数学①　第８回試練　データの分析

1 次のデータの第 1 四分位数，第 2 四分位数，第 3 四分位数を求めよ。

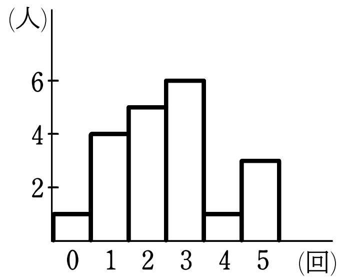
(1) 12，35，47，59，68，73，74，79，87，97

(2) 2，7，10，14，22，33，48，71，84，91，96，98

2 右のヒストグラムは，ある高校の生徒 20 人について，ある 5 日間に校内の売店を利用した回数を調べた結果である。

(1) 利用回数の最頻値，中央値を求めよ。

(2) 利用回数の平均値を求めよ。



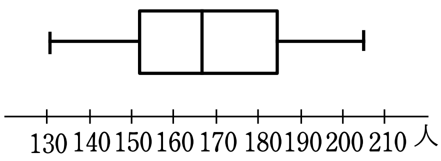
3 次のデータは，ある 10 人の生徒の数学のテストの得点である。ただし，a の値は 0 以上の整数である。

60　74　66　62　82　38　45　41　67　a　(点)

(1) a の値がわからないとき，このデータの中央値として何通りの値があり得るか。

(2) このデータの平均値が 60.0 点のとき，このデータの中央値を求めよ。

4 右の図は，ある店の 30 日間にわたる来客数のデータの箱ひげ図である。
この箱ひげ図から読み取れることとして正しいものを，次の ① ～ ③ から 1 つ選べ。



① 来客数 150 人以上 180 人未満のところには，半分以上の日が分布している。

② 来客数が 160 人以上の日が 15 日以上あった。

③ 少なくとも 2 日は，来客数が 200 人以上である。

5 次のデータは，あるパズルに挑戦した 10 人について，完成させるまでにかかった時間 x (分) をまとめたものである。ただし，x のデータの平均値を $\overline{x}$ で表し，20 分を超えた人はいなかったものとする。次の問いに答えよ。

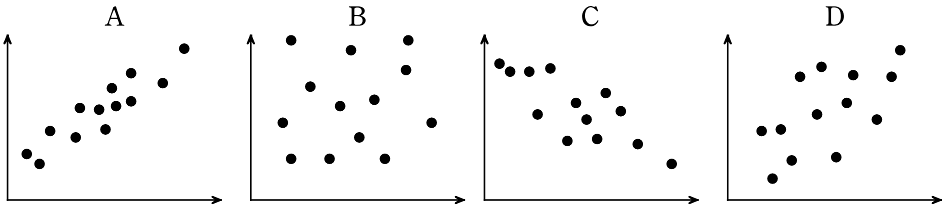
	①	②	③	④	⑤	⑥	⑦	⑧	⑨	⑩
x	13	a	7	3	11	18	7	b	16	3
$(x-\overline{x})^2$	4	c	16	64	0	d	16	1	25	64

(1) a，b，c，d の値を求めよ。

(2) x のデータの分散と標準偏差を求めよ。ただし，小数第 2 位を四捨五入せよ。

6 下の A ～ D の散布図と対応する相関係数 r の値の組について適当なものを，次の ① ～ ④ の中から選べ。

① A：$r=0.8$ B：$r=0.6$ C：$r=-0.6$ D：$r=0.9$
② A：$r=0.9$ B：$r=0.1$ C：$r=-0.8$ D：$r=0.6$
③ A：$r=0.8$ B：$r=-0.1$ C：$r=0.7$ D：$r=0.7$
④ A：$r=-0.9$ B：$r=-0.1$ C：$r=0.8$ D：$r=0.8$



7 下の表は，10 人の生徒に 30 点満点の 2 種類のテスト A，B を行った得点の結果である。テスト A，B の得点をそれぞれ x，y とするとき，x と y の相関係数 r を求めよ。ただし，小数第 3 位を四捨五入せよ。

生徒番号	1	2	3	4	5	6	7	8	9	10
x	29	25	22	28	18	23	26	30	30	29
y	23	23	18	26	17	20	21	20	26	26

8 30 個の値からなるデータがあり，そのうちの 20 個の値の平均値は 7，分散は 5，残り 10 個の値の平均値は 4，分散は 8 である。

(1) このデータ全体の平均値を求めよ。 (2) このデータ全体の分散を求めよ。

9 n 個の値の組として与えられている 2 つの変量 X，Y に対し，新たな変量 X'，Y' を $X'=aX+b$，$Y'=cY+d$ (a，b，c，d は定数で，$a \neq 0$，$c \neq 0$) によって定義する。
次の $\boxed{\text{ア}}$ ～ $\boxed{\text{ウ}}$ に当てはまるものを，下の ⓪ ～ ⑦ のうちからそれぞれ 1 つずつ選べ。

