

相関係数

監修・執筆
湯浅弘一

今回学ぶこと

データの分析の最終回。今までの代表値を複合しながら 2 種類のデータの関係を数値化します。相関係数は、相関がどの程度強いのかを表しています。

学習のポイント

- ① 2 種類のデータの相関関係を 1 つの数値で表す
- ② 共分散と相関係数
- ③ 実際のデータから相関係数を求める

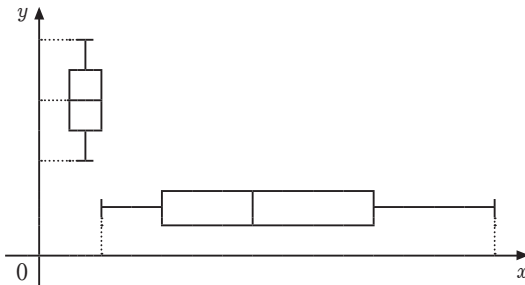
ポイント1 2種類のデータの相関関係を 1 つの数値で表す

2 種類のデータの散らばりは散布図で見ることができました。
これを数値化するには？

散らばりを見るには

- 四分位数 (中央値から見る)
- 分散 (平均値から見る)

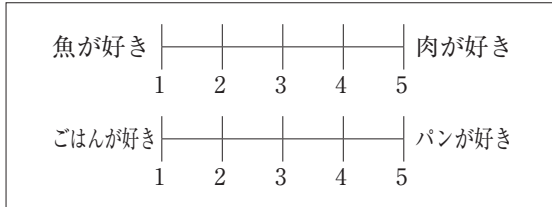
2 種類のデータの関係を四分位数を使って表すと、イメージは、



これだと x と y の関係は見えません !!
そこで、平均値から見てみると…

例

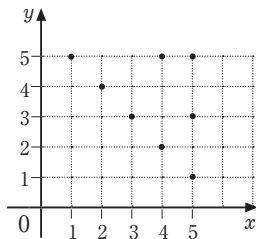
あるグループ 8 人にごはんとパンの好きな度合いと、魚と肉の好きな度合いのアンケートをとりました。アンケートは次の通りです。



回答は、1～5 の整数値に限定しました。結果は以下の通りです。

メンバー	A	B	C	D	E	F	G	H
魚と肉 (x)	1	2	3	4	4	5	5	5
ごはん和パン (y)	5	4	3	2	5	1	3	5

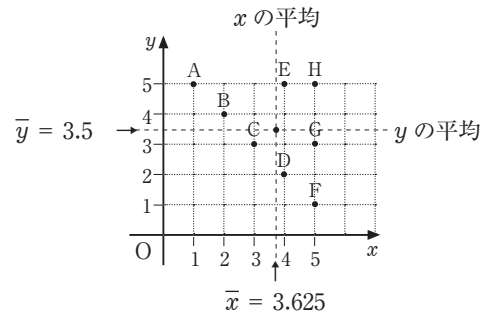
これを散布図で表すと、相関が見える？



ここで平均から見ることにすると

$$\bar{x} = \frac{1}{8} (1+2+3+4+4+5+5+5) = \frac{29}{8} = 3.625$$

$$\bar{y} = \frac{1}{8} (5+4+3+2+5+1+3+5) = \frac{28}{8} = 3.5$$



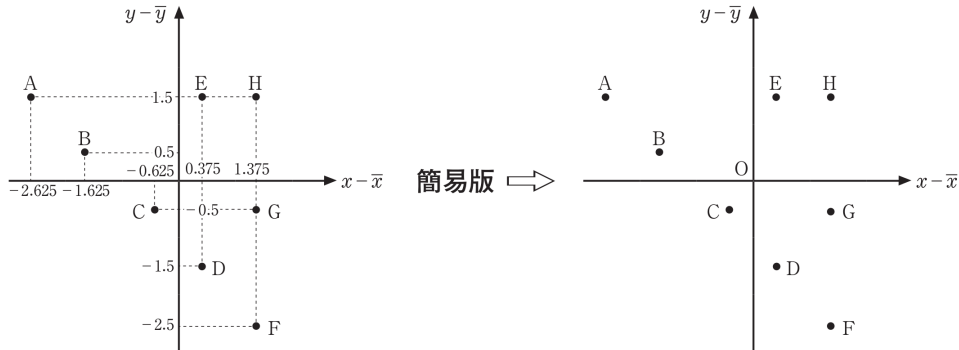
これは見にくいので x の平均と y の平均の交点を O にするように変化を与えます。

