

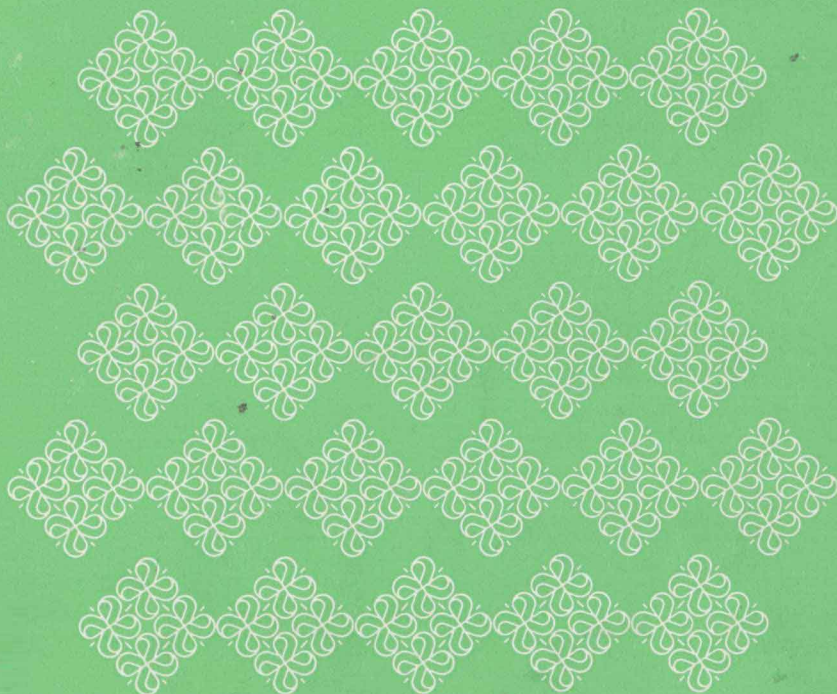


編集 / 田島一郎・島田 茂

教職数学シリーズ 基礎編④

教養統計学

鈴木義一郎 著



共立出版株式会社

編集 / 田島一郎・島田 茂
教職数学シリーズ 基礎編④

教養統計学

鈴木義一郎 著



共立出版株式会社

著者略歴

鈴木義一郎 (すずき・ぎいちろう)

生年月日 1937年9月19日
最終学歴 1962年東北大学理学部数学科修士課程修了
専攻科目 統計学
学位 理学博士
現職 統計数理研究所第三研究部第二研究室長
主要著書 『例解統計入門』(実教出版)
『統計解析術——データを活かして使う法——』(実教出版)
『モデル解析術——OR的考え方の基本原理——』(実教出版)
『データ解析術——“記述統計”のすすめ——』(実教出版)
『基本統計学』(共著)(実教出版)
『統計術指南』(日本経済新聞社——日経新書)
『統計解析法の原理』(朝倉書店)
『「比較」統計学のすすめ』(講談社ブルーバックス)
『統計いろはがた』(東洋経済新報社)
『「勝負」の統計学』(講談社ブルーバックス)

教職数学シリーズ 基礎編 4

教養統計学

定価 2200 円

1982年9月1日 初版1刷発行

検印廃止

NDC 417

ISBN 4-320-01310-7

著者 鈴木義一郎 © 1982

発行 共立出版株式会社 / 南條正男

東京都文京区小日向 4-6-19

電話 東京 (03) 947 局 2511 番 (代表)

郵便番号 112 / 振替口座 東京 1-57035 番

印刷 新日本印刷

製本 中條製本



社団法人
自然科学書協会
会員

Printed in Japan

刊行のこ と ば

豊かな人間性の育成を目ざして、教育全般の改善がいろいろと進められているとき、数学教育もその一環として重要な責任を負わなくてはなりません。この意味において、これからの数学教員が、人間活動の成果としての数学に深い認識をもち、また、生徒の心情や学習過程に暖かい理解をもつということは、絶対に必要であります。

