

無断使用をお断りします。日科技連出版社

# 多 様 性 工 学

個性を活用する  
データサイエンス

中田亨 [著]

日科技連

## まえがき

本屋に置いてある本がすべて同じ本だとしたら、これほど馬鹿馬鹿しいものはないだろう。本にせよ、料理にせよ、人材にせよ、世の中の大抵の物事は、多様なものを取り揃えているからこそ、個々の価値が引き立つのだ。しかし、例えば本を「資産」として帳簿に記録するときには「本の単価×冊数分の金銭的な価値がある」と素っ気ない査定をするしかない。こんな査定は本当の価値からは大きくかけ離れているが、他に有力な査定方法がないがゆえにしばしば実行され、混乱をもたらしている。多様性は大事だと知りつつも、それを分析したり制御することに、我々は慣れていない。

大学で卒業研究を指導していて、研究のセンスがあると感じさせる学生は、多様性の扱い方が巧みである。そもそも世界は多様なものの集積であり、多様であるからこそ価値を引き出すことが研究の目的となる。平均や中央値といった代表値だけを測っても、何もわかったことにはならない。ルーブル美術館を、その代表作であるモナリザ1枚だけで語るのとは不可能である。

研究とは踏み込んで考えることである。「大多数の日本人は味噌汁を好む。だから味噌汁を売ればよい」と荒っぽい議論では浅すぎる。「味噌汁を好まない少数派の人々はどうなっているか？」と観察対象に接近するべきなのだ。センスのある学生なら「少数派のなかにもメジャーな派閥があって、その特徴を調べれば、そこそこ有益な結論を出せるのではないか」と研究の目星をつける。

服装の流行を調べる研究なら、「去年の服装と今年の服装の差を測る」ということになり、平凡にやるなら、「両者の平均値をとって引き算せよ」となる。だが、「服装の平均値」なるものはあるのだろうか？ スカートとズボンの平均とは何だろう？ 禅問答のような話となって頓挫する。

趣味嗜好の世界は多様であることに価値がある。コレクションを収集し、多様なサンプルの一つひとつを尊重することが、分析者として正しい態度である。世の中に存在する何万種類ものスカートとズボンの大集団に素直に対峙する。

「大集団から、何が消え、何が増えたか」で、流行の動きを語るほうが、結論に実りが多い。

多様性を勉強しようとする、なぜか特定の狭い分野の話題に引き込まれてしまう。世間的に「多様性」といえば、エコロジー分野での「生物多様性」か、人間社会での少数派保護という文脈での「社会の多様性」の、2つに話を絞られる。そのせいで多様性を扱う書籍を検索すると、この2つの関連書が占める。一方、これら以外の分野は「多様性」に特段の関心をもってこなかった。むしろ観察対象の「多様性」は分析の邪魔で、平均を計算するなどして、ばらつきを無視することが普通であった。本来、「多様性」は幅広い分野に共通する重要事項であるのに、その利用法を学ぶ機会が少ないことは問題である。

本書では、多様性を「物事の多種性・取り揃えの広がりであり、物事の質や影響力を左右する重要事項」と捉えたうえで、その目的を「多様性をコントロールする技法を集成すること」とした。

多様性は、他者との違いがあってはじめて議論ができる、いわば「関係」の情報である。それは、どのように違うかという質と、どれだけ違うかという量という2つの側面から成り立っており、片方だけに注目しても意味がない。標準偏差などで量を分析する技法程度は学校でも習うが、それだけでは多様性の活用には至らない。本書で多様性の全体像およびその本質を示せたと思う。

本書が想定する読者は、モノ・ヒト・コトの多様性に対応する立場の“実務者”の皆さんである。この場合の“実務”は、学生のアルバイトや日々の家事から、高度な経営判断に至るすべての人間の作業を含んでいる(とはいえ、本書で提示した事例はビジネスパーソンを想定した内容になっている)。

