

# 社会常識としてのデータ解析法

北海道大学 情報基盤センター  
先端データ科学研究室 水田 正弘

## 1. はじめに

私は、毎朝、7時40分に自宅を出発し、勤務先に向かいます。「だいたい」8時10分ころ大学に到着します。大切な会議があるときは、もう少し早く出発します。「たまに」遅刻します。しめ鯖は大好きですが、食べません。「しばしば」体調が悪くなるからです。駅の自動販売機で1040円の切符を買うとき、10000円札を入れるのに躊躇しません。でも、海外に行った時、自動販売機は使いたくありません。「ときどき」おつりが出てこないからです。

私たちは、小学校から算数や数学の勉強をしてきました。論理学の授業を受講した人も多と思います。そこで

もし、○○○なら、△△△である。(IF THEN)

を習いました。1040円の切符を買うのに10000円札を自動販売機に入れるなら、8960円のお釣りができます。これは、IF THENで表現でき、正しいと考えられます。

しかし、生活の多くの場面では、IF THENが成り立ちません。例えば、通勤にかかる時間は、運次第で変化します。でも、多くの場合、遅刻しません。言い換えると「データ解析法」を直感的に実践しているのです。ただ、少し込み入った問題に対しては、直感がうまく働かないことがあります。

「データ解析法」や「統計学」は、高度な数学を使い、帰無仮説とか正規分布などという用語を使った専門家集団の学問であり、普通の生活をしている健全な人は知る必要がないもの・・・、大学入試にもほとんど出題されないもの・・・、というイメージをお持ちかもしれません。

しかし、「データ解析法」のセンスは、普通の生活をする上で、身近な社会常識です。今回は、いくつかの観点から、「データ解析法」について考えてみたいと思います。

## 2. 新聞記事から

はじめに、やさしい問題から扱ってみましょう。2008年10月6日の北海道新聞(夕刊)の「今日の話」に「長者になる法」という記事がありました。みずほ銀行では、1000万円以上の当選金を受け取りに来た高額当選者について「宝くじ長者白書」を発行しています。同記事では、「1回の購入枚数では、「十枚」買った人が全体の13.2%を占め・・・「百枚以上」は、7.1%で、多ければ当るものでもないらしい」とありました。ここから、

宝くじは十枚買ったほうが、百枚買うよりも 1.8 倍ほど当たりやすいことになります。この結論に反論できますか？

読売新聞に「社会保障維持に消費税上げ「やむなし」50%・・・本社世論調査「容認せず」を上回る」と掲載されたそうです([http://job.yomiuri.co.jp/news/jo\\_ne\\_07112102.cfm](http://job.yomiuri.co.jp/news/jo_ne_07112102.cfm)、2007年11月21日)。さらに、「同じ質問をした昨年10月調査では・・・49%だった。」とあります。この記事から2つの課題がおきます。1つは、「やむなし」が50%であったということをどのように理解するかです。もう1つは、49%が50%になったことに意味があるかどうかです。

## 2. PISA にみる社会常識としてのデータ解析法

PISA とは、OECD による国際的な学習到達度調査で、2000 年に第 1 回調査が実施されました。第 2 回調査は 2003 年、第 3 回調査は 2006 年に行われました。調査内容は、「知識や技能等を実生活の様々な場面で直面する課題にどの程度活用できるかを評価」するもので記述式が中心です。2000 年調査では、日本は 32 カ国中、数学は 1 位、読解力は 8 位、科学は 2 位でした。それに対し、2003 年調査では、41 カ国中、数学は 6 位になり、さらに、2006 年調査では、56 カ国中、数学的リテラシーは 10 位になりました。この解釈については「データ解析法」の観点からも多くの課題がありますが、触れないことにします。

例えば、次のような問題があります。

### 盗難事件に関する問題

ある TV レポーターがこのグラフを示して、「1999 年は 1998 年に比べて、盗難事件が激増しています」と言いました。【水田注：1998 年は 508 件、1999 年は 516 件と読み取れる棒グラフがある。ただし、縦軸のメモリは下部が切られ 505 から始まっている】

このレポーターの発言は、このグラフの説明として適切ですか。適切である、適切でない理由を説明してください。

[http://www.mext.go.jp/a\\_menu/shotou/gakuryoku/siryo/05020801.htm](http://www.mext.go.jp/a_menu/shotou/gakuryoku/siryo/05020801.htm)

