確　認　試　験　問　題

データの分析

解　答　例　・　解　説

試験時間： 90 分

問題数： 5 問

配点： 100 点

最終改訂日：2026/04/24

1　第 1 問：代表値・分散・標準偏差

1.1　問題

ある学級の生徒 20 人が 10 点満点の試験を受けたところ，次のような結果であった。

表 1.1: 生徒と試験の得点

生徒	A	B	C	D	E	F	G	H	I	J
得点	1	2	7	8	6	4	4	6	7	8
生徒	K	L	M	N	O	P	Q	R	S	T
得点	9	7	3	6	7	5	6	8	10	6

得点のデータに関して，以下の問いに答えよ。必要ならば，$2.2 < \sqrt{5} < 2.3$ であることを用いてよい。

(1)　このデータの平均値 μ を求めよ。

(2)　このデータの中央値 m を求めよ。

(3)　このデータの最頻値 M を求めよ。

(4)　このデータの分散 V を求めよ。

(5)　このデータの標準偏差 s を求めよ。結論には根号を用いてよい。

(6)　得点が x である生徒の偏差値得点 z を

$$z = \frac{x-\mu}{s} \cdot 10 + 50$$

で定める。このとき，偏差値得点が 60 以上である生徒の人数を求めよ。

1.2　解答

得点を x とおく。各行毎の得点の合計と，$(x-\mu)^2$ とその合計を追記すると次のようになる。

生徒	A	B	C	D	E	F	G	H	I	J	合計
得点	1	2	7	8	6	4	4	6	7	8	53
$(x-\mu)^2$	25	16	1	4	0	4	4	0	1	4	59
生徒	K	L	M	N	O	P	Q	R	S	T	合計
得点	9	7	3	6	7	5	6	8	10	6	67
$(x-\mu)^2$	9	1	9	0	1	1	0	4	16	0	41

(1)　20 人全員の得点の合計は $53 + 67 = 120$ であるから，

$$\mu = \frac{1}{20} \cdot 120 = 6$$

である。

(2)　得点を小さい順に並べると，

$$1, 2, 3, 4, 4, 5, 6, 6, 6, 6, 6, 7, 7, 7, 7, 8, 8, 8, 9, 10$$

となり，10 番目，11 番目の数値はともに 6 である。20 個のデータの中央値は 10 番目と 11 番目の値の平均だから，$m = 6$ である。

(3)　最も多い得点は，6 が 5 人である。よって $M = 6$ である。

(4)　各データの $(x-\mu)^2$ の合計は $59 + 41 = 100$ であるから，

$$V = \frac{1}{20} \cdot 100 = 5$$

である。

(5)

$$s = \sqrt{V} = \sqrt{5}$$

(6)　与式に (1)，(5) で求めた値を代入し，

$$z = \frac{x-6}{\sqrt{5}} \cdot 10 + 50 = 2\sqrt{5}(x-6) + 50$$

である。この式から x が大きいほど z も大きくなることが分かるため，点数 x の大きい生徒から順に z を計算し，何点までが $z \geqq 60$ を満たすかを調べる。

- $x = 10$ のとき，$z = 8\sqrt{5} + 50$ である。$2.2 < \sqrt{5}$ を用い，

  $$8\sqrt{5} + 50 > 8 \cdot 2.2 + 50$$

  より $z > 67.6$ であるから，得点が 10 の生徒の偏差値得点は 60 以上である。

- $x = 9$ のとき，$z = 6\sqrt{5} + 50$ である。

  $$6\sqrt{5} + 50 > 6 \cdot 2.2 + 50$$

  より $z > 63.2$ であるから，得点が 9 の生徒の偏差値得点は 60 以上である。

- $x = 8$ のとき，$z = 4\sqrt{5} + 50$ である。$\sqrt{5} < 2.3$ を用い，

  $$4\sqrt{5} + 50 < 4 \cdot 2.3 + 50$$

  より $z < 59.2$ であるから，得点が 8 の生徒の偏差値得点は 60 未満である。

- x が小さいほど z も小さくなるため，$x < 7$ のときの z は $x = 8$ のときの z よりも小さい。よって得点が 7 以下の生徒の偏差値得点は 60 未満である。

以上から，偏差値得点が 60 以上の生徒の人数は，得点が 9 または 10 の生徒の人数であり，それぞれ 1 人ずつが該当するため，偏差値得点が 60 以上の生徒は 2 人である。

1.3　解説

(1)　平均値は，全ての値を均したときの値で，計算では $\dfrac{(\text{データの合計値})}{(\text{データの個数})}$ で求められます。問題では生徒が 20 人いますが，10 人ごとに合計値を計算し，さらにそれらを足し合わせて最終的な合計値を求めています。紙面に収まる表の大きさにするという意味もありますが，程良い個数でデータを区切ることで，計算ミスを発生しづらくするという工夫も込められています。

