

第5章 データの分析

●Check! (p.280)

1

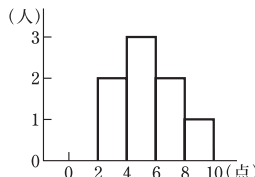
8人の生徒に10点満点のテストを行ったところ、次のような結果になった。

6, 2, 5, 9, 3, 7, 5, 4(点)

- 階級幅を2点として度数分布表を作り、ヒストグラムをかけ、ただし、最初の階級を0点以上2点未満とする。
- 最も度数の多い階級の階級値を求めよ。

(1)

階級(点)	度数(人)
0以上 2未満	0
2以上 4未満	2
4以上 6未満	3
6以上 8未満	2
8以上 10以下	1
合計	8



◀ 得点 x に対して階級幅を

⋮

$2 \leq x < 4$

$4 \leq x < 6$

⋮

ととるか,

⋮

$2 < x \leq 4$

$4 < x \leq 6$

⋮

ととるかで違ってくるが、
解答のように、～以上～未満
とすることが多い。

注 10点は最後の階級に含めた。

- 最も度数の多い階級は、4点以上6点未満で、その階級値は、5点

2

ある小学校の通学団の学年構成を調べたところ、6年生…1人、3年生…1人、2年生…1人、1年生…2人であった。次の表を完成させよ。

階級(学年)	度数(人)	累積度数(人)	相対度数	累積相対度数
1以上3未満				
3以上5未満				
5以上6以下				
合計				

階級(学年)	度数(人)	累積度数(人)	相対度数	累積相対度数
1以上3未満	3	3	0.60	0.60
3以上5未満	1	4	0.20	0.80
5以上6以下	1	5	0.20	1.00
合計	5		1.00	

◀ 相対度数の合計欄には1.00
と書けばよい。

3

次の表は、あるクラスの生徒8名の100m走の計測結果である。

出席番号	1	2	3	4	5	6	7	8
記録(秒)	13.2	19.5	12.7	11.2	13.6	13.8		12.3

出席番号7の生徒のデータの値が消失してしまっているが、平均値が13.8秒であることがわかっている。

- 出席番号7の生徒のデータの値を求めよ。
- このデータの中央値を求めよ。

226 第5章 データの分析

- (1) 出席番号7の生徒のデータの値を
- x
- とおくと、

$$\frac{1}{8}(13.2+19.5+12.7+11.2+13.6+13.8+x+12.3) \\ =13.8 \\ x+96.3=110.4 \quad x=14.1$$

よって、求めるデータの値は、14.1 秒

- (2) 生徒8名のデータの値を小さい順に並べると、

11.2, 12.3, 12.7, 13.2, 13.6, 13.8, 14.1, 19.5
中央値は、4番目と5番目のデータの値の平均値となるから、

$$\frac{13.2+13.6}{2}=13.4 \text{ (秒)}$$

4

次のデータの最頻値と四分位数を求め、箱ひげ図をかけ。

- (1) 6, 5, 4, 7, 8, 6, 6, 8, 5

(2)	階級値	2	6	10	14	計
	度数	2	3	2	1	8

- (1) 最も度数が多い値は6である。

よって、最頻値は、6

データの値を小さい順に並べると、

4, 5, 5, 6, 6, 6, 7, 8, 8

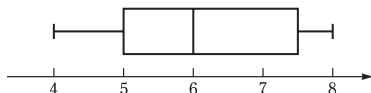
第2四分位数は中央値だから、5番目の値で、6第1四分位数は、2番目の値5と3番目の値5の平均値だから、

$$\frac{5+5}{2}=5$$

第3四分位数は、7番目の値7と8番目の値8の平均値だから、

$$\frac{7+8}{2}=7.5$$

また、最小値は4、最大値は8だから、箱ひげ図は次のようになる。



- (2) 最も度数が多い階級の階級値は6である。

よって、最頻値は、6

データを小さい順に並べると、

2, 2, 6, 6, 6, 10, 10, 14

第2四分位数は中央値だから、4番目の値6と5番目の値6の平均値で、

$$\frac{6+6}{2}=6$$

第1四分位数は、2番目の値2と3番目の値6の平均値だから、

$$\frac{2+6}{2}=4$$

第3四分位数は、6番目の値10と7番目の値10の平

◀ データの値の個数は9個(奇数)

◀ 1～4番目のデータの中央値

◀ 6～9番目のデータの中央値

◀ データの値の個数は8個(偶数)

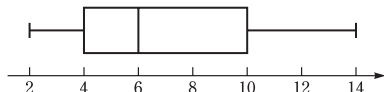
◀ 1～4番目のデータの中央値

◀ 5～8番目のデータの中央値

均値だから、

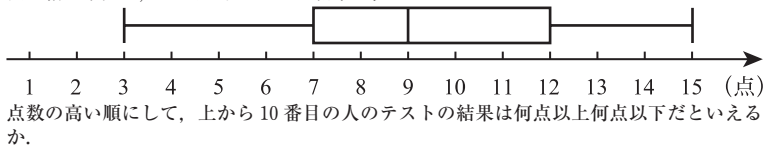
$$\frac{10+10}{2}=10$$

また、最小値は2、最大値は14だから、箱ひげ図は次のようになる。



5

次の箱ひげ図は、36人の小テストの結果を表している。



上から10番目の人は中央値と第3四分位数の間に位置するから、テストの結果は9点以上12点以下

6

あるレストランでランチサービスを始めた。最近5日間の注文数を調べたところ、次の通りであった。

90, 110, 140, 70, 90 (食)

- (1) 注文数の平均値を求めよ。
- (2) 注文数の分散と標準偏差を求めよ。

$$(1) \frac{1}{5}(90+110+140+70+90)=\frac{500}{5}=100$$

より、平均値は、100食

注文数 (食)	偏差 (食)	偏差平方
90	-10	100
110	10	100
140	40	1600
70	-30	900
90	-10	100

よって、分散は、

$$\frac{1}{5}(100+100+1600+900+100)=\frac{2800}{5}=560$$

標準偏差は、 $\sqrt{560}=4\sqrt{35}$ (食)

◀ このように偏差平方の表を作っておくと、計算がラク。

◀ 分散は偏差平方の平均値

◀ (標準偏差) = $\sqrt{(\text{分散})}$

7

変数 x の n 個のデータ x_1, x_2, \dots, x_n がある。

x の平均値 \bar{x} が6であるとき、変数 $y = \frac{1}{4}x + 20$ の平均値 \bar{y} を求めよ。

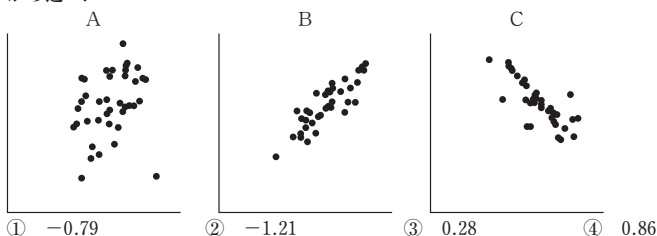
$$\bar{y} = \frac{1}{4}\bar{x} + 20 \text{ より,}$$

$$\bar{y} = \frac{1}{4} \cdot 6 + 20 = 21.5$$

よって、平均値 $\bar{y} = 21.5$

8

次の散布図 A～C のそれぞれに対し、最も近いと思われる相関係数 r の値を、下の①～④から選べ。



A の散布図では、傾向があまりみられず、相関はほとんどない。

B の散布図では、一方が増えると他方も直線的に増える傾向がはっきりとみられるので、強い正の相関がある。

C の散布図では、一方が増えると他方が直線的に減る傾向がはっきりとみられるので、強い負の相関がある。

よって、 A ③ B ④ C ①

◀ 相関係数 r は 0 に近い。

◀ 相関係数 r は 1 に近い。

◀ 相関係数 r は -1 に近い。
 $-1 \leq r \leq 1$ より、②ではない。

9

2 つの変数 x と y が次の表で与えられている。

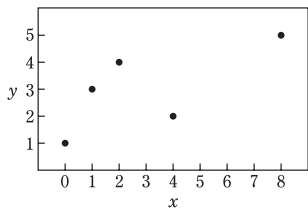
x	4	8	0	2	1
y	2	5	1	4	3

- (1) x を横軸に、 y を縦軸にとって散布図をかけ。
- (2) x と y の平均値をそれぞれ求めよ。
- (3) 次の表を完成させよ。

x	y	$x - \bar{x}$	$y - \bar{y}$	$(x - \bar{x})^2$	$(y - \bar{y})^2$	$(x - \bar{x})(y - \bar{y})$
4	2					
8	5					
0	1					
2	4					
1	3					

- (4) x と y の相関係数 r を求めよ。

(1)



- (2) x の平均値は、 $\frac{1}{5}(4+8+0+2+1) = \frac{15}{5} = 3$
 y の平均値は、 $\frac{1}{5}(2+5+1+4+3) = \frac{15}{5} = 3$

(3)

x	y	$x - \bar{x}$	$y - \bar{y}$	$(x - \bar{x})^2$	$(y - \bar{y})^2$	$(x - \bar{x})(y - \bar{y})$
4	2	1	-1	1	1	-1
8	5	5	2	25	4	10
0	1	-3	-2	9	4	6
2	4	-1	1	1	1	-1
1	3	-2	0	4	0	0

(4) x の分散は, $\frac{1}{5}(1+25+9+1+4) = \frac{40}{5} = 8$

y の分散は, $\frac{1}{5}(1+4+4+1+0) = \frac{10}{5} = 2$

x と y の共分散は, $\frac{1}{5}(-1+10+6-1+0) = \frac{14}{5}$

よって, 相関係数 r は,

$$r = \frac{14}{5} \div (\sqrt{8} \times \sqrt{2}) = \frac{14}{5 \times 4} = \frac{7}{10}$$

◀ $r = \frac{S_{xy}}{S_x S_y}$

●練習

140

花子さんに目分量で10 cmの線分を40本かいてもらい、実際の長さを計測した。それを記録したものが以下の数値である。次の表を完成させ、花子さんがかいた線分の傾向として、①、②の文章が正しいといえるか。

9.3, 8.8, 9.7, 8.8, 9.3, 9.1, 8.9, 8.3, 8.6, 9.4,
9.9, 10.9, 8.7, 10.6, 10.3, 8.4, 9.6, 11.2, 10.8, 10.0,
8.6, 9.4, 10.4, 8.3, 8.8, 10.0, 8.6, 9.4, 9.0, 9.0,
8.8, 9.8, 8.7, 9.6, 9.2, 8.2, 9.5, 8.9, 8.5, 9.9 (cm)

階級 (cm)	度数 (本)	相対度数	累積度数	累積相対度数
8.0 以上 8.5 未満	4	0.10	4	0.10
8.5 以上 9.0 未満				
9.0 以上 9.5 未満				
9.5 以上 10.0 未満				
10.0 以上 10.5 未満				
10.5 以上 11.0 未満				
11.0 以上 11.5 未満				
11.5 以上 12.0 未満				
合計	40	1.00		

