１．データの分析

〔1－1　平均値と分散〕

【例題】

下の表は，5 人の生徒 A，B，C，D，E に 10 点満点のテストを 2 回行ったときの得点の結果です。1 回目，2 回目の得点をそれぞれ x，y とするとき，x，y のデータの平均値 $\bar{x}$，$\bar{y}$，分散 ${s_x}^2$，${s_y}^2$，標準偏差 s_x，s_y をそれぞれ求めよ。

	A	B	C	D	E
1 回目（x）	7	5	6	3	9
2 回目（y）	3	6	2	7	9

まずは、「分散」の定義にしたがって計算してみましょう。

〔1 回目〕

〔2 回目〕

〔2回目〕の計算が少し大変でした。分散には、次のような公式があります。

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

これを使って、〔2回目〕の分散を、もう一度計算してみましょう。

【公式の証明とまとめ】

大きさ n のデータの値を x_1，x_2，……，x_n とすると・・・

・平均値

・分散

・標準偏差

〔1－2　変量の変換〕

	A	B	C	D	E
1回目（x）	7	5	6	3	9

先ほど扱った右上のデータで、平均値 $\overline{x}=6$、分散 $s_x{}^2=4$ でした。

この資料の x の値を一定の規則に基づいて変換したら、平均値や分散がどのように変わるかを調べてみましょう。

それぞれの x の値を5倍し、10をたしてみます。
変換した値を y とすると変数 y は、$y=5x+10$ という関数で表されます。

$$y=5x+10$$

・n 個の変量 x の平均値を $\overline{x}$、分散を ${s_x}^2$ とする。

$y = ax + b\ (a \neq 0)$ としたとき、n 個の変量 y の平均値 $\overline{y}$、分散 ${s_y}^2$ は…

【例題】

次の変量 x のデータについて、

702，732，738，744，750，762，798，822

平均値 $\overline{x}$、分散 ${s_x}^2$ をそれぞれ求めよ。。

まず、計算機を使って答えを先に出しておくと・・・ $\boxed{\overline{x}=756}$

x	702	732	738	744	750	762	798	822
$x-\overline{x}$	-54	-24	-18	-12	-6	6	42	66
$(x-\overline{x})^2$	2916	576	324	144	36	36	1764	4356

