数学演習（データの分析）Ｈ３０．１０．２５　　　　(　　)組(　　)番　名前(　　　　　　　　　　)

1．生徒 7 人のテストの得点が次のようになった。

57，46，37，38，54，48，35　(点)

⑴　このデータの平均値，中央値，分散を求めよ。

⑵　生徒全員の得点に 10 点ずつ足したときの平均値および分散を求めよ。

2．次の 10 個の値からなるデータの平均値が 5，分散が 5 であるとき，a，b の値を求めよ。ただし，$a \leqq b$ とする。

5，2，3，8，6，3，7，6，a，b

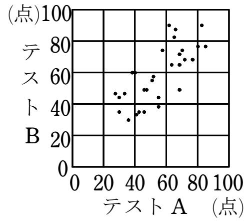
3．40 人の生徒の身長を測定して，右のような度数分布表を得た。この度数分布表から中央値を求めよ。

身　長(cm)	人数(人)
155 以上 ～ 160 未満	2
160　　～ 165	11
165　　～ 170	14
170　　～ 175	10
175　　～ 180	2
180　　～ 185	1
計	40

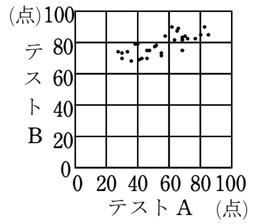
4．30 人の生徒に 100 点満点の 2 種類のテスト A，B を行ったところ，右のような結果となった。テストA とテスト B の点数の散布図として適当なものを下の ①～③から選べ。

	最高得点	標準偏差
テストA	85	16.4
テストB	90	6.5

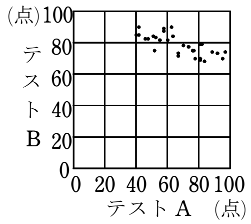
①



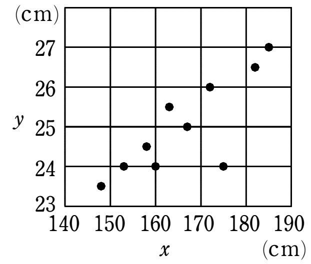
②



③



5．ある生徒 10 人の「身長」と「靴のサイズ」を，それぞれ変量 x，変量 y とする。右の図は，変量 x と変量 y の散布図である。



⑴　変量 x が 175 cm，変量 y が 24 cmとなっている生徒の変量 y の値は誤りであることがわかり，正しい値である 26 cmに修正した。修正前，修正後の変量 y の中央値をそれぞれ求めよ。ただし，変量 y の値は 0.5 cm 単位とする。

⑵　⑴のとき，修正前の x と y の相関係数を r_1，修正後の x と y の相関係数を r_2 とする。値の組 (r_1，r_2) として正しいものを，次の ①～④ から選べ。

①　(0.82，0.96)　②　(0.96，0.82)　③　(-0.82，-0.96)　④　(-0.96，-0.82)

6．次のデータは，高校 1 年生 6 人の 1 年間の身長の伸びを記録したものである。

20，28，19，24，21，26 (単位は mm)

⑴　このデータの平均値を求めよ。

⑵　このデータの一部に誤りがあり，28 mm は正しくは 26 mm，19 mm は正しくは 20 mm，21 mm は正しくは 22 mm であった。この誤りを修正したとき，このデータの平均値，分散は修正前と比べて増加するか，減少するか，変化しないかを答えよ。

7．a，b，c は異なる 3 つの正の整数とする。次のデータは 2 つの科目 X と Y の試験を受けた 10 人の得点をまとめたものである。

	①	②	③	④	⑤	⑥	⑦	⑧	⑨	⑩
科目 X の得点	a	c	a	b	b	a	c	c	b	c
科目 Y の得点	a	b	b	b	a	a	b	a	b	a

科目 X の得点の平均値と科目 Y の得点の平均値とは等しいとする。

(1)　科目 X の得点の分散を ${s_X}^2$，科目 Y の得点の分散を ${s_Y}^2$ とする。$\dfrac{{s_X}^2}{{s_Y}^2}$ を求めよ。

