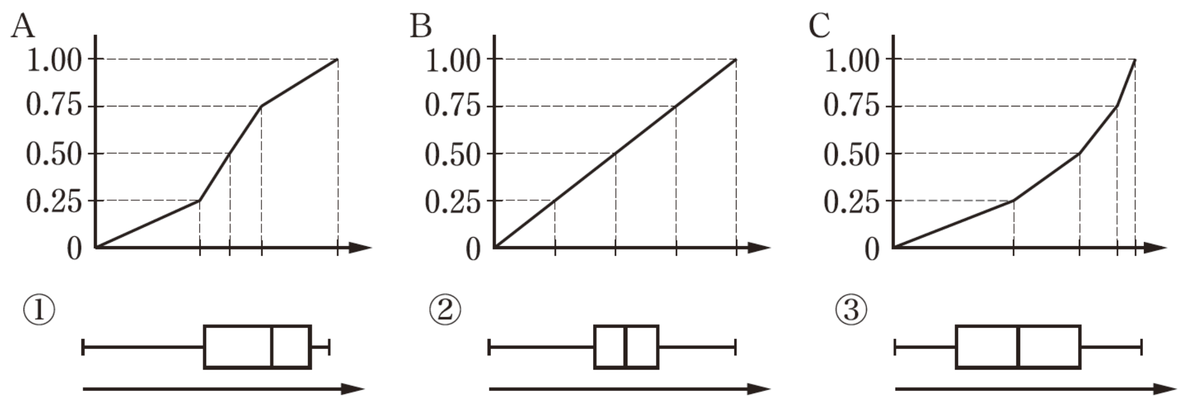
練習問題 A

（教科書 p.176）

1　次の累積相対度数折れ線 A，B，C について，それぞれ対応する箱ひげ図を①，②，③の中から選べ。



2　20 人の生徒に，数学と国語の 5 点満点の小テストを行った。数学の得点を x 点，国語の得点を y 点とする。そのときの結果が次の表である。たとえば，数学が 3 点，国語が 4 点の生徒は 7 人いることがわかる。このとき，次の問に答えよ。

y \ x	0	1	2	3	4	5	計
5				1		3	4
4			3	7	1	1	12
3		3		1			4
2							0
1							0
0							0
計	0	3	3	9	1	4	20

(1)　x の平均値 $\overline{x}$ と，y の平均値 $\overline{y}$ を求めよ。

(2)　x と y の分散は，右の表のようになる。x と y との相関係数 r と値が一致する式を①～③の中からすべて選べ。

	数学	国語
分散	1.6	0.4

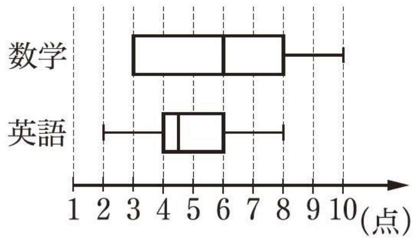
①　$\dfrac{3(5-\overline{x})(3-\overline{y})}{20\sqrt{1.6}\sqrt{0.4}}$　　②　$\dfrac{6(5-\overline{x})(5-\overline{y})}{20\sqrt{1.6}\sqrt{0.4}}$　　③　$\dfrac{3(1-\overline{x})(3-\overline{y})}{\sqrt{16}\sqrt{4}}$

練習問題Ｂ

（教科書p.177）

3 10 人の生徒に，数学と英語の 10 点満点の小テストを行った。数学の得点を x 点，英語の得点を y 点とする。そのときの結果が下の表と箱ひげ図である。また，x，y の平均値を $\overline{x}$，$\overline{y}$ とする。

生徒番号	x	y	$(x-\overline{x})^2$	$(y-\overline{y})^2$	$(x-\overline{x})(y-\overline{y})$
1	8	6	4	1	2
2	6	4	0	1	0
⋮	⋮	⋮	⋮	⋮	⋮
10	3	3	9	4	6
合計	A	B	64	36	43