このように“多様”な読者を想定したのは、どのような仕事でも、結局は多様性の問題に行きつくためだ。老若男女問わず、人は多様性を相手に日々作業や仕事をしているようなものである。多様性ゼロのワンパターンな仕事は味気ない。我々は、「どこまで状況に対応し、どう成果物を取り揃えるか」に心を砕いている。多様性を制するものはすべてを制するのである。

多様性は、実に広汎な分野に顔を出す。事の成否や死活に関わる重要要素で

もある。多様性の制御や競争戦略の技術も必要となる。それを観察し分析するには、平均を計算するだけでは済まず、多様性を扱うための専用のデータサイエンスが求められる。

- 品質管理では、製品のばらつきを減らす、つまり多様性の削減こそが目標となる。
- マーケティングでは、店舗にどれだけ多様な商品を取りそろえるかという、多様性の最適値の探索がテーマである。
- 工業製品の世界では、多様性が競争のきっかけとなる。コンピュータなら Windows と Macintosh、ビデオなら VHS と ベータマックスといった、多様な技術方式が乱立するがゆえに、シェア争いが起こり、企業の栄枯盛衰を織りなしてきた。
- 金融のリスク管理では、一つの投資先に集中すると大損しかねないから、分散投資することが鉄則だ。
- 災害は「1000年に1度の大地震」といった例外的な事象にこそ対策せねばならない。頻度が少ないからといって無視できない。自然の多様性を把握しておく必要がある。

多様性に関する知見は、数多くあるが、さまざまな学問分野に散らばっていて、その全体像は把握しづらい。また、なまじ理工系の実務に詳しいと「多様性は、正規分布を仮定して、エントロピか何かで測れば、それで一丁上がりではないか」という常識的ながら手垢のついた作戦に飛びついてしまって、多様性を丁寧に見ようとする動機は薄らぎがちである。

多様性には、不思議な側面がある。多様性は自己矛盾の傾向があって、「多様であるものは多様でない」という状況が成り立つことが多いのだ。映画や音楽は、既存のもの二番煎じでは価値がなく、「すべてがオリジナルであり、多様でなければいけない」という宿命にある。だが、ハリウッド映画ばかりが幅を利かしていたり、特定の人気歌手ばかりがヒットするというマンネリズムが潜んでいる。電気の街、秋葉原に代表される専門店街という現象も不思議だ。数多くの専門店のなかに、秋葉原でしか売っていないようなレアな商品を見つ

けることができる。一方、同業の店ばかりが集中している点を見れば、街としての業種の多様性は極端に低いといえる。

こうした現象から学べることは、「多様性は単に統計学や生物学の理論だけで説明できるものではなく、人間が多様性をどう感じ、どう制御してきたかを反映している」ということである。本書では、数理から人間の主観までとりまとめて、多様性の世界を巡っていきたい。

2021年1月

中田 亨

#### <本書で想定する読者の要望と、本書での解説箇所>

- 「まだ遭遇していない新種はどれほどいるのか知りたい」 → 3.4 節
- 「多様性の度合いを定量的に測りたい」 → 第1章
- 「多様性からメリットを引き出したい」 → 第1章と第5章
- 「多様性をコントロールしたい」 → 第2章と第5章
- 「多様な天災のなかでも、非常に稀な天災現象についてリスク管理したい」 → 3.4 節と 5.1 節
- 「(本や映画のように)同じ物が2つあっても無意味といえる、個性に価値がある物事を分析したい」 → 5.3 節
- 「やたらと項目が多い、高次元データを分析したい」 → 第3章と第4章
- 「不動産物件のようにやたらと項目が多い商品のなかから、顧客にマッチするものを選びたい」 → 5.2 節
- 「集団全体の平均像よりも個々の差や派閥の様態を分析したい」 → 4.3 節
- 「自らの競争力を高めるやり方を知りたい」 → 第5章
- 「多様性は安全を毀損しないか、あるいは安全に貢献できるかを知りたい」 → 5.5 節

