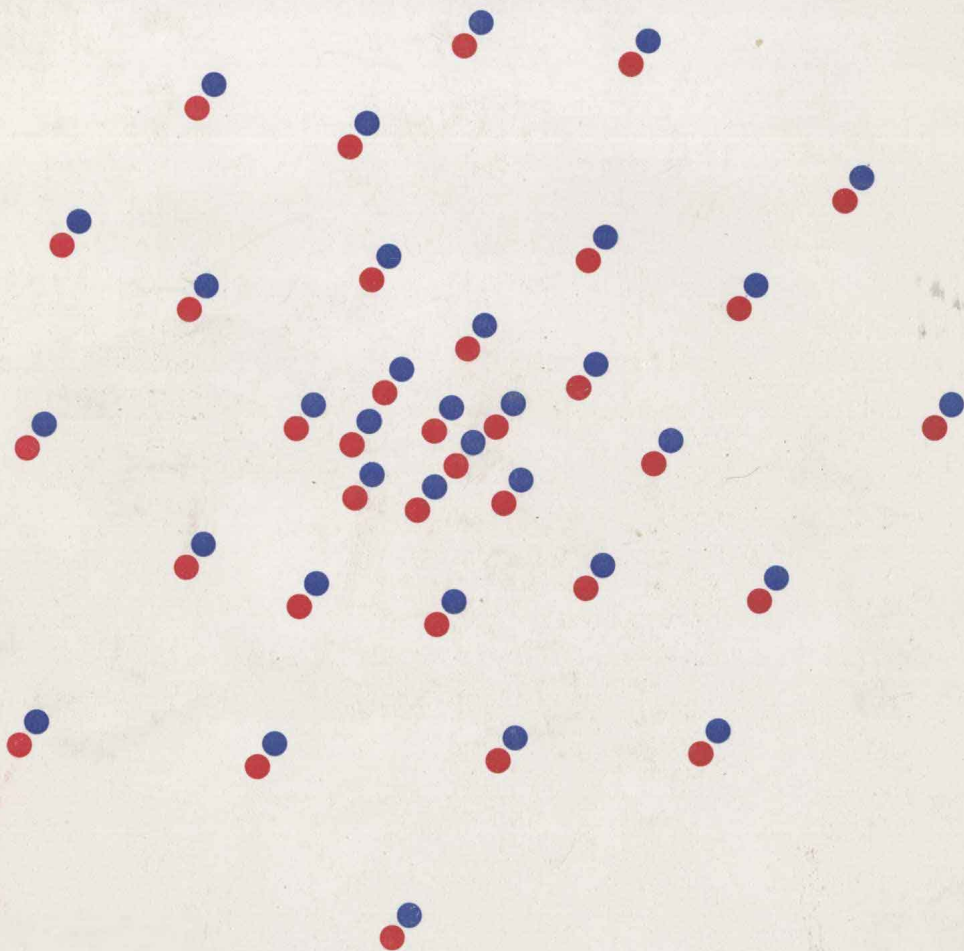


統計学入門

稲葉三男 著



共立出版株式会社

統計学入門

稲葉三男 著

共立出版株式会社

著者紹介

いなばみつお
稲葉三男

熊本大学名誉教授，理学博士

統計学入門

定価 1200 円

NDC 417.5

検印廃止

© 1981

昭和56年4月5日 初版1刷発行

著者 稲葉三男

発行者 南條正男

東京都文京区小日向4丁目6番19号

印刷者 田中春吉

東京都荒川区東日暮里6丁目20番9号

発行所 東京都文京区小日向4丁目6番19号
電話 東京 947 局 2511 番 (代表)
郵便番号 112 振替 東京 1-57035 番

共立出版株式会社

印刷・杜光舎印刷 製本・協栄製本 Printed in Japan

社団法人
自然科学書協会
会員



3041-117170-1371

はじめに

数学関係書はとかく、改まった姿勢でできるだけ一般的・抽象的に述べはじめる。そして、続いて定理・公式を使って次々に一般理論が展開されてゆく。一般の読者は、述べられていることについての具体的なイメージさえ思い浮べる間もなく、方角もことばもわからない異国の中を引きずりまわされている心境に追いこまれてしまうであろう。そのために、読者はしばしば学習を断念させられることになるであろう。特に、統計学関係の学習になると、統計資料という数値の集団を処理するために、数値計算をさけることができない。数値計算に重荷を感じる一般の読者は、せっかく統計学に関心をもっていたり、必要性を感じていたりしても、統計学の学習になじめなくなることもあるであろう。

