2 節 データの相関

1 相関関係

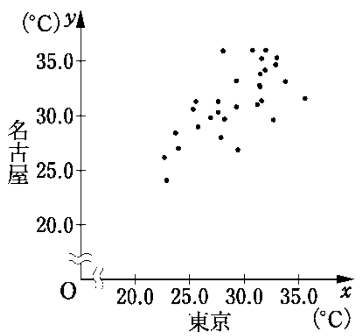
散布図と相関関係

（教科書 p.171）

右の表は，ある年の 7 月の東京と名古屋における各日の最高気温 x(℃)，y(℃) のデータである。

この表から，各日の変量 x，y の値の組を座標とする点を平面上にとると，下の図ができる。

日	東京 x	名古屋 y	日	東京 x	名古屋 y
1	27.9	28.0	16	31.5	33.8
2	29.4	26.9	17	31.2	31.0
3	22.7	26.2	18	32.9	34.7
4	22.9	24.1	19	33.8	33.1
5	29.3	30.8	20	29.3	33.2
6	23.7	28.4	21	30.8	36.0
7	27.6	30.3	22	28.1	35.9
8	27.6	31.3	23	25.8	29.0
9	24.0	27.0	24	26.9	29.8
10	31.6	31.4	25	31.4	32.8
11	32.7	29.6	26	28.2	29.7
12	25.8	29.0	27	35.6	31.6
13	25.3	30.6	28	33.0	35.3
14	25.6	31.3	29	31.6	35.2
15	31.5	32.6	30	32.0	36.0
			31	31.9	34.2



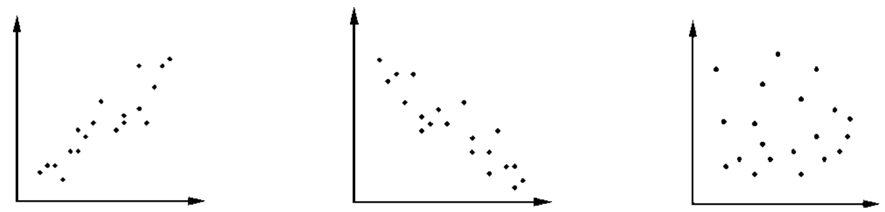
（[1]　　　　　）：上の図のように，縦軸，横軸に 2 つのデータを示した図

上の散布図で，対応する x，y の値の関係について

（[2]　　　　　　　）：一方が増加すれば他方も増加する傾向

（[3]　　　　　　　）：一方が増加すれば他方が減少する傾向

（[4]　　　　　　　　）：正の相関も負の相関もみられない



（[5]　　　　　　　）　（[6]　　　　　　　）　（[7]　　　　　　　　）

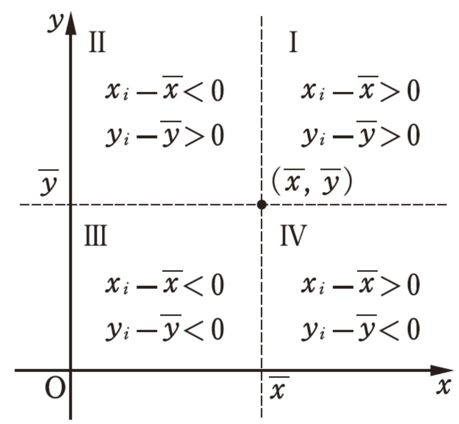
2 相関係数

相関係数

（教科書 p.172）

散布図の n 個のデータを (x_1, y_1)，(x_2, y_2)，$\cdots(x_n, y_n)$ とする。

右の図のように，点$(\overline{x}, \overline{y})$を中心に平面を 4 分割し，各部分をⅠ，Ⅱ，Ⅲ，Ⅳとする。



x と y に正の相関関係があれば，散布図の点はⅠとⅢに多く集まり，負の相関関係があればⅡとⅣに多く集まる傾向がある。

散布図の各点 (x_i, y_i) $(i=1, 2, 3, \cdots, n)$ が

ⅠまたはⅢに属するときは　$(x_i-\overline{x})(y_i-\overline{y})>0$

ⅡまたはⅣに属するときは　$(x_i-\overline{x})(y_i-\overline{y})<0$

である。

したがって，

$(x-\overline{x})(y-\overline{y})$ の平均値

$$\frac{1}{n}\{(x_1-\overline{x})(y_1-\overline{y})+(x_2-\overline{x})(y_2-\overline{y})+\cdots+(x_n-\overline{x})(y_n-\overline{y})\}$$

