2024/11/18

データ
① 合併 ② 変数変換

共80分（4科目2）

英15 数7. 全10分 化

下位データの中央値　　上位

例題 35

定数 $a, b, c, d, e, f, g, h, i, j, k$ が $a<b<c<d<e<f<g<h<i<j<k$ を満たしているとき，次の各データに対して，第 1 四分位数 Q_1，中央値 Q_2，第 3 四分位数 Q_3 をそれぞれ求めよ。

8コ (1) a, b, c, d, e, f, g, h　　$Q_2=\frac{d+e}{2}$　$Q_1=\frac{b+c}{2}$　$Q_3=$

9コ (2) $a, b, c, d, e, f, g, h, i$　　$Q_2=e$，$Q_1=\frac{b+c}{2}$

10コ (3) $a, b, c, d, e, f, g, h, i, j$　　$Q_2=\frac{e+f}{2}$，$Q_1=c$

11コ (4) $a, b, c, d, e, f, g, h, i, j, k$　　$Q_2=f$，$Q_1=c$

$\boxed{\text{偏差の平均は必ず }0}$

$$\frac{1}{n}\sum_{k=1}^{n}(x_k-\overline{x})=\boxed{\frac{1}{n}\sum_{k=1}^{n}x_k}-\frac{1}{n}\overline{\sum_{k=1}^{n}\overline{x}}=0$$

$$\frac{1}{n}\sum_{k=1}^{n}x_k=\overline{x},\quad \sum_{k=1}^{n}\overline{x}=n\overline{x},\quad \frac{1}{n}\sum_{k=1}^{n}\overline{x}=\overline{x}$$

$\boxed{\begin{aligned}\text{分散}&=\text{偏差}^2\text{の平均}\\&=2\text{乗の平均}-\text{平均の}2\text{乗}\end{aligned}}$

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

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

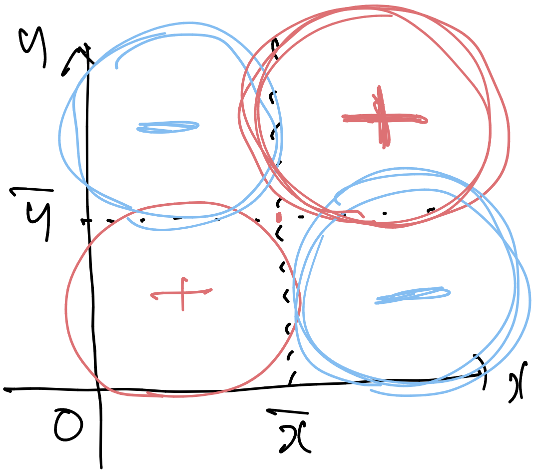
$$\frac{1}{n}\sum_{k=1}^{n}x_k^2=\overline{x^2},\quad \frac{1}{n}\sum_{k=1}^{n}x_k=\overline{x},\quad \sum_{k=1}^{n}\overline{x}^2=n\overline{x}^2$$

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

相関

(英・数)のテスト

$(x_1, y_1), (x_2, y_2), (x_3, y_3), \cdots, (x_n, y_n)$



$\boxed{\text{共分散}}$ co-variance

$$C_{xy} = S_{xy}$$

$$S_{xy} = \frac{1}{n}\sum_{k=1} (x_k - \bar{x}) \times (y_k - \bar{y})$$

$\boxed{\text{相関係数}}$ $$r_{xy} = \frac{S_{xy}}{S_x \cdot S_y}$$

(性質) $$-1 \leqq r_{xy} \leqq 1$$

$\boxed{\text{共分散}}$

$$C_{xy} = \frac{1}{n}\sum_{k=1}^{n} (x_k - \bar{x}) \times (y_k - \bar{y})$$

$\Downarrow$ ($z_k = a x_k + b$ とおく $\bar{z} = a\bar{x} + b$)

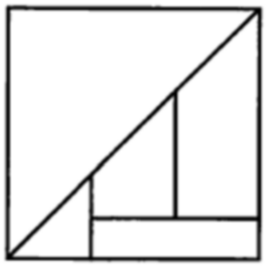
$$C_{zy} = \frac{1}{n}\sum_{k=1}^{n} (z_k - \bar{z}) \times (y_k - \bar{y})$$

$$\boxed{(a x_k + b) - (a\bar{x} + b) = a(x_k - \bar{x})}$$

8-13　右の図のように5つの部分に分けられた正方形を何色かで塗り分けたい。境界線を接する2つの図形は必ず異なる色で塗ることにする。いま，5色の絵の具があるとして，次のような塗り分け方は，それぞれ何通りあるか。

