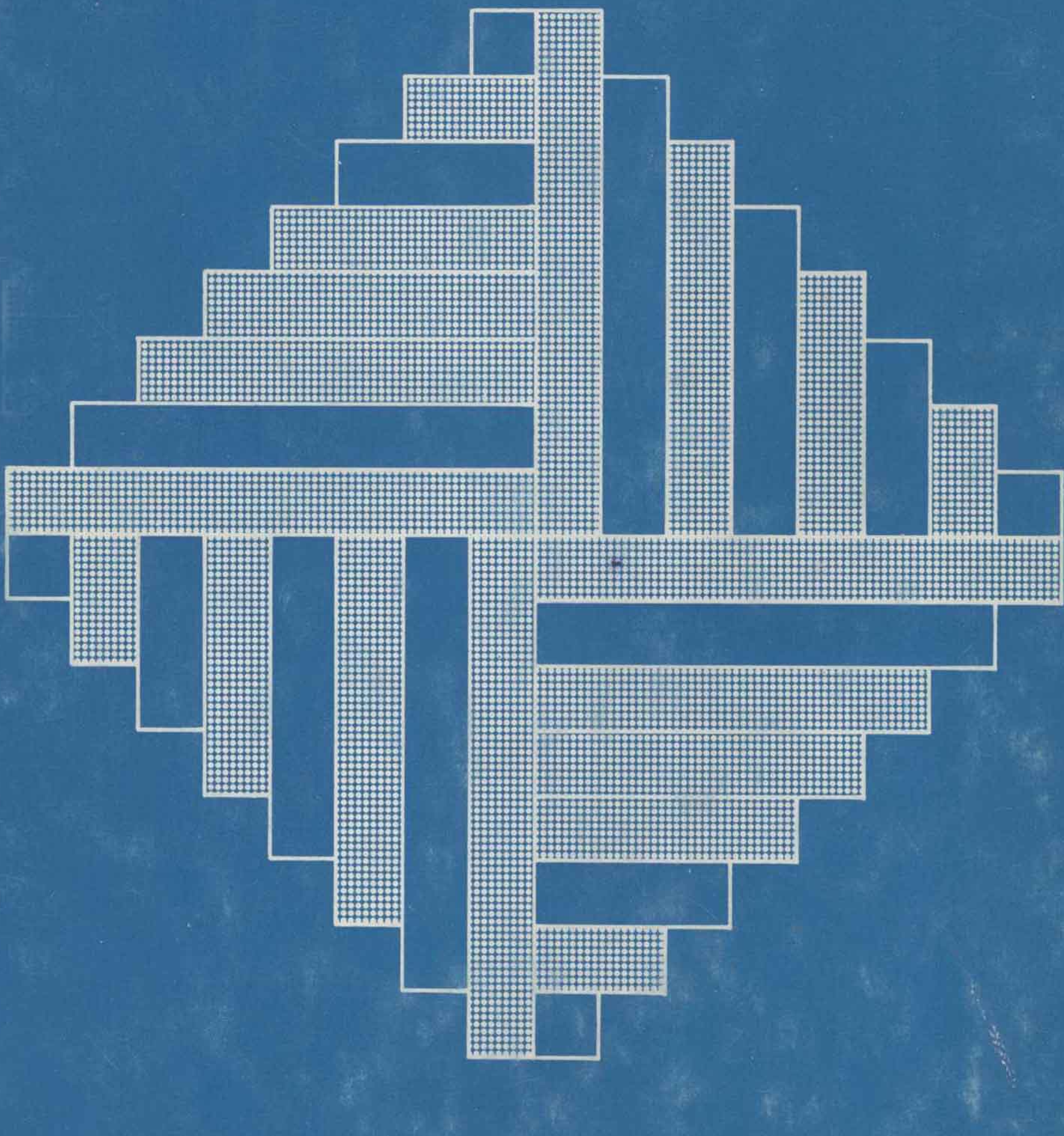
基礎数学

統計学通論

第2版

北川敏男
稲葉三男
共著



共立出版株式会社

基礎数学
統計学通論
第2版

北 川 敏 男
稲 葉 三 男
共著

共立出版株式会社

著者紹介

きたかわとしお
北川敏男　九州大学名誉教授・理学博士

いなばみつお
稲葉三男　熊本大学名誉教授，久留米工業大学教授・理学博士

基礎数学
統計学通論
第2版

定価 1300 円

昭和35年12月20日　初版1刷発行
昭和53年1月20日　初版160刷発行
昭和54年5月1日　第2版1刷発行

著者　北川敏男
　　　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　福田三郎
東京都渋谷区猿楽町19番2号

NDC 417.5

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 振替口座東京1-57035番
共立出版株式会社

印刷・真興社　製本・中條　Printed in Japan

社団法人
自然科学書協会
会員

3041-117025-1371

第 2 版 序

本書初版は，1960 年 12 月の第1刷以来， 1978 年1月の第 160 刷まで，18 年にわたり， 大学教育における統計学の標準的な教科書の一つとしての役割を果し，また広く一般的な入門書として，多くの読者の勉学・教養に役立つことができたのは，著者たちの歓びとするところである.

この間，本書を採用された教師諸氏はじめ，多数の熱心な読者のなかから，本文のみならず演習問題の細部にまでわたって，貴重な校訂ないし忠言を賜った．近年このような修正が少なくなっているが，引続き毎年刊行を見ている本書初版に対し，これからの読者のため，より一層親近感をもって利用していただくため，全般的な改訂を行うことにした.