本書は、このような現実を反省して、読者が接しやすいような具体例からはじめることにし、具体例についての統計学的な見方を述べることによって、必要になってくる諸概念について直観的なイメージを正確につかませることをねらっている。全体がストーリーになっているように、当面必要になってくる統計学の考え方や定理について、直観的にでも理解できるように工夫してある。数値計算については、えんぴつと紙による筆算という伝習的な考え方を脱却して、近年ますます改良され、そして普及してゆく電子式卓上計算器(電卓)——四則計算向きの小型程度でもよい——を利用することをおすすめしたい。そうすることによって、数値計算の重荷から解放されるであろうし、本書の付表の活用と併せて、統計計算の基本をマスターすることができるであろう。本文中にある「問」は、出会いしだいに即席に処理するものとして設けてある。

本書は上のような趣旨のもとに編著されたものであるが、読者を統計学の領域の入口に誘いこむことができたならば、それは本書の成功というべきものである。本書によって学習したことがらについて、さらに理論的な展開によって

iv はじめに

の裏づけの必要を感じる読者は

北川敏男・稲葉三男共著，統計学通論（第2版），共立出版
について，あるいは，同書「参考書」にあげられてある専門書について，学習
をすすめることをおすすめしたい。

昭和 56 年 2 月

著 者

数学関係書より

世界の数学教育…数学教育国際委員会編/B5・382頁
 数学的発見の論理……………佐々木 力訳/A5・230頁
 数学基礎論 増補版……………竹内外史他著/A5・174頁
 現代数学入門事典……………穂刈四三二監訳/B6・248頁
 数学英和・和英辞典……………小松勇作編/B6・350頁
 数学小辞典……………矢野健太郎編/B6・770頁
 共立 数学公式 改訂増補 近藤基吉他編/ポケット判・672頁
 学生 数学ハンドブック…穂刈四三二他編/三六・368頁
 商・経系基礎数学……………小林竜一著/A5・206頁
 数学概説……………赤 瓶也他訳/A5・324頁
 基礎数学……………稲葉三男著/A5・174頁
 一般教育 数学……………金沢 隆著/A5・204頁
 文科の数学……………岡本 勲他著/A5・164頁
 数学序論……………柴田敏男著/A5・206頁
 大学教科 数学序説……………稲葉三男他著/A5・154頁
 数学の序説……………川崎薩男著/A5・228頁
 教養数学概論……………矢野健太郎監修/A5・182頁
 社会科学のための 数学概説…小林竜一著/A5・286頁
 現代科学における 数学概説Ⅱ 後藤憲一著/A5・496頁
 新しい数学……………矢野健太郎訳/A5・368頁
 理工科系一般教育 代数・幾何教科書…占部 実他編/A5・260頁
 大学教養課程 理工科系代数学と幾何学…小松醇郎他著/A5・230頁
 初等代数幾何学……………伊東由文著/A5・360頁
 行 列……………稲葉三男著/A5・204頁
 大学課程 行列とベクトル空間 大学数学教育研究会編/A5・204頁
 ベクトル解析 新版……………友近 晋著/A5・168頁
 基礎課程 線形代数学……………尾野寺 毅他著/A5・234頁
 線形代数序説……………中谷太郎編/A5・198頁
 線形代数 教科書……………佐々間元敬他編/A5・302頁
 大学教科 線形代数概説…小松醇郎他著/A5・162頁
 理工科系一般教育 線形代数…荒木 淳著/A5・204頁
 大学教科 線形代数要論…小松醇郎他著/A5・242頁
 線形代数学通論……………和田秀三他編著/A5・214頁
 基礎数学 線型代数……………那須俊夫他著/A5・192頁
 線形数学……………武貞利隆他編/A5・200頁
 初等線形数学……………平野次郎他著/A5・176頁
 代数学講義 改訂新版……………高木貞治著/A5・400頁
 ブール代数……………三宅彦三郎著/A5・318頁
 りー環論……………松島与三著/A5・196頁
 可換環と微分……………中井喜和著/A5・168頁
 初等整数論講義 第2版……………高木貞治著/A5・432頁
 基礎数学 微分積分……………宮下熊夫他著/A5・182頁
 例題形式 微分積分学……………外岡慶之助編/A5・220頁
 微分積分学通論……………和田秀三他編著/A5・198頁
 微分積分学……………小林幹雄著/A5・172頁
 微積分学……………岡野初男著/A5・212頁