(1)　5色で塗り分ける方法　　(2)　4色で塗り分ける方法

(3)　3色で塗り分ける方法

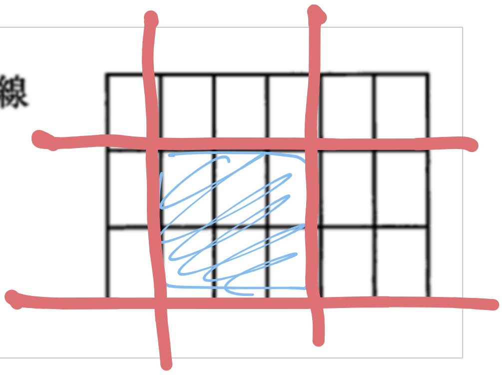


(1) Just 5色　　$5!$

(2) Just 4色　　${}_5C_4 \times (4\times3\times2\times1)\times 2 = 240$　　$2\times$

(3) Just 3色

8-22　右の図のように，横に4本，縦に7本の平行線が並んでいる。この中に長方形はいくつあるか。



${}_4C_2 \times {}_7C_2$

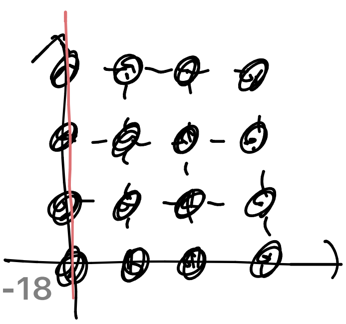
【5】2007 埼玉医科大学医学部

xy 平面上に $(n, m)(n = 0, 1, 2, 3 ; m = 0, 1, 2, 3)$ を座標とする 16 個の点がある．以下の各問に答えなさい．

(1) 3 点を選んでそれらを頂点とする 3 角形をつくるとき，3 角形の数は $\boxed{\text{セソタ}}$ 個である．

(2) 2 点を結んでできる直線の数は $\boxed{\text{チツ}}$ 本である．

(1) ${}_{16}C_3$ ~~1D~~　　${}_{16}C_3 = \frac{16\cdot15\cdot14}{3\cdot2\cdot1} = 16\cdot5\cdot7 = \underline{560}$



${}_4C_3 = 4$ が $\boxed{10}$ 本ある

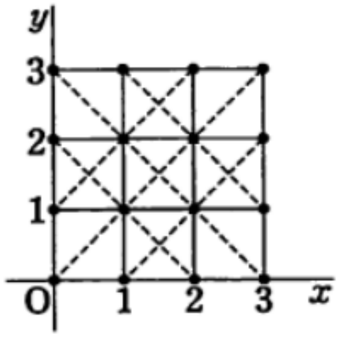
タテヨコナナメ

516

${}_3C_3 = 1$ が 4本ある

過去問めぐり　埼玉医科大学 2007　解答

❹ ［答］ **30.** 5　**31.** 1　**32.** 6　**33.** 6　**34.** 2



［解　説］ (1) 16個の点から3点を選ぶ方法は　${}_{16}C_3=560$ (個)
この中で三角形ができないのは，

(イ) 直線 $x=n$, $y=m$ および $y=x$, $y=-x+3$ の1つの直線から3点を選ぶとき；

(ロ) 直線 $y=x\pm1$, $y=-x+2$, $y=-x+4$ の1つの直線から3点を選ぶとき；

の場合があり，(イ)と(ロ)の場合の数は順に

$${}_4C_3\times10=40\ (\text{個}),\ {}_3C_3\times4=4\ (\text{個})$$

よって，三角形の数は

$$560-(40+4)=516\ (\text{個})$$

(2) 傾きが0以上および傾きをもたない直線の数は

傾き	0	$\frac{1}{3}$	$\frac{1}{2}$	$\frac{2}{3}$	1	$\frac{3}{2}$	2	3	なし
個数	4	3	6	2	5	2	6	3	4

傾きが負のものは，正のものと同じだけあるので

$$(3+6+2+5+2+6+3)\times2+4+4=62\ (\text{個})$$

10-17 1個のバクテリアが10分後に2個，1個，0個になる確率が，それぞれ $\frac{1}{2}$，$\frac{1}{3}$，$\frac{1}{6}$ であるとする。1個のバクテリアが

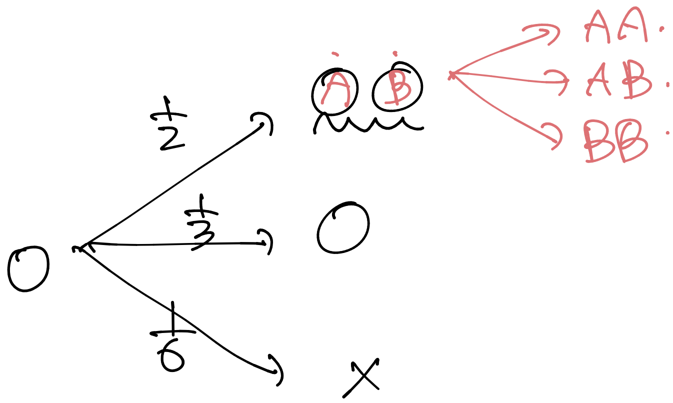
(1) 20分後に2個になっている確率を求めよ。

(2) 30分後に6個になっている確率を求めよ。

（名古屋大）

(1) $11/36$

(2) $5/96$



データの分析　練習問題　（基礎〜標準）

$\boxed{1}$

生徒 10 人の試験結果が次のようになった。

85，80，25，0，65，55，40，55，75，70　(点)

このデータの平均値，中央値，最頻値，分散を求めよ。
また，生徒全員に 10 点ずつ加点したときの平均値および分散を求めよ。

$\boxed{2}$

次のデータは，ある商店における A 弁当と B 弁当の 7 日間の販売数である。

A 弁当　22，28，16，24，33，27，21　(個)
B 弁当　18，22，17，13，28，35，32　(個)

A 弁当と B 弁当のデータの箱ひげ図を並べてかけ。また，データの散らばりの度合いが大きいのは，A 弁当，B 弁当のうちどちらか。

$\boxed{3}$

右の表は，生徒 40 人の国語と数学の小テストの得点をまとめた相関表である。

(1)　国語の得点が 4 点である生徒の数学の得点の中央値と最頻値を求めよ。また，数学の得点が 2 点である生徒の国語の得点の標準偏差を求めよ。

(2)　国語の得点と数学の得点の相関係数を求めよ。

数学(点) ＼ 国語(点)	0	1	2	3	4	5
5				1	3	2
4			3		3	2
3			2	4	5	1
2		2	5		1	
1		3	2	1		
0						

$\boxed{4}$

ある町の 1 日の最低気温と最高気温について，10 日間調べて資料を作成した。最低気温を変量 x，最高気温を変量 y で表すものとする。

最低気温 x(℃)	22.3	22.5	22.7	23.0	23.3	23.5	23.6	23.7	24.1	24.3
最高気温 y(℃)	A	34.8	32.6	28.4	33.6	31.0	31.4	33.1	29.2	B