すなわち，例題として引用する統計資料は，古典的価値の高いもののほかは更新し，また解説の細部にわたって，一層の精確をはかり，趣意の明確なることを期し，所要の修正を行った．ただ，現時点においては，高等学校の学習要領等の現状および将来を考慮して，初版の基本構想はそのまま継承し，大綱においては従来どおりとした.

この第2版の刊行にあたっては，読者諸氏から寄せられた適切な注意に感謝したい． また，校正にあたっても， 数多くの知友の厚意に負うところが多いが， 特に久留米工業大学 富永賢二氏の協力に負うことが大きかったことを記して，感謝の意を表したい.

1979 年 3 月

著　　者

序

統計学に関する授業時間は，現在，週2時間で半年ないし1年という大学が少なくない．この限られた時間数において，しかも数学的教養のまだ十分でない学生諸君を対象にしての講義には，入門的・基礎的な事項の解説だけに力を集中した教科書が望ましいと考えられる場合が少なくない．たまたま，共立出版株式会社の依嘱もあって，年来，そのような教科書の必要を痛感していたので，ここに本書を執筆するにいたったのである．

本書の構成は数理統計学の入門を解説する目的に沿ったものである．しかし他のいくつかの同程度の教科書のように，数理的展開だけに主眼をおくのではなく，むしろ，統計学の基礎概念を明確に説明することに重点をおき，具体的例題を多くとり入れ，それを理解するための数学的な裏付けに留意し，統計概念を体得するよう配慮したつもりである．本書の内容は，基礎的範囲のなかの一部分に限定したから，なお解説すべき多くのことが残されている．しかし本書の克服によって，正確に応用する能力の基礎をつくること，統計学の各方面の学習に進む準備をつくること，この二つの目的が達せられるならば，著者の喜びとするところである．

本書では，基礎的概念の解明に焦点を向けるため，また数学的教養の十分でない読者に無益の抵抗を感じさせないため，本文中には数理的展開を差し控えたものがある．これらは巻末の「補注」にまとめて，数学的な裏付けに備えて読者にいささかも不安の念を残さぬようにした．ただここでも微積分学の初歩程度の数学利用にとどめてある．それ以上高度の数学を必要とする部分については，本文中に【*】印をつけて結果だけを述べておいた．

統計学の学習には，学生自身で，演習問題を解くことが最も効果的と思われるから，できるだけ演習を行なうことをおすすめしたい．このため各節のおわりに演習問題をつけてある．労力および時間を相当必要とするものは，「共同研究」としてある．

本書は，以上のような目標で編著したものであるが，私どもの不才のため，いたらぬことが多いことを恐れる．教授各位ならびに学生諸君からの叱正，注意を乞う次第である．現代統計学の概観を求められる余裕ある場合，あるいは数学の素養によっては，岩波全書「推測統計学 I，II」（北川著）に進まれるよう期待している．

本書の編著には，多くの友人の協力を得たが，とくに細部の検討および校正については熊本女子大学助教授 杉村正彦氏および熊本大学理学部 古川長太氏に負うこと多大であった．記して感謝の意を表明する．

1960 年 10 月

著　　者

目　　　次

第1章　資料の整理

第2章　確　　　率

第3章　確率分布

第4章 標 本 分 布

第5章 推 定

第6章 検 定

付章 統計学の発達

第1章　資料の整理

1·1　度数分布表

（1）**変　量**　学童のある集団の身長について調査しようとするとき，個々の学童の身長を測定し，この測定値を集めると，この集団の学童の身長に関する**統計資料**（statistical data）が得られる．このとき，身長のように，ある特性の度合を数量的に表すものを**変量**（variate, variable）という．変量のうち，身長，体重，温度，雨量などのように，測定または計量されるものを**連続変量**（continuous variate）といい，人口，物の数，ある事件の起こる回数のように，計数として数えられるものを**離散変量**（discrete variate）という．

連続変量の測定値は完全正確を期することはできないもので，測定の精度に限界がある．したがって，ある程度の近似値をもって満足しなければならない．学童の身長の測定値の場合，たとえば，145.5 cm 以上 146.5 cm 未満の測定値は四捨五入して 146 cm と丸めるのがふつうである．このように，学童集団について身長の統計調査を行う目的は，個々の測定値そのものや学童個々の個性に関心があるのではなく，学童の集団全体についての身長に関する情報，知識，または判断を得ることである．測定の精度をも考慮に入れて，測定値を丸めることもあるし，目的によっては，さらに少ない有効数字で丸めることがある．

離散変量の場合にも，たとえば，A市人口 326,000 人，B市人口 729,000 人のように，最初の有効数字3桁で丸めることがある．これは，多くの統計では調査上にいろいろな誤差が伴うので，3桁以上の数字が有効でないときに行う．

（2）**度数分布表**　得られた統計資料を分析して，判断を引き出したり，他の統計資料に対して比較したりするためには，統計資料を整理しなければならない．それにはまずふつう，分類をすることから始まる．**分類**(classification)

をするには**製表** (tabulation) によることが多い．表のうちで最も重要なものが，次に述べる**度数分布表** (table of frequency distribution) である．