(2)　科目 X の得点と科目 Y の得点の相関係数を，四捨五入して小数第 1 位まで求めよ。

(3)　科目 X の得点の中央値が 65，科目 Y の得点の標準偏差が 11 であるとき，a，b，c の組を求めよ。

8．変量 x についてのデータの値が，n 個の値 x_1，x_2，……，x_n であるとし，そのデータの平均値を $\overline{x}$ とする。t の関数 $g(t)$ を

$$g(t)=(t-x_1)^2+(t-x_2)^2+\cdots\cdots+(t-x_n)^2$$

で定義する。このとき，$g(t)$ は $t=\overline{x}$ で最小となることを示せ。

9．変量 x についてのデータの値が，n 個の値 x_1，x_2，……，x_n であるとし，そのデータの中央値を M とする。t の関数 $f(t)$ を

$$f(t)=|t-x_1|+|t-x_2|+\cdots\cdots+|t-x_n|$$

で定義するとき，$f(t)$ の最小値は $f(M)$ であることを示せ。

1．生徒 7 人のテストの得点が次のようになった。

57，46，37，38，54，48，35（点）

(1)　このデータの平均値，中央値，分散を求めよ。

(2)　生徒全員の得点に 10 点ずつ足したときの平均値および分散を求めよ。

解答　(1)　平均値 45 点，中央値 46 点，分散 64

(2)　平均値 55点，分散 64

解説

データを値の大きさの順に並べると，次のようになる。

35，37，38，46，48，54，57

(1)　平均値　$\frac{1}{7}(35+37+38+46+48+54+57)=\frac{1}{7}\times 315=45$（点）

中央値　46（点）

データの値と偏差の 2 乗の値は，次の表のようになる。

x	35	37	38	46	48	54	57	計 315
$(x-\overline{x})^2$	100	64	49	1	9	81	144	計 448

したがって，分散は　$\frac{1}{7}\times 448=64$

(2)　全員の得点に 10 点ずつ足したとき，データの値の総和は 10×7（点）増加する。

よって，平均値は　$\frac{1}{7}\times(315+10\times 7)=\frac{1}{7}\times 385=55$（点）

全員の得点に 10 点ずつ足したとき，データの各値も平均値も 10 点だけ増加するから，偏差はもとのデータと変わらない。

したがって，分散は　64

2．次の 10 個の値からなるデータの平均値が 5，分散が 5 であるとき，a，b の値を求めよ。ただし，$a\leqq b$ とする。　[20点]

5，2，3，8，6，3，7，6，a，b

解答　データの平均値が 5 であるから

$$\frac{1}{10}(5+2+3+8+6+3+7+6+a+b)=5$$

整理すると　$a+b=10$　……①

データの分散が 5 であるから

$$\frac{1}{10}(5^2+2^2+3^2+8^2+6^2+3^2+7^2+6^2+a^2+b^2)-5^2=5$$

整理すると　$a^2+b^2=68$　……②

① から　$b=10-a$

これを ② に代入して整理すると　$a^2-10a+16=0$

これを解いて　$a=2,\ 8$

よって，① と $a\leqq b$ から　$a=2,\ b=8$

解説

データの平均値が 5 であるから

$$\frac{1}{10}(5+2+3+8+6+3+7+6+a+b)=5$$

整理すると　$a+b=10$　……①

データの分散が 5 であるから

$$\frac{1}{10}(5^2+2^2+3^2+8^2+6^2+3^2+7^2+6^2+a^2+b^2)-5^2=5$$

整理すると　$a^2+b^2=68$　……②

① から　$b=10-a$

これを ② に代入して整理すると　$a^2-10a+16=0$

これを解いて　$a=2,\ 8$

よって，① と $a\leqq b$ から　$a=2,\ b=8$

3．40 人の生徒の身長を測定して，右のような度数分布表を得た。この度数分布表から中央値を求めよ。

身　長(cm)	人数(人)
155 以上 ～ 160 未満	2
160 ～ 165	11
165 ～ 170	14
170 ～ 175	10
175 ～ 180	2
180 ～ 185	1
計	40

