1

解答 (1) 2分　(2) 57分　(3) 40 %

解説

(1) 50以上52未満などだから　2分

(2) 56以上58未満だから　57分

(3) 56分以上かかった日数は12日だから

$12 \div 30 \times 100 = 40$ (%)

2

解答 (1) 平均値9，最頻値11，中央値10　(2) 平均値8.5，最頻値12，中央値8.5

解説

(1) このデータの平均値は

$$\frac{1}{9}(3+5+9+9+10+11+11+11+12)=\frac{1}{9}\times 81=9$$

このデータの最頻値は　11

このデータの中央値は　10

(2) このデータを小さい順に並べると

2，3，4，6，7，8，9，12，12，12，13，14

このデータの平均値は

$$\frac{1}{12}(2+3+4+6+7+8+9+12+12+12+13+14)=\frac{1}{12}\times 102=8.5$$

このデータの最頻値は　12

このデータの中央値は　$\frac{8+9}{2}=8.5$

3

解答 (1) $a=542$　(2) 26通り

解説

(1) 平均値が535円であるから

$525+550+498+560+550+555+500+a=535\times 8$

よって　$3738+a=4280$　ゆえに　$a=542$ (円)

(2) 店舗数は8であるから，安い方から4番目と5番目の価格の平均値が中央値になる。

$a$ 以外の価格を安い方から順に並べると

498，500，525，550，550，555，560

[1] $a \leqq 525$ のとき

4番目の価格は525円，5番目の価格は550円であるから，中央値は

$$\frac{1}{2}(525+550)=537.5 \text{ (円)}$$

[2] $a \geqq 550$ のとき

4番目，5番目の価格はともに550円であるから，中央値は　550円

[3] $526 \leqq a \leqq 549$ のとき

4番目の価格は $a$ 円，5番目の価格は550円であり，中央値 $\frac{1}{2}(a+550)$ 円は $a$ の値によってすべて異なる。

$a$ の値は $549-526+1=24$ (通り) であるから，中央値も24通りある。

また，これらの中央値はどれも[1]，[2]の中央値とは異なる。

以上から，中央値として　$1+1+24=26$ (通り)　の値が考えられる。

4

解答 (1) 16，25.5，35.5　(2) 9.5，21，34.5　(3) 20，27.5，44

(4) 18，33，48

解説

(1) 第2四分位数は　$Q_2=\frac{22+29}{2}=25.5$，第1四分位数は　$Q_1=\frac{15+17}{2}=16$，

第3四分位数は　$Q_3=\frac{31+40}{2}=35.5$

(2) 第2四分位数は　$Q_2=21$，第1四分位数は　$Q_1=\frac{7+12}{2}=9.5$，

第3四分位数は　$Q_3=\frac{33+36}{2}=34.5$

(3) 第2四分位数は　$Q_2=\frac{24+31}{2}=27.5$，第1四分位数は　$Q_1=20$，

第3四分位数は　$Q_3=44$

(4) 第2四分位数は　$Q_2=33$，第1四分位数は　$Q_1=18$，

第3四分位数は　$Q_3=48$

5

解答 ②

解説

大きさの順に並べると　12，15，17，21，23，26，29，31，35

このデータの最小値，第1四分位数，中央値，第3四分位数，最大値は，順に

12，$\frac{15+17}{2}=16$，23，$\frac{29+31}{2}=30$，35

これらの値をとっている箱ひげ図は　②

6

解答 ①，②

解説

① 範囲，四分位範囲とも，商品Aの方が商品Bより大きいから，①は正しい。

② 商品Aのデータの中央値は15個より大きいから，販売数が15個以上の日が半数以上，すなわち15日以上あることがわかる。よって，②は正しい。

③ 商品Aのデータの最小値は5個，商品Bのデータの最小値は10個である。よって，商品Aは販売数が10個未満の日があるが，商品Bはないから，③は正しくない。

以上から，正しいものは　①，②

## 第１講　例題演習

1

解答 (1) 20分　(2) 70分　(3) 62.5 %

解説

(1) 0以上20未満などだから　20分

(2) 60 ～ 80だから　70分

(3) 40分以上かかった人数は25人だから　$25\div 40\times 100=62.5$ (%)

2

解答 (1) 平均値5.3，最頻値6，中央値5.5

(2) 平均値7，最頻値8，中央値8

解説

(1) このデータの平均値は

$$\frac{1}{10}(1+1+3+5+5+6+6+6+10+10)=\frac{53}{10}=5.3$$

このデータの最頻値は　6

このデータの中央値は　$\frac{5+6}{2}=5.5$

(2) このデータを小さい順に並べると

2，2，3，5，5，5，8，8，8，8，12，12，13

このデータの平均値は

$$\frac{1}{13}(2+2+3+5+5+5+8+8+8+8+12+12+13)=\frac{91}{13}=7$$

このデータの最頻値は　8

このデータの中央値は　8

3

解答 (1) 7通り　(2) $a=75$

解説

(1) 人数は10人であるから，小さい方から5番目と6番目の得点の平均値が中央値となる。

$a$ 以外の得点を小さい順に並べると

38，41，45，60，62，66，67，74，82

$a \leqq 60$ のとき，5番目の得点は60点，6番目の得点は62点であるから，中央値は

$\frac{60+62}{2}=61$ (点)

$a \geqq 66$ のとき，5番目の得点は62点，6番目の得点は66点であるから，中央値は

$\frac{62+66}{2}=64$ (点)

$61 \leqq a \leqq 65$ のとき，5番目，6番目の得点は $a$ 点か62点のいずれかであるから，中央値は $a$ の値によってすべて異なる。

ゆえに，中央値は，$2+(65-61+1)=7$ (通り) の値があり得る。

(2) 得点の平均値が61.0点であるから

$$\frac{1}{10}(38+41+45+60+62+66+67+74+82+a)=61.0$$

これを解いて　$a=75$

4

解答 第1四分位数，第2四分位数，第3四分位数 の順に

(1) 13，41，53　(2) 20，36，52　(3) 30，52，77　(4) 18，35.5，51.5

解説

第１四分位数 $Q_1$，第２四分位数 $Q_2$，第３四分位数 $Q_3$ は，それぞれ次のようになる。

(1)　$Q_2=41$，　$Q_1=\dfrac{12+14}{2}=13$，　$Q_3=\dfrac{52+54}{2}=53$

(2)　$Q_2=\dfrac{34+38}{2}=36$，　$Q_1=20$，　$Q_3=52$

(3)　$Q_2=52$，　$Q_1=30$，　$Q_3=77$

(4)　$Q_2=\dfrac{32+39}{2}=35.5$，　$Q_1=\dfrac{16+20}{2}=18$，　$Q_3=\dfrac{49+54}{2}=51.5$

5

解答　③

解説

データを大きさの順に並べると

46，48，49，50，50，51，51，52，52，53，54，54，55，57，58

このデータの最小値，第１四分位数，中央値，第３四分位数，最大値は，順に

46，50，52，54，58

これらの値をとっている箱ひげ図は　③

6

解答　①，③

解説

①　範囲，四分位範囲とも，テスト A の方がテスト B より大きいから，① は正しい。

②　テスト A のデータの中央値は 60 点より小さいから，テスト A で 60 点以上の生徒は 200 人以下である。
よって，② は正しくない。

③　テスト B のデータの第３四分位数は 80 点より大きいから，テスト B で 80 点以上の生徒は 100 人以上いる。
よって，③ は正しい。

④　テスト A の最小値，第１四分位数はそれぞれ 20 点，40 点であるから，テスト A に 30 点台がいるかどうかはこの箱ひげ図からはわからない。よって，④ は正しいとはいえない。

