と機械学習の手法

手法		SQC (総合編プレミアム)	機械学習 (機械学習編)
目的			
データ可視化 (第2章)		多変量連関図, モニタリング	濃淡散布図, 密度プロット, 等高線図
層別 (第3章)		階層的クラスター分析 非階層的クラスター分析 (k-means 法)	混合ガウス分布
情報の要約 (第4章)		主成分分析	カーネル主成分分析
予測(回帰) (第5章)		重回帰分析	正則化回帰分析(リッジ回帰, lasso 回帰, Elastic Net)
分類 (第6章)	6.1 節	判別分析	サポートベクターマシン (SVM)
	6.2 節	AID, CAID	ランダムフォレスト
外れ値検出 (第7章)		多変量管理図, MT 法	1 クラス SVM
相関分析 (第8章)		グラフィカルモデリング	glasso

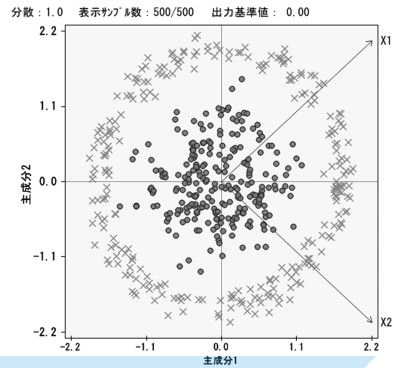
注) JUSE-StatWorks/V5 は Version 5.8 が前提である。

の狙いに違いがある。SQC は品質管理で長年使われてきたように、因果関係を見つけることに主眼を置いている。つまり、ものづくりで扱う製品や技術には何らかの物理法則が働いているため、そのメカニズムを明らかにするために、「作成したモデルが仮説と合っているか」「固有技術で説明できるか」などの技術的な妥当性を重視している。

一方、機械学習は、(やや乱暴に表現すると)一般的には徹底的に予測精度を追求する。仮に作成したモデルが、固有技術で説明できなかりうが、物理法則に反していようが、予測精度が高くなるなら、そのモデルで良しとする特徴がある。つまり、データを正として、どんなモデルでも良いと考えるのが機械学



図表 4.6 内側と外側に分かれたデータ



図表 4.7 主成分分析の結果

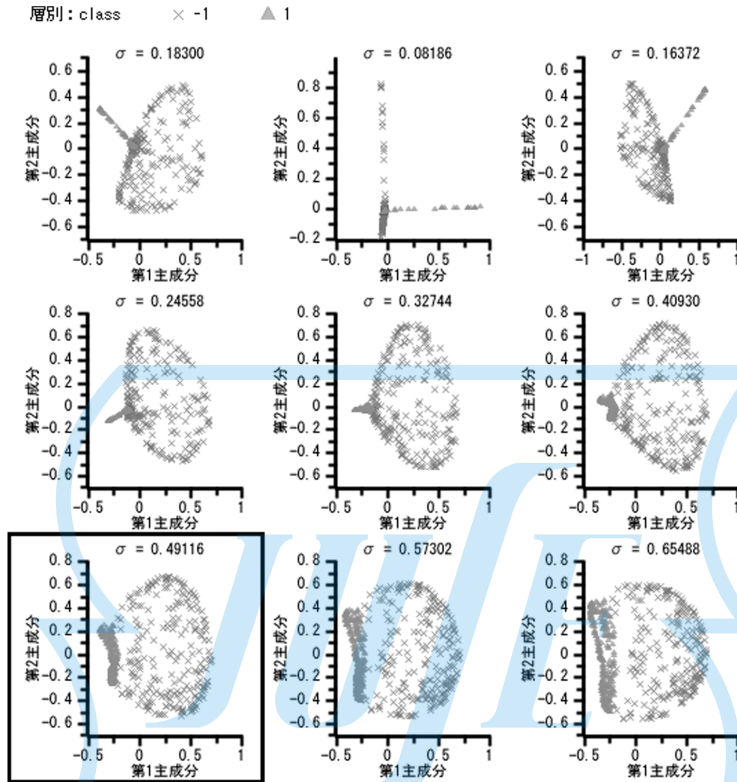
## 4.2 機械学習—カーネル主成分分析

カーネル主成分分析は、カーネル法によりデータを高次元特徴量空間に写像し、その空間で主成分分析を行う手法である。解析手順は、メニューから [手法選択]—[機械学習]—[カーネル主成分分析] をクリックする。表示される [変数選択] ダイアログで量的変数  $X_1$ ,  $X_2$  を選択し次へ進み、[分析パラメータ] のタブをクリックすると、カーネル主成分分析の解析結果が表示される。

