1　度数分布

1.1　統計データ

調査や観測によって得られた数値等の集まりのことをデータという。データはその意味や，どの演算ができるかによって，次の 4 段階に分けられる。

名義尺度（nominal scale）　名義尺度に属する量は，カテゴリーの区別のために用いられ，数値としては意味をもたない。カテゴリーデータともいう。主な代表値のうち，最頻値のみを求めることができる。

1. 陰性　2. 陽性

順序尺度（ordinal scale）　順序尺度に属する量は，大小の区別をもち，数値の順序に意味がある。主な代表値のうち，最頻値，中央値を求めることができる。

1. 不満　2. やや不満　3. やや満足　4. 満足

間隔尺度（interval scale）　間隔尺度に属する量は，足し算や引き算について均質であり，等しい差（等しい間隔）がある一定の量を表している。主な代表値のうち，最頻値，中央値，算術平均を求めることができる。

例：　セ氏温度，　西暦年，　零点を変更できる量

比率尺度（ratio scale）　比率尺度に属する量は，掛け算や割り算についても均質であり，等しい商（等しい比率）がある一定の量を表している。主な代表値のうち，最頻値，中央値，算術平均，幾何平均を求めることができる。ただし比率尺度では正数しか扱えない。

例：　所得，　重量，　零点を変更できない量

このように 4 段階に分けるだけでなく，いくつかの分け方がある。**質的変数**とは量的でない，つまり平均値等が意味をもたないもので，名義尺度と順序尺度が含まれる。**量的変数**には，間隔尺度と比率尺度がが含まれる。**離散型変数（discrete）**とは，とびとびの値（整数値）をとる変数のことで，質的変数に相当する。**連続型変数（continuous）**とは，隙間のない値（実数値）をとる変数のことで，量的変数に相当する。**時系列データ（time series）**とは，ある一対象について，複数の時点を調査して得られたデータのこと，**横断面データ（cross section）**とは，ある一時点において，複数の対象を調査して得られたデータのこと，**パネルデータ**とは，複数時点において，複数の対象を調査して得られたデータのことである。

1.2　測定値の表現

統計で用いる数値の多くは測定値であるため，分数ではなく，小数で表示するのが望ましい。

数値の丸め方　与えられた数値の下位の桁を省略して簡単にすることを「丸める」という。切り捨て，切り上げ，四捨五入等の丸め方がある。通常は四捨五入を用いるが，切り捨てや切り上げをしなければならない場合もある。

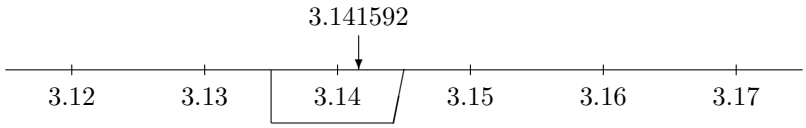
- 27 個のお菓子を 1 人あたり 4 個ずつ配るとき，お菓子を配ることができる人数は 6 人である。（切り捨て）

$$27 \div 4 = 6.75 \xrightarrow{\text{切り捨て}} 6$$

- 25 L の水を容積 4 L の容器に入れて保存するとき，必要な容器の個数は 7 個である。（切り上げ）

$$25 \div 4 = 6.25 \xrightarrow{\text{切り上げ}} 7$$

3.141592 を四捨五入して小数第 2 位まで求めると 3.14 になる。小数第 2 位まで求めるとき，小数第 3 位が 4 以下なら切り捨て，5 以上なら切り上げる。



なお，負数の丸め方は定まっていない。特に指示がない場合，符号を取り去ったものを丸めてから，再び符号を付ける。

$$3.5 \xrightarrow{\text{四捨五入}} 4, \qquad -3.5 \xrightarrow{\text{四捨五入}} -4$$

有効数字　与えられた数値のうち，意味のある桁（値が保証された桁）を有効数字という。データの各値の有効数字が 3 桁なら，それから求めた平均値の有効数字も 3 桁程度になる。末尾が 0 であっても有効数字はすべて表示するべきである。

$$45.6\text{（有効数字 3 桁）}, \qquad 12.0\text{（有効数字 3 桁）}$$

有効数字がわかりにくい場合は，指数表現を用いることができる。

$$\begin{aligned} 47000 &\to 4.7 \times 10^4\text{（有効数字 2 桁）} \\ 47000 &\to 4.70 \times 10^4\text{（有効数字 3 桁）} \\ 47000 &\to 4.700 \times 10^4\text{（有効数字 4 桁）} \end{aligned}$$