以上から，正しいものは　①，③



1

解答　(1)　150 個　(2)　152 個　(3)　142 個以上 162 個未満

解説

(1)　度数が最も大きい階級は 140 個以上 160 個未満であるから，その階級値は　150 個
よって，このデータの最頻値は　150 個

(2)　階級値を用いたデータの平均値は

$\dfrac{1}{30}(110\times3+130\times5+150\times11+170\times8+190\times3)=152$（個）

(3)　データの平均値が最小となるのは，データの各値が階級内の最小の値となるときであるから

$\dfrac{1}{30}(100\times3+120\times5+140\times11+160\times8+180\times3)=142$（個）

また，階級の幅が 20 個であるから，データの平均値のとりうる値の範囲は　142 個以上 162 個未満

2

解答　(1)　(ア)　50　(イ)　59　(2)　(ウ)　61　(エ)　70

解説

(1)　「読み」の得点の中央値が最も小さい（または大きい）値をとるのは，得点の低い方から 5 番目と 6 番目の人の値が最も小さく（または大きく）なるときである。それらは，ともに 50 点以上 59 点以下の階級に属する。

よって，最も小さい値は　$\dfrac{50+50}{2}={}^{ア}50$（点）

最も大きい値は　$\dfrac{59+59}{2}={}^{イ}59$（点）

(2)　「書き取り」の得点の平均値が最も小さい値をとるのは，10 人の得点が階級内の最小の値となるときであるから，最も小さい値は

$\dfrac{1}{10}(50\times2+60\times5+70\times3)={}^{ウ}61$（点）

階級の幅が 9 点であるから，「書き取り」の得点の平均値の最も大きい値は

$61+9={}^{エ}70$（点）

3

解答　(1)　中央値 2.675 t，平均値 2.82 t
(2)　誤っている数値 1.61 t，正しい数値 2.81 t

解説

(1)　データを大きさの順に並べると

1.61，2.13，2.48，2.87，3.71，4.12

データの大きさが 6 であるから，中央値は小さい方から 3 番目と 4 番目の値の平均値である。

よって，中央値は　$\dfrac{1}{2}(2.48+2.87)=2.675$（t）

平均値は

$$\dfrac{1}{6}(1.61+4.12+3.71+2.87+2.48+2.13)=\dfrac{16.92}{6}=2.82\ (\mathrm{t})$$

(2)　正しいデータの平均値は (1) で求めたものより 0.2 t 大きいからどれかが 1.2 t 少ない。データの値から 1 つ選んで 1.2 t を加えた結果，中央値が 2.84 t になるものは，1.61 t のみである。

よって，誤っている数値は　1.61 t
正しい数値は　2.81 t

4

解答　A 君：四分位範囲 37.5 点，四分位偏差 18.75 点
B 君：四分位範囲 18.5 点，四分位偏差 9.25 点
A 君の方がデータの散らばりの度合いが大きいと考えられる

解説

データを小さい方から順に並べると

A 君　42，45，52，58，67，72，83，89，96

B 君　37，60，66，70，72，75，81，82，98

A 君のデータについて，

第１四分位数は　$Q_1=\dfrac{45+52}{2}=48.5$（点）

第３四分位数は　$Q_3=\dfrac{83+89}{2}=86$（点）

よって，四分位範囲は　$Q_3-Q_1=86-48.5=37.5$（点）

四分位偏差は　$\dfrac{Q_3-Q_1}{2}=\dfrac{37.5}{2}=18.75$（点）

B 君のデータについて，

第１四分位数は　$Q_1=\dfrac{60+66}{2}=63$（点）

第３四分位数は　$Q_3=\dfrac{81+82}{2}=81.5$（点）

よって，四分位範囲は　$Q_3-Q_1=81.5-63=18.5$（点）

四分位偏差は　$\dfrac{Q_3-Q_1}{2}=\dfrac{18.5}{2}=9.25$（点）

A 君の方が四分位範囲が大きいから，A 君の方がデータの散らばりの度合いが大きいと考えられる。

5

解答　(1)　C 店　(2)　A 店　(3)　30 日

解説

(1)　中央値が 350 人を超えているものであるから，C 店である。

(2)　第１四分位数が 250 人未満であるものであるから，A 店である。

(3)　B 店において，最小値は 200 人未満，第１四分位数は 200 人を超えているから，1 日の入店者数が 200 人を超えたのは，最も多くて 30 日あったと考えられる。
よって　30 日

## 1

解答 (1)　$x$ のデータについて　平均値 7 個，分散 8，標準偏差 2.8 個

$y$ のデータについて　平均値 7 個，分散 2，標準偏差 1.4 個

(2)　$x$ のデータの方が，平均値からの散らばりの度合いが大きい

解説

(1)　$x$，$y$ のデータの平均値をそれぞれ $\overline{x}$，$\overline{y}$ とすると

$\overline{x}=\dfrac{1}{5}(5+4+8+12+6)=7$（個），　$\overline{y}=\dfrac{1}{5}(6+9+8+5+7)=7$（個）

$x$，$y$ のデータの分散をそれぞれ $s_x{}^2$，$s_y{}^2$ とすると

$$s_x{}^2=\frac{1}{5}\{(5-7)^2+(4-7)^2+(8-7)^2+(12-7)^2+(6-7)^2\}=\frac{40}{5}=8$$

$$s_y{}^2=\frac{1}{5}\{(6-7)^2+(9-7)^2+(8-7)^2+(5-7)^2+(7-7)^2\}=\frac{10}{5}=2$$

よって，標準偏差は　$s_x=\sqrt{8}\fallingdotseq 2.8$（個），$s_y=\sqrt{2}\fallingdotseq 1.4$（個）

別解　分散の求め方

$$s_x{}^2=\frac{1}{5}(5^2+4^2+8^2+12^2+6^2)-7^2=8,\quad s_y{}^2=\frac{1}{5}(6^2+9^2+8^2+5^2+7^2)-7^2=2$$

(2)　(1) から　$s_x>s_y$

ゆえに，$x$ のデータの方が，平均値からの散らばりの度合いが大きいと考えられる。

## 2

解答 (1)　68 点　(2)　4880　(3)　分散 256，標準偏差 16 点

解説

(1)　$\overline{x}=\dfrac{1}{5}(50+70+90+80+50)$

$=\dfrac{1}{5}\times 340=68$（点）

(2)　$\overline{x^2}=\dfrac{1}{5}(50^2+70^2+90^2+80^2+50^2)$

$=\dfrac{1}{5}\times 24400=4880$

(3)　このデータの分散を $s^2$，標準偏差を $s$ とすると

$s^2=\overline{x^2}-(\overline{x})^2=4880-68^2=256$

$s=\sqrt{s^2}=\sqrt{256}=16$（点）

## 3

解答　平均値 8，分散 7

解説

15 個のデータの平均値は　$\dfrac{1}{15}(9\times 10+6\times 5)=\dfrac{120}{15}=8$

10 個の値の 2 乗の平均値を $a$ とすると

$a-9^2=3$　　よって　$a=84$

残りの 5 個の値の 2 乗の平均値を $b$ とすると

$b-6^2=9$　　よって　$b=45$

よって，15 個の値の 2 乗の和は

$a\times 10+b\times 5=84\times 10+45\times 5=1065$

したがって，15 個のデータの分散は　$\dfrac{1065}{15}-8^2=71-64=7$

## 4

解答 (1)　992 m　(2)　標準偏差 12 m，分散 144

解説

(1)　変量 $x$ のデータの各値と仮平均 $x_0=1000$ との差を表にすると，次のようになる。

| $x$ | 1008 | 992 | 980 | 1008 | 984 | 980 | 計 |
|---|---|---|---|---|---|---|---|
| $x-x_0$ | 8 | $-8$ | $-20$ | 8 | $-16$ | $-20$ | $y$ |

$x-x_0$ の合計を $y$ とすると

$y=8+(-8)+(-20)+8+(-16)+(-20)=-48$（m）

よって　$\overline{x}=1000+\dfrac{-48}{6}=992$（m）

(2)　$u=\dfrac{x-1000}{4}$ とおくと，$u$，$u^2$ の値は次のようになる。

| $x$ | 1008 | 992 | 980 | 1008 | 984 | 980 | 計 |
|---|---|---|---|---|---|---|---|
| $u$ | 2 | $-2$ | $-5$ | 2 | $-4$ | $-5$ | $-12$ |
| $u^2$ | 4 | 4 | 25 | 4 | 16 | 25 | 78 |

$u$ のデータの分散は　$\overline{u^2}-(\overline{u})^2=\dfrac{78}{6}-\left(\dfrac{-12}{6}\right)^2=9$