は正の相関関係があるなら正に，負の相関関係があるなら負になる。

この値を x，y の（[8]　　　　　　　）といい，s_{xy} で表す。

x と y に相関関係がないときは，0 に近い値になる。

共分散はデータの単位の取り方や散らばり具合に影響を受ける。

そこで，x と y の共分散

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

を x，y の標準偏差

$$s_x=\sqrt{\frac{1}{n}\{(x_1-\overline{x})^2+(x_2-\overline{x})^2+\cdots+(x_n-\overline{x})^2\}}$$

$$s_y=\sqrt{\frac{1}{n}\{(y_1-\overline{y})^2+(y_2-\overline{y})^2+\cdots+(y_n-\overline{y})^2\}}$$

の積で割った値を考える。

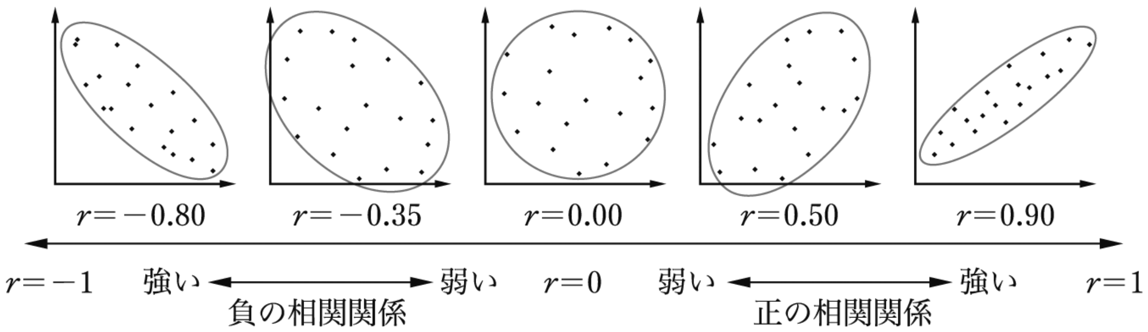
x，y の（[9]　　　　　　　）：x と y の共分散を，x，y の標準偏差の積で割った値。r で表す。

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

一般に，相関係数 r の値については，次の不等式が成り立つ。

$$-1 \leqq r \leqq 1$$

（[10]　　　　）の相関関係が強いほど r の値は 1 に近づき，（[11]　　　　）の相関関係が強いほど r の値は -1 に近づく。



例題 1　下の表は，5 人のフットサルの選手 a，b，c，d，e の身長 $x$cm と体重 $y$kg を調べたものである。このとき，5 人の選手の身長と体重の相関係数を小数第 3 位を四捨五入して，小数第 2 位まで求めよ。

	x(cm)	y(kg)
a	180	68
b	172	72
c	165	58
d	172	65
e	166	52

解

となり，下のような表をつくると，

	x	y	$x-\overline{x}$	$(x-\overline{x})^2$	$y-\overline{y}$	$(y-\overline{y})^2$	$(x-\overline{x})(y-\overline{y})$
a	180	68					
b	172	72					
c	165	58					
d	172	65					
e	166	52					
計	855	315					

したがって，相関係数 r は

問1 右の表は，5 人の生徒の数学と英語の小テストの結果である。このとき，数学の点数と英語の点数の相関係数を小数第 3 位を四捨五入して，小数第 2 位まで求めよ。ただし，$\sqrt{2} = 1.414$ とする。

	数学	英語
a	8	7
b	7	5
c	9	9
d	9	7
e	7	7

問 題

（教科書 p.175）

3 次の散布図 a，b，c，d について，それぞれ対応する相関係数を①，②，③，④の中から選べ。

【散布図】

図 a

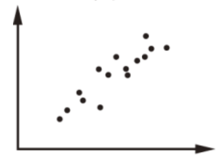


図 b

図 c

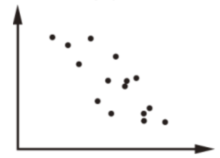
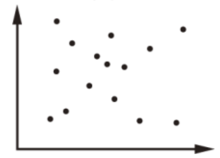


