7 データの分析 [数学Ⅰ]

出題傾向 と対策

「データの分析」は，数学の本流からは切り離された分野で，他の分野の学習をしながら自然に出会うということが少ないため，どうしても"疎遠"になりがちである。問題演習を通して基本用語の定義・意味を1つ1つしっかり理解し，記憶していくことが肝要である。

センター試験攻略のポイント

❶ データ

実験，調査から得られた結果を**データ**という。また，そこで測った数量を**変量**といい，文字 x，y などで表す。

ある集団の傾向・特徴を調べることが**データの分析**の目的である。以下においては，変量 x の n 個の値を小さい順に並べた

x_1，x_2，x_3，$\cdots$，x_n …(＊)

というデータについて考える。(n をこのデータ(＊)の大きさという。)

❷ 代表値

(1) 変量 x の**平均値**は

$$\bar{x}=\frac{x_1+x_2+x_3+\cdots+x_n}{n} \quad \cdots\frac{総和}{個数}$$

「エックスバー」と読む

$$=\frac{1}{n}\sum_{k=1}^{n}x_k$$

〈注意〉 「数列」(数学B)を学んでいる人は，上のようにシグマ記号Σを使うとデータの分析における様々な量を簡潔に表すことができる。ぜひ積極的に使おう。本書では今後，Σ記号を使わない表現と使う表現を併用する。

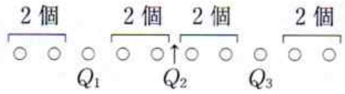
(2) (＊)において中央の順位にくる値を x の**中央値(メジアン)**といい，本書では今後，記号 $\tilde{x}$(「エックスチルダ」と読む)で表す。

〈注意〉 上記以外の代表値である「最頻値」については，❼「度数分布」で触れる。

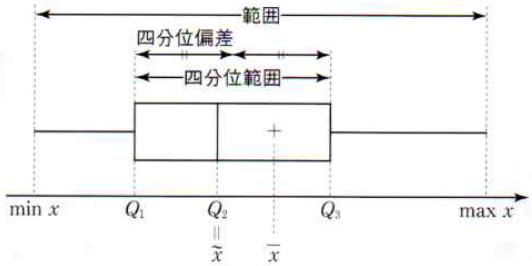
❸ データの散らばり

平均値，中央値だけではわからない「データの散らばり」について考える。

(1) データを，小さい順に並べて4等分するときの境界に位置する変量の値を**四分位数**(Quartile)という。つまり，中央値＝**第2四分位数**(Q_2)を境に同じ個数ずつの2グループに分けたとき，左，右各々のグループの中央値をそれぞれ**第1四分位数**(Q_1)，**第3四分位数**(Q_3)という。データの大きさが10のときは次のようになる。



(2) 最小値(min x)，最大値(max x)，および四分位数(Q_1，Q_2，Q_3)を用いて，データの分布・散らばりを把握することを**五数要約**といい，次の**箱ひげ図**で視覚化できる。



箱ひげ図には，上のように平均値を「＋」の記号で書き入れることもある。

❹ 偏　差

変量 x からその平均値 $\bar{x}$ を引いて得られる**偏差**：$x-\bar{x}$ を用いてデータの散らばりを表すことを考える。

(1) 偏差の2乗の平均値を変量 x の**分散**といい，$s_x{}^2$ と表す。すなわち

$$s_x{}^2=\frac{1}{n}\sum_{k=1}^{n}(x_k-\bar{x})^2 \quad \cdots 偏差平方の平均$$

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

分散が大きいほど，データの散らばりが大きいと考えられる。

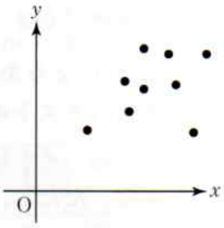
(2) 分散は，次のように計算することもできる。

$s_x{}^2=\overline{x^2}-(\bar{x})^2$ ←(2乗の平均)−(平均の2乗)

(3) 分散の平方根をとってもとの変量と単位を揃えたものを**標準偏差**といい，s_x と表す。

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

❺ データの相関



(1) ある集団における2つの変量 x, y の関係を調べるには，座標平面上に点 $(x,\ y)$ をとって得られる**散布図（相関図）**を描くとよい。

(2) x, y の偏差どうしの積 $(x-\overline{x})(y-\overline{y})$ の平均値を**共分散**といい，s_{xy} で表す。すなわち

$$s_{xy}=\frac{1}{n}\sum_{k=1}^{n}(x_k-\overline{x})(y_k-\overline{y})$$
$$=\frac{1}{n}\{(x_1-\overline{x})(y_1-\overline{y})+(x_2-\overline{x})(y_2-\overline{y})+\cdots+(x_n-\overline{x})(y_n-\overline{y})\}$$

これを x, y の標準偏差の積で割ったものが x と y の**相関係数** r_{xy} である。すなわち

$$r_{xy}=\frac{s_{xy}}{s_x s_y}$$

相関係数の値は，$-1 \leqq r_{xy} \leqq 1$ の範囲にある。r_{xy} が $+1$ に近いときは，偏差どうしの積の値が正に偏っている。つまり，x の増加にともない y も増加する傾向が強い。このとき x, y の間には**正の相関**があるという。逆に r_{xy} が -1 に近いとき，x の増加にともない y は減少する傾向が強く，**負の相関**があるという。また，r_{xy} が0に近いときは「相関がない」という。

❻ 変量の変換

(1) 変量 x と，それから作られる別の変量の平均値，分散，標準偏差の関係は次の通り。(a, b は定数)

変量	平均値	分散	標準偏差	
x	$\overline{x}$	$s_x{}^2$	s_x	
ax	$a\overline{x}$	$a^2s_x{}^2$	$\|a\|s_x$	…①
$x+b$	$\overline{x}+b$	$s_x{}^2$	s_x	…②
$ax+b$	$a\overline{x}+b$	$a^2s_x{}^2$	$\|a\|s_x$	…③

〈注意〉 x, y の相関係数は，上記①～③の変換を施しても(a が正ならば)変化しない。

(2) 変量 x の平均値 $\overline{x}$ を求める際，平均値に近そうで計算しやすい値 X を選び，$x-X$ の平均値を用いて $\overline{x}$ を求めると簡便である。

$$\overline{x}=\frac{1}{n}\sum_{k=1}^{n}(x_k-X)+X \qquad \text{…前記②を利用}$$
$$=\frac{(x_1-X)+(x_2-X)+\cdots+(x_n-X)}{n}+X$$

このとき X を**仮平均**という。

(3) 変量 x から，その平均値と標準偏差を用いて作られる変量 $z=\dfrac{x-\overline{x}}{s_x}$ は，前記①～③より $\overline{z}=0$, $s_z=1$ を満たす。このような変量の変換を**基準化**もしくは**標準化**という。

この基準化された z を用いて任意の平均値・標準偏差をもつ変量を作ることができる。たとえば $Z=10z+50$ は，③より $\overline{Z}=50$, $s_z=10$ を満たす変量である。テストにおける得点 x をこのように変換して得られる変量 Z が，受験生お馴染みの**偏差値**である。

❼ 度数分布

(1) データの大きさ n が大きいとき，変量の値を同じ幅(**階級の幅**)のいくつかの区間(**階級**)に分け，各階級に対してその区間に入っているものの個数(**度数**)を対応させた**度数分布**を考えるとよい。これを表にしたものを**度数分布表**という。また，各階級値の n に対する割合：$\dfrac{\text{度数}}{n}$ を**相対度数**という。

各階級には，実際には異なる値が含まれているが，その違いを無視し，全ての値が階級の真ん中の値(**階級値**)をとるものとみなせば様々な計算を簡略化できる。

〈例〉 通学時間(単位：分)階級の幅：20
データの大きさ：50

階級 以上～未満	階級値	度数	相対度数
0～20	10	3	0.06
20～40	30	14	0.28
40～60	50	23	0.46
60～80	70	8	0.16
80～100	90	2	0.04

(2) 度数を棒グラフで表したものを**ヒストグラム**という。

(3) 度数が最大である階級の階級値を**最頻値(モード)**という。上の例では，最頻値は50である。

基本編　センター試験の必出項目を check!