よって，$u$ のデータの標準偏差は　$\sqrt{9}=3$（m）

ゆえに，$x$ のデータの標準偏差は　$3\times 4=12$（m）

したがって，$x$ のデータの分散は　$12^2=144$

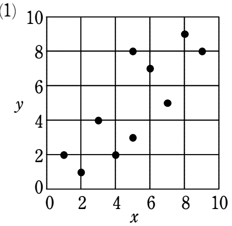
## 5

解答 (1)　［図］，正の相関がある

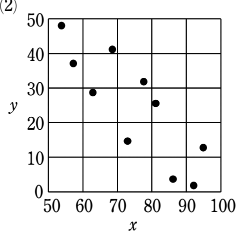
(2)　［図］，負の相関がある

(3)　［図］，相関がない

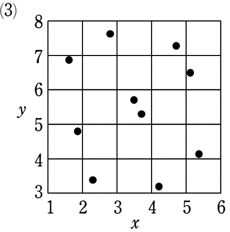
(1)



(2)



(3)



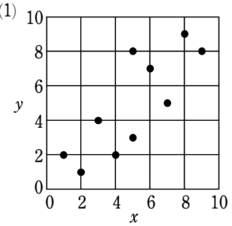
解説

(1)　［図］，正の相関がある

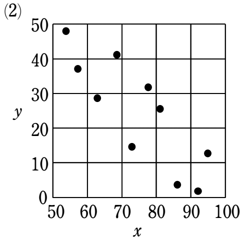
(2)　［図］，負の相関がある

(3)　［図］，相関がない

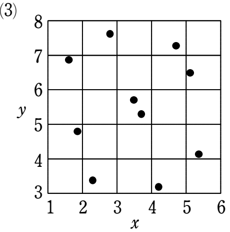
(1)



(2)



(3)



## 6

解答　①　0.03　②　$-0.72$　③　0.88

解説

データの散布図から

①は散らばっているので相関関係はない。

②は一方が増えると他方は減る傾向にあるので，負の相関関係がある。

③は一方が増えると他方も増える傾向にあるので，正の相関関係がある。

よって，各相関係数は　①　0.03，　②　$-0.72$，　③　0.88

## 7

解答　相関係数は　$-0.87$

強い負の相関がある。

解説

テスト A の得点を $x$，テスト B の得点を $y$ として，次のような表を作る。

$\overline{x}=\dfrac{60}{10}=6$

$\overline{y}=\dfrac{50}{10}=5$

| | $x$ | $y$ | $x-\overline{x}$ | $y-\overline{y}$ | $(x-\overline{x})(y-\overline{y})$ | $(x-\overline{x})^2$ | $(y-\overline{y})^2$ |
|---|---|---|---|---|---|---|---|
| ① | 5 | 6 | $-1$ | 1 | $-1$ | 1 | 1 |
| ② | 3 | 9 | $-3$ | 4 | $-12$ | 9 | 16 |
| ③ | 5 | 6 | $-1$ | 1 | $-1$ | 1 | 1 |
| ④ | 5 | 5 | $-1$ | 0 | 0 | 1 | 0 |
| ⑤ | 3 | 7 | $-3$ | 2 | $-6$ | 9 | 4 |
| ⑥ | 8 | 2 | 2 | $-3$ | $-6$ | 4 | 9 |
| ⑦ | 10 | 3 | 4 | $-2$ | $-8$ | 16 | 4 |
| ⑧ | 5 | 8 | $-1$ | 3 | $-3$ | 1 | 9 |
| ⑨ | 8 | 3 | 2 | $-2$ | $-4$ | 4 | 4 |
| ⑩ | 8 | 1 | 2 | $-4$ | $-8$ | 4 | 16 |
| 計 | 60 | 50 | | | $-49$ | 50 | 64 |

上の表から　$r=\dfrac{-49}{\sqrt{50\times64}}=-\dfrac{49\sqrt{2}}{80}=-0.866\cdots\fallingdotseq-0.87$

よって，テスト A とテスト B の得点には強い負の相関がある。

$\boxed{8}$

解答　製品の品質が向上したと判断してよい

解説

仮説 $H_1$ :品質が向上した

と判断してよいかを考察するために，次の仮説を立てる。

仮説 $H_0$ :品質が向上したとはいえず，「品質が向上した」と回答する場合と、そうでない場合がまったくの偶然で起こる

コイン投げの実験結果から，コインを 20 枚投げて表が 15 枚以上出る場合の相対度数は

$\dfrac{3+0+1}{200}=\dfrac{4}{200}=0.02$

すなわち，仮説 $H_0$ のもとでは，15 人以上が「品質が向上した」と回答する確率は 0.02 程度であると考えられる。

これは，0.05 より小さいから、仮説 $H_0$ は正しくなかったと考えられ，仮説 $H_1$ は正しいと判断してよい。

したがって，製品の品質が向上したと判断してよい。

$\boxed{9}$

解答　A は B より強いとは判断できない

解説

仮説 $H_1$ :A は B より強い

と判断してよいかを考察するために，次の仮説を立てる。

仮説 $H_0$ :A と B の強さは同等である

仮説 $H_0$ のもとで，ゲームを 9 回行って，A が 7 回以上勝つ確率は

$${}_9C_9\left(\frac{1}{2}\right)^9\left(\frac{1}{2}\right)^0+{}_9C_8\left(\frac{1}{2}\right)^8\left(\frac{1}{2}\right)^1+{}_9C_7\left(\frac{1}{2}\right)^7\left(\frac{1}{2}\right)^2=\frac{1}{2^9}(1+9+36)=\frac{46}{512}=0.089\cdots\cdots$$

これは 0.05 より大きいから，仮説 $H_0$ は否定できず，仮説 $H_1$ が正しいとは判断できない。

したがって，A は B より強いと判断できない。



$\boxed{1}$

解答　(1)　$x$ : 平均値 5 点，分散 2，標準偏差 1.4 点

$y$ : 平均値 4 点，分散 8，標準偏差 2.8 点

(2)　$y$ の方がデータの平均値からの散らばりの度合いが大きいと考えられる

解説

(1)　$x$ の平均値を $\overline{x}$，分散を $s_x{}^2$，標準偏差を $s_x$ とする。

$$\overline{x}=\frac{1}{10}(5+6+5+5+8+6+3+5+3+4)=\frac{1}{10}\times50=5\ (点)$$

よって　$$s_x{}^2=\frac{1}{10}\{(5-5)^2+(6-5)^2+(5-5)^2+(5-5)^2+(8-5)^2+(6-5)^2+(3-5)^2+(5-5)^2+(3-5)^2+(4-5)^2\}=\frac{1}{10}\times20=2$$

ゆえに　$s_x=\sqrt{2}=1.41\cdots\fallingdotseq1.4$ (点)

$y$ の平均値を $\overline{y}$，分散を $s_y{}^2$，標準偏差を $s_y$ とする。

$$\overline{y}=\frac{1}{10}(4+0+6+3+4+2+10+3+7+1)=\frac{1}{10}\times40=4\ (点)$$

よって　$$s_y{}^2=\frac{1}{10}\{(4-4)^2+(0-4)^2+(6-4)^2+(3-4)^2+(4-4)^2+(2-4)^2+(10-4)^2+(3-4)^2+(7-4)^2+(1-4)^2\}=\frac{1}{10}\times80=8$$

ゆえに　$s_y=\sqrt{8}=2\sqrt{2}=2.82\cdots\fallingdotseq2.8$ (点)

別解　[分散の求め方]

$$\overline{x}=\frac{1}{10}(5+6+5+5+8+6+3+5+3+4)=\frac{1}{10}\times50=5$$

$$\overline{x^2}=\frac{1}{10}(5^2+6^2+5^2+5^2+8^2+6^2+3^2+5^2+3^2+4^2)=\frac{1}{10}\times270=27$$

よって　$s_x{}^2=\overline{x^2}-(\overline{x})^2=27-5^2=2$

$$\overline{y}=\frac{1}{10}(4+0+6+3+4+2+10+3+7+1)=\frac{1}{10}\times40=4$$

$$\overline{y^2}=\frac{1}{10}(4^2+0^2+6^2+3^2+4^2+2^2+10^2+3^2+7^2+1^2)=\frac{1}{10}\times240=24$$

