

1

データの整理と表現

1-1 データの整理

連続変量と離散変量 身長や交通事故の件数などのように、ある集団において観測の対象となる特性を数量で表したものを**変量**という。変量には、身長や体重のように連続的な値をとる**連続変量**と、交通事故の件数や家族の人数のようにとびとびの値しかとらない**離散変量**とがある。

度数分布表 連続変量のデータを整理するときには、データをできるだけ同じ幅の区間に区切って分類する。このとき、各区間を**階級**といい、その中央の値を**階級値**という。また、各階級に分類されたデータの個数を**度数**という。階級の幅はデータ全体の傾向が読みとれるように、適当な大きさに選ぶ。階級または階級値に度数を対応させたものを**度数分布**といい、分類の結果でき上がった表を**度数分布表**という。

離散変量のデータの分類は、その変量をとる値の個数を数えるだけでよいかから簡単であるが、変量のとる値が多いときは、連続変量の場合のようにいくつかの階級を作って分類する。

累積度数分布表 各階級の度数の累計を**累積度数**といい、それらを表にまとめたものを**累積度数分布表**という。

相対度数分布表 各階級の度数をデータの総数で割った値を**相対度数**といい、これを表にしたものを**相対度数分布表**という。

度数分布表, 相対度数分布表, 累積度数分布表

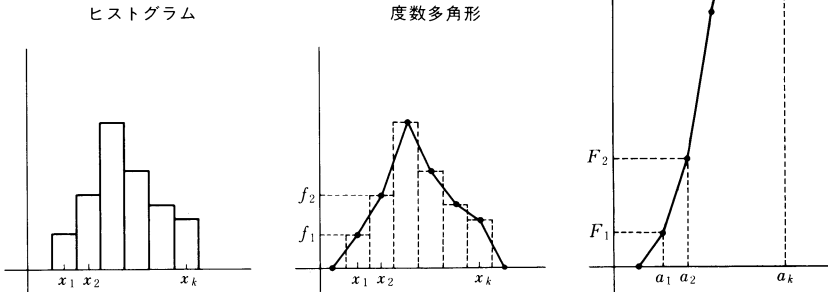
階級	階級値	度数	相対度数	累積度数
$a_0 \sim a_1$	x_1	f_1	f_1/n	$F_1 = f_1$
$a_1 \sim a_2$	x_2	f_2	f_2/n	$F_2 = f_1 + f_2$
\vdots	\vdots	\vdots	\vdots	\vdots
$a_{k-1} \sim a_k$	x_k	f_k	f_k/n	$F_k = f_1 + \dots + f_k = n$
計		n	1	

1-2 グラフによる表現

ヒストグラム 度数分布表を柱状グラフに表したもので, 各階級の上に立
てる長方形の面積は階級の度数に比例させる.

度数多角形 ヒストグラムの各長方形の上辺の
中点を結んで得られる折れ線グラフ.

累積度数多角形 累積度数分布表をグラフに表
したもので, 階級の上方限界値と累積度数を座標に
もつ点を結んで得られる.



茎葉図 たとえば, 10個のデータ (1.67, 1.82, 1.75, 1.63, 1.72, 1.79,
1.84, 1.60, 1.73, 1.75) が与えられたとき, 各データを次のように“茎”に当たる
(1.6, 1.7, 1.8)と“葉”に当たる3桁目の数に分解し, 茎を縦に葉を横に記録す
ることでデータを表現したものを**茎葉図**(または**幹葉図**)という.

茎	葉
1.6	0 3 7
1.7	2 3 5 5 9
1.8	2 4

1-3 代表値と散布度

(a) 生のデータからの場合.

n 個のデータを x_1, x_2, \dots, x_n とする.

(b) 度数分布表からの場合.

n 個のデータが k 個の階級に分類されているとして、階級値を x_1, x_2, \dots, x_k , 度数を f_1, f_2, \dots, f_k , 累積度数を F_1, F_2, \dots, F_k とする.

1-3-1 代表値

データ全体の中心的傾向を表す値で、平均値、メジアン、モードなどがある.

