統計学講義

第 2 版

林　周二
著

丸善株式会社

統計学講義

第 2 版

林　周二
著

丸善株式会社

著者の略歴（1976年現在）

1926年 生れ
1948年 東京大学経済学部卒業
現在 東京大学教養学部教授
著書‘流通研究入門’(1975)
‘現代製品論’(1973)
‘システム時代の流通’(1971) 他

統 計 学 講 義 第2版 ¥2,000

昭和48年3月20日発行
昭和54年9月20日第5刷発行

著作者 林(はやし) 周二(しゅうじ)

発行者 飯泉新吾

発行所 丸善株式会社

郵便番号103 東京都中央区日本橋二丁目3番10号

印刷 日東紙工株式会社・製本 株式会社 松岳社

3033―2133―7924

はしがき

この書物は，大学の教養あるいは前期課程のための統計学教科書，参考書を意図し，同課程の学生に，現代の社会や学問と結びついた統計的方法一般に関する知識・技術を習得させることを目的として書かれたものである．

そのような理由から，本書は特定の社会科学，特定の理工学などの専門分野についての個有な統計的諸方法には深く立ちいらない．また数理統計学の教科書とも趣旨を異にするので，数学的厳密さを 読者は 要求しないことを期待する．したがって特定の専門進学方向の定まった学生の場合には，本書を了えた上にさらにそれらの分野と結びついた斯学の書物に就くことを希望する．本書の巻末にはそういった方向別の文献があげてある．

筆者自身が本書を教室でのテキストに使用する場合には，一週一回一年分の講義を想定し，一回にほぼ一章ずつを充て，終ってなお余裕があれば，学生の進学方向にそくして補論を行うことを予定している．講義のやり方によっては，Ⅲ部（すなわち 20 章以下）について取捨をおこない，対象聴講者の状況に応じて講義を特化させることも可能である．

本書の初版上梓いらい，ほぼ十年を閲するのでこのさい版元とはかり，内容の全改訂と増補を試みた．初版との主要な相違点は，在来の本文を全面的に整序し，巻末に数章を新たに添加した．また各章尾に演習問題をまとめて挿入した．などである．紙幅増加を極力おさえる方向で，その作業を行なった．全くの別構想による新著ではなく，‘第２版’である以上は，初版の思想や構成は，良くも悪くも，そのままこの版へ受け継がれている．

1973 年 2 月

著　　者

凡　　例

1. 本書の全体は，3つの‘部’と 24 の‘章’と附章とからなり，別に目次の後に記号一覧，巻末に参考文献，年表，数値表，附図二葉ならびに索引が附されている．索引は，用語の和英対照表を兼ねている．

2. 各章は，それぞれ幾つかの‘節’に分けられているほか，I 部，II 部の各章の末尾には演習問題が附されている．III 部各章の末尾には附されていない．

3. 数式には，各章内で通し番号が附けられている．たとえば，式 (*3·11*) は‘3 章の 11 番目の式’の意である．また図と表とも各章内で通し番号がふられている．

4. 巻末ならびに本文中の数表類のうち，次のものに関してはそれぞれ原版権者などの，文書による好意ある許可をえた．

表 6·1，表 16·1　R.A. Fisher and Yates, Statistical Tables for Biological, Agricultural and Medical Research (3 rd. rev. ed.), Oliver Boyd,
Sir Ronald A. Fisher, F.R.S., Cambridge および Dr. Frank Yates, F.R.S., Rothamsted ならびに Oliver & Boyd Ltd., Edinburgh の許可により，同書‘第 33 表’‘第 3 表’から抄録.

表 15·1　C.M. Thompson, Biometrika, The Office of Biometrika (1941-2),
Prof. E.S. Pearson ならびに C.M. Thompson の許可により収録.

表 17·1　M. Merrington and C.M. Thompson, Biometrika (1943),
同　上.

表 19·1　A.E. Treloar, Elements of Statistical Reasoning (1939), John Wiley,
John Wiley & Sons, Inc., N.Y. の許可により収録.

表 20·2　日本科学技術連盟，数値表 (A)，(1955)，
日科技連出版社の許可により収録.

