データの分析　基本編　　　　(　　)組(　　)番　名前(　　　　　　　　)

解説

1 (1)　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

$$\frac{1}{25}(40\times2+50\times5+60\times8+70\times7+80\times3)=\frac{1540}{25}=61.6\ (点)$$

また，データの平均値が最大となるのは，データの各値が階級内の最大の値となるとき，すなわち，データの各値が階級内の最小の値となるときより 9 点だけ大きいときである。

よって，平均値も最小となるときより 9 点だけ大きくなるから

$61.6+9=70.6$ (点)

したがって，データの平均値のとり得る範囲は　　61.6 点以上 70.6 点以下

(2)　データの大きさは 25 であるから，中央値は得点が高い方から 13 番目の生徒の得点である。

よって，70 点以上の生徒の人数は $7+3=10$ であるから，60 点以上 69 点以下の階級の中で，得点が高い方から 3 番目の生徒の得点が中央値である。

この階級に含まれる値を大きさの順に並べると

62，63，63，65，66，67，68，68

したがって，データの中央値は　　67 点

別解　(2)　データの大きさは 25 であるから，中央値は得点が低い方から 13 番目の生徒の得点である。

よって，59 点以下の生徒の人数は $2+5=7$ であるから，60 点以上 69 点以下の階級の中で，得点が低い方から 6 番目の生徒の得点が中央値である。

この階級に含まれる値を大きさの順に並べると

62，63，63，65，66，67，68，68

したがって，データの中央値は　　67 点

解説

2 (1)　大きさの順に並べると　　1.2，2.0，2.3，2.4，2.7，3.2

データの大きさは 6 であるから，中央値は 3 番目の値と 4 番目の値の平均値である。

よって，中央値は　　$\frac{1}{2}(2.3+2.4)=2.35$ (kg)

平均値は　　$\frac{1}{6}(1.2+2.0+2.3+2.4+2.7+3.2)=\frac{13.8}{6}=2.3$ (kg)

(2)　実際の平均値は (1) で求めたものより 0.1 kg 大きいから，どれかが 0.6 kg 少ない。

データの値から 1 つ選んで 0.6 kg 加えた結果，中央値が 2.55 kg になるものは，2.3 kg のみである。

よって，誤っている数値は　　2.3 kg，

正しい数値は　　　2.9 kg

参考　2.7 や 3.2 の値を大きくしても中央値は変化しないから，他の 4 つの数値についてのみ考えればよい。

解説

3 (1)　店舗数は 8 であるから，中央値は安い方から 4 番目の価格と 5 番目の価格の平均値である。

a 以外の価格を安い順に並べると

498，500，525，550，550，555，560

$a\leqq525$ 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

$\frac{1}{2}(525+550)=537.5$ (円)

$a\geqq550$ 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

$526\leqq a\leqq549$ のとき，4 番目の価格は a 円，5 番目の価格は 550 円であるから，中央値は　$\frac{1}{2}(a+550)$ 円

この値は a の値によってすべて異なる。

ゆえに，中央値は $2+(549-526+1)=26$ (通り) の値があり得る。

注意　$526\leqq a\leqq549$ のとき $538\leqq\frac{1}{2}(a+550)\leqq549.5$ であるから，

$537.5,\ 550,\ \frac{1}{2}(a+550)\ (526\leqq a\leqq549)$

の値はすべて異なる。

(2)　平均値が 535 円であるから

$$\frac{1}{8}(498+500+525+550+550+555+560+a)=535$$

よって　　$3738+a=4280$　　　ゆえに　　$a=542$ (円)

このとき，中央値は　　$\frac{1}{2}(542+550)=546$ (円)

