医歯薬系のためのデータの分析 (2019〜2015)

ver 2019/10/25
hiraocafe.com

新課程のデータの分析の 2018〜2015 年入試で，主要な医歯薬系の大学で出題された解いておきたいと思うものを挙げ，自分なりに基礎 (ほぼ数 I 範囲のみ) と応用 (他の分野との融合あり) に分けてみました．2018 年はある程度出題されました．2017 年はほとんど出題されず，2016 年は多くの大学でこの単元が出題されました．2015 年は主要大学は，一橋と東邦大医学部のみ出題のはずです (横浜市立大医学部も若干データの分析の背景をもつ大問あり)．

この分野は医歯薬系と文系の社会科学系で出題されているイメージがあります．今のところ理工系はあまり見かけません．

ちまたの多くの問題集でデータの分析が軽んじられていることに鑑み，編集してみました．お役に立てば幸いです．

(基礎)

$\boxed{\quad 1 \quad}$

次の 5 つのデータがあった．

5，2，8，10，5

(1) このデータの第 1 四分位数と中央値を求めよ．

(2) このデータの分散を求めよ．

(2016 神戸薬科大薬学部)

$\boxed{\quad 2 \quad}$

次のデータは，生徒 8 人に 5 点満点の数学の小テストを行った結果である．

3，1，0，4，2，4，2，0　(点)

(1) このデータの四分位範囲を求めよ．

(2) このデータの標準偏差を求めよ．

(2016 明海大歯学部 A 日程)

$\boxed{\quad 3 \quad}$

5 個のデータ 17，x，18，19，9 の平均値が 15 であるとき，分散は $\boxed{\text{(あ)}}.\boxed{\text{(い)}}$ である．

(2016 岩手医科大歯・薬学部前期)

4

70 より大きい 2 桁の素数の値すべてからなる 1 組のデータがある．ただし，同じ値は重複していない．このデータの標準偏差を求めよ．

(2016 藤田保健衛生大医学部前期)

5

6 個の値 5，8，4，2，a，b からなるデータの平均値と中央値がともに 6 であるとき，a，b の値を求めると $a=$ (あ) ，$b=$ (い) である．ただし，$a<b$ とする．

(2016 山梨大医学部後期)

6

右の表は，ある中学校の 5 人の生徒 A，B，C，D、E に 2 つの科目の小テストを行った結果である．2 つの科目の得点をそれぞれ x，y とする．このとき x の分散を求めると (あ) であり，x と y の共分散を求めると (い) である．

	A	B	C	D	E
x	7	3	5	2	3
y	4	5	7	3	6

(2016 福岡大薬学部)

7

次のデータの相関係数を求めよ．

x	8	4	2	6	10
y	4	5	6	3	2

(2016 奈良県立医科大医学部前期)

8

生徒 50 人に行ったテストの得点を x_1，x_2，$\cdots$，x_{50} とする．得点の平均は 42，分散は 36 であった．このとき，$z_i=\dfrac{1}{6}x_i-7\ (i=1,2,\cdots,50)$ とおくと z_1，z_2，$\cdots$，z_{50} の分散を求めよ．

(2018 聖マリアンナ医科大医学部)

$\boxed{9}$

2 つの変量 x，y の 16 個のデータ (x_1, y_1)，(x_2, y_2)，$\cdots$，(x_{16}, y_{16}) が

$$x_1 + x_2 + \cdots + x_{16} = 72$$
$$y_1 + y_2 + \cdots + y_{16} = 120$$
$$x_1^2 + x_2^2 + \cdots + x_{16}^2 = 349$$
$$y_1^2 + y_2^2 + \cdots + y_{16}^2 = 925$$
$$x_1y_1 + x_2y_2 + \cdots + x_{16}y_{16} = 545$$

を満たしているとき，次の問いに小数で答えよ．

(1) 変量 x，y のデータの平均をそれぞれ $\overline{x}$，$\overline{y}$ とすると，$\overline{x} = \boxed{\text{ア}}.\boxed{\text{イ}}$，$\overline{y} = \boxed{\text{ウ}}.\boxed{\text{エ}}$ である．