データの大きさ（個数）がわかっていれば，合計値を求めることで平均値を確実に求められま

すが，例えば

$$5,\ 5,\ 5,\ 4,\ 6$$

のような 5 個の値からなるデータがあるとき，値が 6 のものから 4 のものに 1 だけ移動させると，全ての値が 5 になるため，平均値は 5 である，というような考え方もできます。

(2)　中央値は，値を大きさで順番に並べたときの真ん中の値ですが，データの大きさが奇数のときと偶数のときで求め方が変わります。データの大きさが奇数のときは，ちょうど真ん中として 1 つの値が該当します。例えばデータが 11 個からなるとき，

$$\overbrace{1,\ 2,\ 3,\ 4,\ 5}^{\text{下位}},\ 6,\ \overbrace{7,\ 8,\ 9,\ 10,\ 11}^{\text{上位}}$$

のように，ちょうど 6 番目の値 6 が中央値となります。データの大きさが偶数のとき，ちょうど真ん中が 1 つに定まらないため，真ん中の前後の値を平均を取ります。例えばデータが 10 個からなるとき，

$$\overbrace{1,\ 2,\ 3,\ 4,\ 5}^{\text{下位}},\ \overbrace{6,\ 7,\ 8,\ 9,\ 10}^{\text{上位}}$$

のようになるため，真ん中の前後である 5 番目と 6 番目の間の値の平均をとって $\dfrac{5+6}{2}=5.5$ が中央値となります。

(3)　分散は，値と平均の差の 2 乗の平均値です。s^2 や σ^2 のように表すこともあります。

(4)　標準偏差は，分散に平方根をとったものです。

(5)　学校で受ける試験の成績で用いる「偏差値」がこれに該当します。平均点に該当する偏差値を 50 とし，平均点から標準偏差だけ点数が離れる毎に 10 が加減されます。
また，本設問では，次のように，先に $z \geqq 60$ を満たす x を求めることもできます。

(5) 別解　与式に (1)，(5) で求めた値を代入し，それが 60 以上であるということを式で表すと

$$\frac{x-6}{\sqrt{5}}\cdot 10+50 \geqq 60$$

である。これを整理すると

$$x \geqq 6+\sqrt{5}$$

となる。$2.2<\sqrt{5}<2.3$ より $8.2<6+\sqrt{5}<8.3$ であるから，表の中で上の不等式を満たす得点は 9 か 10 のみで，その人数は 2 人である。

1.4　採点

【第 1 問　20 点】

(1)〜(5) は完答で得点。

(1)　3 点

(2)　3 点

(3)　3 点

(4)　3 点

(5)　3 点

(6)　5 点

- 偏差値得点が 60 以上の生徒が，得点 9，10 の生徒のみであることを求められているか。

2　第 2 問：図の読み取りとデータの分析

2.1　問題

図 2.1 は，ある場所の 16 日間にわたる各日の気温のデータをヒストグラムに，図 2.2 は，ある学校に属する生徒 12 人の体力テストの成績を箱ひげ図に，図 2.3 は，ある 13 人の集団の身長と体重を測定した結果を散布図にして表したものである。集計したデータは，いずれも整数値であったとする。また，図 2.1 のヒストグラムについて，各階級は 15 ℃以上 18 ℃未満，18 ℃以上 21 ℃未満，...，30 ℃以上 33 ℃未満のように区切っている。

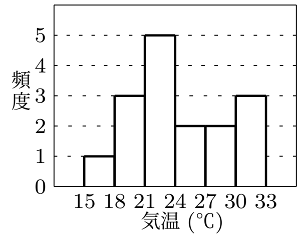

図 2.1

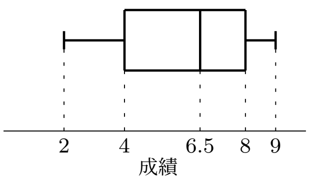

図 2.2

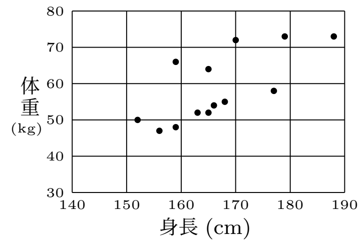

図 2.3

以下の (1) 〜 (8) の記述に対して，それが正しいかどうかを理由をつけて判定せよ。

(1)　図 2.1 から，気温 (℃) のデータの範囲は 18 と確定する。

(2)　図 2.1 から，気温 (℃) のデータの中央値は必ず 21 以上 24 未満の範囲内にあるといえる。

(3)　図 2.1 から，気温 (℃) のデータの平均値は必ず 21 以上 24 未満の範囲内にあるといえる。

(4)　図 2.2 から，成績のデータの範囲は 7 と確定する。

(5)　図 2.2 から，成績のデータの平均値は必ず 6 以上であるといえる。

(6)　図 2.2 から，成績が 7 の生徒が必ず存在するといえる。

(7)　図 2.3 から，身長が 160 cm 以上，体重が 60 kg 以上の生徒が 3 人以上いることが分かる。

(8)　図 2.3 に対応する身長 (cm) と体重 (kg) のデータの共分散を計算し，小数第二位を四捨五入すると 63.2 であった。このとき，身長 (cm) と体重 (kg) の相関係数は正であるといえる。