平均 (または平均値)

$$(a) \quad \bar{x} = \frac{1}{n} \sum_{i=1}^n x_i$$

$$(b) \quad \bar{x} = \frac{1}{n} \sum_{i=1}^k x_i f_i$$

平均の簡便計算法 階級の幅を c , 仮平均を x_0 とする. ここで, 変換 $u_i = \frac{x_i - x_0}{c}$ ($i=1, 2, \dots, k$) によって, x を u に換えれば

$$\bar{x} = x_0 + c\bar{u} = x_0 + c \frac{\sum_{i=1}^k u_i f_i}{n}$$

メジアン

(a) n 個のデータを大きさの順に並べたとき, n が奇数ならば小さい方から $\frac{n+1}{2}$ 番目の値を, n が偶数ならば, 中央にくる 2 つの値の平均をメジアンまたは中央値という.

(b) この場合, メジアン M_e は次式より求められる.

$$M_e = a_{j-1} + c \frac{\frac{n}{2} - F_e}{f_j}$$

ここで, a_{j-1} はメジアンを含む階級の下方限界値,

f_j はメジアンを含む階級の度数,

F_e はメジアンを含む階級より前のすべての階級の度数の和,

すなわち $F_e = f_1 + f_2 + \dots + f_{j-1}$,

c は階級の幅.

モード

(a) データの中で最も多く現れている値をモードまたは最頻値という.

(b) 最大の度数をもつ階級の階級値.

1-3-2 散布度

データの散らばりを表す特性値で、標準偏差、分散、範囲などがある。

分散

$$(a) \quad s^2 = \frac{1}{n} \sum_{i=1}^n (x_i - \bar{x})^2$$

$$= \frac{1}{n} \sum_{i=1}^n x_i^2 - \bar{x}^2$$

$$(b) \quad s^2 = \frac{1}{n} \sum_{i=1}^k (x_i - \bar{x})^2 f_i$$

$$= \frac{1}{n} \sum_{i=1}^k x_i^2 f_i - \bar{x}^2$$

分散の簡便計算法

$$s^2 = c^2 \left(\frac{1}{n} \sum_{i=1}^k u_i^2 f_i - \bar{u}^2 \right)$$

ただし、

$$u_i = \frac{x_i - x_0}{c}, \quad \bar{u} = \frac{1}{n} \sum u_i f_i$$

標準偏差 分散の正の平方根を標準偏差という。

$$(a) \quad s = \sqrt{s^2} = \sqrt{\frac{1}{n} \sum_{i=1}^n (x_i - \bar{x})^2}$$

$$= \sqrt{\frac{1}{n} \sum_{i=1}^n x_i^2 - \bar{x}^2}$$

$$(b) \quad s = \sqrt{s^2} = \sqrt{\frac{1}{n} \sum_{i=1}^k (x_i - \bar{x})^2 f_i}$$

$$= \sqrt{\frac{1}{n} \sum_{i=1}^k x_i^2 f_i - \bar{x}^2}$$

標準偏差の簡便計算法

$$s = c \sqrt{\frac{1}{n} \sum u_i^2 f_i - \bar{u}^2}$$

範囲 データの中の最大値と最小値の差を範囲という。

四分位偏差 n 個のデータを大きさの順に並べたとき、小さい方から $\left[\frac{n}{4} \right] + 1$ 番目の値を第1四分位数、小さい方から $\left[\frac{3n}{4} \right] + 1$ 番目の値を第3四分位数といい、それぞれ Q_1 、 Q_3 で表す。 $Q_3 - Q_1$ を四分位範囲、 $\frac{1}{2}(Q_3 - Q_1)$ を四分位偏差という。(注: $[x]$ は実数 x の整数部分)

例題

例題 1 (離散データの度数分布)

次の数はある町での30日間の救急車の出動回数の記録である。このデータから度数分布表と累積度数分布表を作れ。また、(a) 平均値、(b) メジアン、(c) モードを求めよ。

2 6 1 1 2 0 3 0 2 3
 3 0 4 2 1 1 0 3 4 0
 5 1 4 5 1 0 5 0 2 0

解 データの分類の結果、

x	検数マーク	f	F
0	正下	8	8
1	正 ^一	6	14
2	正	5	19
3	正 ^一	4	23
4	下	3	26
5	下	3	29
6	一	1	30
計		30	