解説

4 (1)　第 2 四分位数は　　$Q_2=22$，

第 1 四分位数は　　$Q_1=15$，

第 3 四分位数は　　$Q_3=31$

(2)　第 2 四分位数は　　$Q_2=\frac{19+21}{2}=20$，

第 1 四分位数は　　$Q_1=\frac{7+12}{2}=9.5$，

第 3 四分位数は　　$Q_3=\frac{30+33}{2}=31.5$

(3)　第 2 四分位数は　　$Q_2=24$，

第 1 四分位数は　　$Q_1=\frac{11+20}{2}=15.5$，

第 3 四分位数は　　$Q_3=\frac{42+44}{2}=43$

(4)　第 2 四分位数は　　$Q_2=\frac{31+33}{2}=32$，

第 1 四分位数は　　$Q_1=18$，

第 3 四分位数は　　$Q_3=45$

解説

5 大きさの順に並べると

A 弁当　　16，17，21，22，23，25，27，28，33，40

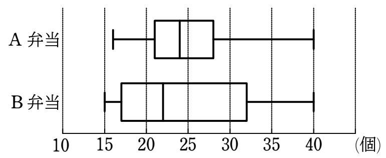
B 弁当　　15，16，17，18，20，24，28，32，35，40

よって，それぞれのデータの最小値，第 1 四分位数，中央値，第 3 四分位数，最大値は，順に

A 弁当　16，21，$\frac{23+25}{2}=24$，28，40

B 弁当　15，17，$\frac{20+24}{2}=22$，32，40

よって，箱ひげ図は下の図のようになる。



また，A 弁当の四分位範囲は　　$Q_3-Q_1=28-21=7$

よって　　$Q_3+1.5\times(Q_3-Q_1)=28+1.5\times7=38.5$

ゆえに，A 弁当のデータの最大値 40 は外れ値である。

次に，B 弁当の四分位範囲は　　$Q_3-Q_1=32-17=15$

よって　　$Q_3+1.5\times(Q_3-Q_1)=32+1.5\times15=54.5$

ゆえに，B 弁当のデータの最大値は外れ値ではない。

解説

6 ①　最小値が 30 点台で，第 1 四分位数が 40 点より大きいから，30 点台の生徒が何人いるかを読み取ることはできない。

よって，① は正しくない。

②　第 1 四分位数が 50 点未満であるから，50 点以上の生徒が 180 人以上いるかを読み取ることはできない。

よって，② は正しくない。

③　中央値が 60 点未満であるから，半数以上の生徒が 60 点未満である。よって，③ は正しい。

以上から，正しいものは　　③

解説

7 ①　数学の範囲は 60 点より大きく，国語と英語の範囲はともに 60 点より小さい。

よって，範囲が最も大きいのは数学である。

ゆえに，① は正しくない。

②　国語の四分位範囲は 20 点より小さく，数学と英語の四分位範囲はともに 20 点より大きい。

よって，四分位範囲が最も小さいのは国語である。

ゆえに，② は正しい。

③　英語の第 1 四分位数が 50 点であるから，60 点以上の生徒が 300 人以上いるかを読み取ることはできない。

よって，③ は正しくない。

④　国語と英語の第 1 四分位数がともに 50 点台であるから，50 点未満の生徒はともに 100 人以下である。

また，数学の第 1 四分位数が 40 点台であるから，50 点未満の生徒は 100 人以上である。

よって，④ は正しい。

⑤　この箱ひげ図からは，数学に 30 点台の生徒がいるかどうかを読み取ることはできない。

よって，⑤ は正しくない。

以上から，正しいものは　　②，④

参考　⑤について，国語と英語は最小値が 30 点台の範囲にあるため，30 点台の生徒がいることがわかる。

解説

[8] ヒストグラムから，A 市，M 市のデータの最小値，第 1 四分位数，中央値，第 3 四分位数，最大値が入る階級は，それぞれ次のようになることがわかる。(単位は ℃)

	A 市	M 市
最小値	4 ～ 6	4 ～ 6
第 1 四分位数	10 ～ 12	8 ～ 10
中央値	12 ～ 14	10 ～ 12 または 12 ～ 14
第 3 四分位数	14 ～ 16	14 ～ 16
最大値	18 ～ 20	18 ～ 20

これらと矛盾しない箱ひげ図は　　A 市：④，　M 市：①

注意　M 市の中央値について，M 市のデータの 15 番目に高い日は 12 ～ 14 の階級にあり，16 番目に高い日は 10 ～ 12 の階級にある。
よって，中央値が入る階級は 10 ～ 12 または 12 ～ 14 となる。