(2) 変量 x, y のデータの標準偏差をそれぞれ s_x, s_y とすると，$s_x = \boxed{\text{オ}}.\boxed{\text{カキ}}$, $s_y = \boxed{\text{ク}}.\boxed{\text{ケコ}}$ である．また，変量 x，y のデータの共分散を s_{xy} とすると，$s_{xy} = \boxed{\text{サ}}.\boxed{\text{シスセソ}}$ である．

(3) 変量 x，y のデータの相関係数を r とすると，$r = \boxed{\text{タ}}.\boxed{\text{チ}}$ である．

(2016 星薬科大薬学部)

$\boxed{10}$

ある大学で N 人の学生が数学を受験した．その得点を x_1，x_2，$\cdots$，x_N とする．平均値 $\overline{x}$ および分散 s^2 は

$$\overline{x} = \frac{x_1 + x_2 + \cdots + x_N}{N}$$
$$s^2 = \frac{(x_1 - \overline{x})^2 + (x_2 - \overline{x})^2 + \cdots + (x_N - \overline{x})^2}{N}$$

で与えられる．標準偏差 $s\ (>0)$ は　$s = \sqrt{s^2}$

となる．このとき x 点を取った学生の偏差値は　$t = 50 + 10 \times \dfrac{x - \overline{x}}{s}$

で与えられる $(x \in \{x_1, x_2, \cdots, x_N\})$．偏差値は無単位であることに注意せよ．

Y 大学で $N = 3n$ 人の学生が数学を受験し，たまたま $2n$ 人の学生が a 点，残りの n 人の学生が b 点を取ったとしよう．簡単にするため，$a < b$ とする．a 点を取った学生および b 点をとった学生の偏差値を求めよ．

(2016 横浜市立大医学部)

(応用)

11

2 つの変量 x，y が右表で与えられているとき，以下の問いに答えよ．

No.	1	2	3	$\cdots$	n
x	1	3	5	$\cdots$	$2n-1$
y	2	4	6	$\cdots$	$2n$

(1) 変量 x の平均値 m_x と分散 s_x^2 を求めよ．

(2) 変量 x と変量 y の相関係数 r を求めよ．

(3) n 個の変量 x に，平均値 $2n$，分散 $4n^2$ からなる n 個のデータを加えた．この $2n$ 個からなるデータの平均値 m_x' と分散 ${s_x'}^2$ を求めよ．

(2016 岐阜薬科大薬学部中期)

12

平均値と中央値は共に代表値であり，求め方は全く異なるが比較的近い値であることが多い．いま，偶数個の身長のデータがあり，その最小値は $m=140$cm，最大値は $M=180$cm である．このデータの中央値が $A=150$cm のとき，半数のデータは m 以上 A 以下の値であり，残る半数のデータは A 以上 M 以下である．このことから平均値 $\overline{x}$ のとる値の範囲は $\boxed{\text{(あ)}}$ である．また，平均値と中央値の関係を用いると，最小値が $m=140$cm，最大値は $M=180$cm である偶数個のデータの平均値が $\overline{x}=170$cm であるとき，中央値 A のとる値の範囲は $\boxed{\text{(い)}}$ である．

(2016 福岡大医学部)

13

次のデータは，ある高校 3 年生 9 人の 100 点満点の試験の結果である．

65，83，64，69，89，68，77，70，81

データを順に，x_1，x_2，x_3，$\cdots$，x_9 と表す．このとき，$\sum_{i=1}^{9}(x_i-\theta)^2$ を最小にする θ の値は $\boxed{\text{アイ}}$ である．また，$\sum_{i=1}^{9}|x_i-\theta|$ を最小にする θ の値は $\boxed{\text{ウエ}}$ である．

(2015 東邦大医学部)

$\boxed{\quad 14 \quad}$

2 つの変量をもつ 100 個のデータ (x_1, y_1)，(x_2, y_2)，$\cdots$，(x_{100}, y_{100}) が，

$$\sum_{i=1}^{100} x_i^2 = 500, \qquad \sum_{i=1}^{100} y_i^2 = 900, \qquad \sum_{i=1}^{100} x_i y_i = 500$$

