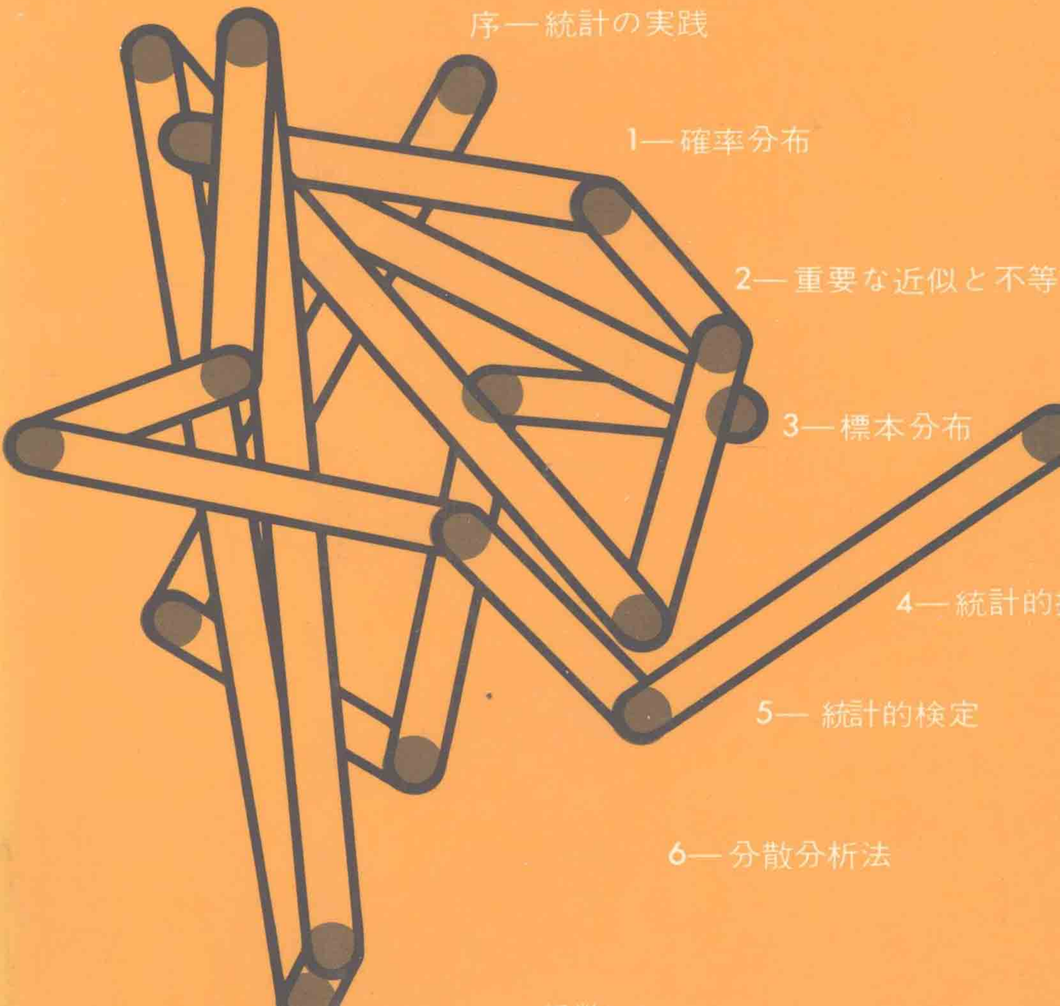
初等統計解析

牧野都治　伊藤正義　道家暎幸　共著

森北出版株式会社



初等統計解析

牧 野 都 治
伊 藤 正 義
道 家 暎 幸
共著

森北出版株式会社

著者略歴

牧野　都治
　1923年　東京生まれ
　1953年　東京理科大学卒業
　専　攻　OR
　現　在　茨城大学教授，理学博士
伊藤　正義
　1935年　北海道生まれ
　1960年　東京理科大学数学科卒業
　現　在　北海道工業大学助教授
道家　暎幸
　1944年　東京生まれ
　1973年　日本大学大学院理工学研究科数学専攻修了
　現　在　九州東海大学経営管理学科専任講師