2.2　解答

(1)　データの値がいずれも整数値であることを考慮すると，ヒストグラムから，考えられる最も小さい値は 15，最も大きい値は 32 である。よって，データの範囲の最大値は $32 - 15 = 17$ であり，データの範囲は 18 とはならない。したがって，設問の内容は正しくない。

(2)　データの大きさは 16 であるため，小さい順に値を並べたとき，8 番目の値と 9 番目の値の平均値がこのデータの中央値である。21 以上 24 未満とそれ以外に区切ってデータを見ると，21 未満が 4 件（つまり小さい順に 1〜4 番目），24 以上が 7 件（つまり小さい順に 10〜16 番目），21 以上 24 未満が 5 件（つまり小さい順に 5〜9 番目）であるため，8 番目と 9 番目の値はいずれも 21 以上 24 未満の範囲にある。よって，ともに 21 以上 24 未満である 8 番目の値と 9

番目の値の平均値は 21 以上 24 未満であり，これは中央値が 21 以上 24 未満であることを示している。したがって，設問の内容は正しい。

(3) データの値がいずれも整数値であるうえで，図に対応するデータで平均値が大きくなるような値の取り方を考えると，

17, 20, 20, 20, 23, 23, 23, 23, 23, 26, 26, 29, 29, 32, 32, 32

という場合が考えられ，このデータの平均値は 24.875 である。よって，データの平均値は 24 以上になり得るといえる。したがって，設問の内容は正しくない。

(4) データの最大値は 9，最小値は 2 であるから，データの範囲は $9-2=7$ である。したがって，設問の内容は正しい。

(5) 図に対応するデータで平均値が小さくなるような値の取り方を考えると，

2, 2, 4, 4, 4, 6, 7, 7, 8, 8, 8, 9

という場合が考えられ，このデータの平均値は $\dfrac{23}{4}$ となり，これは 6 より小さい。よって，平均値は 6 未満になり得る。したがって，設問の内容は正しくない。

(6) 図に対応するデータの取り方として，

2, 3, 3, 5, 5, 5, 8, 8, 8, 8, 8, 9

という場合が考えられ，このデータには値 7 は存在しない。よって，成績 7 の生徒が存在しない場合があり得る。したがって，設問の内容は正しくない。

(7) 身長が 160 cm 以上，体重が 60 kg 以上の場所に位置する点は 4 つあるため，該当する生徒は 4 人，つまり 3 人以上いる。したがって，設問の内容は正しい。

(8) 身長と体重の相関係数 r は

$$r=\frac{(\text{身長と体重の共分散})}{(\text{身長の標準偏差})(\text{体重の標準偏差})}$$

で与えられる。いま，身長と体重の共分散は設問より正，身長と体重の標準偏差は，全てが同じ値でない以上，必ず正であるから，r も正である。したがって，設問の内容は正しい。

2.3　解説

(1) データの範囲は最大値と最小値の差から得られます。解答の通り，そもそもデータの範囲が 18 になることはありません。図から，最小値として考えられるのは 15，16，17 のいずれかであり，最大値は，30，31，32 のいずれかであるため，データの範囲としてはいくつかの候補があります。

(2) 中央値を計算する 8 番目と 9 番目の値がいずれも 21 以上 24 未満に収まっているため，その平均値である中央値もこの範囲内にあります。

(3) 図に対応するデータで，極端な例を考え，平均値を 21 以上 24 未満の範囲から飛び出せないかと考えます。全てを各階級の最大値にすると，平均値を 24 以上にできました。

(4) 箱ひげ図は最大値と最小値が確定するため，データの範囲も確定します。

(5) 図に対応するデータで，値の小さい極端な例を考えています。最小値，四分位数，最大値が図

を満たしつつ，できるだけ小さい値を考えると，平均値を 6 未満にすることができました。

(6) データの大きさが 12 のため，小さい順で 6 番目と 7 番目の平均値が中央値になります。中央値が 6.5 のため，6 番目と 7 番目として考えられるものは，6 と 7 もありますが，5 と 8 も考えられます。あとは 7 が入らないように 12 個のデータを構築すれば反例になります。

(7) 図から，横軸の身長で 160 の線上または右側，縦軸の体重で 60 の線上または上側を見ます。

(8) 変量 x と変量 y の相関係数は，$\dfrac{(x \text{ と } y \text{ の共分散})}{(x \text{ の標準偏差})(y \text{ の標準偏差})}$ で定められます。標準偏差は 0 でない限り必ず正であるため，共分散の正負がそのまま相関係数の正負になります。ただし，変量 x，y のどちらかでも，値がすべて等しい場合，相関係数が 0 になり分母が 0 になるため，相関係数を定義できません。図から明らかではありますが，解答ではこのことに触れました。