解説

[9] (1)　平均値 $\overline{x}$ は

$$\overline{x}=\frac{1}{6}(5+1+8+10+5+7)=\frac{1}{6}\times 36=6 \text{ (個)}$$

よって，分散 s^2 は

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

$$=\frac{1}{6}(1+25+4+16+1+1)=\frac{1}{6}\times 48=8$$

参考　各値の 2 乗の平均値 $\overline{x^2}$ を求めることにより，分散 s^2 を求めると次のようになる。

$$\overline{x^2}=\frac{1}{6}(5^2+1^2+8^2+10^2+5^2+7^2)=\frac{1}{6}\times 264=44$$

よって，分散 s^2 は

$$s^2=\overline{x^2}-(\overline{x})^2=44-6^2=8$$

(2)　標準偏差 s は

$$s=\sqrt{s^2}=\sqrt{8}=2\sqrt{2}=2.82\cdots\fallingdotseq 2.8 \text{ (個)}$$

解説

[10] (1)　$\overline{x}=\frac{1}{8}(7+4+10+1+7+3+10+6)=\frac{1}{8}\times 48=6$ (点)

$$\overline{x^2}=\frac{1}{8}(7^2+4^2+10^2+1^2+7^2+3^2+10^2+6^2)=\frac{1}{8}\times 360=45$$

(2)　$s^2=\overline{x^2}-(\overline{x})^2=45-6^2=9$

$s=\sqrt{s^2}=\sqrt{9}=3$ (点)

参考　偏差の 2 乗の平均値を求めることにより，分散 s^2 を求めると次のようになる。

$$s^2=\frac{1}{8}\{(7-6)^2+(4-6)^2+(10-6)^2+(1-6)^2+(7-6)^2+(3-6)^2+(10-6)^2+(6-6)^2\}$$

$$=\frac{1}{8}(1+4+16+25+1+9+16+0)=\frac{1}{8}\times 72=9$$

解説

[11] (1)　$\frac{1}{20}(3\times 8+8\times 12)=\frac{1}{20}\times 120=6$

(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

$$a\times 8+b\times 12=13\times 8+73\times 12=980$$

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

解説

[12] (1)　$\overline{y}=\overline{x}-10=35-10=25$

$s_y{}^2=1^2s_x{}^2=1\times 16=16$

$s_y=\sqrt{s_y{}^2}=\sqrt{16}=4$

(2)　$\overline{y}=3\overline{x}=3\times 35=105$

$s_y{}^2=3^2s_x{}^2=9\times 16=144$

$s_y=\sqrt{s_y{}^2}=\sqrt{144}=12$

(3)　$\overline{y}=-\frac{1}{2}\overline{x}+6=-\frac{1}{2}\times 35+6=-\frac{23}{2}$

$s_y{}^2=\left(-\frac{1}{2}\right)^2s_x{}^2=\frac{1}{4}\times 16=4$

$s_y=\sqrt{s_y{}^2}=\sqrt{4}=2$

参考　s_y については，$s_y=|a|s_x$ を用いて求めてもよい。(a は x の係数)

(1)　$s_y=|1|s_x=\sqrt{s_x{}^2}=\sqrt{16}=4$

(2)　$s_y=|3|s_x=3\sqrt{s_x{}^2}=3\sqrt{16}=12$

(3)　$s_y=\left|-\frac{1}{2}\right|s_x=\frac{1}{2}\sqrt{s_x{}^2}=\frac{1}{2}\sqrt{16}=2$