初等統計解析

1977年4月20日　第1版第1刷発行
1980年4月10日　第1版第4刷発行

定価はカバーに表示してあります．

著者　牧野都治（まきのとし）
　　　伊藤正義（いとうまさよし）
　　　道家暎幸（どうけひでゆき）

著者との協議により検印は廃止します．

発行者　森北　肇
印刷者　田中春吉

発行所　森北出版株式会社
東京都千代田区富士見 1-4-11
電話 東京 (265) 8341 (代表)
振替 東京 1-34757 郵便番号 102

日本書籍出版協会・自然科学書協会・工学書協会　会員

落丁・乱丁本はお取替えいたします
印刷　壮光舎印刷／製本　正明社

1041-0912-8409

Printed in Japan

ま え が き

「統計は使って生かすものである」——これが統計学に対するわれわれの信条である.

多くの場合, 始めにまず解決しなければならない問題がある. そこで統計調査を企画し, 実施に移し, 得られた資料を集計, 分析する. さらにこのような過程を経て導かれた結果を解釈し, 評価する. これが,「統計を使って生かす」ための標準的な流れである.

統計学では——ことに資料の分析の段階においては, 数学の定理や公式を駆使して得られる多くの知識を活用する. しかし統計の中で, 資料の分析の占める役割りは, 多くの場合, その前の段階である調査の企画, 調査技術, そして後の段階の結果の解釈を凌駕するものではあり得ない. したがって統計学は, 数学と密接なかかわりをもつとはいうものの, 決して数学の1分野にとどまるものではないというのが正しい見方であると思う.

しかし, これはあくまでもたてまえであって, 実際に大学で, とくに低学年の学生に対して講義するとなると別である. われわれの接する多くの学生諸君は, ドロくさい現実の問題にはあまり関心を示さないで, 理論の展開に興味をもつようである. そのようなわけで, 低学年ではやはり, 資料の分析に重点をおいて話を進めることにしている.

ところで近来, 統計学の重要性に対する認識が一段とたかまり, それにともない多数の教科書が出版されている. しかし, あるものは使い方だけに重点をおき, なぜそうすればよいかにふれていなかったり, またあるものは厳密に論証を試みるあまり, 初心者には到底理解しがたいものになっていたり, 時に理論と実際をうまくかみあわせながらときほぐしているものがあっても, ページが少々かさんでしまっていて, 1年間ではとてもこなせそうにない等々, それらはそれなりに良書であるとはいうものの, 教科書として採択しようとなると,

ためらわざるを得ない.

そこで，出版社からのおすすめもあり，理論に飛躍がなく，しかも1年間で十分使いこなすことのできる本を書こうということになった.

この本は，高校での教育課程を履習した人が，さらに進んで,「統計を正しく使いこなす」ために，これだけは知っておいていただきたいという，いわゆるミニマル・エッセンシャルをモットーに組み立てた統計学の教科書である.

この本は序章および第1章～7章までの構成になっている.

序章はデータ解析を通して，ふだん著者らが考えている2，3の事柄について述べたものであるが，通読していただくだけで，著者らの考えをお汲みとりいただけるものと思う．本論は第1章～7章であって，これらの章末には，精選した問題をつけておいた．通年4単位の授業であれば，これらを一通りこなすことができよう．そのうち第1章～3章では確率分布について述べ，第4章～7章が統計解析になっている．もし，通年2単位の授業として統計を扱うのであれば，第4章～7章に重点をおくのがよいと思う.

さらに，学生諸君の実力をたかめるために，補充問題を付しておいた．積極的にとりくまれることを期待している.

この本は以上の趣旨にもとづき，本文を牧野，問題と略解を伊藤・道家が分担して執筆したのであるが,その間十分に意見を交換しあい遺漏なきを期した.したがってもし不備な点があれば，それはひとしく3名の責任である.

統計教育に携わっていらっしゃる諸賢の御批判と，この本によって学ばれる方々から御意見をお寄せいただければ幸である.