- ① 度数が最も多い階級は、9.5 cm 以上 10.0 cm 未満である。
② 半分程度は 10 cm の誤差 5 mm の長さである。

階級 (cm)	度数 (本)	相対度数	累積度数	累積相対度数
8.0 以上 8.5 未満	4	0.10	4	0.10
8.5 以上 9.0 未満	12	0.30	16	0.40
9.0 以上 9.5 未満	9	0.23	25	0.63
9.5 以上 10.0 未満	7	0.18	32	0.81
10.0 以上 10.5 未満	4	0.10	36	0.91
10.5 以上 11.0 未満	3	0.08	39	0.99
11.0 以上 11.5 未満	1	0.03	40	1.00
11.5 以上 12.0 未満	0	0	40	1.00
合計	40	1.00		

度数が最も多い階級は、8.5 cm 以上 9.0 cm 未満。したがって、①は正しくない。

また、10 cm の誤差 5 mm の長さ、つまり 9.5 cm 以上 10.5 cm 未満の割合は、相対度数より、 $0.18 + 0.10 = 0.28$ だから、全体の 28 %。

したがって、②も正しいとはいえない。

◀ (相対度数)
$$= \frac{(\text{その階級の度数})}{(\text{度数の合計})}$$

◀ 四捨五入して得られる相対度数の合計が 1 にならない場合でも、各階級の相対度数を調節する必要はない。

◀ 相対度数の合計欄には 1.00 と書けばよい。

141

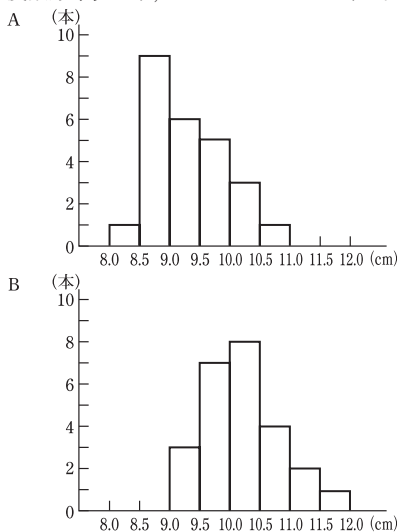
A, B の2人に目分量で 10 cm の線分を 25 本ずつかいてもらい、実際の長さを計測した。それを記録したものが、以下の数値である。

A 8.7, 9.1, 10.3, 8.6, 9.4, B 10.0, 10.7, 11.0, 10.1, 9.8,
9.7, 8.4, 9.0, 8.8, 9.9, 10.5, 9.4, 10.4, 11.5, 11.3,
8.9, 10.7, 8.8, 9.5, 8.7, 9.7, 9.3, 10.5, 9.8, 10.2,
9.8, 9.3, 8.5, 8.9, 10.4, 10.0, 9.9, 10.5, 10.4, 9.8,
8.6, 10.2, 9.7, 9.2, 9.0 (cm) 9.8, 9.3, 10.3, 9.6, 10.0 (cm)

A, B のかく線分を比較したとき、どのようなことがいえるか。

A	階級 (cm)		度数 (本)	B	階級 (cm)		度数 (本)
	8.0 以上 8.5 未満	8.5 以上 9.0 未満			8.0 以上 8.5 未満	8.5 以上 9.0 未満	
	1	9			0	0	
	9.0 以上 9.5 未満	6			9.0 以上 9.5 未満	3	
	9.5 以上 10.0 未満	5			9.5 以上 10.0 未満	7	
	10.0 以上 10.5 未満	3			10.0 以上 10.5 未満	8	
	10.5 以上 11.0 未満	1			10.5 以上 11.0 未満	4	
	11.0 以上 11.5 未満	0			11.0 以上 11.5 未満	2	
	11.5 以上 12.0 未満	0			11.5 以上 12.0 未満	1	
	合計	25			合計	25	

度数分布表より、ヒストグラムをかくと次のようになる。



よって、Aの方がBより短い線分をかく傾向が強い。
10 cm により近い線分を多くかいたのはBの方である。

◀ ヒストグラムをかく際は、最小値を含む階級の左、最大値を含む階級の右にそれぞれ度数 0 の階級を付け加えておくことが多い。ヒストグラムに加えて、度数折れ線(度数分布多角形)とよばれる折れ線グラフをかくときのためである。

◀ 本問においては A, B の比較を行うため、階級のとり方をそろえている。

142

次のデータは、ある商品の20日間の生産数を1日ごとに並べたものである。

20 23 21 20 21 20 19 20 17 22
17 22 18 19 21 20 18 19 19 18 (個)

- (1) このデータについて、下の表を完成させ、仮平均を20個として、販売数の平均値を求めよ。

生産数(個)	17	18	19	20	21	22	23
日数(日)							

- (2) さらに10日を加えた30日間の生産数の平均値を、20個以上とするためには、残り10日間で何個以上生産する必要があるか。

(1)

生産数(個)	17	18	19	20	21	22	23
日数(日)	2	3	4	5	3	2	1

仮平均を20個とすると、20日間のデータの平均値は、

$$\begin{aligned}
 & 20 + \frac{1}{20} \{ (-3) \times 2 + (-2) \times 3 + (-1) \times 4 \\
 & \quad + 0 \times 5 + 1 \times 3 + 2 \times 2 + 3 \times 1 \} \\
 & = 20 + \frac{1}{20} (-6 - 6 - 4 + 0 + 3 + 4 + 3) \\
 & = 20 - \frac{6}{20} = 20 - 0.3 = 19.7 \text{ (個)}
 \end{aligned}$$

- (2) 20日間の生産数は、 $19.7 \times 20 = 394$ (個)
1日に20個生産したときの30日間の生産数は、
 $20 \times 30 = 600$ (個)

したがって、残り10日間で x 個生産するとして、30日間の生産数の平均値が20個以上になるとき、

$$394 + x \geq 600$$

よって、 $x \geq 206$ より、**206個以上**

$$\begin{aligned}
 \bar{x} &= x_0 + \frac{1}{n} \{ (x_1 - x_0) \\
 & \quad + (x_2 - x_0) + \cdots + (x_n - x_0) \}
 \end{aligned}$$

$$\triangleleft (\text{平均値}) \times (\text{総度数}) = (\text{総和})$$

5

143

次の表は、生徒100人の試験の得点(0以上の整数)の累積度数をまとめたもので、各生徒の得点は明らかではない。このとき、次の問いに答えよ。

得点(点)	90以上	80以上	70以上	60以上	50以上	40以上	30以上	20以上	10以上
度数(人)	0	7	20	38	63	85	93	98	100

- (1) 80点以上90点未満を1つの階級として、各階級値に対する度数分布表を作成せよ。
(2) (1)で作成した度数分布表における平均値を求めよ。
(3) 生徒100人の実際の得点の平均値のとり得る値の範囲を求めよ。

(1)

階級値(点)	85	75	65	55	45	35	25	15
度数(人)	7	13	18	25	22	8	5	2

- (2) 仮平均を最頻値55点とし、次のような表を作る。

階級値(点)	85	75	65	55	45	35	25	15
(階級値)-55	30	20	10	0	-10	-20	-30	-40
度数(人)	7	13	18	25	22	8	5	2

表より、平均値は、

$$\begin{aligned}
 & 55 + \frac{1}{100} \{ 30 \times 7 + 20 \times 13 + 10 \times 18 \\
 & \quad + 0 \times 25 + (-10) \times 22 + (-20) \times 8 \\
 & \quad + (-30) \times 5 + (-40) \times 2 \}
 \end{aligned}$$

△ 階級値は、各階級の両端の平均値である。

△ 計算を簡単にするため、仮平均の考え方をを用いた。

$$\begin{aligned}
 &= 55 + \frac{1}{100}(210 + 260 + 180 + 0 - 220 - 160 - 150 - 80) \\
 &= 55 + \frac{40}{100} = 55 + 0.4 = 55.4 \text{ (点)}
 \end{aligned}$$

- (3) 各データの値が各階級の最大値をとるとき、すなわち、各データの値が各階級の階級値より4点だけ大きい値となるとき、平均値は最大となるから、平均値の最大値は、
 $55.4 + 4 = 59.4 \text{ (点)}$

同様に、各データの値が各階級の階級値より5点だけ小さい値となるとき、平均値は最小となるから、平均値の最小値は、

$$55.4 - 5 = 50.4 \text{ (点)}$$

よって、生徒 100 人の実際の平均値のとり得る値の範囲は、50.4 点以上 59.4 点以下である。

◀たとえば、80 点以上 90 点未満の階級は、階級値 85 点
 最大値 89 点
 最小値 80 点
 データの値は整数値をとるので、最大値が存在する。

144

次の表は、生徒 39 人を A 班と B 班に分け、10 点満点の試験を行った結果である。

得点 (点)	0	1	2	3	4	5	6	7	8	9	10	平均値
A 班 (人)	0	0	1	1	3	3	4	2	2	2	1	x
B 班 (人)	0	1	1	2	1	y	z	5	4	2	0	6

- (1) 表中の x, y, z の値を求めよ。また、39 人全員の点数の平均値を求めよ。
 (2) それぞれの班のデータについて、中央値、四分位範囲を求めよ。

- (1) x は A 班の平均値である。

A 班の人数は、 $1 + 1 + 3 + 3 + 4 + 2 + 2 + 2 + 1 = 19$ (人) であるから、

$$\begin{aligned}
 x &= \frac{1}{19}(2 \times 1 + 3 \times 1 + 4 \times 3 + 5 \times 3 \\
 &\quad + 6 \times 4 + 7 \times 2 + 8 \times 2 + 9 \times 2 + 10 \times 1) \\
 &= \frac{1}{19}(2 + 3 + 12 + 15 + 24 + 14 + 16 + 18 + 10) \\
 &= \frac{114}{19} = 6
 \end{aligned}$$

B 班の人数は、 $39 - 19 = 20$ (人) であるから、B 班の人数と平均値より、

$$1 + 1 + 2 + 1 + y + z + 5 + 4 + 2 = 20 \quad \cdots \cdots ①$$

$$\begin{aligned}
 \frac{1}{20}(1 \times 1 + 2 \times 1 + 3 \times 2 + 4 \times 1 + 5 \times y \\
 + 6 \times z + 7 \times 5 + 8 \times 4 + 9 \times 2) &= 6 \quad \cdots \cdots ②
 \end{aligned}$$

- ①, ②より、 $y + z = 4$, $5y + 6z = 22$ であるから、

$$y = 2, z = 2$$

また、39 人全員の点数の平均値は、

$$\frac{1}{39}(6 \times 19 + 6 \times 20) = \frac{6 \times 39}{39} = 6 \text{ (点)}$$

- (2) A 班のデータの値を小さい順に並べると、

2 3 4 4 4 5 5 5 6 6 6 6 6 7 7 8 8 9 9 10

A 班の中央値は、6 点

A 班の第 1 四分位数は 4 点、第 3 四分位数は 8 点であるから、A 班の四分位範囲は、
 $8 - 4 = 4$ (点)

◀ A 班も B 班も平均値が 6 点であることから、39 人全員の平均値も 6 点

◀ A 班の人数は 19 人であるから、A 班のデータの値を小さい順に並べたとき、1 番目から 9 番目までの中央値が第 1 四分位数
 11 番目から 19 番目までの中央値が第 3 四分位数

B班のデータの値を小さい順に並べると、

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

B班の中央値は、 $\frac{7+7}{2}=7$ (点)

B班の第1四分位数は、 $\frac{4+5}{2}=4.5$ (点)、第3四分

位数は、 $\frac{8+8}{2}=8$ (点) であるから、B班の四分位範囲は、

$$8-4.5=3.5 \text{ (点)}$$

◀ B班の人数は20人であるから、B班のデータの値を小さい順に並べたとき、1番目から10番目までの中央値が第1四分位数
11番目から20番目までの中央値が第3四分位数