主な記号，表示

A, A_1	事　象	ν_r, ν_r'	積　率
a_i	要　素	$\mathrm{P}(\ )$	確　率
α	第一種過誤の危険率	$\mathrm{P}(\lambda)$	ポアソン分布
α_r	基準系の積率	${}_n\mathrm{P}_r$	順　列
$\mathrm{B}(n, p)$	二項分布	$P(\ )$	検定力関数
$B(\ )$	ベータ関数	p	母出現率
β	第二種過誤の危険率	q	$1-p$
c	不良個数	R, R_x	範　囲
c_x	変動係数	r, r_{xy}	標本相関係数
$\mathrm{E}(\ )$	期 望 値	$r_{1\cdot 23}$	重相関係数
F	エフ分布する変量	$r_{12\cdot 3}$	偏相関係数
F_i	累積度数	ρ_{xy}	母相関係数
$F(\ )$	累積分布関数	S, S_x	標本変動
f_i	度　数	S_{xy}	標本の共変動
$f(\ )$	密度分布関数	s^2	S/n
$\Phi(z)$	基準正規分布の積分値	σ^2, σ_x^2	母 分 散
ϕ	自 由 度	T, T_x	標 本 和
$\Gamma(\ )$	ガンマ関数	t	ティ分布する変量
H	仮　説	θ	母数一般
χ^2	カイ平方分布する変量	V, V_x	標本分散
$L(\ )$	尤度関数	V_{xy}	標本の共分散
$M(\ )$	積率母関数	$\sqrt{V_x}$	標本の標準偏差
m	標本出現個数（$\leqq n$）	$\mathrm{V}(\ )$	分　散
m/n	標本出現率	W	危 険 域
μ, μ_x	母平均（期望値）	w	統計量一般
$\mu_{\bar{x}}$	平均の期望値	x, x_i, y, z	変数，変量，実現値（ことに z は基準化されたそれをさす）
$\mathrm{N}(\mu, \sigma^2)$	正規分布		
$\mathrm{N}(0, 1^2)$	基準正規分布	$\bar{x}$	標本平均
N	母集団の大きさ	$\bar{\bar{x}}$	同じくその平均
n	標本の大きさ	$\tilde{x}$	同じく中央値

なお一般に太字（たとえば $\boldsymbol{x}, \boldsymbol{y}, \boldsymbol{V}$ など）は確率変数を，＾印（たとえば $\hat{\mu}, \hat{p}$ など）は推定値を，・｜印（たとえば $S_{y\cdot x}$, $y|x$ など）は条件つきの意を，それぞれ表わす.

目　　次

I 序　　論

数表目次

1

統計的データ

1·1 事象とデータ

今日の段階での統計学は，統計的数字にもとづいて記述し，推測し，決定をするための科学であると考えて一応よいであろう．

ところで常識的にいえば‘統計的数字’とは，個々の事実にそくし，じっさいに測った数字的な資料，すなわち質的には相等しい‘データの集まり’のことをいう．このようなデータの集まりは，調査や実験，あるいは繰り返し作業などにさいして得られる．以下かんたんに‘データ’と呼ぶことにし，データについてのこのような通念にたいし，つぎのように多少まとまった言いあらわし方を与えてみよう．

数多くの事物をひとまとめの全体として考えたとき，私たちはこれを **事象** という（数学ふうに‘集合’，‘空間’といってもよい）．これをAであらわす．事象を構成している個々の個体 a_1, a_2, …, a_n を **要素** といい，$A=\{a_1, a_2, \cdots, a_n\}$ あるいは略して $\{a_j\}$ $(j=1, 2, \cdots, n)$ と書く．nを **事象の大きさ** という．

いまAを日本の総人口全体の集まりとし，そのひとりひとりを a_j とする．すると，人口は有限だからAは **有限事象** であり，その大きさは $n \fallingdotseq 1{\cdot}03\times10^8$ である．また自動包装機械が一分間に 80 個の割で包装食品をつくってゆくとする．この包装食品のひとつひとつを $a_1, a_2, \cdots$ であらわすとすれば，それら

の全体の集まりAは，この機械によって製造されるだろう包装食品のすべてにあたる．製造機械の寿命は無限ではないから，Aは有限事象だけれども，このような場合には,実際問題としては **無限事象** を想定したほうが好都合である．

さてつぎに二つの測定結果をながめてみよう．

[例 1·1] 20 人の学生について血液型をしらべたところ，下の結果がえられた．

A	A	O	B	B	B	A	A	B	O
O	A	A	A	AB	O	AB	A	O	O

[例 1·2] 100 個の某カメラ・シャッターを一様に 1/100 秒にセットして，シャッターを切ったところ，つぎのような露出時間をえた．(単位はミリセカンド)

10·5	10·5	11·4	10·8	8·3	9·2	9·1	9·7	8·5	10·1
8·7	10·2	12·6	10·2	9·9	9·5	9·7	9·0	8·4	8·1
9·8	9·6	9·0	11·0	7·0	9·2	8·8	9·4	6·3	9·1
11·2	9·5	10·0	10·6	8·8	7·0	8·7	9·6	9·4	10·3
9·9	8·2	9·2	7·6	9·4	8·8	8·3	9·2	9·3	8·6
10·8	9·8	9·1	9·6	10·7	10·1	9·5	7·9	8·3	10·2
8·3	9·1	8·7	9·5	11·5	8·0	7·3	8·5	9·4	9·3
9·8	9·5	8·9	9·7	9·0	8·3	8·3	9·3	7·5	9·6
10·2	7·9	7·4	8·2	9·4	9·6	10·6	7·6	7·2	12·1
8·7	7·6	10·0	9·5	10·4	9·4	11·8	10·3	8·3	10·1