なおこの機会に，本書を刊行するにあたっていろいろお世話くださった森北出版(株)企画部長の柳澤茂八氏に対し,深甚な感謝の気持ちを表したいと思う.

昭和52年3月12日

著　者

目　　次

序章　統 計 の 実 践

§ 1. 統計調査とデータからの読み

1.1　クセをつかむ

統計では，多くの資料を調査し，将来の推測に役立てているが，そのためには，統計資料は信頼度の高いものであることが大切である．しかしそのような資料を入手するには，えてして経費がかさむことを覚悟してかからなければならない．ごく他愛のないデータでも，ちょっとまじめにとろうとすると，たちまち40～50万円とんでしまう．そのわりに，苦労して入手した情報が，たいしてとるに足らない代物であったりするのは，ごくありきたりの話である．それにもめげず，大まじめにデータをとる．こうして何度か浪費してみて，はじめて統計学の入門書などの冒頭に

「できることなら，本調査を行う前に予備調査をおやりなさい」

と記述されていることのもつ意味の深さが，実感をもってわかってくる．

さて開き直って

「統計調査とは何か」

と問いかけてみると，理論研究家の人たちは，とかく口ごもりながら

「統計学には理論と実践の分野があって，統計調査はまさにその実践面を…」

などから始まり，用意周到なお題目を唱えたりなさるようであるが，実務家のセンスでは，ただ一言，“それはクセをつかむことである”と言い切るかもしれない．

クセがいけなければ，傾向とか特徴といってもよいが，体を張ってデータ集めをしてみると，やはり傾向ではなくて，クセのほうがぴったりくる．

クセをつかむために，いきなり大がかりな本調査をということもある．しかし，できることなら予備調査で大まかなクセをつかんでおく．本調査では，もっとこまかいクセまで見いだそうと努めることになるが，大きなクセをつかん

でおくというのが，安上りの秘訣でもある．

以下，デパートの客を例にとって，このへんの事情を詳述しよう．

1.2　デパートの客

東京の銀座通りに，松屋・三越・松坂屋がある．われわれは，それらの店での客の動向調査，たとえば平均滞留人数とか平均滞留時間などを知るための調査を，しばしば行なってきたが，ここではその１つのM店だけに焦点をしぼって話をすすめることにしよう．

M店には，１階に３つの出入口，地下１階に２つの出入口がある．それらを順にA，B，C，D，E出入口とよぶことにする．

店内に滞留している客の平均人数 L，平均滞留時間 W を知るために，開店時から閉店時まで，たとえば10分キザミに客の到着数および退去数をカウントして，図１のような累積図表が書かれたとする．

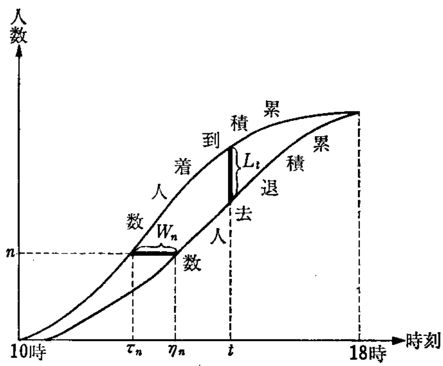


図１　累積人数のグラフ

時刻 t における滞留人数を知りたければ，図１の横軸のところで縦軸 t に平行線をひき，２つの曲線（ほんとうは階段状になるわけであるが）にはさまれた線分の長さ L_t を読みとればよい．

このような線分の長さをいくつも測って，それらを平均すれば，その日の平均滞留人数 L が求まるが，むしろ次のようにして L を求めるとよい．

いま，２つの曲線で囲まれた部分の面積を S，営業時間（図１では８時間）

をTとすると

$$L=\frac{S}{T} \tag{1}$$

また，図 1 を用いて客の平均滞留時間Wを求めることもできる．

n番目に到着した客の到着時刻を τ_n，n番目に退去した客の退去時刻を η_n とすると，一般には η_n はn番目に到着した客の退去時刻にはならないが，平均滞留時間Wを求めるには，その客の滞留時間があたかも W_n であったかのようにみなして，計算してよい．