解答　167.5 cm

解説

累積度数分布表，累積相対度数分布表は右のようになる。中央値を x とする。

データの総数は 40 であるから，中央値 x は階級 165 ～ 170 に入っており

$$\frac{x-165}{\frac{1}{2}-\frac{13}{40}}=\frac{170-165}{\frac{27}{40}-\frac{13}{40}}$$

よって

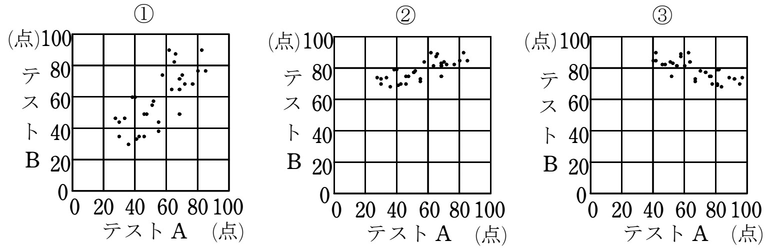
$$x=165+(170-165)\times\frac{7}{14}=167.5\ (\mathrm{cm})$$

階級（cm）	累積度数	累積相対度数
160 未満	2	0.050
165	13	0.325
170	27	0.675
175	37	0.925
180	39	0.975
185	40	1.000

4．30 人の生徒に 100 点満点の 2 種類のテスト A，B を行ったところ，右のような結果となった。テスト A とテスト B の点数の散布図として適当なものを下の ①～③から選べ。　[20点]

	最高得点	標準偏差
テストA	85	16.4
テストB	90	6.5

①　②　③



解答　テスト A の点数の標準偏差が 16.4 点であるのに対して，テスト B の点数の標準偏差は 6.5 点であるから，散布図は，横方向の広がりに比べて縦方向にはあまり広がらない分布となる。この条件を満たすものは　②，③

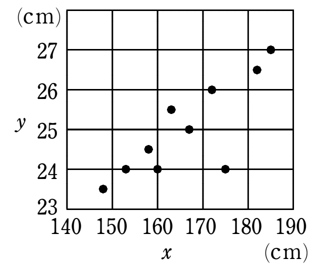
テスト A の最高点が 85 点であるから，②，③ のうち，これを満たすのは ②。

解説

テスト A の点数の標準偏差が 16.4 点であるのに対して，テスト B の点数の標準偏差は 6.5 点であるから，散布図は，横方向の広がりに比べて縦方向にはあまり広がらない分布となる。この条件を満たすものは　②，③

テスト A の最高点が 85 点であるから，②，③ のうち，これを満たすのは ②。

5．ある生徒 10 人の「身長」と「靴のサイズ」を，それぞれ変量 x，変量 y とする。右の図は，変量 x と変量 y の散布図である。　[15点×2＝30点]



(1)　変量 x が 175 cm，変量 y が 24 cmとなっている生徒の変量 y の値は誤りであることがわかり，正しい値である 26 cmに修正した。修正前，修正後の変量 y の中央値をそれぞれ求めよ。ただし，変量 y の値は 0.5 cm 単位とする。

(2)　(1) のとき，修正前の x と y の相関係数を r_1，修正後の x と y の相関係数を r_2 とする。値の組 $(r_1,\ r_2)$ として正しいものを，次の ①～④ から選べ。

①　(0.82，0.96)　②　(0.96，0.82)　③　(−0.82，−0.96)　④　(−0.96，−0.82)

解答　(1)　修正前の中央値は　$\frac{24.5+25}{2}=24.75$ (cm)

修正後の中央値は　$\frac{25+25.5}{2}=25.25$ (cm)

(2)　修正前，修正後とも，散布図から正の相関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

解説

(1)　修正前の中央値は　$\frac{24.5+25}{2}=24.75$ (cm)

修正後の中央値は　$\frac{25+25.5}{2}=25.25$ (cm)