よって　$s_y{}^2=\overline{y^2}-(\overline{y})^2=24-4^2=8$

(2)　$s_x<s_y$ であるから，$y$ の方がデータの平均値からの散らばりの度合いが大きいと考えられる。

$\boxed{2}$

解答　(1)　68 点　(2)　4880　(3)　256，16 点

解説

(1)　$\overline{x}=\dfrac{1}{5}(50+70+90+80+50)=\dfrac{1}{5}\times340=68$ (点)

(2)　$\overline{x^2}=\dfrac{1}{5}(50^2+70^2+90^2+80^2+50^2)=\dfrac{1}{5}\times24400=4880$

(3)　$s^2=\overline{x^2}-(\overline{x})^2=4880-68^2=256$

$s=\sqrt{s^2}=\sqrt{256}=16$ (点)

$\boxed{3}$

解答　(1)　6　(2)　13

解説

(1)　$\dfrac{3\times8+8\times12}{20}=6$

(2)　8 個の値の 2 乗の平均値を $a$ とすると　$a-3^2=4$　よって　$a=13$

残りの 12 個の値の 2 乗の平均値を $b$ とすると　$b-8^2=9$　よって　$b=73$

よって，20 個の値の 2 乗の和は　$a\times8+b\times12=13\times8+73\times12=980$

ゆえに，20 個の値の分散は　$\dfrac{980}{20}-6^2=49-36=13$

$\boxed{4}$

解答　(1)　(ア)　$\overline{x}=858$ (点)　(イ)　標準偏差 21 点，分散 441

(2)　$\overline{z}=0$，$s_z=1$

解説

(1)　(ア)　変量 $x$ のデータの各値と仮平均 $x_0=830$ との差を表にすると，右のようになる。

| $x$ | 844 | 893 | 872 | 844 | 830 | 865 | 計 |
|---|---|---|---|---|---|---|---|
| $x-x_0$ | 14 | 63 | 42 | 14 | 0 | 35 | $y$ |

$x-x_0$ の合計 $y$ の値は

$y=14+63+42+14+0+35=168$ (点)

よって　$\overline{x}=830+\dfrac{168}{6}=858$ (点)

(イ)　$u=\dfrac{x-830}{7}$ とおくと，$u$，$u^2$ の値は右のようになる。

| $x$ | 844 | 893 | 872 | 844 | 830 | 865 | 計 |
|---|---|---|---|---|---|---|---|
| $u$ | 2 | 9 | 6 | 2 | 0 | 5 | 24 |
| $u^2$ | 4 | 81 | 36 | 4 | 0 | 25 | 150 |

$u$ のデータの分散は

$$\overline{u^2}-(\overline{u})^2=\frac{150}{6}-\left(\frac{24}{6}\right)^2=9$$

であるから，$u$ のデータの標準偏差は　$\sqrt{9}=3$

よって，$x$ のデータの標準偏差は　$3\times7=21$ (点)，分散は　$21^2=441$

(2)　$z=\dfrac{1}{s_x}x-\dfrac{\overline{x}}{s_x}$ から　$\overline{z}=\dfrac{\overline{x}-\overline{x}}{s_x}=0$，$s_z=\left|\dfrac{1}{s_x}\right|s_x=1$

$\boxed{5}$

解答　①　$-0.04$　②　$-0.71$　③　0.87

解説

① の散布図において，$x$ と $y$ の間に相関関係がない。

よって，相関係数は 0 に近い値をとる。

② の散布図において，$x$ と $y$ の間に負の相関がある。

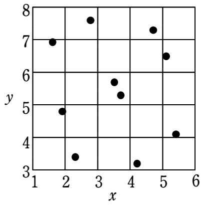
③ の散布図において，$x$ と $y$ の間に正の相関がある。

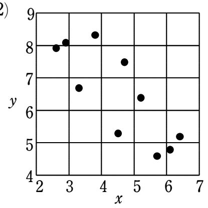
したがって　①　$-0.04$　②　$-0.71$　③　0.87

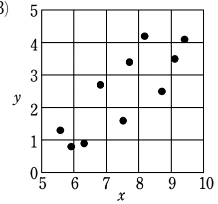
$\boxed{6}$

解答　(1)　[図]，相関関係がない　(2)　[図]，負の相関関係がある

(3)　[図]，正の相関関係がある

(1)



(2)


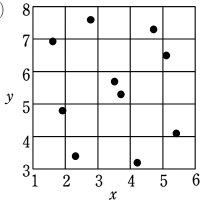

(3)


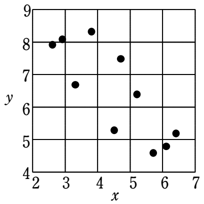

解説

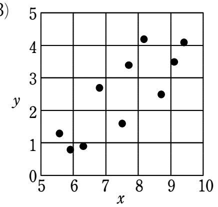
(1)〔図〕，相関関係がない

(2)〔図〕，負の相関関係がある

(3)〔図〕，正の相関関係がある

(1)



(2)



(3)



7

解答　0.88

解説

$x$，$y$ の平均をそれぞれ $\overline{x}$，$\overline{y}$ とすると

$$\overline{x}=\frac{1}{10}(43+41+43+38+39+42+42+39+41+42)=41$$

$$\overline{y}=\frac{1}{10}(49+42+44+36+40+44+45+42+42+46)=43$$

よって，次の表が得られる。

| 番号 | $x-\overline{x}$ | $y-\overline{y}$ | $(x-\overline{x})^2$ | $(y-\overline{y})^2$ | $(x-\overline{x})(y-\overline{y})$ |
|---|---|---|---|---|---|
| 1 | 2 | 6 | 4 | 36 | 12 |
| 2 | 0 | −1 | 0 | 1 | 0 |
| 3 | 2 | 1 | 4 | 1 | 2 |
| 4 | −3 | −7 | 9 | 49 | 21 |
| 5 | −2 | −3 | 4 | 9 | 6 |
| 6 | 1 | 1 | 1 | 1 | 1 |
| 7 | 1 | 2 | 1 | 4 | 2 |
| 8 | −2 | −1 | 4 | 1 | 2 |
| 9 | 0 | −1 | 0 | 1 | 0 |
| 10 | 1 | 3 | 1 | 9 | 3 |
| 計 | | | 28 | 112 | 49 |

ゆえに　　$r=\dfrac{49}{\sqrt{28\times112}}=\dfrac{49}{\sqrt{2^2\cdot7\times4^2\cdot7}}=\dfrac{49}{2\times4\times7}=0.875\fallingdotseq0.88$

8

解答　(1)　支持率は上昇したと判断してよい　　(2)　支持率は上昇したと判断できない

解説

仮説 $H_1$：支持率は上昇した

と判断してよいかを考察するために、次の仮説を立てる。

仮説 $H_0$：支持率は上昇したとはいえず，「支持する」と回答する確率は $\dfrac{2}{3}$ である

さいころを 1 個投げて 1 から 4 までのいずれかの目が出る確率は $\dfrac{2}{3}$ である。さいころ投げの実験結果から，さいころを 30 個投げて 1 から 4 までのいずれかの目が 25 個以上出る場合の相対度数は $\dfrac{4+2+1}{200}=\dfrac{7}{200}=0.035$

すなわち，仮説 $H_0$ のもとでは，25 人以上が「支持する」と回答する確率は 0.035 程度であると考えられる。

(1)　0.035 は基準となる確率0.05より小さい。よって，仮説 $H_0$ は正しくなかったと考えられ，仮説 $H_1$ は正しいと判断してよい。
したがって，支持率は上昇したと判断してよい。

(2)　0.035 は基準となる確率 0.01 より大きい。よって，仮説 $H_0$ は否定できず，仮説 $H_1$ が正しいとは判断できない。したがって，支持率が上昇したとは判断できない。