解説

[13] もとの得点を x (点)，調整後の得点を y (点) とすると　　$y=2.5x+30$

よって，y の平均値 $\overline{y}$，分散 $s_y{}^2$，標準偏差 s_y は

$$\overline{y}=2.5\overline{x}+30=2.5\times 68+30=200 \text{ (点)}$$

$$s_y{}^2=2.5^2s_x{}^2=\frac{25}{4}\times 36=225$$

$$s_y=\sqrt{s_y{}^2}=\sqrt{225}=15 \text{ (点)}$$

解説

[14] (1)　$u=\frac{x-644}{7}$ から

u のデータ　　4，7，0，3，-2，0

u^2 のデータ　　16，49，0，9，4，0

よって，変量 u のデータの平均値は

$$\overline{u}=\frac{1}{6}\{4+7+0+3+(-2)+0\}=\frac{1}{6}\times 12=2$$

また　　$\overline{u^2}=\frac{1}{6}(16+49+0+9+4+0)=\frac{1}{6}\times 78=13$

よって，変量 u のデータの分散は　　$s_u{}^2=13-2^2=9$

変量 u のデータの標準偏差は　　$s_u=\sqrt{9}=3$

(2)　$u=\frac{x-644}{7}$ から　　$x=7u+644$

よって，変量 x のデータの平均値 $\overline{x}$，分散 $s_x{}^2$，標準偏差 s_x は

$$\overline{x}=7\overline{u}+644=7\times 2+644=658$$

$$s_x{}^2=7^2s_u{}^2=49\times 9=441$$

$$s_x=\sqrt{s_x{}^2}=\sqrt{441}=21$$

解説

[15] 番号 6 の得点 (漢字 3 点，英単語 7 点) の位置に点がある散布図は ③ だけである。

よって　　③

解説

[16] ①　散布図から，漢字テストの点数が増えるにつれて，50 m 走のタイムが短くなる傾向があることが読み取れる。

よって，漢字テストの点数と 50 m 走のタイムには負の相関関係がある。

②，③　散布図からは，漢字テストと 50 m 走のタイムとの因果関係は読み取ることはできない。

以上から，正しいものは　①

参考　一般に，因果関係があるかどうかをデータのみから判断することはできない。

解説

[17] 散布図から

① は相関関係がない。

② は負の相関関係がある。

③ は正の相関関係がある。

よって　　① 0.04　　② -0.71　　③ 0.87

解説

[18] 求める相関係数は

$$\frac{13.77}{5.05\times 4.26}=\frac{13.77}{21.513}=0.640\cdots\fallingdotseq 0.64$$

解説

[19] 右手の握力を x(kg)，左手の握力を y(kg) とする。

x のデータの平均値は　　$\overline{x}=\frac{370}{10}=37$

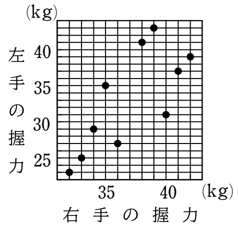
y のデータの平均値は　$\overline{y}=\dfrac{330}{10}=33$

番号	x	$x-\overline{x}$	$(x-\overline{x})^2$
1	36	-1	1
2	42	5	25
3	35	-2	4
4	33	-4	16
5	38	1	1
6	32	-5	25
7	39	2	4
8	40	3	9
9	34	-3	9
10	41	4	16
計	370		110

番号	y	$y-\overline{y}$	$(y-\overline{y})^2$	$(x-\overline{x})(y-\overline{y})$
1	27	-6	36	6
2	39	6	36	30
3	35	2	4	-4
4	25	-8	64	32
5	41	8	64	8
6	23	-10	100	50
7	43	10	100	20
8	31	-2	4	-6
9	29	-4	16	12
10	37	4	16	16
計	330		440	164

この表から，求める相関係数は　$\dfrac{164}{\sqrt{110\times440}}=\dfrac{164}{220}=0.745\cdots\cdots\fallingdotseq0.75$

参考　このデータの散布図は次のようになる。



解説

20 ①　散布図から，最低気温が上がるにつれて最高気温も上がる傾向にあることが読み取れる。

よって，① は正しい。

②　最高気温が 14 ℃ 未満の部分に，点が 8 個確認できる。

よって，② は正しい。

③　それぞれの最大値と最小値から，最低気温の範囲は $16-1=15$ (℃) 未満，最高気温の範囲は $26-8=18$ (℃) 以上であることがわかる。

よって，最低気温の範囲より最高気温の範囲の方が大きいから，③ は正しくない。

④　最低気温が 10 ℃ を超える部分のうち，最高気温が 18 ℃ 未満である部分に，点が 1 個確認できる。

よって，④ は正しくない。

⑤　正しい。

⑥　正しい内容である ① から，最低気温と最高気温の間には正の相関関係がある。

よって，⑥ は正しくない。

以上から，正しくないものは　③，④，⑥

参考　相関係数は 0.68 である。

解説

21 x のデータ x_k $(k=1,\ 2,\ \cdots\cdots,\ 50)$ に対応する z のデータを z_k とし，y のデータ y_k $(k=1,\ 2,\ \cdots\cdots,\ 50)$ に対応する w のデータを w_k とする。

x，y，z，w のデータの平均値を，それぞれ $\overline{x}$，$\overline{y}$，$\overline{z}$，$\overline{w}$ とする。