大学教養微分積分……………竹之内 脩他著/A5・180頁
 大学 微分積分学……………田中明雄他著/A5・296頁
 大学課程 微分積分学概説 新版……………大学数学教育研究会編/A5・268頁
 理工科系一般教育 微分・積分教科書…占部 実他編/A5・280頁
 大学教養課程 理工科系微分学と積分学 福原満洲雄他著/A5・252頁
 微分積分学演習上 改訂版 坂井英太郎著/A5・464頁
 微分積分学演習下 改訂版 坂井英太郎著/A5・418頁
 大学教科 解析学概説……………竹之内 脩著/A5・214頁
 大学教科 解析学要論……………福原満洲雄著/A5・226頁
 現代解析学……………近藤基吉他訳/A5・324頁
 位相解析……………加藤敏夫著/A5・336頁
 確率とBASIC……………池浦孝雄他著/A5・254頁
 2値系列……………伊理正夫他訳/A5・168頁
 リースナジー 関数解析学下 秋月康夫監訳/A5・336頁
 複素関数概説……………黒田 正著/A5・206頁
 大学テキスト 関数論……………小松勇作他著/A5・178頁
 函数概論 改訂版……………竹内端三著/B6・284頁
 微分方程式の解法……………田中静男著/A5・248頁
 基礎数学 微分方程式通論 福原満洲雄他著/A5・226頁
 微分方程式概論……………田中明雄他著/A5・202頁
 応用微分方程式論……………永宮健夫著/A5・274頁
 工学基礎 微分方程式……………林 英輔他著/A5・238頁
 ユークリッド原論……………中村幸四郎他訳・解説/B5・572頁
 ボンリヤーゲン 常微分方程式 新版……………千葉克裕訳/A5・344頁
 幾何学通論……………朝長康郎著/A5・228頁
 新版 集合論……………小松勇作改訂/A5・192頁
 統計学入門……………稲葉三男著/A5・144頁
 統計学概論……………岡田泰幸著/A5・190頁
 基礎数学 統計学通論 第2版…北川敏男他著/A5・240頁
 大学教科 統計学要論……………竹之内 脩監修/A5・218頁
 基礎数理統計……………関口愷夫他著/A5・196頁
 調査統計入門……………船津好明著/A5・294頁
 医学・生物学の統計学……………阪谷太一著/A5・260頁
 数表と計算……………黒田孝郎編/A5・208頁
 電子計算機による 計算法入門…松 信訳/A5・280頁
 数値計算における誤差……………戸川隼人編/A5・196頁
 近似理論入門……………一松 信他訳/A5・304頁
 電子計算機 数値計算法とプログラミング…小郷 寛他著/A5・138頁
 FORTRAN 数値計算とプログラミング…国井利泰監訳/A5・224頁
 基礎 図学……………玉腰芳夫他著/A5・194頁
 大学教育 図学……………中根孝治著/A5・200頁
 要説 図学……………田中政夫著/A5・150頁

目 次

1 章 統計資料を読む	1
1. 1 統計資料を整理する	1
(1) 資料から度数分布表をつくる	1
(2) 資料のグラフをつくる	5
1. 2 分布の特性を表す特性値	8
(1) 分布の代表値	8
(2) 度数分布表から平均値を計算する	10
(3) 分布のちらばり (ばらつき)	12
1. 3 2 変量の場合の分布	16
(1) 相関の状態	16
(2) 相関の度合	18
(3) 相関表をつくり, そして読む	22
共同研究	28
2 章 標本調査の考え	29
2. 1 標本調査の意味	29
(1) 全部を調査することは	29
(2) 標本調査することは	31
(3) 無作為ということは	32
2. 2 確率の考え	35
(1) 確率を割り当てる	35
(2) 事象のいろいろ	37
(3) 確率の基本的な性質	39
(4) 確率の実践的な意味は	42
2. 3 ばらばらな値をとる変数	46
(1) 試行をくり返す	46
(2) 事象を変数で表す	49
(3) 二項分布	50
(4) 平均値, 分散	53