145

次の表は、ある野球チームの最近30試合の得点の結果である。

得点 (点)	0	1	2	3	4	5	6	7	8
試合数 (試合)	3	6	4	2	8	4	2	0	1

このデータについて、四分位範囲を求めよ。また、このデータの箱ひげ図をかけ。ただし、箱ひげ図には平均値もかき入れるものとする。

第2四分位数は中央値で、15番目の値3と16番目の値4

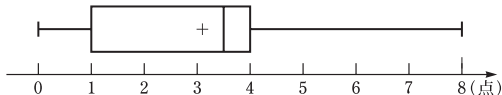
の平均値だから、 $\frac{3+4}{2}=3.5$ (点)

第1四分位数は、8番目の値だから、1点

第3四分位数は、23番目の値だから、4点

よって、四分位範囲は、 $4-1=3$ (点)

箱ひげ図は、



◀ 最小値 0 点、最大値 8 点

$$\begin{aligned} \text{平均値} &= \frac{1}{30} \times 92 \\ &= 3.06\cdots \approx 3.1 \text{ (点)} \end{aligned}$$

146

次のデータは、16日間、毎日の血圧を測定した値である。このデータの箱ひげ図をかけ。ただし、外れ値がある場合は外れ値を示し、箱ひげ図をかけ。

50, 55, 58, 64, 65, 66, 66, 67, 68, 69, 70, 72, 73, 78, 80, 89 (mmHg)

与えられたデータを小さい順に並べると、次のようになる。

50, 55, 58, 64, 65, 66, 66, 67,

68, 69, 70, 72, 73, 78, 80, 89

よって、最大値は89 mmHg、最小値は50 mmHg

第1四分位数は、 $\frac{64+65}{2}=64.5$ (mmHg)

第2四分位数は、 $\frac{67+68}{2}=67.5$ (mmHg)

第3四分位数は、 $\frac{72+73}{2}=72.5$ (mmHg)

四分位範囲は、 $72.5-64.5=8$ (mmHg)

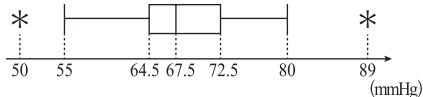
$8 \times 1.5 = 12$ より、外れ値の目安は、

$64.5-12=52.5$ より、52.5 以下か、

$72.5+12=84.5$ より、84.5 以上のデータである。

したがって、50 mmHg, 89 mmHg が外れ値となる。

よって、箱ひげ図は次のようになる。

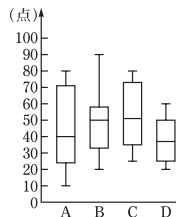


◀ 外れ値を除いて最大の値が最大値、最小の値が最小値となる。

147

右の図は、生徒数がいずれも40人のA組、B組、C組、D組に、100点満点の同じテストを行った結果を箱ひげ図に表したものである。

- (1) 60点以上の生徒が10人以下の組はどれか。
- (2) 上位10人が70点以上で、その散らばりが最も小さい組はどれか。
- (3) 全体の散らばりが最も大きい組はどれか。



- (1) 各組の生徒数は40人であるから、第3四分位数は点数の高い方から10番目と11番目の得点の平均値である。60点以上の生徒が10人以下の組は、第3四分位数が60点未満となるから、**B組とD組**である。

- (2) 上位10人は、第3四分位数以上の生徒に相当する。上位10人の散らばりは、箱ひげ図の上側のひげの長さで判断できる。

第3四分位数が70点以上であるのはA組とC組で、さらに、箱ひげ図の上側のひげの長さが短い方であるから、**C組**である。

- (3) 範囲(箱ひげ図の全長)が大きいのはA組とB組で、**A組とB組の範囲は同じである。**

また、**四分位範囲(箱の長さ)はB組よりA組の方が大きい。**

つまり、集団全体の散らばりが最も大きい組は、**A組**である。

◀ 点数の低い方から30番目と31番目

◀ 範囲が同じ場合、四分位範囲の大小から散らばりを考える。

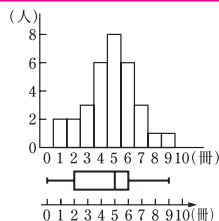
5

148

右のヒストグラムは、生徒32人のA組の1か月の読書冊数を表したものである。

- (1) A組について、箱ひげ図を作れ。
- (2) 右の図は、生徒32人のB組の1か月の読書冊数の箱ひげ図である。A組とB組の箱ひげ図から読み取れることとして次の①～③は正しいといえるか。

- ① A組とB組の合計64人の生徒の25%は6冊以上の本を読んでいる。
- ② A組の下位25%の人数は、B組の下位25%の人数よりも多い。
- ③ A組とB組の最頻値はどちらも5冊である。

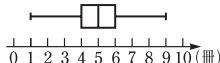


- (1) 中央値は、 $\frac{5+5}{2}=5$ (冊)

第1四分位数は、 $\frac{4+4}{2}=4$ (冊)

第3四分位数は、 $\frac{6+6}{2}=6$ (冊)

よって、箱ひげ図は右のようになる。



- (2) ① A組とB組はともに第3四分位数は6冊であるから、ともに25%は1か月に6冊以上の本を読んでいる。

したがって、A組とB組の合計64人のうち25%

◀ 32人いるので、読書冊数が少ない順に並べたとき、中央値は、16番目と17番目の平均値
第1四分位数は、8番目と9番目の平均値
第3四分位数は、24番目と25番目の平均値

は1か月に6冊以上の本を読んでいるから、正しい。

- ② A組とB組はともに32人であるから、その25%の人数も同じになり、正しくない。

- ③ B組の最頻値は箱ひげ図からは読み取れないので、正しくない。

149

次の表は、P組とQ組で同じテストを行った結果である。

得点(点)	0	1	2	3	4	5	6	7	8	9	10	合計
P組(人)	0	2	2	3	6	7	5	3	1	1	0	30
Q組(人)	1	6	2	3	2	5	6	1	3	1	0	30

この表から、P組とQ組の標準偏差をそれぞれ求めよ。また、P組とQ組の得点の散らばり进行比较するとどのようなことがいえるか。

P組の平均値は、

$$\begin{aligned} & \frac{1}{30}(1 \times 2 + 2 \times 2 + 3 \times 3 + 4 \times 6 \\ & \quad + 5 \times 7 + 6 \times 5 + 7 \times 3 + 8 \times 1 + 9 \times 1) \\ &= \frac{1}{30}(2 + 4 + 9 + 24 + 35 + 30 + 21 + 8 + 9) \\ &= \frac{142}{30} = \frac{71}{15} \text{ (点)} \end{aligned}$$

したがって、P組の分散は、

$$\begin{aligned} & \frac{1}{30}(1^2 \times 2 + 2^2 \times 2 + 3^2 \times 3 + 4^2 \times 6 + 5^2 \times 7 \\ & \quad + 6^2 \times 5 + 7^2 \times 3 + 8^2 \times 1 + 9^2 \times 1) - \left(\frac{71}{15}\right)^2 \\ &= \frac{1}{30}(2 + 8 + 27 + 96 + 175 + 180 + 147 + 64 + 81) - \left(\frac{71}{15}\right)^2 \\ &= \frac{780}{30} - \left(\frac{71}{15}\right)^2 = 26 - \frac{71^2}{15^2} = \frac{5850 - 5041}{15^2} = \frac{809}{15^2} \end{aligned}$$

よって、P組の標準偏差は、 $\sqrt{\frac{809}{15^2}} = \frac{\sqrt{809}}{15}$ (点)

Q組の平均値は、

$$\begin{aligned} & \frac{1}{30}(0 \times 1 + 1 \times 6 + 2 \times 2 + 3 \times 3 + 4 \times 2 \\ & \quad + 5 \times 5 + 6 \times 6 + 7 \times 1 + 8 \times 3 + 9 \times 1) \\ &= \frac{1}{30}(0 + 6 + 4 + 9 + 8 + 25 + 36 + 7 + 24 + 9) \\ &= \frac{128}{30} = \frac{64}{15} \text{ (点)} \end{aligned}$$

したがって、Q組の分散は、

$$\begin{aligned} & \frac{1}{30}(0^2 \times 1 + 1^2 \times 6 + 2^2 \times 2 + 3^2 \times 3 + 4^2 \times 2 + 5^2 \times 5 \\ & \quad + 6^2 \times 6 + 7^2 \times 1 + 8^2 \times 3 + 9^2 \times 1) - \left(\frac{64}{15}\right)^2 \\ &= \frac{1}{30}(0 + 6 + 8 + 27 + 32 + 125 + 216 + 49 + 192 + 81) \\ & \quad - \left(\frac{64}{15}\right)^2 \\ &= \frac{736}{30} - \left(\frac{64}{15}\right)^2 = \frac{368}{15} - \frac{64^2}{15^2} = \frac{5520 - 4096}{15^2} = \frac{1424}{15^2} \end{aligned}$$

$$\begin{aligned} & \triangleleft (x \text{ の分散}) \\ &= (x^2 \text{ の平均値}) \\ & \quad - (x \text{ の平均値})^2 \end{aligned}$$

よって、Q組の標準偏差は、 $\sqrt{\frac{1424}{15^2}} = \frac{4\sqrt{89}}{15}$ (点)

標準偏差はQ組の方が大きいので、Q組はP組よりデータの散らばりが大きいといえる。

◀ 標準偏差が大きいほどデータの散らばりは大きい。

150

次の変数 x のデータについて、 $u = \frac{x-60}{5}$ において得られる新しい変数 u の標準偏差 s_u を計算し、変数 x の標準偏差 s_x を求めよ。
40, 40, 55, 50, 60, 65, 75, 75, 80, 85

$u = \frac{x-60}{5}$ より、右のような表を作る。

これより、 u の平均値 \bar{u} は、
$$\bar{u} = \frac{1}{10} \{(-4) + (-4) + (-1) + (-2) + 0 + 1 + 3 + 3 + 4 + 5\}$$

$$= \frac{5}{10}$$

u^2 の平均値 $\bar{u^2}$ は、

$$\begin{aligned} \bar{u^2} &= \frac{1}{10} (16 + 16 + 1 + 4 + 0 + 1 + 9 + 9 + 16 + 25) \\ &= \frac{97}{10} \end{aligned}$$

u の分散 s_u^2 は、

$$s_u^2 = \bar{u^2} - (\bar{u})^2 = \frac{97}{10} - \left(\frac{5}{10}\right)^2 = \frac{970 - 25}{100} = \frac{945}{100}$$

u の標準偏差 s_u は、 $s_u = \sqrt{\frac{945}{100}} = \frac{3\sqrt{105}}{10}$

以上より、 x の標準偏差 s_x は、

$$s_x = 5s_u = 5 \times \frac{3\sqrt{105}}{10} = \frac{3\sqrt{105}}{2}$$

x	$x-60$	u	u^2
40	-20	-4	16
40	-20	-4	16
55	-5	-1	1
50	-10	-2	4
60	0	0	0
65	5	1	1
75	15	3	9
75	15	3	9
80	20	4	16
85	25	5	25

◀ 表を作ると考えやすい。

◀ $x = au + b$, すなわち、
 $u = \frac{x-b}{a}$ のとき、
 $s_x = |a|s_u$

151

例題 151 において、さらに生徒 10 人がテスト C を行ったところ、C の分散が $\frac{23}{5}$ で、A と C の共分散が $-\frac{27}{5}$ であった。A と C の相関係数を求めよ。また、A と C の散布図としてふさわしいものを、例題 151 の①～③から選べ。

