２節 データの相関

❶ 相関関係

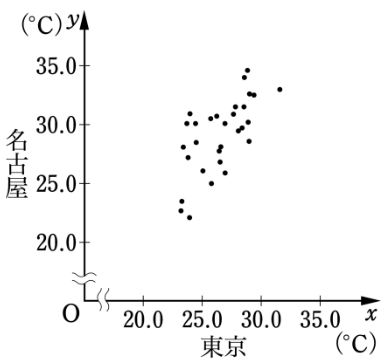
散布図

（教科書 p.175）

次の表は，ある年の 9 月の日ごとの東京，名古屋それぞれの最高気温を x℃，y℃として示したものである。

日	東京 x(℃)	名古屋 y(℃)	日	東京 x(℃)	名古屋 y(℃)	日	東京 x(℃)	名古屋 y(℃)
1	31.5	33.0	11	26.9	30.1	21	23.1	28.1
2	24.0	30.9	12	23.3	23.5	22	25.8	25.0
3	24.5	28.5	13	28.9	28.6	23	26.9	25.9
4	26.2	30.7	14	26.4	27.8	24	27.6	30.9
5	29.0	32.6	15	23.9	22.1	25	28.5	31.5
6	28.5	34.0	16	28.3	29.6	26	27.8	31.5
7	29.3	32.5	17	26.5	28.1	27	24.4	30.1
8	28.8	34.6	18	23.8	27.2	28	26.5	26.8
9	23.7	30.1	19	25.7	30.5	29	25.1	26.1
10	28.8	30.2	20	28.1	29.5	30	23.2	22.7

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

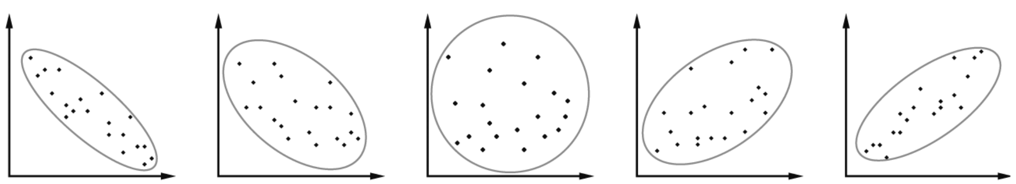


（[1]　　　　　）：右の図のように，変量 x，y の値の組を座標とする点を平面上にとった図。

（[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]　　　　　　　　　）：正の相関関係も負の相関関係もみられない



強い ←→ 弱い　　　　弱い ←→ 強い

（[5]　　　　　）　　　　（[6]　　　　　）

2 つの変量の間に相関関係があるとき

（[7]　　　　　　　　　）：散布図の点の分布が直線状に近づく

（[8]　　　　　　　　　）：散布図の点の分布が広く散らばる

問1 下の表は，ある高校の 1 年生女子 10 人の体力テストにおける 50m 走の記録を x 秒，ハンドボール投げの記録を ym として示したものである。このとき，x と y の散布図を作成せよ。

50m 走 x（秒）	ハンドボール投げ y (m)	50m 走 x（秒）	ハンドボール投げ y (m)
8.5	11	9.4	11
8.0	16	10.3	8
8.6	18	8.1	21
8.8	14	9.2	10
7.8	19	8.9	9

❷ 相関係数

相関係数

（教科書 p.177）

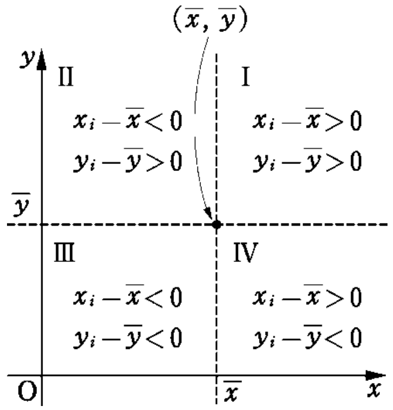
2 つの変量 x，y のデータの値の組を

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

とし，x，y のデータの平均値をそれぞれ $\overline{x}$，$\overline{y}$ とする。

この平均値を座標とする点 $(\overline{x},\ \overline{y})$ は，散布図の点の代表的な位置を表すと考えられ，散布図の点はこのまわりに分布している。

右の図のように，$(\overline{x},\ \overline{y})$ 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 の間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。