(2)　修正前，修正後とも，散布図から正の相関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

6．次のデータは，高校 1 年生 6 人の 1 年間の身長の伸びを記録したものである。

20，28，19，24，21，26（単位は mm）

(1)　このデータの平均値を求めよ。

(2)　このデータの一部に誤りがあり，28 mm は正しくは 26 mm，19 mm は正しくは 20 mm，21 mm は正しくは 22 mm であった。この誤りを修正したとき，このデータの平均値，分散は修正前と比べて増加するか，減少するか，変化しないかを答えよ。

解答　(1)　23 mm

(2)　平均値は修正前と比べて変化しない，分散は修正前と比べて減少する

解説

(1)　このデータの平均値を $\overline{x}$ とすると

$$\overline{x}=\frac{1}{6}(20+28+19+24+21+26)=\frac{138}{6}=23\ (\mathrm{mm})$$

(2)　データの修正によって，2 mm 減少するものが 1 つと，1 mm 増加するものが 2 つある。

よって，データの総和は変化しないから，平均値は修正前と比べて変化しない。

また，修正した 3 つのデータの偏差の 2 乗の和について

修正前は　$(28-23)^2+(19-23)^2+(21-23)^2=45$

修正後は　$(26-23)^2+(20-23)^2+(22-23)^2=19$

ゆえに，偏差の 2 乗の総和は減少するから，分散は修正前と比べて減少する。

7．a，b，c は異なる 3 つの正の整数とする。次のデータは 2 つの科目 X と Y の試験を受けた 10 人の得点をまとめたものである。

	①	②	③	④	⑤	⑥	⑦	⑧	⑨	⑩
科目Xの得点	a	c	a	b	b	a	c	c	b	c
科目Yの得点	a	b	b	b	a	a	b	a	b	a

科目Xの得点の平均値と科目Yの得点の平均値とは等しいとする。

(1) 科目Xの得点の分散を $s_X{}^2$，科目Yの得点の分散を $s_Y{}^2$ とする。$\dfrac{s_X{}^2}{s_Y{}^2}$ を求めよ。

(2) 科目Xの得点と科目Yの得点の相関係数を，四捨五入して小数第1位まで求めよ。

(3) 科目Xの得点の中央値が65，科目Yの得点の標準偏差が11であるとき，a，b，c の組を求めよ。

解答 (1) $\dfrac{3}{5}$　(2) 0.3　(3) $(a,\ b,\ c)=(76,\ 54,\ 65)$，$(54,\ 76,\ 65)$

解説

(1) Xの得点の平均値とYの得点の平均値が等しいから

$$3a+3b+4c=5a+5b$$

よって　$c=\dfrac{a+b}{2}$

また，Yの得点の平均値は　$\dfrac{5a+5b}{10}=\dfrac{a+b}{2}$

ゆえに　$s_X{}^2=\dfrac{3a^2+3b^2+4c^2}{10}-\left(\dfrac{a+b}{2}\right)^2=\dfrac{4a^2+2ab+4b^2}{10}-\dfrac{a^2+2ab+b^2}{4}$

$=\dfrac{3}{20}(a-b)^2$

$s_Y{}^2=\dfrac{5a^2+5b^2}{10}-\left(\dfrac{a+b}{2}\right)^2=\dfrac{a^2+b^2}{2}-\dfrac{a^2+2ab+b^2}{4}$

$=\dfrac{1}{4}(a-b)^2$

よって　$\dfrac{s_X{}^2}{s_Y{}^2}=\dfrac{3}{5}$

参考　$s_X{}^2=\dfrac{3a^2+3b^2+4c^2}{10}-c^2=\dfrac{3}{10}(a^2+b^2-2c^2)$，

$s_Y{}^2=\dfrac{5a^2+5b^2}{10}-c^2=\dfrac{5}{10}(a^2+b^2-2c^2)$ として $\dfrac{s_X{}^2}{s_Y{}^2}$ を求めてもよい。

(2) Xの得点とYの得点の共分散を s_{XY} とする。

ここで　$\left(a-\dfrac{a+b}{2}\right)\left(a-\dfrac{a+b}{2}\right)=\dfrac{1}{4}(a-b)^2$

$\left(a-\dfrac{a+b}{2}\right)\left(b-\dfrac{a+b}{2}\right)=-\dfrac{1}{4}(a-b)^2$