【解答　p.10】

以下，小数の形で解答する場合，指定された桁数の一つ下の桁を四捨五入し，解答せよ。途中で割り切れた場合，指定された桁まで⓪にマークすること。

1　平均値　(目標時間2分)

下の表は，あるクラスの生徒10人に対して行われた国語と英語の小テスト(各10点満点)の得点をまとめたものである。

生徒番号	1	2	3	4	5	6	7	8	9	10	平均値
国語	9	10	4	7	10	5	5	7	6	7	A
英語	9	9	8	6	8	B	8	9	C	7	8.0

10人の国語の得点の平均値Aは $\boxed{ア}.\boxed{イ}$ 点である。また，生徒6と生徒9の英語の得点B，Cは，関係式 B＋C＝$\boxed{ウエ}$ を満たす。

(13年本試・改)

2　仮平均　(目標時間3分)

ゲームに参加した15人の得点は

33　44　30　38　29　26　43　23　28
34　33　26　36　30　27

であった。この15人の得点の平均値は

$\boxed{アイ}.\boxed{ウ}$ 点である。　(11年本試・改)

3　中央値(その1)　(目標時間1分)

あるクラスの生徒10人に対して行われた国語の小テスト(10点満点)の得点は次の通りであった。

9　10　4　7　10　5　5　7　6　7

このデータを値の小さいものから順に並べて考えることにより，国語の得点の中央値は

$\boxed{ア}.\boxed{イ}$ 点である。　(13年本試・改)

4　中央値(その2)　(目標時間4分)

10名からなるあるクラスで実施した100点満点の英語のテストの得点は

43　55　A　64　36　48　46　71　65　50

であった。ただし，Aは0以上100以下の整数とする。

(1) Aを除く9人の得点の中央値は $\boxed{アイ}$ 点である。

(2) A＝47，53のとき，10人の得点の中央値はそれぞれ $\boxed{ウエ}$ 点，$\boxed{オカ}.\boxed{キ}$ 点である。

(3) 10人の得点の中央値として $\boxed{ク}$ 通りの値があり得る。

(09年本試・改)

5　四分位数　(目標時間2分)

あるクラスで，女子10人と男子9人に対して数学のテストを行った。女子の得点 x と男子の得点 y をそれぞれ値の小さいものから順に並べると次のようになった。

x：50　51　54　56　65　67　71　72　75　79
y：45　47　50　51　59　60　62　79　96

データの第1四分位数，第2四分位数(中央値)，第3四分位数を Q_1，Q_2，Q_3 と表すと，変量 x については

$Q_1=\boxed{アイ}$，$Q_2=\boxed{ウエ}$，$Q_3=\boxed{オカ}$

である。また変量 y については

$Q_1=\boxed{キク}.\boxed{ケ}$，$Q_2=\boxed{コサ}$，
$Q_3=\boxed{シス}.\boxed{セ}$

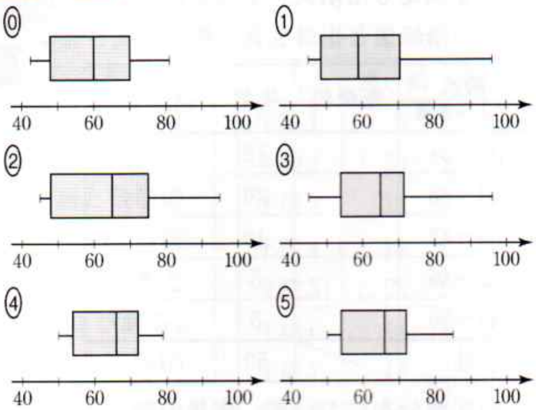
である。

6　箱ひげ図　(目標時間3分)

[5]のデータを下の表のようにまとめた。[5]の結果も用いて表を完成させよ。

	女子(x)	男子(y)
最小値		
第1四分位数(Q_1)		
中央値(Q_2)		
第3四分位数(Q_3)		
最大値		
範囲		
四分位範囲		

この表をもとに，x のデータ，y のデータを表す箱ひげ図を，下の⓪～⑤から選ぶと，順に ア ， イ となる。



7 偏差・分散

目標時間 2 分

生徒番号	国語 x	偏差 $x-\overline{x}$	偏差平方 $(x-\overline{x})^2$
1	9		
2	10		
3	4		
4	7		
5	10		
6	5		
7	5		
8	7		
9	6		
10	7		
平均値	$\overline{x}=7.0$	ア . イ	ウ . エ

上の表はあるクラスの生徒 10 人に対して行われた国語のテストの結果をまとめたものであり，点数 x の平均値 $\overline{x}$ を求めると

$$\overline{x}=7.0$$

であった。これをもとに表を完成させると，偏差の平均値は ア . イ ，偏差平方の平均値は ウ . エ である。また，変量 x の分散，標準偏差は，それぞれ

$s_x{}^2=$ オ . カ ，$s_x=$ キ . ク

である。

8 分散

目標時間 2 分

大きさ 6 のデータがあり，変量 x の値は

5　0　2　5　6　3

である。このデータの平均値は $\overline{x}=$ ア . イ であり，分散は $s_x{}^2=$ ウ . エオ である。

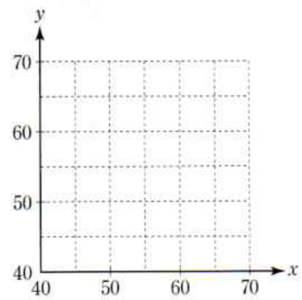
9 散分図・相関係数

目標時間 5 分

あるクラスの生徒 6 人が，数学と英語のテスト(いずれも 100 点満点)を受けた。数学の得点 x，英語の得点 y をまとめると下の表のようになった。$(x,\ y)$ の各組に対応する点を下の座標平面に「•」として書き入れ，散布図を作れ。なお，x，y の平均値 $\overline{x}$，$\overline{y}$ はすでに求めて表に書き込んである。

生徒番号	数学 x	英語 y	$x-\overline{x}$	$y-\overline{y}$	$(x-\overline{x})(y-\overline{y})$
1	43	55			
2	50	48			
3	53	60			
4	55	70			
5	64	58			
6	65	69			
平均値	55	60	0	0	アイ

上の表を完成させることにより，x の偏差 $x-\overline{x}$ と y の偏差 $y-\overline{y}$ の積の平均値を求めると アイ となる。また，x の分散，y の分散，x と y の共分散は，それぞれ



$s_x{}^2=$ ウエ ，$s_y{}^2=$ オカ ，$s_{xy}=$ キク

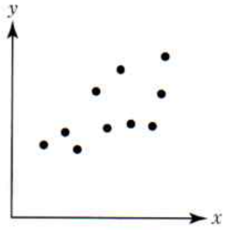
である。よって，x と y の相関係数は，

$r_{xy}=$ ケ . コサ である。

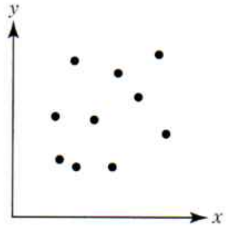
10 相関図

（目標時間 1 分）

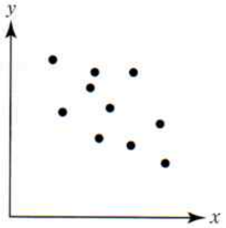
次の散布図（相関図）で表される分布における，x と y の相関係数 r の値 ア ～ エ としてもっとも近いものを，下の ⓪～⑥ から 1 つずつ選べ。



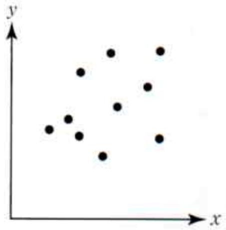
$r=$ ア



$r=$ イ



$r=$ ウ



$r=$ エ

⓪ -1　① -0.7　② -0.3　③ 0
④ 0.3　⑤ 0.7　⑥ 1

11 変量の変換

（目標時間 4 分）

2 つの変量 x, y があり，それぞれの平均値，分散は

$$\overline{x}=4.0,\ s_x^2=0.4,\ \overline{y}=8.0,\ s_y^2=2.0$$

とする。

(1) 変量 x を使って新しい変量 t を $t=5x+30$ で定めると，変量 t の平均値，分散はそれぞれ $\overline{t}=$ アイ ，$s_t^2=$ ウエ となる。

(2) 変量 y を使って新しい変量 u を $u=y-$ オ で定めると，変量 u の平均値は 0 になる。

(3) 変量 y を使って新しい変量 v を $v=\dfrac{\sqrt{\boxed{カ}}}{\boxed{キ}}y$ で定めると，変量 v の分散は x の分散と同じになる。

(4) x と y の相関係数を r_{xy} とし，(1)の t と(2)の u の相関係数を r_{tu} とする。このとき，次の ⓪～② のうち ク が成り立つ。

⓪ $r_{tu}=r_{xy}$　① $r_{tu}>r_{xy}$　② $r_{tu}<r_{xy}$

（06 年本試・改）

12 度数分布

（目標時間 3 分）

あるクラブの 50 人に対して，右手の握力 x（単位 kg）を測定した結果，下の度数分布表を得た。

表中の階級値と相対度数の欄を完成させよ。

階級 以上～未満	階級値	度数	相対度数
30～34		10	
34～38		20	
38～42		10	
42～46		5	
46～50		5	
計		50	1.0

この度数分布において，階級の幅は， ア kg である。階級値を用いて x の平均値を求めると

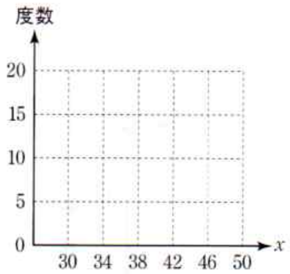
$\overline{x}=$ イウ (kg)

である。また，このデータの中央値は エオ kg であり，最頻値は カキ kg である。

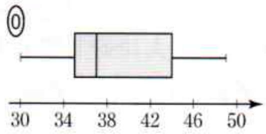
13 ヒストグラム

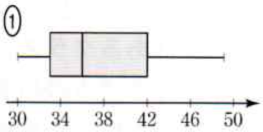
（目標時間 2 分）

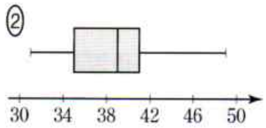
12 のデータのヒストグラムを右図の中に完成させよ。

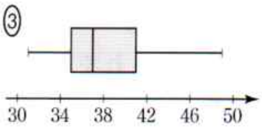


また，12 のデータを表した箱ひげ図としてもっとも適切なものは，次の ⓪～③ のうち ア である。

⓪


①


②


③


実戦編 センター試験の過去問に challenge!