(1) 最低気温の平均値を求めよ。また，最高気温の平均値が 31.2 ℃ であり，最低気温と最高気温の相関係数がちょうど 0 であるとき，A，B の値を求めよ。

(2) (1) のとき，1 日の気温差を表す変量を $w\,(=y-x)$ とする。変量 x と変量 w の散布図をかけ。また，次の (ア) ～ (ウ) のうち，このデータについて正しいものをすべて答えよ。

(ア) 最低気温が高いほど最高気温も高いという傾向がある。

(イ) 最低気温が高いほど 1 日の気温差が大きいという傾向がある。

(ウ) 最高気温が高いほど 1 日の気温差が大きいという傾向がある。

$\boxed{5}$

右の表は，あるクラスの生徒の身長を測定した結果をまとめたものである。

階級 (cm)	度数
150 以上 155 未満	2
155 以上 160 未満	3
160 以上 165 未満	5
165 以上 170 未満	14
170 以上 175 未満	8
175 以上 180 未満	7
180 以上 185 未満	1

(1) 階級値 x に対し，$u=\dfrac{x-167.5}{5}$ とする。

u の平均値 $\overline{u}$ と分散 ${S_u}^2$ の値を求めよ。

(2) x の平均値 $\overline{x}$ と分散 ${S_x}^2$ は

$$\overline{x}=167.5+5\overline{u},\qquad {S_x}^2=25{S_u}^2$$

で与えられることを示し，$\overline{x}$ と ${S_x}^2$ の値を求めよ。

$\boxed{6}$

x_1，x_2，……，x_n はいずれも 0，1，2 のうちのどれかの値をとり，x_1，x_2，……，x_n のうち，値が 2 であるものが p 個，1 であるものが q 個（ただし，$0 \leqq p \leqq n$，$0 \leqq q \leqq n$）であるとする。x_1，x_2，……，x_n の平均値を $\overline{x}$，分散を V で表す。

(1) $\overline{x}$ および V を n，p，q を用いて表せ。

(2) p，q が動くとき，V の最大値を求めよ。

$\boxed{7}$

(1) 変量 x のデータの値を x_1，x_2，……，x_n とし，x の平均値を $\overline{x}$，標準偏差を $s\ (s \neq 0)$ とする。$p \geqq 1$ を満たす p について，$|x_k - \overline{x}| \geqq ps$ を満たす x_k の個数を $K(p)$ とするとき，$K(p) \leqq \dfrac{n}{p^2}$ が成り立つことを示せ。

(2) あるテストを受けた生徒 300 人について調べたところ，得点の平均値が 63 点，標準偏差が 8 点であった。得点が 47 点より高く 79 点より低い生徒は少なくとも何人いるか，(1) を利用して答えよ。

$\boxed{8}$

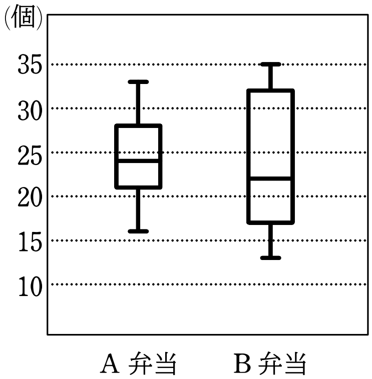
10 個の値からなるデータがあり，そのうちの 4 個の値の平均値は 30，分散は 16 であり，残りの 6 個の値の平均値は 40，分散は 25 であった。このデータの平均値と分散を求めよ。

$\boxed{1}$

解答　(1)　順に 55 点，60 点，55 点，640；65 点，640

$\boxed{2}$

解答　〔図〕，B 弁当



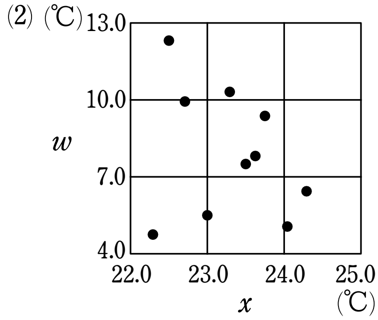
$\boxed{3}$

解答　(1)　順に　3.5 点，3 点，$\dfrac{\sqrt{3}}{2}$ 点　　(2)　0.65625

$\boxed{4}$

解答　(1)　23.3 ℃，A＝27.1，B＝30.8

(2)　〔図〕，(ウ)



$\boxed{5}$

解答　(1)　$\overline{u}=0.2$，$S_u{}^2=1.96$　　(2)　証明略，$\overline{x}=168.5$，$S_x{}^2=49$

$\boxed{6}$

解答　(1)　$\overline{x}=\dfrac{2p+q}{n}$，$V=\dfrac{4p+q}{n}-\left(\dfrac{2p+q}{n}\right)^2$