2. 4	連続した値をとる変数	56
(1)	確率のアイデアを一般的にする	56
(2)	ルーレットの場合	58
(3)	一様分布	60
(4)	正規分布	62
2. 5	確率と標本調査とのつながり	67
(1)	標本をとる実験の試み	67
(2)	標本分布	71
3 章	統計的に推測する	77
3. 1	推定の考え	77
(1)	母平均と母分散を推定する	77
(2)	区間を推定する	82
3. 2	検定の考え	86
(1)	疑いの心理から数理に	86
(2)	二項分布を正規分布で近似する	88
(3)	正規母集団を検定する	91
(4)	正規確率紙による検定	92
3. 3	カイ二乗分布	95
(1)	正規分布からカイ二乗分布に	95
(2)	カイ二乗分布で検定する	97
3. 4	F 分布	103
(1)	正規分布から t 分布, F 分布に	103
(2)	F 分布で検定する	106
問題解答		111
付 表		113
付 録	正規確率紙	129
索 引		131

1 章

統計資料を読む

1.1 統計資料を整理する

(1) 資料から度数分布表をつくる

S 中学校 1 年女生徒の身長はどんなものかを知りたいので、同校の定期健康診断（測定）の身長の部を書き写すことにする。測定の数字は、たとえば

148.5 cm 141.2 cm 155.6 cm …

のようにくわしく記載されている。ところで、わたくしたちの関心は、1 年女生徒個々の身長の正確な測定値を知ることにあるのではなくて、1 年女生徒全体の身長の状態を知ることにある。そこで、1 cm 以下は四捨五入して

149 cm 141 cm 156 cm …

のように丸めることにする。そうすると、女生徒の身長の測定値は表 1.1 のように書き改められる。

このように、S 中学校 1 年女生徒の身長の測定値を得たとき、これら女生徒の集合を**集団**ということにし、個々の女生徒を集団の**個体**という。このときの身長のように、測定される特性を**変量**といい、表 1.1 のような身長の測定値の集合をこの変量に関する**統計資料**または単に**資料（データ）**という。わたくしたちの関心は、統計資料を得ることによって、集団の特性（変量）に関する情報または判断を得ることにある。

ところで、表 1.1 の統計資料そのままでは、S 中学校 1 年女生徒の集団の身

2 1章 統計資料を読む

表 1.1 S中学校1年女生徒 92 人の身長 (昭和55年4月) (単位 cm)

1	組	149	141	156	166	151	154	145	156	152	154
		143	143	149	149	148	146	149	155		
2	組	159	148	144	146	156	147	157	151	161	153
		155	151	155	153	161	155	157	145	152	
3	組	151	155	157	143	149	146	144	154	147	156
		151	151	149	154	151	153	159	153	158	
4	組	157	154	149	151	155	156	146	153	150	141
		152	155	154	137	144	141	148	144		
5	組	152	147	154	155	150	153	150	153	151	142
		144	147	155	151	159	152	150	148		

長については、なんら見通しのよい情報も得られないであろう。そこで、この資料を分類することにする。この資料は 135 (cm) から 166 (cm) までの範囲にはいっている。この範囲を、たとえば、4 (cm) 刻みの小範囲

135-138 139-142 143-146 147-150 151-154

155-158 159-162 163-166

に分割し、これらの小範囲を級(階級)とよぶことにする。表 1.1 の資料のおのおのの数値がどの級に属するかをチェックして、これらの級に属する個数を数えあげる (表 1.2)。この個数をその級に対応する度数という。各級に対応する度数の系列を統計資料に対する度数分布といい、これを表の形式に表したものを度数分布表という (表 1.3)。いまのところ、級の幅を 4 (cm) にとったのであるが、この幅を級間隔という。級間隔は、ほかに特別の必要がない場