1.3　度数分布表とヒストグラム

次のデータはあるクラスの学生 40 人の試験の結果である。

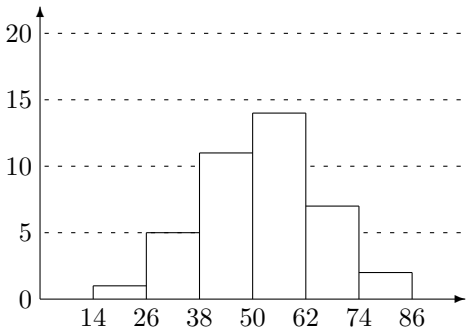
39	21	60	53	60	36	66	74	58	70	61	56	47	33	42
49	59	37	65	54	66	61	60	65	45	39	66	46	55	60
44	53	45	43	60	40	74	36	66	33					

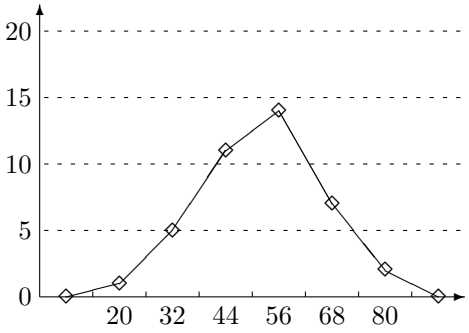
幹葉図（stem-and-leaf plot）　左側の幹（みき）は十の位以上，右側の葉（は）は一の位以下。

2	1
3	3 3 6 6 7　9 9
4	0 2 3 4 5　5 6 7 9
5	3 3 4 5 6　8 9
6	0 0 0 0 0　1 1 5 5 6　6 6 6
7	0 4 4

度数分布表・ヒストグラム・度数折れ線　上のデータから度数分布表を作成し，その表からヒストグラムと度数折れ線を作成した。[*i] [*ii]

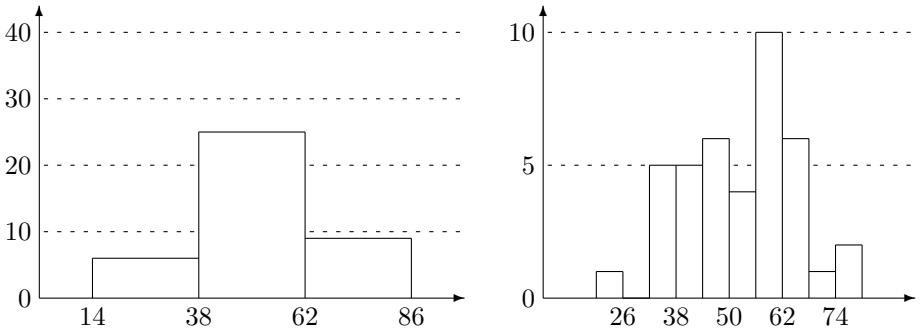
階級	階級値	度数
$14^{以上}$ $26^{未満}$	20	1
26 ～ 38	32	5
38 ～ 50	44	11
50 ～ 62	56	14
62 ～ 74	68	7
74 ～ 86	80	2
計	–	40





[*i]「階級」を級,「階級の幅」を級間隔,「階級値」を級代表値,「度数」を頻度ということもある。
[*ii]「ヒストグラム」を柱状図,「度数折れ線」を度数分布多角形ということもある。

階級の数・階級の幅の定め方　階級の数は多すぎても少なすぎても分布の形が見えにくくなるため，適切な数の階級に分けなければならない。



データの大きさ n に適する階級の数 k を知るための公式がいくつか知られている。

$$k = \sqrt{n}, \qquad k = 1 + \log_2 n \tag{1}$$

$n = 40$ の場合，1 番目の公式なら，$6 = \sqrt{36} \leqq \sqrt{40} \leqq \sqrt{49} = 7$ から，$6 \leqq k \leqq 7$，2 番目の公式なら，$5 = \log_2 32 \leqq \log_2 40 \leqq \log_2 64 = 6$ から，$6 \leqq k \leqq 7$ になる。

観測値 $x_1, x_2, \cdots, x_n$ は小さいほうから順に並べられているとする。

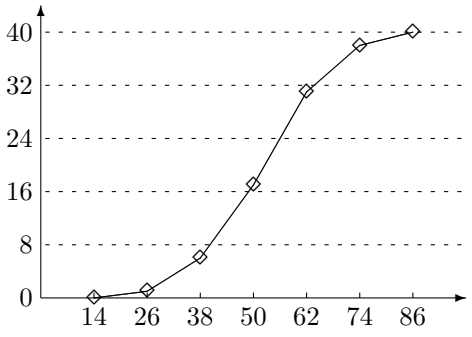
$$x_1 \leqq x_2 \leqq \cdots \leqq x_n$$