(1) X' の分散は，X の分散の $\boxed{\text{ア}}$ 倍になる。

(2) X' と Y' の共分散は，X と Y の共分散の $\boxed{\text{イ}}$ 倍である。

(3) X' と Y' の相関係数は，X と Y の相関係数の $\boxed{\text{ウ}}$ 倍である。

⓪ a ① a^2 ② ac ③ $\dfrac{ac}{|ac|}$ ④ b ⑤ b^2 ⑥ bd ⑦ $|bd|$

$\boxed{1}$ 解答 第 1 四分位数，第 2 四分位数，第 3 四分位数の順に

(1) 70.5，47，79 (2) 40.5，12，87.5

$\boxed{2}$ 解答 (1) 最頻値 3 回，中央値 2.5 回 (2) 2.55 回

$\boxed{3}$ 解答 (1) 7 通り (2) 63.5 点

$\boxed{4}$ 解答 ②

$\boxed{5}$ 解答 (1) $a=20$，$b=12$，$c=81$，$d=49$ (2) 32，5.6 分

$\boxed{6}$ 解答 ②

$\boxed{7}$ 解答 0.77

$\boxed{8}$ 解答 (1) 6 (2) 8

$\boxed{9}$ 解答 (ア) ⓵ (イ) ⓶ (ウ) ⓷

$\boxed{1}$ 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)　$Q_2=\dfrac{68+73}{2}=70.5$，$Q_1=47$，$Q_3=79$

(2)　$Q_2=\dfrac{33+48}{2}=40.5$，$Q_1=\dfrac{10+14}{2}=12$，$Q_3=\dfrac{84+91}{2}=87.5$

$\boxed{2}$ (1)　ヒストグラムから，最頻値は　3 回

利用回数が少ない方から 10 番目の生徒の利用回数は 2 回，11 番目の生徒の利用回数は 3 回である。

よって，中央値は　$\dfrac{1}{2}(2+3)=2.5$（回）

(2)　平均値は　$\dfrac{1}{20}(0\times1+1\times4+2\times5+3\times6+4\times1+5\times3)=\dfrac{51}{20}=2.55$（回）

$\boxed{3}$ (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

[1]　$a\leqq60$ 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

$$\frac{60+62}{2}=61\ (\text{点})$$

[2]　$a\geqq66$ 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

$$\frac{62+66}{2}=64\ (\text{点})$$

[3]　$61\leqq a\leqq65$ のとき

5 番目と 6 番目の得点は a 点と 62 点であるから，中央値は　$\dfrac{a+62}{2}$（点）

これは，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 60.0 点であるとき

$$\frac{1}{10}(38+41+45+60+62+66+67+74+82+a)=60.0$$

これを解いて　$a=65$

$61\leqq a\leqq65$ であるから，得点の中央値は　$\dfrac{65+62}{2}=63.5$（点）

$\boxed{4}$ ①　第 1 四分位数は 150 人以上であるが，第 3 四分位数は180 人より大きいから，150 人以上 180 人未満のところに半分以上の日が分布していることは読み取れない。

よって，① は正しくない。

②　中央値が 160 人以上であるから，半分以上の日にちで来客数が 160 人以上であることが読み取れる。

よって，② は正しい。

③　最大値が 200人以上であるから，200人以上の日が少なくとも 1 日あることが読み取れるが，2日あるかどうかは読み取れない。

よって，③は正しくない。

以上から，正しいものは②

$\boxed{5}$ (1) 番号5のデータに着目すると，$(11-\overline{x})^2=0$ であるから，平均値 $\overline{x}$ は11(分)

x のデータの総和は　$11\times10=110$

よって　$13+a+7+3+11+18+7+b+16+3=110$

したがって　$a=32-b$

(3) 番号6のデータに着目すると，

$d=(18-11)^2=7^2=49$

番号8のデータに着目すると，$(b-11)^2=1$ より　$b-11=\pm1$

よって　$b=12$ または $b=10$

$b=12$ のとき，$a=32-12=20$ となり，$a\leqq20$ に適する。

$b=10$ のとき，$a=32-10=22$ となり，$a\leqq20$ に適さない。

よって　$a=20$，$b=12$

$c=(20-11)^2=9^2=81$

したがって　$a=20$，$b=12$，$c=81$，$d=49$

(4) 偏差の2乗の和は

$4+81+16+64+0+49+16+1+25+64=320$

よって，分散は　$\dfrac{1}{10}\times320=32$

標準偏差は　$\sqrt{32}=4\sqrt{2}\fallingdotseq5.6$ (分)