表 1.2 表 1.1 に対する度数のチェック

級	チェック	度数
135-138	一	1
139-142	正	4
143-146	正 正 正	14
147-150	正 正 正 正	19
151-154	正 正 正 正 正 正	29
155-158	正 正 正 正	19
159-162	正	5
163-166	一	1
計		92

表 1.3 表 1.1 に対する度数分布表

級	級中値	度数
135-138	136.5	1
139-142	140.5	4
143-146	144.5	14
147-150	148.5	19
151-154	152.5	29
155-158	156.5	19
159-162	160.5	5
163-166	164.5	1
計		92

合には、等しくとることが多い。

級間隔について見直してみよう。たとえば、級 135-138 の 135 と 138 は四捨五入によって丸めた数字であるから、級 135-138 は 134.5 (cm) 以上 138.5 (cm) 未満の範囲である。この範囲の中央の値は

$$\frac{134.5+138.5}{2} = \frac{135+138}{2} = 136.5 \text{ (cm)}$$

で、級 135-138 の級中値とよばれる。この級にはいるのは 4 組の 137 (cm) だけであるが、これはこれからはこの級の級中値 136.5 (cm) に丸めることにする。同じように、級 139-142 は 138.5 (cm) 以上 142.5 (cm) 未満の範囲で、この級中値は

$$\frac{138.5+142.5}{2} = \frac{139+142}{2} = 140.5 \text{ (cm)}$$

である。この級にはいるのは、1 組の 141, 4 組の 141, 141 および 5 組の 142 の 4 個で、これらはこれからは級中値 140.5 (cm) に丸めることにする。

以下の級についても、同じようにして、表 1.3 の度数分布表が得られる。

統計資料から度数分布表をつくるときに、級の数、したがって、級間隔をどのくらいにとったらよいかは、資料の個数、性質および統計利用の目的に従って定められるべきものである。また、級の度数の代わりに、級の度数の集団の総個数に対する百分率（パーセント）でおきかえることがある。このときの度数分布表を相対的度数分布表という。

級 135-138 に対して、134.5 および 138.5 をそれぞれ級下限界および級上限界といい、両者を併せて級限界という。級 139-142 に対しては、138.5 および 142.5 がそれぞれ級下限界および級上限界である。以下の級についても、同じようである。表 1.3 に対して、各級の級限界以下の級の度数を加え合わせて累積度数をつくると、表 1.4 の累積度数分布表が得られる。級限界以下の級の度数の代わりに級限界以上の級の度数とすると、もう一つの累積度数分布表が得られ

表 1.4 表 1.1 に対する
累積度数分布表

級	累積度数
cm 以下	
134.5	0
138.5	1
142.5	5
146.5	19
150.5	38
154.5	67
158.5	86
162.5	91
166.5	92
計	92

4 1章 統計資料を読む

表 1.5 S中学校1年生
の性別分布表（昭和 50 年）

性別	男	女	計
人数	93	92	185

る。
S中学校1年生185人について、性別を調べると、
男生徒93人、女生徒92人である。このように、集
団の特性が質的属性であって、数えられるときも、属
性による階級分けをすることによって、度数分布表を
つくることができる。S中学校1年生の性別は表1.5

の性別分布表で表される。また、たとえば、東京都の法定伝染病患者数は、病
名による階級分けをすることによって、表1.6の度数分布表で表される。

表 1.6 東京都の法定伝染病患者数（昭和 48 年）

病 名	赤 痢	腸チフス	パ ラ チフス	猩紅熱	ジフテ リ ア	流行性脳 脊髄膜炎	日本脳炎	計
患者数	186	24	5	2 310	18	6	0	2 550

（第 25 回日本統計年鑑による）

表をつくる上での注意

1° 表題は簡潔で、一見して内容が明らかであるようにする。

表 1.7 日本人の年齢および男女別
死亡表（昭和 48 年）

年齢階級	男	女
0- 4	18 188	13 161
5- 9	2 258	1 366
10-14	1 393	860
15-19	4 180	1 545
20-24	6 245	3 213
25-29	5 686	3 345
30-34	6 465	3 956
35-39	9 338	5 105
40-44	13 367	7 135
45-49	15 511	9 617
50-54	16 251	12 371
55-59	23 415	16 064
60-64	35 377	23 183
65-69	46 656	30 331
70-74	58 118	43 561
75-79	55 840	51 833
80-84	39 957	49 943
85-	25 041	49 164
不詳	306	71
計	383 592	325 824