それが **偏差** です。偏差 = (個々のデータ) - 平均値 でした。

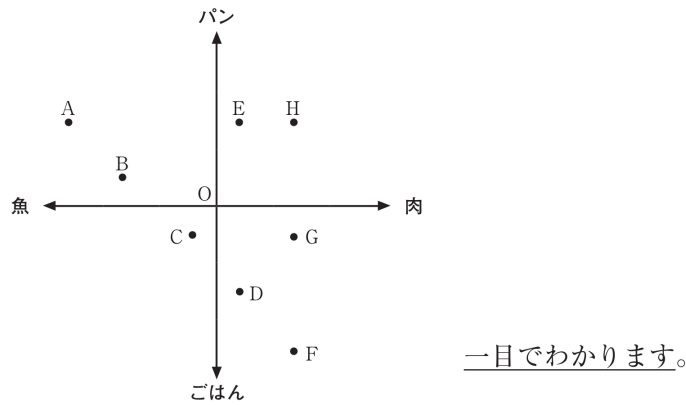
先ほどのデータを使って

メンバー	A	B	C	D	E	F	G	H
x	1	2	3	4	4	5	5	5
y	5	4	3	2	5	1	3	5
$x - \bar{x}$	-2.625	-1.625	-0.625	0.375	0.375	1.375	1.375	1.375
$y - \bar{y}$	1.5	0.5	-0.5	-1.5	1.5	-2.5	-0.5	1.5

下の2段を図示すると、

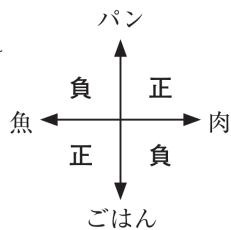


これにアンケート項目を入れると…、



これを数値化すると、

$(x - \bar{x})(y - \bar{y})$ を作る



⇨ エリアで符号が決まります。この符号は、 $x > 0$ (肉好き) と $y > 0$ (パン好き) の積 (かけ算) で決まります。

$(x - \bar{x})(y - \bar{y})$ の値の絶対値が好きな度合いです。

メンバー	A	B	C	D	E	F	G	H
$x - \bar{x}$	-2.625	-1.625	-0.625	0.375	0.375	1.375	1.375	1.375
$y - \bar{y}$	1.5	0.5	-0.5	-1.5	1.5	-2.5	-0.5	1.5
$(x - \bar{x})(y - \bar{y})$	-3.9375	-0.8125	0.3125	-0.5625	0.5625	-3.4375	-0.6875	2.0625

↑
Aさんはとても魚とパンが好き

↑
Fさんはとても肉とごはんが好き

↑
Hさんはとても肉とパンが好き

ポイント2 共分散と相関係数

上の積 $(x - \bar{x})(y - \bar{y})$ の平均値を x と y の **共分散** といいます。

平均値を中心に考えたときの上下左右への散らばり具合の目安(イメージ)がわかります。

そして、その共分散をもとにして

$$\text{相関係数} = \frac{x \text{ と } y \text{ の 共分散}}{\sqrt{x \text{ の 分散}} \sqrt{y \text{ の 分散}}} = r \quad \text{と定めます。}$$

この値 r は必ず、 $-1 \leq r \leq 1$ を満たします。そして、

$$\begin{cases} r \geq 0.7 \\ r \leq -0.7 \end{cases} \quad \text{: 強い相関がある}$$

$$\begin{cases} 0.4 \leq r < 0.7 \\ -0.7 < r \leq -0.4 \end{cases} \quad \text{: 中程度の相関がある (やや相関がある)}$$

$$\begin{cases} 0.2 < r < 0.4 \\ -0.4 < r < -0.2 \end{cases} \quad \text{: 弱い相関がある}$$