連続変量の場合には，変量の範囲をいくつかの**級** (**階級**, class) に分け，統計資料の測定値のこれらの級に属する個数を**度数** (frequency) という．各級に対応する度数の系列は，この特性に対する集団の構造を表すもので，この系列を**度数分布** (frequency distribution) といい，これを表の形式で表したものが度数分布表である．級の幅を**級間隔** (class width) といい，特に他の必要がなければ一般には等しくとる．しかし，等間隔でないほうがつごうのよい場合もある．級の定義はあいまいさを残さないように明確にしなければならない．たとえば，学科の成績の点数の場合，級 30-39 は 29.5 以上 39.5 未満の区間を示すことにする[1]．級の中央の値を**級中値**または**階級値** (class mark) といい，この級にはいる測定値を丸めてこの級中値とする．級の限界の値を**級限界** (class limit) といい，その大きいほうを**級上限界** (class upper limit)，小さいほうを**級下限界** (class lower limit) という．級中値は級上限界と級下限界の相加平均に等しい．たとえば，級 30-39 の級上限界は 39.5 で，級下限界は 29.5 である．級中値は

$$\frac{29.5+39.5}{2}=34.5$$

である．

表 1·1　K高等学校 1 年 100 人の生徒の数学の基礎テスト (1977 年)

25	72	43	10	76	84	47	36	16	50
74	30	48	54	49	30	15	40	15	24
28	65	47	60	52	27	55	69	82	84
58	68	62	86	42	15	79	96	95	58
15	72	99	42	20	92	25	10	49	30
25	74	30	59	35	30	40	52	64	26
30	95	72	40	10	60	20	62	50	20
30	45	30	25	67	42	76	84	11	20
39	54	65	84	52	41	78	34	86	15
76	15	38	24	20	40	42	40	35	51

1) この場合，級のとり方は 30-39 だけに限ることはない．たとえば，級 30-40 として，30 点以上 40 点未満を表すとしてもよい．

表 1·1 によって与えられた学科成績の統計資料から度数分布表をつくると，表 1·2 または表 1·3 が得られる．

表 1·2 表 1·1 に対する度数分布表

級	度数計算	度数
10-19	正 正 一	11
20-29	正 正 正	14
30-39	正 正 正	14
40-49	正 正 正 丁	17
50-59	正 正 丁	12
60-69	正 正	10
70-79	正 正	10
80-89	正 丁	7
90-99	正	5
計		100

表 1·3 表 1·1 に対する度数分布表

級	級中値	度数
10-19	14.5	11
20-29	24.5	14
30-39	34.5	14
40-49	44.5	17
50-59	54.5	12
60-69	64.5	10
70-79	74.5	10
80-89	84.5	7
90-99	94.5	5
計		100

級の数がどのくらいがよいかは，資料の性質，統計利用の目的などを考慮して定められるべきである．実際上の経験からすると， 10～20 個くらいまでがつごうがよい．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．

離散変量の場合にも，属性による階級分けをすることによって，同じように度数分布が得られる（表 1·4, 表 1·5).

表 1·4 サイコロを 300 回投げたとき出た目の度数分布表

サイコロの目	1	2	3	4	5	6	計
度数	58	64	39	45	51	43	300

表 1·5 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）

（3） 累積度数分布（表）　各級の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，**累積度数分布（表）**（cumulative frequency distribution）が得られる．表1·6は表1·3の度数分布表からつくられた累積度数分布表である． 級限界以下の代わりに級限界以上とすると，もう一つの累積度数分布（表）が得られる．

表1·6　表1·1に対する累積度数分布表

階　級	累積度数
点以下 9.5	0
19.5	11
29.5	25
39.5	39
49.5	56
59.5	68
69.5	78
79.5	88
89.5	95
99.5	100
計	100

製表上の注意　1°．表題は簡潔で，一見して内容が明らかになるようにする．

2°．説明書または注を表の適当な場所につける．

3°．表題の説明として，表の内容に関係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）を区別するものが何であるか，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．

表1·7　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

表1·8　日本人の年齢および男女別死亡表（昭和48年）

年令階級	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）

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．

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，右から 4 字ごとに区切って 468 5732 のように表す．

問題 1. 下の表について累積度数分布表をつくれ．

某クラスの国語科成績分布表

成 績	人 数	成 績	人 数
30-39	1	70-79	15
40-49	6	80-89	8
50-59	13	90-99	2
60-69	23		

問題 2. 表 1·8 について男女別に累積度数分布表をつくれ．

1·2 度数分布図

統計資料を視覚に訴えて直観的にとらえるために，いろいろな **統計図表** (statistical diagram) が工夫される．次には，度数分布に対する図表，すなわち，**度数分布図** について述べる．

（1） **度数折線**　度数分布の i 番目の級中値を x_i とし，度数を f_i とするとき，平面上に座標軸をとって点 (x_i, f_i) を記す．これらの点 (x_i, f_i) を，x_i の小さいほうから大きいほうに次々に線分で結んで得られた折線が**度数折線** (frequency poligon) である．このとき，横軸の単位長さと縦軸の単位長さとが等しいことは必ずしも必要でない．