9

解答　このコインは裏が出やすいと判断してよい

解説

仮説 $H_1$ :このコインは裏が出やすい

と判断してよいかを考察するために，次の仮説を立てる。

仮説 $H_0$ :このコインは公正である

仮説 $H_0$ のもとで，コインを 8 回投げて，裏が 7 回以上出る確率は

$${}_8C_8\left(\frac{1}{2}\right)^8\left(\frac{1}{2}\right)^0+{}_8C_7\left(\frac{1}{2}\right)^7\left(\frac{1}{2}\right)^1=\frac{1}{2^8}(1+8)=\frac{9}{256}=0.035\cdots\cdots$$

これは 0.05 より小さいから，仮説 $H_0$ は正しくなかったと考えられ，仮説 $H_1$ は正しいと判断してよい。

したがって，このコインは裏が出やすいと判断してよい。

$\boxed{1}$

解答　(1)　$\overline{x}=18$　(2)　$a=23$，$b=20$，$c=25$，$d=36$

(3)　分散 18，標準偏差 4.2 個

解説

(1)　A のデータに着目すると

$x=22$，$(x-\overline{x})^2=16$ から　$(22-\overline{x})^2=16$

よって　$22-\overline{x}=\pm4$

ゆえに　$\overline{x}=18,\ 26$　……①

B のデータに着目すると

$x=21$，$(x-\overline{x})^2=9$ から　$(21-\overline{x})^2=9$

よって　$21-\overline{x}=\pm3$

ゆえに　$\overline{x}=18,\ 24$　……②

①，② から　$\overline{x}=18$ (個)

別解　① と $x<25$ から　$\overline{x}=18$ (個)

(2)　G のデータに着目すると　$d=(24-18)^2=36$

$x$ のデータの総和は　$18\times10=180$

よって　$22+21+16+a+13+15+24+b+12+14=180$

ゆえに　$a=43-b$　……③

H のデータに着目すると

$(b-18)^2=4$ から　$b-18=\pm2$

よって　$b=16,\ 20$

$b=16$ のとき，③ から $a=27$ となり，$x<25$ に適さない。

$b=20$ のとき，③ から $a=23$ となり，$x<25$ に適する。

よって　$a=23$，$b=20$

ゆえに　$c=(23-18)^2=25$

したがって　$a=23$，$b=20$，$c=25$，$d=36$

(3)　偏差の 2 乗の和は

$16+9+4+25+25+9+36+4+36+16=180$

よって，分散は　$\dfrac{1}{10}\times180=18$

ゆえに，標準偏差は　$\sqrt{18}=3\sqrt{2}\fallingdotseq4.2$ (個)

$\boxed{2}$

解答　平均値 18 g，分散 11

解説

ボールの重さの平均値は

$$\frac{1}{30}(20\times15+18\times5+15\times10)=18\ (\mathrm{g})$$

A，B，C の組のボールの重さの 2 乗の平均値をそれぞれ $a$，$b$，$c$ とすると

$a-20^2=1^2$　よって　$a=401$

$b-18^2=5^2$　よって　$b=349$

$c-15^2=2^2$　よって　$c=229$

よって，30 個のボールの重さの 2 乗の和は

$$a\times15+b\times5+c\times10=401\times15+349\times5+229\times10=10050$$

ゆえに，30 個のボールの重さの分散は

$$\frac{10050}{30}-18^2=335-324=11$$

$\boxed{3}$

解答　$-0.6$

解説

相関係数を $r$ とすると　$r=\dfrac{-270}{\sqrt{250\times810}}=-\dfrac{270}{\sqrt{5^2\times9^2\times10^2}}=-\dfrac{270}{5\times9\times10}=-0.6$

$\boxed{4}$

解答　(1)　A 選手の評価は高まったとは判断できない

(2)　A 選手の評価は高まったとして判断してよい

解説

仮説 $H_1$ : A 選手の評価は高まった

と判断してよいかを考察するために，次の仮説を立てる。

仮説 $H_0$ : A 選手の評価は変わらない

すなわち，ヒットを打つ確率は $\dfrac{1}{4}$ である。

(1)　仮説 $H_0$ のもとで，5 打席中，ヒットを打つ打席が 3 打席以上である確率は

$${}_5C_5\left(\frac{1}{4}\right)^5\left(\frac{3}{4}\right)^0+{}_5C_4\left(\frac{1}{4}\right)^4\left(\frac{3}{4}\right)^1+{}_5C_3\left(\frac{1}{4}\right)^3\left(\frac{3}{4}\right)^2=\frac{1+15+90}{4^5}=\frac{106}{1024}=0.10\cdots\cdots$$

これは 0.05 より大きいから，仮説 $H_0$ は否定できず，仮説 $H_1$ が正しいとは判断できない。

したがって，A 選手の評価は高まったとは判断できない。

(2)　仮説 $H_0$ のもとで，10 打席中，ヒットを打つ打席が 5 打席以下である確率は，与えられた表から

$0.056+0.188+0.282+0.250+0.146+0.058=0.98$

よって，10 打席中，ヒットを打つ確率が 6 打席以上である確率は

$1-0.98=0.02$

これは 0.05 より小さいから，仮説 $H_0$ は正しくなかったと考えられ，仮説 $H_1$ は正しいと判断してよい。

したがって，A 選手の評価は高まったと判断してよい。



$\boxed{1}$

解答　(1)　(ア)　6　(イ)　4　(ウ)　6

(2)　(エ)　14　(オ)　2　(カ)　6　(キ)　8　(3)　(ク)　0.2

解説

(1)　10 人の国語の得点の平均値 $A$ は

$$A=\frac{1}{10}(6+5+6+4+4+9+6+3+9+8)={}^{\text{ア}}6\ (\text{点})$$

各生徒の国語の得点の偏差を並べると

$0,\ -1,\ 0,\ -2,\ -2,\ 3,\ 0,\ -3,\ 3,\ 2$

よって，求める分散 $B$ は

$$B=\frac{1}{10}\{0^2+(-1)^2+0^2+(-2)^2+(-2)^2+3^2+0^2+(-3)^2+3^2+2^2\}={}^{\text{イ}}4$$

国語の得点を小さいものから順に並べると

3，4，4，5，6，6，6，8，9，9

よって，中央値は　$\dfrac{6+6}{2}={}^{\text{ウ}}6$ (点)

(2)　10 人の英語の得点の平均値が 7 点であるから

$$\frac{1}{10}(6+C+8+7+D+7+5+7+8+8)=7$$

よって　$C+D={}^{\text{エ}}14$

各生徒の英語の得点の偏差を並べると

$-1,\ C-7,\ 1,\ 0,\ D-7,\ 0,\ -2,\ 0,\ 1,\ 1$

分散が 1 であるから

$$\frac{1}{10}\{(-1)^2+(C-7)^2+1^2+0^2+(D-7)^2+0^2+(-2)^2+0^2+1^2+1^2\}=1$$

よって　$(C-7)^2+(D-7)^2={}^{\text{オ}}2$

$C-7$，$D-7$ は整数であるから　$(C-7)^2=1$，$(D-7)^2=1$

$C<D$ であるから　$C-7=-1$，$D-7=1$

よって　$C={}^{\text{カ}}6$ (点)，$D={}^{\text{キ}}8$ (点)

(3)　国語と英語の共分散は

$$\frac{1}{10}\{0\cdot(-1)+(-1)\cdot(-1)+0\cdot1+(-2)\cdot0+(-2)\cdot1+3\cdot0+0\cdot(-2)+(-3)\cdot0+3\cdot1+2\cdot1\}=0.4$$