【解答　p.16】

制限時間 9 分

1 右の表は，あるクラスの生徒 9 人に対して行われた英語と数学のテスト(各 20 点満点)の得点をまとめたものである。ただし，テストの得点は整数値である。また，表の数値はすべて正確な値であり，四捨五入されていないものとする。

	英　語	数　学
生徒 1	9	15
生徒 2	20	20
生徒 3	18	14
生徒 4	18	17
生徒 5	A	8
生徒 6	18	C
生徒 7	14	D
生徒 8	15	14
生徒 9	18	15
平均値	16.0	15.0
分　散	B	10.00
相関係数	0.500	

以下，小数の形で解答する場合，指定された桁(けた)数の一つ下の桁を四捨五入し，解答せよ。途中で割り切れた場合，指定された桁まで⓪にマークすること。

(1) 生徒 5 の英語の得点 A は $\boxed{\text{アイ}}$ 点であり，9 人の英語の得点の分散 B の値は $\boxed{\text{ウエ}}.\boxed{\text{オカ}}$ である。また，9 人の数学の得点の平均値が 15.0 点であることと，英語と数学の得点の相関係数の値が 0.500 であることから，生徒 6 の数学の得点 C と生徒 7 の数学の得点 D の関係式

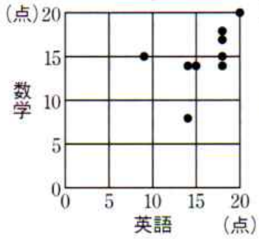
$\mathrm{C}+\mathrm{D}=\boxed{\text{キク}}$

$\mathrm{C}-\mathrm{D}=\boxed{\text{ケ}}$

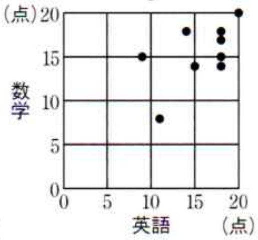
が得られる。したがって，C は $\boxed{\text{コサ}}$ 点，D は $\boxed{\text{シス}}$ 点である。

(2) 9 人の英語と数学の得点の散布図(相関図)として適切なものは $\boxed{\text{セ}}$ である。$\boxed{\text{セ}}$ に当てはまるものを，次の⓪～③のうちから一つ選べ。

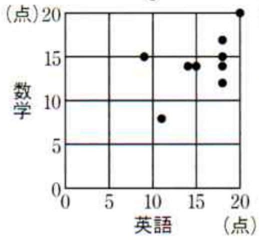
⓪



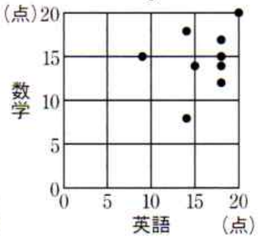
①



②



③



(3) 生徒 10 が転入して来たので，その生徒に対して同じテストを行ったところ，英語が 16 点，数学が 15 点であった。次の $\boxed{\text{ソ}}$，$\boxed{\text{タ}}$，$\boxed{\text{チ}}$ に当てはまるものを，下の⓪～⑧のうちから一つずつ選べ。ただし，同じものを 2 回以上選んでもよい。

転入前における 9 人の生徒について，英語の得点の平均値を m，英語の得点の標準偏差を s，英語と数学の得点の相関係数を r とする。また，転入後における 10 人の生徒の英語の得点の平均値を m'，英語の得点の標準偏差を s'，英語と数学の得点の相関係数を r' とする。このとき，

$$\frac{m'}{m}=\boxed{\text{ソ}},\quad \frac{s'}{s}=\boxed{\text{タ}},\quad \frac{r'}{r}=\boxed{\text{チ}}$$

である。

⓪ -1　　① 0　　② 1

③ $\dfrac{9}{10}$　　④ $\left(\dfrac{9}{10}\right)^2$　　⑤ $\dfrac{3\sqrt{10}}{10}$

⑥ $\dfrac{10}{9}$　　⑦ $\left(\dfrac{10}{9}\right)^2$　　⑧ $\dfrac{\sqrt{10}}{3}$

(14 年本試・改)

制限時間９分

2 ある高等学校のＡクラスには全部で 20 人の生徒がいる。次の表は，その 20 人の生徒の国語と英語のテストの結果をまとめたものである。表の横軸は国語の得点を，縦軸は英語の得点を表し，表中の数値は，国語の得点と英語の得点の組み合わせに対応する人数を表している。ただし，得点は 0 以上 10 以下の整数値をとり，空欄は 0 人であることを表している。たとえば，国語の得点が 7 点で英語の得点が 6 点である生徒の人数は 2 である。

英語（点）＼国語（点）	0	1	2	3	4	5	6	7	8	9	10
10											
9											
8							1		1		
7						5					
6					4	1	1	2			
5						2					
4				1	1						
3				1							
2											
1											
0											

また，右の表は，Ａクラスの 20 人について，上の表の国語と英語の得点の平均値と分散をまとめたものである。ただし，表の数値はすべて正確な値であり，四捨五入されていない。

	国　語	英　語
平均値	B	6.0
分　散	1.60	C

以下，小数の形で解答する場合，指定された桁（けた）数の一つ下の桁を四捨五入し，解答せよ。途中で割り切れた場合，指定された桁まで⓪にマークすること。

(1) Ａクラスの 20 人のうち，国語の得点が 4 点の生徒は［ア］人であり，英語の得点が国語の得点以下の生徒は［イ］人である。

(2) Ａクラスの 20 人について，国語の得点の平均値Ｂは［ウ］.［エ］点であり，英語の得点の分散Ｃの値は［オ］.［カキ］である。

(3) 国語の得点の中央値は［ク］.［ケ］点であり，英語の得点の第 1 四分位数は［コ］.［サ］点である。

(4) Ａクラスの 20 人のうち，国語の得点が平均値［ウ］.［エ］点と異なり，かつ，英語の得点も平均値 6.0 点と異なる生徒は［シ］人である。

Ａクラスの 20 人について，国語の得点と英語の得点の相関関係の値は［ス］.［セソタ］である。

(5) 英語のテストで採点ミスが見つかった，国語が 5 点である生徒の英語の成績は，正しくは 5 点が 1 人，6 点が 3 人，7 点が 4 人であった。得点を修正した後，次の 3 つの値は修正前と比べてどのようになったか。［チ］，［ツ］，［テ］に当てはまるものを，下の⓪～③のうちから 1 つずつ選べ。

英語の得点の平均値は［チ］。

英語の得点の分散は［ツ］。

国語の得点と英語の得点の相関係数は［テ］。

⓪ 変化しない　① 増加する

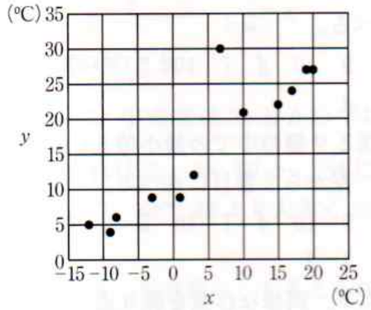
② 減少する　③ どうなるかわからない

（12 年本試・改）

制限時間９分

3 ある都市におけるある年の月ごとの最低気温を変量 x，最高気温を変量 y とする。ただし，単位は℃とし，最低気温と最高気温は，一日の最低気温と最高気温について月ごとに平均をとり，小数第１位を四捨五入したものとする。

次の図は，変量 x と変量 y の散布図（相関図）である。



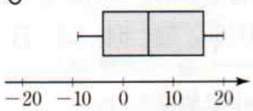
以下，小数の形で解答する場合は，指定された桁(けた)数の一つ下の桁を四捨五入し，解答せよ。途中で割り切れた場合は，指定された桁まで⓪にマークすること。

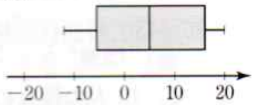
(1) １月から12月までの変量 x は次のとおりであった。

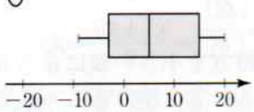
−12　−9　−3　3　10　17　20　19　15　7　1　−8（単位は℃）

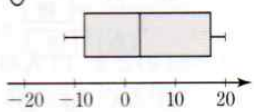
変量 x の平均値は，$\overline{x}=$ ア . イ ℃，中央値は ウ . エ ℃である。

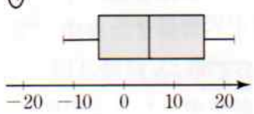
また，変量 x の分布を表した箱ひげ図は次の⓪〜⑤のうち オ である。

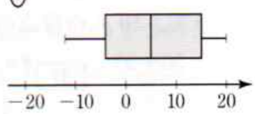
⓪


①


②


③


④


⑤


(2) １月から12月までの12か月を，変量 x が0℃未満の四つの月からなるＡグループと，0℃以上の八つの月からなるＢグループとに分けて分析したところ，Ａグループにおける変量 y の平均値は6.0℃で，Ｂグループにおける変量 y の平均値は21.5℃であった。このとき，１月から12月までの変量 y の平均値は カキ . ク ℃である。

変量 x と変量 y の相関図のデータの中で，入力ミスが見つかった。変量 x の値が7℃，変量 y の値が30℃となっている月の変量 y の値は，正しくは18℃であった。

(3) この誤りを修正すると，変量 y の平均値は ケ . コ ℃減少する。また，変量 y の分散は サ する。そして，x と y の相関係数は シ する。ただし， サ と シ については，当てはまるものを，次の⓪〜②のうちから一つずつ選べ。