A の得点の標準偏差は $\sqrt{\frac{46}{5}}$ 点

C の得点の標準偏差は $\sqrt{\frac{23}{5}}$ 点

A と C の得点の共分散は $-\frac{27}{5}$ より、求める相関係数は、

$$\frac{-\frac{27}{5}}{\sqrt{\frac{46}{5}} \times \sqrt{\frac{23}{5}}} = -\frac{27}{23\sqrt{2}} = -\frac{27\sqrt{2}}{46}$$

相関係数は負で、負の相関があるから、散布図は①か③のいずれかである。

$$-\frac{27\sqrt{2}}{46} \div -0.83 \text{ だから、散布図は、 } \textcircled{3}$$

◀ 負の相関があるとき、散布図上の点は全体的に右下がりになる。

152

右の図は、8人の生徒に行った漢字の読みと書きを問うテスト（ともに10点満点）の得点の散布図で、読みの得点を横軸に、書きの得点を縦軸にとったものである。

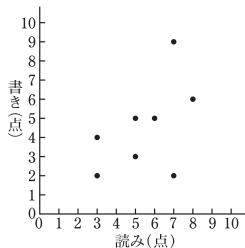
- (1) 次の表は、8人の生徒の出席番号、読みと書きの得点をまとめた表である。空欄をうめ、表を完成させよ。

出席番号	1	2	3	4	5	6	7	8	平均値
読み(点)	3	8	3			7		5	
書き(点)	2		4	5	3	2	9	5	

- (2) この8人の読みと書きの得点の標準偏差がそれぞれ $\sqrt{3}$ 点、 $\frac{\sqrt{19}}{2}$ 点で、共分散が $\frac{15}{8}$ である。

読みと書きの相関係数 r を求めよ。

- (3) 読みの採点にはミスがなかったが、書きの採点にミスがあり、出席番号1, 2, 3, 4の生徒の書きの点数がそれぞれ1点ずつ加算された。変更後の読みと書きの相関係数 R を求めよ。



(1)

出席番号	1	2	3	4	5	6	7	8	平均値
読み(点)	3	8	3	6	5	7	7	5	$\frac{11}{2}$
書き(点)	2	6	4	5	3	2	9	5	$\frac{9}{2}$

(2) $r = \frac{15}{8} \div \left(\sqrt{3} \times \frac{\sqrt{19}}{2} \right) = \frac{15}{4\sqrt{57}} = \frac{5\sqrt{57}}{76}$

- (3) 変更前と変更後の書きの点数を表にすると、次のようになる。

出席番号	1	2	3	4	5	6	7	8	平均値
書き(変更前)	2	6	4	5	3	2	9	5	$\frac{9}{2}$
書き(変更後)	3	7	5	6	3	2	9	5	5

変更後の書きの平均値は、5点になる。

変更後の書きの分散は、

$$\begin{aligned} & \frac{1}{8} \{ (3-5)^2 + (7-5)^2 + (5-5)^2 + (6-5)^2 \\ & \quad + (3-5)^2 + (2-5)^2 + (9-5)^2 + (5-5)^2 \} \\ & = \frac{1}{8} (4+4+0+1+4+9+16+0) \\ & = \frac{38}{8} = \frac{19}{4} \end{aligned}$$

◀ $r = \frac{s_{xy}}{s_x s_y}$

◀ 平均値が変化しているので、例題152(3)のような計算の工夫はできない。

したがって、変更後の書きの標準偏差は、

$$\sqrt{\frac{19}{4}} = \frac{\sqrt{19}}{2} \text{ (点)}$$

変更後の読みと書きの共分散は、

$$\begin{aligned} & \frac{1}{8} \left\{ \left(3 - \frac{11}{2}\right)(3-5) + \left(8 - \frac{11}{2}\right)(7-5) + \left(3 - \frac{11}{2}\right)(5-5) \right. \\ & + \left(6 - \frac{11}{2}\right)(6-5) + \left(5 - \frac{11}{2}\right)(3-5) + \left(7 - \frac{11}{2}\right)(2-5) \\ & \quad \left. + \left(7 - \frac{11}{2}\right)(9-5) + \left(5 - \frac{11}{2}\right)(5-5) \right\} \\ & = \frac{1}{8} \left\{ \left(-\frac{5}{2}\right) \times (-2) + \frac{5}{2} \times 2 + \left(-\frac{5}{2}\right) \times 0 + \frac{1}{2} \times 1 \right. \\ & \quad \left. + \left(-\frac{1}{2}\right) \times (-2) + \frac{3}{2} \times (-3) + \frac{3}{2} \times 4 + \left(-\frac{1}{2}\right) \times 0 \right\} \\ & = \frac{1}{8} \left(5 + 5 + 0 + \frac{1}{2} + 1 - \frac{9}{2} + 6 + 0 \right) = \frac{13}{8} \end{aligned}$$

よって、変更後の読みと書きの相関係数 R は、

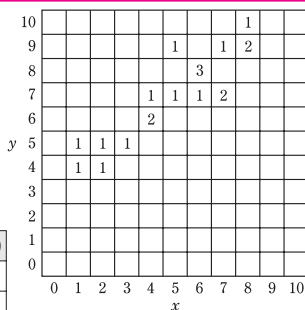
$$R = \frac{13}{8} \div \left(\sqrt{3} \times \frac{\sqrt{19}}{2} \right) = \frac{13}{4\sqrt{57}} = \frac{13\sqrt{57}}{228}$$

153

右の図は、生徒 20 人に行った漢字の読みと書きのテスト (ともに 10 点満点) の結果をまとめたものである。読みの得点 x を横軸に、書きの得点 y を縦軸にとっている。図の中の数値は x, y の値の組に対応する人数を表している。

- (1) 各生徒の得点について、 $x+y$ の最大値と、 $|x-y|$ の最大値を求めよ。
- (2) 図をもとに、次の表を完成させよ。また、各テストの得点の平均値を求めよ。

得点 (点)	0	1	2	3	4	5	6	7	8	9	10
読み (人)											
書き (人)											



- (3) (2) の表を使って各テストの標準偏差を求めると、読みは $\sqrt{5}$ 点、書きは $\sqrt{3}$ 点で、読みと書きの得点の共分散は 3.45 であった。読みと書きの得点の相関係数 r の値を四捨五入して小数第 2 位まで求めよ。ただし、 $\sqrt{15} = 3.873$ とする。
- (4) 右の表は、別の 5 人の生徒 A, B, C, D, E に同じ問題のテストを行った結果である。5 人の読みと書きの得点の平均値は、それぞれ 20 人の得点の平均値と同じであった。20 人にこの 5 人を加えた合計 25 人の生徒に関する読みと書きの得点の相関係数 R の値を四捨五入して小数第 2 位まで求めよ。ただし、 $\sqrt{5} = 2.236$ とする。
- (5) これらのテストの結果について、次の①～③は正しいといえるか。
 - ① 生徒 25 人の得点について、読みと書きの平均値からの散らばり具合は同じである。
 - ② 生徒 20 人の読みと書きの得点の正の相関はやや強いが、A～E の 5 人が加わると正の相関は少し弱まる。
 - ③ 生徒 25 人の読みの得点が一律に 1 点下がれば、25 人の読みと書きの得点の相関係数 R の値はより大きくなる。

5 人の生徒	A	B	C	D	E
読みの得点	4	2	2	8	9
書きの得点	8	4	6	7	10

- (1) $x+y$ が最大となる生徒の x, y の値の組は、
 $(x, y) = (8, 10)$ より、 $x+y$ の最大値は 18 である。
 $|x-y|$ が最大となる生徒の x, y の値の組は、

$(x, y) = (1, 5), (5, 9)$ より, $|x - y|$ の最大値は 4 である.

(2)

得点 (点)	0	1	2	3	4	5	6	7	8	9	10
読み (人)	0	2	2	1	3	2	4	3	3	0	0
書き (人)	0	0	0	0	2	3	2	5	3	4	1

読みの得点の平均値は,

$$\begin{aligned} & \frac{1}{20}(1 \times 2 + 2 \times 2 + 3 \times 1 + 4 \times 3 \\ & \quad + 5 \times 2 + 6 \times 4 + 7 \times 3 + 8 \times 3) \\ &= \frac{1}{20}(2 + 4 + 3 + 12 + 10 + 24 + 21 + 24) \\ &= \frac{100}{20} = 5 \text{ (点)} \end{aligned}$$

書きの得点の平均値は,

$$\begin{aligned} & \frac{1}{20}(4 \times 2 + 5 \times 3 + 6 \times 2 + 7 \times 5 \\ & \quad + 8 \times 3 + 9 \times 4 + 10 \times 1) \\ &= \frac{1}{20}(8 + 15 + 12 + 35 + 24 + 36 + 10) \\ &= \frac{140}{20} = 7 \text{ (点)} \end{aligned}$$

(3) $r = \frac{3.45}{\sqrt{5} \sqrt{3}} = \frac{3.45}{\sqrt{15}} = \frac{3.45 \times \sqrt{15}}{15} = \frac{3.45 \times 3.873}{15}$
 $= 0.23 \times 3.873 \div 0.89$

- (4) 5 人の読みと書きの得点の平均値がそれぞれ 20 人の平均値と同じであることから, 生徒 25 人の読みと書きの平均値も, これらと同じになる.

20 人の読みの得点の偏差平方の和は,

$$(\sqrt{5})^2 \times 20 = 100$$

また, 5 人の読みの得点の偏差平方の和は,

$$\begin{aligned} & (4-5)^2 + (2-5)^2 + (2-5)^2 + (8-5)^2 + (9-5)^2 \\ &= 1 + 9 + 9 + 9 + 16 = 44 \end{aligned}$$

であるから, 生徒 25 人の読みの得点の分散 s_x^2 は,

$$s_x^2 = \frac{1}{25}(100 + 44) = \frac{144}{25}$$

20 人の書きの得点の偏差平方の和は,

$$(\sqrt{3})^2 \times 20 = 60$$

また, 5 人の書きの得点の偏差平方の和は,

$$\begin{aligned} & (8-7)^2 + (4-7)^2 + (6-7)^2 + (7-7)^2 + (10-7)^2 \\ &= 1 + 9 + 1 + 0 + 9 = 20 \end{aligned}$$

であるから, 生徒 25 人の書きの得点の分散 s_y^2 は,

$$s_y^2 = \frac{1}{25}(60 + 20) = \frac{80}{25}$$

20 人の得点の偏差積の和は,

$$3.45 \times 20 = 69$$

また, 5 人の得点の偏差積の和は,

$$\begin{aligned} & (4-5)(8-7) + (2-5)(4-7) + (2-5)(6-7) \\ & \quad + (8-5)(7-7) + (9-5)(10-7) \\ &= -1 + 9 + 3 + 0 + 12 = 23 \end{aligned}$$

◀ (20 人の偏差平方の和)
 $= (20 \text{ 人の分散}) \times 20$

◀ $\frac{1}{25} \{ (20 \text{ 人の得点の}$
 偏差平方の和)
 $+ (5 \text{ 人の得点の}$
 偏差平方の和) $\}$