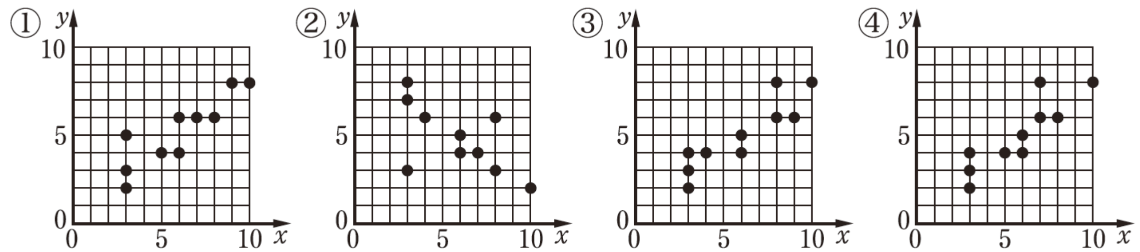


(1)　表の A，B の値を求めよ。

(2)　x の分散 $s_x{}^2$ と，y の分散 $s_y{}^2$ を求めよ。

(3)　x と y の相関係数 r を小数第 3 位を四捨五入して，小数第 2 位まで求めよ。

(4)　x と y の散布図として，適当なものを次の①～④の中から選べ。



(5)　11 人目の生徒が，同じ数学と英語の小テストを受けた。この生徒の得点は，数学が 6 点，英語が 5 点であった。次の記述のうち，適当でないものをすべて選べ。

①　11 人の数学の平均値は，10 人のときの平均値 $\overline{x}$ と等しい。

②　11 人の英語の分散は，10 人のときの分散 $s_y{}^2$ より小さい。

③　11 人の数学の得点と英語の得点の相関係数は，10 人のときの相関係数 r より大きい。

参 考　はずれ値

（教科書 p.178）

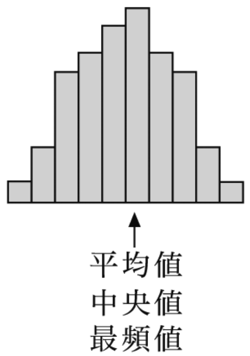
データには，他と比べて値が大きくなる「**はずれ値**」といわれるものが含まれる場合がある。

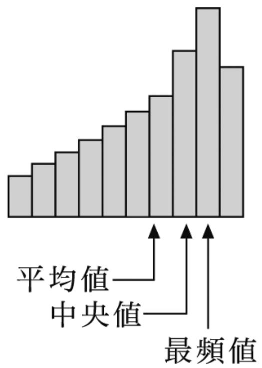
たとえば，ある5人のテスト結果が 15 点，13 点，12 点，16 点，98 点のとき，平均値は 30.8 点，中央値は 15 点である。

98 点は他の点数より著しく高いため，98 点の値を「はずれ値」として除いて考えると，平均値は 14 点と大きく変化するが，中央値は 14 点とあまり変化しない。

このことから，平均値のほうが中央値よりも「はずれ値」からの影響を受けやすいといえる。

また，平均値，中央値，最頻値は，右の図のように，データが 1 つの山でほぼ左右対称に分布しているとき，互いに近い値となる。



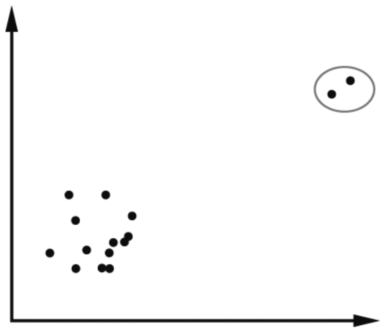


一方，データが 1 つの山で左右対称に分布していないときは，平均値よりも中央値や最頻値のほうが，データの代表的な位置を表すのに適切である。

さらに，相関係数も「はずれ値」の影響を受けやすい。

右の散布図で表されたデータの相関係数は 0.885 であるが，円の中の 2 つの「はずれ値」を除いて相関係数を求めると 0.023 となる。

このため，相関を考えるときは相関係数だけではなく，散布図も調べることが重要である。



発 展　$y=ax+b$ における平均値と分散・標準偏差

（教科書 p.179）

2 つの変数 x と y の間に，a，b を定数として $y=ax+b$ という関係があるとき，x の平均値 $\overline{x}$ と y の平均値 $\overline{y}$ の関係について調べてみよう。

変量 x の n 個の値を x_1，x_2，⋯，x_nとし，変量 y の n 個の値をy_1，y_2，⋯，y_nとするとき，次の式が成り立つ。

$$\overline{y}=\frac{1}{n}\{(ax_1+b)+(ax_2+b)+\cdots(ax_n+b)\}=\frac{1}{n}\{a(x_1+x_2+\cdots+x_n)+nb\}$$