を満たす場合を考える．$X = \dfrac{1}{100}\displaystyle\sum_{i=1}^{100} x_i$ および $Y = \dfrac{1}{100}\displaystyle\sum_{i=1}^{100} y_i$ とするとき，点 (X, Y) の存在範囲は不等式 $\dfrac{(Y-X)^2}{\boxed{\text{ア}}} + \dfrac{X^2}{\boxed{\text{イ}}} \leqq 1$ の表す領域である．また $|X+Y|$ のとり得る値の範囲は $0 \leqq |X+Y| \leqq \boxed{\text{ウ}}\sqrt{\boxed{\text{エ}}}$ である．

(2016 東邦大医学部)

$\boxed{\quad 15 \quad}$

n は 2 以上の整数とする．変量 x についてのデータの値を x_k $(1 \leqq k \leqq n)$ とし，変量 y についてのデータの値を y_k $(1 \leqq k \leqq n)$ とする．変量 z はデータの値が $x_k y_k$ $(1 \leqq k \leqq n)$ である変量を表す．

(1) 変量 x と y の n 個の値の組を (x_k, y_k) $(1 \leqq k \leqq n)$ としたときの x と y の共分散 s_{xy} (偏差の積の平均) について $s_{xy} = \overline{z} - \overline{x}\,\overline{y}$ が成り立つことを証明せよ．ここで，$\overline{x}$，$\overline{y}$，$\overline{z}$ はそれぞれ変量 x，y，z についてのデータの値の平均値を表す．

0 以上の整数 a と 1 以上の整数 b に対し，a を b で割った余りを $R_b(a)$ と表す．l，m は 2 以上 n 以下の整数とする．変量 x と変量 y の n 個の値の組を

$$(x_k, y_k) = (R_l(k-1)+1, R_m(k-1)+1) \ (1 \leqq k \leqq n)$$

としたときの x と y の相関係数を r とする．

(2) l は n の約数とし，$m = n$ であるとき，r を求めよ．

(3) $n = l(l+1)$ とし，$m = l+1$ であるとき，r を求めよ．

(2016 京都府立医科大医学部前期)

応用で他に良問を挙げるとすれば，2018 福岡大医学部，2016 広島大文系，2015 一橋大を解いておくといいと思います．2019 聖マリアンナ医科大は $\boxed{\quad 13 \quad}$ の簡単版．

解答

$\boxed{1}$

小さい順に並べると，2,5,5,8,10

(1) $Q_1 = \dfrac{2+5}{2} = \mathbf{3.5}$，(中央値) $= \mathbf{5}$

(2) 平均が 6(整数) なので定義を使う．

(分散)

$= \dfrac{1}{5}\{(2-6)^2+(5-6)^2+\cdots+(10-6)^2\}$

$= \mathbf{7.6}$

$\boxed{2}$

小さい順に並べると，0,0,1,2,2,3,4,4

(1) $Q_1 = 0.5$，$Q_3 = 3.5$ より，

(四分位範囲)$= 3.5 - 0.5 = \mathbf{3}$

(2) 平均が 2 だが 0 点あるので定理を使う．

(分散)

$= \dfrac{3^2+1^2+\cdots+0^2}{8} - 2^2$

$= 2.25$

$\therefore$ (標準偏差) $= \mathbf{1.5}$

補足　定義と定理どちら使うかは好み．

$\boxed{3}$

$\dfrac{17+x+18+19+9}{5} = 15$ より，$x = 12$

平均が 15 なので定義を使う．

(分散)

$= \dfrac{1}{5}\{(17-15)^2+(12-15)^2+\cdots+(9-15)^2\}$

$= \mathbf{14.8}$

$\boxed{4}$

70 より大きい 2 桁の素数は

71,73,79,83,89,97

なので，平均は 82.

(分散)

$= \dfrac{1}{6}\{(71-82)^2+(73-82)^2+\cdots+(97-83)^2\}$

$= 81$

$\therefore$ (標準偏差) $= \mathbf{9}$

$\boxed{5}$

$\dfrac{5+8+4+2+a+b}{6} = 6$ より，$a+b = 17$.