であるから、生徒 25 人の読みと書きの得点の共分散 s_{xy} は、

$$s_{xy} = \frac{1}{25}(69+23) = \frac{92}{25}$$

よって、生徒 25 人の読みと書きの得点の相関係数 R は、

$$\begin{aligned} R &= \frac{s_{xy}}{s_x s_y} = \frac{92}{25} \div \left(\sqrt{\frac{144}{25}} \times \sqrt{\frac{80}{25}} \right) \\ &= \frac{92}{12 \times 4\sqrt{5}} = \frac{23}{12\sqrt{5}} = \frac{23\sqrt{5}}{60} \\ &= \frac{23 \times 2.236}{60} = \frac{51.428}{60} \div 0.86 \end{aligned}$$

◀ $\frac{1}{25}$ {(20 人の得点の
偏差積の和)
+ (5 人の得点の
偏差積の和)}

- (5) ① 生徒 25 人の読みと書きの得点の標準偏差は、それぞれ $\sqrt{5.76}$ 、 $\sqrt{3.2}$ より、正しくない。
 ② 生徒 20 人のときの相関係数は 0.89、生徒 25 人のときの相関係数は 0.86 であり、正の相関は少し弱まるから、正しい。
 ③ 読みの得点が一様に 1 点下がっても、読みの標準偏差 s_x の値、共分散 s_{xy} の値が変わらないので、読みと書きの得点の相関係数 R の値は変わらず、正しくない。

◀ 読みの得点が一様に 1 点下がっても、読みの得点の偏差は変化しないことから、 s_x 、 s_{xy} の値は変わらない。

154

袋の中には赤玉と白玉が合計 7 個入っている。

袋の中には一方の色が 2 個、他方が 5 個入っていることがわかっている。

A さんが袋の中から 1 個ずつ、2 回取り出したところ、2 回とも赤玉であった。

A さんはこのことから「袋の中には赤玉が 5 個、白玉が 2 個入っている」と思った。このことを検証したい。

- (1) 帰無仮説を答えよ。
 (2) 赤玉を 2 個取り出す確率が 5% 以下であれば帰無仮説を棄却するとして、玉の取り出し方が次の(i)、(ii)の場合について検証せよ。
 (i) 1 回取り出すごとに取り出した玉を袋に戻さない場合
 (ii) 1 回取り出すごとに取り出した玉を袋に戻す場合

- (1) 袋の中には赤玉が 2 個、白玉が 5 個入っている

- (2) (i) 2 個の玉の取り出し方は、全部で 42 通り。

その中で、2 個とも赤玉である確率は、

$$\frac{2}{42} = \frac{1}{21} = 0.0476\cdots$$

したがって、2 個とも赤玉である確率は約 4.8% で、5% より小さいから、この仮説は棄却される。
 よって、A さんの仮説が正しいと判断される。

- (ii) 2 個の玉の取り出し方は、全部で 49 通り。

その中で、2 個とも赤玉である確率は、

$$\frac{4}{49} = 0.0816\cdots$$

したがって、2 個とも赤玉である確率は約 8.2% で、5% より大きいから、この仮説は正しいとは判断できない。

●Step Up (p.309)

1

右の度数分布表は、ある植物の種子 100 個の重さを測定した結果である。

- (1) この度数分布表から、階級ごとの累積相対度数分布表を作成せよ。
- (2) (1)の累積相対度数分布表を用いて、累積相対度数の折れ線グラフをかけ。ただし、 x g 未満の累積相対度数 y に対して、点 (x, y) をとること。
- (3) (2)のグラフを利用して、この種子 100 個の重さの中央値を求めよ。

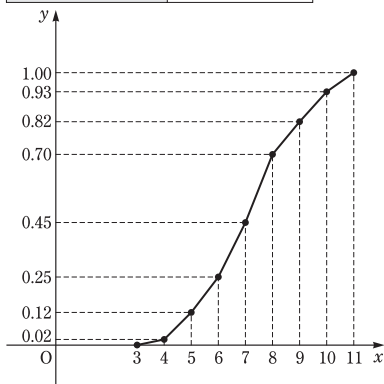
階級 (g)	度数 (個)
3 以上 4 未満	2
4 以上 5 未満	10
5 以上 6 未満	13
6 以上 7 未満	20
7 以上 8 未満	25
8 以上 9 未満	12
9 以上 10 未満	11
10 以上 11 未満	7
合計	100

＜考え方＞ (3) データの値を大きさの順に並べたとき、中央にくる値が中央値である。
したがって、グラフ上で $y=0.50$ となる x の値を求める。

(1)

階級 (g)	累積相対度数
3 以上 4 未満	0.02
4 以上 5 未満	0.12
5 以上 6 未満	0.25
6 以上 7 未満	0.45
7 以上 8 未満	0.70
8 以上 9 未満	0.82
9 以上 10 未満	0.93
10 以上 11 未満	1.00

(2)



◀ 階級ごとの相対度数は次のようになる。

階級 (g)	相対度数
3 以上 4 未満	0.02
4 以上 5 未満	0.10
5 以上 6 未満	0.13
6 以上 7 未満	0.20
7 以上 8 未満	0.25
8 以上 9 未満	0.12
9 以上 10 未満	0.11
10 以上 11 未満	0.07
合計	1.00

- (3) 中央値は、グラフ上で
 $y=0.50$ のときの x の値で
 あり、グラフより、 $7 \leq x < 8$
 の区間にある。

2点 $(7, 0.45)$, $(8, 0.70)$
 を通る直線の傾きは、

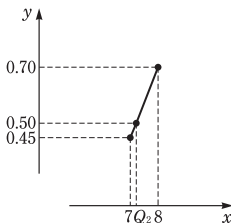
$$\frac{0.70-0.45}{8-7}=0.25$$

中央値を Q_2 とすると、点 $(Q_2, 0.50)$ がこの直線上に
 あるから、2点 $(7, 0.45)$, $(Q_2, 0.50)$ を通る直線の傾き
 は 0.25 となる。

$$\text{したがって、} \frac{0.50-0.45}{Q_2-7}=0.25 \text{ より、}$$

$$Q_2 = \frac{0.05}{0.25} + 7 = 0.2 + 7 = 7.2$$

よって、中央値は、 7.2 g



◀ 直線の傾きを考える。

2

次の表は、ある試験とその追試の結果である。試験後には、直しのレポートを提出させ、よく復習してから追試を受けるように指示した。箱ひげ図をかき、それを比較し、復習はよくなされたといえるかどうか答えよ。

得点(点)	0	1	2	3	4	5	6	7	8	9	10	合計
試験(人)	0	0	0	1	3	2	6	4	6	2	1	25
追試(人)	0	0	0	0	1	3	2	4	6	7	2	25

<考え方> 平均値、四分位数、最小値、最大値を調べ、箱ひげ図をかく。
 試験と追試のデータの違いを図から読み取る。

<試験について>

平均値は、

$$\begin{aligned} & \frac{1}{25}(3 \times 1 + 4 \times 3 + 5 \times 2 + 6 \times 6 + 7 \times 4 + 8 \times 6 \\ & \qquad \qquad \qquad + 9 \times 2 + 10 \times 1) \\ & = \frac{1}{25}(3 + 12 + 10 + 36 + 28 + 48 + 18 + 10) \\ & = \frac{165}{25} = 6.6 \text{ (点)} \end{aligned}$$

データの値を小さい順に並べると、

3, 4, 4, 4, 5, 5, 6, 6, 6, 6, 6, 6, 7,
 7, 7, 7, 8, 8, 8, 8, 8, 8, 9, 9, 10

中央値は13番目の値であるから、 7 点

第1四分位数は、1番目から12番目までのデータの中央
 値であるから、6番目の値5と7番目の値6の平均値で、

$$\frac{5+6}{2}=5.5 \text{ (点)}$$

第3四分位数は、14番目から25番目までのデータの中央
 値であるから、19番目の値8と20番目の値8の平均値で、

$$\frac{8+8}{2}=8 \text{ (点)}$$

また、最小値は3点、最大値は10点である。

◀ データの値の個数が奇数
 $(2n+1)$ 個の場合、中央値は、
 $(n+1)$ 番目の値となる。

○○○○ ○ ○○○○
 n 個 中央値 n 個

◀ データ全体の中央値より前の
 データと後のデータにおいて、
 それぞれの中央値が第1四分
 位数と第3四分位数である。

244 第5章 データの分析

<追試について>

平均値は、

$$\begin{aligned} & \frac{1}{25}(4 \times 1 + 5 \times 3 + 6 \times 2 + 7 \times 4 + 8 \times 6 + 9 \times 7 + 10 \times 2) \\ &= \frac{1}{25}(4 + 15 + 12 + 28 + 48 + 63 + 20) \\ &= \frac{190}{25} = 7.6 \text{ (点)} \end{aligned}$$

データの値を小さい順に並べると、

4, 5, 5, 5, 6, 6, 7, 7, 7, 7, 8, 8, 8,
8, 8, 8, 9, 9, 9, 9, 9, 9, 10, 10

中央値は13番目の値であるから、8点

第1四分位数は、1番目から12番目までのデータの中央値であるから、6番目の値6と7番目の値7の平均値で、

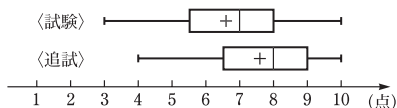
$$\frac{6+7}{2} = 6.5 \text{ (点)}$$

第3四分位数は、14番目から25番目までのデータの中央値であるから、19番目の値9と20番目の値9の平均値で、

$$\frac{9+9}{2} = 9 \text{ (点)}$$

また、最小値は4点、最大値は10点である。

以上から、試験と追試のデータの箱ひげ図は、次のようになる。



箱ひげ図を比較すると、追試の方が、箱の位置は点数の高い方へ移動しており、最小値、平均値、中央値も上がっている。

よって、復習はよくなされたといえる。

◀ 第1四分位数から第3四分位数までで箱をかく。そして、最小値、最大値までひげをのばす。

3

次の表は、生徒10人のテストの点数で、下はその箱ひげ図である。平均値が62点であるとき、表の a , b , c , d の値をそれぞれ求めよ。ただし、 $a < b < c < d$ とする。

出席番号	1	2	3	4	5	6	7	8	9	10
点数(点)	42	33	a	52	b	d	76	c	82	68



<考え方> データの値の個数は10個であるから、データの値を小さい順に並べたとき、第1四分位数は3番目、第2四分位数は5番目と6番目の平均値、第3四分位数は8番目となる。

それぞれのデータの値が何番目になるのかを考える。

データの値を小さい順に並べるとする。

箱ひげ図から、最小値は33で、この値は表の中にあり、1番目となる。

箱ひげ図から、最大値は85であるが、表の中にこの値はな

い.

したがって, a, b, c, d の中で最も大きい値が 85 となり,
 $a < b < c < d$ より, $d = 85$

箱ひげ図から, 第 1 四分位数は 46 で, これは 3 番目となる.
表にはこれより小さい 33 と 42 はあり, 46 はない.

したがって, a, b, c の中で最も小さい値が 46 となるので,
 $a = 46$

箱ひげ図から, 第 3 四分位数は 76 で, これは 8 番目となる.
表にはこれより大きい 82 と d , すなわち, 82 と 85 があり,
76 も表の中にある.

以上より, 4 ~ 7 番目は, 52, 68, b, c のいずれかとなる.
箱ひげ図から, 中央値は 66 点である.

ここで, 52 は中央値より小さいから, 4 番目か 5 番目となる.

