

第 42 講 相関係数

学習目標

- ・ 相関係数が -1 以上 1 以下の値をとることを理解できる。
- ・ 相関係数の符号から、正の相関・負の相関を判断できる。
- ・ 相関係数の絶対値から、直線的な相関の強さを読み取れる。
- ・ 小さな 2 変量データについて、偏差・標準偏差を使って相関係数を求められる。

散布図の傾向を 1 つの数で表す

散布図だけでなく数値でも表す

前講では、2 変量データを散布図に表し、点の並びから正の相関・負の相関・相関がほとんど見られない状態を読み取りました。

たとえば、点が全体として右上がりなら正の相関、右下がりなら負の相関と判断しました。

しかし、散布図を見ただけでは「どの程度はっきりした直線的な傾向なのか」を数値では表していません。

そこで、2 つの数量の直線的な関係を -1 から 1 までの数で表したものが**相関係数**です。

相関係数

2 つの数量の直線的な相関の向きと強さを表す数を**相関係数**といいます。

相関係数を r とすると、

$$-1 \leq r \leq 1$$

です。

まず符号を見る

相関係数 r の符号は、散布図の向きと対応します。

- ・ $r > 0$: 正の相関
- ・ $r < 0$: 負の相関
- ・ r が 0 に近い : 直線的な相関は弱い

まず正か負かを見て、相関の向きを判断します。

絶対値で相関の強さを見る

1 や -1 に近いほど直線的な傾向が強い

相関係数では、符号だけでなく絶対値 $|r|$ に注目します。

たとえば、

- ・ $r = 0.9$
- ・ $r = 0.3$

はどちらも正の相関ですが、 0.9 の方が 1 に近く、より強い正の相関を表します。

また、

- ・ $r = -0.8$
- ・ $r = -0.2$

はどちらも負の相関ですが、 $|-0.8| = 0.8$ の方が大きいため、 $r = -0.8$ の方が強い負の相関です。

符号と絶対値の役割

相関係数では、

- ・ 符号を見ると、相関の向きが分かる。
- ・ 絶対値を見ると、直線的な相関の強さが分かる。

$|r|$ が 1 に近いほど強く、0 に近いほど弱い直線的な相関を表します。

負の数では大小だけで比べない

$r = -0.9$ と $r = -0.4$ を比べると、数としては $-0.9 < -0.4$ です。

しかし、相関の強さを比べるときは絶対値を使います。

$$|-0.9| = 0.9, \quad |-0.4| = 0.4$$

なので、 $r = -0.9$ の方が強い負の相関です。

散布図と相関係数を対応させる

右上がりなら正の値

次の 2 変量データを考えます。

$$(1, 2), (2, 4), (3, 6), (4, 8)$$

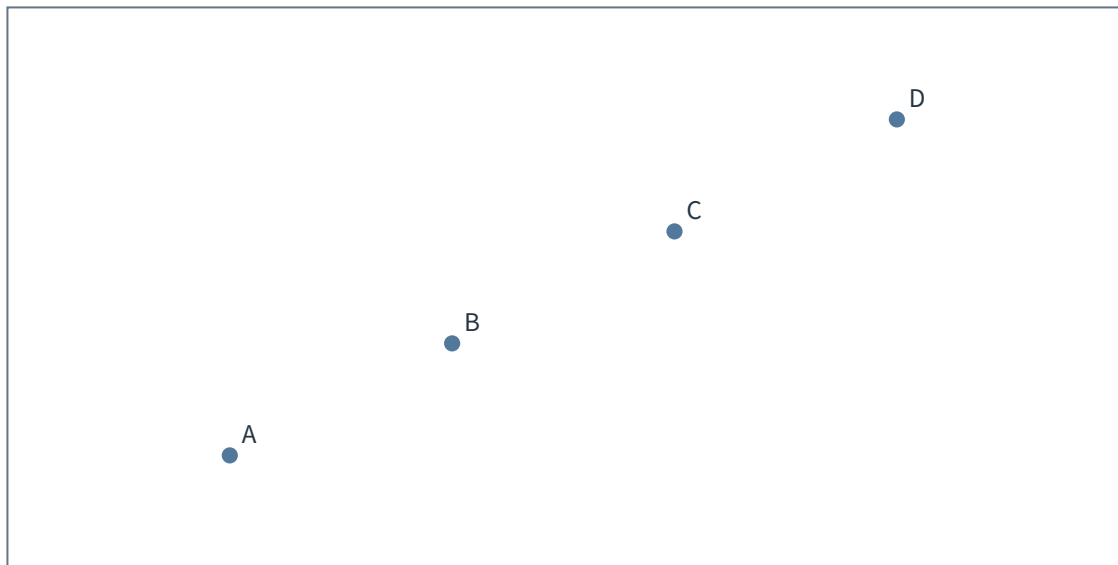


図 1 点が一直線上に右上がり並ぶ散布図

点は一直線上に右上がり並んでいます。

このような場合、相関係数は $r = 1$ です。

これは最も強い正の直線的な相関を表します。

右下がりなら負の値

次のデータでは、

$$(1, 8), (2, 6), (3, 4), (4, 2)$$

x が大きくなるにつれて y が一定の割合で小さくなります。

点は一直線上に右下がりであり、相関係数は $r = -1$ です。
これは最も強い負の直線的な相関を表します。

$$r = 1 \text{ と } r = -1$$

- $r = 1$: 点が右上がりの 1 本の直線上に並ぶ。
 - $r = -1$: 点が右下がりの 1 本の直線上に並ぶ。
- どちらも $|r| = 1$ なので、直線的な相関の強さは最大です。