よって，相関係数は　$\dfrac{0.4}{\sqrt{4}\sqrt{1}}={}^{\text{ク}}0.2$

$\boxed{2}$

解答　(ア)　70　(イ)　0.81

解説

データの番号 1 の $x-\overline{x}=-1$ から　$\overline{x}=6$　……①

よって，データの番号 10 の $x$ の値は　$10\overline{x}-(5+7+4+5+8+3+8+7+6)=7$

データの番号 1 の $y-\overline{y}=-1$ から　$\overline{y}=7$

よって　$a=10\overline{y}={}^{\text{ア}}70$

変量 $x$ の標準偏差を $s_x$，変量 $y$ の標準偏差を $s_y$，$x$，$y$ の共分散を $s_{xy}$ とすると，表から

$$s_y=\sqrt{\frac{1}{10}\cdot46}=\sqrt{\frac{23}{5}}$$

$$s_{xy}=\frac{1}{10}\cdot 28=\frac{14}{5}$$

表と①から，データの番号3から10までの $(x-\overline{x})^2$ の値は順に

$(-2)^2=4,\quad (-1)^2=1,\quad (8-6)^2=4,\quad (3-6)^2=9,$

$(8-6)^2=4,\quad (7-6)^2=1,\quad (6-6)^2=0,\quad (7-6)^2=1$

よって　$s_x=\sqrt{\frac{1}{10}(1+1+4+1+4+9+4+1+0+1)}=\sqrt{\frac{26}{10}}=\sqrt{\frac{13}{5}}$

以上から，相関係数 $r$ は

$$r=\frac{s_{xy}}{s_x s_y}=\frac{\frac{14}{5}}{\sqrt{\frac{13}{5}}\cdot\sqrt{\frac{23}{5}}}=\frac{14}{\sqrt{299}}=\frac{14\sqrt{299}}{299}=\frac{14\cdot 17.3}{299}=0.810\cdots\cdots$$

したがって，小数第3位を四捨五入して　$r=^{イ}0.81$

参考　この問題では，相関係数 $r$ を計算するときに，分母を有理化せず

$$r=\frac{14}{\sqrt{299}}=\frac{14}{17.3}=0.809\cdots\cdots$$

と計算しても答えは変わらない。

3

解答　(1)（ア）20　(2)（イ）90
(3)（ウ）22　（エ）66　（オ）2　（カ）157

解説

(1)　求める標準偏差を $s_1$ とすると　$62=50+\frac{10(84-60)}{s_1}$

すなわち　$s_1=^{ア}20$

(2)　求める条件は，(1)の結果を用いると　$65\leqq 50+\frac{10(x-60)}{20}$

整理すると　$x\geqq^{イ}90$

(3)　150人の受験者の得点の標準偏差を $s$ とすると　$60=50+\frac{10(84-62)}{s}$

すなわち　$s=^{ウ}22$

新たに受験した50人の受験者の得点の平均値を $\overline{x}$ とすると

$60\cdot 100+50\cdot\overline{x}=62\cdot 150$

ゆえに　$\overline{x}=^{エ}66$

ここで，最初に受験した100人の受験者の得点の平 均値を $\overline{y}$ とすると

$s_1{}^2=\overline{y^2}-(\overline{y})^2$

すなわち　$20^2=\overline{y^2}-60^2$　　ゆえに　$\overline{y^2}=4000$

100人の試験の得点の2乗の和を $A$ とすると　$\frac{A}{100}=4000$

ゆえに　$A=400000$　……①

また，150人の受験者の得点の平均値を $\overline{z}$ とすると　$s^2=\overline{z^2}-(\overline{z})^2$

すなわち　$22^2=\overline{z^2}-62^2$　　ゆえに　$\overline{z^2}=4328$

新たに受験した50人の受験者の得点の2乗の和を $B$ とすると　$\frac{A+B}{150}=4328$

①から　$B=249200$

よって，新たに受験した50人の受験者の標準偏差を $s_2$ とすると

$$s_2{}^2=\frac{B}{50}-(\overline{x})^2=\frac{249200}{50}-66^2=4984-4356=628$$

したがって　$s_2=\sqrt{628}=^{オ}2\sqrt{^{カ}157}$

4

解答　（ア）③　（イ）②

解説

(1)　散布図⓪には，X＝80，Y＝59の点がない。
散布図①には，X＝59，Y＝50の点がない。
散布図②には，X＝59，Y＝50の点がない。
よって，科目Xと科目Yの得点を散布図にしたものは　$^{ア}$③

(2)　科目Xのもとの得点の偏差を $d_{X_1}$，$d_{X_2}$，……，$d_{X_{30}}$，換算後の偏差を $d_{X_1}{}'$，$d_{X_2}{}'$，……，$d_{X_{30}}{}'$ とすると，$d_X{}'=\frac{1}{2}d_X$ が成り立つ。

科目Xのもとの標準偏差を $s_X$，換算後の標準偏差を $s_X{}'$ とすると

$s_X{}'=\frac{1}{2}s_X$　……①

科目Yのもとの得点の偏差を $d_{Y_1}$，$d_{Y_2}$，……，$d_{Y_{30}}$，換算後の偏差を $d_{Y_1}{}'$，$d_{Y_2}{}'$，……，$d_{Y_{30}}{}'$ とすると，$d_Y{}'=\frac{1}{2}d_Y$ が成り立つ。

科目Yのもとの標準偏差を $s_Y$，換算後の標準偏差を $s_Y{}'$ とすると，同様にして，

$s_Y{}'=\frac{1}{2}s_Y$ が成り立つ。

科目Xと科目Yのもとの得点の共分散を $s_{XY}$，換算後の共分散を $s_{XY}{}'$ とすると

$$s_{XY}{}'=\frac{1}{30}(d_{X_1}{}'\cdot d_{Y_1}{}'+d_{X_2}{}'\cdot d_{Y_2}{}'+\cdots\cdots+d_{X_{30}}{}'\cdot d_{Y_{30}}{}')$$

$$=\frac{1}{30}\left\{\left(\frac{1}{2}d_{X_1}\right)\cdot\left(\frac{1}{2}d_{Y_1}\right)+\left(\frac{1}{2}d_{X_2}\right)\cdot\left(\frac{1}{2}d_{Y_2}\right)+\cdots\cdots+\left(\frac{1}{2}d_{X_{30}}\right)\cdot\left(\frac{1}{2}d_{Y_{30}}\right)\right\}$$

$$=\frac{1}{4}\cdot\frac{1}{30}(d_{X_1}\cdot d_{Y_1}+d_{X_2}\cdot d_{Y_2}+\cdots\cdots+d_{X_{30}}\cdot d_{Y_{30}})\quad=\frac{1}{4}s_{XY}$$

以上から，換算後の標準偏差の値はもとの値の $\frac{1}{2}$ になり，共分散の値はもとの値の $\frac{1}{4}$ になる。　（$^{イ}$②）

参考　（等式①について）

$a$，$b$ を定数とする。変量 $x$ のデータから $y=ax+b$ によって新しい変量 $y$ のデータが得られるとき，$x$，$y$ のデータの平均値を $\overline{x}$，$\overline{y}$，分散を $s_x{}^2$，$s_y{}^2$，標準偏差を $s_x$，$s_y$ とすると，次のことが成り立つ。

$\overline{y}=a\overline{x}+b,\quad s_y{}^2=a^2s_x{}^2,\quad s_y=|a|s_x$

証明　変量 $x$ についてのデータを $x_1$，$x_2$，……，$x_n$ とする。

変量 $y$ のデータの平均値 $\overline{y}$ は

$$\overline{y}=\frac{1}{n}(y_1+y_2+\cdots\cdots+y_n)=\frac{1}{n}\{(ax_1+b)+(ax_2+b)+\cdots\cdots+(ax_n+b)\}$$

$$=\frac{1}{n}\{a(x_1+x_2+\cdots\cdots+x_n)+nb\}=a\cdot\frac{1}{n}(x_1+x_2+\cdots\cdots+x_n)+b$$

よって　$\overline{y}=a\overline{x}+b$

また，$y_k-\overline{y}=ax_k+b-(a\overline{x}+b)=a(x_k-\overline{x})$ であることから，変量 $y$ のデータの分散 $s_y{}^2$ は

$$s_y{}^2=\frac{1}{n}\{(y_1-\overline{y})^2+(y_2-\overline{y})^2+\cdots\cdots+(y_n-\overline{y})^2\}$$