最近の社会情勢の変化にともない、数学教員を志望する学生が急増しているのは誠に喜ばしいことではありますが、上記のような必要性に応えるための基礎的な教職数学の書物は、残念ながら乏しいように思われます。すなわち、数学の専門書はおびただしくあり、数学教育の専門書も数多くありますが、数学と数学教育との関連をふまえたもの、とくに数学教育の背景にあることを意識して書かれた数学の専門書となると殆どないといってもよいでしょう。

このような状況にかんがみ、まったく新たな構想のもとに、

○ 教育学部や理工系学部で中学・高校の数学科教職課程を履修する学生のためのモデル・テキスト

○ 現職数学教員の日常研修のためのリファレンス・ブック

となることを目ざして本シリーズを企画いたしました。前にも述べたように、数学教員としては少なくとも次の二つの能力が不可欠であります。

I. 教員としてもつべき数学そのものの能力

II. 効果的な授業を展開するための指導能力

本シリーズは基礎編と実践編とに分けてありますが，前者は主としてⅠを，後者は主としてⅡを目標とし，この両者を合わせてⅠ，Ⅱの能力の調和のとれた伸長を期しています。

本シリーズがこれからの数学教育の発展に大いに寄与することを願って止みません。

1981 年 10 月

田 島 一 郎

島 田 茂

は じ め に

世の中には、不完全な知識にもとづく判断で行動をとらなければならないことが数多くあるから、大なり小なりの“失敗することによる危険”をはらんでいる。たとえば、飛行機を利用すれば目的地に早く到達できるかわりに、墜落でもして死を余儀なくされるかもしれないといった危険が伴う。神経質な人なら、そんな危険を犯すのはいやだからと陸路を選ぶかもしれない。またある人は、なァに人間一度は死ぬものよ、くよくよしていたら文明の利器は何も使えなくなると飛行機を多用する。

このように危険の程度は同じでも、それを無視しうる程度のものとみなすかどうかは、人によってまちまちである。多少の危険を覚悟のうえでとる行動に、何らかの客観的な目安を与えてやろうとするときに、確率とか統計といった概念が必要になってくる。つまり統計の役割とは、不確実な状況下での判断に、有効適切な情報を提供してやることにある。ただ、そのような情報を使っ

ての判断には、必ずしも一通りの答えがでるとは限らない。

統計を利用する際に生じるこの種の“あいまいさ”が、すっきり好みの数学的思考法とは、ひと味違ったニュアンスを与えるところとなる。これが数学担当教師などに、どうも統計を教えるのは苦手であるといった印象を与えてしまう。ところで、粉れのない数値解が有無をいわず説得力をもつといったケースが、この世はどのくらいあるだろう。偏差値にしても物価指数などにしても、非常に限られた側面を認識する際に有用なだけである。つまり、現象認識

のほうが先にあって、計測値のようなものはその便法として集められているにすぎない。結局、現象認識のための一つの補助手段として、数学、特に数式運用法を学ぶ目的がある。数学担当の教師は、篤とこの点をわきまえて数学教育の任にあたらなければならない。要するに、数字バカでは数学教師としての資質に欠けるということである。

さてこの書では、数学とはひと味違う統計学の基本的な側面を、いろいろな角度から紹介する。記述法を、従来類書とは大幅に変えてみた。100冊は優に超すであろうと思われる統計学の入門書の中に、似たような一冊を付け加えてみても意味がないと考えたからである。各章はほとんど独立しているので、どこから読みはじめても構わない。またすべてを読破できなくても、それなりの統計的素養が身につけられるよう配慮してある。忙しい読者もおられようから、とりあえず各人の好みの章から目を通して頂きたい。

序章では、統計のイメージと題して、統計的な考え方についての概観を与える。1章では、これまでの筆者の体験をもとに、統計を教育する際の勘どころを被露するとともに、統計教育のあり方に関する筆者なりの見解のようなものを論述した。