（第 25 回日本統計年鑑による）

- 2° 注または説明書きを表の適当な場所につけ
る。たとえば、表 1.6 および 表 1.7 では
出所が欄外下に示されている。
- 3° 表題の説明として、表の内容に関する日
時および場所を記入する。
- 4° 列または行の題目には、何が記入されてい
るか、特に列と列（または行と行）を区別す
るものは何であるか、を簡潔に示す。
- 5° 行と列は、比較を容易にするために、たと
えば、大きさの順に並べるように、それぞれ
系統的な順序に並べる。
- 6° 表に使用されている単位を記入する。たと
えば、表 1.8 では 1 000 戸が単位である。
- 7° 数字の長い列を並べるときは、5 行または
10 行ごとに間隔をおく（表 1.7）。
- 8° たとえば、4685732 のように桁数の多い数
字は、区切って書く。西洋式ならば、右から

3字ごとに区切って4 685 732のように、日本式ならば4字ごとに区切って468 5732のように書く。

表 1.8 全国借家延べ面積別住宅数（昭和 48 年）（単位 1 000 戸）

延べ面積 (m^2)	20 未満	20-29	30-49	40-49	50-69	70-99	100-149	150-199	200-249	250 以上
借 家 数	1 659.1	1 997.4	2 654.0	2 212.6	1 877.2	451.1	125.9	28.5	6.7	5.5

（第 25 回日本統計年鑑による）

（2）資料のグラフをつくる

統計資料の度数分布を目に見えるように、直観的にとらえやすくするために、資料をグラフで表すことが工夫される。このグラフを統計図表という。統計図表にはさまざまあるが、ここではこれからさきに必要なものだけに止めることにしよう。

まず、表 1.3 の度数分布表について考えてみよう。級中値を横座標、対応する度数を縦座標とする点

(136.5, 1), (140.5, 4), (144.5, 14), (148.5, 19),
(152.5, 29), (156.5, 19), (160.5, 5), (164.5, 1)

を座標平面上にとって、次々に線分で結んで得られる折線が表 1.3 の度数分布表に対する度数折線である（図 1.1）。このとき、横軸の単位長さと縦軸の単位長さとを等しくとることは必要でなく、グラフが見やすいように適宜にとる工夫をするといふ。

次に、表 1.3 の度数分布について、級 135-138 の級下限界、級上限界を端点とする区間 [134.5, 138.5]

を底とし、度数 1 を高さとする長方形をつくり、級 139-142 の級下限界、級上限界を端点とする区間 [138.5, 142.5] を底とし、度数 4 を高さとする長方形をつくり、

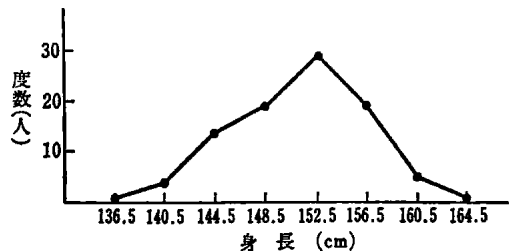


図 1.1 表 1.3 に対する度数折線

6 1章 統計資料を読む

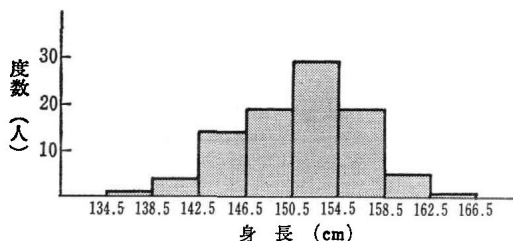


図 1.2 表 1.3 に対するヒストグラム

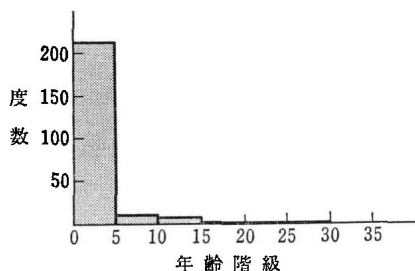


図 1.3 昭和 50 年麻疹での死亡者の年齢別分布のヒストグラム

以下同じようにして次々に長方形をつくる。これらの長方形を結びつけたものが、表 1.3 の度数分布表に対するヒストグラムまたは柱状図である (図 1.2)。この場合にも、座標軸の単位長さを適宜にとる工夫をする。度数が急激に変化したり、級間隔が全部は等しくないような場合には、ヒストグラムはぐあいがいよい (表 1.10, 表 1.11)。