$(x_1-\overline{x})(y_1-\overline{y})$，$(x_2-\overline{x})(y_2-\overline{y})$，$\cdots$，$(x_n-\overline{x})(y_n-\overline{y})$

の平均値

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

を変量 x，y の（[1]　　　　　　）といい，正の相関関係があるときには（[2]　　　　）に，負の相関関係があるときは（[3]　　　　）となる傾向がある。相関関係がないときは，正の項と負の項が消し合って（[4]　　　）に近い値となる。

共分散は変量 x，y の単位の取り方や散らばり具合に影響を受けるから，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 の標準偏差と 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 s_xs_y で割った値を考える。

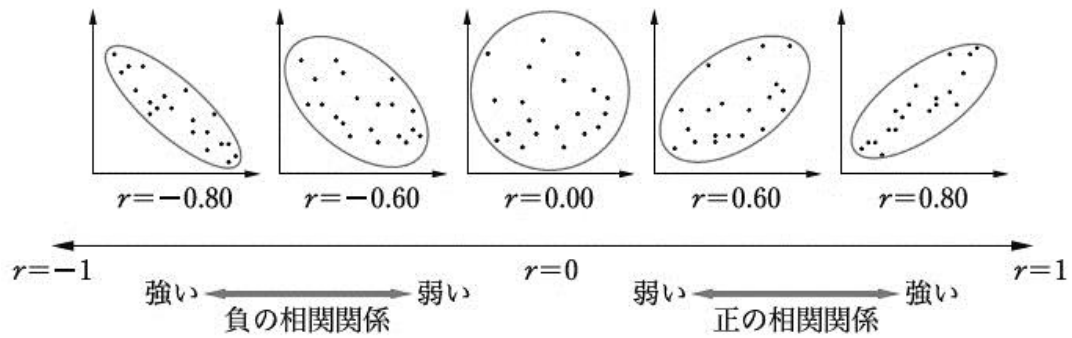
この値を変量 x と y の（[5]　　　　　　）といい r で表す。

相関係数
$$r=\frac{s_{xy}}{s_xs_y}$$ $$(x\text{と}y\text{の相関係数})=\frac{(x\text{と}y\text{の共分散})}{(x\text{の標準偏差})\times(y\text{の標準偏差})}$$

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 人の生徒 a，b，c，d，e が受けた，数学と国語のテストの点数である。
数学（x 点）と国語（y 点）の点数の相関係数 r を，小数第 3 位を四捨五入して答えよ。

	x (点)	y (点)
a	72	69
b	76	76
c	93	88
d	84	75
e	85	87

解 数学の平均点 $\overline{x}$，国語の平均点 $\overline{y}$ は

となり，下の表のようにして計算すると，

	x	y	$x-\overline{x}$	$(x-\overline{x})^2$	$y-\overline{y}$	$(y-\overline{y})^2$	$(x-\overline{x})(y-\overline{y})$
a	72	69					
b	76	76					
c	93	88					
d	84	75					
e	85	87					
計							

問2 下の表は 4 人の生徒 a，b，c，d の数学と英語の小テストの点数である。数学と英語の点数の相関係数を求めよ。

	a	b	c	d
数学	4	7	5	4
英語	6	9	8	9

	x	y	$x-\overline{x}$	$(x-\overline{x})^2$	$y-\overline{y}$	$(y-\overline{y})^2$	$(x-\overline{x})(y-\overline{y})$
a	4	6					
b	7	9					
c	5	8					
d	4	9					
計							

参考　相関表

（教科書 p.180）

2 つの変量 x と y の間の相関関係を調べるとき，散布図のほかに，x と y の度数分布表を組み合わせた表を用いることがある。そのような表を（[1]　　　　　　　）という。

教科書 175 ページの東京と名古屋それぞれの最高気温を相関表に表すと右のようになる。ただし，この相関表は，たとえば東京の最高気温が 26℃ 以上 28℃ 未満で，名古屋の最高気温が 30℃ 以上 32℃ 未満の日が 4 日あることを表している。

名古屋(℃) ＼ 東京(℃)	22～24	24～26	26～28	28～30	30～32
34～36				2	
32～34				2	1
30～32	1	3	4	2	
28～30	1	1	1	3	
26～28	1	1	2		
24～26		1	1		
22～24	3				