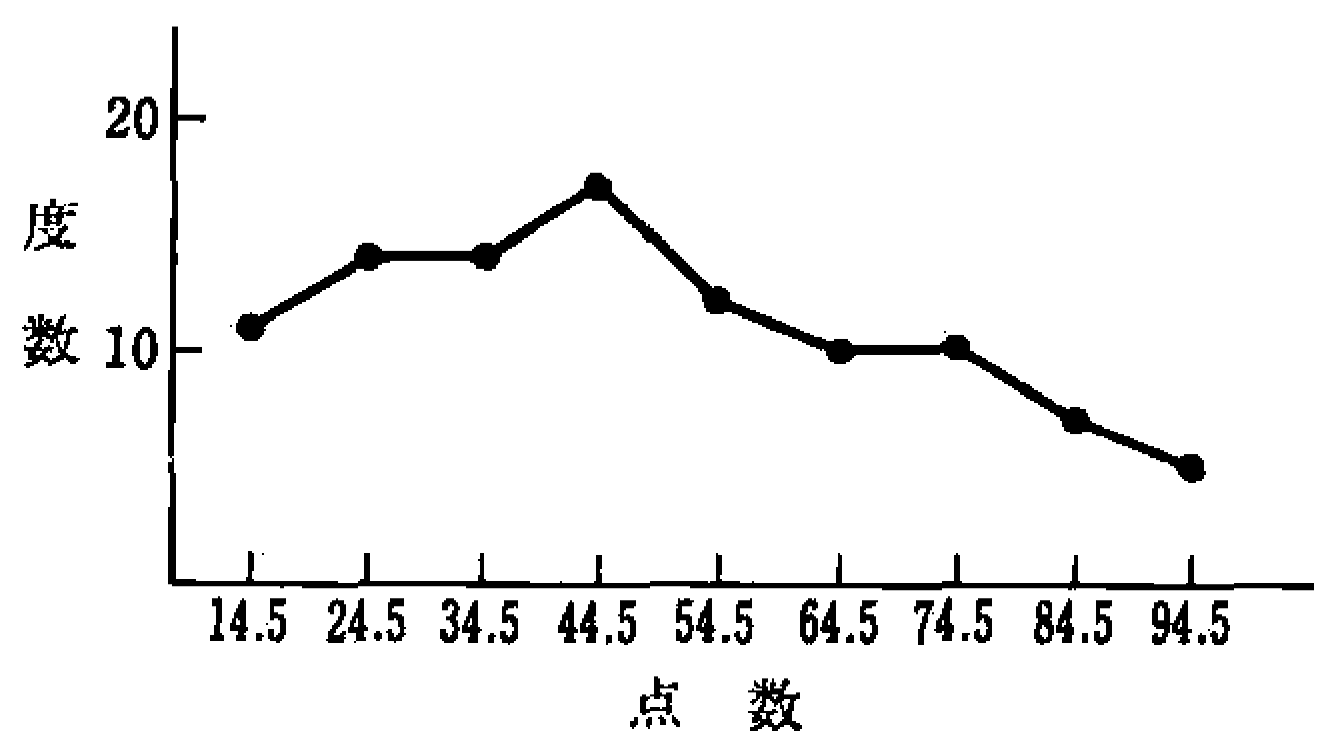


図 1·1 表 1·1 に対する度数折線

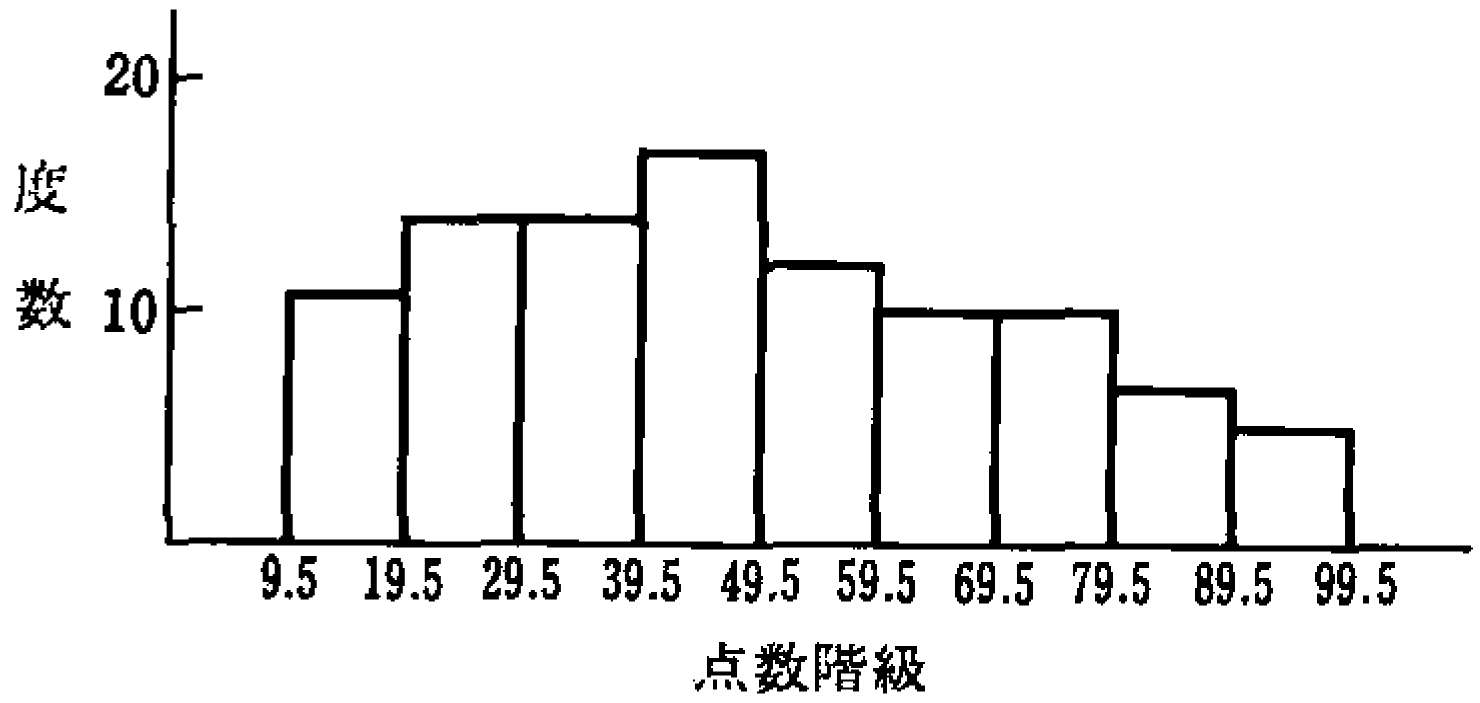


図 1·2 表 1·2 に対するヒストグラム

（2） **ヒストグラム**　度数

分布の i 番目の級を表す区間を底とし，級の度数 f_i を高さとする長方形をつくり，これらの長方形をすべての級に対して結びつけたものが**ヒストグラム**または**柱状図**(histogram) である．座標軸の単位長さを適宜の割合にとることは，度数折線の場合と同じである．度数が急激に変化したり，級間隔がことごとくは等しくないような場合には，ヒストグラムはぐあいがよい．