多様性工学  
—個性を活用するデータサイエンス—  
目 次

まえがき	iii
<b>第 1 部 多様性工学の理論</b>	
<b>第 1 章 多様性の意味と意義</b>	3
1.1 多様性の効用—多彩なことはいいことだ	3
1.2 多様性の弊害	14
1.3 多様性の度合いを表す指標	21
1.4 「同じ」とは何か？	38
1.5 多様度は観点に依存する	43
<b>第 2 章 多様性増減の原理と法則</b>	47
2.1 多様性を生み出す原動力	47
2.2 ゆらぎが拡大される誘因	59
2.3 多様性を減らす原動力	65
2.4 多様性を保つ力	78
<b>第 3 章 モデル化—正規分布当てはめを超えて</b>	87
3.1 正規分布は特別な地位を占める	87
3.2 ブラック・スワンと街灯効果	89
3.3 平均を気にしないという新発想	92
3.4 その存在自体以外に存在証明がない存在	95
3.5 多様度の真値に迫る	97

3.6	多様性をとりこぼさない計測は難しい	105
3.7	無用な多様性と、それを捨てる能力	107
3.8	深層学習が多様性の見極めで人間を圧倒	109
3.9	解の個数の爆発とスパースデータの台頭	112
3.10	名義尺度が世の中を動かす	116
3.11	巨大な図書館—希少例は希少なまま呼び出せ	118
<b>第2部 多様性工学の用途</b>		
<b>第4章 多様性の分析</b> 123		
4.1	分析技法の概況	123
4.2	標本全体についてデータの相関関係を調べる技法	128
4.3	標本をクラスタに分割する発想	139
4.4	高多様度大規模データの樹形図分析	151
<b>第5章 多様性への戦略</b> 157		
5.1	品質管理	157
5.2	販売戦略	161
5.3	ポジショニング戦略	172
5.4	取引相手の多様性	178
5.5	安全とセキュリティのための多様性	183
あとがき		193
参考文献		195
索引		197

## 1.2 多様性の弊害

---

多様であること、ばらばらであることには、弊害もある。実はそれらは多様性の効用と表裏一体ともいえる。

### (1) 幅広い対応が必要でコストがかかる

メニューが多い食堂は、それだけ原材料を手広く準備せねばならず、また多数のメニューごとに調理法を学ばねばならない。調理に時間がかかる料理と、すぐにでき上がる料理とが併存していると、ある客には早く出せるが、その連れの客はずいぶん待たせてしまうといった不都合も生じる。コスト圧縮や効率性を考えるなら、牛丼専門店やハンバーガー専門店といった、メニューの多様性を減らした業態のほうが有利である。

工業製品の世界では、工業標準に従うことが定石である。標準に指定されている形と大きさだけに絞り込んで、部品を作り、売り、使う。全員がこれに従うことで、部品の仕様をゼロから取り決める手間が減り、部品の供給体制を低コストに維持できる(標準でない特注の部品は、調達コストが跳ね上がる)。

規格から外れた特注品には、ホールドアップという問題もある。特定の納入先の1社でしか使わない特注部品を製造するための専用ラインを構築してしまうと、納入先がその部品を買わなくなればラインへの投資が無駄になるという弱みが生じる。納入先がその弱みにつけ込んで購買停止をちらつかせて、部品の価格を強引に値切るのである。

逆に、売り手がホールドアップを仕掛けてくる事例もあり、例えばコンピュータシステムの「1円入札」という現象が挙げられる。最初に、自社の手による特注のコンピュータを納入してしまえば、特注ゆえの互換性の欠如が他社への参入障壁(2.3節(1)脚注23(p.65))となって、その後の保守や拡張工事のすべてで、自社システムを選ばせ続けることが期待できる。よって、最初の入札ではどんなに安くても勝つべきであり、1円の入札する事例が実際に起こった<sup>4)</sup>。

取り扱い商品があまりに多様すぎることはブランドイメージをぼやかす。百



貨店という業態は近年苦戦を強いられている。百貨とはいえずべての商品を売っているわけではない。かつては生活全体をカバーする広汎な商品をそろえることができたが、現代人の多様な生活をカバーすることはもはや無理である。そうになると、何を売っているかがわかりにくいという欠点が出てしまう。客は欲しい商品があれば、その筋の専門店に足を向ける。昭和の頃は、企業でも学界でも「総合研究所」や「中央研究所」というネーミングが流行したが、これも部外者には印象が不明確である。いっそ「睡眠研究所」や「鉄鋼研究所」と限定した名前のほうが、断然覚えてもらいやすい。

