

## データの分析

**1**

右の度数分布表は、A 高校の 20 人について、1 日にみたインターネットの時間を記入したものである。

次の問いに答えよ。

- (1) インターネットをみた時間が 95 分未満の生徒は何人いるか。
- (2) インターネットをみた時間が 105 分以上の生徒は全体の何%であるか。

階級 (分)	階級値 (分)	度数 (人)	相対度数
65 以上～75 未満	70	1	0.05
75 ～85	80	3	0.15
85 ～95	90	4	0.20
95 ～105	100	5	0.25
105 ～115	110	3	0.15
115 ～125	120	2	0.10
125 ～135	130	2	0.10
合計		20	1.00

### 解答

- (1)  $1+3+4=8$  (人)
- (2)  $0.15+0.10+0.10=0.35$   
したがって **35%**

**2**

(1) 次のデータは、ある 7 人の家にある観葉植物の個数  $x$  を調べたものである。

2, 5, 0, 6, 4, 1, 3 (個)

このデータの平均値  $\bar{x}$  を求めよ。

(2) 右の表から、インターネットをみた時間  $x$  の平均値を求めよ。

階級 (分)	階級値 $x$ (分)	度数 $f$ (人)
65 以上～75 未満	70	1
75 ～85	80	3
85 ～95	90	4
95 ～105	100	5
105 ～115	110	3
115 ～125	120	2
125 ～135	130	2
合計		20

**解答**

$$(1) \bar{x} = \frac{1}{7}(2+5+0+6+4+1+3) = \frac{21}{7} = 3 \text{ (個)}$$

$$(2) \bar{x} = \frac{1}{20}(70 \times 1 + 80 \times 3 + 90 \times 4 + 100 \times 5 + 110 \times 3 + 120 \times 2 + 130 \times 2)$$
$$= \frac{1}{20} \times 2000 = 100 \text{ (分)}$$

**3**

次のデータは、ある8人の昨年1年間のスポーツ観戦の回数である。

2, 0, 4, 1, 1, 9, 4, 1 (回)

- (1) このデータの中央値を求めよ。
- (2) このデータの最頻値を求めよ。

**解答**

(1) 小さい方から順に並べると 0, 1, 1, 1, 2, 4, 4, 9

これより、中央値は  $\frac{1+2}{2} = 1.5$  (回)

(2) 最頻値は 1 (回)

**4**

次のデータは、T社の従業員10人の年収を調べたものである。

510, 400, 430, 630, 520, 450, 420, 580, 540, 520 (万円)

次の問いに答えよ。

- (1) このデータの範囲を求めよ。
- (2) このデータの四分位数  $Q_1$ ,  $Q_2$ ,  $Q_3$  を求めよ。
- (3) このデータの四分位範囲と四分位偏差を求めよ。

**解答**

(1) 最大値は 630 万円，最小値は 400 万円であるから，範囲は  $630 - 400 = 230$  (万円)

(2) 小さい方から順に並べると 400, 420, 430, 450, 510, 520, 520, 540, 580, 630

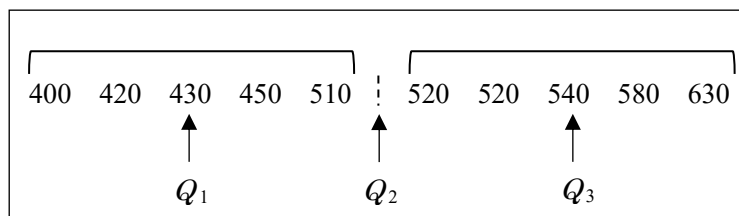
中央値から  $Q_2 = \frac{510 + 520}{2} = 515$  (万円)

前半部分の中央値から

$Q_1 = 430$  (万円)

後半部分の中央値から

$Q_3 = 540$  (万円)



(3)  $Q_1 = 430$ ,  $Q_3 = 540$  であるから

四分位範囲は  $540 - 430 = 110$  (万円)

四分位偏差は  $\frac{110}{2} = 55$  (万円)

**5**

次のデータは，T社の従業員 10 人，N社の従業員 11 人の年収を調べたものである。それぞれの箱ひげ図をかき，散らばりの度合いを比較せよ。

T社： 510, 400, 430, 630, 520, 450, 420, 580, 540, 520 (万円)

N社： 400, 600, 360, 420, 520, 350, 700, 400, 480, 570, 380 (万円)

**解答**

T社の 最小値， $Q_1$ ， $Q_2$ ， $Q_3$ ，最大値は，**4** から 400, 430, 515, 540, 630 (万円)