ヒストグラムの形によって、分布の型を次のように名づける。ヒストグラムが対称の形をなしているとき、分布は**対称型**であるという。図 1.2 のヒストグラムは正確には対称ではないが、分布は割に対称型に近いものとみなされる。対称型からのずれがはっきりしているときは、分布は**非対称型**であるという。図 1.3 では、年齢階級は、たとえば、級 5-10 は 5 歳以上 10 歳未満を表しているのであるが、ヒストグラムは非対称型の極端な場合で、この分布は**J 型**である (問題 2 の表 1.10 のヒストグラムをも参照)。図 1.4 のように、中央部の階級の度数が小さくて、両端に度数が集中しているとき、分布は**U 字型**である (問題 3 の表 1.11 のヒストグラムをも参照)。

実際に現れる分布は、これらの分布型のいずれかに属す

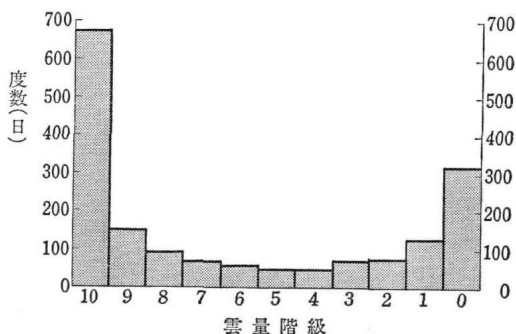


図 1.4 イギリス・グリニッチの 6 月 (1890-1904) の 1715 回観測の雲量分布のヒストグラム

ることもあるが、これらの分布型の混合とみられるものも少なくない。試みに、表 1.7 の度数分布のヒストグラムをつくってみよ。

度数分布表から度数折線をつくったと同じ手続きで、累積度数分布表から**累積度数折線**が得られる。たとえば、表 1.4 の累積度数分布表については、級下限界を横座標、対応する累積度を縦座標とする点

(134.5, 0), (138.5, 1), (142.5, 5), ..., (166.5, 92)

を座標平面上にとって、次々に線で結んで得られる折線が表 1.3 に対する累積度数折線である (図 1.5)。

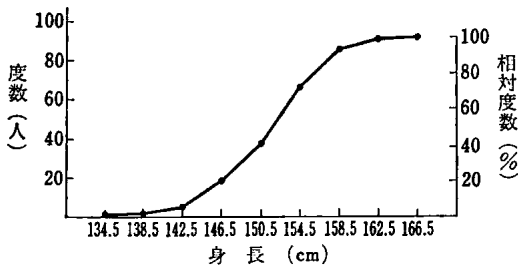


図 1.5 表 1.3 に対する累積度数折線

表 1.9 K 小学校 6 年生 (11 歳) の
男女別身長分布表 (昭和 52
年 4 月) (単位 cm)

階 級	男	女
130-132	2	3
133-135	9	3
136-138	8	9
139-141	16	14
142-144	22	9
145-147	14	20
148-150	5	15
151-153	5	9
154-156	2	3
157-159	0	1
160-162	1	1
計	84	87

【問題 1】 表 1.9 の度数分布表に対して、男生徒、女生徒別に、度数折線、ヒストグラムおよび累積度数折線をつくれ。

【問題 2】 次の表 1.10 の死亡者数について、ヒストグラムをつくれ。(年齢階級 5-14 は 5 歳以上 15 歳未満を表している。)

表 1.10 高血圧性疾患による年齢階級別の女性死亡者線 (昭和 40-52 年)

年齢階級	0-4	5-14	15-24	25-34	35-44	45-54	55-64	65-74	75 歳以上
死亡者数	0	0	5	21	67	188	497	2 020	8 280