## (2) 品質管理がおろそかである印象を与える

そろっていないもの、混じりつけのあるものからは、雑に管理されているという印象を感じる。大型店では野菜は、同じような大きさや、色・形のものばかりが売られている。植物の造形がそんなにそろはずはなく、厳しく選別した結果、そうなっているのである。ばらばらすぎでは質が悪いように見える。

大学入試のランキングは、合格者のなかで一番成績の良い人ではなく、一番悪い人の「偏差値」や「合格可能最低点」で測られる。よって、あまりに多様な人材に入学を許すと、簡単に入れるレベルの低い学校と受けとられてしまう。

珍芸をもつ人材すら集めるといふ鶏鳴狗盗について、北宋(960～1127年)の政治家・学者である王安石(1021～1056年)は、「そういう妙な人材が周りにいるから、本当に有能な人材が集まらなかったのだ」と批判している。

---

4) 産経新聞社「富士通やヤフー、「1円入札」過去にも問題に」(2019年11月20日)には、以下の事例が掲載されている(<https://www.sankei.com/affairs/news/191120/afr1911200029-n1.html>)。

- 1988年：富士通が広島市水道局のシステム開発を1円で落札し、NECなど大手IT企業の1円入札が相次ぎ表面化。
- 2005年：財務省が国保有の近代金貨を売却するオークションの補助業務をめぐってヤフーなど2社が1円で入札し、受注。
- 2013年：林野庁が実施した衛星携帯電話の調達に絡み、KDDIとソフトバンクグループが相次ぎ1円入札で契約したことが発覚。

## 4.2 標本全体についてデータの相関関係を調べる技法

### (1) データの下見—散布図行列

多様な分析対象の例として、プロ野球チームのシーズン成績を挙げてみよう。架空のデータではあるが、表 4.3 に示す「勝利」「打率」「二塁打」「三塁打」「本塁打」「押出得点」「得点」「盗塁」「防御率」「奪三振」「セーブ」「完封勝」「完投」「失点」の、14 項目のデータがあるとする。

これら数字の羅列だけを見て、「勝つチームの特徴は〇〇だ」とすぐに見抜くことはできない。そのためにはまず、何らかの傾向を発見する必要がある。そのためには、視覚に訴えるようグラフ化するのが正攻法である。しかし、14 次元のデータは、2 次元の紙上に表現できない。

高次元のデータを 2 次元(紙上)に落とし込むため、統計手法を使い、無駄な次元を捨てて、重要な情報だけを濃縮したい。しかし、「情報の取捨選択にどの方法を使うのが最善であるか」は一度データを視覚化しなければ判断できない。そこで、データ項目のペアを総当たりで散布図である「散布図行列」(scatter plot matrix)を描いて観察する(図 4.1、図 4.2)。

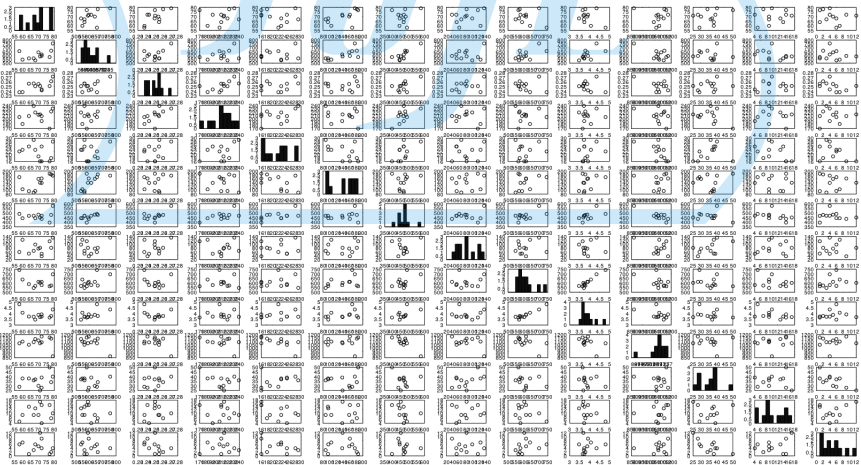
「散布図行列」にある 2 つのデータ項目の散布図のなかに、人間の目だと「両者にはいかにも関係なり法則性がありそう」と思っても、統計手法によってはそれがわからない場合がある。というのも、2 つのデータの関係の強さを積率相関係数(しばしば単に「相関係数」とよばれる)で測る評価の事例が非常に多いからである。