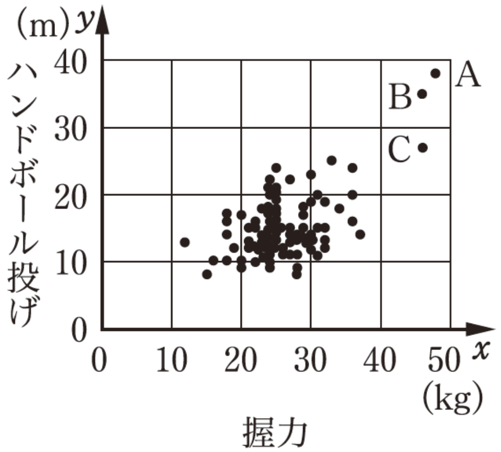
図 d



【相関係数】

①：$r = 0.9$　②：$r = -0.8$　③：$r = 0$　④：$r = 0.6$

4 右の散布図は，ある高校の１年生女子の握力の記録 x kg とハンドボール投げの記録 y m の測定結果である。x と y の相関係数 r は 0.60 であった。ところが，集計ミスで，散布図中の点 A，B，C は男子生徒の結果であることがわかった。この３人の男子生徒の測定結果を除くと，r はどうなるか。①～③のうちから適するものを選べ。



① 0.60 より大きくなる

② 0.60 とほぼ変わらない

③ 0.60 より小さくなる

2 節 データの相関

1 相関関係

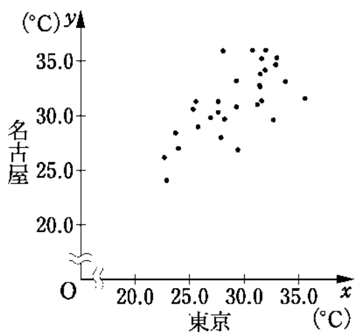
散布図と相関関係

（教科書 p.171）

右の表は，ある年の７月の東京と名古屋における各日の最高気温 x(℃)，y(℃) のデータである。

この表から，各日の変量 x，y の値の組を座標とする点を平面上にとると，下の図ができる。

日	東京 x	名古屋 y	日	東京 x	名古屋 y
1	27.9	28.0	16	31.5	33.8
2	29.4	26.9	17	31.2	31.0
3	22.7	26.2	18	32.9	34.7
4	22.9	24.1	19	33.8	33.1
5	29.3	30.8	20	29.3	33.2
6	23.7	28.4	21	30.8	36.0
7	27.6	30.3	22	28.1	35.9
8	27.6	31.3	23	25.8	29.0
9	24.0	27.0	24	26.9	29.8
10	31.6	31.4	25	31.4	32.8
11	32.7	29.6	26	28.2	29.7
12	25.8	29.0	27	35.6	31.6
13	25.3	30.6	28	33.0	35.3
14	25.6	31.3	29	31.6	35.2
15	31.5	32.6	30	32.0	36.0
			31	31.9	34.2



