

第II部 統計

8 直線上に並んだ値の相関係数

ポイント

- 一次変換した値
- 変数変換後の平均と分散

- 変数変換後の共分散と相関係数

8.1 はじめに

- 変数を定数倍し、ある定数を加えてみます。
- 一見別の数字ですが、実は元とあんまり変わっていません。
- 同じ数字同士なので強い関係を持っています。

8.2 平方根

二乗すると a になる数を、 a の平方根という。

$a > 0$ のとき a の平方根は正と負の二つあり、それらの絶対値は等しい。

その正の平方根を $\sqrt{a}$ であらわす。負の平方根は $-\sqrt{a}$ である。

これら二つを合わせて $\pm\sqrt{a}$ とあらわす。

0 の平方根は 0 だけであり、 $\sqrt{0} = 0$ と定義する。

$$a > 0 \text{ のとき } \sqrt{a}^2 = (-\sqrt{a})^2 = a, \quad \sqrt{a} > 0 \quad (1)$$

$$a = 0 \text{ のとき } \sqrt{0} = 0 \quad (2)$$

8.2.1 b^2 の平方根

$b^2 \geq 0$ である。 $b^2 > 0$ ならば、二乗すると b^2 になる数は、 $-b$ と b である。それらのうち正の値が $\sqrt{b^2}$ である。

$$b < 0 \implies -b > 0 \quad (3)$$

である。 $b^2 = 0$ ならば、 $b = 0$ なので定義により $\sqrt{b^2} = b$ である。したがって、

$$b \geq 0 \implies \sqrt{b^2} = b \quad (4)$$

$$b < 0 \implies \sqrt{b^2} = -b \quad (5)$$

である。これらをまとめて、

$$\sqrt{b^2} = |b| \quad (6)$$

である。

8.2.2 根号を含む計算

根号を含む計算では以下の関係が成り立つ。ここで $a > 0$, $b > 0$ とする。

$$\sqrt{ab} = \sqrt{a}\sqrt{b} \quad (7)$$

$$\sqrt{\frac{a}{b}} = \frac{\sqrt{a}}{\sqrt{b}} \quad (8)$$

8.3 総和

定義

n 個の数値があり、 x_i を i 番目の値とする。全ての x_i を足し合わせた値を『総和』と呼び、 $\sum_{i=1}^n x_i$ という記号であらわす。

$$\sum_{i=1}^n x_i = x_1 + x_2 + \cdots + x_n \quad (9)$$

8.3.1 総和の公式

総和記号の定義から以下の公式を導き出すことができる。

$$\sum_{i=1}^n c = nc \quad ; c \text{ は定数} \quad (10)$$

$$\sum_{i=1}^n ax_i = a \sum_{i=1}^n x_i \quad ; a \text{ は定数} \quad (11)$$

$$\sum_{i=1}^n (x_i + y_i) = \sum_{i=1}^n x_i + \sum_{i=1}^n y_i \quad (12)$$

8.3.2 総和の計算において自明な事

総和の公式を組み合わせることにより以下の関係は容易に確認することができる。

$$\sum_{i=1}^n (ax_i + c) = a \sum_{i=1}^n x_i + nc \quad (13)$$

$$\sum_{i=1}^n (x_i - y_i) = \sum_{i=1}^n x_i - \sum_{i=1}^n y_i \quad (14)$$

$$\sum_{i=1}^n (x_i + y_i + z_i) = \sum_{i=1}^n x_i + \sum_{i=1}^n y_i + \sum_{i=1}^n z_i \quad (15)$$

問題 II-8-1

以下の式を計算しなさい。

$$\sum_{i=1}^n (ax_i + by_i + c)$$

解例 II-8-1

$$\begin{aligned}\sum_{i=1}^n (ax_i + by_i + c) &= \sum_{i=1}^n ax_i + \sum_{i=1}^n by_i + \sum_{i=1}^n c \\ &= a \sum_{i=1}^n x_i + b \sum_{i=1}^n y_i + nc\end{aligned}$$

8.4 総和と平均

x_i の総和を n で割った値を $\bar{x}$ であらわし平均と呼ぶ。

$$\bar{x} = \frac{1}{n} \sum_{i=1}^n x_i \quad (16)$$

平均と総和には以下の関係が有る。

$$n\bar{x} = \sum_{i=1}^n x_i \quad (17)$$

8.5 分散・標準偏差・共分散・相関係数

確認のため各統計量の定義を示す。

$$V_x = \frac{1}{n} \sum_{i=1}^n (x_i - \bar{x})^2 \quad (18)$$

$$\sigma_x = \sqrt{V_x} \quad (19)$$

$$Cov_{xy} = \frac{1}{n} \sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y}) \quad (20)$$

$$\rho_{xy} = \frac{Cov_{xy}}{\sigma_x \sigma_y} \quad (21)$$

8.6 一次変換

定義

任意の定数 α と $\beta \neq 0$ を用い、変数 x_i を、

$$y_i = \alpha + \beta x_i \quad (22)$$

とする変換を『一次変換』とよぶ。

問題 II-8-2

n 組のデータがあり、 i 番目のデータの組を (x_i, y_i) とする。 x_i と y_i の関係が

$$y_i = \alpha + \beta x_i$$

であるとき、以下の値を求めなさい。

1. y_i の平均 ($\bar{y}$) を $\bar{x}$ を用いてあらわしなさい。
2. y_i の分散 (V_y) を x_i の分散 (V_x) を用いてあらわしなさい。ただし、 $\beta \neq 0$ とする。