2章では、統計学にでてくる基本的な概念を12個選んで、身近な事例をまじえながら解説を行った。不必要な数式を極力排除したのは、ほどほどに数式を入れると数式の展開部分を理解しただけでその概念を判ったと錯覚しがちになるからである。この章だけでも熟読玩味されれば、統計的な考え方の基本をマスターすることができるだろう。

3章は、この書のハイライトというべき部分である。身近に手に入るデータを用いて、データ記述の基本的なテクニック、さらには整理された結果をみてデータの背景をどう解釈していくか、など統計教育の実践面で参考となるであろう事例をいろいろと扱ってみた。なおここで使われている手法は、加・減・乗・除の四則演算だけで、統計的解釈には数学的素養が必ずしも必要ではないといった事情も理解頂けたらと思う。

逆に数学的な記号や式が入っていないと満足できないといった読者のためには、次の4章がうってつけである。確率分布という概念が、近代の推測統計学の分野では必須知識となる。数式にそう弱くない読者は、教養としてこの程度の範囲まで身につけておかれば十分であろう。

最後の5章で扱った統計の歴史、これもほどほどには教養として身につけておかれると、教育の仕方にも一段と幅がつくこと請合いである。ここでの叙述には、歴史の専門家にいわせると厳密さに欠けると指摘されるかもしれないが、要は統計的方法の発展のプロセスを大雑把にでも認識されることで、大抵は十分であると思われる。なお巻末の付表として、身近に得られて、しかも統計教育を実際に行う際に有用となるようなデータ類を掲載しておいた。3章に記した分析事例などを参考にされて、読者がせいぜいご利用頂けたらと思う。

以上概観したように、この書は統計を実際に教育しなければならない立場の読者を想定してまとめあげている。しかし書名が示すように、統計学に関する基礎知識というよりは基本的な考え方を、教養として身につけたいと考えられている読者のために書き下ろしたものである。特に独習書として利用できるよう意を用いてあるので、一般社会人のための、やや風変りな入門書としても十分お役に立てるのではないかと自負している。ともあれこの種の本を、世に出す機会を与えてくれた共立出版編集部の英断に敬意を表したい。

1982 年 7 月

鈴木義一郎

目 次

序章 統計のイメージ

1 章 統 計 と 教 育

- 1.1 なぜ「統計学」は面白くないのか 11
- 1.2 人間が「りこう」になる法 19
- 1.3 急がれる統計教育の充実 23
- 1.4 指導要領によらざる統計教育のすすめ（その1）
——平均に関する不等式をめぐって 25
- 1.5 指導要領によらざる統計教育のすすめ（その2）
——2項分布をめぐって 31

2 章 統 計 語 典

- 2.1 平 均 41
- 2.2 デ ー タ 45
- 2.3 確 率 48

x	目	次
2.4	分	布 51
2.5	独	立 54
2.6	推	定 57
2.7	誤	差 61
2.8	検	定 64
2.9	比	較 68
2.10	判	別 71
2.11	予	測 75
2.12	モ	デ ル 78

3 章 データの語る規則性

3.1	文章の長さを調べる——データの整理と平均，分散の求め方——	81
3.2	高校野球とプロ野球——相関関係の捉え方——	90
3.3	百貨店の売上高——移動平均と季節変動の捉え方——	104
3.4	大相撲——勝星数の分布パターン——	113
3.5	男 と 女	122
3.6	野球における得点経過	128

4 章 確率分布の概念

4.1	確率変数の概念	133
4.2	確率変数の期待値と分散と積率母関数	136
4.3	2次元の確率変数	141
4.4	確率変数の独立性	147
4.5	多次元分布	150
4.6	標 本 分 布	153