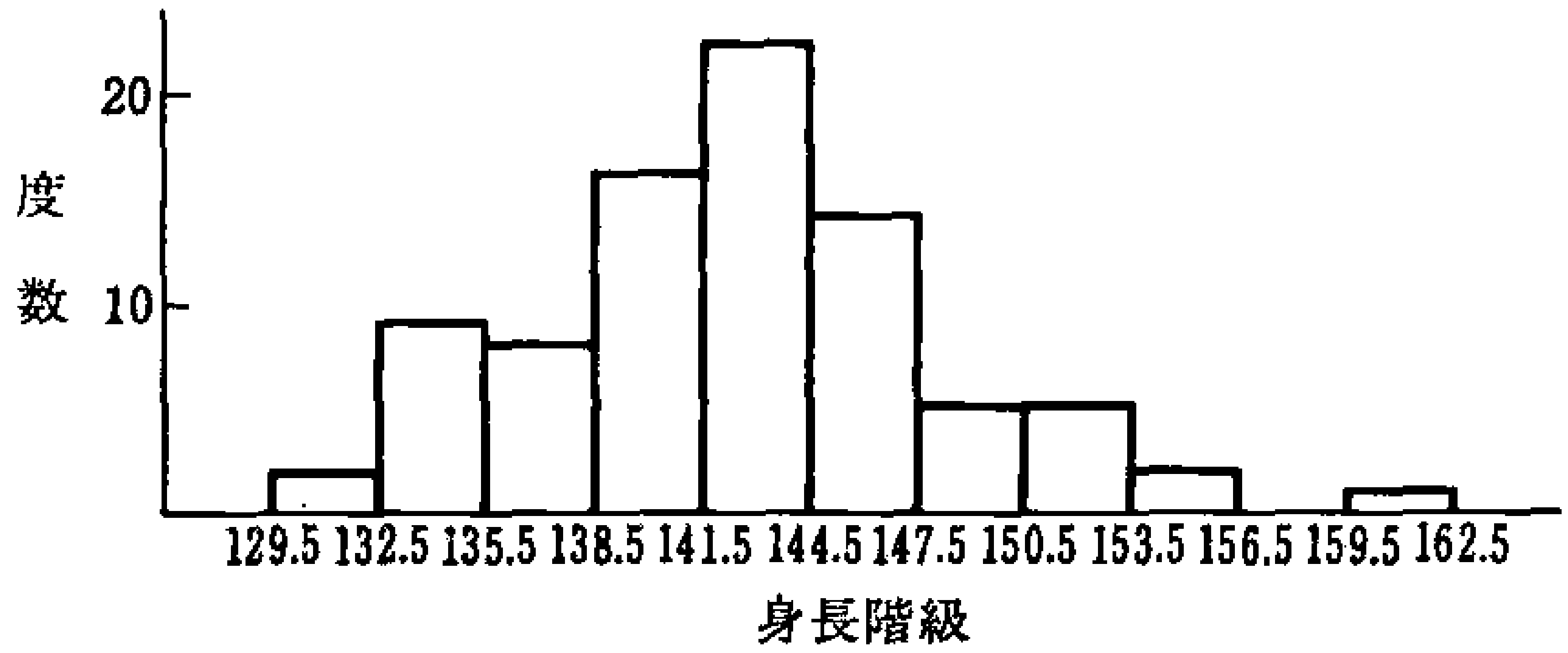


図 1·3　表 1·7 (男生徒の場合) に対するヒストグラム

（3） **分布の型**　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．対称の形をなすときは，分布は**対称型** (symmetrical) であるという．図 1·3 は正確には対称ではないが，分布は対称型に割合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** (asymmetrical) であるという．図 1·4 (年齢階級は，たとえば，5–10 は5歳以上 10 歳未満を表す) は，非対称型の極端になった場合で，**J 型**(J-shaped) であるという．図 1·5 のように，中央部の度数が小さくて，両端に度数が集中しているとき，分布は**U字型** (U-shaped) であるとい