順番は 2,4,5,a,8,b になるので

(中央値) $= \dfrac{5+a}{2} = 6$ より，$a = \mathbf{7}$，$b = \mathbf{10}$

$\boxed{6}$

$\overline{x} = 4$ になるので，

$s_x^2 = \dfrac{1}{5}\{(7-4)^2+(3-4)^2+\cdots+(3-4)^2\} = \mathbf{\dfrac{16}{5}}$

$\overline{y} = 5$ になるので，

(共分散)

$= \dfrac{1}{5}\{(7-4)(4-5)+(3-4)(5-5)+$

$\cdots+(3-4)(6-5)\}$

$= \mathbf{\dfrac{2}{5}}$

$\boxed{7}$

$\overline{x} = 6$，$\overline{y} = 4$ になるので，定義通り計算すると

$s_x^2 = 8$，$s_y^2 = 2$.

s_{xy}

$= \dfrac{1}{5}\{(8-6)(4-4)+(4-6)(5-4)+$

$\cdots+(10-6)(2-4)\}$

$= -\dfrac{18}{5}$

$\therefore r = \dfrac{s_{xy}}{s_x s_y} = -\dfrac{18}{5\sqrt{8}\sqrt{2}} = \mathbf{-\dfrac{9}{10}}$

$\boxed{8}$

すべての値から 7 引いても分散は変わらない．変数を $\dfrac{1}{6}$ 倍すると，分散は $\dfrac{1}{36}$ 倍より，z_i の分散は **1**

※ (上の証明)

$s_z^2 = \displaystyle\sum_{k=1}^{50}\left\{z_i - \overline{z}\right\}^2$

$= \displaystyle\sum_{k=1}^{50}\left\{\frac{1}{6}x_i - 7 - \overline{\frac{1}{6}x - 7}\right\}^2$

$= \displaystyle\sum_{k=1}^{50}\left\{\frac{1}{6}x_i - 7 - \left(\frac{1}{6}\overline{x} - 7\right)\right\}^2$

$= \displaystyle\frac{1}{36}\sum_{k=1}^{50}\left\{x_i - \overline{x}\right\}^2$

$$= \frac{1}{36}s_x^2 = \mathbf{1}$$

$\boxed{\quad 9 \quad}$

(1) $\overline{x} = \dfrac{72}{16} = \mathbf{4.5},\ \ \overline{y} = \dfrac{120}{16} = \mathbf{7.5}$

(2)

$$s_x^2 = \overline{x^2} - (\overline{x})^2$$
$$= \frac{349}{16} - \left(\frac{9}{2}\right)^2$$
$$= \frac{25}{16} \qquad \therefore s_x = \frac{5}{4} = \mathbf{1.25}$$

$$s_y^2 = \overline{y^2} - (\overline{y})^2$$
$$= \frac{925}{16} - \left(\frac{15}{2}\right)^2$$
$$= \frac{25}{16} \qquad \therefore s_y = \frac{5}{4} = \mathbf{1.25}$$

$s_{xy} = \overline{xy} - \overline{x} \cdot \overline{y}$ ←定理（要暗記！）

$$= \frac{545}{16} - \frac{9}{2} \cdot \frac{15}{2}$$
$$= \frac{5}{16} = \mathbf{0.3125}$$

(3)

$$r = \frac{s_{xy}}{s_x s_y} = \frac{\frac{5}{16}}{\frac{5}{4} \cdot \frac{5}{4}} = \mathbf{0.2}$$

$\boxed{\quad 10 \quad}$

$$\overline{x} = \frac{2na+nb}{3n} = \frac{2a+b}{3}$$

$\overline{x^2} = \dfrac{2na^2+nb^2}{3n} = \dfrac{2a^2+b^2}{3}$ より

$$s^2 = \overline{x^2} - (\overline{x})^2$$
$$= \frac{2a^2+b^2}{3} - \left(\frac{2a+b}{3}\right)^2$$
$$= \frac{2a^2-4ab+2b^2}{9}$$
$$= \frac{2}{9}(a-b)^2$$

$\therefore s = \dfrac{\sqrt{2}}{3}(b-a)\ \ (\because a < b)$

となるので

$$t_a = 50 + 10 \times \frac{a - \frac{2a+b}{3}}{s}$$
$$= 50 + 10 \times \frac{1}{3}(a-b) \times \frac{3}{\sqrt{2}(b-a)}$$
$$= \mathbf{50 - 5\sqrt{2}}$$