2.4　採点

【第 2 問　24 点】

(1)〜(8) について各 3 点。理由が正しいか，結論が正しいかを見る。

3　第 3 問：平均値と分散の関係式

3.1　問題

問い A，B に答えよ。

A

(1)　変量 x についての n 個のデータの値において，x のデータの平均値を $\overline{x}$，x のデータの各値をそれぞれ 2 乗したものの平均値を $\overline{x^2}$ とおくとき，x の分散 v に関して，

$$v = \overline{x^2} - \left(\overline{x}\right)^2$$

が成り立つ。$n = 4$ の場合について，この等式が成り立つことを計算によって確かめよ。例えば，4 つの値 x_1，x_2，x_3，x_4 について，その分散 v と $\overline{x^2} - \left(\overline{x}\right)^2$ をそれぞれ計算し，それらから同じ式を得られることを確かめるとよい。

(2)　2 つの集団 A，B があり，それぞれが果樹園で果物の収穫を行った。集団の人数と，収穫した果物の個数について，集団内の平均値と分散を調べたものを次の表に示した。

表 3.1: 集団と収穫した果物の個数

集団	人数	平均値	分散
A	36	29	31
B	54	24	36

このとき，A と B を合わせた集団全体について，収穫した果物の個数の平均値と分散を求めよ。

B

(3)　変量 x のデータの大きさが n であるとする。a，b を実数の定数とし，変量 x について，変量 y を $y = ax + b$ と定める。このとき，変量 x のデータの平均値を $\overline{x}$，変量 y のデータの平均値を $\overline{y}$，変量 x のデータの分散を v_x，変量 y のデータの分散を v_y とおくと，

$$\overline{y} = a\overline{x} + b, \quad v_y = a^2 v_x$$

が成り立つ。$n = 4$ の場合について，この等式が成り立つことを計算によって確かめよ。例えば，変量 x についての 4 つの値 x_1，x_2，x_3，x_4 について，対応する変量 y の値を y_1，y_2，y_3，y_4 とおき，それぞれデータの平均値と分散を求め，それらから関係式を得られることを確かめるとよい。

(4)　次の 6 つの値からなるデータの平均値と分散を求めよ。

$$\begin{gathered} 314.15926535, \quad 514.15926535, \quad 1014.15926535, \\ 764.15926535, \quad -285.84073465, \quad -135.84073465 \end{gathered}$$

3.2　解答

(1)　4 つの値 x_1，x_2，x_3，x_4 について，分散 v は，その定義から

$$\begin{aligned} v &= \frac{1}{4}\left\{\left(x_1-\overline{x}\right)^2+\left(x_2-\overline{x}\right)^2+\left(x_3-\overline{x}\right)^2+\left(x_4-\overline{x}\right)^2\right\} \\ &= \frac{1}{4}\Bigl\{x_1^2-2x_1\overline{x}+\left(\overline{x}\right)^2+x_2^2-2x_2\overline{x}+\left(\overline{x}\right)^2 \\ &\qquad\qquad +x_3^2-2x_3\overline{x}+\left(\overline{x}\right)^2+x_4^2-2x_4\overline{x}+\left(\overline{x}\right)^2\Bigr\} \\ &= \frac{1}{4}\left\{x_1^2+x_2^2+x_3^2+x_4^2-2\overline{x}(x_1+x_2+x_3+x_4)+4\left(\overline{x}\right)^2\right\} \\ &= \frac{1}{4}\left\{x_1^2+x_2^2+x_3^2+x_4^2-8\left(\overline{x}\right)^2+4\left(\overline{x}\right)^2\right\} \\ &= \frac{1}{4}\left\{x_1^2+x_2^2+x_3^2+x_4^2-4\left(\overline{x}\right)^2\right\} \\ &= \frac{1}{4}\left(x_1^2+x_2^2+x_3^2+x_4^2\right)-\left(\overline{x}\right)^2 \end{aligned} \tag{3.1}$$

である。また，

$$\overline{x^2}=\frac{1}{4}\left(x_1^2+x_2^2+x_3^2+x_4^2\right)$$

より，

$$\overline{x^2}-\left(\overline{x}\right)^2=\frac{1}{4}\left(x_1^2+x_2^2+x_3^2+x_4^2\right)-\left(\overline{x}\right)^2 \tag{3.2}$$

である。(3.1)，(3.2)から，v と $\overline{x^2}-\left(\overline{x}\right)^2$ を計算すると同じ式が得られたため，$n=4$ においては $v=\overline{x^2}-\left(\overline{x}\right)^2$ が成り立つことが確かめられた。