変量 x，y の組の個数が多くなると，散布図では点が重なり，相関関係が把握しにくくなる場合がある。そのような場合，相関表を用いるとよい。

問 1　下の表は，ある高校の 1 年生男子 10 人の体力テストにおける握力を $x$kg，ハンドボール投げの記録を ym として示したものである。このとき，x と y の相関表を作成せよ。ただし，x は 34kg 以上36kg 未満，y は 26m 以上 28m 未満という区間から，幅を一定にとるものとする。

握力 x (kg)	ハンドボール投げ y (m)	握力 x (kg)	ハンドボール投げ y (m)
37	31	43	30
39	33	38	26
34	28	37	31
38	30	38	32
40	37	36	31

Training

（教科書 p.181）

4　次の a, b, c, d の散布図に対応する相関係数を下の 4 つから選べ。

図 a

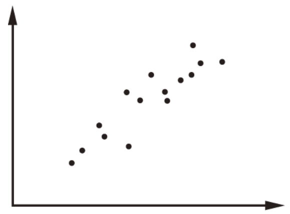


図 b

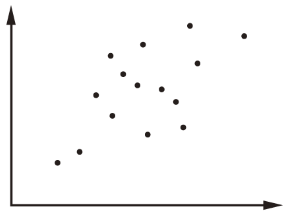


図 c

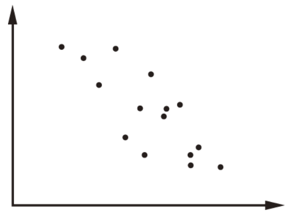
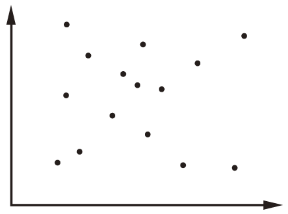


図 d



$r = 0.6,\ r = 0,\ r = -0.8,\ r = 0.9$

5　次の表は，あるクラスの生徒 10 人に，10 日間の合計読書時間x（時間）と読んだ本の冊数y（冊）を調査し，その結果をまとめたものである。このとき，次の問に答えよ。

	x	y	$x-\overline{x}$	$(x-\overline{x})^2$	$y-\overline{y}$	$(y-\overline{y})^2$	$(x-\overline{x})(y-\overline{y})$
生徒 1	14	7	A	B	C	D	E
生徒 2	8	3	-2	4	-2	4	4
⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮
生徒 9	5	3	-5	25	-2	4	10
生徒 10	11	8	1	1	3	9	3
計	100	50	F	160	G	40	58

(1)　表中の A ～ G の値を求めよ。

(2)　読書時間と本の冊数の相関係数 r を，小数第 3 位を四捨五入して答えよ。

２節 データの相関

❶ 相関関係

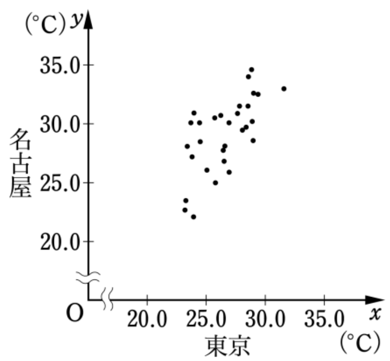
散布図

（教科書 p.175）

次の表は，ある年の 9 月の日ごとの東京，名古屋それぞれの最高気温を x℃，y℃ として示したものである。

日	東京 x(℃)	名古屋 y(℃)	日	東京 x(℃)	名古屋 y(℃)	日	東京 x(℃)	名古屋 y(℃)
1	31.5	33.0	11	26.9	30.1	21	23.1	28.1
2	24.0	30.9	12	23.3	23.5	22	25.8	25.0
3	24.5	28.5	13	28.9	28.6	23	26.9	25.9
4	26.2	30.7	14	26.4	27.8	24	27.6	30.9
5	29.0	32.6	15	23.9	22.1	25	28.5	31.5
6	28.5	34.0	16	28.3	29.6	26	27.8	31.5
7	29.3	32.5	17	26.5	28.1	27	24.4	30.1
8	28.8	34.6	18	23.8	27.2	28	26.5	26.8
9	23.7	30.1	19	25.7	30.5	29	25.1	26.1
10	28.8	30.2	20	28.1	29.5	30	23.2	22.7