52 が 5 番目であるとする, 中央値は 5 番目と 6 番目の平均値であることから, 6 番目の値は, $66 \times 2 - 52 = 80$

これは第 3 四分位数, すなわち, 8 番目の値より大きくなるから, 不適.

したがって, 52 は 4 番目である.

平均値が 62 であることと, わかっている値から,

$$\begin{aligned} b + c &= 62 \times 10 - (33 + 42 + 46 + 52 + 68 + 76 + 82 + 85) \\ &= 620 - 484 = 136 \end{aligned}$$

これと $b < c$ より, $b < 68 < c$

したがって, b は 5 番目, 68 は 6 番目の値で,

$$b = 66 \times 2 - 68 = 64$$

c は 7 番目の値で, $c = 136 - b = 136 - 64 = 72$

よって, $a = 46, b = 64, c = 72, d = 85$

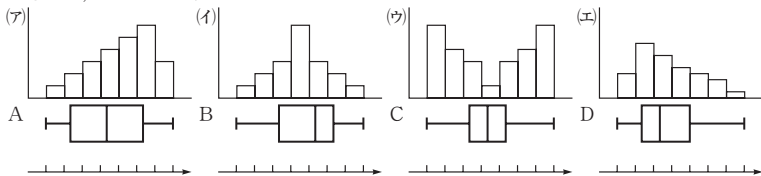
◀ 1 番目 33, 2 番目 42,
3 番目 46, 8 番目 76,
9 番目 82, 10 番目 85

◀ $b + c = 136$ より, $c = 136 - b$
これと $b < c$ より, $b < 136 - b$
 $2b < 136, b < 68$
同様に, $c > 68$

5

4

次の(ア)~(エ)のヒストグラムについて, 同じデータを使って表した箱ひげ図としてふさわしいものを, A~D から選べ.



<考え方> ヒストグラムが左右対称なのか左や右に偏っているのか, データが集まっているのか散らばっているのか, といったことから, 対応する箱ひげ図を考える.

(イ)と(ウ)のヒストグラムはデータの中央部分に関して対称なので, 対応する箱ひげ図は A か C になる.

(イ)のデータが中央に集まっているから, 箱の長さはより短い方になる.

一方, (ウ)のヒストグラムの山は両端にあり, 広い区間にデータが散らばっているから, 箱の長さは長くなる.

また, (ア)のヒストグラムはデータが右に偏って左に裾をひいているから, 箱ひげ図の左側のひげが長くなる.

(エ)のヒストグラムはデータが左に偏って右に裾をひいてい

◀ 対称性に着目

◀ データの集まり具合に着目

◀ データの偏りに着目

246 第5章 データの分析

るので、箱ひげ図の右側のひげが長くなる。

よって、 (ア) B (イ) C (ウ) A (エ) D

5

8 個のデータの値 x_1, x_2, \dots, x_8 の平均値は 3, 分散は 4 である。これに $x_9=6, x_{10}=5$ をつけ加えた 10 個のデータの値 x_1, x_2, \dots, x_{10} の平均値と分散を求めよ。

<考え方> 平均値や分散の定義の式から、まず $x_1+x_2+\dots+x_8, x_1^2+x_2^2+\dots+x_8^2$ の値を求めておく。

$$\frac{1}{8}(x_1+x_2+\dots+x_8)=3 \text{ より,}$$

$$x_1+x_2+\dots+x_8=24$$

よって、求める平均値は、

$$\begin{aligned} & \frac{1}{10}\{(x_1+x_2+\dots+x_8)+x_9+x_{10}\} \\ &= \frac{1}{10}(24+6+5)=\frac{35}{10}=\frac{7}{2} \end{aligned}$$

$$\frac{1}{8}(x_1^2+x_2^2+\dots+x_8^2)-3^2=4 \text{ より,}$$

$$x_1^2+x_2^2+\dots+x_8^2=104$$

よって、求める分散は、

$$\begin{aligned} & \frac{1}{10}\{(x_1^2+x_2^2+\dots+x_8^2)+x_9^2+x_{10}^2\}-\left(\frac{7}{2}\right)^2 \\ &= \frac{1}{10}(104+36+25)-\frac{49}{4} \\ &= \frac{165}{10}-\frac{49}{4}=\frac{33}{2}-\frac{49}{4}=\frac{17}{4} \end{aligned}$$

◀ (x の分散)
 $= (x \text{ の平均値})^2$
 $- (x \text{ の平均値})^2$

6

10 個のデータの値がある。そのうちの 6 個のデータの値の平均値は 3, 分散は 9 である。残りの 4 個のデータの値の平均値は 8, 分散は 14 である。この 10 個のデータの値の平均値と分散を求めよ。

<考え方> 平均値や分散の定義の式から、まず 6 個のデータの値と 4 個のデータの値それぞれについて、値の和と値の平方の和を求めておく。

10 個の値を x_1, x_2, \dots, x_{10} とし、

x_1, x_2, \dots, x_6 の平均値が 3, 分散が 9

x_7, x_8, x_9, x_{10} の平均値が 8, 分散が 14

とする。

$$\frac{1}{6}(x_1+x_2+\dots+x_6)=3 \text{ より, } x_1+x_2+\dots+x_6=18$$

$$\frac{1}{4}(x_7+x_8+x_9+x_{10})=8 \text{ より, } x_7+x_8+x_9+x_{10}=32$$

よって、求める平均値は、

$$\begin{aligned} & \frac{1}{10}\{(x_1+x_2+\dots+x_6)+(x_7+x_8+x_9+x_{10})\} \\ &= \frac{1}{10}(18+32)=5 \end{aligned}$$

$$\frac{1}{6}(x_1^2 + x_2^2 + \cdots + x_6^2) - 3^2 = 9 \text{ より,}$$

$$x_1^2 + x_2^2 + \cdots + x_6^2 = 108$$

$$\frac{1}{4}(x_7^2 + x_8^2 + x_9^2 + x_{10}^2) - 8^2 = 14 \text{ より,}$$

$$x_7^2 + x_8^2 + x_9^2 + x_{10}^2 = 312$$

よって、求める分散は、

$$\begin{aligned} & \frac{1}{10}\{(x_1^2 + x_2^2 + \cdots + x_6^2) + (x_7^2 + x_8^2 + x_9^2 + x_{10}^2)\} - 5^2 \\ &= \frac{1}{10}(108 + 312) - 25 = 42 - 25 = 17 \end{aligned}$$

◀ (x の分散)
 $= (x^2 \text{ の平均値})$
 $- (x \text{ の平均値})^2$

7

任意の連続する5個の自然数の分散 s^2 を求めよ。

<考え方> n を自然数とすると、任意の連続する5個の自然数は、 $n, n+1, n+2, n+3, n+4$ とおけるので、この5個の自然数の分散を求める。

n を自然数とすると、任意の連続する5個の自然数は、 $n, n+1, n+2, n+3, n+4$ とおける。

この5個の自然数の平均値は、

$$\frac{1}{5}\{n + (n+1) + (n+2) + (n+3) + (n+4)\} = n+2$$

よって、この5個の自然数の分散は、

$$\begin{aligned} & \frac{1}{5}\{n^2 + (n+1)^2 + (n+2)^2 + (n+3)^2 + (n+4)^2\} - (n+2)^2 \\ &= \frac{1}{5}(5n^2 + 20n + 30) - (n^2 + 4n + 4) = 2 \end{aligned}$$

【参考】 連続する任意の m 個の自然数の分散は、記号 Σ (数学B) を使うと、

$$\begin{aligned} & \frac{1}{m} \sum_{k=0}^{m-1} (n+k)^2 - \left\{ \frac{1}{m} \sum_{k=0}^{m-1} (n+k) \right\}^2 \\ &= \frac{1}{12}(m-1)(m+1) \end{aligned}$$

となる。

◀ (x の分散)
 $= (x^2 \text{ の平均値})$
 $- (x \text{ の平均値})^2$
 ◀ n の値にかかわらず一定の値となる。

8

変数 x の n 個のデータの値 x_1, x_2, \dots, x_n がある。 x の平均値が a (定数) であるとき、このデータ x の分散 s^2 の最小値を求めよ。

<考え方> $s^2 = \frac{1}{n}\{(x_1-a)^2 + (x_2-a)^2 + \cdots + (x_n-a)^2\}$ であることから、 s^2 の最小値を考える。

$$s^2 = \frac{1}{n}\{(x_1-a)^2 + (x_2-a)^2 + \cdots + (x_n-a)^2\}$$

ここで、

$$(x_1-a)^2 \geq 0, (x_2-a)^2 \geq 0, \dots, (x_n-a)^2 \geq 0$$

であり、それぞれの不等式は、 $x_k = a$ ($k=1, 2, \dots, n$) のとき等号が成り立つ。

よって、 s^2 は、 $x_1 = x_2 = \cdots = x_n = a$ のとき、最小値0をとる。

◀ (実数) $^2 \geq 0$

9

変量 x の n 個のデータの値 x_1, x_2, \dots, x_n がある. x の平均値 \bar{x} が 6, 標準偏差 s_x が $2\sqrt{2}$ であるとき, 変量 $y = \frac{1}{4}x + 20$ の平均値 \bar{y} と標準偏差 s_y を求めよ.

<考え方> $y_1 = \frac{1}{4}x_1 + 20, y_2 = \frac{1}{4}x_2 + 20, \dots, y_n = \frac{1}{4}x_n + 20$ と表せることから, \bar{y} と s_y を \bar{x} や s_x で表す.

変量 y の n 個のデータの値は,

$$y_1 = \frac{1}{4}x_1 + 20, y_2 = \frac{1}{4}x_2 + 20, \dots, y_n = \frac{1}{4}x_n + 20$$

と表せる.

$$\bar{x} = \frac{1}{n}(x_1 + x_2 + \dots + x_n)$$

$$s_x^2 = \frac{1}{n}\{(x_1 - \bar{x})^2 + (x_2 - \bar{x})^2 + \dots + (x_n - \bar{x})^2\}$$

よって, 変量 y の平均値は,

$$\begin{aligned}\bar{y} &= \frac{1}{n}(y_1 + y_2 + \dots + y_n) \\ &= \frac{1}{n}\left\{\left(\frac{1}{4}x_1 + 20\right) + \left(\frac{1}{4}x_2 + 20\right) + \dots + \left(\frac{1}{4}x_n + 20\right)\right\} \\ &= \frac{1}{n}\left\{\frac{1}{4}(x_1 + x_2 + \dots + x_n) + 20n\right\} \\ &= \frac{1}{4} \cdot \frac{1}{n}(x_1 + x_2 + \dots + x_n) + \frac{1}{n} \cdot 20n \\ &= \frac{1}{4}\bar{x} + 20 = \frac{1}{4} \cdot 6 + 20 = \frac{43}{2}\end{aligned}$$