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

また，x の分散 ${s_x}^2$と y の分散 ${s_y}^2$ の間には，次の式が成り立つ。

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

$$=\frac{1}{n}[\{(ax_1+b)^2-(a\overline{x}+b)^2\}+\{(ax_2+b)-(a\overline{x}+b)\}^2+\cdots+\{(ax_n+b)-(a\overline{x}+b)\}^2]$$

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

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

これより，y の標準偏差 s_y は　$s_y=\sqrt{{s_y}^2}=\sqrt{a^2{s_x}^2}=|a|s_x$

よって，$y=ax+b$ という関係があるとき，次の性質が成り立つ。

平均値　$\boldsymbol{\overline{y}=a\overline{x}+b}$

分散　$\boldsymbol{{s_y}^2=a^2{s_x}^2}$

標準偏差　$\boldsymbol{s_y=|a|s_x}$

問1　あるクラスで数学のテストを実施したところ，平均値 55 点，標準偏差 7 点であった。このとき，次の値を求めよ。

(1)　全員の得点に 10 点ずつ加えたときの平均値と標準偏差

(2)　全員の得点から 5 点ずつ引いて 2 倍したときの平均値と標準偏差

COLUMN　偏差値

（教科書 p.180）

数学の試験を 2 回行ったところ，1 回目の得点は 75 点，2 回目の得点も 75 点でしたが，これを同じ結果と見てもいいのでしょうか。2 つの試験の難易度が異なるとすると，得点を単純に比較することはできません。それぞれの試験を受けた人が同じであれば，順位を比較するという方法もありますが，そうではない場合にはどうしたら比較できるでしょうか。

変量 x の各データの値 x_i の偏差を $x_i - \overline{x}$ の標準偏差 s_x で割った値をとると，次のような変量 z を考えられます。

$$z = \frac{x_i - \overline{x}}{s_x}$$

この変量 z は平均値 0，標準偏差 1 をもつ単位のない数値になります。このような変量の変換方法を**標準化**といいます。

試験の得点を標準化してから，比較しやすいように，10 を掛けて 50 を加えた値を求めると，平均値が 50，標準偏差が 10 である変量になります。この値を**偏差値**といい，データの中での位置を示したものです。

$$\text{得点の偏差値} = 50 + 10 \times \frac{\text{得点} - \text{得点の平均値}}{\text{得点の標準偏差}}$$

たとえば，2 回のテストの得点がともに 75 点であっても平均値や標準偏差によって，偏差値は次のようになります。

	得点	平均値	標準偏差	偏差値
第 1 回	75	51	20	62
第 2 回	75	63	6	70

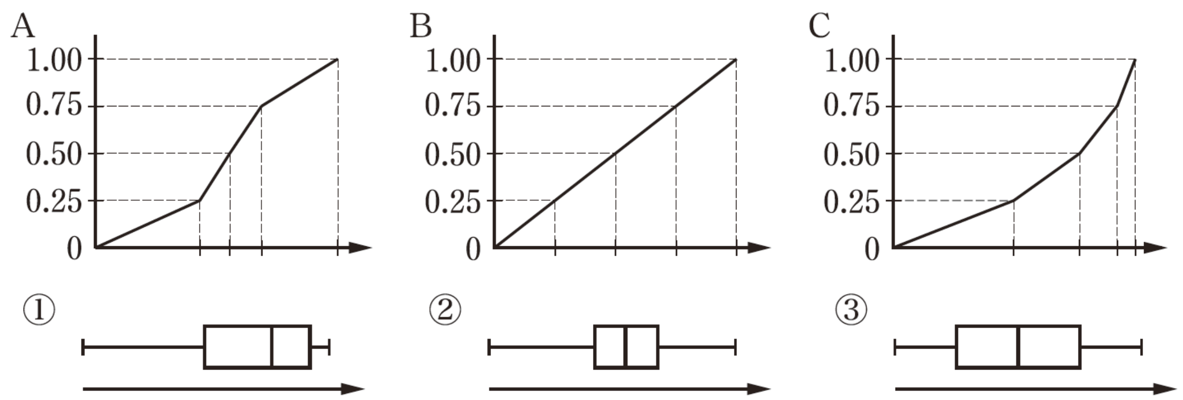
課題 1　偏差値が 100 より大きくなることはあるでしょうか。あるとしたら，それはどのような場合でしょうか。