解例 II-8-2

$$\begin{aligned}
 \bar{y} &= \frac{1}{n} \sum_{i=1}^n y_i \\
 &= \frac{1}{n} \sum_{i=1}^n (\alpha + \beta x_i) \\
 &= \frac{1}{n} \sum_{i=1}^n \alpha + \frac{1}{n} \sum_{i=1}^n \beta x_i \\
 &= \frac{1}{n} \cancel{n} \alpha + \beta \frac{1}{n} \sum_{i=1}^n x_i \\
 &= \alpha + \beta \bar{x}
 \end{aligned}$$

$$\begin{aligned}
 V_y &= \frac{1}{n} \sum_{i=1}^n (y_i - \bar{y})^2 \\
 &= \frac{1}{n} \sum_{i=1}^n ((\alpha + \beta x_i) - (\alpha + \beta \bar{x}))^2 \\
 &= \frac{1}{n} \sum_{i=1}^n (\cancel{\alpha} + \beta x_i - \cancel{\alpha} - \beta \bar{x})^2 \\
 &= \frac{1}{n} \sum_{i=1}^n (\beta(x_i - \bar{x}))^2 \\
 &= \beta^2 \frac{1}{n} \sum_{i=1}^n (x_i - \bar{x})^2 = \beta^2 V_x
 \end{aligned}$$

定理 II-8-1 (一次変換した変数の平均と分散)

x_i と y_i の関係が

$$y_i = \alpha + \beta x_i, \quad \beta \neq 0$$

であるとき、 x_i と y_i の間には以下の関係がある。

$$\bar{y} = \alpha + \beta \bar{x} \quad (23)$$

$$V_y = \beta^2 V_x \quad (24)$$

問題 II-8-3

$$\alpha = \frac{-\bar{x}}{\sigma_x}$$
$$\beta = \frac{1}{\sigma_x}$$

とする。 x_i と z_i の関係が

$$z_i = \alpha + \beta x_i$$

であらわされるとき、 z_i の平均と分散を求めなさい。ただし $\sigma_x \neq 0$ とする。

解例 II-8-3

$$\bar{z} = \alpha + \beta \bar{x} = \frac{-\bar{x}}{\sigma_x} + \frac{1}{\sigma_x} \bar{x} = \frac{-\bar{x} + \bar{x}}{\sigma_x} = 0$$

$$V_z = \beta^2 V_x = \left(\frac{1}{\sigma_x} \right)^2 V_x = \frac{1}{\sigma_x^2} V_x = \frac{1}{V_x} V_x = 1$$

問題 II-8-4

n 組のデータがあり、 i 番目のデータの組を (x_i, y_i) とする。 x_i と y_i の関係が

$$y_i = \alpha + \beta x_i$$

であるとき、 x_i と y_i の共分散を求めなさい。ただし、 $\beta \neq 0$ とする。

解例 II-8-4

$$\begin{aligned} Cov_{xy} &= \frac{1}{n} \sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y}) \\ &= \frac{1}{n} \sum_{i=1}^n (x_i - \bar{x})((\alpha + \beta x_i) - (\alpha + \beta \bar{x})) \\ &= \frac{1}{n} \sum_{i=1}^n (x_i - \bar{x})(\cancel{\alpha} + \beta x_i - \cancel{\alpha} - \beta \bar{x}) \\ &= \beta \frac{1}{n} \sum_{i=1}^n (x_i - \bar{x})^2 \\ &= \beta V_x \end{aligned}$$

問題 II-8-5

n 組のデータがあり、 i 番目のデータの組を (x_i, y_i) とする。 x_i と y_i の関係が

$$y_i = \alpha + \beta x_i$$

であるとき、 x_i と y_i の相関係数を求めなさい。ただし、 $V_x \neq 0$, $\beta \neq 0$ とする。

解例 II-8-5

$$\rho_{xy} = \frac{Cov_{xy}}{\sigma_x \sigma_y} = \frac{Cov_{xy}}{\sqrt{V_x} \sqrt{V_y}} = \frac{Cov_{xy}}{\sqrt{V_x V_y}} = \frac{\beta V_x}{\sqrt{V_x \beta^2 V_x}} = \frac{\beta V_x}{\sqrt{\beta^2} \sqrt{V_x^2}}$$

$V_x \geq 0$ なので $\sqrt{V_x^2} = V_x$

$$= \frac{\beta V_x}{\sqrt{\beta^2} V_x} = \frac{\beta \cancel{V_x}}{\sqrt{\beta^2} \cancel{V_x}} = \frac{\beta}{|\beta|}$$

したがって

$$\rho_{xy} = \begin{cases} +1 & ; \beta > 0 \\ -1 & ; \beta < 0 \end{cases}$$

定理 II-8-2 (一次変換した変数の共分散と相関係数)

x_i と y_i の関係が

$$y_i = \alpha + \beta x_i, \quad \beta \neq 0$$

であって、 $\sigma_x \neq 0$ とき、 x_i と y_i の間には以下の関係がある。

$$Cov_{xy} = \beta V_x \tag{25}$$

$$\rho_{xy} = \begin{cases} +1 & ; \beta > 0 \\ -1 & ; \beta < 0 \end{cases} \tag{26}$$

8.7 まとめ

- 正の傾きがある直線上に並んだデータの相関係数は $+1$ である。
- 負の傾きがある直線上に並んだデータの相関係数は -1 である。
- 傾きの大小は相関係数に影響しない。
- 標準化すると $\bar{z} = 0, \sigma_z = 1$ になる。
- $y_i = \alpha + \beta x_i$; $\beta \neq 0$ のとき、

$$\bar{y} = \alpha + \beta \bar{x}$$

$$V_y = \beta^2 V_x$$

$$Cov_{xy} = \beta V_x$$