(2)　n が偶数のとき 1，n が奇数のとき $1-\dfrac{1}{n^2}$

$\boxed{7}$

解答　(1)　略　　(2)　225 人

$\boxed{8}$

解答　平均値は 36，分散は 45.4

1

このデータの平均値は

$$\frac{1}{10}(85+80+25+0+65+55+40+55+75+70)=55\ (点)$$

データを値の小さい順に並べると　0，25，40，55，55，65，70，75，80，85

このデータの中央値は，小さい方から 5 番目と 6 番目の値の平均値であるから

$$\frac{55+65}{2}=60\ (点)$$

このデータの最頻値は　55 点

このデータの分散は

$$\frac{1}{10}\{(85-55)^2+(80-55)^2+(25-55)^2+(0-55)^2+(65-55)^2+(55-55)^2+(40-55)^2+(55-55)^2+(75-55)^2+(70-55)^2\}$$

$$=\frac{6400}{10}=640$$

また，生徒全員に 10 点ずつ加点したときの平均値は

$$\frac{1}{10}(95+90+35+10+75+65+50+65+85+80)=65\ (点)$$

分散は

$$\frac{1}{10}\{(95-65)^2+(90-65)^2+(35-65)^2+(10-65)^2+(75-65)^2+(65-65)^2+(50-65)^2+(65-65)^2+(85-65)^2+(80-65)^2\}$$

$$=\frac{6400}{10}=640$$

別解　生徒全員に 10 点ずつ加点したときの平均値は，加点する前の平均値に 10 点加えた点と等しいから　65 点

また，分散は，データの散らばりの度合いを表す。

生徒全員に 10 点ずつ加点しても，データの散らばりの度合いは加点する前と変わらないから，加点後の分散は　640

2

A 弁当，B 弁当のデータを小さい順に並べると

A 弁当　16，21，22，24，27，28，33

B 弁当　13，17，18，22，28，32，35

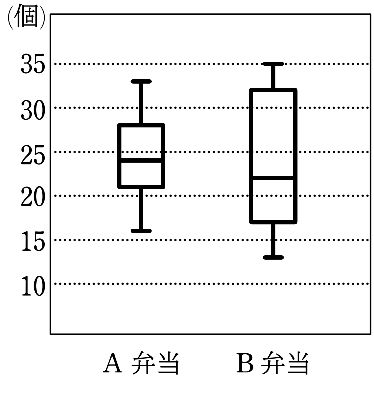
したがって，それぞれのデータの最小値，第 1 四分位数，中央値，第 3 四分位数，最大値は，順に

A 弁当　16，21，24，28，33

B 弁当　13，17，22，32，35

よって，箱ひげ図は [図] のようになる。

また，箱の長さ，全体の長さともに B 弁当の方が長いから，B 弁当のデータの方が散らばりの度合いが大きいと考えられる。



$\boxed{3}$

(1)　国語の得点が 4 点である生徒について，数学の得点 u と人数は右の表のようになる。

u	2	3	4	5	計
人数	1	5	3	3	12

よって，中央値は　$\dfrac{3+4}{2}=3.5$（点）

最頻値は　3 点

また，数学の得点が 2 点である生徒について，国語の得点 v と人数は右の表のようになる。

v	1	2	4	計
人数	2	5	1	8

よって，平均値は　$\dfrac{1\cdot2+2\cdot5+4\cdot1}{8}=2$（点）

ゆえに，標準偏差は　$\sqrt{\dfrac{\{(1-2)^2\cdot2+(2-2)^2\cdot5+(4-2)^2\cdot1\}}{8}}=\dfrac{\sqrt{3}}{2}$（点）

(2)　生徒 40 人の国語の得点 x の平均値 $\overline{x}$ は

$$\overline{x}=\frac{1\cdot5+2\cdot12+3\cdot6+4\cdot12+5\cdot5}{40}=3\text{（点）}$$

分散 $s_x{}^2$ は　$s_x{}^2=\dfrac{2^2\cdot(5+5)+1^2\cdot(12+12)}{40}=\dfrac{8}{5}$

同様に，生徒 40 人の数学の得点 y の平均値 $\overline{y}$ は

$$\overline{y}=\frac{1\cdot6+2\cdot8+3\cdot12+4\cdot8+5\cdot6}{40}=3\text{（点）}$$

分散 $s_y{}^2$ は　$s_y{}^2=\dfrac{2^2\cdot(6+6)+1^2\cdot(8+8)}{40}=\dfrac{8}{5}$

右の図のアミが掛かっていないマスに着目して，x と y の共分散 s_{xy} を計算すると

$$s_{xy}=\frac{1}{40}\{1\cdot(3+5)+2\cdot(2+3+2+2)$$
$$+4\cdot(2+3)+(-1)\cdot(1+3)\}$$
$$=\frac{21}{20}$$

数学＼国語	0	1	2	3	4	5
5				1	3	2
4			3		3	2
3			2	4	5	1
2		2	5		1	
1		3	2	1		
0						

よって，求める相関係数は

$$\frac{s_{xy}}{s_x\cdot s_y}=\frac{\dfrac{21}{20}}{\sqrt{\dfrac{8}{5}}\sqrt{\dfrac{8}{5}}}=\frac{21}{32}=0.65625$$

$\boxed{4}$

(1)　最低気温の平均値は

$$\frac{1}{10}(22.3+22.5+22.7+23.0+23.3+23.5+23.6+23.7+24.1+24.3)=23.3\text{（℃）}$$

参考　仮平均を 22.0 ℃ などに設定して計算するとよい。

また，最高気温の平均値が 31.2 ℃ であるから

$$\frac{1}{10}(\mathrm{A}+34.8+32.6+28.4+33.6+31.0+31.4+33.1+29.2+\mathrm{B})=31.2$$

ゆえに　　$\mathrm{A}+\mathrm{B}=57.9$　……①

さらに，最低気温と最高気温の相関係数が 0 であるから，その共分散は 0 である。

よって
$$\frac{1}{10}\{(22.3-23.3)(\mathrm{A}-31.2)+(22.5-23.3)(34.8-31.2)$$
$$+(22.7-23.3)(32.6-31.2)+(23.0-23.3)(28.4-31.2)$$
$$+(23.3-23.3)(33.6-31.2)+(23.5-23.3)(31.0-31.2)$$
$$+(23.6-23.3)(31.4-31.2)+(23.7-23.3)(33.1-31.2)$$
$$+(24.1-23.3)(29.2-31.2)+(24.3-23.3)(\mathrm{B}-31.2)\}$$
$$=0$$