データを視覚化して下見するための要点は下記のとおりである。

- 散布図行列を作ってみて、データが全体として、一つの直線状(や細長い楕円状)に分布している部分は、積率相関係数にもとづく分析方法が有効である。その代表は主成分分析である。
- それ以外の形の分布、例えば図 4.3 に示すような曲線的な分布や、分散にかかわる法則性があるものは、積率相関係数では関係の存在を見抜かず、関係なしと誤答しがちである。「一つの直線・楕円として捉えるこ

表 4.3 架空の野球 12 球団の成績

チーム	勝利	得点	打率	二塁打	三塁打	本塁打	押出得点	盗塁	失点	防御率	奪三振	セーブ	完封勝	完投
A	79	756	0.272	231	28	176	577	134	695	4.35	872	40	10	4
B	78	662	0.262	222	18	182	479	81	573	3.78	1140	34	7	2
C	77	579	0.251	170	16	186	396	115	565	3.65	1157	50	15	1
D	72	613	0.244	193	25	138	470	50	578	3.73	1084	39	4	4
E	72	594	0.242	208	16	161	436	37	613	3.92	1172	28	15	9
F	71	588	0.252	213	16	139	451	81	602	3.69	1070	25	17	12
G	70	640	0.256	212	19	159	482	77	613	3.88	1112	30	9	5
H	68	536	0.251	219	23	96	444	98	569	3.47	1134	37	13	6
I	67	562	0.257	239	29	88	470	65	547	3.71	1087	36	9	3
J	64	562	0.256	208	25	95	469	51	588	3.74	1047	38	14	1
K	60	546	0.238	184	23	105	443	123	638	4.06	1095	38	9	7
L	58	656	0.238	205	22	168	490	59	742	4.80	1069	29	5	6



注) 2つの項目の間に相関関係があるかを一覧して調べる(Gnu Octave の plotmatrix 関数で作図)。

図 4.1 野球データ(表 4.3)の散布図行列

## ●著者紹介

中田 亨(なかた とおる)

2001年、東京大学大学院工学系研究科先端学際工学専攻修了。博士(工学)。

現在、国立研究開発法人 産業技術総合研究所 人工知能研究センター 副連携研究室長。国際電気標準会議(IEC)の「ヒューマンファクターと機能安全」に関する規格検討グループ(TC65 AHG16)のメンバー。内閣府消費者安全調査委員会専門委員、中央大学大学院理工学研究科客員教授などを兼務。

研究テーマは、人工知能を活用した産業のデジタルトランスフォーメーション(DX)による安全の向上。特にヒューマンエラーの抑止を専門とし、事故を分析し、人間の弱点をカバーする人工知能の技術について研究開発を進めている。

### 【主な著書】

『防げ 現場のヒューマンエラー』(日科技連出版社)、『ヒューマンエラーを防ぐ知恵』(化学同人)、『速攻 理系のための卒業論文術』(講談社)、『超入門ヒューマンエラー対策』(日科技連出版社)など。

## 多様性工学

個性を活用するデータサイエンス

2021年2月28日 第1刷発行

検印  
省略

著者 中田 亨  
発行人 戸羽 節文

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 (株)三秀舎

© Toru Nakata 2021

URL <https://www.juse-p.co.jp/>

ISBN 978-4-8171-9732-0

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。