相関係数を計算する考え方

平均からのずれを 2 つの数量で比べる

相関係数を計算するときは、2 つの数量がそれぞれ平均値からどのようにずれているかを見ます。
 x の平均値を $\bar{x}$ 、 y の平均値を $\bar{y}$ とします。
各組について、

$$x - \bar{x}$$

と

$$y - \bar{y}$$

を求めます。
2 つの偏差が同じ符号になりやすければ、その積は正になりやすく、正の相関につながります。
反対に、一方が正のときもう一方が負になりやすければ、積は負になりやすく、負の相関につながります。

共分散

x の偏差と y の偏差の積を求め、その平均をとったものを**共分散**といいます。
 n 組のデータについて、共分散は

$$\frac{(x_1 - \bar{x})(y_1 - \bar{y}) + \cdots + (x_n - \bar{x})(y_n - \bar{y})}{n}$$

です。

共分散を標準偏差で調整する

共分散は 2 つの数量の動き方を表しますが、その値はデータの単位や散らばり方の影響を受けます。
そこで、 x の標準偏差と y の標準偏差を使って調整します。
 x の標準偏差を s_x 、 y の標準偏差を s_y 、共分散を s_{xy} とすると、相関係数は

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

で求めます。

相関係数を求める手順

小さなデータで相関係数を求めるときは、

1. x と y の平均値を求める。
2. x と y の偏差を求める。
3. 偏差の積の平均から共分散を求める。
4. x と y の標準偏差を求める。
5. 共分散を 2 つの標準偏差の積で割る。

という順に進めます。

簡単なデータで計算する

完全な正の相関を計算で確認する

次の 3 組のデータを考えます。

$$(1, 2), (2, 4), (3, 6)$$

x の平均値は 2、 y の平均値は 4 です。

x	$x - 2$	y	$y - 4$	偏差の積
1	-1	2	-2	2
2	0	4	0	0
3	1	6	2	2

共分散は

$$s_{xy} = \frac{2 + 0 + 2}{3} = \frac{4}{3}$$

です。

x の分散は $\frac{2}{3}$ なので、

$$s_x = \sqrt{\frac{2}{3}}$$

です。

y の分散は $\frac{8}{3}$ なので、

$$s_y = \sqrt{\frac{8}{3}}$$

です。

したがって、

$$r = 1$$

となります。

散布図で見た「一直線上の右上がり」という特徴と一致します。

相関係数は直線的な関係を見る

相関係数は、2 つの数量の直線的な関係の向きと強さを表す値です。

r が 0 に近いときは、右上がりまたは右下がりの直線的な傾向が弱いと読み取ります。

例題

例題 1

次の相関係数について、正の相関・負の相関のどちらかを答え、相関の強さを比べなさい。

(1) $r = 0.82$

(2) $r = -0.35$

(3) $r = -0.91$

例題 1 の解答・解説

(1) $r > 0$ なので正の相関です。

(2) $r < 0$ なので負の相関です。

(3) $r < 0$ なので負の相関です。

強さは絶対値で比べます。

$$0.91 > 0.82 > 0.35$$

なので、強い順に (3)、(1)、(2) です。

例題 2

次の 3 つの散布図について相関係数が、それぞれ 0.9、 -0.8 、0.1 のいずれかであるとします。

- ・ A：点が全体としてはっきり右上がりに並ぶ。
- ・ B：点が全体としてはっきり右下がりに並ぶ。
- ・ C：点にほとんど右上がり・右下がりの傾向が見られない。