(2)　A と B を合わせた $36+54=90$ 人全員の集団を U とし，集団 G（G には A，B，U が入る）について，人数を h_{G}，平均値を $\overline{x_{\mathrm{G}}}$，分散を v_{G}，各値の 2 乗の平均値を $\overline{x_{\mathrm{G}}^2}$ とおく。まず，U の平均値 $\overline{x_{\mathrm{U}}}$ を求める。

$$\begin{aligned} \overline{x_{\mathrm{U}}} &= \frac{1}{h_{\mathrm{U}}}(\text{U の合計値}) \\ &= \frac{1}{h_{\mathrm{U}}}\{(\text{A の合計値})+(\text{B の合計値})\} \\ &= \frac{1}{h_{\mathrm{U}}}\left\{\overline{x_{\mathrm{A}}}\cdot h_{\mathrm{A}}+\overline{x_{\mathrm{B}}}\cdot h_{\mathrm{B}}\right\} \\ &= \frac{1}{90}(29\cdot 36+24\cdot 54) \\ &= 26 \end{aligned}$$

次に U の分散 v_{U} を求める。まず，(1) で得られた等式を変形した

$$\overline{x^2}=v-\left(\overline{x}\right)^2$$

を A，B に適用した

$$\overline{x_{\mathrm{A}}^2}=v_{\mathrm{A}}+\left(\overline{x_{\mathrm{A}}}\right)^2,\quad \overline{x_{\mathrm{B}}^2}=v_{\mathrm{B}}+\left(\overline{x_{\mathrm{B}}}\right)^2$$

に値を代入し，

$$\overline{x_{\mathrm{A}}^2}=31+29^2=872,$$

$$\overline{x_{\mathrm{B}}^2} = 36 + 24^2 = 612$$

を得る。これを用いて v_{U} を求めると，

$$\begin{aligned}
v_{\mathrm{U}} &= \overline{x_{\mathrm{U}}^2} - \left(\overline{x_{\mathrm{U}}}\right)^2 \\
&= \frac{1}{h_{\mathrm{U}}}(\text{U の各値の 2 乗の合計値}) - 26^2 \\
&= \frac{1}{90}\{(\text{A の各値の 2 乗の合計値}) + (\text{B の各値の 2 乗の合計値})\} - 26^2 \\
&= \frac{1}{90}\left(\overline{x_{\mathrm{A}}^2}\cdot h_{\mathrm{A}} + \overline{x_{\mathrm{B}}^2}\cdot h_{\mathrm{B}}\right) - 26^2 \\
&= \frac{1}{90}(872\cdot 36 + 612\cdot 54) - 26^2 \\
&= 40
\end{aligned}$$

である。したがって，A と B を合わせた集団 U の個数の平均値 $\overline{x_{\mathrm{U}}}$ は 26，分散 v_{U} は 40 である。

(3)　x のデータを x_1，x_2，x_3，x_4 とし，y のデータを

$$y_1 = ax_1 + b,\quad y_2 = ax_2 + b,\quad y_3 = ax_3 + b,\quad y_4 = ax_4 + b$$

とする。x と y それぞれについてデータの平均値を求めると，

$$\overline{x} = \frac{1}{4}(x_1 + x_2 + x_3 + x_4), \tag{3.3}$$

$$\begin{aligned}
\overline{y} &= \frac{1}{4}(y_1 + y_2 + y_3 + y_4) \\
&= \frac{1}{4}(ax_1 + b + ax_2 + b + ax_3 + b + ax_4 + b) \\
&= a\cdot\frac{1}{4}(x_1 + x_2 + x_3 + x_4) + b
\end{aligned} \tag{3.4}$$

となり，(3.3)と (3.4)を比べて $\overline{y} = a\overline{x} + b$ が得られる。また，分散を求めると，

$$v_x = \frac{1}{4}\left\{\left(x_1 - \overline{x}\right)^2 + \left(x_2 - \overline{x}\right)^2 + \left(x_3 - \overline{x}\right)^2 + \left(x_4 - \overline{x}\right)^2\right\}, \tag{3.5}$$

$$\begin{aligned}
v_y &= \frac{1}{4}\left\{\left(y_1 - \overline{y}\right)^2 + \left(y_2 - \overline{y}\right)^2 + \left(y_3 - \overline{y}\right)^2 + \left(y_4 - \overline{y}\right)^2\right\} \\
&= \frac{1}{4}\left\{\left(ax_1 + b - \left(a\overline{x} + b\right)\right)^2 + \left(ax_2 + b - \left(a\overline{x} + b\right)\right)^2 + \right. \\
&\qquad \left.\left(ax_3 + b - \left(a\overline{x} + b\right)\right)^2 + \left(ax_4 + b - \left(a\overline{x} + b\right)\right)^2\right\} \\
&= \frac{1}{4}\left\{\left(ax_1 - a\overline{x}\right)^2 + \left(ax_2 - a\overline{x}\right)^2 + \left(ax_3 - a\overline{x}\right)^2 + \left(ax_4 - a\overline{x}\right)^2\right\} \\
&= \frac{1}{4}\left\{a^2\left(x_1 - \overline{x}\right)^2 + a^2\left(x_2 - \overline{x}\right)^2 + a^2\left(x_3 - \overline{x}\right)^2 + a^2\left(x_4 - \overline{x}\right)^2\right\} \\
&= a^2\cdot\frac{1}{4}\left\{\left(x_1 - \overline{x}\right)^2 + \left(x_2 - \overline{x}\right)^2 + \left(x_3 - \overline{x}\right)^2 + \left(x_4 - \overline{x}\right)^2\right\}
\end{aligned} \tag{3.6}$$