$\boxed{6}$ AとDの散布図では，正の相関関係があるから　$r>0$

また，その傾向はAの方が著しいから，Aの相関係数の方がDの相関係数よりも1に近い。

Cの散布図では，負の相関関係があるから　$r<0$

Bの散布図では，相関が認められないから，相関係数は0に近い。

したがって　②

$\boxed{7}$ x，y の平均をそれぞれ $\overline{x}$，$\overline{y}$ とすると

$$\overline{x}=\frac{1}{10}(29+25+22+28+18+23+26+30+30+29)=26$$

$$\overline{y}=\frac{1}{10}(23+23+18+26+17+20+21+20+26+26)=22$$

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	3	1	9	1	3
2	-1	1	1	1	-1
3	-4	-4	16	16	16
4	2	4	4	16	8
5	-8	-5	64	25	40
6	-3	-2	9	4	6
7	0	-1	0	1	0
8	4	-2	16	4	-8
9	4	4	16	16	16
10	3	4	9	16	12
計			144	100	92

ゆえに　　$r=\dfrac{92}{\sqrt{144\times100}}=\dfrac{92}{12\times10}\fallingdotseq0.77$

$\boxed{8}$ (1)　30 個の値の和は　　$7\times20+4\times10=180$

よって，このデータ全体の平均値は　　$\dfrac{180}{30}=6$

(2)　20 個の値の 2 乗の平均値を a とすると　　$a-7^2=5$　　よって　$a=54$

残りの 10 個の値の 2 乗の平均値を b とすると　$b-4^2=8$　　よって　$b=24$

よって，30 個の値の 2 乗の和は　　$a\times20+b\times10=54\times20+24\times10=1320$

したがって，このデータ全体の分散は　　$\dfrac{1320}{30}-6^2=44-36=8$

$\boxed{9}$ X，Y，X'，Y' のそれぞれについて，

平均値を $\overline{X}$，$\overline{Y}$，$\overline{X'}$，$\overline{Y'}$，

分散を　$s_X{}^2$，$s_Y{}^2$，$s_{X'}{}^2$，$s_{Y'}{}^2$　とする。

また，X と Y の共分散を s_{XY}，相関係数を r，X' と Y' の共分散を $s_{X'Y'}$，相関係数を r' とする。

(1)　$s_{X'}{}^2=a^2s_X{}^2$ が成り立つから，X' の分散は，X の分散の a^2 倍になる。　(ア ①)

(2)　$s_{X'Y'}=acs_{XY}$ が成り立つから，X' と Y' の共分散は，X と Y の共分散の ac 倍である。　(イ ②)

(3)　$r'=\dfrac{s_{X'Y'}}{s_{X'}s_{Y'}}=\dfrac{acs_{XY}}{|a|s_X\cdot|c|s_Y}=\dfrac{ac}{|ac|}r$

よって，X' と Y' の相関係数は，X と Y の相関係数の $\dfrac{ac}{|ac|}$ 倍である。　(ウ ③)

参考　(変量の変換)

変量 X，Y が，n 個の値の組 $(X_1,\ Y_1)$，$(X_2,\ Y_2)$，……，$(X_n,\ Y_n)$ であるとし，$k=1$，……，n に対して，$X_k'=aX_k+b$，$Y_k'=cY_k+d$ が成り立つとき

$$ns_{X'}{}^2=(X_1'-\overline{X'})^2+\cdots\cdots+(X_n'-\overline{X'})^2$$
$$=\{(aX_1+b)-(a\overline{X}+b)\}^2+\cdots\cdots+\{(aX_n+b)-(a\overline{X}+b)\}^2$$
$$=a^2\{(X_1-\overline{X})^2+\cdots\cdots+(X_n-\overline{X})^2\}$$
$$=a^2ns_X{}^2$$

よって，$s_{X'}{}^2=a^2s_X{}^2$ が成り立つ。

$$ns_{X'Y'}=(X_1'-\overline{X'})(Y_1'-\overline{Y'})+\cdots\cdots+(X_n'-\overline{X'})(Y_n'-\overline{Y'})$$
$$=\{(aX_1+b)-(a\overline{X}+b)\}\{(cY_1+d)-(c\overline{Y}+d)\}$$
$$+\cdots\cdots+\{(aX_n+b)-(a\overline{X}+b)\}\{(cY_n+d)-(c\overline{Y}+d)\}$$
$$=ac\{(X_1-\overline{X})(Y_1-\overline{Y})+\cdots\cdots+(X_n-\overline{X})(Y_n-\overline{Y})\}$$
$$=acns_{XY}$$

よって，$s_{X'Y'}=acs_{XY}$ が成り立つ。