したがって、度数分布表と累積度数分布表は

x	f	xf	x	F
0	8	0	0	8
1	6	6	1	14
2	5	10	2	19
3	4	12	3	23
4	3	12	4	26
5	3	15	5	29
6	1	6	6	30
計	30	61		

(a) 平均値は、 $\bar{x} = \frac{61}{30} \approx 2$ (回)。

(b) 累積度数分布表より小さい方から15番目と16番目の値はともに2で

あるから、メジアンは2(回).

(c) 最大度数は8だから、モードは0.

例題 2 (ヒストグラム・度数多角形・累積度数多角形)

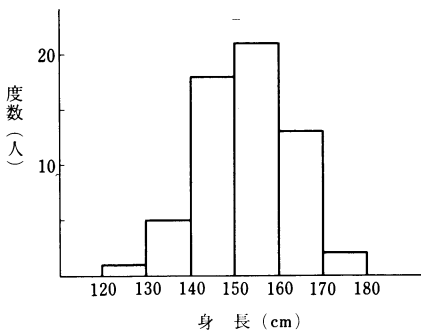
次の度数分布において、階級の真の限界、階級値および相対度数を与えよ。また、ヒストグラム、度数多角形、および累積度数多角形を図示せよ。

階級	度数
120-129	1
130-139	5
140-149	18
150-159	21
160-169	13
170-179	2
計	60

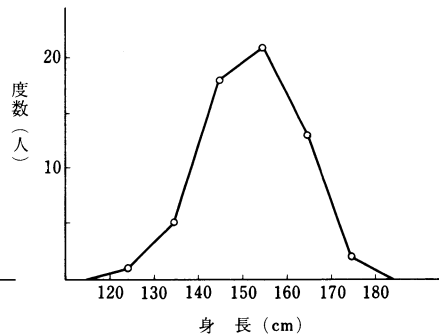
解

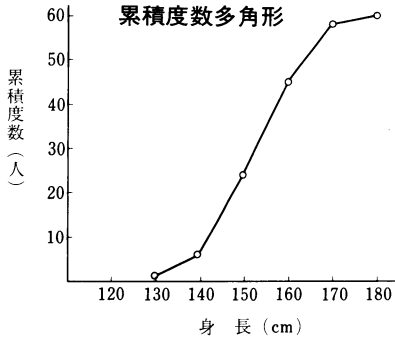
階級	階級の真の限界	階級値	度数	累積度数
120-129	119.5-129.5	124.5	1	1
130-139	129.5-139.5	134.5	5	6
140-149	139.5-149.5	144.5	18	24
150-159	149.5-159.5	154.5	21	45
160-169	159.5-169.5	164.5	13	58
170-179	169.5-179.5	174.5	2	60

ヒストグラム



度数多角形





例題 3 (茎葉図と度数分布)

男子学生 70 人の身長 (cm) を測って、次の結果を得た。

169	176	169	169	167	170	172
163	182	173	175	163	168	173
171	164	168	162	167	163	177
159	161	167	164	176	164	159
172	169	164	174	165	177	175
157	175	154	169	155	176	183
167	168	159	168	171	162	173
172	170	173	171	166	168	168
171	164	161	165	168	164	156
174	175	168	161	180	181	166

このデータから、身長

- (a) 茎葉図,
- (b) 度数分布,
- (c) 平均と標準偏差,
- (d) メジアンとモード

を求めよ。

解 (a) 茎葉図

茎	葉
15	4 5 6 7 9 9 9
16	1 1 1 2 2 3 3 3 4 4 4 4 4 5 5 6 6 7 7 7 7 8 8 8 8 8 8 8 8 9 9 9 9 9
17	0 0 1 1 1 1 2 2 2 3 3 3 3 4 4 5 5 5 5 6 6 6 7 7
18	0 1 2 3

(b) 最大値は 183 で、最小値は 154 であるから、範囲は $183 - 154 = 29$. 階級の幅を 5 cm にとれば、7 個の階級となる.