ゆえに
$$(-10)\cdot(10\mathrm{A}-312)+(-8)\cdot 36+(-6)\cdot 14+(-3)\cdot(-28)+2\cdot(-2)+3\cdot 2$$
$$+4\cdot 19+8\cdot(-20)+10\cdot(10\mathrm{B}-312)=0$$

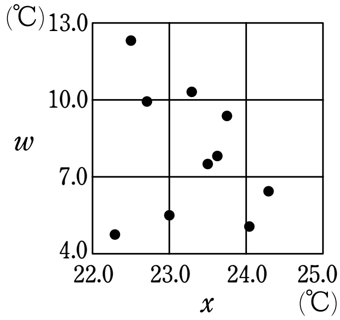
すなわち　$100(\mathrm{B}-\mathrm{A})=370$　　したがって　$\mathrm{B}-\mathrm{A}=3.7$ ……②

①，② から　$\mathrm{A}=27.1$，$\mathrm{B}=30.8$

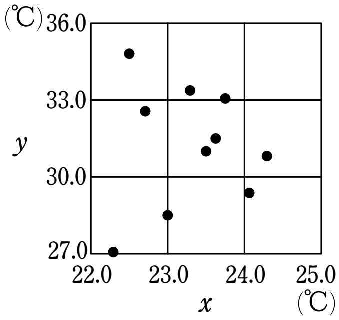
(2)　1 日の気温差 $w\ (=y-x)$ の行を加えた表は次のようになる。

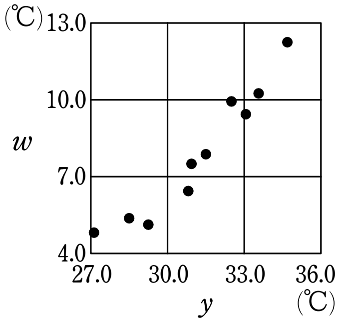
x	22.3	22.5	22.7	23.0	23.3	23.5	23.6	23.7	24.1	24.3
y	27.1	34.8	32.6	28.4	33.6	31.0	31.4	33.1	29.2	30.8
w	4.8	12.3	9.9	5.4	10.3	7.5	7.8	9.4	5.1	6.5

よって，変量 x と変量 w の散布図をかくと，右の図のようになる。



また，変量 x と変量 y の散布図，および変量 y と変量 w の散布図をかくと，次のようになる。





よって，変量 y と変量 w の間には正の相関があり，変量 x と変量 y，または変量 x と変量 w については相関がないといえる。

すなわち，(ア) ～ (ウ) のうち，「最高気温が高いほど 1 日の気温差が大きいという傾向がある」のみが正しいと言える。

したがって　(ウ)

5

(1) $u=\dfrac{x-167.5}{5}$ の値に対する度数は，右の表のようになる。

u	-3	-2	-1	0	1	2	3	計
度数	2	3	5	14	8	7	1	40

よって，u の平均値 $\overline{u}$ と分散 ${S_u}^2$ は

$$\overline{u}=\frac{1}{40}\{(-3)\cdot 2+(-2)\cdot 3+(-1)\cdot 5+0\cdot 14+1\cdot 8+2\cdot 7+3\cdot 1\}=\frac{1}{5}=0.2,$$

$${S_u}^2=\frac{1}{40}\{3^2\cdot(2+1)+2^2\cdot(3+7)+1^2\cdot(5+8)\}-\left(\frac{1}{5}\right)^2=\frac{49}{25}=1.96$$

(2) 階級値 $x_1=152.5$，……，$x_7=182.5$ に対して，

$$u_1=\frac{x_1-167.5}{5},\quad \cdots\cdots,\quad u_7=\frac{x_7-167.5}{5}$$

とし，その度数を f_1，……，f_7 とすると，

$$x_i=167.5+5u_i\quad (i=1,\ \cdots\cdots,\ 7),\qquad \sum_{i=1}^{7} f_i=40$$

であるから
$$\overline{x}=\frac{1}{40}\sum_{i=1}^{7} f_i x_i=\frac{1}{40}\sum_{i=1}^{7} f_i(167.5+5u_i)=167.5\cdot\frac{1}{40}\sum_{i=1}^{7} f_i+5\cdot\frac{1}{40}\sum_{i=1}^{7} f_i u_i$$
$$=167.5+5\overline{u},$$
$${S_x}^2=\frac{1}{40}\sum_{i=1}^{7} f_i(x_i-\overline{x})^2=\frac{1}{40}\sum_{i=1}^{7} f_i\{(167.5+5u_i)-(167.5+5\overline{u})\}^2$$
$$=\frac{1}{40}\sum_{i=1}^{7} f_i\{5(u_i-\overline{u})\}^2=5^2\cdot\frac{1}{40}\sum_{i=1}^{7} f_i(u_i-\overline{u})^2=25{S_u}^2$$

また，(1) より，$\overline{u}=\dfrac{1}{5}$，${S_u}^2=\dfrac{49}{25}$ であるから

$$\overline{x}=167.5+5\cdot\frac{1}{5}=168.5,\qquad {S_x}^2=25\cdot\frac{49}{25}=49$$

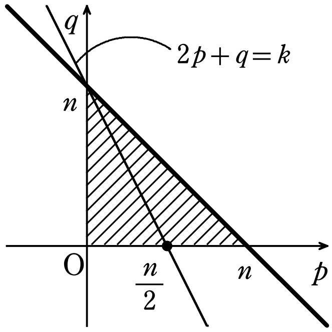
$\boxed{6}$

(1) $\overline{x}=\dfrac{2\cdot p+1\cdot q+0\cdot(n-p-q)}{n}=\dfrac{2p+q}{n}$,

$$V=\frac{2^2\cdot p+1^2\cdot q+0^2\cdot(m-p-q)}{n}-(\overline{x})^2=\frac{4p+q}{n}-\left(\frac{2p+q}{n}\right)^2$$