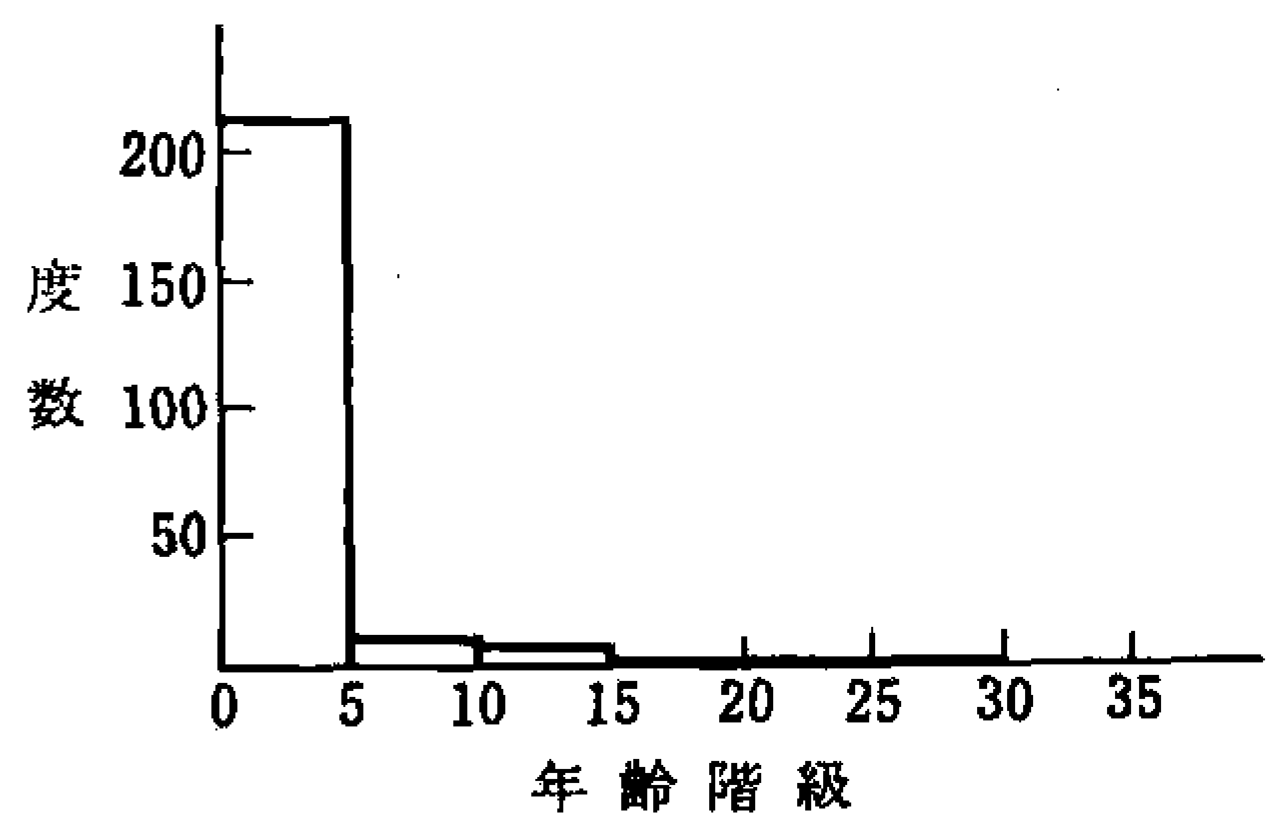


図 1·4　昭和 50 年麻疹での死亡者の年齢別分布のヒストグラム

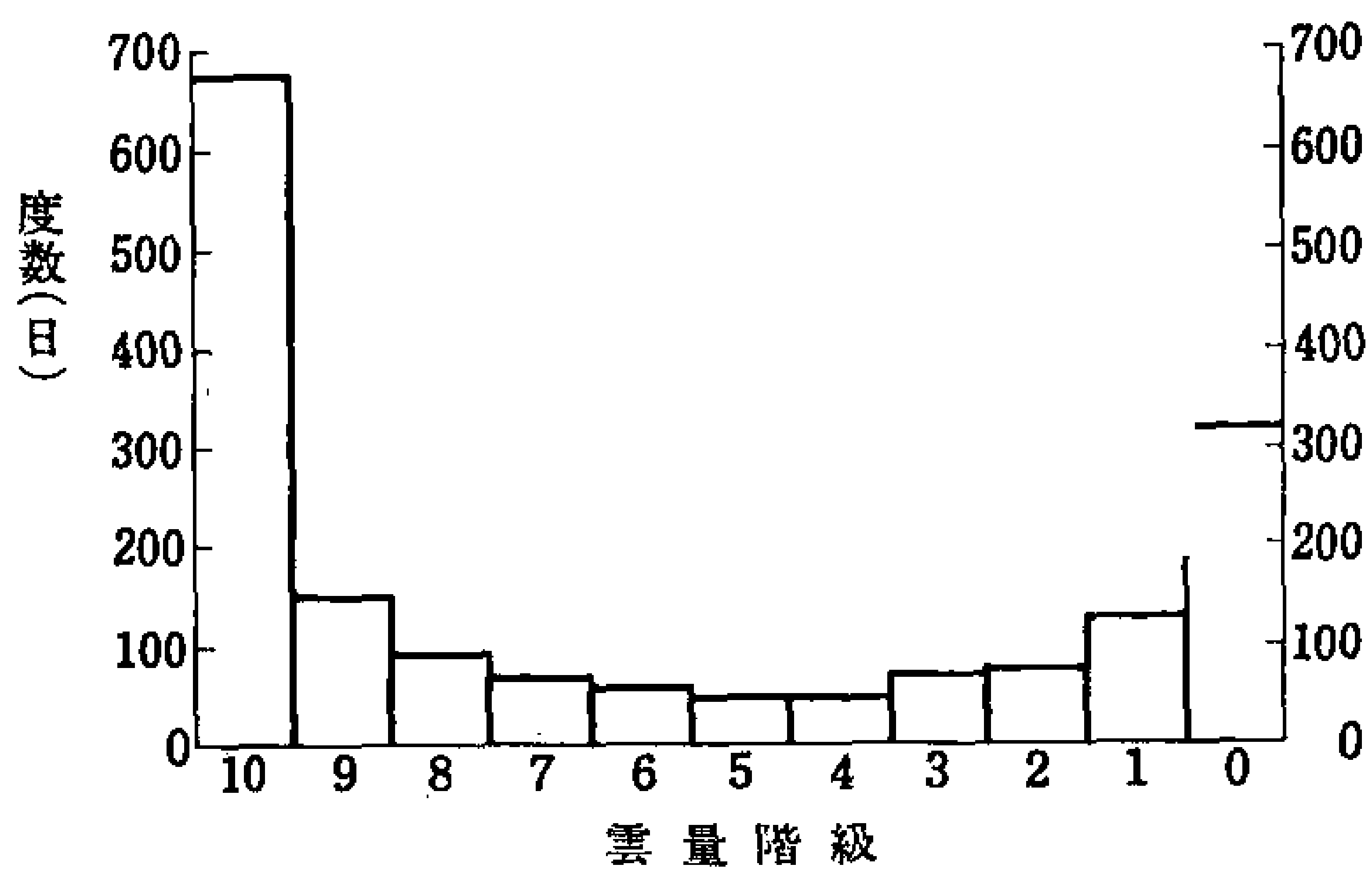


図 1·5　英国グリニッチの6月 (1890–1904) の 1715 回観測の雲量分布のヒストグラム

う．実際に現われる分布はこれらの分布型のいずれかに属することがあるが，これらの分布型の混合と見られるものも少なくない（表1·8の度数分布のヒストグラムを考えてみよ）．

（4） **累積度数分布図**　度数分布表から度数分布折線をつくったと同じ手続きで，累積度数分布表から**累積度数分布図**が得られる．