また，x と y の共分散を s_{xy}，z と w の共分散を s_{zw} とし，x，y，z，w の標準偏差をそれぞれ s_x，s_y，s_z，s_w とする。

(1)　$\overline{z}=\overline{x}+3$，$\overline{w}=4\overline{y}$ から

z の偏差は　$z_k-\overline{z}=(x_k+3)-(\overline{x}+3)=x_k-\overline{x}$

w の偏差は　$w_k-\overline{w}=4y_k-4\overline{y}=4(y_k-\overline{y})$

よって

$$s_{zw}=\frac{1}{50}\{(z_1-\overline{z})(w_1-\overline{w})+(z_2-\overline{z})(w_2-\overline{w})+\cdots\cdots+(z_{50}-\overline{z})(w_{50}-\overline{w})\}$$
$$=\frac{1}{50}\{(x_1-\overline{x})\cdot4(y_1-\overline{y})+(x_2-\overline{x})\cdot4(y_2-\overline{y})+\cdots\cdots+(x_{50}-\overline{x})\cdot4(y_{50}-\overline{y})\}$$
$$=4\cdot\frac{1}{50}\{(x_1-\overline{x})(y_1-\overline{y})+(x_2-\overline{x})(y_2-\overline{y})+\cdots\cdots+(x_{50}-\overline{x})(y_{50}-\overline{y})\}$$
$$=4\cdot s_{xy}=4\cdot192=768$$

また　$s_z=|1|s_x=s_x$，$s_w=|4|s_y=4s_y$

よって，z と w の相関係数は　$\dfrac{s_{zw}}{s_zs_w}=\dfrac{4s_{xy}}{s_x\cdot4s_y}=\dfrac{s_{xy}}{s_xs_y}=0.55$

(2)　$\overline{z}=\dfrac{1}{2}\overline{x}$，$\overline{w}=2\overline{y}-5$ から

z の偏差は　$z_k-\overline{z}=\dfrac{1}{2}x_k-\dfrac{1}{2}\overline{x}=\dfrac{1}{2}(x_k-\overline{x})$

w の偏差は　$w_k-\overline{w}=(2y_k-5)-(2\overline{y}-5)=2(y_k-\overline{y})$

よって

$$s_{zw}=\frac{1}{50}\{(z_1-\overline{z})(w_1-\overline{w})+(z_2-\overline{z})(w_2-\overline{w})+\cdots\cdots+(z_{50}-\overline{z})(w_{50}-\overline{w})\}$$
$$=\frac{1}{50}\left\{\frac{1}{2}(x_1-\overline{x})\cdot2(y_1-\overline{y})+\cdots\cdots+\frac{1}{2}(x_{50}-\overline{x})\cdot2(y_{50}-\overline{y})\right\}$$
$$=\frac{1}{50}\{(x_1-\overline{x})(y_1-\overline{y})+\cdots\cdots+(x_{50}-\overline{x})(y_{50}-\overline{y})\}=s_{xy}=192$$

また　$s_z=\left|\dfrac{1}{2}\right|s_x=\dfrac{1}{2}s_x$，$s_w=|2|s_y=2s_y$

よって，z と w の相関係数は　$\dfrac{s_{zw}}{s_zs_w}=\dfrac{s_{xy}}{\dfrac{1}{2}s_x\cdot2s_y}=\dfrac{s_{xy}}{s_xs_y}=0.55$

(3)　$\overline{z}=-2\overline{x}-2$，$\overline{w}=\dfrac{2}{3}\overline{y}$ から

z の偏差は　$z_k-\overline{z}=(-2x_k-2)-(-2\overline{x}-2)=-2(x_k-\overline{x})$

w の偏差は　$w_k-\overline{w}=\dfrac{2}{3}y_k-\dfrac{2}{3}\overline{y}=\dfrac{2}{3}(y_k-\overline{y})$

