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。

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。
解答

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

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

$= \frac{1}{20} \times 2000 = 100$ (分)

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)

次のデータは、C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 \( Q_2 = \frac{520+500}{2} = 510 \) （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 であるからの四分位範囲は 540−430=110 （万円）
四分位偏差は \( \frac{110}{2} = 55 \) （万円）

次のデータは，C 社の従業員 10 人，D 社の従業員 9 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。

C 社 : 510，400，430，630，520，450，420，580，540，520 （万円）
D 社 : 430，500，520，450，450，380，370，1000，400 （万円）

解答
C 社の 最小値，\( Q_1，Q_2，Q_3，最大値は \( Q_4 \) から 400，430，515，540，630 （万円）
D 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 370，380，400，430，450，500，520，1000
これから，最小値，最大値は 370，1000 （万円）
また \( Q_2 = 450 \) （万円） \( Q_1 = \frac{380+400}{2} = 390 \) （万円） \( Q_1 = \frac{500+520}{2} = 510 \) （万円）
以上から，C 社と D 社の箱ひげ図は次のようにになる。

箱ひげ図から読み取れる範囲や四分位範囲から，D 社よりも C 社の方が散らばりの度合いが小さい。
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 \[ \overline{x} = \frac{1}{7} (2+5+0+6+4+1+3) = \frac{21}{7} = 3 \text{（個）} \]
偏差は -1, 2, -3, 3, 1, -2, 0（個）
よって、分散は \[ s^2 = \frac{1}{7} \left( (-1)^2 + 2^2 + (-3)^2 + 3^2 + 1^2 + (-2)^2 + 0^2 \right) = \frac{28}{7} = 4 \]

別解

平均値は \[ \overline{x} = 3 \text{（個）} \]
偏差は -1, 2, -3, 3, 1, -2, 0（個）
よって、標準偏差は \[ s = \sqrt{\frac{1}{7} \left( (-1)^2 + 2^2 + (-3)^2 + 3^2 + 1^2 + (-2)^2 + 0^2 \right)} = \sqrt{\frac{28}{7}} = 2\sqrt{2} = 2.8 \text{（回）} \]

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、\[ \sqrt{2} = 1.4 \text{とする。} \]

解答

平均値は \[ \overline{x} = \frac{1}{7} (2+0+4+1+1+9+4) = \frac{21}{7} = 3 \text{（回）} \]
偏差は -1, -3, 1, -2, -2, 6, 1（回）
よって、標準偏差は \[ s = \sqrt{\frac{1}{7} \left( (-1)^2 + (-3)^2 + 1^2 + (-2)^2 + 6^2 + 1^2 \right)} \]
\[ = \frac{\sqrt{56}}{7} \approx 2.8 \text{（回）} \]

別解

平均値は \[ \overline{x} = 3 \text{（回）} \]
偏差は -1, -3, 1, -2, -2, 6, 1（回）
よって、標準偏差は \[ s = \sqrt{17-3^2} \approx 2\sqrt{2} = 2.8 \text{（回）} \]

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。

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。

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。

\( x \) 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
\( \bar{x} = \frac{1}{7}(2+5+0+6+4+1+3) = \frac{21}{7} = 3 \)、\( \bar{y} = \frac{1}{7}(2+0+4+1+9+4) = \frac{21}{7} = 3 \)から、次のような表を作成する。

<table>
<thead>
<tr>
<th>7人</th>
<th>( x )</th>
<th>( y )</th>
<th>( x - \bar{x} )</th>
<th>( y - \bar{y} )</th>
<th>( (x - \bar{x})^2 )</th>
<th>( (y - \bar{y})^2 )</th>
<th>( (x - \bar{x})(y - \bar{y}) )</th>
</tr>
</thead>
<tbody>
<tr>
<td>A</td>
<td>2</td>
<td>2</td>
<td>-1</td>
<td>-1</td>
<td>1</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>B</td>
<td>5</td>
<td>0</td>
<td>2</td>
<td>-3</td>
<td>4</td>
<td>9</td>
<td>-6</td>
</tr>
<tr>
<td>C</td>
<td>0</td>
<td>4</td>
<td>-3</td>
<td>1</td>
<td>9</td>
<td>1</td>
<td>-3</td>
</tr>
<tr>
<td>D</td>
<td>6</td>
<td>1</td>
<td>3</td>
<td>-2</td>
<td>9</td>
<td>4</td>
<td>-6</td>
</tr>
<tr>
<td>E</td>
<td>4</td>
<td>1</td>
<td>1</td>
<td>-2</td>
<td>1</td>
<td>4</td>
<td>-2</td>
</tr>
<tr>
<td>F</td>
<td>1</td>
<td>9</td>
<td>-2</td>
<td>6</td>
<td>4</td>
<td>36</td>
<td>-12</td>
</tr>
<tr>
<td>G</td>
<td>3</td>
<td>4</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>1</td>
<td>0</td>
</tr>
<tr>
<td>合計</td>
<td>21</td>
<td>21</td>
<td></td>
<td></td>
<td>28</td>
<td>56</td>
<td>-28</td>
</tr>
</tbody>
</table>

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。