変量 y の分散は,

$$\begin{aligned}s_y^2 &= \frac{1}{n}\{(y_1 - \bar{y})^2 + (y_2 - \bar{y})^2 + \dots + (y_n - \bar{y})^2\} \\ &= \frac{1}{n}\left\{\left(\frac{1}{4}x_1 + 20 - \frac{1}{4}\bar{x} - 20\right)^2 + \left(\frac{1}{4}x_2 + 20 - \frac{1}{4}\bar{x} - 20\right)^2\right. \\ &\quad \left.+ \dots + \left(\frac{1}{4}x_n + 20 - \frac{1}{4}\bar{x} - 20\right)^2\right\} \\ &= \frac{1}{n}\left\{\left(\frac{1}{4}\right)^2(x_1 - \bar{x})^2 + \left(\frac{1}{4}\right)^2(x_2 - \bar{x})^2\right. \\ &\quad \left.+ \dots + \left(\frac{1}{4}\right)^2(x_n - \bar{x})^2\right\} \\ &= \left(\frac{1}{4}\right)^2 \cdot \frac{1}{n}\{(x_1 - \bar{x})^2 + (x_2 - \bar{x})^2 + \dots + (x_n - \bar{x})^2\} \\ &= \left(\frac{1}{4}\right)^2 s_x^2\end{aligned}$$

$$\text{よって, } s_y = \frac{1}{4}s_x = \frac{1}{4} \cdot 2\sqrt{2} = \frac{\sqrt{2}}{2}$$

◀ 一般に, $y = ax + b$ のとき,
(y の平均値)
 $= a \times (x \text{ の平均値}) + b$

◀ 一般に, $y = ax + b$ のとき,
(y の標準偏差)
 $= |a| \times (x \text{ の標準偏差})$

10

次の表は、高校生の兄弟9組の身長を計測した結果である。

番号	1	2	3	4	5	6	7	8	9
兄の身長 (cm)	166	170	179	173	184	172	169	163	172
弟の身長 (cm)	165	170	a	174	176	171	166	166	167

弟の身長の平均値は、兄の身長の平均値より 2 cm 小さいという。

- 変量 x と変量 u の間に $x = mu + n$ (m, n は定数) という関係があるとき、 x の平均値 \bar{x} と u の平均値 \bar{u} の間には $\bar{x} = m\bar{u} + n$ という関係が成り立つ。兄の身長 x cm を $x = u + 170$ と考えて、その平均値 \bar{x} を求めよ。
- a の値を求めよ。
- 兄の身長 x cm と弟の身長 y cm の散布図をかけ。
- 兄の身長 x cm と弟の身長 y cm の相関係数 r を求めよ。また、兄の身長と弟の身長の間には、相関関係があるといえるか。

<考え方> (4) x, y の相関係数 r は、 $r = \frac{S_{xy}}{S_x S_y}$ で表される。

兄の身長と弟の身長の間相関関係は、相関係数や散布図から判断する。

$$\begin{aligned} (1) \quad \bar{x} &= \frac{1}{9}\{(-4)+0+9+3+14+2+(-1)+(-7)+2\}+170 \\ &= \frac{18}{9}+170=2+170=172 \text{ (cm)} \end{aligned}$$

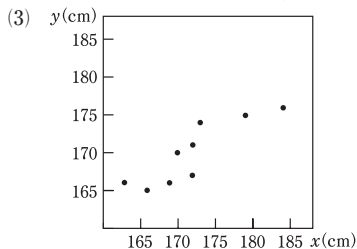
$$(2) \quad \text{弟の身長の平均値 } \bar{y} \text{ は, } \bar{y}=172-2=170 \text{ (cm)}$$

また,

$$\begin{aligned} \bar{y} &= \frac{1}{9}\{(-5)+0+(a-170)+4+6+1+(-4) \\ &\quad +(-4)+(-3)\}+170 \\ &= \frac{a-175}{9}+170 \end{aligned}$$

$$\text{したがって, } \frac{a-175}{9}+170=170 \text{ より, } \frac{a-175}{9}=0$$

$$\text{よって, } a-175=0 \text{ より, } a=175$$



◀たとえば、166 は、
 $166 = (-4) + 170$ と考える。

◀弟の身長の平均値も同様に
して計算する。このように、平
均値に近い値を基準として平
均値を求めるとき、基準とな
る値を「仮平均」という。

◀右上がりの直線状に点が並ぶ
傾向が見られる。

(4)

	x	y	$x-\bar{x}$	$y-\bar{y}$	$(x-\bar{x})^2$	$(y-\bar{y})^2$	$(x-\bar{x})(y-\bar{y})$
	166	165	-6	-5	36	25	30
	170	170	-2	0	4	0	0
	179	175	7	5	49	25	35
	173	174	1	4	1	16	4
	184	176	12	6	144	36	72
	172	171	0	1	0	1	0
	169	166	-3	-4	9	16	12
	163	166	-9	-4	81	16	36
	172	167	0	-3	0	9	0
合計			0	0	324	144	189
平均値	72	70	0	0	36	16	21

 x, y の標準偏差は、それぞれ、

$$s_x = \sqrt{36} = 6, \quad s_y = \sqrt{16} = 4$$

 x と y の共分散は、 $s_{xy} = 21$

$$x \text{ と } y \text{ の相関係数は, } r = \frac{21}{6 \times 4} = \frac{7}{8}$$

$\frac{7}{8} = 0.875$ より、相関係数が 1 に近いことと、(3) の散布図から、兄の身長と弟の身長の間には、正の相関があるといえる。

◀ 表を作ると、標準偏差や共分散の値を計算しやすい。

◀ 相関係数

$$r = \frac{s_{xy}}{s_x s_y}$$

11

2つの変量 x, y をもつ n 個のデータ $(x_1, y_1), (x_2, y_2), \dots, (x_n, y_n)$ がある。 x と y の間に $y = 7x + 8$ の関係があるとき、 x と y の相関係数 r を求めよ。

<考え方> 2つの変量 x, y の間に $y = ax + b$ の関係があるとき、

$$\bar{y} = a\bar{x} + b, \quad s_y = |a|s_x$$

が成り立つことを利用する。

x, y の平均値をそれぞれ \bar{x}, \bar{y} とすると、

$$\bar{y} = 7\bar{x} + 8$$

x, y の標準偏差をそれぞれ s_x, s_y とすると、

$$s_y = |7|s_x = 7s_x$$

$k = 1, 2, \dots, n$ について、

$$(x_k - \bar{x})(y_k - \bar{y}) = (x_k - \bar{x})\{(7x_k + 8) - (7\bar{x} + 8)\}$$

$$= (x_k - \bar{x})\{7(x_k - \bar{x})\} = 7(x_k - \bar{x})^2$$

が成り立つから、 x と y の共分散 s_{xy} は、

$$\begin{aligned} s_{xy} &= \frac{1}{n}\{(x_1 - \bar{x})(y_1 - \bar{y}) + (x_2 - \bar{x})(y_2 - \bar{y}) \\ &\quad + \dots + (x_n - \bar{x})(y_n - \bar{y})\} \end{aligned}$$

$$= \frac{1}{n}\{7(x_1 - \bar{x})^2 + 7(x_2 - \bar{x})^2 + \dots + 7(x_n - \bar{x})^2\}$$

$$= 7 \cdot \frac{1}{n}\{(x_1 - \bar{x})^2 + (x_2 - \bar{x})^2 + \dots + (x_n - \bar{x})^2\}$$

◀ $y = ax + b$ のとき、
 $(y \text{ の平均値})$
 $= a \times (x \text{ の平均値}) + b$
 $(y \text{ の標準偏差})$
 $= |a| \times (x \text{ の標準偏差})$
 (例題 150 を参照)

$$=7s_x^2$$

よって、 x と y の相関係数 r は、

$$r = \frac{s_{xy}}{s_x s_y} = \frac{7s_x^2}{s_x \times 7s_x} = 1$$

◀ 一般に、 $y=ax+b$ の関係がある2つの変量 x と y の相関係数 r は、
 $a > 0$ のとき、 $r=1$
 $a < 0$ のとき、 $r=-1$

12

右の表は、芸術大学の学生10人がそれぞれ作品A(絵画)と作品B(彫刻)を制作し、それらに対して0, 1, 2の3段階で評価を行った際の得点を2次元の度数分布表にまとめたものである。Aの得点 x と Bの得点 y の相関係数 r を小数第2位まで求めよ。ただし、 $\sqrt{41}=6.403$ とする。

A \ B	0	1	2	合計
0	0	1	0	1
1	1	3	1	5
2	0	2	2	4
合計	1	6	3	10

<考え方> x, y の相関係数 r は、 $r = \frac{s_{xy}}{s_x s_y}$ で表される。

標準偏差や共分散の計算は、表を作るとよい。

	x	y	$x-\bar{x}$	$y-\bar{y}$	$(x-\bar{x})^2$	$(y-\bar{y})^2$	$(x-\bar{x})(y-\bar{y})$
	0	1	-1.2	-0.3	1.44	0.09	0.36
	1	0	-0.2	-1.3	0.04	1.69	0.26
	1	1	-0.2	-0.3	0.04	0.09	0.06
	1	1	-0.2	-0.3	0.04	0.09	0.06
	1	1	-0.2	-0.3	0.04	0.09	0.06
	1	2	-0.2	0.7	0.04	0.49	-0.14
	1	2	-0.2	0.7	0.04	0.49	-0.14
	2	1	0.8	-0.3	0.64	0.09	-0.24
	2	2	0.8	0.7	0.64	0.49	0.56
	2	2	0.8	0.7	0.64	0.49	0.56
合計	12	13	0	0	3.6	4.1	1.4
平均値	1.2	1.3	0	0	0.36	0.41	0.14

x, y の標準偏差は、それぞれ、

$$s_x = \sqrt{0.36} = \sqrt{\frac{9}{25}} = \frac{3}{5}$$

$$s_y = \sqrt{0.41} = \sqrt{\frac{41}{100}} = \frac{\sqrt{41}}{10}$$

x と y の共分散は、 $s_{xy} = 0.14 = \frac{7}{50}$

よって、 x と y の相関係数 r は、

$$r = \frac{7}{50} \div \left(\frac{3}{5} \times \frac{\sqrt{41}}{10} \right) = \frac{7}{3\sqrt{41}} = \frac{7\sqrt{41}}{123}$$

$$= \frac{7 \times 6.403}{123} = \frac{44.821}{123} \div 0.36$$

◀ 表を作ると標準偏差や共分散の値を計算しやすい。

◀ 相関係数

$$r = \frac{s_{xy}}{s_x s_y}$$

13

99 個の観測値からなるデータがある。次の四分位数について書かれた文章のうち、どのようなデータでも成り立つものをすべて選べ。

- ① 平均値は第 1 四分位数と第 3 四分位数の間にある。
- ② 四分位範囲は標準偏差より大きい。
- ③ 中央値より小さい観測値の個数は 49 個である。
- ④ 最大値に等しい観測値を 1 個削除しても第 1 四分位数は変わらない。
- ⑤ 第 1 四分位数より小さい観測値と、第 3 四分位数より大きい観測値とをすべて削除すると、残りの観測値の個数は 51 個である。
- ⑥ 第 1 四分位数より小さい観測値と、第 3 四分位数より大きい観測値とをすべて削除すると、残りの観測値からなるデータの範囲はもとのデータの四分位範囲に等しい。

- ① 98 個が 0, 1 個が 99 のデータを例として考える。

この場合、平均値は 1 となり、第 3 四分位数の 0 より大きいから、正しくない。

- ② ①の例で考えると、四分位範囲は、 $0 - 0 = 0$ 、標準偏差は、

$$s = \sqrt{\frac{1}{99} \{(0-1)^2 \times 98 + (99-1)^2\}} > 0$$

であるから、正しくない。

- ③ ①の例で考えると、中央値は 0 であり、それより小さい観測値はないから、正しくない。

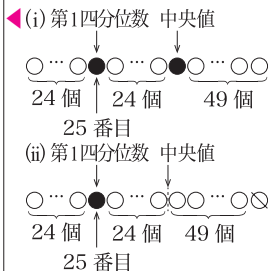
- ④ 99 個のデータを小さい方から順に並べた場合と、右側のデータを 1 個削除した場合をくらべると、第 1 四分位数はともに、左から 25 番目となり等しいから、正しい。