5 章 統 計 の 歴 史

5.1	統計学の誕生	157
5.2	18 世紀の統計学	159
5.3	19 世紀の統計史概観	163
5.4	マルサスの人口論	164
5.5	社会統計学の展開	165
5.6	生物統計学の誕生	168
5.7	統計力学の誕生	170
5.8	ガウスの誤差論	170
5.9	20 世紀（前半）の統計史概観	171
5.10	数理統計学の誕生	173
付 表		177
あ と が き		181
索 引		185

序章 統計のイメージ

「統計」についてのイメージ

統計と聞くと、国や地方公共団体などによって集計された膨大な数字の羅列を思いかけて、一般庶民にとってはあまり縁なきものと判断される方もいるであろう。またある人は、いかめしそうな数式を運用して理屈をこねまわす学問が統計で、数学アレルギー患者などはなまじ近よらぬほうが賢明とばかりに、頭ごなしに敬遠してしまうかもしれない。そのような判断の大方は正しいと思う。ただそのような“いわゆる統計”は、肝心な判断の場面では存外効いてない代物なのである。

ここで、判断に直接効いているものこそ“ほんとうの統計”と呼ぶべきである、などと主張するつもりは毛頭ない。ただこの本の、気に入りの話題がきっかけになって、“いかめしそうな統計が有用である”といった誤れるイメージを払拭していただけるなら、望外の幸せといえよう。

まずは“使いで”のあるものなら、ことさら使えなどといわずとも、誰もがそれを使っているという話から、おかみさんが朝げに味噌汁でも作ろうと、鍋の中に味噌を入れ、よくかき混ぜてからおたまですくって塩梅をみる。これは鍋全体の味噌汁の味と、おたまですくった一部分の味とが一致することを期待しての行為にはかならない。いわば**抜取検査**の一種である。しかもよくかき混ぜてからすくったということは、**ランダム・サンプル**を得たことに対応している。つまり、おかみさんも立派に統計を使っている、ということになる。この

場合、何気なく無意識に使っていることと、特別に数量化を試みてはいないことが特徴的である。

この種の例に限ったことではないが、意識せずとも使えるということは習熟した使い方ができるということである。習熟した使い方をするから、事はスムーズに運ぶのである。無意識のうちに使われた事柄は、当人はもちろん周囲の人にも気づかれることが少ないから、概して目立たない。やたら〇〇的といった形容句が付されるケースほど、その〇〇は意識的に使われている、したがってその〇〇は存外効いていないということになる。この〇〇の中に、数学、統計、理論といったような言葉を挿入してみれば、なるほどと感じられる読者も多かろう。

特別に数量化を試みない、これまた判断に直接効いているものについての基本条件なのである。物事の是非を論ずる際、数量化して紛れを一掃できるならそのほうがよさそうに思える。だがそうは問屋が卸してくれない。どのようにして数量化すべきかという段階では、無数の紛れが生じていたはずである。そのいい例が、昨今さかんに論争されている“偏差値”問題。業者テストに基づく偏差値が、有名高校への入学可能性を的確に判別できる優秀な統計値であるのに、その是非論が云々されるのはなぜか。偏差値の高い生徒は有名高校から一流大学へ、そして世のエリートへと進出する可能性が約束される。人間、エリートだけが一流ではないとの主張者から、偏差値の「非」が強弁されることになる。だが、利用できるものならどんなものでも使ったほうがよい、と筆者なら考える。偏差値を使わせないようにしたければ、偏差値を無能力者にしてしまうのがよい。それには、少なくとも現行の入学試験制度を変えていく必要がある。試験などやらずに、誰でも好きな学校へ入れるようにしてしまうのも一法だろう。人間に格づけを与える、それも万人の納得するような数値などは、あるはずもないからである。

要するに、表面的な数値に振り回されることなく、何気なく使っている事柄が、判断の「決め手」になっているということである。これを“ほんとうの統計”と呼んでいいようにも思うし、他の適切な言葉で代用したほうがよいとい