となり，(3.5)と (3.6)を比べて $v_y = a^2 v_x$ が得られる。

(4)　与えられたデータを変量 x のものとし，変量 y を

$$y = \frac{1}{100}x - 0.1415926535 \tag{3.7}$$

とすると，y のデータは

$$3,\quad 5,\quad 10,\quad 7.5,\quad -3,\quad -1.5$$

となる。y のデータの平均値 $\overline{y}$ は

$$\overline{y} = \frac{1}{6}(3+5+10+7.5-3-1.5) = \frac{7}{2} = 3.5$$

である。また，分散 v_y は

$$\begin{aligned} v_y &= \frac{1}{6}\{(3-3.5)^2+(5-3.5)^2+(10-3.5)^2+ \\ &\qquad (7.5-3.5)^2+(-3-3.5)^2+(-1.5-3.5)^2\} \\ &= \frac{1}{6}(0.25+2.25+42.25+16+42.25+25) \\ &= \frac{128}{6} = \frac{64}{3} \end{aligned}$$

である。(3.7)より，(3) において $a = \dfrac{1}{100}$，$b = -0.1415926535$ とした場合を考え，

$$\overline{x} = \frac{1}{a}\left(\overline{y}-b\right) = 100\cdot(3.5+0.1415926535) = 364.15926535,$$

$$v_x = \frac{v_y}{a^2} = 10000\cdot\frac{64}{3} = \frac{640000}{3}$$

である。よって，元データの平均値 $\overline{x}$ は 364.15926535，分散 v_x は $\dfrac{640000}{3}$ である。

3.3　解説

(1)　分散の定義と，与えられた式を計算し，比較しています。v の計算時に，平均を求める式の分母を払った

$$x_1+x_2+x_3+x_4 = 4\overline{x}$$

を用いて式変形しています。

(2)　(1) を使います。このように，平均値と分散がわかっているグループが複数あるとき，そのグループを合わせた全体の平均値と分散も求められます。計算時に，

(データの合計値) = (データの平均値)(データの大きさ)

を用いています。

(3)　設問の通りに値をとり，平均値と分散の定義に従って計算しています。設問で提示された等式は，

- データの値すべてを a 倍すると，平均値も a 倍になり，分散は a^2 倍になる。
- データの値すべてを b だけずらすと，平均値も b ずれるが，分散には影響しない。

ということを表しています。

(4)　(3) を使います。小数点以下が第 8 位まであるため計算時に排除したいのと，1，10 の位も同じあるいは似た値が並んでいるため，まとめて簡単にしたいです。まず全体を 100 分の 1 倍にして，1 つめが 3.1415926535 となるようにします。ここから計算を複雑にしうる小数点以下を引いてしまいましょう。マイナスが付いている後ろ 2 つの値も，小数点以下を引くと計算しやすい値になります。こうして計算しやすい値にして平均値と分散を計算した後には，もとの

平均値と分散に戻すことを忘れずに行います。

3.4　採点

【第 3 問　20 点】

(1)　5 点

(2)　5 点

(3)　5 点

(4)　5 点

4　第 4 問：共分散・相関係数

4.1　問題

あなたは，シミュレーションソフトを使って仮想的な国家を構築できるとし，構築した国家では自動的に人々が経済活動を行うものとする。

いま，1 つ国家を構築したとする。この国家には首都近辺に地方が 8 つあり，各地方の首都からの距離 (km) と，その地方で生活する人々の収入 (Mvy) が次の表に示した通りであったとする。ただし，収入の単位は仮想国家独自のものである。

表 4.1: 各地方の首都からの距離と収入

地方	A	B	C	D	E	F	G	H
距離 (km)	17	23	26	27	38	43	50	56
収入 (Mvy)	19	14	11	12	15	13	12	8

以下の問いに答えよ。

(1)　距離 (km) の平均値，分散をそれぞれ求めよ。

(2)　収入 (Mvy) の平均値，分散をそれぞれ求めよ。

(3)　距離 (km) と収入 (Mvy) の共分散を求めよ。

(4)　距離 (km) と収入 (Mvy) の相関係数を求めよ。

(5)　仮に，B の収入 (Mvy) が 14 から 15 に増加し，E の収入が 15 から 14 に減少した場合，距離 (km) と収入 (Mvy) の相関係数は (4) で求めた値より大きくなるか，小さくなるか，変化しないか，理由をつけて判定せよ。

