

本川 裕 著

統計データはおもしろい！

～相関図でわかる経済・文化・世相・社会情勢のウラ側～

技術評論社、2010年、255ページ、定価1,659円



渡辺 美智子

近年、情報化の進展とともに統計への社会的関心が高まってきており、ここで取り上げる本書をはじめとして、統計データの活用に関する一般の書籍が多数出版されるようになった。とくに海外ではその傾向が顕著で、日本でも翻訳された「ヤバい経済学 [増補改訂版]」スティーヴン・D・レヴィット/スティーヴン・J・ダブナー(著)；東洋経済新報社(2007)や「その数学が戦略を決める」イアン エアーズ(著)；文春文庫(2010)、「ヤバい統計学」カイザー・ファンク(著)；阪急コミュニケーションズ(2011)などが注目されている。これらの内容は、統計学の学習書ではなく、統計分析の手法がいかにかに現実のデータに駆使され、それが実社会でどのように活用されているのか、その結果、私たちの暮らしや仕事のシステムが驚くほど便利になっていることを読み物として提供している。

このような中で、本書は、散布図(書籍の中では相関図と呼んでいる)を中心として、様々な分野の実際のデータが示す相関や連関関係から、社会や経済の実態を観察する方法や具体的な事例を提供し、著書のいう「統計グラフのもつ“威力”」を、読者に知らしめることを目的とした内容構成になっている。ヒストグラムや箱ひげ図と並んで統計3大グラフの一つである散布図を使うことで、どこまで社会や経済・文化・世相の実態に迫ることができるのかを実に興味深く刺激的に具体例で示している。

本書は、以下に挙げる12の章
 第1章：典型的な相関図
 第2章：散布図によるグルーピング、
 第3章：散布図のさまざまな使い方
 第4章：バラツキを比較する
 第5章：相関度の高くない相関図
 第6章：相関を掘り下げる
 第7章：男女の相関、第8章：反転する相関
 第9章：片相関、第10章：意味のある無相関
 第11章：相関からはずれたものに注目する
 第12章：時代とともに移り変わる相関

と 付録：相関図をめぐる話題

1. 統計グラフの種類、
2. 日本人は相関図がお好き、
3. 相関図のはじまり

で構成されている。

相関関係と因果関係の区別を知ることが、統計教育の視点からも、一般市民が身につけるべき重要な要素となっている(Utts(2003))。一般には、本書で取り扱われているような観察研究データ(非実験データ)から、安直な因果関係への直接的言及は控えるべきであるが、これは単に因果関係へ考察を諦めるという意味ではない。星野(2009)でも詳しく議論されているように、一定の背景知識に基づく層別などのデータ調整の工夫により、因果関係への考察は可能となる。

本書は、観察研究や公的統計等のデータが示す散布図上の相関や連関関係から、因果関係への考

察をどのように試みればよいのかを具体的に学ぶことができる良書である。単に、相関関係と因果関係が異なることを強調するだけではなく、その違いを踏まえた上で現状で手に入る情報で因果関係にどこまで接近できるのか、そのための工夫が随所に見られる。また、代替の変数の用い方や残された課題などについても、丁寧に解説されている。

典型的な例として、第5章の第3節で取り扱われている「花粉症の発症」と「スギ人工林」の関係を紹介しよう。タイトルは、「花粉症の人が多いのはスギ林が多いからか？—スギ人工林比率と花粉症有病率—」である。本書は、各節のタイトルの付け方が明確で面白いのも特徴であるが、散布図を作成してそこから傾向を読み取る・・・という流れではなく、最初に、明らかにしたい仮説、とくに因果関係を示す仮説が提示され、そのエビデンスとしてどう散布図を構成していくのかが語られていく。

ここでは、都道府県を単位とした、日本アレルギー協会が実施した疫学調査での耳鼻咽喉科医とその家族の花粉症罹患率とスギ人工林比率が最初に用いられている。なぜ耳鼻咽喉科医とその家族の罹患率なのか？ 一般に公表される花粉症（と思っている）患者数ではいけないのか・・・など、データ選択に際して払う専門家の注意深さが窺われ、参考になる。この指標で最初に作成された散布図では、スギ人工林比率がほぼゼロである北海道と沖縄を例外として除くと、期待されるような正の相関関係は見られない。そこで簡単に仮説を結果にすり合わせて変更するのではなく、更に仮説立証に向けた考察が始まる。先ず、風による花粉の拡散を考慮すると、そもそも都道府県単位の指標を使ったことに問題があること、ヒノキの花粉飛散量や大気汚染等の他の要因も花粉症の罹患

率に関係している問題などもあげて、データ加工と推論を進めていく。都道府県単位の集計データを使用した問題については、地域ブロックに集計を変更して、再度、散布図を作成し、更に西風が基本であることについては、散布図を読み取る際に地域ブロックの位置情報を組み込んで解釈をする必要性も解説している。勿論、この内容だけで十分とはいえないため、どの部分の議論や情報が不足しているかについての言及もなされており、最終的に筆者が提示する仮説に沿った散布図の形状とあわせて見ると大変に興味深い。