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

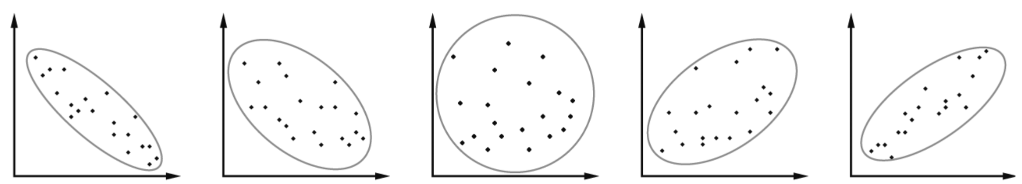


（[1] 散布図 ）：右の図のように，変量 x，y の値の組を座標とする点を平面上にとった図。

（[2] 正の相関関係がある ）：一方が増加すれば他方も増加する傾向

（[3] 負の相関関係がある ）：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 ）：正の相関関係も負の相関関係もみられない



強い ←→ 弱い　（[5] 負の相関関係 ）

弱い ←→ 強い　（[6] 正の相関関係 ）

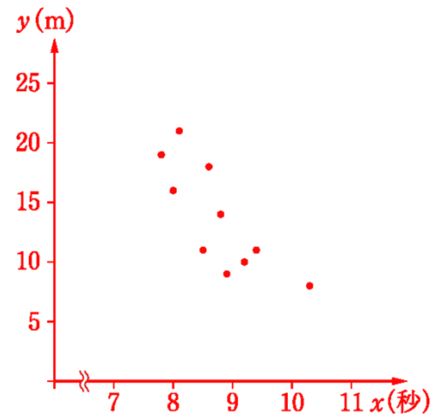
2 つの変量の間に相関関係があるとき

（[7] 相関関係が強い ）：散布図の点の分布が直線状に近づく

（[8] 相関関係が弱い ）：散布図の点の分布が広く散らばる

問1 下の表は，ある高校の 1 年生女子 10 人の体力テストにおける 50m 走の記録を x 秒，ハンドボール投げの記録を ym として示したものである。このとき，x と y の散布図を作成せよ。

50m 走 x（秒）	ハンドボール投げ y (m)	50m 走 x（秒）	ハンドボール投げ y (m)
8.5	11	9.4	11
8.0	16	10.3	8
8.6	18	8.1	21
8.8	14	9.2	10
7.8	19	8.9	9



❷ 相関係数

相関係数

（教科書 p.177）

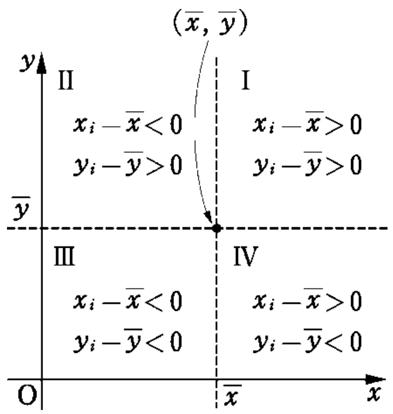
2 つの変量 x，y のデータの値の組を

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

とし，x，y のデータの平均値をそれぞれ $\overline{x}$，$\overline{y}$ とする。

この平均値を座標とする点 $(\overline{x},\ \overline{y})$ は，散布図の点の代表的な位置を表すと考えられ，散布図の点はこのまわりに分布している。

右の図のように，$(\overline{x},\ \overline{y})$ 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 の間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β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。

$(x_1-\overline{x})(y_1-\overline{y})$，$(x_2-\overline{x})(y_2-\overline{y})$，$\cdots$，$(x_n-\overline{x})(y_n-\overline{y})$

の平均値

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

を変量 x，y の（[1]　共分散　）といい，正の相関関係があるときには（[2]　正　）に，負の相関関係があるときは（[3]　負　）となる傾向がある。相関関係がないときは，正の項と負の項が消し合って（[4]　0　）に近い値となる。

共分散は変量 x，y の単位の取り方や散らばり具合に影響を受けるから，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 の標準偏差と 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 s_xs_y で割った値を考える。