偏差値の平均は 50 なので

$t_b = \mathbf{50 + 10\sqrt{2}}$

$\boxed{\quad 11 \quad}$

(1)

$$m_x = \frac{1}{n}\sum_{k=1}^{n}(2k-1)$$
$$= \frac{1}{n}\{n(n+1) - n\} = \boldsymbol{n}$$

$$s_x^2$$
$$= \overline{x^2} - (\overline{x})^2$$
$$= \frac{1}{n}\sum_{k=1}^{n}(2k-1)^2 - n^2$$
$$= \frac{1}{n}\sum_{k=1}^{n}(4k^2-4k+1) - n^2$$
$$= \frac{1}{n}\left\{4\frac{1}{6}n(n+1)(2n+1) - 4\frac{1}{2}n(n+1) + n\right\} - n^2$$
$$= \boldsymbol{\frac{n^2-1}{3}}$$

(2)

$$m_y = \frac{1}{n}\sum_{k=1}^{n}2k = n+1$$

$$s_y^2$$
$$= \overline{y^2} - (\overline{y})^2$$
$$= \frac{1}{n}\sum_{k=1}^{n}(2k)^2 - (n+1)^2$$
$$= \frac{4}{6}(n+1)(2n+1) - (n+1)^2$$
$$= \frac{n^2-1}{3}$$

$$s_{xy}$$
$$= \frac{1}{n}\sum_{k=1}^{n}(2k-1-m_x)(2k-m_y)$$
$$= \frac{1}{n}\sum_{k=1}^{n}(2k-1)2k - m_y\frac{1}{n}\sum_{k=1}^{n}(2k-1) - m_x\frac{1}{n}\sum_{k=1}^{n}2k + m_x m_y$$
$$= \frac{1}{n}\sum_{k=1}^{n}(4k^2-2k) - m_y m_x - m_x m_y + m_x m_y$$
$$= \frac{4}{6}(n+1)(2n+1) - (n+1) - m_x m_y$$

$$= (n+1)\left(\frac{4n+2}{3} - 1 - n\right) = \frac{n^2-1}{3}$$

$$\therefore r = \frac{s_{xy}}{s_x s_y} = \frac{\frac{n^2-1}{3}}{\sqrt{\frac{n^2-1}{3}}\sqrt{\frac{n^2-1}{3}}} = \mathbf{1}$$

補足　減点の恐れを除くために念のため，共分散の定理を定義から導くような形で解いた．試験時間がない場合はいきなり $s_{xy} = \overline{xy} - \overline{x}\cdot\overline{y}$ を使うとよい．

(3)

$$m'_x = \frac{nm_x + n\cdot 2n}{2n} = \boldsymbol{\frac{3n}{2}}$$

$$\begin{aligned}&(\text{追加した } x \text{ の 2 乗の和})\\ &= n\times(\text{追加した } x \text{ の 2 乗の平均})\\ &= n\{4n^2 + (2n)^2\}\\ &= 8n^3\end{aligned}$$

より

$$\begin{aligned}&s'^2_x\\ &= \frac{\text{もとの } x \text{ の 2 乗の和} + \text{追加した } x \text{ の 2 乗の和}}{2n} - (m'_x)^2\\ &= \frac{n\overline{x^2} + 8n^3}{2n} - \left(\frac{3n}{2}\right)^2\\ &= \frac{n(s_x^2 + m_x^2) + 8n^3}{2n} - \frac{9n^2}{4}\\ &= \boldsymbol{\frac{29n^2-2}{12}}\end{aligned}$$

$\boxed{12}$

データを $2n$ $(n\in\mathbb{N})$ 個とする．

$\overline{x}$ が最大となるとき | で下位と上位のデータを区切ると

$$mA\cdots A|AM\cdots M$$

となるので

$$\begin{aligned}&\overline{x}\\ &= \frac{m + nA + (n-1)M}{2n}\\ &= \frac{A+M}{2} - \frac{M-m}{2n}\\ &\therefore \overline{x} < \frac{A+M}{2} = 165\end{aligned}$$