したがって，1 時間あたりの平均到着人数をλとすると，Wは

$$W=\frac{S}{\lambda T} \tag{2}$$

によって求めることができる．

また，式 (*1*)，(2) からつぎの関係式が成り立つこともわかる．

$$L=\lambda\cdot W \tag{3}$$

平均滞留人数

平均滞留時間

平均到着人数

そういうわけで，図 1 が書ければ，L や W がとにかくすんなり求められることになる．しかしそれには，開店時刻から閉店まで，M 店の 5 つの出入口に調査員を配して，客の入・退数をカウントさせなければならない．そこでクセを使う．

出入口にはそれぞれ特有のクセがある．

比較的たくさん入りやすい出入口とか，むしろ出ていく客が多い出入口といったクセである．

しかも，このクセが実にはっきりしていて，M 店の場合，5 つの出入口のうちの A，D 2 つだけについて，客の到着をカウントすれば，時間帯にあまり関係なく，いつも全体の出入口からの入場者数の約 69 ％を占めている．また退去数については，B，D 出入口に注目したほうがよくて，それだけで全体の 50 ％という安定した値が得られることもわかった．

このようなクセがつかめればしめたもので，以後は半数以下の調査員を配置することによって，十分活用しうるデータの入手が可能になった．

表1，図2は，このような調査にもとづく集計結果の1部である．すなわち，昭和45年11月14日，15日，17日における1日の延べ到着人数N，平均滞留人数L，平均滞留時間Wなどについては表1のようになり，滞留人数の時間的変動については図2のようになった．

表 1 集計結果の一例

項目 ＼ 年月日(昭和)	45. 11. 14 (土)	45. 11. 15 (日)	45. 11. 17 (火)
平均滞留人数 (L)	2 984	4 580	2 393
平均滞留時間(分) (W)	33	41	30
延べ到着人数 (N)	43 692	53 829	38 030

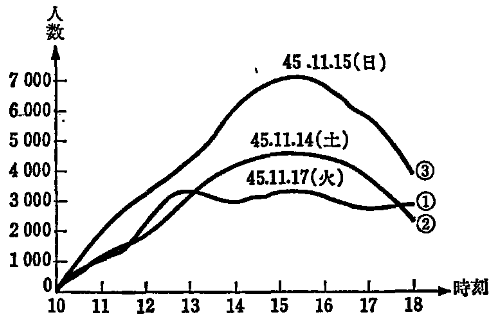


図 2 滞留人数の変動

その後，数回にわたって，調査して確信を得たことであるが，平日の滞留人数は図2の①のように2つのヤマ，土・日は②，③のような1つのヤマになるクセがある．ちなみに，図2で，ヤマの下の部分の面積から判断すると，日曜のほうが土曜よりずっと大きくなっているので，その割合だけ日曜のほうが延べ到着人数が多かったと考えがちであるが，表1によれば，実は日曜は土曜のせいぜい1.23倍であったに過ぎないことを注意しておこう．それでは，図2から，なぜそのように誤まって読まれがちなのであろうか．それは，ヤマの下

の部分の面積が延べ到着人数を表わすものではなくて，時間要素が加味されたものだということを読みおとしがちであることにほかならない．つまり，土・日の延べ到着人数がほとんど同じであったとしても，客の平均滞留時間が 1:2 であれば，ヤマの下の部分の面積も 1:2 になるという，しごく当然のことをうっかり見過ごしてしまってはいけないということである．

ふたたび

$$L=\lambda W$$

をもち出してみよう．かりに延べ到着人数が同じ（したがって単位時間あたりの到着人数も同じ）であったとすれば，平均滞留人数は平均滞留時間に比例することになる．上での見誤りは，ただそれだけの ことを勘違いしたにすぎない．

次に，$L=\lambda W$ を使った，もう一つの解釈をご紹介しよう．

1.3　結果の読み

やはり前述のデパートに関してのことであるが，こんどは歩行者天国の日のデパートへの人出についてである．この日は，車の心配なしにショッピング・ムードを楽しめるとあって，銀座への人出は毎回相当なものになっている．表2は，昭和47年7月～8月の歩行者天国の日のM店への客をカウントして得た調査結果の1部である．