			度数分布	
階級	検数マーク	階級の真の限界	階級値	度数
150-154	—	149.5-154.5	152	1
155-159	正—	154.5-159.5	157	6
160-164	正正正	159.5-164.5	162	14
165-169	正正正正—	164.5-169.5	167	21
170-174	正正正	169.5-174.5	172	15
175-179	正正	174.5-179.5	177	9
180-184	正	179.5-184.5	182	4
			計	70

(c) 簡便計算法によって、平均と標準偏差を求める. 階級の幅は $c=5$, 仮平均を $x_0=167$ にとり

$$u = \frac{x - 167}{5}$$

によって, x を u に変換する.

階級値	度数				
x	f	u	uf	$u^2 f$	
152	1	-3	-3	9	
157	6	-2	-12	24	
162	14	-1	-14	14	
167	21	0	0	0	
172	15	1	15	15	
177	9	2	18	36	
182	4	3	12	36	
計	70		16	134	

$$\bar{u} = \frac{\sum u_i f_i}{n}$$

$$= \frac{16}{70} \doteq 0.229$$

$$\bar{x} = x_0 + c\bar{u}$$

$$= 167 + 5 \times 0.229 \doteq \mathbf{168.1 \text{ (cm)}}$$

$$s = c \sqrt{\frac{\sum u_i^2 f_i}{n} - \bar{u}^2}$$

$$= 5 \sqrt{\frac{134}{70} - \left(\frac{16}{70}\right)^2} \doteq \mathbf{6.8 \text{ (cm)}}$$

(d) メジアンは公式より直接求める $n=70$, $c=5$, $a_{j-1}=164.5$, $F_e=1+6+14=21$, $f_j=21$ であるから

$$M_e = 164.5 + 5 \times \frac{35 - 21}{21} \doteq \mathbf{167.8 \text{ (cm)}}$$

モードは度数分布表より, $\mathbf{167 \text{ (cm)}}$.

例題 4 (階級の幅が等しくないときのヒストグラム)

ある日、図書館で本を借りた人の年齢の分布は次のようであった。

年齢	8-12	13-20	21-60	61-64
人数	9	22	35	14

- (a) このデータをヒストグラムで表せ。
 (b) 80人の年齢の平均と標準偏差を求めよ。

解 (a) ここでの変量は年齢であるから階級の真の限界を定めるときに注意が必要である。たとえば、階級8-12は8歳ちょうどから13歳未満の人を含む(下表参照)この問題は階級の幅が等間隔でないから、ヒストグラムをかくとき、柱の面積が各度数に比例するようにしなければならない。これは、

$$\text{柱の高さ} \times \text{階級の幅} = \text{柱の面積} \propto \text{階級の度数}$$

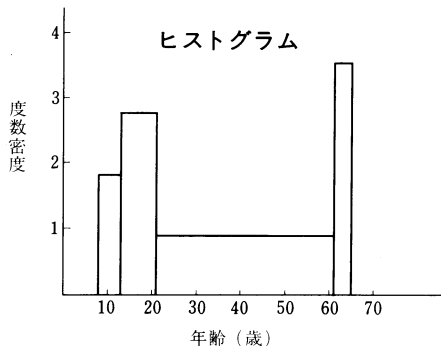
の関係より、

$$\text{柱の高さ} \propto \frac{\text{階級の度数}}{\text{階級の幅}}$$

よって、まず与えられた度数分布より $\frac{\text{階級の度数}}{\text{階級の幅}}$ (これを**度数密度**という) を求め、柱の高さをこの値に比例するようにとればよい。

階級	階級の真の限界	階級の幅	度数	度数密度
8-12	8-13	5	9	1.8
13-20	13-21	8	22	2.75
21-60	21-61	40	35	0.875
61-64	61-65	4	14	3.5

よって、



(b) 前の表より,

階級の真の限界	階級値 x	度数 f	xf	x^2f
8-13	10.5	9	94.5	992.25
13-21	17.0	22	374.0	6358.00
21-61	41.0	35	1435.0	58835.00
61-65	63.0	14	882.0	55566.00
計		80	2785.5	121751.25