よって

$$s_{zw}=\frac{1}{50}\{(z_1-\overline{z})(w_1-\overline{w})+(z_2-\overline{z})(w_2-\overline{w})+\cdots\cdots+(z_{50}-\overline{z})(w_{50}-\overline{w})\}$$
$$=\frac{1}{50}\left\{-2(x_1-\overline{x})\cdot\frac{2}{3}(y_1-\overline{y})+\cdots\cdots+(-2)\cdot(x_{50}-\overline{x})\cdot\frac{2}{3}(y_{50}-\overline{y})\right\}$$
$$=-\frac{4}{3}\cdot\frac{1}{50}\{(x_1-\overline{x})(y_1-\overline{y})+\cdots\cdots+(x_{50}-\overline{x})(y_{50}-\overline{y})\}$$
$$=-\frac{4}{3}\cdot s_{xy}=-\frac{4}{3}\cdot192=-256$$

また　$s_z=|-2|s_x=2s_x$，$s_w=\left|\dfrac{2}{3}\right|s_y=\dfrac{2}{3}s_y$

よって，z と w の相関係数は　$\dfrac{s_{zw}}{s_zs_w}=\dfrac{-\dfrac{4}{3}s_{xy}}{2s_x\cdot\dfrac{2}{3}s_y}=-\dfrac{s_{xy}}{s_xs_y}=-0.55$

参考　a，b，c，d を定数とし，2 つの変量 x，y から $z=ax+b$，$w=cy+d$ によって新しい変量 z，w が得られたとする。

x と y の共分散を s_{xy}，z と w の共分散を s_{zw}，x と y の相関係数を r_{xy}，z と w の相関係数を r_{zw} とすると，次のことが成り立つ。

$s_{zw}=acs_{xy}$，$ac>0$ のとき　$r_{zw}=r_{xy}$，

$ac<0$ のとき　$r_{zw}=-r_{xy}$

証明　z の偏差は

$z_k-\overline{z}=(ax_k+b)-(a\overline{x}+b)=a(x_k-\overline{x})$

w の偏差は

$w_k-\overline{w}=(cy_k+d)-(c\overline{y}+d)=c(y_k-\overline{y})$

よって，各変量のデータの大きさを n とすると

$$s_{zw}=\frac{1}{n}\{(z_1-\overline{z})(w_1-\overline{w})+\cdots\cdots+(z_n-\overline{z})(w_n-\overline{w})\}$$
$$=\frac{1}{n}\{a(x_1-\overline{x})\cdot c(y_1-\overline{y})+\cdots\cdots+a(x_n-\overline{x})\cdot c(y_n-\overline{y})\}$$
$$=ac\cdot\frac{1}{n}\{(x_1-\overline{x})(y_1-\overline{y})+\cdots\cdots+(x_n-\overline{x})(y_n-\overline{y})\}=acs_{xy}$$

$$r_{zw}=\frac{s_{zw}}{s_zs_w}=\frac{acs_{xy}}{|a|s_x\cdot|c|s_y}=\frac{acs_{xy}}{|ac|s_xs_y}$$

よって，$ac>0$ のとき　$r_{zw}=\dfrac{acs_{xy}}{acs_xs_y}=\dfrac{s_{xy}}{s_xs_y}=r_{xy}$

$ac<0$ のとき　$r_{zw}=\dfrac{acs_{xy}}{-acs_xs_y}=-\dfrac{s_{xy}}{s_xs_y}=-r_{xy}$

解説

22 変量 x のデータを x_1，x_2，$\cdots\cdots$，x_{30} とし，データの平均値を $\overline{x}$ とする。

y，z，w のデータについても同様に定め，平均値をそれぞれ $\overline{y}$，$\overline{z}$，$\overline{w}$ とすると

$\overline{z}=2\overline{x}+10$，$\overline{w}=3\overline{y}-20$

よって，z の偏差は　$z_k-\overline{z}=(2x_k+10)-(2\overline{x}+10)=2(x_k-\overline{x})$

w の偏差は　$w_k-\overline{w}=(3y_k-20)-(3\overline{y}-20)=3(y_k-\overline{y})$

よって，x と y の共分散を s_{xy}，z と w の共分散を s_{zw} とすると

$$s_{zw}=\frac{1}{30}\{(z_1-\overline{z})(w_1-\overline{w})+(z_2-\overline{z})(w_2-\overline{w})+\cdots\cdots+(z_{30}-\overline{z})(w_{30}-\overline{w})\}$$
$$=\frac{1}{30}\{2(x_1-\overline{x})\cdot3(y_1-\overline{y})+2(x_2-\overline{x})\cdot3(y_2-\overline{y})+\cdots\cdots+2(x_{30}-\overline{x})\cdot3(y_{30}-\overline{y})\}$$
$$=6\cdot\frac{1}{30}\{(x_1-\overline{x})(y_1-\overline{y})+(x_2-\overline{x})(y_2-\overline{y})+\cdots\cdots+(x_{30}-\overline{x})(y_{30}-\overline{y})\}$$
$$=6\cdot s_{xy}=6\cdot217=1302$$