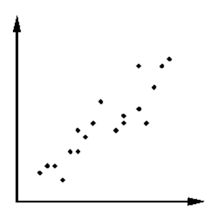
（1　散布図　）：上の図のように，縦軸，横軸に２つのデータを示した図

上の散布図で，対応する x，y の値の関係について

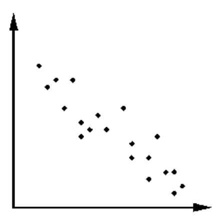
（2　正の相関関係　）：一方が増加すれば他方も増加する傾向

（3　負の相関関係　）：一方が増加すれば他方が減少する傾向

（4　相関関係がない　）：正の相関も負の相関もみられない



（5　正の相関関係　）



（6　負の相関関係　）

（7　相関関係がない　）

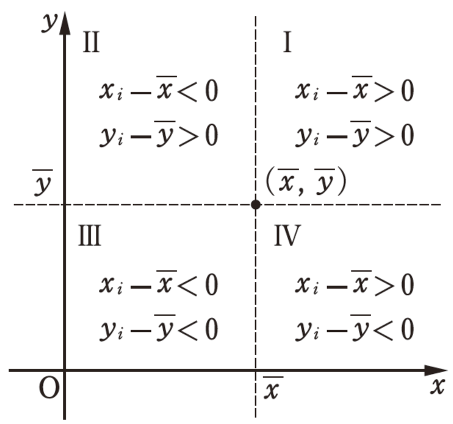
2 相関係数

相関係数

（教科書 p.172）

散布図の n 個のデータを (x_1, y_1)，(x_2, y_2)，$\cdots(x_n, y_n)$ とする。

右の図のように，点$(\overline{x}, \overline{y})$を中心に平面を４分割し，各部分をⅠ，Ⅱ，Ⅲ，Ⅳとする。



x と y に正の相関関係があれば，散布図の点はⅠとⅢに多く集まり，負の相関関係があればⅡとⅣに多く集まる傾向がある。

散布図の各点 (x_i, y_i) $(i = 1, 2, 3, \cdots, n)$ が

ⅠまたはⅢに属するときは　$(x_i-\overline{x})(y_i-\overline{y})>0$

ⅡまたはⅣに属するときは　$(x_i-\overline{x})(y_i-\overline{y})<0$

である。

したがって，

$(x-\overline{x})(y-\overline{y})$ の平均値

$$\frac{1}{n}\{(x_1-\overline{x})(y_1-\overline{y})+(x_2-\overline{x})(y_2-\overline{y})+\cdots+(x_n-\overline{x})(y_n-\overline{y})\}$$

は正の相関関係があるなら正に，負の相関関係があるなら負になる。

この値を x，y の（8　共分散　）といい，s_{xy} で表す。

x と y に相関関係がないときは，0 に近い値になる。

共分散はデータの単位の取り方や散らばり具合に影響を受ける。

そこで，x と y の共分散

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

を x，y の標準偏差

$$s_x=\sqrt{\frac{1}{n}\{(x_1-\overline{x})^2+(x_2-\overline{x})^2+\cdots+(x_n-\overline{x})^2\}}$$

$$s_y=\sqrt{\frac{1}{n}\{(y_1-\overline{y})^2+(y_2-\overline{y})^2+\cdots+(y_n-\overline{y})^2\}}$$

の積で割った値を考える。

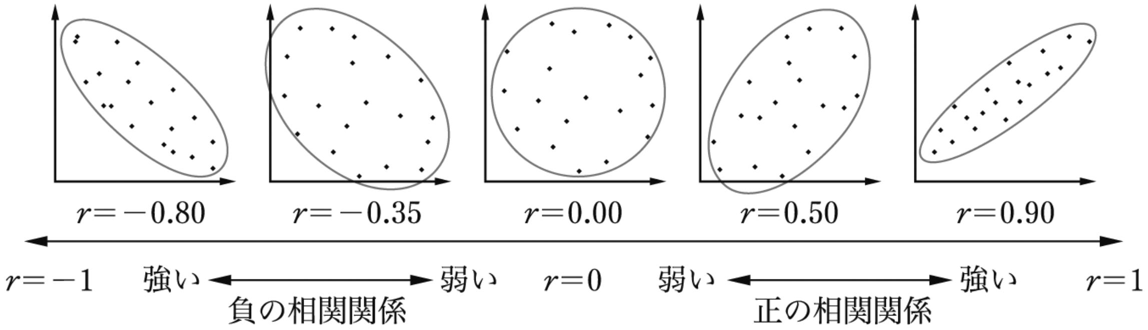
x，y の（9　相関係数　）：x と y の共分散を，x，y の標準偏差の積で割った値。r で表す。

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

一般に，相関係数 r の値については，次の不等式が成り立つ。

$$-1 \leqq r \leqq 1$$

（[10]　正　）の相関関係が強いほど r の値は 1 に近づき，（[11]　負　）の相関関係が強いほど r の値は -1 に近づく。



例題 1　下の表は，5 人のフットサルの選手 a，b，c，d，e の身長 $x$cm と体重 $y$kg を調べたものである。このとき，5 人の選手の身長と体重の相関係数を小数第 3 位を四捨五入して，小数第 2 位まで求めよ。

	x(cm)	y(kg)
a	180	68
b	172	72
c	165	58
d	172	65
e	166	52

解

$\overline{x} = \frac{1}{5}\cdot 855 = 171$ (cm)

$\overline{y} = \frac{1}{5}\cdot 315 = 63$ (kg)

となり，下のような表をつくると，

	x	y	$x-\overline{x}$	$(x-\overline{x})^2$	$y-\overline{y}$	$(y-\overline{y})^2$	$(x-\overline{x})(y-\overline{y})$
a	180	68	9	81	5	25	45
b	172	72	1	1	9	81	9
c	165	58	−6	36	−5	25	30
d	172	65	1	1	2	4	2
e	166	52	−5	25	−11	121	55
計	855	315	0	144	0	256	141