うのなら、それでも結構。統計という言葉に限らず、この世にはいろんな言葉が氾濫している。それらの各言葉についてのイメージが、各人各様であって一向にさしつかえはない。だが判断に直接効いているものという概念については、皆がもっと共通のイメージをもつ必要があるのではなからうか。

データについてのイメージ

さて統計的判断の大元は何かと考えてみると、それはデータということになる。したがって、統計がうまく使われるための第一義的なことは、良い素材データが利用できるかどうかである。利用対象となるデータを大きく分けてみると、**既存データ**とあらたに収集すべき**追加データ**とが考えられる。

既存データに関して問題となるのは、利用可能なデータの中から何を用いるべきかを選択することである。さらに、それが十分使用に耐え得る素材なのかどうかを点検することも忘れてはならない。

追加データを収集する場合には、何をどれだけ追加することが必要かつ十分であるかを見定めることである。その際に留意すべき点は、そのデータがいかなる判断に供すべきものであるかを問い続けること、つまり**当該データの利用目的を明確に把握しておく**ことである。また、データ収集の号令さえかければ、後はデータが自動的に集積されるものと安直に考えてもいけない。データ収集の機構を、常によく管理された状態にしておくよう留意することが肝心なのである。

次に、収集されたデータを巧く分析するにはどうするか。われわれが誠心誠意努力してみても、良質のデータがどうしても得られないという場合が多々ある。そのような場合、当該データの“分”をよくわきまえておくことが肝心である。データの不足分を補い得るのは、統計手法ではなくして担当者の常識的な考え方である。つまり、データからこうなったという表現の代わりに、データをもとに私はこう考えるという表現ができるようにならなければならない。このように、データは人に考える素材を提供するだけだから、データにあまり深入りしすぎるのは禁物である。データを介して人が語る、これこそ正しくデ

ータを分析する場合の基本的態度といえよう。

この世に出まわっているデータで、すべてを信ずるに足ると期待できるものは皆無である。データの分をはるか越えた結論を導出せんと息まくのは、錬金術を夢見るに等しい。ピンボケのネガから鮮明な写真がプリントできるはずもないから。もっとも、データはことごとく疑ってながめるべきものとの境地に達すれば、いかなる類のデータからでもそれなりの情報は汲みとれることにもなる。データがあればコタエがある、これが大方の人のデータについて抱いている、誤ったイメージではなからうか。

平均についてのイメージ

統計とは「平均」の学であるといった表現もなされているくらい、平均は統計的方法の中心的概念である。データ全体の情報を端的に縮約したものが**算術平均**なら、**標準偏差**はデータの散らばり具合の平均値。2次元データの平均的直線が**回帰直線**なら、直線的傾向の平均値が**相関係数**ということになる。

ところで、平均という概念は実に親しみやすい。データの総和をデータ数で割るだけで求められるから、小学生の子供にだって「平均ぐらいいは計算できろあ」ということになる。しかしことほどさように、その意味合いを正しく理解して使われているかどうかは、疑問なしとしない。平均賃金1割アップといっても、自分の取り分についてそうなるとの保証はない。すべての母親が、己が赤ん坊を平均体重以上にしようと躍起になっている。零歳時における平均余命が、成人の場合についてもそのままあてはまるものと錯覚している人が存外多い。

そもそも、平均は何のために求める必要があるのだろう。データ全体の中から、標準的目安になる“情報”が欲しいからではなからうか。まず判断の足しにできそうなデータが与えられる。生のデータのままであれば繁雑にすぎて判断の便に供しにくい。そこでデータをひとまとめにして表現したい、ということになる。元のデータを判断しやすいところまで変形した“縮約値”，その一つとして（算術）平均というものが考えられる。前期は100点だったが後期には50