$\left(b-\dfrac{a+b}{2}\right)\left(b-\dfrac{a+b}{2}\right)=\dfrac{1}{4}(a-b)^2$

ゆえに，Xの得点とYの得点の相関係数は

$$\frac{s_{XY}}{s_Xs_Y}=\frac{\dfrac{1}{10}\times\dfrac{1}{4}(a-b)^2\times(4-2)}{\sqrt{\dfrac{3}{20}(a-b)^2}\sqrt{\dfrac{1}{4}(a-b)^2}}=\frac{\dfrac{1}{20}}{\sqrt{\dfrac{3}{80}}}=\frac{\sqrt{15}}{15}$$

$3.8\times3.8=14.44<15<15.21=3.9\times3.9$ であるから

$$\frac{3.8}{15}<\frac{\sqrt{15}}{15}<\frac{3.9}{15}$$

$\dfrac{3.8}{15}=0.25\cdots$，$\dfrac{3.9}{15}=0.26$ であるから，求める相関係数は　0.3

(3) $c=\dfrac{a+b}{2}$ より，a，b，c の大小関係は

$a<c<b$　または　$b<c<a$

であるから，Xの得点の中央値は c である。

Xの得点の中央値が65であるから　$c=65$

よって　$\dfrac{a+b}{2}=65$　ゆえに　$b=130-a$

Yの得点の標準偏差が11であるから

$\dfrac{1}{4}(a-b)^2=11^2$　すなわち　$(a-b)^2=4\cdot11^2$

$b=130-a$ を代入して　$(2a-130)^2=22^2$

よって　$2a-130=\pm22$　ゆえに　$a=76,\ 54$

$b=130-a$ であるから，a，b，c の組は

$(a,\ b,\ c)=(76,\ 54,\ 65)$，$(54,\ 76,\ 65)$

8．変量 x についてのデータの値が，n 個の値 x_1，x_2，……，x_n であるとし，そのデータの平均値を $\overline{x}$ とする。t の関数 $g(t)$ を

$$g(t)=(t-x_1)^2+(t-x_2)^2+\cdots\cdots+(t-x_n)^2$$

で定義する。このとき，$g(t)$ は $t=\overline{x}$ で最小となることを示せ。

解答 略

解説

$g(t)=(t-x_1)^2+(t-x_2)^2+\cdots\cdots+(t-x_n)^2$

$=(t^2-2x_1t+x_1{}^2)+(t^2-2x_2t+x_2{}^2)+\cdots\cdots+(t^2-2x_nt+x_n{}^2)$

$=nt^2-2(x_1+x_2+\cdots\cdots+x_n)t+(x_1{}^2+x_2{}^2+\cdots\cdots+x_n{}^2)$

$=n\left(t^2-2\cdot\dfrac{x_1+x_2+\cdots\cdots+x_n}{n}t\right)+(x_1{}^2+x_2{}^2+\cdots\cdots+x_n{}^2)$

ここで，$\overline{x}=\dfrac{x_1+x_2+\cdots\cdots+x_n}{n}$ であるから

$g(t)=n(t^2-2\overline{x}t)+(x_1{}^2+x_2{}^2+\cdots\cdots+x_n{}^2)$

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

$=n\left\{(t-\overline{x})^2+\dfrac{x_1{}^2+x_2{}^2+\cdots\cdots+x_n{}^2}{n}-(\overline{x})^2\right\}$

したがって，$g(t)$ は $t=\overline{x}$ で最小となり，最小値

$g(\overline{x})=n\left\{\dfrac{x_1{}^2+x_2{}^2+\cdots\cdots+x_n{}^2}{n}-(\overline{x})^2\right\}$

をとる。

9．変量 x についてのデータの値が，n 個の値 x_1，x_2，……，x_n であるとし，そのデータの中央値を M とする。t の関数 $f(t)$ を

$$f(t)=|t-x_1|+|t-x_2|+\cdots\cdots+|t-x_n|$$

で定義するとき，$f(t)$ の最小値は $f(M)$ であることを示せ。

解答 略

解説

データ x_1，x_2，……，x_n を大きさの順(小さい順)に並べ替えたものを a_1，a_2，……，a_n とする。このとき　$a_1\leqq a_2\leqq\cdots\cdots\leqq a_n$