よって, 平均 $\bar{x} = \frac{2785.5}{80} \doteq 34.8$ (歳)

標準偏差 $s = \sqrt{\frac{121751.25}{80} - 34.8^2} \doteq 17.6$ (歳)

例題 5 (平均, 標準偏差, メジアン, モード)

次の表は 1988 年全国高校野球選手権大会での全試合の勝敗の結果を示したものである。この表より

- 得点差の度数分布と累積度数分布を作れ。
- 得点差の平均と標準偏差を求めよ。
- 得点差のメジアンとモードを求めよ。

3-0	10-1	6-2	2-1
7-4	4-1	5-1	4-1
8-0	9-0	3-2	4-3
6-0	3-2	10-1	6-3
4-3	7-4	3-2	3-2
9-3	4-3	9-3	5-0
4-0	5-4	8-1	7-3
2-1	4-1	4-0	2-1
19-1	2-1	4-3	9-3
5-2	6-4	12-1	4-2
4-2	5-3	1-0	5-1
8-4	7-5	4-2	1-0

解 得点差 x を求めると

3 3 8 6 1 6 4 1 18 3 2 4
 9 3 9 1 3 1 1 3 1 2 2 2
 4 4 1 9 1 6 7 4 1 11 1 2
 1 3 1 3 1 5 4 1 6 2 4 1

度数を f , 累積度数を F で表し, 次のようにデータを分類して, 度数分布と累積度数分布を作る.

(a) 度数分布 累積度数分布

x	検数マーク	f	F	x	f	x	F
1	正正正	15	15	1	15	1	15
2	正 ^一	6	21	2	6	2	21
3	正下	8	29	3	8	3	29
4	正 ^丁	7	36	4	7	4	36
5	一	1	37	5	1	5	37
6	正 ^フ	4	41	6	4	6	41
7	一	1	42	7	1	7	42
8	一	1	43	8	1	8	43
9	下	3	46	9	3	9	46
11	一	1	47	11	1	11	47
18	一	1	48	18	1	18	48
計		48		計	48		

(b) 度数分布表より

x	f	xf	x^2f
1	15	15	15
2	6	12	24
3	8	24	72
4	7	28	112
5	1	5	25
6	4	24	144
7	1	7	49
8	1	8	64
9	3	27	243
11	1	11	121
18	1	18	324
計	48	179	1193

よって,

$$\text{平均 } \bar{x} = \frac{179}{48} = 3.7(\text{点})$$

$$\text{標準偏差 } s = \sqrt{\frac{1193}{48} - \left(\frac{179}{48}\right)^2} = 3.3(\text{点})$$

(c) 累積度数分布表より、小さい方から 24 番目と 25 番目の値はともに 3 だから、メジアンは 3 (点).

モードは、度数分布表から 1 (点).

例題 6 (メジアン)

次の数の平均値は 7 で、モードは 6 である。これら数のメジアンはいくつか。

$$8, 10, x, 6, y, z, 8, 21$$

解 モードが 6 であることから、 x, y, z のうち少なくとも 2 つは 6 でなければならない。3 つとも 6 とすると、平均値が 7 にならないから 2 つが 6 である。いま、 x, y を 6 とすると平均値は 7 だから

$$7 = \frac{8+10+6+6+6+z+8+21}{8} = \frac{65+z}{8} \Rightarrow z = -9$$

よって、これらの数を大きさの順に並べると

$$-9, 6, 6, 6, 8, 8, 10, 21$$

 ↑ ↑

ゆえに、メジアンは $\frac{6+8}{2} = 7$.