⓪　修正前より増加　　①　修正前より減少

②　修正前と一致

(4) 修正後のデータをもとに，気温の単位を摂氏（単位℃）から華氏（単位°F）に変更した。すなわち，新しい変量 X(°F)，Y(°F) を

$$X=\frac{9}{5}x+32,\quad Y=\frac{9}{5}y+32$$

によって定めた。

変量 X の平均値は，$\overline{X}=$ スセ . ソ である。

変量 x，X の標準偏差をそれぞれ s_x，s_X とすると，$\dfrac{s_X}{s_x}=$ タ . チ である。

x と y の相関係数，X と Y の相関係数をそれぞれ r_{xy}，r_{XY} とすると $\dfrac{r_{XY}}{r_{xy}}=$ ツ . テ である。

7 データの分析の解答

基本編

1 $\overset{\text{平均値}}{\mathrm{A}}=\dfrac{9+10+4+7+10+5+5+7+6+7}{10}$ ←$\dfrac{\text{総和}}{\text{個数}}$

$=\dfrac{70}{10}=\boxed{7}^{ア}.\boxed{0}^{イ}$(点)

(英語の得点合計)$=\underbrace{8.0}_{\text{平均値}}\times\underbrace{10}_{\text{個数}}=80$(点)

よって　$\mathrm{B}+\mathrm{C}=80-(9+9+8+6+8+8+9+7)$

$\mathrm{B}+\mathrm{C}=\boxed{16}^{ウエ}$

参考 (平均値)$=\dfrac{(\text{総和})}{(\text{個数})}$

(総和)=(平均値)×(個数)

のどちらの形でも使えるようにしましょう。

2 仮平均を 30 点として，得点 x と 30 の差を考えると次のようになる。

x	33	44	30	38	29	26	43	23	28	34	33	26	36	30	27
$x-30$	3	14	0	8	−1	−4	13	−7	−2	4	3	−4	6	0	−3

よって，$x-30$ の平均値は

$$\overline{x-30}=\frac{1}{15}(3+14+8-1-4+13-7-2+4+3-4+6-3)$$

$=2$(点)

したがって，求める平均値は

$\overline{x}=30+2=\boxed{32}^{アイ}.\boxed{0}^{ウ}$(点)

参考 ・**1** に比べて，データの各数値が大きいですね。このようなときは「仮平均」の利用が有効です。仮平均は，平均値に近そうで，しかもなるべくキレイな値を選びます。

・変量 x の平均値 $\overline{x}$ が，仮平均 X を用いて

$\overline{x}=\overline{x-X}+X$

と求まることを示します。変量 x の値を x_1, x_2, …, x_n とすると

$$\overline{x-X}=\frac{1}{n}\sum_{k=1}^{n}(x_k-X)$$

$$=\frac{1}{n}\cdot\sum_{k=1}^{n}x_k-\frac{1}{n}\cdot nX=\overline{x}-X$$

よって　$\overline{x}=\overline{x-X}+X$

(…本冊と同様やや高度な参考事項を示す)

3 データを小さいものから順に並べ直すと

$\underbrace{4\ 5\ 5\ 6\ 7}_{5\text{個}}\ \underbrace{7\ 7\ 9\ 10\ 10}_{5\text{個}}$

よって，求める中央値は

$\dfrac{7+7}{2}=\boxed{7}^{ア}.\boxed{0}^{イ}$(点)

参考 ・データを小さい順に並べる際，筆者は次のようにしています。

1° 10 個の中での最小値 4 を斜線で消し，左端に並べる。

9　10　~~4~~　7　10　5　5　7　6　7

4

2° 残る 9 個の中での最小値 5 (1 個)を斜線で消し，左から 2 番目に並べる。

9　10　~~4~~　7　10　~~5~~　5　7　6　7

4　5

3° 以下，同様な作業を繰り返す。

・データの大きさが偶数(本問では 10 個)のとき，“中央” にはデータがありませんから，その両隣を平均したものを「中央値」と定めます。

中央値：$\dfrac{□+△}{2}$

4 (1) A 以外の 9 人の得点を小さい順に並べると

$\underbrace{36\ 43\ 46\ 48}_{4\text{個}}\ 50\ \underbrace{55\ 64\ 65\ 71}_{4\text{個}}$ …①

よって，この 9 人の得点の中央値は $\boxed{50}^{アイ}$ 点である。

(2) A＝47 のとき，10 人の得点を小さい順に並べると

$\underbrace{36\ 43\ 46\ \overset{}{\underset{\mathrm{A}}{47}}\ 48}_{5\text{個}}\ \underbrace{50\ 55\ 64\ 65\ 71}_{5\text{個}}$

よって，10 人の得点の中央値は

$\dfrac{48+50}{2}=\boxed{49}^{ウエ}$(点) …②

A＝53 のとき，10 人の得点を小さい順に並べると

$\underbrace{36\ 43\ 46\ 48\ 50}_{5\text{個}}\ \underbrace{\underset{\mathrm{A}}{53}\ 55\ 64\ 65\ 71}_{5\text{個}}$

よって，10 人の得点の中央値は

$\dfrac{50+53}{2}=\boxed{51}^{オカ}.\boxed{5}^{キ}$(点) …③

(3) ①より，A の値に応じて，次の 3 タイプに分けられる。

(i) A≦48 のとき

10 人の得点を小さい順に並べると

・・・・・48 ┆ 50・・・・・

A　5 個　　　5 個

よって 10 人の得点の中央値は②と同じく 49 点

(ii) A≧55 のとき　　　　　　　　Aによらない

下のようになるから中央値は $\dfrac{50+55}{2}=52.5$(点)

・・・・・50 ┆ 55・・・・・

5 個　　　5 個　A

(iii) 49≦A≦54 のとき

③と同様に，中央値は 50 と A の平均 $\dfrac{50+A}{2}$ である。これは A＝49, 50, …, 54 に応じて 6 個の値をとり，全て 49 より大きく 52.5 より小さい。

以上(i)〜(iii)より，中央値として考えられる値は全部で　$1+1+6=\boxed{8}^{\text{ク}}$(通り)

5

x について

50　51　54　56　65 ┆ 67　71　72　75　79

5 個　　　5 個

$Q_2=\dfrac{65+67}{2}=\boxed{66}^{\text{ウエ}}$(点)　←中央値 $\tilde{x}$ でもある

50　51　54　56　65 ┆ 67　71　72　75　79

$Q_1=\boxed{54}^{\text{アイ}}$(点)，$Q_3=\boxed{72}^{\text{オカ}}$(点)

y について

45　47　50　51　59　60　62　79　96

4 個　　　4 個

$Q_2=\boxed{59}^{\text{コサ}}$(点)

45　47 ┆ 50　51 ┆ 60　62 ┆ 79　96

2 個　2 個　2 個　2 個

$Q_1=\dfrac{47+50}{2}=\boxed{48}^{\text{キク}}.\boxed{5}^{\text{ケ}}$(点)

$Q_3=\dfrac{62+79}{2}=\boxed{70}^{\text{シス}}.\boxed{5}^{\text{セ}}$(点)

参考　四分位数を求める手順は次の通りです。

1°　中央値を求める。これが Q_2 である。

2°　中央値を境に，下位，上位の 2 つのグループ(同じ個数)に分ける。元のデータの大きさが奇数のときは，中央値をとる 1 つを除いた 2 グループに分ける。

3°　下位グループの中央値，上位グループの中央値を求める。これらがそれぞれ Q_1, Q_3 である。

6

下表のようになる。

		女子(x)	男子(y)
(*)	最小値	**50**	**45**
(*)	第 1 四分位数(Q_1)	54	48.5
(*)	中央値(Q_2)	66	59
(*)	第 3 四分位数(Q_3)	72	70.5
(*)	最大値	**79**	**96**
	範囲	**29**	**51**
	四分位範囲	**18**	**22**

(範囲)＝(最大値)−(最小値)

(四分位範囲)＝Q_3-Q_1

を用いている。このデータを表す箱ひげ図は，

$x\cdots\boxed{④}^{\text{ア}}$，$y\cdots\boxed{⓪}^{\text{イ}}$

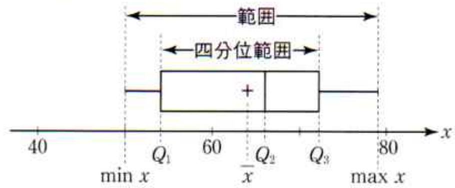
である。

参考　• 箱ひげ図を選ぶとき，たとえば次のようにします。

x と y とで顕著に違うものとして「最大値」と「範囲」が挙げられます。この 2 つに注目すると，y を表すものは①，②，③に絞られ，四分位数 Q_1, Q_2, Q_3 も考えると，①が正しいことがわかります。x については，残りの⓪，④，⑤の中から最大値が 79 である④を選びます。念のため，四分位数などもチェックしておきましょう。

• 上表の(*)部のことを「五数要約」といいます。

• 箱ひげ図に関して，女子の得点 x を例として説明を加えます。



左のひげ「—」，箱の左「□ Q_2」，箱の右「□ Q_2」，右のひげ「—」の部分には，それぞれ全体の約 $\dfrac{1}{4}$ の人数が分布していることになります。

箱ひげ図には平均値($\bar{x}=64$ となります)を「＋」で書き入れることもあります。

単元別学習

7 各生徒の偏差および偏差平方は下のようになる。

生徒番号	国語 x	偏差 $x-\overline{x}$	偏差平方 $(x-\overline{x})^2$
1	9	2	4
2	10	3	9
3	4	−3	9
4	7	0	0
5	10	3	9
6	5	−2	4
7	5	−2	4
8	7	0	0
9	6	−1	1
10	7	0	0
平均値	7.0	0.0	4.0