点しかとれなかった生徒と、二回とも 75 点をとった生徒とでは、どちらが“できる”のか。どちらの生徒の平均点も 75 点であるから、両者の異なるパターンを平均値だけで考えるのはちと問題である。データのもつ情報をすべて吸収しつくしている縮約値など、一般には存在し得ないから、平均値もしょせん「試み」の値でしかない。平均についての異常すぎる信頼感、データとみれば何でも信じこんだり、〇〇的に導出された表現を鵜呑みにしたりする風潮と「根」を同じくするものではなかろうか。

グラフ化についてのイメージ

“目は口ほどにものを言い”といった表現もあるように、データから情報を吸収するのに視覚を援用させると意外に効果的なことが多い。否、効果的であるからこそ、使い方を間違うととんでもない誤情報を提供することにもなりかねない。以下において、データをグラフ化するときを生ずるいくつかの問題点を列挙してみよう。

まずはグラフ化の“いろは”に属する基本問題。図 1 の左側のほうの図は、

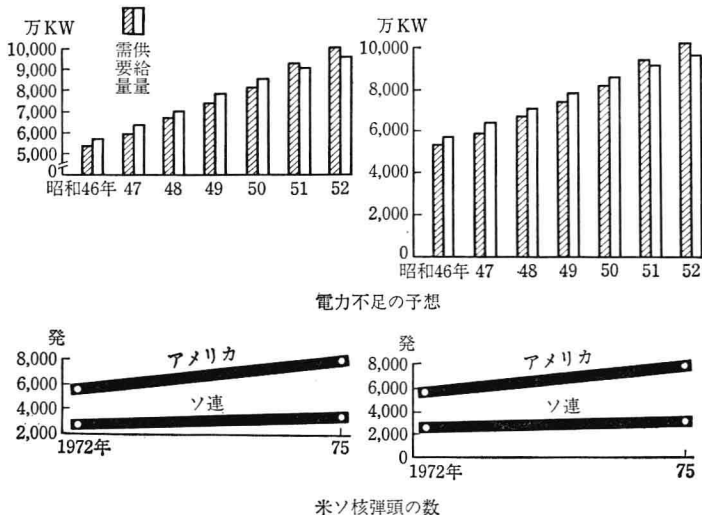


図 1

某一流新聞の記事としてとりあげられたグラフである。不注意からか、それとも意図的なものかはわからないが、前者では途中の目盛の部分が短縮されているし、後者では基線が欠如している。そのような省略をしないで描いてみた図が右側である。このデータの例に関する限り、そのような省略はさほどこスペースの節約にもなっていない。左側の図だけを見ていると、電気の需要量と供給量の“のび”が顕著で将来の電力不足が深刻であること、さらにアメリカの核弾頭数がソ連の場合をかなり上まわっていることが、実際以上に増幅されてグラフ化されていることがわかる。この種の誤用例を数え上げていたら際限がないので、われわれがグラフ化されたデータからの情報を汲みとるときには

グラフはデータと目盛の関数である

といった事情を銘記しておくことが肝要となる。

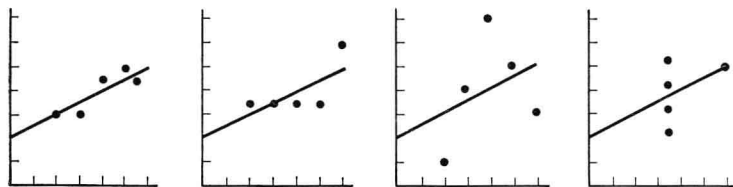


図 2

次はデータの散布図に直線をあてはめるときに心掛けておかなければならない基礎知識。回帰直線とは、2次元データの平均的直線であることは前に触れた。つまり、データの各点から直線までの偏差（の平方値）の総和を最小にするものとして求めることができる。その導出法に関する手続きについては省略して、図2に4種類のデータに対する回帰直線が描かれている。この図をみて気づくことは、4種のデータが明らかに異なるパターンをしているのに、回帰直線は全く同一である。このような事実から学ぶべき教訓は

回帰直線を求めることが最終解とは限らない

ことを認識する必要がある。つまり“回帰分析”といったような手法がうまく