

exercise 【データの相関①】

●データの関係性をあらわそう

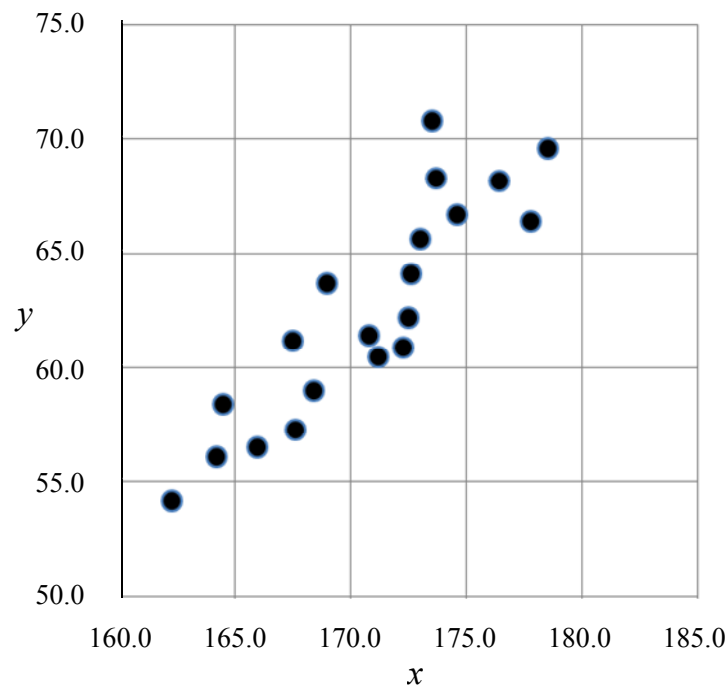
2つの変量からなるデータの間には、一方が増加するともう一方も増加する場合や、一方が増加するともう一方は減少する場合がある。このとき2つのデータには相関がある、または相関関係があるという。その関係が見られないときには相関がない、または相関関係がないという。

例) 下の表と図は、ある学校の陸上部員 20 人について、2つの変量 x , y からなるデータである。
 x は身長, y は体重である。

	1	2	3	4	5	6	7	8	9	10
x	168.4	164.5	171.2	173.0	162.3	170.8	172.5	164.2	169.0	167.6
y	59.0	58.4	60.5	65.6	54.2	61.4	62.2	56.1	63.7	57.3

	11	12	13	14	15	16	17	18	19	20
x	172.6	166.0	173.7	176.4	178.5	167.5	177.8	174.6	172.3	173.5
y	64.1	56.5	68.3	68.2	69.6	61.2	66.4	66.7	60.9	70.8

(x の単位はcm, y の単位はkg)



表を見ても関係がわかりにくいデータも、変数を座標とする点を取ると、関係性を視覚的にとらえることができわかりやすくなる。この図のように2つのデータを平面上に図示したものを散布図という。