(2) (1) より　　$V=-\dfrac{1}{n^2}\{(2p+q)-n\}^2+1-\dfrac{q}{n}$　$\cdots\cdots$ ①

$0\leqq p\leqq n$，$0\leqq q\leqq n$ かつ $p+q\leqq n$ であるから，点 $(p,\ q)$ が右の図の斜線部分の領域に含まれる格子点上を動くときの V の最大値を求めればよい。



ここで，$2p+q=k$ とおくと　$q=-2p+k$

pq 平面において，直線 $q=-2p+k$ は，傾きが -2 で q 切片が k の直線を表す。

① より，k の値が n に近くなるほど V の値が大きくなり，さらに，そのとき，q の値が小さいほど V の値は大きくなる。

[1]　n が偶数の場合

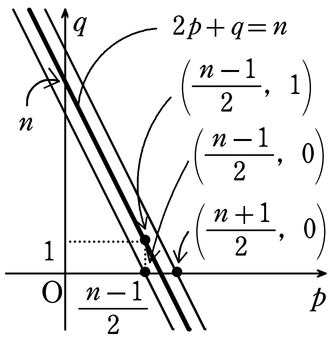
$(p,\ q)=\left(\dfrac{n}{2},\ 0\right)$ のとき，V は最大となり，その最大値は　1

[2]　n が奇数の場合

V が最大となる p，q の組の候補は，

$(p,\ q)=\left(\dfrac{n\pm1}{2},\ 0\right)$，$\left(\dfrac{n-1}{2},\ 1\right)$

の 3 組である。



$(p,\ q)=\left(\dfrac{n\pm1}{2},\ 0\right)$ のとき　$V=1-\dfrac{1}{n^2}$

$(p,\ q)=\left(\dfrac{n-1}{2},\ 1\right)$ のとき　$V=1-\dfrac{1}{n}$

$1-\dfrac{1}{n}\leqq1-\dfrac{1}{n^2}$ であるから，V は，

$(p,\ q)=\left(\dfrac{n\pm1}{2},\ 0\right)$ のとき最大値 $1-\dfrac{1}{n^2}$

をとる。

[1]，[2] より，n が偶数のとき，最大値　1

n が奇数のとき，最大値　$1-\dfrac{1}{n^2}$

$\boxed{7}$

(1)　必要ならば x_1，……，x_n の順番を並びかえて，

$$k=1,\ \cdots\cdots,\ K(p) \text{ のとき } \quad |x_k-\overline{x}| \geqq ps,$$

$$k=K(p)+1,\ \cdots\cdots,\ n \text{ のとき } \quad |x_k-\overline{x}| < ps$$

が成り立つとしてよい。

このとき，x_1，……，x_n の分散 s^2 について

$$s^2=\frac{1}{n}\sum_{k=1}^{n}(x_k-\overline{x})^2 \geqq \frac{1}{n}\sum_{k=1}^{K(p)}|x_k-\overline{x}|^2 \geqq \frac{1}{n}\sum_{k=1}^{K(p)}(ps)^2=\frac{1}{n}p^2s^2K(p)$$

$s^2>0$ であるから，$s^2 \geqq \dfrac{1}{n}p^2s^2K(p)$ より　　$K(p) \leqq \dfrac{n}{p^2}$

(2)　(1) において，$n=300$，$\overline{x}=63$，$s=8$，$p=2$ と考える。

$|x_k-\overline{x}|<ps$ を満たす x_k の個数を N とすると，(1) より

$$N=n-K(p) \geqq n-\frac{n}{p^2}=n\left(1-\frac{1}{p^2}\right)$$

よって，$47<x_k<79$ すなわち $|x_k-63|<2\times 8$ を満たす x_k の個数 N は

$$N \geqq 300\times\left(1-\frac{1}{2^2}\right)=225$$

を満たす。

したがって，得点が 47 点より高く 79 点より低い生徒は少なくとも 225 人いる。

$\boxed{8}$

このデータの平均値は　　$\dfrac{30\times 4+40\times 6}{10}=\dfrac{360}{10}=36$

また，4 個の値の 2 乗の平均値を a とすると　　$a=30^2+16=916$

残りの 6 個の値の 2 乗の平均値を b とすると　　$b=40^2+25=1625$

したがって，このデータの分散は　　$\dfrac{916\times 4+1625\times 6}{10}-36^2=45.4$

データの分析 Σで表すと….

nコのデータ $x_1, x_2, x_3, \cdots, x_n$

平均値 $\bar{x} = \frac{1}{n}\sum_{k=1}^{n} x_k$

分散 ① $S_x^2 = \frac{1}{n}\sum_{k=1}^{n}(\underbrace{x_k - \bar{x}}_{\text{偏差}})^2$