メニューボタンの [層別] から質的変数 class を選択すると、図表 4.8 に示す、カーネルパラメータ  $\sigma$  ごとの第1主成分、第2主成分の散布図が表示される。このデータでは、内側と外側の分類が最もできているカーネルパラメータを選択すればよいため、 $\sigma = 0.49116$  を採用する。このように、カーネルパラメータを調整することで、非線形データに対しても識別できることがカーネル主成分分析の強みとなる。

メニューボタンの [パラメータ  $\sigma$ ] から  $\sigma$  の値を 0.49116 に変更し、[固有値] のタブに移動すると、図表 4.9 に示す固有値や寄与率を、[固有ベクトル] のタブでは固有ベクトルをそれぞれ確認することができる。

ここで、固有値および寄与率は主成分分析と扱いが異なることに注意が必要



図表 4.8 カーネル主成分分析の結果

である。カーネル主成分分析では固有値の合計は変数の数にならないなど、主成分分析と同じ基準で扱うことができない。つまり、主成分分析のように解釈に取り上げる主成分を累積寄与率 70~80% 以上や固有値 1 以上という基準がないことに注意したい。また、軸の解釈についても高次元特徴量空間に写像した軸を解釈することは、困難であるといわれている。

つまり、カーネル主成分分析では通常の主成分分析と違い、主成分に意味をもたせることができないという欠点がある。その一方、先に示したようにサンプルの層別や次に示す異常値の発見に期待ができる手法となる。

異常値の発見例として、主成分分析で解析した 5 教科のテスト結果をカーネ

## 7.3 本章のまとめ

ものづくりによく見られる正規分布に対して外れ値検出を行うのであれば、まずは多変量管理図や MT 法(基本原理としては両者は同じ)で解析し、マハラノビス距離で外れ値を発見すればよく、結果も解釈しやすい。

一方、1クラス SVM は偽陽性(%)の設定が必要で、ガウスクーネルの場合はカーネルパラメータの値によっても検出される点が異なる。そもそも、「分析対象データ=標準空間」とし、データの外側  $\alpha\%$  ( $\alpha$  は任意)は外れ値とする考え方のため、それらを留意して使う手法となる。また、前述のとおり1クラス SVM で検出されても、すぐに異常だと鵜呑みにせず、「なぜ高次元特徴量空間に写像することで外れ値と判定されたのか」「本当に外れ値として扱うべきか」を十分に吟味することが大切である。

### コラム 12 調査に必要なサンプルサイズ

実践支援では「どれくらいの数を調査すべきか」というサンプルサイズに関する相談はよくある。本コラムでは、特に相談が多い「①アンケート」「②工程能力指数算出」「③従っている分布の決定」に必要なサンプルサイズについて紹介していく。

まず、「①アンケート」の調査人数だが、例えば「新製品を購入した10万人を対象に満足度調査を実施したい。500人分の予算しかないが、もう500人追加して1000人調査したいという意見が出ている。しかし、本当に追加する必要があるのだろうか?」という相談があったとする。この場合、どのように回答すればよいだろうか?

調査人数を決める式として、一般的には次が使われている。

●著者紹介

渡邊 克彦 (わたなべ かつひこ)

【経歴・専門分野】

東京理科大学基礎工学研究科修士課程修了。トヨタ自動車(株)第3電子技術部に入社。その後、TQM推進部、SQCグループのグループ長として応答曲面法、信頼性、品質工学、感性評価などの社内講座における質向上の取り組みを行う。現在は、SQCおよび機械学習におけるデータ分析の普及推進や実践活用の支援に従事。

社内およびトヨタグループのデータ分析推進・指導アドバイザー、日本規格協会講師、中部品質管理協会講師。

【受賞歴・著作】

- ・2011年 (一社)日本品質管理学会 第1回 Activity Acknowledgment 賞。
- ・2013年 (株)日本科学技術研修所 第10回 JUSE-StatWorks 活用エキスパート賞。
- ・『開発・設計に必要な統計的品質管理』(共著、日本規格協会、2015年)。  
※2016年度日経品質管理文献賞受賞

上手な機械学習と統計的品質管理の使い方入門

JUSE-StatWorks によるこれからのものづくりに必要な両利きのデータ分析

2021年3月22日 第1刷発行

著者 渡邊 克彦  
発行人 戸羽 節文

検 印  
省 略

発行所 株式会社 日科技連出版社  
〒151-0051 東京都渋谷区千駄ヶ谷5-15-5  
DSビル  
電話 出版 03-5379-1244  
営業 03-5379-1238

Printed in Japan

印刷・製本 東港出版印刷

© Katsuhiko Watanabe 2021 URL <https://www.juse-p.co.jp/>  
ISBN 978-4-8171-9727-6

本書の全部または一部を無断でコピー、スキャン、デジタル化などの複製をすることは著作権法上での例外を除き禁じられています。本書を代行業者等の第三者に依頼してスキャンやデジタル化することは、たとえ個人や家庭内での利用でも著作権法違反です。