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

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

よって　$s_y{}^2=a^2s_x{}^2$

これより，変量 $y$ のデータの標準偏差は　$s_y=|a|s_x$　終

□1

解答　(ア) ⓪　(イ) ④　(または(ア) ④　(イ) ⓪)　(ウ) ②

解説

(1) ⓪　ひげが最も右側まで伸びている箱ひげ図は，B 組のものである。
よって，この文は正しい。

① 箱の長さが最も長い箱ひげ図は，C 組のものである。
よって，この文は誤っている。

② 箱ひげ図全体の長さが最も長い箱ひげ図は，A 組のものである。
よって，この文は正しい。

③ 箱の左端から箱の中に引いてある線までの長さが最も短い箱ひげ図は，B 組のものである。
よって，この文は正しい。

④ A 組の得点の第 1 四分位数は 60 点，第 3 四分位数は 80 点である。
すなわち，A 組では，60 点未満の人数は 7 人以下，80 点以上の人数は 8 人以上いる。
よって，この文は誤っている。

⑤ A 組の得点の中央値は 70 点より高いから，A 組の 70 点以下の人数は多くても 15 人である。
一方，C 組の得点の中央値は 70 点であるから，C 組の 70 点以下の人数は少なくとも 15 人いる。
よって，この文は正しい。

したがって　ア⓪，イ④　(または　ア④，イ⓪)

(2) C 組の箱ひげ図から，C 組の第 1 四分位数は 60 点以上 70 点未満，第 3 四分位数は 80 点以上 90 点未満であることがわかる。
すなわち，C 組の得点を低い方から順に並べたとき，8 番目の生徒の得点は 60 点以上 70 点未満，23 番目の生徒の得点は 80 点以上 90 点未満である。
この両方を満たすヒストグラムは　ウ②

□2 [2015 センター本試]

解答　(ア) ④　(イ),(ウ),(エ),(オ) ⓪，②，③，⑤ (順不同)
(カ),(キ) ⓪，② (順不同)

解説

(1) ヒストグラムより，大きい方から 10 番目の記録と 11 番目の記録は 25 m 以上 30 m 未満の階級に含まれることがわかるから，第 3 四分位数もこの階級に含まれる。
よって　ア④

(2) (1) と同様に考えると，第 1 四分位数は 15 m 以上 20 m 未満の階級に含まれることがわかる。これと (1) の結果に矛盾する箱ひげ図は
イ，ウ，エ，オ⓪，②，③，⑤

(3) ⓪ a の箱ひげ図は最初のデータよりも第 1 四分位数が大きくなっているから，矛盾している。

① b の箱ひげ図は最初のデータよりも最大値，第 3 四分位数，中央値，第 1 四分位数，最小値すべてが大きくなっているから，矛盾しているとはいえない。

② c の箱ひげ図は最初のデータよりも最大値が小さくなっているから，矛盾している。

③ d の箱ひげ図は最初のデータよりも最大値と第 3 四分位数が大きくなっており，第 1 四分位数と最小値が小さくなっているから，矛盾しているとはいえない。

よって，分析結果と箱ひげ図が矛盾するものは　カ，キ⓪，②

□3

解答　(ア).(イ) 3.5　(ウ).(エ) 2.8　(オ).(カキ) 1.01　(ク) 5　(ケ) 2
(コ) ②　(サ).(シ) 1.5　(ス).(セ) 0.7　(ソ) ④　(タ) ②　(チ) ⑤

解説

(1) 教科ごとの順位の度数分布表は右のようになる。

| 順位 | 1 | 2 | 3 | 4 | 5 |
|---|---|---|---|---|---|
| 教科 1 | 3 | 1 | 1 | 4 | 1 |
| 教科 2 | 0 | 1 | 4 | 2 | 3 |
| 教科 3 | 1 | 5 | 0 | 1 | 3 |
| 教科 4 | 3 | 1 | 4 | 1 | 1 |
| 教科 5 | 3 | 2 | 1 | 2 | 2 |

教科 1 について，順位の高い方から 5 番目の順位は 3，6 番目の 順位は 4 であるから，中央値 A は

$$\frac{3+4}{2}={}^{ア}3.{}^{イ}5$$

教科 5 の順位の平均値 B は

$$\frac{1}{10}(1\cdot3+2\cdot2+3\cdot1+4\cdot2+5\cdot2)=\frac{28}{10}={}^{ウ}2.{}^{エ}8$$

教科 2 の順位の分散 C は

$$\frac{1}{10}(1^2\cdot0+2^2\cdot1+3^2\cdot4+4^2\cdot2+5^2\cdot3)-3.7^2=14.7-13.69={}^{オ}1.{}^{カキ}01$$

(2) 教科 1 の順位が教科 3 の順位より上位である生徒は
生徒 3，生徒 5，生徒 7，生徒 8，生徒 9
教科 2 の順位が教科 4 の順位より上位である生徒は
生徒 3，生徒 6
よって　D$={}^{ク}5$，E$={}^{ケ}2$
また，同じ順位がないように順位をつけるから，すべての生徒は教科 $j$ と教科 $k$ のいずれか一方をもう一方よりも上位に順位づけする。
したがって　$w_1+w_2=10$ (コ②)
教科 2 の中央値は 3.5，平均値は 3.7，教科 3 の中央値は 2.0，平均値は 3.0 であるから，
$(j,\ k)=(2,\ 3)$ のとき　$u=3.5-2.0={}^{サ}1.{}^{シ}5$，$v=3.7-3.0={}^{ス}0.{}^{セ}7$
$(j,\ k)=(2,\ 4)$ のとき　$u=0.5$，$v=1.1$
また　$(j,\ k)=(2,\ 3)$ のとき $w=4$，$(j,\ k)=(2,\ 4)$ のとき $w=2$
相関図の中で，$(u,\ w)=(1.5,\ 4)$，$(0.5,\ 2)$ を満たす点があるのは ③，④ であり，
$(v,\ w)=(0.7,\ 4)$，$(1.1,\ 2)$ を満たす点があるのは ①，② である。
さらに，$(j,\ k)=(4,\ 2)$ のとき　$w=10-2=8$，$u=-0.5$，$v=-1.1$
よって，$u$ と $w$ の相関図は ソ④，$v$ と $w$ の相関図は タ② である。また，$u$ と $w$，$v$ と $w$ はともに負の相関があり，$v$ と $w$ の方が負の相関が強いことが読み取れる。
したがって　$r_2<r_1<0$ (チ⑤)



□1

解答　(ア) ⑤　(イ),(ウ) ①，③　または　③，①　(エ) ⑨　(オ) ⑧
(カ) ⑦

解説

(1) 各都市のヒストグラムから，最低気温を読み取ると
東京は　0℃ ～ 5℃
N 市は　−10℃ ～ −5℃
M 市は　5℃ ～ 10℃
よって，都市名と箱ひげ図の組み合わせは
東京― c，N 市― b，M 市― a　(ア⑤)

(2) ⓪，①，②　散布図から，東京と N 市の最高気温の間には正の相関があり，東京と M 市の最高気温の間には負の相関があることが読み取れる。① が正しい。
③，④　散布図から，東京と O 市の散布図の点の方が，東京と N 市の散布図の点より，右上がりの直線に沿って分布する傾向が強いことが読み取れる。
すなわち，東京と O 市の最高気温の間の相関の方が，東京と N 市の最高気温の間の相関より強いことが読み取れる。③ が正しい。
よって　イ①，ウ③　(または　イ③，ウ①)