$= \cdots\cdots$

② $= \underbrace{\frac{1}{n}\sum_{k=1}^{n} x_k^2}_{\overline{x^2}} - \bar{x}^2$

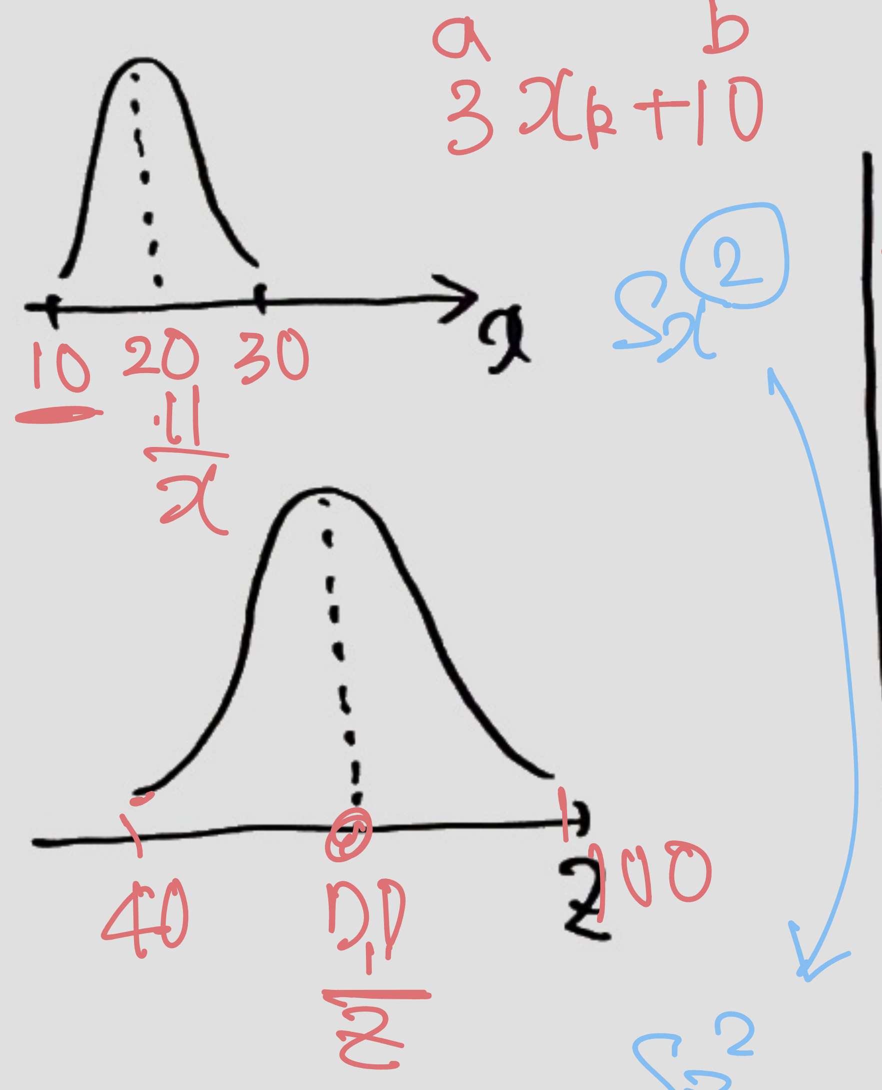
S_x：標準偏差

変数変換 $z_k = ax_k + b \quad (a \neq 0)$

平均値 $\bar{z} = \boxed{a \cdot \bar{x} + b}$ ← $\bar{x}\ \bar{z}$

分散 $S_z^2 = \boxed{a^2 \cdot S_x^2}$ ← S_x^2 z"

★標準偏差 $S_z = |a| S_x$

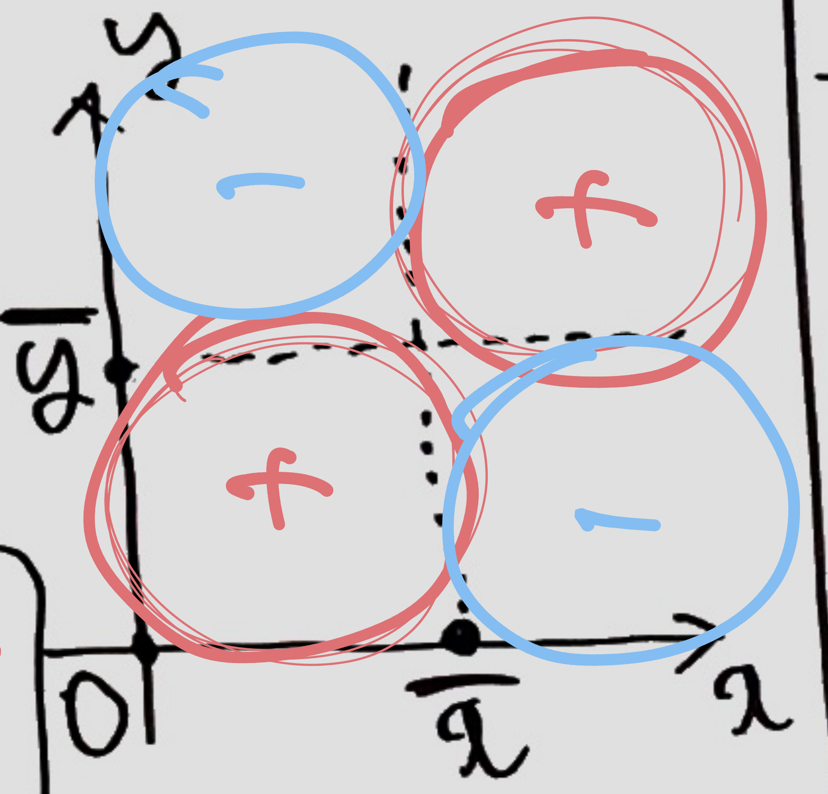


nコのデータの組 $(x_1, y_1), (x_2, y_2), \cdots (x_n, y_n)$

共分散 $C_{xy} = \frac{1}{n}\sum_{k=1}^{n}(x_k - \bar{x})(y_k - \bar{y})$

相関係数 $r_{xy} = \frac{C_{xy}}{S_x S_y}$

性質 $-1 \leqq r \leqq 1$



変数変換の影響 数り……英 $x \longrightarrow z$

共分散

$C_{zy} = \boxed{a}\, C_{xy}$

$r_{zy} = \boxed{1}\, r_{xy}$

類 2018 IA 本

次回 データの合併 例 2020 岩手医

$\{x_1, x_2, \cdots, x_m\} \cup \{y_1, y_2, \cdots, y_n\}$ mコ+nコ

$\Rightarrow \{z_1\ z_2, \cdots z_m, z_{m+1}, z_{m+2}, \cdots, z_{m+n}\}$

データの分析　Σで表すと....