$\overline{x}$ が最小となるとき | で下位と上位のデータを区切ると

$$m\cdots mA|A\cdots AM$$

となるので

$$\begin{aligned}&\overline{x}\\ &= \frac{(n-1)m + nA + M}{2n}\\ &= \frac{m+A}{2} + \frac{M-m}{2n}\\ &\therefore \overline{x} > \frac{m+A}{2} = 145\end{aligned}$$

以上より

$$\therefore \mathbf{145 < \overline{x} < 165}$$

また中央値の範囲に関しては

$$\frac{m+A}{2} < \overline{x} < \frac{A+M}{2}$$

より

$$\frac{140+A}{2} < 170 < \frac{A+180}{2}$$

$$\Longleftrightarrow 160 < A < 200$$

しかし当然 $A \leqq M$ なので

$$\therefore \mathbf{160 < A \leqq 180}$$

$\boxed{13}$

$$\begin{aligned}&\sum_{i=1}^{9}(x_i-\theta)^2\\ &= \sum_{i=1}^{9}(x_i^2 - 2\theta x_i + \theta^2)\\ &= 9\theta^2 - 2\left(\sum_{i=1}^{9}x_i\right)\theta + \sum_{i=1}^{9}x_i^2\\ &= 9\left(\theta - \frac{1}{9}\sum_{i=1}^{9}x_i\right)^2 - \frac{1}{9}\left(\sum_{i=1}^{9}x_i\right)^2 + \sum_{i=1}^{9}x_i^2\\ &= 9\left(\theta - \overline{x}\right)^2 + 9\left\{\frac{1}{9}\sum_{i=1}^{9}x_i^2 - \left(\frac{1}{9}\sum_{i=1}^{9}x_i\right)^2\right\}\\ &= 9\left(\theta - \overline{x}\right)^2 + 9s_x^2\end{aligned}$$

よって，$\theta = \overline{x} = \frac{1}{9}(64 + \cdots + 89) = \mathbf{74}$ のとき最小．

補足　一般に $\frac{1}{n}\sum_{i=1}^{n}(x_i-\theta)^2$ を最小にしそうな θ は中央値なども候補にあがるはず．しかし平方完成すれば，θ は $\overline{x}$ のときに最小になるとわかり，そのときの $\frac{1}{n}\sum_{i=1}^{n}(x_i-\theta)^2$ を分散という．この知識が

あれば平方完成せずともすぐ答えられるでしょう．

データを小さい順に並べたものを X_i とおく．

$f(\theta)=\sum_{i=1}^{9}|x_i-\theta|=\sum_{i=1}^{9}|X_i-\theta|$ とおくと

(i) $\theta \leqq X_1$ のとき

$$f(\theta)=\sum_{i=1}^{9}(X_i-\theta)$$

$$=-9\theta+\sum_{i=1}^{9}X_i$$

(ii) $X_1 \leqq \theta \leqq X_2$ のとき

$$f(\theta)=\theta-X_1+\sum_{i=2}^{9}(X_i-\theta)$$

$$=-7\theta+\sum_{i=2}^{9}X_i-X_1$$

(iii) $X_2 \leqq \theta \leqq X_3$ のとき

$$f(\theta)=\theta-X_1+\theta-X_2+\sum_{i=3}^{9}(X_i-\theta)$$

$$=-5\theta+\sum_{i=3}^{9}X_i-X_1-X_2$$

同様に

(iv) $X_3 \leqq \theta \leqq X_4$ のとき

$$f(\theta)=-3\theta+\sum_{i=4}^{9}X_i-\sum_{i=1}^{3}X_i$$

(v) $X_4 \leqq \theta \leqq X_5$ のとき

$$f(\theta)=-\theta+\sum_{i=5}^{9}X_i-\sum_{i=1}^{4}X_i$$

(vi) $X_5 \leqq \theta \leqq X_6$ のとき

$$f(\theta)=\theta+\sum_{i=6}^{9}X_i-\sum_{i=1}^{5}X_i$$

(vii) $X_6 \leqq \theta \leqq X_7$ のとき

$$f(\theta)=3\theta+\sum_{i=7}^{9}X_i-\sum_{i=1}^{6}X_i$$

$$\vdots$$

と続く．$f(\theta)$ は折れ線グラフになり，1 つ 1 つの場合分けでは θ の 1 次関数である．傾きが負から正に変わる，$\theta=X_5=$ (中央値) $=\mathbf{70}$ のとき最小になる．