この問題の「完全正答」は、「グラフのごく一部が示されているにすぎないので、適切ではない」ことを指摘することです。「データ解析法」における非常に初歩的な問題です。つまり、得られた情報を的確に把握し、それを使って判断する能力が問われています。

その他、PISA では、「二つの病棟における分娩 100 件当たりの、産褥熱による死者数(1841-1846)」の線グラフを示し、産褥熱の原因を説明させる問題や、「北アフリカのサハラ砂漠にあるチャド湖の水位変化」のグラフなどを示し、それを読解させる問題など、多数のデータ解析法に関係した問題が出題されています。つまり、「知識や技能等を実生

活の様々な場面で直面する課題にどの程度活用できるかを評価」するためにデータ解析法のセンスを評価することが重要とされているわけです。

### 3. 全国学力調査の出題におけるデータ解析法

平成19年度から、全国的に子ども達の学力状況を把握する「全国学力・学習状況調査」が実施されています。これは、小学校第6学年、中学校第3学年の原則として全児童生徒を対象として、教科に関する調査（国語、算数・数学）および生活習慣や学習環境等に関する質問紙調査をしています。特に教科に関する調査では、主として「知識」に関する調査と、主として「活用」に関する調査の2種類の問題を利用しています。

例えば、平成20年4月22日に実施された調査の小学校算数B（活用）は、5問から構成されています。そのうち、問2は、農業生産額の棒グラフと円グラフを示して、それに関する問題、問5は、身長に関する線グラフや棒グラフを示し、その会社に関する問題です。小学校国語A（知識）、B（活用）も含め、多くの統計グラフやデータ解析法に関係する設問があります。全ての問題はWebページから入手できます。

<http://www.nier.go.jp/08tyousa/08tyousa.htm>

### 4. 新しい学習指導要領から－「生きる力」

本年に入り小学校学習指導要領および中学校学習指導要領が公示されました。現行の学習指導要領も、新学習指導要領も重要なキーワードとして「生きる力」が強調されています。21世紀を生きる子どもたちの教育の充実はみんなの願いです。

新学習指導要領において、今回のお話に関係する重要な事項に「資料の活用」があります。例えば、中学校数学科の改善の具体的事項として以下のような記述があります。

中学校においては、数学的活動に主体的に取り組み、基礎的・基本的な知識・技能を習得し、数学的に考える力をはぐくむとともに、数学のよさを知り、数学が生活に役立つことや数学と科学技術との関係などについての理解を深め、事象を数理的に考察する能力と態度を養うことを重視して、次のような改善を図る。

(ア) 領域構成については、現行の「数と式」、「図形」、「数量関係」の3領域から、確率・統計に関する領域「資料の活用」を新設するとともに、「数量関係」を「関数」と改め、「数と式」、「図形」、「関数」、「資料の活用」の4領域とする。

(キ) 「資料の活用」の領域では、資料に基づいて集団の傾向や特徴をとらえ、それをもとに判断することを重視する。例えば、従来から指導している確率に加え、ヒストグラムや代表値を用いて全体の傾向をとらえたり、標本を取り出して調べることで母集団の傾向をとらえたりすることを指導する。

「数学が生活に役立つ」ものとして、データ解析法は強力な方法だといえます。その言い換えである「資料の活用」の領域が義務教育の学習指導要領で明示されたことにより、データ解析法が社会常識になっていくと期待しております。

## 5. おわりに

データ解析法の考え方やセンスが社会生活で役に立つ例をいくつか紹介してきました。データ解析法の理論や応用方法を学んでいくことにより、専門的な能力が得られるだけでなく、普通の生活においても役に立つ社会常識が身につくと思います。合理的な判断をするためには、統計に騙されないことが大切です。

以下に、本稿の参考書として啓発書を中心に紹介しておきます。

## 参考文献

田栗正章・藤越康祝・柳井晴夫・C.R.ラオ(2007), やさしい統計入門－視聴率調査から多変量解析まで－, BLUE BACKS, 講談社.

田村 秀(2006), データの罨－世論はこうしてつくられる－, 集英社新書.

ダレル・ハフ著／高木秀玄訳(1968), 統計でウソとつく法－数式を使わない統計学入門－, BLUE BACKS, 講談社.

仲本秀四郎(1993), 情報を考える, 丸善ライブラリー.

福井幸男(1997), 知の統計学2－ケインズからナイチンゲール, 森鷗外まで－, 共立出版.

水田正弘・山本義郎・南 弘征・田澤 司(2005), S-PLUSによるデータマイニング入門, 森北出版.