nコのデータ　$x_1, x_2, x_3, \cdots, x_n$

平均値　$\overline{x} = \frac{1}{n}\sum_{k=1}^{n} x_k$

偏差

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

定義

$= \cdots\cdots$

$\boxed{\overline{x^2} - \overline{x}^2}$

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

性質

s_x：標準偏差

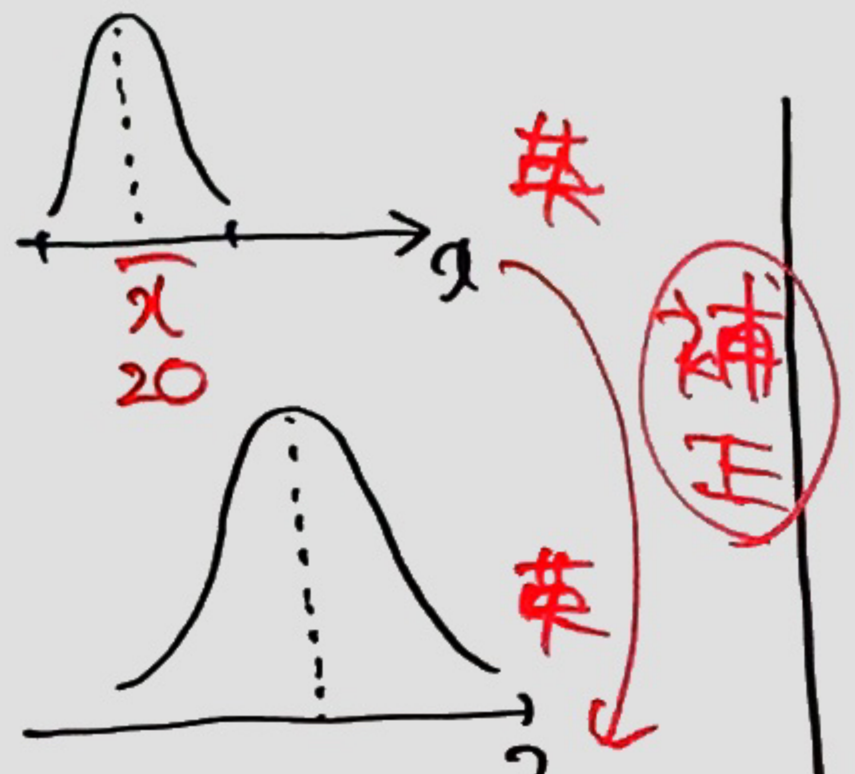
変数変換

$z_k = \overset{3x}{a} x_k + \overset{+10}{b}$　$(a \neq 0)$とおく

平均値　$\overline{z} = \boxed{a \cdot \overline{x} + b}$ ← $\overline{x}$で

分散　$s_z^2 = \boxed{a^2 \cdot s_x^2}$ ← s_x^2で

$s_z = |a|\, s_x$



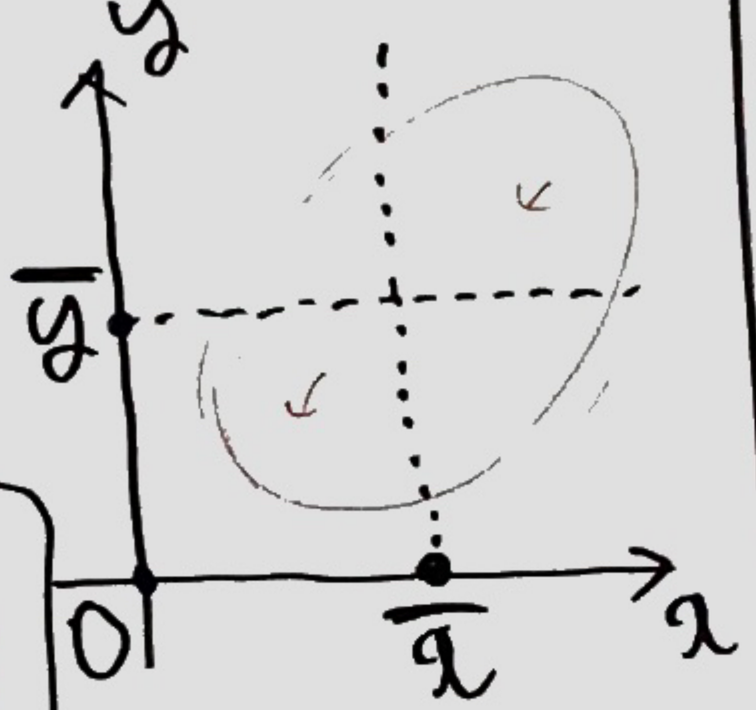
nコのデータの組　$(x_1, y_1), (x_2, y_2), \cdots (x_n, y_n)$

共分散　$C_{xy} = \frac{1}{n}\sum_{k=1}^{n}(x_k - \overline{x})(y_k - \overline{y})$

相関係数　$r_{xy} = \frac{C_{xy}}{s_x \cdot s_y}$　性質 $\boxed{-1 \le r \le 1}$

証　CS不等式

$z_k - \overline{z}$



変数変換の影響

スキージャンプ

$C_{zy} = \boxed{a} \cdot C_{xy}$

$r_{zy} = \boxed{1}\, r_{xy}$

（相関係数は 1次の変数変換に 依存しない）

類　2018 IA 本

次回　データの合併　例　2020 岩手医

$\{x_1, x_2, \cdots, x_m\} \cup \{y_1, y_2, \cdots, y_n\}$　mコ+nコ

$\Rightarrow \{z_1, z_2, \cdots z_m, z_{m+1}, z_{m+2}, \cdots, z_{m+n}\}$