補足　一般に $\frac{1}{n}\sum_{i=1}^{n}|x_i-\theta|$ を最小にしそうな θ は平均値なども候補にあがるはず．しかし上のように場合分けして折れ線グラフで考えれば，n が奇数のときは θ は中央値が最小にし，n が偶数のときは，θ は下位のデータの最大値以上，上位のデータの最小値以下の値であればよいことがわかる．つまり，n が偶数のときも θ は中央値でよい．そのときの $\frac{1}{n}\sum_{i=1}^{n}|x_i-\theta|$ のことを中央値絶対偏差の和と呼びます．この知識があればこんな場合分けなどせずともすぐ答えられるでしょう．

分散も中央値絶対偏差の和もどちらもデータの散らばりを表す統計量です．しかし，それぞれに統計学的な良さがあり，特に後者の方は，外れ値 (極端な異常値) の影響を受けにくいというメリットがあります．

$\boxed{14}$

$\overline{x^2}=5,\ \overline{y^2}=9,\ \overline{xy}=5$

$s_{xy}=\overline{xy}-\overline{x}\cdot\overline{y}$　←定理（要暗記！）

$\quad=5-XY$

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

$\quad=5-X^2$

$s_y^2=9-Y^2.$

これより

$$|r|=\left|\frac{5-XY}{\sqrt{5-X^2}\sqrt{9-Y^2}}\right|\leqq 1$$

$$\iff |5-XY|\leqq\left|\sqrt{5-X^2}\sqrt{9-Y^2}\right|$$

2 乗すると

$$25-10XY+X^2Y^2\leqq 45-9X^2-5Y^2+X^2Y^2$$

$$\iff 9X^2-10XY+5Y^2-20\leqq 0$$

$$\iff 5(X-Y)^2+4X^2-20\leqq 0$$

$$\therefore\ \frac{(Y-X)^2}{\mathbf{4}}+\frac{X^2}{\mathbf{5}}\leqq 1$$

$X+Y=k$ とおくと，$Y=-X+k$.

$5(X-Y)^2+4X^2-20=0$　←斜めの楕円

と連立すると

$$5(2X-k)^2+4X^2-20$$

$$=24X^2-20kX+5k^2-20=0$$

$$\frac{D}{4}=100k^2-24(5k^2-20)$$

$$=-20k^2+480\geqq 0$$

$$\iff k^2\leqq 24$$

$$\iff -2\sqrt{6}\leqq k\leqq 2\sqrt{6}$$

$\therefore 0 \leqq |X+Y| \leqq \mathbf{2\sqrt{6}}$

15

(1)

$s_{xy} = \dfrac{1}{n}\displaystyle\sum_{k=1}^{n}(x_k - \overline{x})(y_k - \overline{y})$

$= \dfrac{1}{n}\displaystyle\sum_{k=1}^{n}(x_k y_k - \overline{y}x_k - \overline{x}y_k + \overline{x}\cdot\overline{y})$

$= \overline{xy} - \overline{y}\dfrac{1}{n}\displaystyle\sum_{k=1}^{n}x_k - \overline{x}\dfrac{1}{n}\sum_{k=1}^{n}y_k + \dfrac{1}{n}n\overline{x}\cdot\overline{y}$

$= \overline{z} - \overline{y}\cdot\overline{x} - \overline{x}\cdot\overline{y} + \overline{x}\cdot\overline{y}$

$= \overline{z} - \overline{x}\cdot\overline{y}$

(2)

データを順に (l 個ごとに改行して) 並べてみると

$(1,1)$, $(2,2)$, $\cdots$, (l,l),

$(1,l+1)$, $(2,l+2)$, $\cdots$, $(l,2l)$,

$(1,2l+1)$, $(2,2l+2)$, $\cdots$, $(l,3l)$,

$\vdots$

$(1,(k-1)l+1)$, $(2,(k-1)l+2)$, $\cdots$, (l,kl),

$\vdots$

$(1,n-l+1)$, $(2,n-l+2)$, $\cdots$, (l,n)