課題 2　偏差値が 0 より小さくなることはあるでしょうか。あるとしたら，それはどのような場合でしょうか。

練習問題 A

（教科書 p.176）

1　次の累積相対度数折れ線 A，B，C について，それぞれ対応する箱ひげ図を①，②，③の中から選べ。



A は②，B は③，C は①

2　20 人の生徒に，数学と国語の 5 点満点の小テストを行った。数学の得点を x 点，国語の得点を y 点とする。そのときの結果が次の表である。たとえば，数学が 3 点，国語が 4 点の生徒は 7 人いることがわかる。このとき，次の問に答えよ。

y \ x	0	1	2	3	4	5	計
5				1		3	4
4			3	7	1	1	12
3		3		1			4
2							0
1							0
0							0
計	0	3	3	9	1	4	20

(1)　x の平均値 $\overline{x}$ と，y の平均値 $\overline{y}$ を求めよ。

$\overline{x}=\frac{1}{20}\{1\cdot3+2\cdot3+3\cdot9+4\cdot1+5\cdot4\}$

$=\frac{60}{20}=3$（点）

$\overline{y}=\frac{1}{20}\{3\cdot4+4\cdot12+5\cdot4\}=\frac{80}{20}=4$（点）

(2)　x と y の分散は，右の表のようになる。x と y との相関係数 r と値が一致する式を①～③の中からすべて選べ。

	数学	国語
分散	1.6	0.4

①　$\frac{3(5-\overline{x})(3-\overline{y})}{20\sqrt{1.6}\sqrt{0.4}}$　　②　$\frac{6(5-\overline{x})(5-\overline{y})}{20\sqrt{1.6}\sqrt{0.4}}$　　③　$\frac{3(1-\overline{x})(3-\overline{y})}{\sqrt{16}\sqrt{4}}$

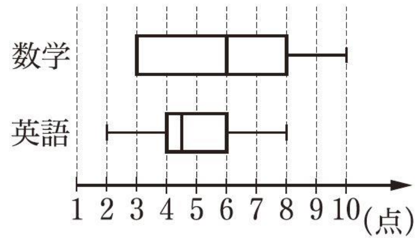
r と値が一致するのは②と③

練習問題 B

（教科書p.177）

3 10 人の生徒に，数学と英語の 10 点満点の小テストを行った。数学の得点を x 点，英語の得点を y 点とする。そのときの結果が下の表と箱ひげ図である。また，x，y の平均値を $\overline{x}$，$\overline{y}$ とする。

生徒番号	x	y	$(x-\overline{x})^2$	$(y-\overline{y})^2$	$(x-\overline{x})(y-\overline{y})$
1	8	6	4	1	2
2	6	4	0	1	0
⋮	⋮	⋮	⋮	⋮	⋮
10	3	3	9	4	6
合計	A	B	64	36	43



(1) 表の A，B の値を求めよ。

1 番の生徒のデータから

$(8-\overline{x})(6-\overline{y})=2$ ……①

2 番の生徒のデータから

$(6-\overline{x})^2=0$ ……②

②より $\overline{x}=6$

これを①に代入して $\overline{y}=5$

この結果は 1 番，2 番，10 番の生徒のデータに適する。

よって

$A=10\overline{x}=60$

$B=10\overline{y}=50$

(2) x の分散 $s_x{}^2$と，y の分散 $s_y{}^2$を求めよ。

表より

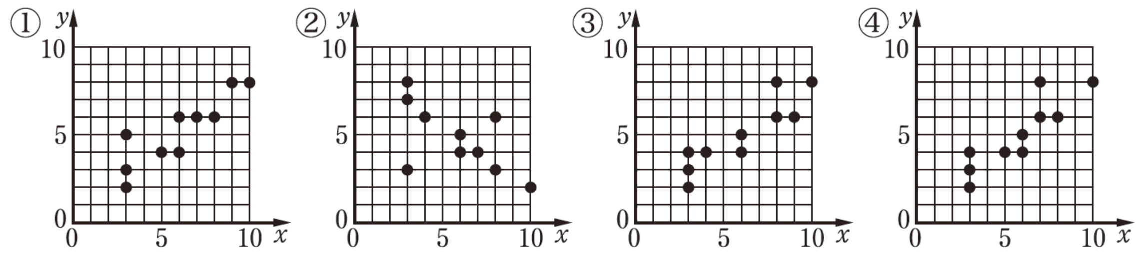
$s_x{}^2=\frac{64}{10}=6.4$，$s_y{}^2=\frac{36}{10}=3.6$