例題 7 (メジアンと四分位偏差)

高速道路のある地点で、走行中の車 400 台の時速 (km/h) を測って、次の度数分布を得た。

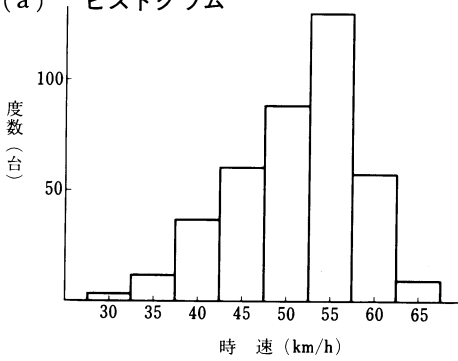
階級値	30	35	40	45	50	55	60	65
度数	3	12	37	61	89	130	58	10

この結果を

- (a) ヒストグラム,
- (b) 累積度数多角形

で表せ。求めた累積度数多角形からメジアンと四分位偏差を求めよ。

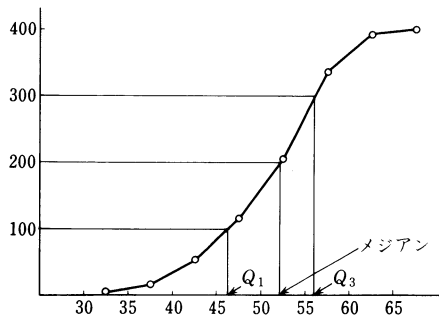
解 (a) ヒストグラム



(b) 累積度数分布表は、

階級値	度数	階級の真の限界	累積度数
30	3	27.5-32.5	3
35	12	32.5-37.5	15
40	37	37.5-42.5	52
45	61	42.5-47.5	113
50	89	47.5-52.5	202
55	130	52.5-57.5	332
60	58	57.5-62.5	390
65	10	62.5-67.5	400
計	400		

累積度数多角形



この図からメジアンは 52.

Q_1 は 46, Q_3 は 56. よって, 四分位偏差は $Q_3 - Q_1 = 56 - 46 = 10$.

例題 8 (合併データの平均と標準偏差)

公式

$$\sum_{i=1}^n (x_i - \bar{x})^2 = \sum_{i=1}^n x_i^2 - n\bar{x}^2$$

を証明せよ。

ある中学校の3年生は女子生徒263人と男子生徒282人からなり、女子の身長は平均155.5 cm、標準偏差4.0 cmで、男子の身長は平均163.0 cm、標準偏差4.4 cmである。この学校の3年生全員の身長の平均と標準偏差を求めよ。

解 公式の証明：

$$\begin{aligned} \sum (x_i - \bar{x})^2 &= \sum (x_i^2 - 2x_i\bar{x} + \bar{x}^2) \\ &= \sum x_i^2 - 2\bar{x} \sum x_i + n\bar{x}^2 \\ &= \sum x_i^2 - 2n\bar{x}^2 + n\bar{x}^2 \quad (\sum x_i = n\bar{x} \text{ より}) \\ &= \sum x_i^2 - n\bar{x}^2 \end{aligned}$$

3年生全員の身長の平均は

$$\bar{x} = \frac{263 \times 155.5 + 282 \times 163.0}{545} = 159.4 \text{ (cm)}$$

標準偏差の計算には上記の公式を使う。標準偏差を s とすると、この公式より

$$\sum x_i^2 = n(\bar{x}^2 + s^2)$$

女子生徒の場合、この値は

$$\sum x_i^2 = 263(155.5^2 + 4.0^2) = 6363613.75$$

男子生徒の場合、この値は

$$\sum x_i^2 = 282(163.0^2 + 4.4^2) = 7497917.52$$

よって、3年生全員の身長の標準偏差は

$$\begin{aligned} s &= \sqrt{\frac{1}{n} \sum (x_i - \bar{x})^2} \\ &= \sqrt{\frac{\sum x_i^2}{n} - \bar{x}^2} \\ &= \sqrt{\frac{6363613.75 + 7497917.52}{545} - 159.4^2} = \sqrt{25.64} = 5.1 \text{ (cm)} \end{aligned}$$

1章の問題

1.1 次の各階級に対して、階級の真の限界、階級値、階級の幅を示せ。

- (a) 20-29
- (b) 150-, 180-
- (c) 1.6-3.5
- (d) (-8)-(-5)