すなわち，級限界を横座標とし，これに対応する累積度数を縦座標とする点を記し，これらの点を級限界の小さいほうから大きいほうに次々に線分で結んで得られる折線が累積度数分布図である．

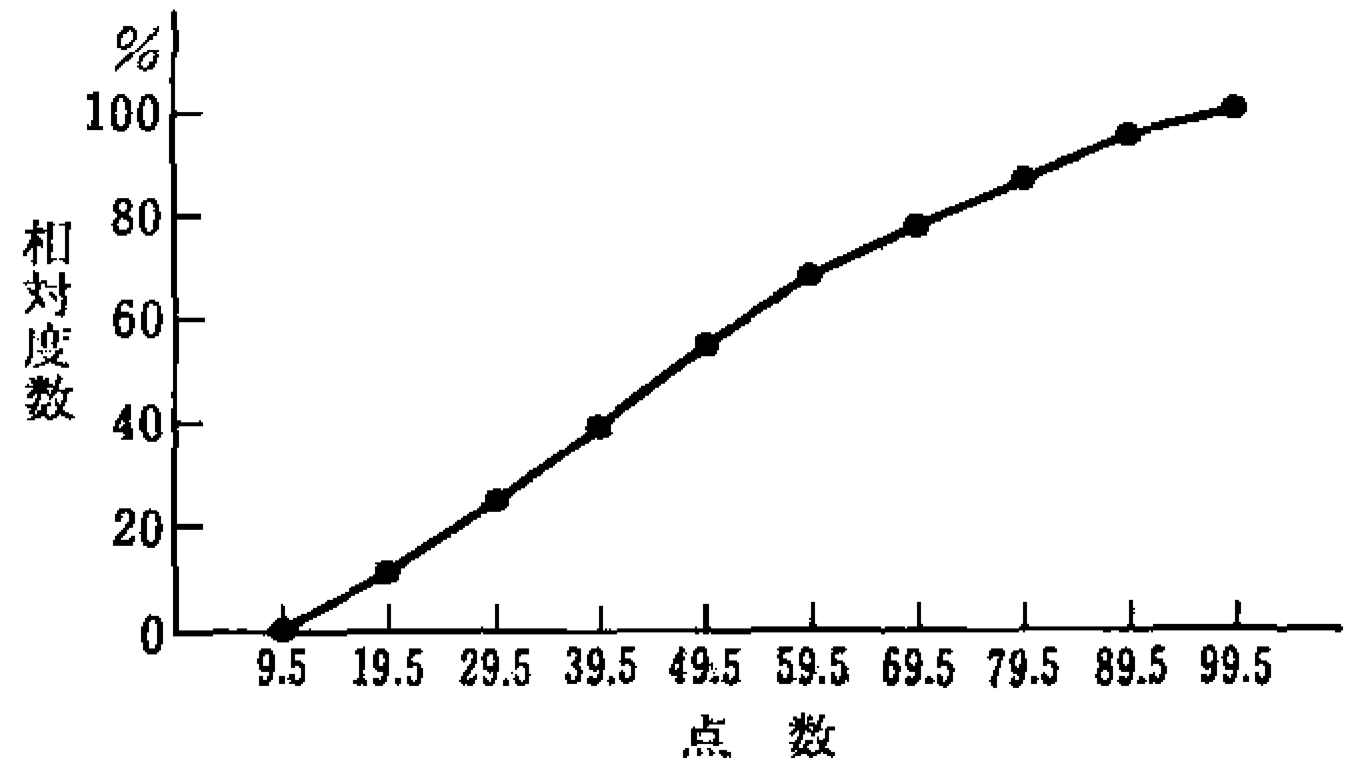


図 1·6　表 1·2 に対する累積度数分布図

問題 1. 1·1 節問題1の度数分布表に対して，度数折線，ヒストグラムおよび累積度数分布図をつくれ．

問題 2. 次の度数分布表に対してヒストグラム，累積度数分布図をつくれ．

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	30–39	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）