(3) xと y の相関係数 r を小数第 3 位を四捨五入して，小数第 2 位まで求めよ。

$$r=\frac{\frac{43}{10}}{\sqrt{\frac{64}{10}}\sqrt{\frac{36}{10}}}=\frac{43}{\sqrt{64}\sqrt{36}}=\frac{43}{8\cdot 6}$$

$$=0.895\cdots \fallingdotseq 0.90$$

(4) x と y の散布図として，適当なものを次の①～④の中から選べ。



③

①は英語の中央値が異なる。したがって，不適。

②は負の相関関係と考えられる。したがって，不適。

③は与えられた条件に矛盾しない。

④は数学の第 3 四分位数が異なる。したがって，不適。

(5) 11 人目の生徒が，同じ数学と英語の小テストを受けた。この生徒の得点は，数学が 6 点，英語が 5 点であった。次の記述のうち，適当でないものをすべて選べ。

① 11 人の数学の平均値は，10 人のときの平均値 $\overline{x}$ と等しい。

② 11 人の英語の分散は，10 人のときの分散 $s_y{}^2$ より小さい。

③ 11 人の数学の得点と英語の得点の相関係数は，10 人のときの相関係数 rより大きい。

③

①②は正しい。

11 人の数学と英語の得点の分散を，それぞれ $t_x{}^2$，$t_y{}^2$ とし，共分散を t_{xy} とする。

$$t_x{}^2=\frac{10}{11}s_x{}^2$$

$$t_y{}^2=\frac{10}{11}s_y{}^2$$

$$t_{xy}=\frac{10}{11}s_{xy}$$

よって，11 人の数学と英語の得点の相関係数 r' は

$$r'=\frac{t_{xy}}{\sqrt{t_x{}^2}\sqrt{t_y{}^2}}=\frac{\frac{10}{11}s_{xy}}{\sqrt{\frac{10}{11}s_x{}^2}\sqrt{\frac{10}{11}s_y{}^2}}$$

$$=\frac{s_{xy}}{\sqrt{s_x{}^2}\sqrt{s_y{}^2}}=r$$

したがって，相関係数は変わらない。

参考　はずれ値

（教科書 p.178）

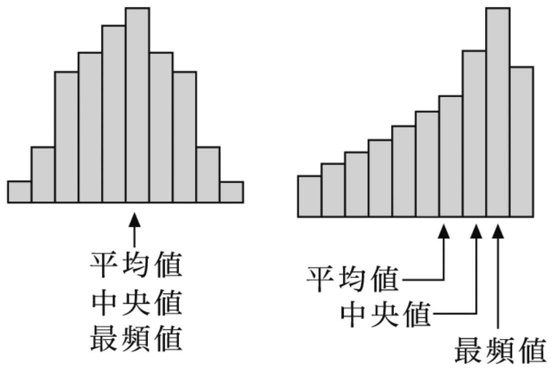
データには，他と比べて値が大きくなる「**はずれ値**」といわれるものが含まれる場合がある。

たとえば，ある5人のテスト結果が 15 点，13 点，12 点，16 点，98 点のとき，平均値は 30.8 点，中央値は 15 点である。

98 点は他の点数より著しく高いため，98 点の値を「はずれ値」として除いて考えると，平均値は 14 点と大きく変化するが，中央値は 14 点とあまり変化しない。

このことから，平均値のほうが中央値よりも「はずれ値」からの影響を受けやすいといえる。

また，平均値，中央値，最頻値は，右の図のように，データが 1 つの山でほぼ左右対称に分布しているとき，互いに近い値となる。

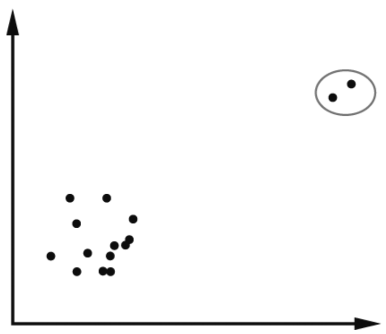


一方，データが 1 つの山で左右対称に分布していないときは，平均値よりも中央値や最頻値のほうが，データの代表的な位置を表すのに適切である。

さらに，相関係数も「はずれ値」の影響を受けやすい。

右の散布図で表されたデータの相関係数は 0.885 であるが，円の中の 2 つの「はずれ値」を除いて相関係数を求めると 0.023 となる。

このため，相関を考えるときは相関係数だけではなく，散布図も調べることが重要である。



発 展　$y=ax+b$ における平均値と分散・標準偏差

（教科書 p.179）

2 つの変数 x と y の間に，a，b を定数として $y=ax+b$ という関係があるとき，x の平均値 $\overline{x}$ と y の平均値 $\overline{y}$ の関係について調べてみよう。

