

相関係数

監修・執筆
 湯浅弘一

今回学ぶこと

データ分析の最終回。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

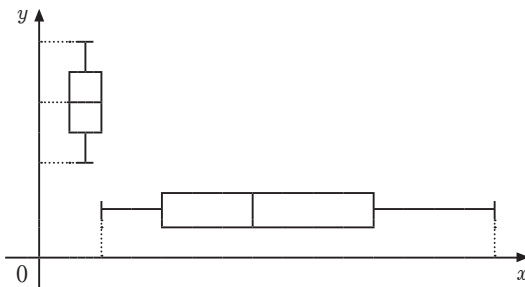
ポイント1 2種類のデータの相関関係を 1 つの数値で表す

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、

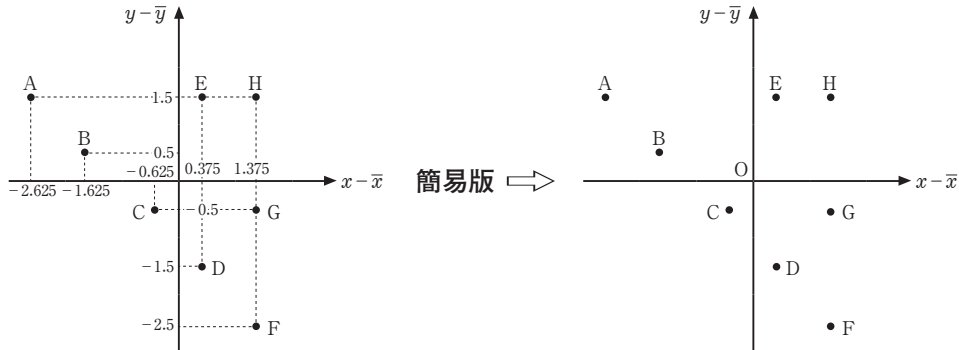


これだと x と y の関係は見えません!!
 そこで、平均値から見てみると…

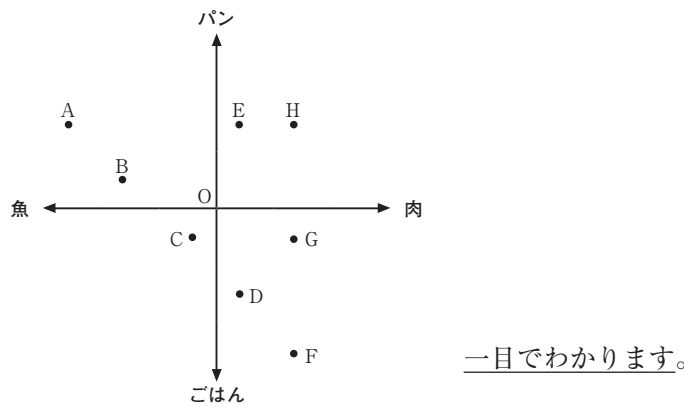
先ほどのデータを使って

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、

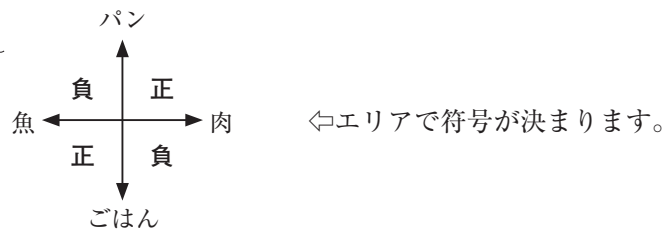


これにアンケート項目を入れると…、



これを数値化すると、

$(x - \bar{x})(y - \bar{y})$ を作る



■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					

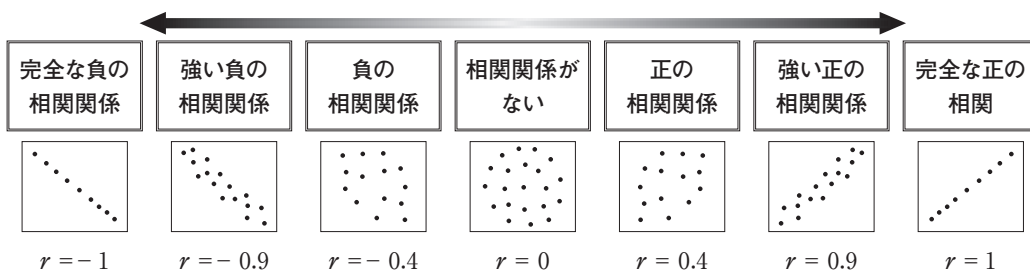
- 作業 1 : x の平均値 \bar{x} を求めて①に書き込む。
- 作業 2 : y の平均値 \bar{y} を求めて②に書き込む。
- 作業 3 : x のデータから平均値 \bar{x} をひいて③に書き込む。
- 作業 4 : y のデータから平均値 \bar{y} をひいて④に書き込む。
- 作業 5 : ③の各データを2乗して⑤に書き込む。
- 作業 6 : ④の各データを2乗して⑥に書き込む。
- 作業 7 : ③と④のデータの積を計算して⑦に書き込む。
- 作業 8 : ⑤の平均が x の分散。
- 作業 9 : ⑥の平均が y の分散。
- 作業 10 : ⑦の平均を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

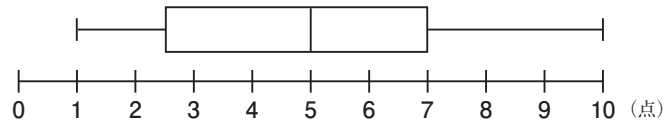
$$\bar{x} = \frac{1}{6}(10 + 7 + 6 + 12 + 9 + 10) = \frac{54}{6} = 9$$

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

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

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

- (1) 生徒⑩⑪の英語の平均なので、5 (点)
 (2) 英語の第一四分位数は、⑤と⑥の平均 2.5 (点)
 第三四分位数は⑮と⑯の平均 7 (点)
 第二四分位数は (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) $\frac{5.75}{\sqrt{6.8} \times \sqrt{6.8}} = \frac{5.75}{6.8} = \frac{575}{680} = 0.845\cdots \approx \underline{0.85}$