上記は一例に過ぎず、読者は本書を通して、データを活用した議論の進め方、いわゆる統計的な思考の進め方を学ぶことができる。統計の活用力を付けるうえでは、いきなり多変量解析の諸手法をひろく学ぶよりは、最初に、散布図というサンプルで基本的な道具の意味をきちんと理解し、現実を活用する上での注意点や応用範囲などを勉強しておくが大変に効果的と思う。とくに、今回の学習指導要領の改訂に伴い、小中学校・高校を通して統計教育の拡充が明示されたが、そこでは統計量の計算や統計グラフ作成の技術の習得だけではなく、統計分析を通して推論を進める思考力の育成が求められている。とくに、高校1年生の共通必修の内容として、散布図・相関係数が含まれているが、その具体的な指導で悩まれている高校の先生方には、本書は是非、一読していただきたい。散布図を活用する統計的思考力やいわゆる統計的センスを養成するための格好の材料が提供されている。

勿論、本書のアプローチに対して、より詳細なマイクロデータの使用や他の統計分析法の適用なども有効になることは間違いない。しかし、「統計学」は限られた情報や時間の中で、いかに各自の判断や決定のために有益な情報を提供できるかを

追求する学問のはずである。限られた情報から背景知識を活用し、判断や決断に有益となる情報を得る工夫にも力を入れなければならない。

また、11章の「**相関からはずれたものに注目する**」で取り扱われている内容も興味深い。一般に統計では全体的なバラツキの傾向を探るために手法を駆使するが、ここで扱われているような外れ値に着目しその特徴を考えることでも有効な情報を得ることができる。この章の第1節では、日本が抱える豊かさとやせ過ぎの女性の問題を人口規模と食料自給率の散布図から分析し、相関傾向からのかい離という観点から新たな尺度を考えている。この扱いは、そのまま主成分分析の概念にも繋がるであろう。

さらに付録の章「**付録-2 日本人は相関図が好き**」も興味深い。ここでは、「データのダビンチ」とも呼ばれる米国イェール大学の統計政治学者エドワード・タフティによる各国の新聞記事における統計グラフの洗練度の調査結果が紹介されているが、日本は他国に比べ新聞に散布図が多く掲載されていることが示されている。さらに、その理由として、日本人は幼少のころから学校教育において統計グラフに親しんでいることがあげられている。ただし、この調査は、1974年から80年の世界の新聞雑誌に対するもので、当時は、戦後日本が学校教育でも企業教育でも、統計教育を非常に重要視していた頃で、統計教育研究指定校も100を超えるころではなかったかと推察する。指導要領が最近、改訂になったとはいえ、その後の20年にも渡る日本の統計教育の低迷ぶりは皮肉でもある。

本書の内容について、課題が全くないわけではない。ここで取り扱われる散布図を作成するデータは、その多くが国内外の公的統計であり、本来バラツキを考慮すべき個体（たとえば、世帯や個人）ではないことに関する注意書きがないことは

残念である。時折盛り込まれているコラムや本文の中でも、本来の個体のバラツキでないことが示唆されることがあるが、専門的な知識を持っていない者にとってはその違いを明確に区別し理解することは難しいと思われる。また、本書で使用されている統計的用語が、専門的には必ずしも正確でないことが気になる。これは、本書が専門書としての位置付けではなく、一般の読者を対象としたためであると思われる。一例をあげると、第7章のタイトルである「**男女の相関**」である。読み物としての刺激的なタイトルなのかもしれないが、実際は国別の男女別の統計指標を散布図でみてみるという意味である。具体的な節は、「7.1 身長の高い国・低い国から見える傾向」、「7-2 自殺率はその国でも男性の方が高い」、「7-3 自殺を許す雰囲気があるか?」、「7-4 平均寿命はなぜ女性の方が長いのか?」などで、7.1節では、国別の男女の平均身長の散布図、7.2節や7.3節では自殺率、7.4節では平均寿命が取り上げられている。

統計に関する関心が高まり、一般向けの統計的思考力や統計的センスを高めるための教材開発や教育方法、学習環境の提供が求められている中、本書のようなデータ活用に主眼を置いた書籍が今後も多数出版されることを期待したい。2011年度から、日本統計学会が公認する統計検定がスタートする。統計検定のサイト (<http://www.toukei-kentei.jp>) では、データに基づいて客観的に判断し、科学的に問題を解決する能力は、仕事や研究をするための21世紀型スキルとして国際社会で広く認められていると謳っている。本書を通じ統計の面白さや有益性を知る人々が増えることと並行して、統計検定などを通じ、日本で高度な統計的思考力や統計的センスを身に付けた国民が増えることを期待したい。

<参考文献>

イアン エアーズ (2010)、その数学が戦略を決める (山形 浩生
訳)、文春文庫
カイザー・ファンク (2011)、ヤバい統計学 (矢羽野 薫 翻訳)、
阪急コミュニケーションズ
スティーヴン・D・レヴィット/スティーヴン・J・ダブナー

(2007)、ヤバい経済学 [増補改訂版] (望月衛 訳)、東洋経済新
報社 (2007)
星野崇宏 (2009)、調査観察データの統計科学 - 因果推論・選択バ
イアス・データ融合、岩波書店
Utts, J. (2003). What Educated Citizens Should Know about Sta-
tistics and Probability, *American Statistician*, 57(2), 74-79.

・(わたなべ みちこ・東洋大学経済学部教授)