問題 3. 次の度数分布表に対してヒストグラムをつくれ．

100–999 人規模サービス業の年齢別給与額（昭和 48 年）（単位 1 000 円）

年齢階級	18歳未満	18–19	20–24	25–29	30–34	35–39	40–44	45–49	50–54	55–59	60–64	65歳以上
給 与 額	44.6	53.8	66.7	84.2	101.3	118.6	128.5	141.3	133.9	124.0	109.7	94.1

給与額の数字は平均月間にきまって支給する額である（第 25 回日本統計年鑑による）．

1·3 代　表　値

分布の特性を少数の**特性値**によって表現することが考えられる．これらの特性値のうちで，最も基本的なものは**代表値**である．これはなんらかの意味で分布の中心となる変量の値を表す．

(1) 代表値の種類　(a) 変量の N 個の測定値を $x_1, x_2, \cdots, x_N$ とするとき，これらの値の総和を測定値の個数 N で割ったものを，これら測定値(分布)の**算術平均** (arithmetical mean) または単に**平均値** (mean) といい，$\bar{x}$ で表す．

$$\bar{x}=\frac{x_1+x_2+\cdots+x_N}{N}=\frac{\sum_{i=1}^{N}x_i}{N}\text{[1)]}$$

N 個の測定値のうち，値 x_1 が f_1 個，値 x_2 が f_2 個,…，値 x_k が f_k 個 $(f_1+f_2+\cdots+f_k=N)$ あるときは，算術平均 $\bar{x}$ は

$$\bar{x}=\frac{f_1x_1+f_2x_2+\cdots+f_kx_k}{N}=\frac{\sum_{i=1}^{k}f_ix_i}{N}\qquad\left(\sum_{i=1}^{k}f_i=N\right)$$

によって与えられる．算術平均は，定義が簡明で，計算が容易である点で，次に述べる他の代表値に比してすぐれている．

(b) N 個の測定値 $x_1, x_2, \cdots, x_N$ が正の数であるとき，これらの積の N 乗根(正の数)をこれら測定値(分布)の**幾何平均** (geometrical mean) という．

$$G=\sqrt[N]{x_1x_2\cdots x_N}$$

例 1. 5個の測定値 3, 6, 12, 24, 48 の算術平均および幾何平均を求める．

解　算術平均は

$$\bar{x}=\frac{3+6+12+24+48}{5}=\frac{93}{5}=18.6$$

で，幾何平均は

$$G=\sqrt[5]{3\cdot6\cdot12\cdot24\cdot48}=\sqrt[5]{3^5\cdot2^{10}}=3\cdot2^2=12$$

である．

この例では，算術平均は幾何平均より大きいことがみられるが，この大小関係は一般に成り立つ．すなわち，$x_1, x_2, \cdots, x_N$ が正の数であるとき，不等式

$$\frac{x_1+x_2+\cdots+x_N}{N}\geqq\sqrt[N]{x_1x_2\cdots x_N}$$

---

1) 和 $x_1+x_2+\cdots+x_N$ を x_i の1から N までの和の略記号 $\sum_{i=1}^{N}x_i$，または簡略して $\sum_i x_i$ または $\sum x_i$ で表す．

が成り立つ [補注 1·1].

幾何平均 G の対数をとると

$$\log G = \frac{\log x_1 + \log x_2 + \cdots + \log x_N}{N}$$

となる．すなわち，幾何平均の対数は測定値の対数の算術平均に等しい．また，この関係式によって，幾何平均を対数計算によって計算することもできる．

N 個の測定値のうち，値 x_1 が f_1 個，値 x_2 が f_2 個，…，値 x_k が f_k 個 $(f_1+f_2+\cdots+f_k=N)$ あるとき，幾何平均は

$$G = \sqrt[N]{x_1^{f_1} x_2^{f_2} \cdots x_k^{f_k}} \quad \left(\sum_{i=1}^{k} f_i = N\right)$$

によって与えられる．幾何平均は，ある時点を基準点としたときの他の時点での価格の比（物価指数）とか，人口増加率とか，一般に比率などの代表値に適している．

(c)　N 個の測定値を大きさの順に並べて

$$x_1 \leqq x_2 \leqq \cdots \leqq x_N$$

としたとき，中央にくる値を**中位数**または**中央値**という．N が奇数であるときは，$(N+1)/2$ 番目の測定値 $x_{(N+1)/2}$ が中位数で，N が偶数であるときは，$N/2$ 番目の測定値と $(N/2)+1$ 番目の測定値の算術平均 $(x_{N/2}+x_{(N/2)+1})/2$ をもって中位数とする．中位数は，分布の端にかけ離れた値が現れた場合などは，そういう値にあまり左右されないという点で，算術平均よりすぐれている．

(d)　変量の測定値のうちで，度数の最も大きいものを**最頻値**(ひん)または**モード** (mode) という．これはまた**流行値**ともいわれ，需要，売買，賃金，生計費などについての分布での代表値として適することが多い．

（2） 度数分布からの平均値計算　度数分布の級の級中値を x_1, x_2, …, x_k とし，度数を f_1, f_2, …, f_k とすると，値 x_1 が f_1 個，値 x_2 が f_2 個，…，値 x_k が f_k 個であるとみなされるから，平均値（算術平均）は

$$\bar{x} = \frac{f_1x_1+f_2x_2+\cdots+f_kx_k}{N} = \frac{\sum_{i=1}^{k} f_i x_i}{N} \quad \left(\sum_{i=1}^{k} f_i = N\right)$$