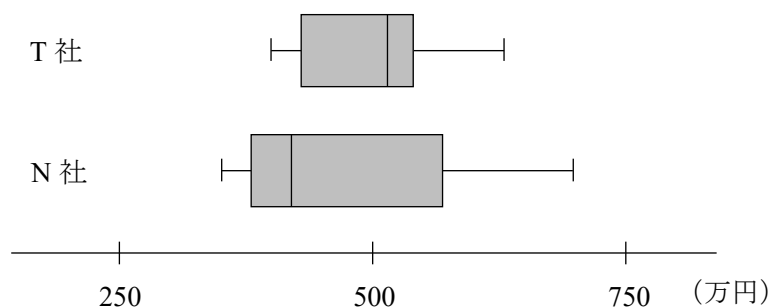
N社の 最小値， $Q_1$ ， $Q_2$ ， $Q_3$ ，最大値を求める。

小さい方から順に並べると 350, 360, 380, 400, 400, 420, 480, 520, 570, 600, 700

これから，最小値，最大値は 350, 700 (万円)

また  $Q_2 = 420$  (万円)       $Q_1 = 380$  (万円)       $Q_3 = 570$  (万円)

以上から，T社とN社の箱ひげ図は次のようになる。



箱ひげ図から読み取れる範囲や四分位範囲から，N社よりもT社の方が散らばりの度合いが小さい。

6

次のデータは、H社の従業員9人の年収を調べたものである。外れ値があれば求めて、箱ひげ図をかけ。

H社：370, 380, 400, 430, 450, 450, 500, 520, 1000（万円）

**解答**

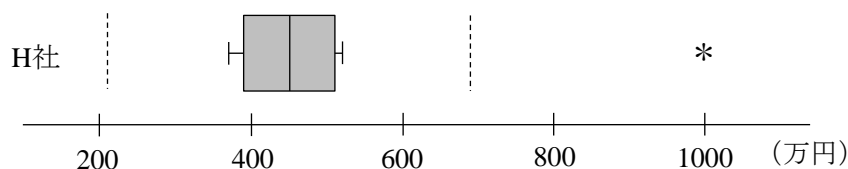
中央値（第2四分位数  $Q_2$ ）は 450（万円）

第1四分位数  $Q_1$  は  $\frac{380+400}{2} = 390$  (万円)      第3四分位数  $Q_3$  は  $\frac{500+520}{2} = 510$  (万円)

また、四分位範囲は  $510-390=120$ （万円）であるから（四分位範囲） $\times 1.5 = 120 \times 1.5 = 180$   
 よって、箱の両端から180離れた値、すなわち210や690より外側にある値は外れ値である。

したがって、**外れ値は1000（万円）**

箱ひげ図は次の図のようになる。



7

次のデータは、ある7人の家にある観葉植物の個数  $x$  を調べたものである。

2, 5, 0, 6, 4, 1, 3（個）

このデータの分散  $s^2$  を求めよ。

**解答**

平均値は  $\bar{x} = \frac{1}{7}(2+5+0+6+4+1+3) = \frac{21}{7} = 3$ （個）

偏差は -1, 2, -3, 3, 1, -2, 0（個）

よって、分散は  $s^2 = \frac{1}{7}\{(-1)^2+2^2+(-3)^2+3^2+1^2+(-2)^2+0^2\} = \frac{28}{7} = 4$

**別解** 平均値は  $\bar{x} = 3$ （個）       $\overline{x^2} = \frac{1}{7}(2^2+5^2+0^2+6^2+4^2+1^2+3^2) = \frac{91}{7} = 13$

したがって  $s^2 = 13 - 3^2 = 4$

8

次のデータは、ある7人の昨年1年間のスポーツ観戦の回数  $x$  である。

2, 0, 4, 1, 1, 9, 4 (回)

このデータの標準偏差  $s$  を求めよ。ただし、 $\sqrt{2} = 1.4$  とする。

**解答**

平均値は  $\bar{x} = \frac{1}{7}(2+0+4+1+1+9+4) = \frac{21}{7} = 3$  (回)

偏差は  $-1, -3, 1, -2, -2, 6, 1$  (回)

よって、標準偏差は  $s = \sqrt{\frac{1}{7}\{(-1)^2+(-3)^2+1^2+(-2)^2+(-2)^2+6^2+1^2\}} = \sqrt{\frac{56}{7}} = 2\sqrt{2} = 2.8$  (回)

**別解** 平均値は  $\bar{x} = 3$  (回)  $\bar{x^2} = \frac{1}{7}(2^2+0^2+4^2+1^2+1^2+9^2+4^2) = \frac{119}{7} = 17$

したがって  $s = \sqrt{17-3^2} = 2\sqrt{2} = 2.8$  (回)