また，x，y，z，w の標準偏差をそれぞれ s_x，s_y，s_z，s_w とすると

$s_z=|2|s_x=2s_x$,　$s_w=|3|s_y=3s_y$

よって，z と w の相関係数は　$\dfrac{s_{zw}}{s_z s_w}=\dfrac{6s_{xy}}{2s_x\cdot 3s_y}=\dfrac{s_{xy}}{s_x s_y}=0.78$

解説

23 (1)　変量 x のデータの平均値は

$$\frac{1}{10}(13+17+20+23+28+34+36+40+44+45)=\frac{300}{10}=30\ (点)$$

変量 x のデータの中央値は　$\dfrac{1}{2}(28+34)=31$ (点)

(2)　修正前の変量 y のデータの小さい方から 5，6 番目の値が，それぞれ 15，20 であるから，中央値は　$\dfrac{1}{2}(15+20)=17.5$ (点)

修正後の変量 y のデータの小さい方から 5，6 番目の値が，それぞれ 20，25 であるから，中央値は　$\dfrac{1}{2}(20+25)=22.5$ (点)

(3)　修正前，修正後とも，散布図から正の相関関係があることがわかる。

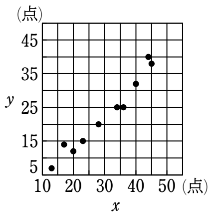
よって　$r_1>0$，$r_2>0$

また，修正後の方が，散布図が右上がりの直線に沿って分布する傾向がより強くなる。

よって　$r_1<r_2$

これらを満たす $(r_1,\ r_2)$ の組は　①

参考　修正後の散布図は次のようになる。



解説

24 平均値 $\overline{x}$ と分散 s^2 は

$$\overline{x}=\frac{1}{5n}(a\cdot 4n+b\cdot n)=\frac{4a+b}{5}$$

$$s^2=\frac{1}{5n}(a^2\cdot 4n+b^2\cdot n)-\left(\frac{4a+b}{5}\right)^2$$

$$=\frac{4a^2+b^2}{5}-\frac{16a^2+8ab+b^2}{25}=\frac{4}{25}(a-b)^2$$

$a-b<0$ から，標準偏差 s は　$s=\dfrac{2}{5}(b-a)$

よって，a 点をとった生徒，b 点をとった生徒の偏差値をそれぞれ T_a，T_b とすると

$$T_a=10\times\frac{a-\dfrac{4a+b}{5}}{\dfrac{2}{5}(b-a)}+50=45$$

$$T_b=10\times\frac{b-\dfrac{4a+b}{5}}{\dfrac{2}{5}(b-a)}+50=70$$

解説

25 ①　箱ひげ図から，範囲は数学の方が大きいが，四分位範囲は理科の方が大きいことがわかる。

よって，① は正しくない。

②　散布図から，数学が 50 点未満である生徒は，全員理科が 60 点未満であることがわかる。

よって，② は正しい。

③　散布図から，理科が 60 点未満である部分のうち，数学が 70 点以上である部分に，点が 1 個確認できる。

よって，③ は正しくない。

④　散布図から，数学の得点が最も低い生徒を示す点より下側に，点が 1 個確認できる。

よって，④ は正しくない。

⑤　箱ひげ図から，第 3 四分位数は理科の方が大きいことがわかる。

よって，⑤ は正しい。

⑥　散布図から，数学の得点が増えるにつれて理科の得点も増える傾向にあることが読み取れる。

すなわち，数学と理科の間には正の相関関係があることがわかる。

よって，⑥ は正しくない。

⑦　散布図から，数学も理科も 90 点以上の部分に，点が 2 個確認できる。

よって，⑦ は正しい。

以上から，正しいものは　②，⑤，⑦

参考　相関係数は 0.87 である。