中央値 M について，n が奇数のとき　$M=a_{\frac{n+1}{2}}$，

n が偶数のとき　$M=\dfrac{1}{2}\left(a_{\frac{n}{2}}+a_{\frac{n}{2}+1}\right)$

$|t-a_k|=\begin{cases}t-a_k & (t\geqq a_k)\\ -(t-a_k) & (t<a_k)\end{cases}$　$(k=1,\ 2,\ \cdots\cdots,\ n)$ であるから

[1] $t<a_1$ のとき

$f(t)=|t-a_1|+|t-a_2|+\cdots\cdots+|t-a_n|=-(t-a_1)-(t-a_2)-\cdots\cdots-(t-a_n)$

$=-nt+(a_1+a_2+\cdots\cdots+a_n)$

[2] $a_k\leqq t<a_{k+1}$ $(k=1,\ 2,\ \cdots\cdots,\ n-1)$ のとき

$f(t)=|t-a_1|+|t-a_2|+\cdots\cdots+|t-a_k|+|t-a_{k+1}|+\cdots\cdots+|t-a_n|$

$=(t-a_1)+(t-a_2)+\cdots\cdots+(t-a_k)-(t-a_{k+1})-\cdots\cdots-(t-a_n)$

$=\{kt-(a_1+a_2+\cdots\cdots+a_k)\}-\{(n-k)t-(a_{k+1}+\cdots\cdots+a_n)\}$

$=(-n+2k)t-(a_1+a_2+\cdots\cdots+a_k)+(a_{k+1}+\cdots\cdots+a_n)$

[3] $t\geqq a_n$ のとき

$f(t)=|t-a_1|+|t-a_2|+\cdots\cdots+|t-a_n|=(t-a_1)+(t-a_2)+\cdots\cdots+(t-a_n)$

$=nt-(a_1+a_2+\cdots\cdots+a_n)$

したがって，$y=f(t)$ のグラフは折れ線となる。

[A] n が奇数のとき，$k=\dfrac{n-1}{2}$ とすると，

$-n+2k=-1<0$ となるから

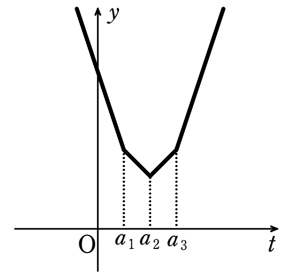
$t\leqq a_{\frac{n-1}{2}+1}$　すなわち　$t\leqq a_{\frac{n+1}{2}}$ のとき，$f(t)$ は減少する。

また，$t\geqq a_{\frac{n+1}{2}}$ のとき，$-n+2\cdot\dfrac{n+1}{2}=1>0$ であるから，$f(t)$ は増加する。

したがって，$t=a_{\frac{n+1}{2}}$ のとき $f(t)$ は最小値をとる。

$a_{\frac{n+1}{2}}=M$ であるから，$f(M)$ は最小値である。

$n=3$ のとき



[B] n が偶数のとき，$k=\dfrac{n}{2}$ とすると，$-n+2k=0$ であるから，$a_{\frac{n}{2}}\leqq t\leqq a_{\frac{n}{2}+1}$ のとき $f(t)$ は定数となる。

また，$t\leqq a_{\frac{n}{2}}$ のとき $f(t)$ は減少し，$t\geqq a_{\frac{n}{2}+1}$ のとき $f(t)$ は増加するから，$a_{\frac{n}{2}}\leqq t\leqq a_{\frac{n}{2}+1}$ の範囲の任意の t で $f(t)$ は最小値をとる。

ここで，$M=\dfrac{1}{2}\left(a_{\frac{n}{2}}+a_{\frac{n}{2}+1}\right)$ から　$a_{\frac{n}{2}}\leqq M\leqq a_{\frac{n}{2}+1}$

したがって，$f(M)$ は最小値である。

$n=4$ のとき