表 2　歩行者天国の日のデパートへの客

年　月　日	昭和47年 7月23日	7月30日	8月6日	8月13日
延べ到着人数	188 521	208 629	188 334	190 897

これを，まだ歩行者天国が実施されていなかった45年11月15日（日曜）の到着数 53 829 人と比べてみると，桁違いの繁昌ぶりである．

45年11月17日（火）の 到着数は 38 030 人であり， 同じ火曜日でも 47年 7月18日（火）は 47 764 人であった．そこで季節変動や伸びを加味して

$$47\,764 \div 38\,030 = 1.262 \text{（倍）}$$

という数値を，かりに日曜日にも適用してよいことにすれば，47年 7月に歩行

者天国の影響がみられなかったとするときの，M店への到着数は大体

$$53\,829\times1.262=67\,932\ (\text{人})$$

したがって，歩行者天国による影響がM店では，実に2.7倍を上回わる客の入場となって表われている．デパートにとっては，まさに天国さまさまである．

しかし，ここでちょっとした疑問がわく．――というのは，45年11月15日の到着数53 829人をもってしても，ピーク時には，かなりの混雑を呈していた．
(注：ピーク時の平均滞留人数 $L=6\,820$ 人，平均滞留時間 $W=44.8$分)

$L=6\,820$ でさえも，相当の混雑なのに，到着人数が2.7倍以上になったという事実を，どう解釈したらよいであろうか．

その1つの，もっともらしい解釈はこうである．――延べ到着人数が2.7倍になったといっても，それはLが2.7倍になったことを意味しない．むしろ，平均滞留時間Wのほうが 1/2.7 になったと考えるべきである．（実際には L もかなり増大しているが，Wはもっとはるかに減少したとみるのが正しいであろう）．

このように，歩行者天国の日のM店への客は，全体の到着人数という面では大きくふくれあがり，個々の客の滞留時間という点では，ずっと短縮されていると解釈するならば，それだけでもおのずから，販売政策における手の打ち方が考えられようというものである．

§2. 図式統計学1つの考え

2.1　ヒストグラムと累積図表

統計資料による解釈を容易にする目的から，各種の図示法が随時用いられるが，その中でヒストグラム（柱状図表）はいちばん基本的なものであるといってよかろう．

さて，統計図表の作り方については，小学校のときからいろいろ学んできているし，中学校の数学教材としても，中1で“資料の整理”という項があって，ヒストグラム，累積度数のグラフなどが扱われている．しかし，ヒストグラム

1つをとってみても，いろいろむずかしい問題がある．たとえば，これは筆者の友人で数学の指導主事をしている人のはなしであるが，中学校の先生方に統計教材の扱い方を説明していたところ，次のような質問があったそうである．

「階級が不等間隔になっている度数表をもとにしてヒストグラムを書くのには，どうしたらよいのですか.」

すかさず彼は

「原則として，そのような場合を中学校では扱いません」

と答えたところ

「数学では扱わなくとも，社会科の方ではそんなのを扱う必要もあるのではないでしょうか．現に官庁統計などを見ますと，そういった度数表がふんだんにのっているのですから」

「それはそうですね」

ということになり，指導主事は彼の考えているような解釈を示したところ，その先生も納得されたというのであるが，筆者は

「それでよかったのだろうね」

と彼からたずねられて，われわれも同意見だと答えたことがあった．

いまの話を，もうすこし具体的に例題の形で述べてみよう．

例題 1．ある営業所に100人のセールスマンが所属していて，それらの人による1期間での売上高が表3のようになった．ヒストグラムをつくれ．

表3をながめてみると，階級1から10までは100万円の幅になっているが，階級11から先は，いずれも1000万円の幅になっている．ヒストグラムの上で，これをどう処理するか，である．これに対する指導主事の回答は，図3のように階級幅が10倍になっている階級の度数は1/10にしてグラフを書けばよい，ということにあった．

このように扱うと，次のような反論が出るかもしれない．

「ヒストグラムの上から，たとえば0～100（万円）の売り上げをもたらしたセールスマンが30人いたというのは結構だが，1000～2000（万円）は0.4人，5000～6000（万円）が0.5人と誤って読まれてしまうではないか.」と．