4.2　解答

距離 (km) を変量 x，収入 (Mvy) を変量 y とする。x のデータの平均値を m_x，分散を v_x，標準偏差を s_x とおき，y のデータの平均値を m_y，分散を v_y，標準偏差を s_y とおく。

x，y それぞれのデータの合計値，$(x-m_x)^2$，$(y-m_y)^2$，$(x-m_x)(y-m_y)$ の各値とその合計値を表に追記した。

表 4.2: 距離 x と収入 y

地方	A	B	C	D	E	F	G	H	合計
x	17	23	26	27	38	43	50	56	280
y	19	14	11	12	15	13	12	8	104
$(x-m_x)^2$	324	144	81	64	9	64	225	441	1352
$(y-m_y)^2$	36	1	4	1	4	0	1	25	72
$(x-m_x)(y-m_y)$	-108	-12	18	8	6	0	-15	-105	-208

(1)　$m_x = \dfrac{280}{8} = 35$，$v_x = \dfrac{1352}{8} = 169$ より，距離 (km) の平均値は 35，分散は 169 である。

(2)　$m_y = \dfrac{104}{8} = 13$，$v_y = \dfrac{72}{8} = 9$ より，収入 (Mvy) の平均値は 13，分散は 9 である。

(3)　x と y の共分散 c は $\dfrac{-208}{8} = -26$ である。

(4)　$s_x = \sqrt{v_x} = 13$，$s_y = \sqrt{v_y} = 3$ である。x と y の相関係数 r は

$$r = \frac{c}{s_x s_y} = \frac{-26}{13 \cdot 3} = -\frac{2}{3}$$

である。

(5)　表 4.2 の y の行において，B の列が 14 から 15 に，E の列が 15 から 14 に変化しても，y の行の合計は変わらないため，m_y は変化しない。また，$(y - m_y)^2$ の行において，B の列は 1 から 4 に，E の列が 4 から 1 に変化するが，合計の列は変化しない。よって v_y も変化しないため，s_y も変化しない。$(x - m_x)(y - m_y)$ の行において，B の列は -12 から -24 に，E の列は 6 から 3 に変化する。いずれも減少しているため，合計も小さくなる。よって c は小さくなる。x についてはデータのどの値も変化しないため，s_x は変化しない。よって，

$$r = \frac{c}{s_x s_y}$$

について，s_x，s_y は変化せず，c は小さくなるため，r は元より小さくなる。

4.3　解説

(4)　変量 x と変量 y の相関係数は,「x が大きくなるにつれて y も大きくなる傾向がある」と正,「x が大きくなるにつれて y が小さくなる傾向がある」と負になり，それぞれ正の相関関係，負の相関関係があると言います。

(5)　相関係数を構成する x の標準偏差，y の標準偏差，x と y の共分散がどのように変化するかを考えます。今回は共分散のみしか変化しなかったため，その変化がそのまま相関係数の変化を導きました。

4.4　採点

【第 4 問　20 点】

(1)　4 点

- 平均値と分散にそれぞれ 2 点

(2)　4 点

- 平均値と分散にそれぞれ 2 点

(3)　3 点

(4)　4 点

(5)　5 点

- 相関係数を構成する各値について，それぞれどのように変化するかを議論できたか。

5　第 5 問：仮説検定

5.1　問題

問い A，B に答えよ。

A

「当たり」または「はずれ」のどちらかが出るくじを 10 回引いたところ，「当たり」が 8 回出た。このくじは「当たり」が「はずれ」より出やすいといえるか。仮説検定の考え方を用い，有意水準 0.05 で考察せよ。ただし，各回のくじの結果は，別の回のくじの結果に影響を及ぼさないとする。また，表と裏の出る確率がどちらも $\frac{1}{2}$ である硬貨を 10 回投げたとき，ちょうど n 回表が出る確率 $P(n)$ を表 5.1 に示したため，必要ならばこれを用いてもよい。

表 5.1: n 回表が出る確率 $P(n)$

n	0	1	2	3	4	5
$P(n)$	$\frac{1}{1024}$	$\frac{10}{1024}$	$\frac{45}{1024}$	$\frac{120}{1024}$	$\frac{210}{1024}$	$\frac{252}{1024}$
n	6	7	8	9	10	
$P(n)$	$\frac{210}{1024}$	$\frac{120}{1024}$	$\frac{45}{1024}$	$\frac{10}{1024}$	$\frac{1}{1024}$	

B

ある企業が，年に 1 度行われる総会の後，従業員全員に対して，業務時間の快適さに関するアンケートを取った。「快適である」，「快適でない」の 2 つの選択肢を提示して回答を得たところ，「快適である」と回答したのは全体の $\frac{1}{3}$ であった。後日，その企業がデジタル化の一環で事務作業の一部を自動化するサービスを契約し，しばらくして従業員 8 人に同様のアンケートを取ったところ，「快適である」と回答したのは 6 人であった。この結果から，業務について「快適である」と回答した人は増えたといえるか。仮説検定の考え方を用い，有意水準 7% で考察せよ。ただし，1，2，3，4，5，6 の目が等確率 $\frac{1}{6}$ で出る立方体のさいころを 8 回投げたとき，3 以上の目がちょうど n 回出る確率 $P(n)$ を表 5.2 に示したため，必要ならばこれを用いてもよい。