偏差 $x-\overline{x}$ の平均値は $\boxed{0}^{ア}.\boxed{0}^{イ}$

偏差平方の平均値は

$$\frac{1}{10}(4+9+9+9+4+4+1)=\frac{40}{10}=\boxed{4}^{ウ}.\boxed{0}^{エ}$$

x の分散 $s_x{}^2$ とは，偏差平方の平均値であり

$$s_x{}^2=\boxed{4}^{オ}.\boxed{0}^{カ}$$

x の標準偏差は

$$s_x=\sqrt{s_x{}^2}=\sqrt{4.0}=\boxed{2}^{キ}.\boxed{0}^{ク}$$

参考 偏差 $x-\overline{x}$ の平均値は，x のとる値を x_1, x_2, …，x_n として

$$\frac{1}{n}\sum_{k=1}^{n}(x_k-\overline{x})=\frac{1}{n}\sum_{k=1}^{n}x_k-\frac{1}{n}\cdot n\overline{x}=\overline{x}-\overline{x}$$

より，つねに 0 となります。

8

$$\overline{x}=\frac{5+0+2+5+6+3}{6}=\frac{21}{6}$$

$$=\boxed{3}^{ア}.\boxed{5}^{イ}$$

$$s_x{}^2=\overline{x^2}-(\overline{x})^2 \quad \leftarrow(2\text{乗の平均})-(\text{平均の}2\text{乗})$$

$$=\frac{25+0+4+25+36+9}{6}-\left(\frac{21}{6}\right)^2$$

$$=\frac{99}{6}-\frac{441}{6^2}=\frac{594-441}{6^2}=\frac{17}{4}$$

$$=\boxed{4}^{ウ}.\boxed{25}^{エオ}$$

参考 • 平均値 $\overline{x}$ の値が整数でないので，各偏差の値も整数でなくなります。そこで，分散の公式：

(分散)=(2 乗の平均)−(平均の 2 乗)

を用いました。

• 分散の公式を証明します。変量 x が x_1, x_2, …, x_n の n 個の値をとるとき，x の分散は

$$s_x{}^2=\frac{1}{n}\sum_{k=1}^{n}(x_k-\overline{x})^2 \quad \leftarrow\text{偏差平方の平均値}$$

$$=\frac{1}{n}\sum_{k=1}^{n}\{x_k{}^2-2\overline{x}x_k+(\overline{x})^2\}$$

$$=\underbrace{\frac{1}{n}\sum_{k=1}^{n}x_k{}^2}_{x^2\text{の平均値}}-2\overline{x}\cdot\frac{1}{n}\sum_{k=1}^{n}x_k+\frac{1}{n}\cdot n(\overline{x})^2$$

$$=\overline{x^2}-2\overline{x}\cdot\overline{x}+(\overline{x})^2$$

$$\therefore\quad s_x{}^2=\overline{x^2}-(\overline{x})^2$$

• 分散を，公式を用いず定義に従って求めると，次のようになります。

$$s_x{}^2=\frac{1}{6}\left\{\left(5-\frac{7}{2}\right)^2+\left(0-\frac{7}{2}\right)^2+\left(2-\frac{7}{2}\right)^2+\left(5-\frac{7}{2}\right)^2+\left(6-\frac{7}{2}\right)^2+\left(3-\frac{7}{2}\right)^2\right\}$$

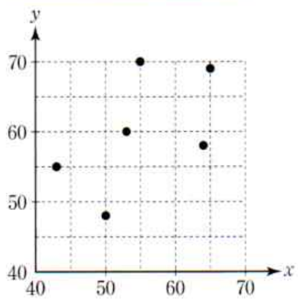
$$=\frac{1}{6\cdot4}(9+49+9+9+25+1)=\frac{102}{6\cdot4}=\frac{17}{4}$$

$$=4.25$$

本問においては，公式を用いる場合と所要時間はあまり変わりませんが，状況に応じて 2 つの方法を使い分けられるようにしましょう。

9 表と散布図を完成させると，次のようになる。

生徒番号	x	y	$x-\overline{x}$	$y-\overline{y}$	$(x-\overline{x})(y-\overline{y})$
1	43	55	−12	−5	60
2	50	48	−5	−12	60
3	53	60	−2	0	0
4	55	70	0	10	0
5	64	58	9	−2	−18
6	65	69	10	9	90
平均値	55	60	0	0	$\boxed{アイ}$



$$\boxed{アイ}=\frac{1}{6}(60+60-18+90)=\frac{192}{6}=\boxed{32}^{アイ}\text{(点)}$$

$s_x^2=(x-\bar{x})^2$ の平均値

$=\dfrac{1}{6}(144+25+4+81+100)=\dfrac{354}{6}=\boxed{59}$ ウエ

$s_y^2=(y-\bar{y})^2$ の平均値　　　　同じ値

$=\dfrac{1}{6}(25+144+100+4+81)=\boxed{59}$ オカ

x と y の共分散 s_{xy} とは，$(x-\bar{x})(y-\bar{y})$ の平均値に他ならないから

$s_{xy}=\boxed{アイ}=\boxed{32}$ キク

$\therefore\quad r_{xy}=\dfrac{s_{xy}}{s_x s_y}$

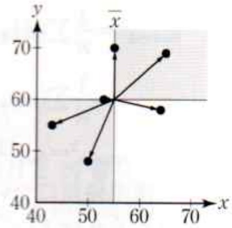
$=\dfrac{32}{\sqrt{59}\sqrt{59}}=\dfrac{32}{59}=0.542\cdots$

小数第 2 位までが問われているから，小数第 3 位を四捨五入して

$r_{xy}=\boxed{0}$ ケ $.\boxed{54}$ コサ

参考 •「相関係数」の意味について簡単に説明します。

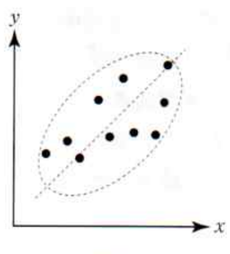
散布図の座標平面を，直線 $x=\bar{x}$，直線 $y=\bar{y}$ によって，右のように 4 つの領域に分けます。本問のデータでは，影をつけた領域に多くの点があり，



偏差の積 $\underbrace{(x-\bar{x})}\underbrace{(y-\bar{y})}$ の平均値は大きくなる。
同符号ゆえ積は正
x が大きいとき y も大きい傾向にある。

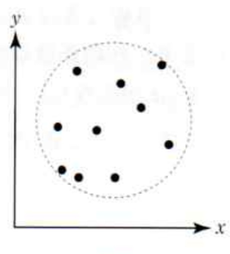
したがって，相関係数 r_{xy} の値が大きい（$+1$ に近い）ほど，x の増加にともない y も増加する傾向が強くなります。このことを，「x と y の間に正の相関がある」といいます。上記と逆のことを「負の相関がある」といいます。

10 相関係数は，散布図の点が右上がりの直線近くに並ぶほど 1 に近く，右下がりの直線近くに並ぶほど -1 に近くなる。



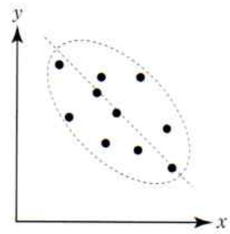
$r=0.7$

$\boxed{ア}=\boxed{⑤}$ ア



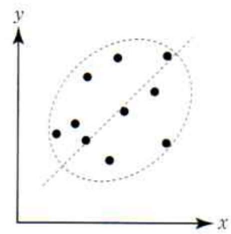
$r=0$

$\boxed{イ}=\boxed{③}$ イ



$r=-0.7$

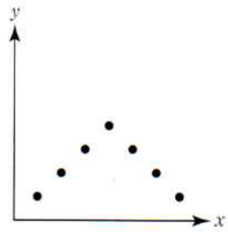
$\boxed{ウ}=\boxed{①}$ ウ



$r=0.3$

$\boxed{エ}=\boxed{④}$ エ

注意 •「相関係数」は，変量 x，y の散布図における直線的な関係の強さを表します。たとえば右の散布図において相関係数はほぼ 0 ですが，「x と y は無関係である」とは言えませんね。



• 変量 x と y の間に「相関関係がある」からといって，x と y の間に「因果関係がある」とは限りません。たとえば，別の変量 z が x，y 双方に影響を与え，その結果として x と y の間に相関が生まれた可能性も考えられますね。

参考 相関係数 r が $-1\leqq r\leqq 1$ を満たすことは，次のようにして証明されます。技巧を駆使しますから，観賞する気分でどうぞ。

変量 x，y の組として $(x_1,\ y_1)$，$(x_2,\ y_2)$，…，$(x_n,\ y_n)$ があるとき，x，y の偏差をそれぞれ $X_k=x_k-\bar{x}$，$Y_k=y_k-\bar{y}$ とおくと，x と y の相関係数 r_{xy} は

$$r_{xy}=\frac{s_{xy}}{s_x s_y}=\frac{\dfrac{1}{n}\displaystyle\sum_{k=1}^{n}X_kY_k}{s_x s_y}$$

（以下，「$\displaystyle\sum_{k=1}^{n}$」を「$\Sigma$」と記す）

$$\underset{\sim\sim\sim\sim\sim\sim\sim\sim}{\frac{1}{n}\Sigma\left(\frac{X_k}{s_x}\pm\frac{Y_k}{s_y}\right)^2}$$

$$=\frac{1}{s_x^{\,2}}\cdot\frac{1}{n}\Sigma X_k^{\,2}\pm 2\cdot\frac{\dfrac{1}{n}\Sigma X_kY_k}{s_x s_y}+\frac{1}{s_y^{\,2}}\cdot\frac{1}{n}\Sigma Y_k^{\,2}$$

$$=\frac{s_x^{\,2}}{s_x^{\,2}}\pm 2r_{xy}+\frac{s_y^{\,2}}{s_y^{\,2}}\text{（複号同順）}$$