表3 度 数 表

売上高 階級	売上高 金額 (万円)	セールスマン人数	売上高 階級	売上高 金額 (万円)	セールスマン人数
1	以上 0 ～ 未満 100	30	9	以上 800 ～ 未満 900	1
2	100 ～ 200	22	10	900 ～ 1 000	1
3	200 ～ 300	12	11	1 000 ～ 2 000	4
4	300 ～ 400	3	12	2 000 ～ 3 000	3
5	400 ～ 500	3	13	3 000 ～ 4 000	7
6	500 ～ 600	2	14	4 000 ～ 5 000	5
7	600 ～ 700	1	15	5 000 ～ 6 000	5
8	700 ～ 800	1	計	計	100

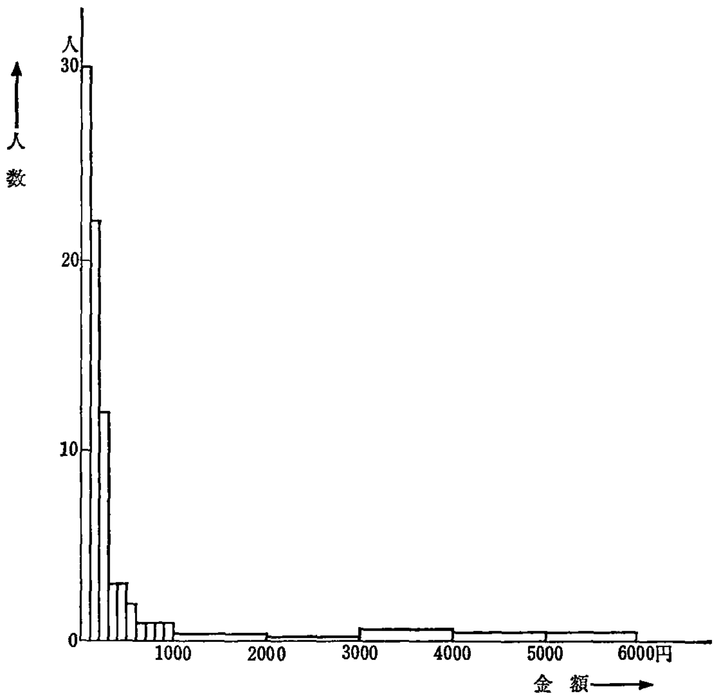


図3 ヒストグラム

しかし，この反論はナンセンスである．“なぜヒストグラムを作るのか”という，かんじんな点を考慮していない質問だからである．ヒストグラムは，全体の様子を端的に把握するために利用されるものであり，本来個々の値を読み取ることが第一義的なものではないからである．したがってそのような観点からすれば，図3の縦軸の目盛りはとり外しておいたほうがよいかもしれないが，それではグラフを読みにくくするので，単に便宜的に記してあるにすぎないものであって，その絶対的な数値を読み取ってはいけない．

一方，同じく中学校1年で扱われる累積度数の表やグラフについて，学習指導要領は

「着目する階級より下位または上位の階級の度数全部とその階級の度数を加えたもの（累積度数）を求めて，それを表にしたり，グラフをかいたりすると，ある対象が全体の中でどのような位置を占めるかということがはっきりすることを理解させる.」

とうたっている．

累積図表については，例題1のような度数表をもとにして書く場合であっても，何ら疑義が生じないであろう．これを例題2として扱ってみよう．

例題 2.

例題1の表3にもとづいて，累積度数のグラフをつくり，折れ線で結べ．

（解） 表3から累積度数表をつくると表4のようになるので，図4が得られる．

表4 累積度数表

金額（万円）	累積人数	金額（万円）	累積人数
100 未満	30	900 未満	75
200 〃	52	1 000 〃	76
300 〃	64	2 000 〃	80
400 〃	67	3 000 〃	83
500 〃	70	4 000 〃	90
600 〃	72	5 000 〃	95
700 〃	73	6 000 〃	100
800 〃	74		