表 5.2: 3 以上の目が n 回出る確率 $P(n)$

n	0	1	2	3	4
$P(n)$	$\frac{1}{6561}$	$\frac{16}{6561}$	$\frac{112}{6561}$	$\frac{448}{6561}$	$\frac{1120}{6561}$
n	5	6	7	8	
$P(n)$	$\frac{1792}{6561}$	$\frac{1792}{6561}$	$\frac{1024}{6561}$	$\frac{256}{6561}$	

5.2 解答

A　次のように仮説を定める。

仮説 H_1　このくじは，「当たり」が出やすい。

仮説 H_1 に反する仮説を次のように定める。

仮説 H_0　このくじは，「当たり」と「はずれ」が出る確率が変わらず $\dfrac{1}{2}$ である。

硬貨を 10 回投げて 8 回以上表が出る確率 p は，表 5.1 より

$$p = \frac{45}{1024} + \frac{10}{1024} + \frac{1}{1024} = \frac{7}{128}$$

である。すなわち，仮説 H_0 のもとでは，「当たり」が 8 回以上出る確率は $\dfrac{7}{128}$ である。$\dfrac{7}{128} = 0.054\cdots$ より $p > 0.05$ であるから，仮説 H_0 は有意水準 0.05 のもとで否定できず，仮説 H_1 が正しいとは判断できない。したがって，このくじは「当たり」が出やすいとは判断できない。

B　アンケートの選択肢は「快適である」，「快適でない」の 2 つであるから，「快適である」と回答した人が増えるということと，「快適でない」と回答した人が減るということは同義である。最初に，「快適である」と回答した人は $\dfrac{1}{3}$ であったため，「快適でない」と回答した人は $\dfrac{2}{3}$ である。2 回目に 8 人にアンケートを取ったとき，「快適でない」と回答した人は $8 - 6 = 2$ 人である。以上から，仮説を次のように定める。

仮説 H_1　「快適でない」と回答した人は減った。

仮説 H_1 に反する仮説を次のように定める。

仮説 H_0　「快適でない」と回答した人は，依然として全体の $\dfrac{2}{3}$ のままである。

題のさいころが 3 以上（すなわち 3，4，5，6）の目のいずれかが出る確率は，$\dfrac{1}{6} \cdot 4 = \dfrac{2}{3}$ であることに注意すると，確率 $\dfrac{2}{3}$ の事象が起きる回数が 2 回以下となる確率 p は，表 5.2 より

$$p = \frac{1}{6561} + \frac{16}{6561} + \frac{112}{6561} = \frac{43}{2187}$$

である。すなわち，仮説 H_0 のもとでは，「快適でない」と回答する人数が 2 人以下である確率は $\dfrac{43}{2187}$ である。$\dfrac{43}{2187} = 0.019\cdots$ より，$p < 0.07$ であるから，仮説 H_0 は有意水準 7% のもとで否定でき，仮説 H_1 は正しいと判断してよい。したがって，「快適でない」と回答した人は減ったといえ，これは「快適である」と回答した人が増えたといえることと同義である。

5.3 解説

A　基本的な仮説検定です。考察すべき仮説を対立仮説とし，それに反する帰無仮説を立て，それを棄却できるか考察します。

B　元のアンケートで「快適である」と回答した人が全体の $\dfrac{1}{3}$ であったため，$\dfrac{1}{3}$ からどう変化したかを確認したいですが，与えられた表は $\dfrac{2}{3}$ が n 回起きる確率であるため，元のアンケート

で $\dfrac{2}{3}$ の方を考えるか，表から $\dfrac{1}{3}$ が n 回起きる確率を考える必要があります。表から $\dfrac{1}{3}$ が起きる確率を考えた解答が次のようになります。

- 「快適である」と回答した人が増えた，という仮説 H_1 と，それに反する仮説 H_0 として，「快適である」と回答した人は全体の $\dfrac{1}{3}$ のままである，というものを考える。さいころで 2 以下（すなわち 1，2）の目のいずれかが出る確率は $\dfrac{1}{6}+\dfrac{1}{6}=\dfrac{1}{3}$ である。また，さいころを 8 回投げるとき，さいころで 3 以上の目が n 回出る確率と，2 以下の目が $8-n$ 回出る確率は等しいため，表 5.2 より，さいころを 8 回投げて 2 以下の目が 6 回以上出る確率は，3 以上の目が 2 回以下出る確率
$$p=\frac{1}{6561}+\frac{16}{6561}+\frac{112}{6561}=\frac{43}{2187}$$
で 0.07 より小さい。よって仮説 H_0 は否定でき，「快適である」と回答した人は増えたといえる。

5.4　採点

【第 5 問　16 点】

A　8 点

- 正しく仮説を定められているか。
- 確率の計算を正しく行えているか。

B　8 点

- 正しく仮説を定められているか。
- 確率の計算を正しく行えているか。