- ⑤ ①の例で考えると、第 1 四分位数は 0 で、それより小さい観測値はなく、第 3 四分位数の 0 より大きい観測値は 1 個である。これを削除すると、残りの観測値の個数は 98 個であるから、正しくない。

- ⑥ 第 1 四分位数を Q_1 、第 3 四分位数を Q_3 とすると、四分位範囲は、 $Q_3 - Q_1$ である。

Q_1 より小さい観測値と Q_3 より大きい観測値をすべて削除したときのデータの範囲は、 $Q_3 - Q_1$ と表せるので、正しい。

以上より、どのようなデータでも成り立つのは、④と⑥。



14

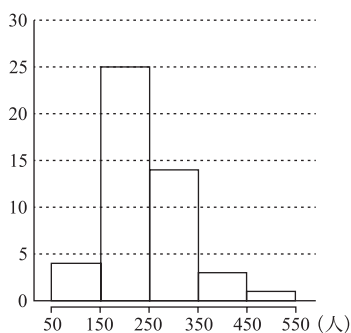
右の図は、2008 年における 47 都道府県の人口 1 万人あたりの旅券取得者数のヒストグラムである。なお、ヒストグラムの各階級の区間は、左側の数値を含み、右側の数値を含まない。

このヒストグラムに対して、各階級に含まれるデータの値がすべてその階級値に等しいと仮定するとき、次の問いに答えよ。

- (1) この変量の平均値 \bar{x} を小数第 1 位を四捨五入して求めよ。
- (2) この変量の分散 s^2 の値に最も近いものを次の①～⑧から 1 つ選べ。

- ① 3900 ② 4900 ③ 5900
- ④ 6900 ⑤ 7900 ⑥ 8900
- ⑦ 9900 ⑧ 10900

(都道府県数)



2008 年における旅券取得者数のヒストグラム

(出典：外務省の Web ページにより作成)

ヒストグラムより、階級値と度数を読み取り度数分布表を作ると、次のようになる。

階級値(人)	100	200	300	400	500	計
度数	4	25	14	3	1	47

$$(1) \quad \bar{x} = \frac{1}{47}(100 \times 4 + 200 \times 25 + 300 \times 14 + 400 \times 3 + 500 \times 1) \\ = 240.42 \cdots$$

よって、 $\bar{x} = 240$ (人)

別解 仮平均を最頻値 200 人すると、平均値は、

$$\begin{aligned} \bar{x} &= 200 + \frac{1}{47}\{(-100) \times 4 + 0 \times 200 \\ &\quad + 100 \times 14 + 200 \times 3 + 300 \times 1\} \\ &= 200 + \frac{1900}{47} \\ &= 240.42 \cdots \end{aligned}$$

よって、 $\bar{x} = 240$ (人)

(2) 変量 x の分散 s^2 は、 $s^2 = \bar{x}^2 - (\bar{x})^2$ で求められる。

$$\begin{aligned} s^2 &= \frac{1}{47}(100^2 \cdot 4 + 200^2 \cdot 25 + 300^2 \cdot 14 \\ &\quad + 400^2 \cdot 3 + 500^2 \cdot 1) - 240^2 \\ &= 6868.08 \cdots \end{aligned}$$

よって、最も近い値は 6900 なので、④となる。

15

次の問いに答えよ。

- (1) ある硬貨Aがある。恵さんが、この硬貨Aを投げて表が出るか裏が出るかという実験を行い、次のような結果が得られた。
恵さん：硬貨Aを10回投げて、表が7回出た。
この結果から、「硬貨Aは表が出やすく加工されている」と恵さんは考えた。

起こる割合が5%以下であればほとんど起こりえないと判断するものとし、恵さんの考えが正しいと判断できるか答えよ。なお、硬貨を10回投げたときの表裏の出方は1024通りで、表が7回以上出る出方は176通りである。

- (2) 今度は、光さんが(1)の硬貨Aを投げて表裏の出方の実験を行ったところ、100回中70回表が出た。100回投げて70回以上の表が出る確率を計算すると、約0.004%になることから、光さんは、「硬貨Aは表が出やすく加工されている」と考えた。

恵さんと光さんの考えを利用して、次の文の内容が正しいと判断できるかどうかを考えた。(ア)、(イ)の○×の組合せが正しいものを、下の選択肢①～④より1つ選べ。

(ア) 1人の当選者を選ぶ選挙に、S候補とT候補の2名が立候補した。第1投票所での出口調査で無作為に選んだ10人のうち7人がS候補に投票したので、S候補が当選すると判断した。

(イ) 新発売の商品のパッケージデザインをA案にするかB案にするかの消費者アンケートを100人選んで行い、投票数の多いほうの案を消費者全体に好まれそうな案だと判断した。

	①	②	③	④
(ア)	○	○	×	×
(イ)	○	×	○	×

○：正しいと判断できる、×：正しいと判断できない

254 第5章 データの分析

- (1) 「硬貨Aは表が出やすく加工されていない」と仮定する。

硬貨Aを投げて、表が7回以上出る確率は、

$$\frac{176}{1024} = 0.1718\cdots$$

より、約17.2%

5%より大きいから、恵子さんの仮説は正しいとは判断できない。

- (2) (ア) 1人の当選者を選ぶ選挙で、S候補が当選するかT候補が当選するかは、硬貨Aを投げて表が出るか裏が出るかの考え方と同じである。

したがって、恵さんの考え方を利用することができるので、(1)の結果より、正しいとは判断できないから、×。

- (イ) A案、B案のどちらの案が消費者全体に好まれるかどうかは、恵さんと光さんの考え方を利用しても、判断できないから、×。
よって、正しい組合せは④。

◀ 「硬貨Aは表が出やすく加工されている」の否定。

◀ A案を51人、B案を49人が選んだ場合、投票数の多いA案の方が消費者全体に好まれるような案かどうかはわからない。

● 章末問題 (p.314)

1

新商品のパッケージデザインとしてAとBの2案を作成し、街頭アンケートで消費者30人に好みを選んでもらったところ、B案を選んだ人が23人いた。一般に、B案の方が消費者に好まれる案だと判断してもよいだろうか。もしA案とB案を何も見ずに選ぶ場合、それぞれが選ばれる確率は0.5とし、起こる割合が5%以下であればほとんど起こりえないと判断するものとする。また、右の表は、表と裏が同じ割合で出るコイン30枚を同時に投げたときの、表が出た枚数を記録することを1000回行った結果である。必要に応じて用いてもよい。

表の枚数	回	表の枚数	回
0~15	526	23	5
16	132	24	4
17	131	25	3
18	95	26	3
19	46	27	1
20	32	28	0
21	14	29	0
22	8	30	0

「B案の方が消費者に好まれるとはいえない」と仮定する。

A案とB案を何も見ずに選ぶ場合のB案の選ばれ方が0.5なので、コインの表裏の出方とおき換えて考えることができる。

コインの表裏の出方において、30枚中23枚以上が表となるのは、

$$\frac{5}{1000} + \frac{4}{1000} + \frac{3}{1000} + \frac{3}{1000} + \frac{1}{1000} = \frac{16}{1000} = 0.016$$

より、1.6%

5%より小さいから、この仮説は棄却される。

よって、もとの仮説「B案の方が消費者に好まれる案」が正しいと判断される。

●思考力問題 (p.314)

2

総務省が実施している国勢調査では都道府県ごとの総人口が調べられており、その内訳として日本人人口と外国人人口が公表されている。また、外務省では旅券(パスポート)を取得した人数を都道府県ごとに公表している。加えて、文部科学省では都道府県ごとの小学校に在籍する児童数を公表している。

そこで、47 都道府県の、人口1万人あたりの外国人人口(以下、外国人数)、人口1万人あたりの小学校児童数(以下、小学生数)、また、日本人1万人あたりの旅券を取得した人数(以下、旅券取得者数)を、それぞれ計算した。

次の図1は、2010年における47都道府県の、旅券取得者数(横軸)と小学生数(縦軸)の関係を黒丸で、また、旅券取得者数(横軸)と外国人数(縦軸)の関係を白丸で表した散布図である。

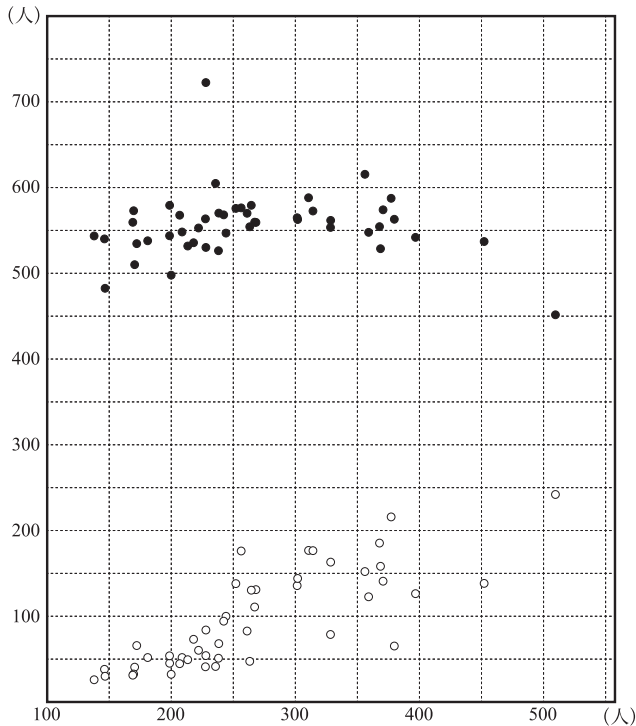


図1 2010年における、旅券取得者数と小学生数の散布図(黒丸)、
旅券取得者数と外国人数の散布図(白丸)

(出典：外務省、文部科学省および総務省のWebページにより作成)

次の(I)、(II)、(III)は図1の散布図に関する記述である。

(I)~(III)の文が正しいかどうか答えよ。

- (I) 小学生数の四分位範囲は、外国人数の四分位範囲より大きい。
- (II) 旅券取得者数の範囲は、外国人数の範囲より大きい。
- (III) 旅券取得者数と小学生数の相関係数は、旅券取得者数と外国人数の相関係数より大きい。

- (I) 小学生数について、第1四分位数は550人付近、第3四分位数は575人付近なので、四分位範囲は25人程度である。

一方、外国人数については、第1四分位数は50人付近、第3四分位数は150人付近なので、四分位範囲は100人程度である。

したがって、外国人数の四分位範囲のほうが大きいといえるので、正しくない。

- (II) 旅券取得者数は、最小値が150人未満で、最大値が500人より大きいことから、範囲は350人より大きい。

一方、外国人数は、最小値が25人付近で、最大値が250人付近であることから、範囲は225人程度である。

したがって、旅券取得者の範囲のほうが大きいといえるので、正しい。

- (III) 旅券取得者数と小学生数については、旅券取得者数が多いところでも小学生数はあまり大きな変化はみられない。

一方、旅券取得者数と外国人数については、旅券取得者数が多いところは、外国人数は多くなっている傾向がみられる。

したがって、旅券取得者数と外国人数の相関係数のほうが大きいといえるので、正しくない。