まず

$\overline{x} = \dfrac{1}{n}\times\dfrac{1}{2}l(l+1)\times\dfrac{n}{l} = \dfrac{l+1}{2}$

$\overline{y} = \dfrac{1}{n}\times\dfrac{1}{2}n(n+1) = \dfrac{n+1}{2}$

$\overline{x^2} = \dfrac{1}{n}\times\dfrac{1}{6}l(l+1)(2l+1)\times\dfrac{n}{l}$

$= \dfrac{(l+1)(2l+1)}{6}$

$\overline{y^2} = \dfrac{1}{n}\times\dfrac{1}{6}n(n+1)(2n+1)$

$= \dfrac{(n+1)(2n+1)}{6}$

$\displaystyle\sum_{k=1}^{n}x_k y_k$

$= \displaystyle\sum_{k=1}^{\frac{n}{l}}\sum_{m=1}^{l}m\{(k-1)l+m\}$

$= \displaystyle\sum_{k=1}^{\frac{n}{l}}\left\{(k-1)l\frac{l}{2}(l+1) + \frac{l}{6}(l+1)(2l+1)\right\}$

$= \displaystyle\sum_{k=1}^{\frac{n}{l}}\left\{\frac{l^2}{2}(l+1)k + \frac{l}{6}(l+1)(2l+1) - \frac{l^2}{2}(l+1)\right\}$

$= \dfrac{l^2}{2}(l+1)\dfrac{1}{2}\dfrac{n}{l}\left(\dfrac{n}{l}+1\right) + \dfrac{n}{6}(l+1)(2l+1) - \dfrac{nl}{2}(l+1)$

$= \dfrac{n(l+1)}{12}(3n+l+2)$

$\therefore \overline{xy} = \dfrac{l+1}{12}(3n+l+2)$

これらより

$s_x^2 = \dfrac{(l+1)(2l+1)}{6} - \dfrac{(l+1)^2}{4}$

$= \dfrac{l^2-1}{12}$

$s_y^2 = \dfrac{n^2-1}{12}$

$s_{xy} = \dfrac{n(l+1)}{12}(3n+l+2) - \dfrac{(l+1)(n+1)}{4}$

$= \dfrac{l^2-1}{12}$

$\therefore r = \dfrac{s_{xy}}{s_x s_y} = \dfrac{\frac{l^2-1}{12}}{\sqrt{\frac{l^2-1}{12}}\sqrt{\frac{n^2-1}{12}}}$

$= \sqrt{\boldsymbol{\dfrac{l^2-1}{n^2-1}}}$

(3)

例として，$l=3$ のときを 書き出してみると $(1,1)$，$(2,2)$，$(3,3)$，$(1,4)$，$(2,1)$，$(3,2)$，$(1,3)$，$(2,4)$，$(3,1)$，$(1,2)$，$(2,3)$，$(3,4)$ となる．

(2) より

$\overline{x} = \dfrac{l+1}{2}$

$\overline{y} = \dfrac{1}{n}\times\dfrac{1}{2}(l+1)(l+2)\times l = \dfrac{l+2}{2}$

$\overline{x^2} = \dfrac{(l+1)(2l+1)}{6}$

$\overline{y^2} = \dfrac{1}{n}\times\dfrac{1}{6}(l+1)(l+2)(2l+3)\times l$

$= \dfrac{(l+2)(2l+3)}{6}$

そして $\displaystyle\sum_{k=1}^{n}x_k y_k$ は，上の例では x は 1〜3，y は 1〜4 から選んでできる数字の積の和より $(1+2+3)(1+2+3+4)$ となる．つまり

$\displaystyle\sum_{k=1}^{n}x_k y_k = (1+2+\cdots+l)(1+2+\cdots+(l+1))$

$= \dfrac{1}{2}l(l+1)\dfrac{1}{2}(l+1)(l+2)$

$= \dfrac{l}{4}(l+1)^2(l+2)$

$\therefore \overline{xy} = \frac{1}{4}(l+1)(l+2)$

$\therefore s_{xy} = \overline{xy} - \overline{x} \cdot \overline{y} = 0$

$\therefore r = \mathbf{0}$