1.2 (a) 数 $\{3, 5, 4, 1, 7\}$ の平均と標準偏差を求めよ。この結果を用いて、計算によらずに次の数の平均と標準偏差を求めよ。

- (i) 13, 15, 14, 11, 17
- (ii) -2, 0, -1, -4, 2
- (iii) $a+3, a+5, a+4, a+1, a+7$
- (iv) 0.3, 0.5, 0.4, 0.1, 0.7
- (v) $3a+b, 5a+b, 4a+b, a+b, 7a+b$

(b) 次のデータから、平均、メジアン、モードを求めよ。ただし、 $a < b < c < d < e$ 。

$$a, a, a, a, a, b, b, b, b, b, b, c, c, c, c, d, d, d, e, e$$

1.3 次の数はある電話交換局が30秒間隔で延べ25分間に受けた電話呼び出し数の記録である。

2	3	5	2	2	0	1	3	2	4
3	2	2	3	1	1	3	3	4	6
2	4	1	3	5	4	5	2	3	6
3	5	2	5	3	0	2	0	5	4
4	2	1	2	4	3	6	4	2	0

このデータから呼び出し数の度数分布を作り、度数多角形をかけ。また、呼び出し数の平均値、メジアン、モードを求めよ。

1.4 以下の数値は100人の生徒のIQのデータである。55-64, 65-74, …を階級に選び、これらデータを分類して

- (a) 度数分布表を作れ。
- (b) IQの平均と標準偏差を求めよ。
- (c) IQのメジアンを求めよ。

81	106	81	116	105	107	110	84	78	91
109	98	106	133	108	109	105	102	97	77
100	101	120	73	90	99	90	143	100	102
82	109	90	97	101	116	103	84	104	119
107	102	96	101	88	80	85	124	117	100
86	81	91	91	124	111	108	82	97	99
108	101	58	95	106	106	91	118	107	66
107	121	108	79	94	82	93	104	128	100
101	100	102	94	89	90	108	114	92	111
81	94	72	118	93	103	104	103	100	92

1.5 観測値 x , 5, y , 13 は大きさの順に並んでいる. これらの平均値は 7 で, メジアンは 6 である. 分散を求めよ.

1.6 3つのかごに, それぞれ 6 個, 5 個, 4 個のりんごが入っている. 第 1 のかごのりんごの重さの平均値は 220 g で, 第 2 のかごのりんごの重さの平均値は 280 g で, 第 3 のかごのりんごの重さは

$$247 \text{ g}, \quad 250 \text{ g}, \quad 239 \text{ g}, \quad 264 \text{ g}$$

である. 第 3 のかごのりんごはどれもが第 1 のかごのりんごより重く, 第 2 のかごのりんごより軽い. そのとき, これら 15 個のりんごの重さの平均値とメジアンとを求めよ. 次に, これら 3 つのかごの重さを p , q , r とする.

$$p+q+r=240, \quad p^2+q^2+r^2=20000$$

のとき, かごの重さの標準偏差を求めよ.

1.7 2 個のサイコロを同時に 100 回投げ, 出た目の和 x を観測して次の度数分布を得た. x の平均と標準偏差を求めよ.

x	2	3	4	5	6	7	8	9	10	11	12	計
f	2	4	9	10	16	17	15	11	8	6	2	100

1.8 ある中学校の遅刻者 100 人の遅刻時間(分)の度数分布は次のようであった.

遅刻時間(分)	0-2	3-7	8-12	13-17	18-27	計
度数	40	37	13	5	5	100

- (a) この分布のヒストグラムをかけ。
(b) 遅刻時間の平均と標準偏差を求めよ。

1.9 ある試験を受けた200人の生徒の得点の度数分布が、次のように与えられたとき、

- (a) この分布の平均と標準偏差を求めよ。
(b) 累積度数分布表を求め、累積度数多角形をかけ。

階級	1-20	21-30	31-40	41-50	51-60	61-70	71-80	81-100
度数	4	15	43	55	39	31	10	3

1.10 4個の観測値の平均は3.13、標準偏差は0.15である。これに、さらに6個の観測値

3.19 2.86 2.93 3.15 3.14 3.21

が追加された。これら10個の観測値の平均と標準偏差を求めよ。

1.11 4個の数の平均は5、分散は2で、別な6個の数の平均は7、分散は3である。これら10個の数の平均と分散を求めよ。