$-0.2 \leq r \leq 0.2$ は、相関がないとされています。

■相関係数を求める手順

手順 1 与えられたデータから相関係数を求めやすいように表を作る。

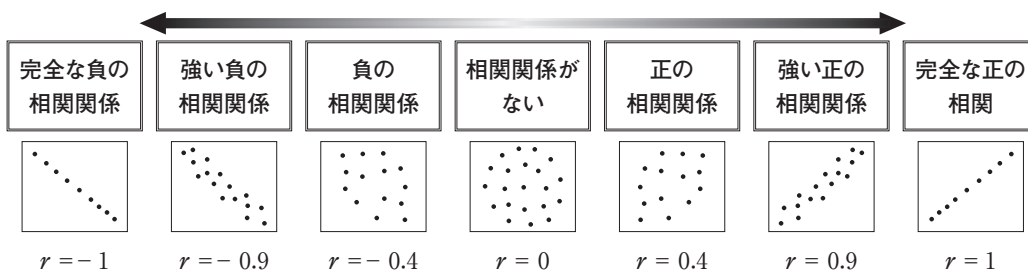
x	y	$x - \bar{x}$	$y - \bar{y}$	$(x - \bar{x})^2$	$(y - \bar{y})^2$	$(x - \bar{x})(y - \bar{y})$
		③	④	⑤	⑥	⑦
		③	④	⑤	⑥	⑦
		③	④	⑤	⑥	⑦
.....
①	②					

- 作業 1 : x の平均値 \bar{x} を求めて①に書き込む。そして、 x の平均値 \bar{x} を求める。
- 作業 2 : y の平均値 \bar{y} を求めて②に書き込む。そして、 y の平均値 \bar{y} を求める。
- 作業 3 : x のデータから平均値 \bar{x} をひいて③に書き込む。
- 作業 4 : y のデータから平均値 \bar{y} をひいて④に書き込む。
- 作業 5 : ③の各データを 2 乗して⑤に書き込む。
- 作業 6 : ④の各データを 2 乗して⑥に書き込む。
- 作業 7 : ③と④のデータの積を計算して⑦に書き込む。
- 作業 8 : ⑤の平均が x の分散。
- 作業 9 : ⑥の平均が y の分散。
- 作業 10 : ⑦の平均を計算する。(共分散という)

手順 2 作成した表をもとにして相関係数 r を求める。

$$\frac{\text{共分散⑩}}{\sqrt{x\text{の分散⑧}} \times \sqrt{y\text{の分散⑨}}}$$

相関係数の読み取り方



分析 1 : r の範囲は、 $-1 \leq r \leq 1$ である。

分析 2 : r の値が 1 に近いほど正の相関が強くなる。(右上がりの直線上)

分析 3 : r の値が -1 に近いほど負の相関が強くなる。(右下がりの直線上)

分析 4 : r の値が 0 に近いほど相関は弱くなる。(散らばった状態)

 使ってみよう!

例題

下表はある月の連続 6 日間の最高気温とある店のおでんの売り上げのデータである。

このデータの最高気温 x (°C) とおでんの売り上げ y (万円) の相関係数 r を小数第 3 位を四捨五入して求めよ。

	1 日目	2 日目	3 日目	4 日目	5 日目	6 日目
x (°C)	10	7	6	12	9	10
y (万円)	13	19	19	12	15	12

答え

	x	y	$x - \bar{x}$	$y - \bar{y}$	$(x - \bar{x})^2$	$(y - \bar{y})^2$	$(x - \bar{x})(y - \bar{y})$
1 日目	10	13	1	-2	1	4	-2
2 日目	7	19	-2	4	4	16	-8
3 日目	6	19	-3	4	9	16	-12
4 日目	12	12	3	-3	9	9	-9
5 日目	9	15	0	0	0	0	0
6 日目	10	12	1	-3	1	9	-3
合計	54	90	0	0	24	54	-34

$$\bar{x} = \frac{1}{6}(10+7+6+12+9+10) = \frac{54}{6} = 9$$

$$\bar{y} = \frac{1}{6}(13+19+19+12+15+12) = \frac{90}{6} = 15$$

$$x \text{ の分散は、} \frac{(x - \bar{x})^2 \text{ の総和}}{6} = \frac{24}{6} = 4$$

$$y \text{ の分散は、} \frac{(y - \bar{y})^2 \text{ の総和}}{6} = \frac{54}{6} = 9$$

$$x \text{ と } y \text{ の共分散は、} \frac{(x-\bar{x})(y-\bar{y})\text{の総和}}{6} = \frac{-34}{6} = -\frac{17}{3}$$

$$\begin{aligned} \text{相関係数 } r &= \frac{x \text{ と } y \text{ の共分散}}{\sqrt{x \text{ の分散}} \sqrt{y \text{ の分散}}} \\ &= \frac{\left(-\frac{17}{3}\right)}{\sqrt{4} \sqrt{9}} \\ &= -\frac{17}{18} \\ &= -0.9444\cdots \doteq -0.94 \\ &\quad \text{(強い負の相関がある)} \end{aligned}$$