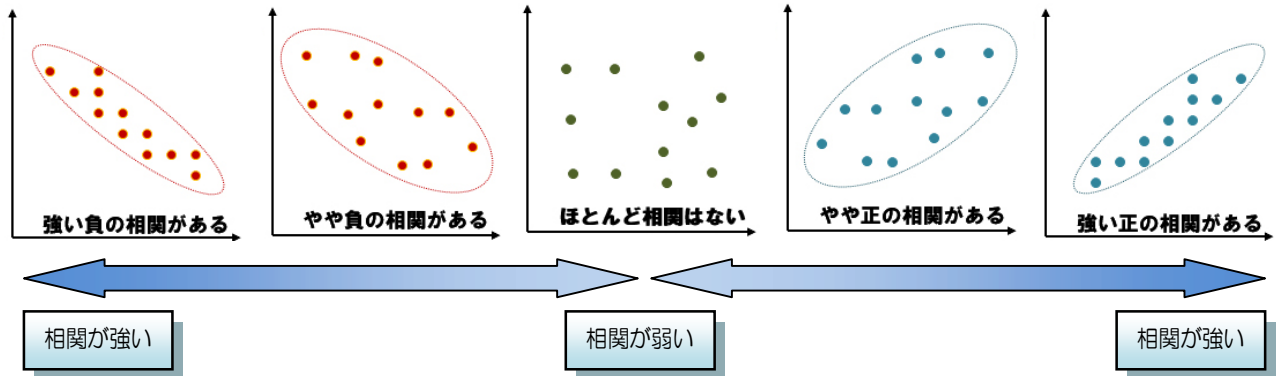
練習) 11~20 までの部員のデータを上の散布図に追加しなさい。

◎ 散布図からデータの傾向を読み取ろう

正の相関、負の相関

2つの変量からなるデータにおいて、一方が増加すると他方も増加する傾向が見られるとき、2つの変量には**正の相関**があるという。また、一方が増加すると他方は減少する傾向が見られるとき、2つの変量には**負の相関**があるという。どちらの傾向も見られないときには、**相関がない**または**相関関係がない**という。

2つの変量の間に関係があるとき、散布図における点の分布の様子が1つの直線に接近しているほど**相関が強い**といい、散らばっているほど**相関が弱い**という。



練習 11) 次のような2つの変量 x, y からなるデータがある。これらについて x と y の間に相関があるかどうかを調べよ。また、相関がある場合には、正か負のどちらの相関であるかをいえ。

(1)

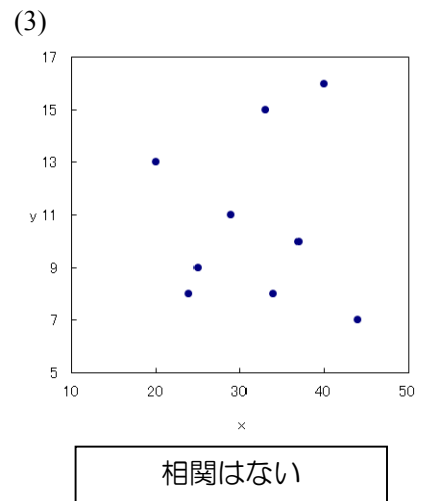
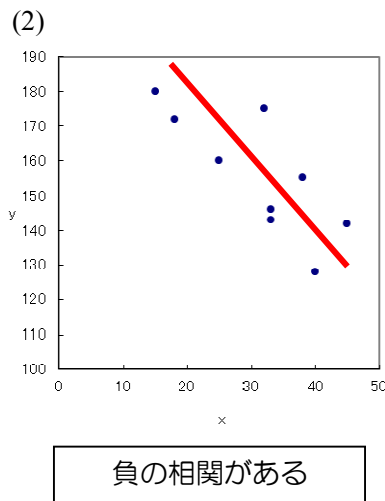
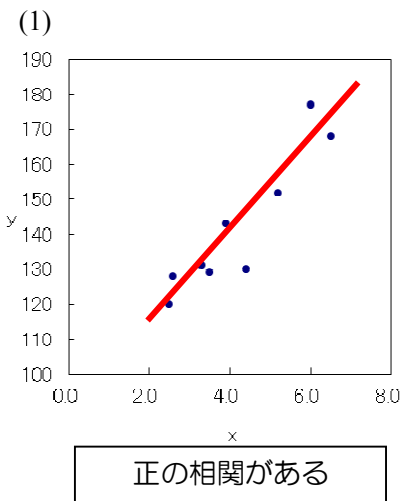
x	3.5	2.6	5.2	2.5	3.9	6.5	3.3	6.0	4.4	3.5
y	129	128	152	120	143	168	131	177	130	129

(2)

x	15	33	18	25	45	33	38	40	32	15
y	180	143	172	160	142	146	155	128	175	180

(3)