9

右のデータは、ある7人の家にある観葉植物の個数  $x$  と、昨年1年間のスポーツ観戦の回数  $y$  を調べたものである。

7人	A	B	C	D	E	F	G
観葉植物 (個)	2	5	0	6	4	1	3
スポーツ観戦 (回)	2	0	4	1	1	9	4

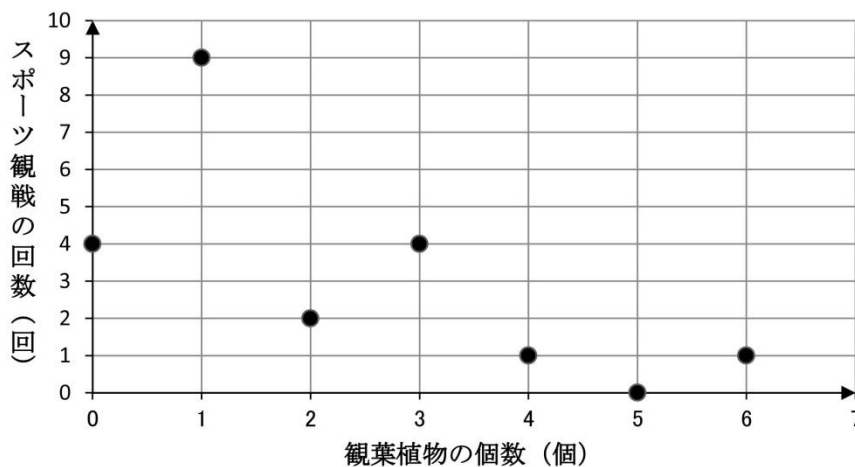
観葉植物の個数  $x$  を横軸、昨年1年間の

スポーツ観戦の回数  $y$  を縦軸として散布図をかけ。また、 $x$  と  $y$  の間には、どのような相関関係があるといえるか。

**解答**

散布図は右のようになる。

右の散布図から、 $x$  と  $y$  の間には負の相関関係があるといえる。



10

右のデータは、ある7人の家にある観葉植物の個数  $x$  と、昨年1年間のスポーツ観戦の回数  $y$  を調べたものである。

7人	A	B	C	D	E	F	G
観葉植物(個)	2	5	0	6	4	1	3
スポーツ観戦(回)	2	0	4	1	1	9	4

$x$  と  $y$  の相関係数  $r$  を求めよ。ただし、

$\sqrt{2} = 1.4$  とする。また、 $x$  と  $y$  の間には、どのような相関関係があるといえるか。

解答

$$\bar{x} = \frac{1}{7}(2+5+0+6+4+1+3) = \frac{21}{7} = 3, \quad \bar{y} = \frac{1}{7}(2+0+4+1+1+9+4) = \frac{21}{7} = 3 \text{ から、次のような表を作る。}$$

7人	$x$	$y$	$x - \bar{x}$	$y - \bar{y}$	$(x - \bar{x})^2$	$(y - \bar{y})^2$	$(x - \bar{x})(y - \bar{y})$
A	2	2	-1	-1	1	1	1
B	5	0	2	-3	4	9	-6
C	0	4	-3	1	9	1	-3
D	6	1	3	-2	9	4	-6
E	4	1	1	-2	1	4	-2
F	1	9	-2	6	4	36	-12
G	3	4	0	1	0	1	0
合計	21	21			28	56	-28

$$\text{したがって } r = \frac{-28}{\sqrt{28} \sqrt{56}} = -\frac{1}{\sqrt{2}} = -\frac{\sqrt{2}}{2} = -0.7$$

このことから、 $x$  と  $y$  の間には **強い負の相関関係がある** といえる。

11

高校生の3人に1人が知っているウェブサイトがある。最近になってSNSで話題になっているのを目にするなど、少し知名度が上がっている気がしたので、あらためて高校生10人にそのウェブサイトを知っているか聞いたところ、6人が知っていると回答した。

このことから、そのウェブサイトは知名度が上がったと判断してよいか。仮説検定の考え方をを用い、基準となる確率を0.05として考察せよ。

ただし、さいころ10個を投げ1, 2の目が出た個数を記録する実験を1000回くり返したところ、右の表のようになった。考察にはこの結果を用いよ。

1, 2の目が出た個数	度数
0	22
1	89
2	186
3	256
4	243
5	131
6	56
7	15
8	1
9	1
10	0
合計	1000

### 解答

正しいかどうか判断したい主張を、「そのウェブサイトの知名度は上がった。」とし、その主張に反する仮説を「そのウェブサイトの知名度は3人に1人の割合から変わらない。」とする。

さいころ投げの結果から、1000回中1, 2の目が6個以上出た度数は

$$56+15+1+1+0=73$$

よって、仮説のもとでそのウェブサイトを10人中6人以上が知っていると回答する確率は

$$\frac{73}{1000} = 0.073 \text{ 程度であると考えられる。}$$

これは0.05より大きいから、仮説「そのウェブサイトの知名度は3人に1人の割合から変わらない。」が正しくなかったとは考えられない。

したがって、そのウェブサイトは知名度が上がったかどうか判断できない。