〰〰 部分は 0 以上だから

$2\pm 2r_{xy}\geqq 0$　　よって　$-1\leqq r_{xy}\leqq 1$

また，$r_{xy}=\dfrac{\frac{1}{n}\sum X_kY_k}{\sqrt{\frac{1}{n}\sum X_k{}^2}\sqrt{\frac{1}{n}\sum Y_k{}^2}}=\dfrac{\sum X_kY_k}{\sqrt{\sum X_k{}^2}\sqrt{\sum Y_k{}^2}}$

と $r_{xy}{}^2\leqq 1$ より

$(\sum X_kY_k)^2\leqq(\sum X_k{}^2)(\sum Y_k{}^2)$

これは，「コーシー・シュワルツの不等式」と呼ばれる有名不等式の一般形です。

11 (p. 3 ❻の公式を用います。)

(1) $\overline{t}=\overline{5x+30}=5\overline{x}+30=5\cdot4+30=\boxed{50}$ アイ

$s_t{}^2=5^2\cdot s_x{}^2=5^2\times0.4=\boxed{10}$ ウエ

(2) $\boxed{\text{オ}}$ を a とおくと　$u=y-a$

$\overline{u}=\overline{y-a}=\overline{y}-a$

これと，$\overline{u}=0$，$\overline{y}=8.0$ より

$0=8-a$　$a=\boxed{8}$ オ

(3) $\dfrac{\sqrt{\boxed{\text{カ}}}}{\boxed{\text{キ}}}$ を $b(>0)$ とおくと　$v=by$

$s_v{}^2=b^2s_y{}^2$

これと，$s_v{}^2=s_x{}^2=0.4$，$s_y{}^2=2.0$ より

$0.4=b^2\cdot2$　$b^2=0.2=\dfrac{1}{5}$

$b=\dfrac{1}{\sqrt{5}}=\dfrac{\sqrt{\boxed{5}}}{\boxed{5}}$ カ キ

(4) 変量を変換しても相関係数は変化しないから

$r_{tu}=r_{xy}$　　よって　$\boxed{\text{ク}}=\boxed{0}$ ク

参考　• p. 3 ❻の公式のうち，分散に関する公式は覚えづらいかもしれませんね。直観的に説明するなら，次の通りです。

公式①：分散は偏差を2乗して得られるので，x を定数 a 倍した ax の分散は，もとの x の分散の a^2 倍になる。

公式②：分散は変量の平均値からの隔り（偏差）によって定まる。x に定数 b を加えて $x+b$ にしたとき，変量の各値と平均値はともに b ずつ増加するから，「隔り」は変化しない。よって，分散も変わらない。

• p. 3 ❻の公式を証明します。

x_1，x_2，…，x_n の n 個の値をとるときを考え，a，b は定数とする。「$\sum_{k=1}^{n}$」を「$\sum$」と書くと

$$\overline{ax}=\frac{1}{n}\sum ax_k=a\cdot\frac{1}{n}\sum x_k=a\overline{x}$$

$$\overline{x+b}=\frac{1}{n}\sum(x_k+b)=\frac{1}{n}\sum x_k+\frac{1}{n}\cdot nb=\overline{x}+b$$

$$\therefore\quad \overline{ax+b}=\overline{ax}+b=a\overline{x}+b$$

$$s_{ax}{}^2=\frac{1}{n}\sum(ax_k-\overline{ax})^2$$
$$=\frac{1}{n}\sum(ax_k-a\overline{x})^2=a^2\cdot\frac{1}{n}\sum(x_k-\overline{x})^2=a^2s_x{}^2$$

$$s_{x+b}{}^2=\frac{1}{n}\sum\{(x_k+b)-(\overline{x+b})\}^2$$
$$=\frac{1}{n}\sum\{(x_k+b)-(\overline{x}+b)\}^2=\frac{1}{n}\sum(x_k-\overline{x})^2$$
$$=s_x{}^2$$

$$s_{ax+b}{}^2=\frac{1}{n}\sum(ax_k+b-\overline{ax+b})^2$$
$$=\frac{1}{n}\sum(ax_k+b-a\overline{x}-b)^2=\frac{1}{n}\sum(ax_k-a\overline{x})^2$$

$$\therefore\quad s_{ax+b}{}^2=s_{ax}{}^2=a^2s_x{}^2$$

標準偏差については，上記の両辺に対して平方根をとればよい。

次に，2つの変量 x，y の組として $(x_1,\ y_1)$，$(x_2,\ y_2)$，…，$(x_n,\ y_n)$ があるとき，x を用いて変量 $z=ax+b$ を作り，$z_k=ax_k+b$ とすると，z と y の共分散は

$$s_{zy}=\frac{1}{n}\sum(z_k-\overline{z})(y_k-\overline{y})$$
$$=\frac{1}{n}\sum\{(ax_k+b)-(a\overline{x}+b)\}(y_k-\overline{y})$$
$$=a\cdot\frac{1}{n}\sum(x_k-\overline{x})(y_k-\overline{y})=as_{xy}$$

よって，z と y の相関係数は

$$r_{zy}=\frac{s_{zy}}{s_zs_y}=\frac{as_{xy}}{|a|s_x\cdot s_y}$$

$\therefore$　$a>0$ のとき　$r_{zy}=r_{xy}$

（$a<0$ のときは $r_{zy}=-r_{xy}$ となる。）

つまり，相関係数は変量を①～③のように変換しても（a が正ならば）変化しない。

12

階級 以上～未満	階級値	度数	相対度数
30～34	32	10	0.2
34～38	36	20	0.4
38～42	40	10	0.2
42～46	44	5	0.1
46～50	48	5	0.1
計		50	1.0

階級の幅は

$34-30=\boxed{4}^{ア}$(kg)

38－34 などでもよい。

握力 x の平均値は

$$\bar{x}=\frac{総和}{個数}$$

$$=\frac{1}{50}(32\cdot10+36\cdot20+40\cdot10+44\cdot5+48\cdot5)\cdots①$$

$$=\frac{1}{5}(32+72+40+22+24)$$

$$=\frac{190}{5}=\boxed{38}^{イウ}\text{(kg)}$$

$\frac{50}{2}$

x の値の小さい方から 25 番目と 26 番目はいずれも階級 34～38 に属するから，x の中央値は

$\tilde{x}=$階級 34～38 の階級値$=\boxed{36}^{エオ}$(kg)

度数が最大である階級は 34～38 だから，x の最頻値は，階級 34～38 の階級値 $\boxed{36}^{カキ}$ kg である。

参考 •①式を書きかえると，次のようになります。

$$\bar{x}=32\cdot\frac{10}{50}+36\cdot\frac{20}{50}+40\cdot\frac{10}{50}+44\cdot\frac{5}{50}+48\cdot\frac{5}{50}$$

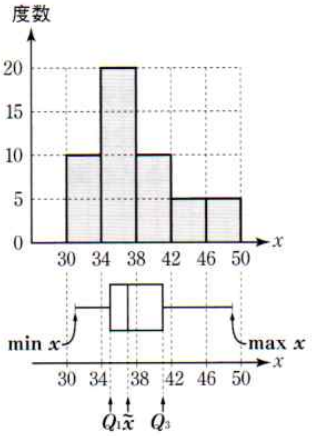
$$=32\times0.2+36\times0.4+40\times0.2+44\times0.1+48\times0.1$$

階級値×相対度数の総和

つまり，平均値は初めから相対度数を用いて求めることもできるわけです。ただし，「0.2」などの小数が現れるので得策とは限りませんが。

• 度数分布で与えられているデータの代表値(平均値，中央値，最頻値)は，各階級に属する変量の値が全てその階級の階級値をとると仮定して考えます。もちろん実際の変量の値はそれとは多少異なるのですが，このようにすることで計算処理が簡略化できますね。

13 このデータのヒストグラムは下図の通り。



また，このデータの「五数要約」を考えると

min$x$…階級 30～34 にある。

四分位数

- Q_1…小さい方から 12 番目と 13 番目を考えて，階級 34～38 にある。
- $\tilde{x}$…**12** より，階級 34～38 にある。
- Q_3…大きい方から，12 番目と 13 番目を考えて，階級 38～42 にある。

max$x$…階級 46～50 にある。

これらの条件を全て満たすものを選んで

$\boxed{ア}=\boxed{③}^{ア}$

注意 min$x$＝30 を満たしている⓪，①の中から選びたくなった人もいるかもしれませんね。しかし，データを度数分布にまとめたものからわかることは，最小値 min$x$ が「階級 30～34 のどこかにある」ことだけであり，その階級の下の限界：30(kg)が最小値になるとは限りません。

参考 図に示したように，ヒストグラムと箱ひげ図を対応づけて考えられるようにしておきましょう。

実戦編

1 やや難

着眼 平均値，分散，相関係数のいずれにおいても鍵となる量：「偏差」に注目しましょう。まず，問題文の表に偏差を書き込む欄を追加してしまいます！

解答 (1) 英語得点の偏差に注目し，その総和が0であることを利用する。→基本編 7

生徒番号	英語 x	数学 y	英語偏差 $x-\overline{x}$	数学偏差 $y-\overline{y}$
1	9	15	−7	0
2	20	20	4	5
3	18	14	2	−1
4	18	17	2	2
5	A	8	a	−7
6	18	C	2	c
7	14	D	−2	d
8	15	14	−1	−1
9	18	15	2	0
平均値	$\overline{x}=16$	$\overline{y}=15$	0	0