変量 x の n 個の値を x_1，x_2，…，x_nとし，変量 y の n 個の値をy_1，y_2，…，y_nとするとき，次の式が成り立つ。

$$\overline{y}=\frac{1}{n}\{(ax_1+b)+(ax_2+b)+\cdots(ax_n+b)\}=\frac{1}{n}\{a(x_1+x_2+\cdots+x_n)+nb\}$$

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

また，x の分散 ${s_x}^2$と y の分散 ${s_y}^2$ の間には，次の式が成り立つ。

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

$$=\frac{1}{n}[\{(ax_1+b)^2-(a\overline{x}+b)^2\}+\{(ax_2+b)-(a\overline{x}+b)\}^2+\cdots+\{(ax_n+b)-(a\overline{x}+b)\}^2]$$

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

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

これより，y の標準偏差 s_y は　$s_y=\sqrt{{s_y}^2}=\sqrt{a^2{s_x}^2}=\ |a|s_x$

よって，$y=ax+b$ という関係があるとき，次の性質が成り立つ。

平均値　$\boldsymbol{\overline{y}=a\overline{x}+b}$

分散　$\boldsymbol{{s_y}^2=a^2{s_x}^2}$

標準偏差　$\boldsymbol{s_y=|a|s_x}$

問1　あるクラスで数学のテストを実施したところ，平均値 55 点，標準偏差 7 点であった。このとき，次の値を求めよ。

(1)　全員の得点に 10 点ずつ加えたときの平均値と標準偏差

(2)　全員の得点から 5 点ずつ引いて 2 倍したときの平均値と標準偏差

数学の得点を x，平均値を $\overline{x}$，分散を ${s_x}^2$，標準偏差を s_x とする。
また，条件より，$\overline{x}=55$，$s_x=7$ である。

(1)　$y=x+10$とする。

y の平均値 $\overline{y}$ は

$\overline{y}=\overline{x}+10=55+10=65$（点）

y の標準偏差 s_y は

$s_y=s_x=7$（点）

(2)　$z=2(x-5)$とすると　$z=2x-10$

z の平均値 $\overline{z}$ は

$\overline{z}=2\overline{x}-10=2\cdot55-10=100$（点）

z の標準偏差 s_z は

$s_z=\ |2|S_x=2\cdot7=14$（点）

COLUMN　偏差値

（教科書 p.180）

数学の試験を 2 回行ったところ，1 回目の得点は 75 点，2 回目の得点も 75 点でしたが，これを同じ結果と見てもいいのでしょうか。2 つの試験の難易度が異なるとすると，得点を単純に比較することはできません。それぞれの試験を受けた人が同じであれば，順位を比較するという方法もありますが，そうではない場合にはどうしたら比較できるでしょうか。

変量 x の各データの値 x_i の偏差を $x_i - \overline{x}$ の標準偏差 s_x で割った値をとると，次のような変量 z を考えられます。

$$z = \frac{x_i - \overline{x}}{s_x}$$

この変量 z は平均値 0，標準偏差 1 をもつ単位のない数値になります。このような変量の変換方法を**標準化**といいます。

試験の得点を標準化してから，比較しやすいように，10 を掛けて 50 を加えた値を求めると，平均値が 50，標準偏差が 10 である変量になります。この値を**偏差値**といい，データの中での位置を示したものです。

$$\text{得点の偏差値} = 50 + 10 \times \frac{\text{得点} - \text{得点の平均値}}{\text{得点の標準偏差}}$$

たとえば，2 回のテストの得点がともに 75 点であっても平均値や標準偏差によって，偏差値は次のようになります。

	得点	平均値	標準偏差	偏差値
第 1 回	75	51	20	62
第 2 回	75	63	6	70

課題 1 偏差値が 100 より大きくなることはあるでしょうか。あるとしたら，それはどのような場合でしょうか。

課題 2 偏差値が 0 より小さくなることはあるでしょうか。あるとしたら，それはどのような場合でしょうか。