A、B、C に対応する相関係数を答えなさい。

例題 2 の解答・解説

A は強い正の相関なので 0.9 です。

B は強い負の相関なので -0.8 です。

C は直線的な相関が弱いので、0 に近い 0.1 です。

例題 3

次の 2 変量データの相関係数を求めなさい。

$$(1, 3), (2, 2), (3, 1)$$

例題 3 の解答・解説

x の平均値は 2、 y の平均値も 2 です。

x の偏差は $-1, 0, 1$ 、 y の偏差は $1, 0, -1$ です。

偏差の積は $-1, 0, -1$ なので、共分散は

$$s_{xy} = -\frac{2}{3}$$

です。

x と y の分散はいずれも $\frac{2}{3}$ なので、標準偏差も等しくなります。

したがって、

$$r = -1$$

です。

点が右下がりの 1 本の直線上に並ぶことと一致します。

演習問題

練習問題 1

次の相関係数について、正の相関か負の相関か答えなさい。

(1) $r = 0.64$

(2) $r = -0.72$

(3) $r = 0.08$

練習問題 2

次の相関係数を、直線的な相関が強いものから順に並べなさい。

0.25, -0.93 , 0.71

練習問題 3

散布図の点が全体として右下がりに強くまとまっていました。

次のうち、この散布図の相関係数として最も適切なものを選びなさい。

- (ア) 0.88
- (イ) -0.86
- (ウ) 0.04

練習問題 4

散布図の点に、はっきりした右上がり・右下がりの傾向がほとんど見られませんでした。

次のうち、この散布図の相関係数として最も適切なものを選びなさい。

- (ア) 0.97
- (イ) -0.95
- (ウ) 0.06

練習問題 5

2 つのデータ A、B の相関係数が次のように与えられています。

データ A : $r = -0.78$

データ B : $r = 0.52$

- (1) それぞれ相関の向きを答えなさい。
- (2) 直線的な相関が強いのはどちらですか。

練習問題 6

次の 2 変量データについて、 x と y の平均値と、各組の偏差の積を求めなさい。

$(1, 2), (2, 4), (3, 6)$

練習問題 7

次の 2 変量データの相関係数を求めなさい。

$(1, 1), (2, 2), (3, 3)$

練習問題 8

次の 2 変量データについて、共分散が 0 になることを確かめ、相関係数を求めなさい。

$(1, 1), (2, 3), (3, 3), (4, 1)$

練習問題 9

ある生徒は、 $r = -0.9$ と $r = 0.6$ を比べて、

「 $0.6 > -0.9$ だから、 $r = 0.6$ の方が相関が強い」

と考えました。

この考えの誤りを説明し、どちらの方が直線的な相関が強いのか答えなさい。

演習問題の解答・解説

練習問題 1 の解答・解説

- (1) $r > 0$ なので正の相関です。
- (2) $r < 0$ なので負の相関です。
- (3) $r > 0$ なので向きとしては正ですが、0 に近いので直線的な相関は弱いと読み取れます。

練習問題 2 の解答・解説

相関の強さは絶対値で比べます。

それぞれの絶対値は 0.25, 0.93, 0.71 です。

したがって、強い順に

$-0.93, 0.71, 0.25$

です。

練習問題 3 の解答・解説

右下がりなので相関係数は負です。

さらに強くまとまっているので、絶対値は 1 に近いと考えます。

したがって、最も適切なのは (イ) の -0.86 です。

練習問題 4 の解答・解説

はっきりした直線的な傾向が見られないので、相関係数は 0 に近い値が適切です。

したがって、(ウ) の 0.06 です。

練習問題 5 の解答・解説

(1) データ A は $r < 0$ なので負の相関、データ B は $r > 0$ なので正の相関です。

(2) 強さは絶対値で比べます。

$$|-0.78| = 0.78, \quad |0.52| = 0.52$$

なので、データ A の方が直線的な相関が強いです。

練習問題 6 の解答・解説

x の平均値は 2、 y の平均値は 4 です。

偏差は、

- (1, 2) では $-1, -2$
- (2, 4) では $0, 0$
- (3, 6) では $1, 2$

です。

したがって、偏差の積は順に

$$2, 0, 2$$

です。

練習問題 7 の解答・解説

x と y の平均値はいずれも 2 です。

x と y の偏差はいずれも $-1, 0, 1$ です。

したがって、共分散は

$$\frac{1 + 0 + 1}{3} = \frac{2}{3}$$

です。

x と y の標準偏差はいずれも $\sqrt{\frac{2}{3}}$ です。

よって、

$$r = \frac{\frac{2}{3}}{\sqrt{\frac{2}{3}}\sqrt{\frac{2}{3}}} = 1$$

です。

練習問題 8 の解答・解説

x の平均値は $\frac{5}{2}$ 、 y の平均値は 2 です。

x の偏差は、

$$-\frac{3}{2}, -\frac{1}{2}, \frac{1}{2}, \frac{3}{2}$$

です。

y の偏差は、

$$-1, 1, 1, -1$$

です。

偏差の積の合計は

$$\frac{3}{2} - \frac{1}{2} + \frac{1}{2} - \frac{3}{2} = 0$$

です。

したがって、共分散は 0 です。

標準偏差はどちらも 0 ではないので、

$$r = 0$$

です。

練習問題 9 の解答・解説

相関の強さを比べるときは、相関係数そのものの大小ではなく絶対値を比べます。

$$|-0.9| = 0.9, \quad |0.6| = 0.6$$

です。

したがって、 $r = -0.9$ の方が直線的な相関が強いです。

符号が負であることは相関が弱いことを意味せず、負の相関であることを表します。