したがって，相関係数 r は

$$r = \frac{\frac{1}{5}\cdot 141}{\sqrt{\frac{1}{5}\cdot 144}\sqrt{\frac{1}{5}\cdot 256}} = \frac{141}{\sqrt{144}\sqrt{256}} = \frac{141}{12\cdot 16} = 0.734\cdots \fallingdotseq 0.73$$

問1 右の表は，5 人の生徒の数学と英語の小テストの結果である。このとき，数学の点数と英語の点数の相関係数を小数第 3 位を四捨五入して，小数第 2 位まで求めよ。ただし，$\sqrt{2}=1.414$ とする。

	数学	英語
a	8	7
b	7	5
c	9	9
d	9	7
e	7	7

5 人の生徒の数学の点数を x，英語の点数を y とする。また，数学の点数の平均値を $\overline{x}$，英語の点数の平均値を $\overline{y}$ とすると

$$\overline{x}=\frac{1}{5}(8+7+9+9+7)=\frac{40}{5}=8\text{（点）}$$

$$\overline{y}=\frac{1}{5}(7+5+9+7+7)=\frac{35}{5}=7\text{（点）}$$

※：$(x-\overline{x})(y-\overline{y})$

	x	y	$x-\overline{x}$	$(x-\overline{x})^2$	$y-\overline{y}$	$(y-\overline{y})^2$	※
a	8	7	0	0	0	0	0
b	7	5	−1	1	−2	4	2
c	9	9	1	1	2	4	2
d	9	7	1	1	0	0	0
e	7	7	−1	1	0	0	0
計	40	35	0	4	0	8	4

数学の点数と英語の点数の相関係数rは

$$r=\frac{\frac{1}{5}\cdot 4}{\sqrt{\frac{1}{5}\cdot 4}\sqrt{\frac{1}{5}\cdot 8}}=\frac{1}{\sqrt{2}}=\frac{\sqrt{2}}{2}$$

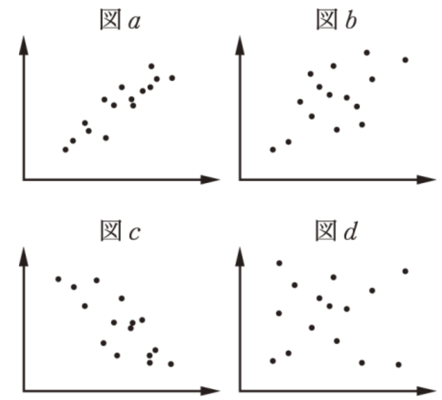
$$=\frac{1.414}{2}=0.707\cdots\fallingdotseq 0.71$$

問　題

（教科書 p.175）

3 次の散布図 a，b，c，d について，それぞれ対応する相関係数を①，②，③，④の中から選べ。

【散布図】

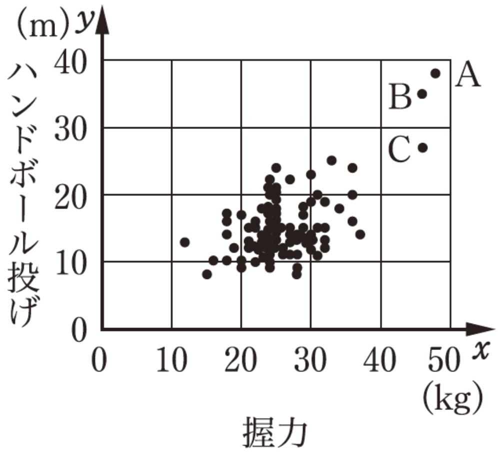


【相関係数】

①：$r=0.9$　　②：$r=-0.8$　　③：$r=0$　　④：$r=0.6$

図 a は①，図 b は④，図 c は②，図 d は③

4 右の散布図は，ある高校の１年生女子の握力の記録 x kg とハンドボール投げの記録 y m の測定結果である。x と y の相関係数 r は 0.60 であった。ところが，集計ミスで，散布図中の点 A，B，C は男子生徒の結果であることがわかった。この３人の男子生徒の測定結果を除くと，r はどうなるか。①～③のうちから適するものを選べ。



① 0.60 より大きくなる

② 0.60 とほぼ変わらない

③ 0.60 より小さくなる

相関関係は 0.60 であるから，x と y の間にやや強い正の相関関係がある。点 A，B，C が除かれると，散布図の形状はより円形に近づくことから，相関関係は弱くなる。したがって，相関係数は 0.60 より小さくなる。

したがって，正解は③