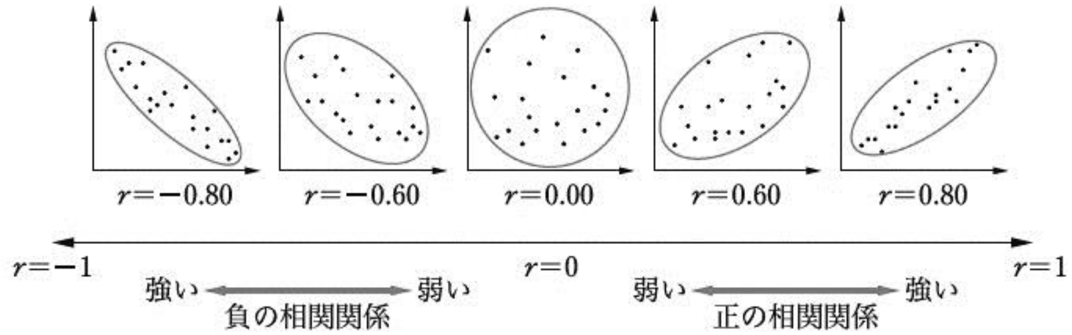
この値を変量 x と y の（[5]　相関係数　）といい r で表す。

相関係数
$r=\dfrac{s_{xy}}{s_xs_y}$ $(x\text{ と }y\text{ の相関係数})=\dfrac{(x\text{ と }y\text{ の共分数})}{(x\text{ の標準偏差})\times(y\text{ の標準偏差})}$

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 人の生徒 a, b, c, d, e が受けた，数学と国語のテストの点数である。
数学（x 点）と国語（y 点）の点数の相関係数 r を，小数第 3 位を四捨五入して答えよ。

	x (点)	y (点)
a	72	69
b	76	76
c	93	88
d	84	75
e	85	87

解 数学の平均点 $\overline{x}$，国語の平均点 $\overline{y}$ は

$\overline{x} = \frac{410}{5} = 82$，$\overline{y} = \frac{395}{5} = 79$

となり，下の表のようにして計算すると，

	x	y	$x-\overline{x}$	$(x-\overline{x})^2$	$y-\overline{y}$	$(y-\overline{y})^2$	$(x-\overline{x})(y-\overline{y})$
a	72	69	-10	100	-10	100	100
b	76	76	-6	36	-3	9	18
c	93	88	11	121	9	81	99
d	84	75	2	4	-4	16	-8
e	85	87	3	9	8	64	24
計	410	395	0	270	0	270	233

$$r = \frac{\frac{1}{5}\cdot 233}{\sqrt{\frac{1}{5}\cdot 270}\sqrt{\frac{1}{5}\cdot 270}} = \frac{233}{\sqrt{270}\sqrt{270}} = \frac{233}{270} = 0.862\cdots \fallingdotseq 0.86$$

問2 下の表は 4 人の生徒 a, b, c, d の数学と英語の小テストの点数である。数学と英語の点数の相関係数を求めよ。

	a	b	c	d
数学	4	7	5	4
英語	6	9	8	9

数学の点数を x 点，英語の点数を y 点とする。
x と y の値を，表の x と y の欄に転記して計算すると

$\overline{x} = \frac{20}{4} = 5$，$\overline{y} = \frac{32}{4} = 8$

となり，下の表のようにして計算すると

$$r = \frac{\frac{1}{4}\cdot 3}{\sqrt{\frac{1}{4}\cdot 6}\sqrt{\frac{1}{4}\cdot 6}} = \frac{3}{4} \div \frac{3}{2} = \frac{1}{2} = 0.5$$

	x	y	$x-\overline{x}$	$(x-\overline{x})^2$	$y-\overline{y}$	$(y-\overline{y})^2$	$(x-\overline{x})(y-\overline{y})$
a	4	6	-1	1	-2	4	2
b	7	9	2	4	1	1	2
c	5	8	0	0	0	0	0
d	4	9	-1	1	1	1	-1
計	20	32	0	6	0	6	3

参考　相関表

（教科書 p.180）

2 つの変量 x と y の間の相関関係を調べるとき，散布図のほかに，x と y の度数分布表を組み合わせた表を用いることがある。そのような表を（[1]　　相関表　　）という。

教科書 175 ページの東京と名古屋それぞれの最高気温を相関表に表すと右のようになる。ただし，この相関表は，たとえば東京の最高気温が 26℃ 以上 28℃ 未満で，名古屋の最高気温が 30℃ 以上 32℃ 未満の日が 4 日あることを表している。

名古屋(℃) ＼ 東京(℃)	22～24	24～26	26～28	28～30	30～32
34～36				2	
32～34				2	1
30～32	1	3	4	2	
28～30	1	1	1	3	
26～28	1	1	2		
24～26		1	1		
22～24	3				