私たちは帳簿記録類を転記したり，またじっさいに調査や実験を行なったりすることで，このような資料の集まりを得ることがしばしばある．上の2つの例のように，一般に，事象A——この場合はそれぞれ学生の集まり，シャッターの集まり——が与えられたとして，これについて要素 a_j のそれぞれに相当する標識となるようなそれの指標xを考えることができる場合，私たちはこのxをさして，仮りに **特性** とよぶことにしよう．(事象の要素 a_j については，特性 x 以外に他の特性 y, z, … を添えてこれを考えることも，もちろんできる．たとえば‘例 1·2’の場合に関していえば，各シャッター a_j については，露出時間xのほかにも，その耐久度 y，シャッターを切るときの堅さ z,

合格・不合格の決定 w, … などの特性を付与して考察することが可能である.)

一般に，ある統計的な事象が与えられたとき，これらの要素に対して，単数または複数個の特性を付与し，これをひとまとめにして考えたとき，これらを **統計的データ**（以下，略して **データ** という）とよぶ. データを構成している個々の観察値をもデータという. 特性の個数が一つのときこれを **一重分類のデータ**，そうでないときこれを **多重分類のデータ** という.

‘例 1·1’ ‘例 1·2’ はいずれも一重分類のデータである. そして，たとえば ‘例 1·2’ の場合，露出時間とともに仮りに耐久度をしらべ，それを露出時間にそえて同時計測したとすれば，二重分類のデータを考えることができる.

データの個数 n を **データの大きさ** という.（‘データの大きさ’ と混同しやすい言葉に **データの数** というのがある. 後者は特定の大きさをもったデータが幾組も存在しているとき，その組の個数をさすに当って用いる言葉である.）‘例 1·1’ ‘例 1·2’ では，データの大きさは，それぞれ $n=20$, $=100$ である.

特性のうちには，定量的に言いあらわすことのできるものと，ただ言語的にだけしか言いあらわすことのできない定性的なものとがある. たとえば酒のコク，食品の風味，衣料の肌ざわりのような官能特性は，たやすくは定量化できない. あらゆる物ごとや現象を定量化して考察することは，科学的な研究にとっては重要な手続きであるが，ただし定量化の方向が科学的アプローチの唯一のみちだと考えることは，必ずしも正しいとは言えないだろう.

さて A の要素 a_j が取りうる標識を x, y, z, … などであらわしたとして，もしこれら x, y, z, … などが数的な特性であるとすれば，x, y, z, … などはこれを **変数** と見なすことができるから，a_j における x, y, z, … の値はこれを関数表現によって $x(a_j)$, $y(a_j)$, $z(a_j)$, … などと書くことができよう.

また本来ならば非数的な特性によって表示される定性的な標識であっても，適当な数値を賦与し数量化を行うならば，これを変数として取扱うことは可能である. たとえば ‘例 1·1’ の場合，A型の学生については $x=1$, それ以

外の学生については $x=0$ と賦値すれば，x はひとつの変数と見なすことができる．このような特性を，変数という見方から，**定義変数** とよぶこともできよう．変数的なデータにあっては，データをとくに **実測値** とよぶことがある．なお以下かんたんのために $x(a_j)$, $y(a_j)$, … などを，略して単に x_j, y_j, … などと記すことにする．

また変数は，これを実数軸のうえのあらゆる連続的な値をとる **連続変数** と，そのうちの特定の値だけをトビトビにとる **離散変数** とに分けておくと便利である．'例 1·2' 以下の実例にそくしていえば，シャッターの露出時間などは連続変数である．これに対し仕切り（ロット）中の不良品の個数などは離散変数である．なお本来ならば連続変数であるものでも，現実には，ある桁位において四捨五入などして表示をする結果，見かけのうえでは離散変数の姿をとるものもある．定義変数はもちろん離散変数である．

データの集まりを考える場合には，変数は必ずしも一つでなくてよい．すなわち，A_j が変数 $x, y,$ … などのうちの単一個のものについてだけ考えられるとき，私たちはこれら $\{x_j\}$, $\{y_j\}$ を **一変数のデータ** とよぶ．またこれらのうちのどれか二つ（または三つ以上）の同時表示によってあらわされるとき，私たちは $\{(x_j, y_j)\}$ を **二変数のデータ**，$\{(x_j, y_j, \cdots, w_j)\}$ を **多変数のデータ** などという．

1·2　統計データの取扱い

統計的な数値計算は，すべて，ある具体的な推測ないしは決定目的をもった実用計算を立て前とする．実用計算の原則は，そのために必要な正確さをもつ値（近似値でもよい）を，速く（ないし安く）出すことにある．そのためには，たとえばつぎのような心得が必要であろう．

第1は，計算の手段を吟味することである．計算には，道具を使わない