ポイント3 実際のデータから相関係数を求める

問題 総復習です！

20人に数学と英語のテストを実施しその結果をまとめたところ、下のような相関表になった。表中の数値は、数学の得点と英語の得点の組み合わせに対応する人数を表している。

英語(点) 数学(点)	10	9	8	7	6	5	4	3	2	1
10			1							
9	1									
8		1	1							
7				2						
6						1	2			
5					1	1				
4					2					1
3								1		
2									3	
1							1			1

以下の値を求めなさい。

- (1) 英語の中央値は？
- (5) 数学の平均値は？
- (2) 英語の箱ひげ図を書け。
- (6) 数学の分散は？
- (3) 英語の平均値は？
- (7) 英語と数学の共分散は？
- (4) 英語の分散は？
- (8) 英語と数学の相関係数は？

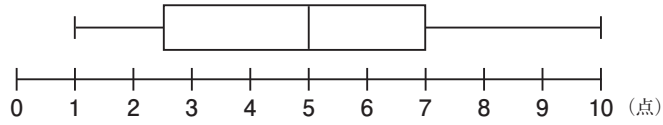
答え

英語と数学の得点をそれぞれ x 、 y として次の表を作る。

ここで $\bar{x} = \frac{100}{20} = 5$ 、 $\bar{y} = \frac{100}{20} = 5$ 、 x は小さい順に並べます。

生徒	x	y	$x - \bar{x}$	$y - \bar{y}$	$(x - \bar{x})^2$	$(y - \bar{y})^2$	$(x - \bar{x})(y - \bar{y})$
①	1	1	-4	-4	16	16	16
②	1	4	-4	-1	16	1	4
③	2	2	-3	-3	9	9	9
④	2	2	-3	-3	9	9	9
⑤	2	2	-3	-3	9	9	9
⑥	3	3	-2	-2	4	4	4
⑦	4	1	-1	-4	1	16	4
⑧	4	6	-1	1	1	1	-1
⑨	4	6	-1	1	1	1	-1
⑩	5	5	0	0	0	0	0
⑪	5	6	0	1	0	1	0
⑫	6	4	1	-1	1	1	-1
⑬	6	4	1	-1	1	1	-1
⑭	6	5	1	0	1	0	0
⑮	7	7	2	2	4	4	4
⑯	7	7	2	2	4	4	4
⑰	8	8	3	3	9	9	9
⑱	8	10	3	5	9	25	15
⑲	9	8	4	3	16	9	12
⑳	10	9	5	4	25	16	20
計	100	100	--	--	136	136	115

- (1) 生徒⑩⑪の英語の平均なので、5 (点)
 (2) 英語の第 1 四分位数は、⑤と⑥の平均 2.5 (点)
 第 3 四分位数は⑮と⑯の平均 7 (点)
 第 2 四分位数は (1) より 5 (点) なので



- (3) $\bar{x} = \frac{x \text{の総和}}{20} = \frac{100}{20} = \underline{5}$ (点)
 (4) $\frac{(x - \bar{x})^2 \text{の総和}}{20} = \frac{136}{20} = \underline{6.8}$ (点)
 (5) $\bar{y} = \frac{y \text{の総和}}{20} = \frac{100}{20} = \underline{5}$ (点)
 (6) $\frac{(y - \bar{y})^2 \text{の総和}}{20} = \frac{136}{20} = \underline{6.8}$ (点)
 (7) $\frac{(x - \bar{x})(y - \bar{y}) \text{の総和}}{20} = \frac{115}{20} = \underline{5.75}$ (点)
 (8) $\frac{5.75}{\sqrt{6.8} \times \sqrt{6.8}} = \frac{5.75}{6.8} = \frac{575}{680} = 0.845\cdots \approx \underline{0.85}$