生徒5の英語の偏差を $a=\mathrm{A}-16$ とおくと

$-7+4+2+2+a+2-2-1+2=0$ より

$a=-2$ …①

$\therefore\ \ \mathrm{A}=16-2=\boxed{14}^{アイ}$(点)

英語得点 x の分散は

$$s_x{}^2=\frac{1}{9}(49+16+4+4+\underbrace{4}_{a^2}+4+4+1+4)\leftarrow\text{偏差平方の平均}$$

$$=\frac{90}{9}=\boxed{10}^{ウエ}.\boxed{00}^{オカ}$$

数学得点の偏差を考える。生徒6，生徒7の数学の偏差をそれぞれ $c=\mathrm{C}-15$，$d=\mathrm{D}-15$ とおくと，①と同様に，

$5-1+2-7+c+d-1=0$ より $c+d=2$

$(\mathrm{C}-15)+(\mathrm{D}-15)=2$

$\therefore\ \ \mathrm{C}+\mathrm{D}=\boxed{32}^{キク}$ …②

数学得点 y の分散を $s_y{}^2$，x と y の共分散を s_{xy}，x と y の相関係数を r_{xy} とすると，

$r_{xy}=\dfrac{s_{xy}}{s_xs_y}$ より $0.5=\dfrac{s_{xy}}{\sqrt{10}\sqrt{10}}$ $s_{xy}=5$

よって，英語と数学の偏差の積について

総和$=5\times9=45$ ←共分散$=\dfrac{\text{偏差の積の総和}}{\text{個数}}$ より

$\therefore\ \ \underbrace{20}_{生徒2}-2+4+\underbrace{14}_{-7a\,(a=-2)}+2c-2d+\underbrace{1}_{生徒8}=45$

$c-d=4$ $(\mathrm{C}-15)-(\mathrm{D}-15)=4$

$\therefore\ \ \mathrm{C}-\mathrm{D}=\boxed{4}^{ケ}$ …③

②，③より $\mathrm{C}=\boxed{18}^{コサ}$(点)，$\mathrm{D}=\boxed{14}^{シス}$(点)

(2) 散布図において，他の点から大きく隔った点に注目する。

数学の得点が最小の8点である生徒5は，英語が A＝14(点)であるから⓪か③に絞られる。そして，英語が14(＝A)である生徒5と生徒7に注目する。2人の数学の得点は8点と D＝14(点)であるから，⓪が正しい。よって

$\boxed{セ}=\boxed{⓪}^{セ}$

(3) 生徒10の得点は，英語，数学とも生徒1～9の9人の平均点に等しいから，生徒10が加入しても英語，数学の平均点は変化しない。

よって

$m=m'$ つまり $\dfrac{m'}{m}=1$ $\boxed{ソ}=\boxed{②}^{ソ}$

また，英語得点の偏差平方の総和 S は変化しないから

$$s=\sqrt{\frac{S}{9}},\ s'=\sqrt{\frac{S}{10}}$$

$$\therefore\ \ \frac{s'}{s}=\sqrt{\frac{S}{10}}\sqrt{\frac{9}{S}}=\frac{3}{\sqrt{10}}=\frac{3\sqrt{10}}{10}$$

$\boxed{タ}=\boxed{⑤}^{タ}$

数学得点の偏差平方の総和 T，数学と英語の偏差の積の総和 U も変化しないから

$$r=\frac{\dfrac{U}{9}}{\sqrt{\dfrac{S}{9}}\sqrt{\dfrac{T}{9}}}=\frac{U}{\sqrt{S}\sqrt{T}}$$

$$r'=\frac{\dfrac{U}{10}}{\sqrt{\dfrac{S}{10}}\sqrt{\dfrac{T}{10}}}=\frac{U}{\sqrt{S}\sqrt{T}}$$

$\therefore\ \ r=r'$ つまり $\dfrac{r'}{r}=1$

$\boxed{チ}=\boxed{②}^{チ}$

参考 •(1)で与えられた英語得点の平均値を用いて A を求める際，単純に得点の総和を求めるより，**解答**のように「偏差」に注目する方が計算が楽です。これは，与えられたデータから未知なる平均値を求める際に「仮平均」を用いるのとよく似た方法論ですね。

なお，「偏差の平均値が0」（基本編 7）と「偏差の総和が0」（本問）は，結局同じことを言っています。

• (3)の $\boxed{ソ}$ では，「平均値は変化しない」と直観的に判断してしまいましたが，このことを計算で示すと次の通りです。

生徒1，2，…，10の英語の得点をそれぞれ x_1, x_2, …, x_{10} とすると

$$m'=\frac{x_1+x_2+\cdots+x_9+x_{10}}{10}$$

$$=\frac{(x_1+x_2+\cdots+x_9)+x_{10}}{10}$$

$$=\frac{\overbrace{m\times 9}^{\text{生徒1～9の得点の総和}}+m}{10}=\frac{10m}{10}=m$$

• 逆に，(3)の $\boxed{タ}$，$\boxed{チ}$ を，次のように直観的に答えてしまうことも可能です。生徒10が加入しても，平均値，偏差平方の総和，偏差の積の総和に変化はない。変わるのは人数(データの大きさ)が9人→10人と増えることのみ。よって

標準偏差は $\sqrt{\frac{\bigcirc}{9}}\to\sqrt{\frac{\bigcirc}{10}}$ と $\frac{\sqrt{9}}{\sqrt{10}}$ 倍になり，

相関係数 $\dfrac{\frac{\square}{9}}{\sqrt{\frac{\bigcirc}{9}}\sqrt{\frac{\triangle}{9}}}$ は，データの大きさ「9」が約分されて消えるので，10人に増えても変化しない。

• 上で述べた相関係数の特性は，次のように一般化して覚えておく価値があります。

2つの変量 x, y の組として $(x_1,\ y_1)$, $(x_2,\ y_2)$, …, $(x_n,\ y_n)$ があるとき，x と y の相関係数 r_{xy} は $\left(\sum_{k=1}^{n} \text{を} \sum \text{と書くと}\right)$

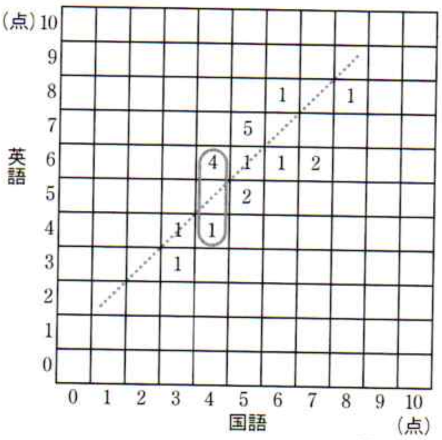
$$r_{xy}=\frac{\frac{1}{n}\sum(x_k-\bar{x})(y_k-\bar{y})}{\sqrt{\frac{1}{n}\sum(x_k-\bar{x})^2}\sqrt{\frac{1}{n}\sum(y_k-\bar{y})^2}}$$

$$=\frac{\overbrace{\sum(x_k-\bar{x})(y_k-\bar{y})}^{\text{偏差の積の総和}}}{\underbrace{\sqrt{\sum(x_k-\bar{x})^2}\sqrt{\sum(y_k-\bar{y})^2}}_{\text{偏差平方の総和}}}$$

2 やや難

着眼 散布図(相関図)と似た機能をもつ“相関表”が与えられています。この表の意味をしっかり読み取った上で，必要に応じて国語のみ，英語のみの分布も明示しておきましょう。

解答 (1)



国語の得点が4点である生徒数は，上図 ( ) の人数を求めて

$1+4=\boxed{5}^{ア}$(人)

英語の得点が国語の得点以下の生徒数は，上図 ⋰ より下の人数を考えて

$1+1+2+1+2+1=\boxed{8}^{イ}$(人)

(2) 国語の得点 x のみに注目してまとめると，次のようになる。

x	3	4	5	6	7	8	計
人数	2	5	8	2	2	1	20

よって，x の平均値は

$$B=\bar{x}=\frac{3\cdot2+4\cdot5+5\cdot8+6\cdot2+7\cdot2+8\cdot1}{20} \leftarrow \frac{\text{総和}}{\text{個数}}$$

$$=\frac{6+20+40+12+14+8}{20}=\frac{100}{20}$$

$$=\boxed{5}^{ウ}.\boxed{0}^{エ}\text{(点)}$$

英語の得点 y のみに注目してまとめると，次のようになる。

y	3	4	5	6	7	8	計
人数	1	2	2	8	5	2	20

よって y の分散は

$$s_y{}^2=\frac{1}{20}\{1\cdot(3-6)^2+2\cdot(4-6)^2+2\cdot(5-6)^2+5\cdot(7-6)^2+2\cdot(8-6)^2\}$$

$$=\frac{9+8+2+5+8}{20} \leftarrow \frac{\text{偏差平方の和}}{\text{個数}}$$

$$=\frac{32}{20}=\boxed{1}^{オ}.\boxed{60}^{カキ}$$

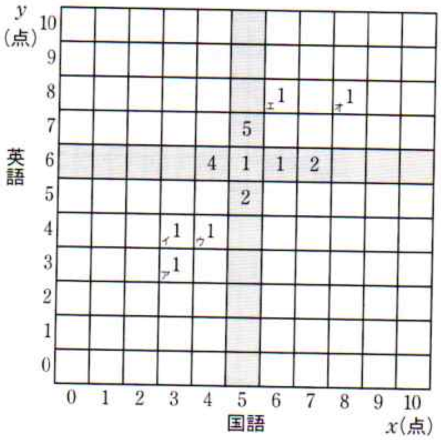
(3) 国語の得点が下から $\overset{\frac{20}{2}}{10}$ 番目の生徒と11番目の生徒はいずれも5点だから，国語の得点 x の中央値 $\tilde{x}$ は