(第 29 回日本統計年鑑による)

【問題 3】 次の表 1.11 の死亡者数について、ヒストグラムをつくれ。

8 1章 統計資料を読む

表 1.11 肺炎による年齢階級別の男性死亡者数（昭和 40-52 年）

年齢階級	0	1-4	5-14	15-24	25-34	35-44	45-54	55-64	65-74	75 歳以上
死亡者数	566	184	108	97	130	234	485	1 061	3 630	7 840

（第 29 回日本統計年鑑による）

【問題 4】 表 1.8 の度数分布表に対してヒストグラムをつくれ。

1.2 分布の特性を表す特性値

（1）分布の代表値

S 中学校 1 年女生徒の身長と K 小学校 6 年女生徒の身長を比較することを考えてみよう。個々の生徒についていうならば、中学校の女生徒の身長が小学校の女生徒の身長より大きいこともあり、反対に、小学校の女生徒の身長が中学校の女生徒の身長より大きいこともある。しかし、集団としてみると、S 中学校の女生徒の身長は K 小学校の女生徒の身長より大きいとみられる。このことがらを数値で表現するために、統計資料について中心となる変量の値としての**代表値**が導入される。

たとえば、1.1 節の表 1.1 の資料のうち、1 組の生徒の集団を考えると、その身長（cm）の資料

149 141 156 166 151 154 145 156 152 154
143 143 149 149 148 146 149 155

のおのおのを同等なものとして扱って、これらの総和をつくって資料の個数 18 で割って得られる

$$(149+141+156+\cdots+155)\div 18=2\,706\div 18=150.3^{1)}$$

は資料の代表値で、**算術平均**または**平均値**とよばれる。

【問 1】 1.1 節の表 1.1 の資料のうち、2 組の女生徒の身長（cm）の資料

159 148 144 146 156 147 157 151 161 153
155 151 155 153 161 155 157 145 152

の算術平均（平均値）を求め、1 組の生徒の身長の算術平均（平均値）と比較せよ。

1) 本来ならば、 $\div 150.3$ （四捨五入）とすべきであるが、今後簡略してこのように表すことにする。

(答 152.9 (cm))

上のことを一般化すると、集団の変量 x の統計資料を $x_1, x_2, \dots, x_n$ とするとき、その算術平均 (平均値) は

$$\frac{x_1 + x_2 + \dots + x_n}{n}$$

である。この値を $\bar{x}$ で表すことにする ($\bar{x}$ は x バー bar と読む)。ここで、和 $x_1 + x_2 + \dots + x_n$ は、 x_i で i を $1, 2, \dots, n$ としたときの和 Sum、すなわち、 x_i の $i=1$ から $i=n$ までの和 Sum で、短縮して

$$\sum_{i=1}^n x_i$$

で表される (Σ は、ローマ字 S に対応するギリシア文字 (シグマ) で、Sum の略である)。そうすると、算術平均 (平均値) $\bar{x}$ は

$$\bar{x} = \frac{1}{n} \sum_{i=1}^n x_i = \frac{x_1 + x_2 + \dots + x_n}{n}$$

で表される。

統計資料の代表値としては、一般には算術平均 (平均値) がよく用いられるものであるが、そのほかに次のものがあげられる。1.1 節の表 1.1 の資料のうち、2 組の女生徒の身長 の測定値 (問 1) を大きさの順に並べると

144 145 146 147 148 151 151 152 153 153 155
155 155 156 157 157 159 161 161

となる。中央にくる値 153 を代表値とすることもある。このような代表値を **中位数** または **中央値** という。この中位数は、組のうちの身長があまりにも小さいものや、あまりにも大きいものがある場合、そういう極端な値には影響されないという点では、つごうがよい。2 組のように、個体の数が奇数であるときは、中央にくるのはただ一つであるが、1 組のように個体の数が偶数 (18) であるときは、中央にくるのは二つ (9 番目と 10 番目) あるので、この両者の算術平均をとって中位数とする。

【問 2】 1.1 節の表 1.1 の資料のうち、1 組の女生徒の身長の中位数を求めよ。

(答 149.0 (cm))