度数分布表は次の手順に従って作成するが，階級は切りの良い値になるように微調整してもよい。

1. 最小値 x_1，最大値 x_n，範囲 $R = x_n - x_1$ を求める。
2. データの大きさ n に適する階級の数 k を求める。
3. 階級の幅 w を，R/k よりわずかに大きくする。
4. 最初の階級の下限を x_1 よりわずかに小さくして，階級を決定する。
5. 度数を求める。

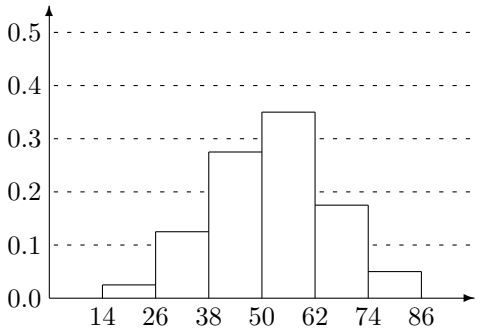
累積度数分布表　値が小さいほうからの度数の和を累積度数という。累積度数は常に増加（非減少）する。

階級	度数	累積度数
14以上 26未満	1	1
26 ～ 38	5	6
38 ～ 50	11	17
50 ～ 62	14	31
62 ～ 74	7	38
74 ～ 86	2	40
計	40	–



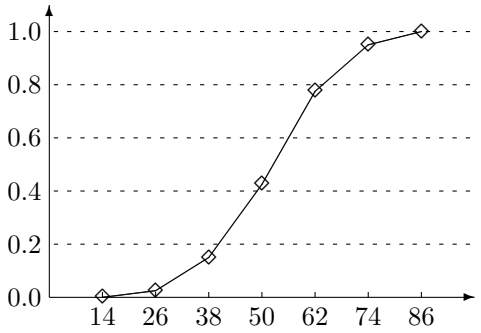
相対度数分布表・累積相対度数分布表　各度数を，度数の総和で割った値のことを相対度数という。相対度数の総和は 1 になる。

階級	度数	相対度数
14 ～ 26	1	0.025
26 ～ 38	5	0.125
38 ～ 50	11	0.275
50 ～ 62	14	0.350
62 ～ 74	7	0.175
74 ～ 86	2	0.050
計	40	1.000



値が小さいほうからの相対度数の和を累積相対度数という。累積相対度数は常に増加（非減少）で，0 以上 1 以下の値になる。

階級	相対度数	累積相対度数
14 ～ 26	0.025	0.025
26 ～ 38	0.125	0.150
38 ～ 50	0.275	0.425
50 ～ 62	0.350	0.775
62 ～ 74	0.175	0.950
74 ～ 86	0.050	1.000
計	1.000	–



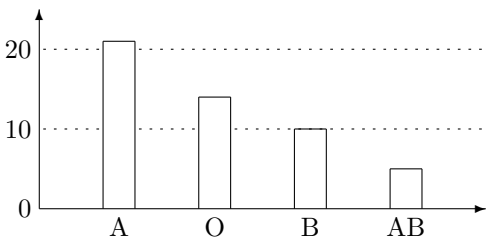
1.4　棒グラフ

次のデータはあるクラスの学生 50 人の血液型である。

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

このような名義尺度では階級を設定できないため，ヒストグラムではなく棒グラフ（bar chart）を用いる。棒グラフはヒストグラムと異なり，棒の太さや間隔を任意に決めてよい。

値	度数
A	21
O	14
B	10
AB	5
計	50



参考文献

統計学入門（基礎統計学）
東京大学教養学部統計学教室（編）東京大学出版会 978-4-13-042065-5

統計学
久保川 達也 （著） 東京大学出版会 978-4-13-062921-8

はじめての統計学
道家 暎幸（共著） コロナ社 978-4-339-06113-0

確率統計 新版 （新版数学シリーズ）
岡本 和夫 （ほか著） 実教出版 978-4-407-32171-5

統計学序論 改訂版
山本 義郎 （著） 東海大学出版部 978-4-486-02133-9

確率統計（高専テキストシリーズ）
上野 健爾（監修） 森北出版 978-4-627-05561-2

基本統計学 第４版
宮川 公男 （著） 有斐閣 978-4-641-16455-0

新統計入門
小寺 平治 （著） 裳華房 978-4-7853-1099-8

Schaum's Outline of Introduction to Probability and Statistics
Seymour Lipschutz （著） McGraw-Hill Education 978-0-07-176249-6

A Dictionary of Statistics
Graham Upton （著） Oxford Univ Pr 978-0-19-967918-8

www5e.biglobe.ne.jp/~emm386/statistics/