$\tilde{x}=\boxed{5}^{ク}.\boxed{0}^{ケ}$(点)

英語の得点が下から5番目の生徒は5点，下から6番目の生徒は6点だから，英語の得点 y の第1四分位数は

$$\frac{5+6}{2}=\boxed{5}^{コ}.\boxed{5}^{サ}(点)$$

(4)



上図において，影の部分以外にある人数を数えて

$\boxed{シ}=\boxed{5}^{シ}$(人)

x と y の偏差の積 $(x-\overline{x})(y-\overline{y})$ は，図の影の部分にある生徒については全て0である。よって x と y の共分散は

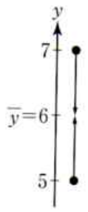
$$s_{xy}=\frac{\overset{ア}{6}+\overset{イ}{4}+\overset{ウ}{2}+\overset{エ}{2}+\overset{オ}{6}}{20}=1$$ ←上の表を参照

よって，x と y の相関係数は

$$r_{xy}=\frac{1}{\sqrt{1.6}\sqrt{1.6}}=\frac{1}{1.6}=\frac{5}{8}=\boxed{0}^{ス}.\boxed{625}^{セソタ}$$

(5) 英語 $\begin{cases} 7点：5人\to 4人 \\ 6点：1人\to 3人 \\ 5点：2人\to 1人 \end{cases}$

上の得点修正において，得点が変わったのは2人であり，右図のように変化した。



よって英語の得点合計に変化はないから，英語の得点の平均値は変わらない。

∴ $\boxed{チ}=\boxed{0}^{チ}$

得点が変わった2人について，その偏差平方は，いずれも

$1^2\to 0^2$

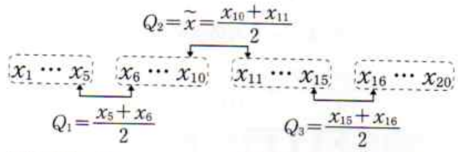
と減少したから，y の分散 s_y^2，すなわち偏差平方の平均値は減少する。すなわち

$\boxed{ツ}=\boxed{2}^{ツ}$

x と y の相関係数 $r_{xy}=\dfrac{s_{xy}}{s_x s_y}$ において，国語の得点変化はないから s_x は不変。上記より s_y は減少。

前記の2人について，国語の得点は $\overline{x}=6$ と一致しているので，x と y の偏差の積は得点修正の前，後いずれにおいても0である。したがって，x と y の共分散 s_{xy} も変化しない（符号は正）。以上より，r_{xy} は増加するから $\boxed{テ}=\boxed{1}^{テ}$

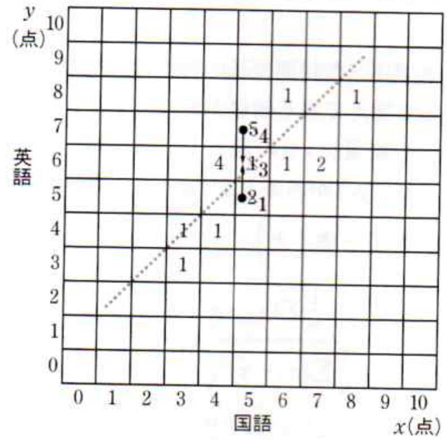
参考 •データの大きさが20であるとき，四分位数は次のようになります。（下記において，$x_1\leqq x_2\leqq\cdots\leqq x_{20}$ とする。）



•(4)の図では，x，y の平均値を表す影の部分によって右上，左上，右下の4つの領域に分けられ，そのうち右上の左下だけに生徒が分布しています。これらの生徒については，x と y の偏差の積は全て正となりますから，x と y の相関係数は正である（つまり正の相関がある）ことがわかります。

これぞセンター流 解法！

(5)の解答では，相関係数の各要素：s_x，s_y，s_{xy} の変化をきちんと調べましたが，もっと直観的，視覚的に答えることもできます。



もともとこの分布では，上図点線のまわりにデータがかなり集まっていたのですが，得点の修正によりさらにその傾向が強まったことがわかります。よって，相関係数は増加するはずです。

3 （標準）

着眼 データが散布図によって与えられていますが，前半の大半は x，y 各々について考えれば解答できます。(3)，(4)では平均値などの定義と意味

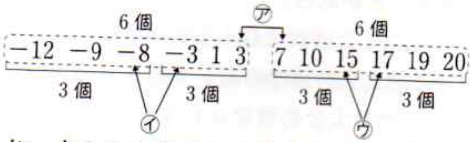
を理解しているかどうかが問われます。

解答 (1) x の平均値は

$$\frac{1}{12}(-12-9-3+3+10+17+20+19+15+7+1-8)$$

$$=\frac{60}{12}=\boxed{5}^{ア}.\boxed{0}^{イ}(℃)$$

x の値を小さい順に並べると，次の通り。



小さい方から 6 番目と 7 番目(上図の㋐)を用いて，x の中央値は

$$\tilde{x}=\frac{3+7}{2}=\boxed{5}^{ウ}.\boxed{0}^{エ}(℃)$$

次に四分位数を求める。下位 6 個のグループ，上位 6 個のグループそれぞれにおいて，小さい方から 3 番目と 4 番目(上図の㋑，㋒)を用いて，第 1 四分位数 Q_1，第 3 四分位数 Q_3 は

$$Q_1=\frac{-8-3}{2}=-5.5(℃),\ Q_3=\frac{15+17}{2}=16(℃)$$

変量 x の五数要約は

$\min x=-12,\quad Q_1=-5.5\quad \tilde{x}=5.0\quad Q_3=16$

$\max x=20$

これに適する箱ひげ図を選んで

$\boxed{オ}=\boxed{0}^{オ}$

(2) A グループにおける y の和は

$6.0\times4=24(℃)$

B グループにおける y の和は

$21.5\times8=172(℃)$

よって，1 月から 12 月までの y の平均値は

$$\bar{y}=\frac{24+172}{12}\ \leftarrow\frac{総和}{個数}$$

$$=\frac{196}{12}=\frac{49}{3}=16.333\cdots$$

小数第 2 位を四捨五入して

$$\bar{y}=\boxed{16}^{カキ}.\boxed{3}^{ク}(℃)$$

(3) y の値の 1 つを 30 ℃から 18 ℃に変えると，y の総和は $30-18=12$(℃)減少する。

よって，y の平均値は

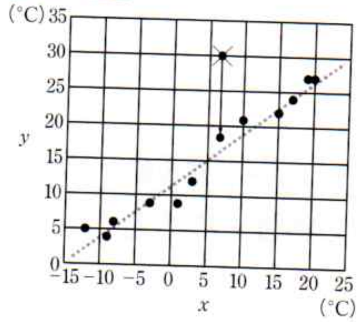
$$\frac{12}{12}=\boxed{1}^{ケ}.\boxed{0}^{コ}(℃)減少する。$$

修正前 $\bar{y}=16.3$(℃)から大きく隔っていた 30(℃)が，修正後には $\bar{y}=16.3-1.0=15.3$(℃)に近い 18 ℃になったので，その分偏差平方は減少する。他の月についての偏差平方の合計はあまり変わらないから，12 か月分の平方偏差の総和は減少する。よって，y の分散も減少する。

$\therefore\ \boxed{サ}=\boxed{1}^{サ}$

$(x,\ y)=(7,\ 30)$ だった点を $(x,\ y)=(7,\ 18)$ に移すと，散布図上の点は，移動前よりもある直線の近くに集まる。よって，x と y の相関係数は増加して 1 に近づく。

$\therefore\ \boxed{ス}=\boxed{0}^{シ}$



(4) $X=\frac{9}{5}x+32$ より，X と x の平均値に関して，

$$\bar{X}=\frac{9}{5}\bar{x}+32=\frac{9}{5}\times5.0+32=\boxed{41}^{スセ}.\boxed{0}^{ソ}(°F)$$

標準偏差については

$$s_X=\frac{9}{5}s_x\qquad\therefore\ \frac{s_X}{s_x}=\frac{9}{5}=\boxed{1}^{タ}.\boxed{8}^{チ}$$

変数変換をしても相関係数は変わらないから

$$r_{XY}=r_{xy}\qquad つまり\ \frac{r_{XY}}{r_{xy}}=\boxed{1}^{ツ}.\boxed{0}^{テ}$$

参考 • (1)で x の平均値を求める際に，仮平均は使いませんでした。値の符号が正，負にまたがっているのでうまく消し合ってくれることが予想されたからです。見方によっては「0」を仮平均として選んでいると言えなくもありません。

• (2)で y の平均値を求める式は，次のように書くこともできます。

$$\bar{y}=\frac{6.0\times4+21.5\times8}{12}$$

$$=6.0\times\frac{1}{3}+21.5\times\frac{2}{3}$$

これは 4 個の平均値 6.0 と，8 個の平均値 21.5 を，それぞれの個数に比例した "重み" $\frac{1}{3}$，$\frac{2}{3}$ を掛けて加えたもので，「加重平均」と言われたりします。

• (4)で用いた変数変換時の公式は，(できれば，理由を理解した上で)記憶してしまいましょう。