変量 x，y の組の個数が多くなると，散布図では点が重なり，相関関係が把握しにくくなる場合がある。そのような場合，相関表を用いるとよい。

問 1　下の表は，ある高校の 1 年生男子 10 人の体力テストにおける握力を $x$kg，ハンドボール投げの記録を ym として示したものである。このとき，x と y の相関表を作成せよ。ただし，x は 34kg 以上36kg 未満，y は 26m 以上 28m 未満という区間から，幅を一定にとるものとする。

握力 x (kg)	ハンドボール投げ y (m)	握力 x (kg)	ハンドボール投げ y (m)
37	31	43	30
39	33	38	26
34	28	37	31
38	30	38	32
40	37	36	31

ハンドボール投げ y(m) ＼ 握力 x(kg)	34～36	36～38	38～40	40～42	42～44
36～38				1	
34～36					
32～34			2		
30～32		3	1		1
28～30	1				
26～28			1		

Training

（教科書 p.181）

4 次の a, b, c, d の散布図に対応する相関係数を下の 4 つから選べ。

図 a

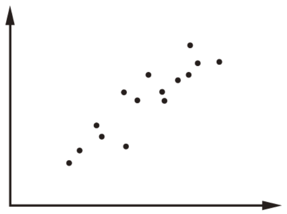


図 b

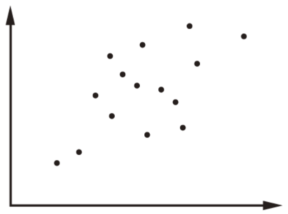


図 c

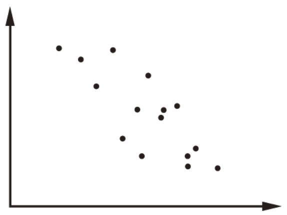
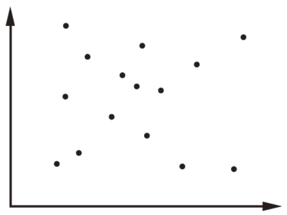


図 d



$r = 0.6$, $r = 0$, $r = -0.8$, $r = 0.9$

図 a と b は正の相関関係があり，b よりも a の方が直線状に近く，相関関係が強い。

図 c は負の相関関係がある。

図 d は相関関係がない。

図 a は　$r = 0.9$

図 b は　$r = 0.6$

図 c は　$r = -0.8$

図 d は　$r = 0$

5 次の表は，あるクラスの生徒 10 人に，10 日間の合計読書時間 x（時間）と読んだ本の冊数 y（冊）を調査し，その結果をまとめたものである。このとき，次の問に答えよ。

	x	y	$x-\overline{x}$	$(x-\overline{x})^2$	$y-\overline{y}$	$(y-\overline{y})^2$	$(x-\overline{x})(y-\overline{y})$
生徒 1	14	7	A	B	C	D	E
生徒 2	8	3	−2	4	−2	4	4
⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮
生徒 9	5	3	−5	25	−2	4	10
生徒 10	11	8	1	1	3	9	3
計	100	50	F	160	G	40	58

(1) 表中の A ～ G の値を求めよ。

$\overline{x} = \frac{100}{10} = 10$，$\overline{y} = \frac{50}{10} = 5$ であるから

$\mathrm{A} = 14 - 10 = 4$

$\mathrm{B} = 4^2 = 16$

$\mathrm{C} = 7 - 5 = 2$

$\mathrm{D} = 2^2 = 4$

$\mathrm{E} = 4 \times 2 = 8$

$\mathrm{F} = (14-10) + (8-10) + \cdots + (5-10) + (11-10)$

$= (14 + 8 + \cdots + 5 + 11) - 10 \times 10$

$= 100 - 100 = 0$

$\mathrm{G} = (7-5) + (3-5) + \cdots + (3-5) + (8-5)$

$= (7 + 3 + \cdots + 3 + 8) - 5 \times 10$

$= 50 - 50 = 0$

(2) 読書時間と本の冊数の相関係数 r を，小数第 3 位を四捨五入して答えよ。

$$r = \frac{\frac{1}{10}\cdot 58}{\sqrt{\frac{1}{10}\cdot 160}\sqrt{\frac{1}{10}\cdot 40}} = \frac{58}{\sqrt{160}\sqrt{40}}$$

$$= \frac{58}{80} = \frac{29}{40} = 0.725 \fallingdotseq 0.73$$