(3) N 市の摂氏での最高気温 $x_N$ のデータを $x_{N_1}$，$x_{N_2}$，……，$x_{N_{365}}$，華氏での最高気温 $y_N$ のデータを $y_{N_1}$，$y_{N_2}$，……，$y_{N_{365}}$ と表す。
$x_N$ と $y_N$ の間には　$y_N=\frac{9}{5}x_N+32$ ……①　の関係がある。
このとき，$x_N$，$y_N$ の分散を $X$，$Y$ で表すと　$Y=\left(\frac{9}{5}\right)^2X$ ……②
よって　$\frac{Y}{X}=\frac{81}{25}$　(エ⑨)
東京 (摂氏) の最高気温 $x_T$ のデータを $x_{T_1}$，$x_{T_2}$，……，$x_{T_{365}}$，平均値を $\overline{x_T}$，N 市の摂氏での平均値を $\overline{x_N}$，華氏での平均値を $\overline{y_N}$ と表す。
ここで，① の関係から　$\overline{y_N}=\frac{9}{5}\overline{x_N}+32$ ……③
共分散の定義から

$$\begin{aligned}W&=\frac{1}{365}\{(x_{T_1}-\overline{x_T})(y_{N_1}-\overline{y_N})+(x_{T_2}-\overline{x_T})(y_{N_2}-\overline{y_N})\\&\qquad+\cdots\cdots+(x_{T_{365}}-\overline{x_T})(y_{N_{365}}-\overline{y_N})\}\\&=\frac{1}{365}\left\{(x_{T_1}-\overline{x_T})\cdot\frac{9}{5}(x_{N_1}-\overline{x_N})+(x_{T_2}-\overline{x_T})\cdot\frac{9}{5}(x_{N_2}-\overline{x_N})\right.\\&\qquad\left.+\cdots\cdots+(x_{T_{365}}-\overline{x_T})\cdot\frac{9}{5}(x_{N_{365}}-\overline{x_N})\right\}\\&=\frac{9}{5}\cdot\frac{1}{365}\{(x_{T_1}-\overline{x_T})(x_{N_1}-\overline{x_N})+(x_{T_2}-\overline{x_T})(x_{N_2}-\overline{x_N})\\&\qquad+\cdots\cdots+(x_{T_{365}}-\overline{x_T})(x_{N_{365}}-\overline{x_N})\}\\&=\frac{9}{5}Z\end{aligned}$$

ゆえに　$\frac{W}{Z}=\frac{9}{5}$　(オ⑧)
東京 (摂氏) の分散を $s_T{}^2$ と表すと，相関係数の定義から

$$V=\frac{W}{\sqrt{s_{\mathrm{T}}{}^2}\sqrt{Y}}=\frac{\frac{9}{5}Z}{\sqrt{s_{\mathrm{T}}{}^2}\sqrt{\left(\frac{9}{5}\right)^2X}}=\frac{Z}{\sqrt{s_{\mathrm{T}}{}^2}\sqrt{X}}=U$$

よって　$\frac{V}{U}=1$　（カ ⑦）

参考　相関係数は単位の取り方によらないから，$\frac{V}{U}=1$ となることは明らかである。

参考　（等式②，③について）

$a$，$b$ を定数とする。変量 $x$ のデータから $y=ax+b$ によって新しい変量 $y$ のデータが得られるとき，$x$，$y$ のデータの平均値を $\overline{x}$，$\overline{y}$，分散 $s_x{}^2$，$s_y{}^2$，標準偏差を $s_x$，$s_y$ とすると，次のことが成り立つ。

$\overline{y}=a\overline{x}+b$，　$s_y{}^2=a^2s_x{}^2$，　$s_y=|a|s_x$

証明　変量 $x$ についてのデータを $x_1$，$x_2$，……，$x_n$ とする。

変量 $y$ のデータの平均値 $\overline{y}$ は

$$\overline{y}=\frac{1}{n}(y_1+y_2+\cdots\cdots+y_n)=\frac{1}{n}\{(ax_1+b)+(ax_2+b)+\cdots\cdots+(ax_n+b)\}$$

$$=\frac{1}{n}\{a(x_1+x_2+\cdots\cdots+x_n)+nb\}=a\cdot\frac{1}{n}(x_1+x_2+\cdots\cdots+x_n)+b$$

よって　$\overline{y}=a\overline{x}+b$

また，$y_k-\overline{y}=ax_k+b-(a\overline{x}+b)=a(x_k-\overline{x})$ であることから，変量 $y$ のデータの分散 $s_y{}^2$ は

$$s_y{}^2=\frac{1}{n}\{(y_1-\overline{y})^2+(y_2-\overline{y})^2+\cdots\cdots+(y_n-\overline{y})^2\}$$

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

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

よって　$s_y{}^2=a^2s_x{}^2$

これより，変量 $y$ のデータの標準偏差は　$s_y=|a|s_x$　終

2

解答　（ア）③　（イ）④　（ウ），（エ）④，⑦（順不同）　（オ）⓪
（カ）⓪　（キ）①　（ク）②

解説

(1)　図1より，2013年の開花日の最大値は135以上である。

よって，2013年のヒストグラムは　ア ③

図1より，2017年の開花日の最大値は120以上125未満である。

よって，2017年のヒストグラムは　イ ④

(2)　⓪　図3より，モンシロチョウの初見日の最小値はツバメの初見日の最小値と同じである。

よって，正しい。

①　図3より，モンシロチョウの初見日の最大値はツバメの初見日の最大値より大きい。

よって，正しい。

②　図3より，モンシロチョウの初見日の中央値はツバメの初見日の中央値より大きい。

よって，正しい。

③　図3より，モンシロチョウの初見日の四分位範囲はおよそ　$104-84=20$（日）

ツバメの初見日の四分位範囲はおよそ　$97-88=9$（日）

したがって，モンシロチョウの初見日の四分位範囲はツバメの初見日の四分位範囲の3倍より小さい。

よって，正しい。

④　モンシロチョウの初見日の四分位範囲はおよそ20日であり，15日より大きい。

よって，正しくない。

⑤　ツバメの初見日の四分位範囲はおよそ9日であり，15日以下である。

よって，正しい。

⑥　図4において，原点を通り傾き1の直線（実線）上にある点が，モンシロチョウとツバメの初見日が同じ所を表している。

重なった点を考慮すると実線上に少なくとも4点あるから，モンシロチョウとツバメの初見日が同じ所は少なくとも4地点ある。

よって，正しい。

⑦　図4において，破線は切片が $-15$ および15で傾きが1の直線である。

モンシロチョウの初見日を $X$，ツバメの初見日を $Y$ とおくと，2本の破線の方程式は $Y=X+15$，$Y=X-15$ と書ける。

図4において，直線 $Y=X+15$ の上側の部分に点があるから，その座標を $(x,\ y)$ とおくと　$y>x+15$　すなわち　$y-x>15$

したがって，モンシロチョウの初見日とツバメの初見日の差が15日より大きい地点がある。

よって，正しくない。

以上から　ウ ④，エ ⑦（または ウ ⑦，エ ④）

(3)　$X$ の偏差 $x_1-\overline{x}$，$x_2-\overline{x}$，……，$x_n-\overline{x}$ の平均値は　0　（オ ⓪）

また，各 $x_i$ に対して $x_i'=\frac{1}{s}x_i-\frac{\overline{x}}{s}$ と書けるから，データ $X'$ の平均値を $\overline{x'}$，標準偏差を $s'$ とすると

$$\overline{x'}=\frac{1}{s}\overline{x}-\frac{\overline{x}}{s}=0\quad（カ ⓪）$$

$s>0$ であるから　$s'=\left|\frac{1}{s}\right|s=1$　（キ ①）

また，$M'$ と $T'$ の散布図を考える。

変換 $x_i'=\frac{1}{s}x_i-\frac{\overline{x}}{s}$ において，$\frac{1}{s}>0$ であるから，変換後の散布図は，変換前の散布図を縦，横に拡大，縮小および平行移動したものである。

よって，①，③ は適さない。

また，⓪ において，散布図上のすべての点は $M'$，$T'$ ともに $-1$ から1の間にある。

したがって，$M'$，$T'$ ともに偏差の平方の平均値，すなわち分散が1より小さくなり，標準偏差も1より小さくなるから，⓪ は適さない。

以上から　ク ②

参考　（偏差の平均値）

一般に，データ $X$ の偏差 $x_1-\overline{x}$，$x_2-\overline{x}$，……，$x_n-\overline{x}$ の平均値は，次の計算からわかるように常に0になる。

$$\frac{1}{n}\{(x_1-\overline{x})+(x_2-\overline{x})+\cdots\cdots+(x_n-\overline{x})\}=\frac{1}{n}\{(x_1+x_2+\cdots\cdots+x_n)-n\overline{x}\}$$

$$=\frac{1}{n}(x_1+x_2+\cdots\cdots+x_n)-\overline{x}=\overline{x}-\overline{x}=0$$

参考　（標準化）

一般に，各 $x_i$ について $x_i'=\frac{x_i-\overline{x}}{s}$ と変換することで，変換後のデータの平均値を0，標準偏差を1にすることができる。この操作を標準化という。