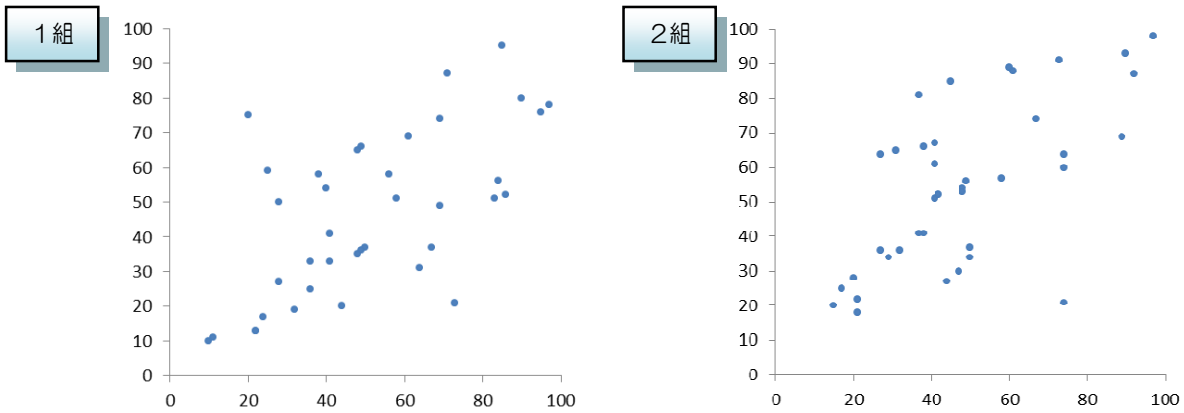
x	29	34	25	20	40	24	37	33	44	29
y	11	8	9	13	16	8	10	15	7	11



exercise 【データの相関 ②】

下の図は、ある2クラスの国語と英語のテストの散布図である。どの程度強い相関関係であるかを他の人に伝える際にどうすれば的確に伝えられるだろうか？

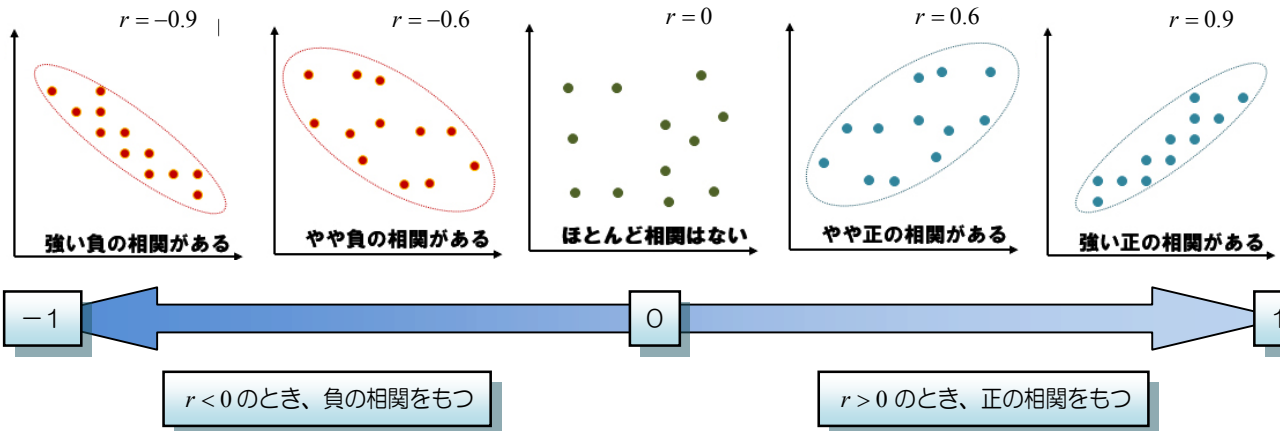
また微妙な散らばり具合の散布図の相関の強さを比較するとき、どうすれば判断することができるだろうか。