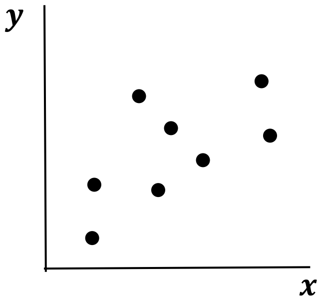
よって $\boxed{{s_x}^2=1269}$

・・・これを、工夫して求めることを考えてみましょう。

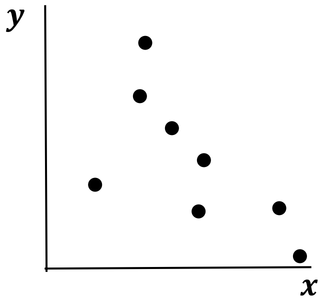
☆

〔1－3　相関係数〕

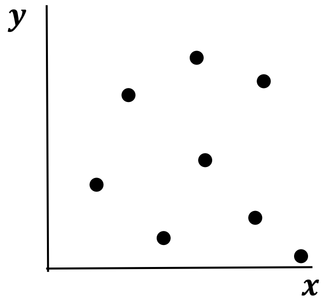
・相関関係



正の相関関係



負の相関関係

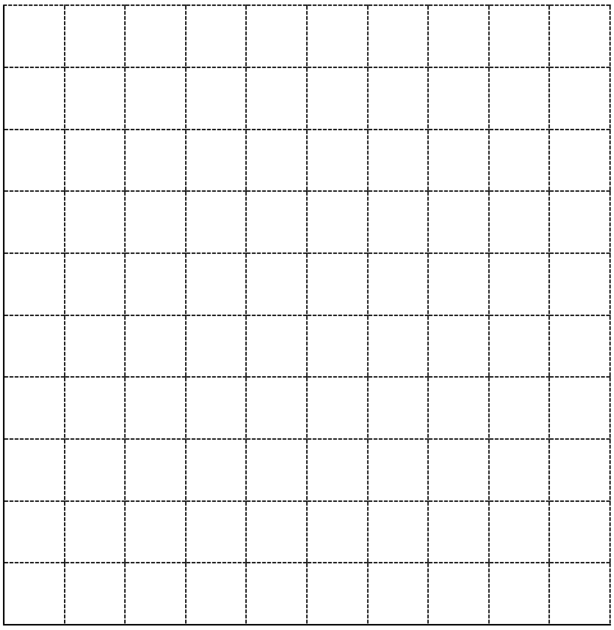


相関関係なし

【例題】

次の表は、10 人の生徒に 10 点満点のテストを 2 回行ったときの得点の結果である。1 回目、2 回目の得点をそれぞれ x、y とし、散布図を作成しましょう。

	A	B	C	D	E	F	G	H	I	J
1回目（x）	2	2	6	4	3	3	5	7	4	4
2回目（y）	3	7	9	7	2	4	8	5	6	9



☆　共分散　s_{xy}　・・・x の偏差と y の偏差の積 $(x_k-\overline{x})(y_k-\overline{y})$ の平均値

$$s_{xy}=\frac{1}{n}\left\{(x_1-\overline{x})(y_1-\overline{y})+(x_2-\overline{x})(y_2-\overline{y})+\cdot\cdot\cdot\cdot\cdot+(x_n-\overline{x})(y_n-\overline{y})\right\}$$

例）

	A	B	C	D	E	F	G	H	I	J
１回目（x）	2	2	6	4	3	3	5	7	4	4
２回目（y）	3	7	9	7	2	4	8	5	6	9

☆　相関係数　r

・・・2つの変量データにおいて、その相関関係の強弱を表すもの。

x、y の標準偏差をそれぞれ s_x、s_y とし、共分散を s_{xy} とすると

$$r=\frac{s_{xy}}{s_x s_y}$$

$$=\frac{\frac{1}{n}\left\{(x_1-\overline{x})(y_1-\overline{y})+(x_2-\overline{x})(y_2-\overline{y})+\cdots\cdots+(x_n-\overline{x})(y_n-\overline{y})\right\}}{\sqrt{\frac{1}{n}\{(x_1-\overline{x})^2+(x_2-\overline{x})^2+\cdots+(x_n-\overline{x})^2\}\times\frac{1}{n}\{(y_1-\overline{y})^2+(y_2-\overline{y})^2+\cdots+(y_n-\overline{y})^2\}}}$$

$-1 \leqq r \leqq 1$　であり

〔1〕　r の値が 1 に近いとき、強い正の相関関係

〔2〕　r の値が -1 に近いとき、強い負の相関関係

〔3〕　r の値が 0 に近いとき、相関関係はない

	A	B	C	D	E	F	G	H	I	J
1回目（x）	2	2	6	4	3	3	5	7	4	4
2回目（y）	3	7	9	7	2	4	8	5	6	9

このデータの相関係数 r を調べてみましょう。

小数第 3 位を四捨五入し、小数第 2 位まででよいです。

次のような問題を考えてみましょう。

問）次の相関係数に関する①～③の記述について、それぞれ正しいか正しくないかを答えなさい。

① 2つの変量のどちらを散布図の縦軸・横軸にするかで、相関係数の値は変わる。
② もとのデータの一方の変量に定数を加えると、相関係数の値は変わる。
③ 一方の変量がもう一方の変量に比例するとき、相関係数は1である。

①、②について・・・

$$r=\frac{(x_1-\overline{x})(y_1-\overline{y})+(x_2-\overline{x})(y_2-\overline{y})+\cdot\cdot\cdot\cdot\cdot+(x_n-\overline{x})(y_n-\overline{y})}{\sqrt{\{(x_1-\overline{x})^2+(x_2-\overline{x})^2+\cdots+(x_n-\overline{x})^2\}\{(y_1-\overline{y})^2+(y_2-\overline{y})^2+\cdots+(y_n-\overline{y})^2\}}}$$

③について・・・

☆　$r = -1$　について