2つの変量からなるデータが与えられたとき、データの値から相関関係を調べる方法として、どの程度直線的であるかを数値で表すことで比較することができる。

相関係数(r)とは

2変数間にどの程度直線的な関係があるかを数値で表す方法として、相関係数を調べる方法がある。



相関係数 r については、 $-1 \leq r \leq 1$ であることが知られている。また、 r が 1 に近いほど正の相関が強く、 -1 に近いほど負の相関が強い。相関がないとき、 r は 0 に近い値を取る。

【参考】

相関係数	相関関係
0	相関がない
0.0~±0.2	ほとんど相関がない
±0.2~±0.4	やや相関がある (低い相関)
±0.4~±0.7	相関がある
±0.7~±0.9	強い相関がある (高い相関)
±0.9~±1.0	きわめて強い相関がある
±1.0	完全な相関

ちなみに前述の1組、2組の散布図の相関係数は 0.65 と 0.62 となるので、微妙ながら1組の方が正の相関が強いということがわかる。

◎相関係数を計算してみよう。

	x	y	$x-\bar{x}$	$y-\bar{y}$	$(x-\bar{x})(y-\bar{y})$	$(x-\bar{x})^2$	$(y-\bar{y})^2$
①	21	15	-5	-3	15	25	9
②	27	17	1	-1	-1	1	1
③	29	19	3	1	3	9	1
④	23	17	-3	-1	3	9	1
⑤	30	22	4	4	16	16	16
計	130	90	計		36 ~○	60 ~△	28 ~□
平均	26	18	平均		7.2	12	5.6

x と y の共分散
x の分散
y の分散

●計算方法 1 (合計の値で計算する)

$$r = \frac{\bigcirc}{\sqrt{\triangle \times \square}} = \frac{36}{\sqrt{60 \times 28}} = \frac{36}{4\sqrt{105}} = 0.8783\dots \approx 0.88$$

●計算方法 2 (共分散や分散、標準偏差の値で計算する)

$$r = \frac{(\text{共分散})}{\sqrt{(x \text{ の分散}) \times (y \text{ の分散})}} = \frac{(\text{共分散})}{(x \text{ の標準偏差}) \times (y \text{ の標準偏差})} = \frac{7.2}{\sqrt{12} \times \sqrt{5.6}} = \frac{7.2}{\sqrt{67.2}} = 0.8783\dots \approx 0.88$$

練習) 下の表は生徒 10 人の数学の小テストの結果 x 点と、国語の小テストの結果 y 点です。
 x と y の相関係数を求めなさい。また、これらの間にはどのような相関があると考えられるか。

	x	y	$x-\bar{x}$	$y-\bar{y}$	$(x-\bar{x})(y-\bar{y})$	$(x-\bar{x})^2$	$(y-\bar{y})^2$
①	10	7	3	1	3	9	1
②	5	5	-2	-1	2	4	1
③	9	6	2	0	0	4	0
④	9	9	2	3	6	4	9
⑤	8	5	1	-1	-1	0	1
⑥	4	3	-3	-3	9	9	9
⑦	8	10	1	4	4	1	16
⑧	4	3	-3	-3	9	9	9
⑨	7	8	0	2	0	0	4
⑩	6	4	-1	-2	2	1	4
計	70	60	計		34	42	54
平均	7	6	平均		3.4	4.2	5.4

$$r = \frac{34}{\sqrt{42 \times 54}} = 0.713932 \approx 0.71$$

よって正の強い相関がある

【参考】相関表

練習の x と y を散布図にすると重なる点があり、正しく表すことができない。またデータの数が多すぎると点だらけになってしまう。そのようなときは相関表を用いると良い。

	0 以上 20 未満	20 ~ 40	40 ~ 60			
80 以上 100 未満			2	3	6	11
60 ~ 80		1	3	14	4	22
40 ~ 60	2		6	12	2	22
20 ~ 40		4	2			6
0 ~ 20	5	1		1		7
計	7	6	13	30	12	68