教養科目／B 自然の摂理の探求

# 2026 年度 統計学（集中１期）

## 第５回

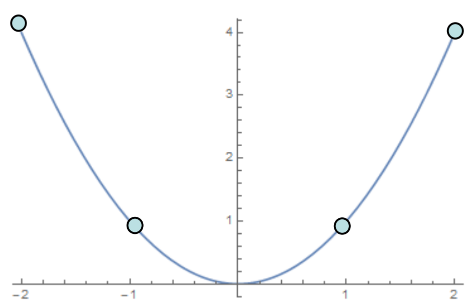
# 前回の内容

- **2 次元データ**
  - 散布図　・共分散　・相関係数
- 相関係数は直線的な関係の強さを表している.

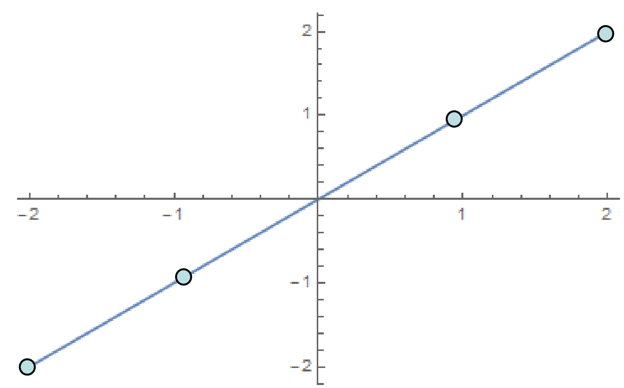
相関係数の値が $0$ に近い

$x$ と $y$ の直線的な関係がほとんどない

$x$ と $y$ の関係がないとは言っていないので注意



相関係数：0



相関係数：1

# 1．8. クロス集計表

**例 1.12**

80 人のクラスで試験を行い, 試験は 2 問あり，次の結果が得られた.

**表 1.13　例 1.12 の試験結果**

| | 問題 1 | 問題 2 |
|---|---|---|
| 1 | 正 | 正 |
| 2 | 誤 | 誤 |
| ⋮ | ⋮ | ⋮ |
| 80 | 正 | 誤 |

- 問題 1, 問題 2 が何人の学生に正しく解答されているか？
- 問題 1 の出来具合と問題 2 の出来具合の関係は？

⇩

**クロス集計表が便利**

# 1．8　クロス集計表

問題 1 も問題 2 も正しく解答した学生の人数

問題 1 は正しく解答し, 問題 2 は間違えた学生の人数

**表 1.14　例 1.12 のクロス集計表**

| 問題 1 ＼問題 2 | 正 | 誤 | 計 |
|---|---|---|---|
| 正 | 32 | 12 | 44 |
| 誤 | 8 | 28 | 36 |
| 計 | 40 | 40 | 80 |

問題 1 は間違え, 問題 2 は正しく解答した学生の人数

問題 1 も 2 も間違えた学生の人数

- このような表を**クロス集計表**あるいは 2 × 2 **の分割表**という.

# 1．8．クロス集計表

表 1.14　例 1.12 のクロス集計表

| 問題 1 ＼問題 2 | 正 | 誤 | 計 |
|---|---|---|---|
| 正 | 32 | 12 | 44 |
| 誤 | 8 | 28 | 36 |
| 計 | 40 | 40 | 80 |

問題 1 を正しく解答した学生の人数

問題 1 を間違えた学生の人数

問題 2 を正しく解答した学生の人数

問題 2 を間違えた学生の人数

・問題 1 を正しく解けることと問題 2 を正しく解けることには関係がありそうである．(主対角の数字に注目)

・それでは関係の強さを表す代表値はどう考えればよいのか?

# 1．9．独立性

一般的に話をするために，次のようなクロス集計表を考える．より詳しくは **4 章 (p.123∼126)** の検定で述べる．

表 1.15　一般のクロス集計表

| $A \diagdown B$ | $B_1$ | $B_2$ | 計 |
|---|---|---|---|
| $A_1$ | $n_{11}$ | $n_{12}$ | $n_{1\cdot}$ |
| $A_2$ | $n_{21}$ | $n_{22}$ | $n_{2\cdot}$ |
| 計 | $n_{\cdot 1}$ | $n_{\cdot 2}$ | $n$ |

ただし，

$$n_{1\cdot} = n_{11} + n_{12}, \quad n_{2\cdot} = n_{21} + n_{22},$$

$$n_{\cdot 1} = n_{11} + n_{21}, \quad n_{\cdot 2} = n_{12} + n_{22}.$$

・**Note :** $n_{1\cdot} + n_{2\cdot} = n_{\cdot 1} + n_{\cdot 2} = n$

# 1．9．独立性

・A と B が独立である，つまり，A と B に関係がないときに期待される度数を**独立期待度数**という．

・**独立期待度数**

$$e_{11} = \frac{n_{1\cdot}\,n_{\cdot 1}}{n}, \quad e_{12} = \frac{n_{1\cdot}\,n_{\cdot 2}}{n}, \quad e_{21} = \frac{n_{2\cdot}\,n_{\cdot 1}}{n}, \quad e_{22} = \frac{n_{2\cdot}\,n_{\cdot 2}}{n}$$

**表 1.16**　一般の独立期待度数表

| $A \diagdown B$ | $B_1$ | $B_2$ | 計 |
|---|---|---|---|
| $A_1$ | $e_{11}$ | $e_{12}$ | $n_{1\cdot}$ |
| $A_2$ | $e_{21}$ | $e_{22}$ | $n_{2\cdot}$ |
| 計 | $n_{\cdot 1}$ | $n_{\cdot 2}$ | $n$ |

# 1．9．独立性

表 1.15　一般のクロス集計表

| $A \diagdown B$ | $B_1$ | $B_2$ | 計 |
|---|---|---|---|
| $A_1$ | $n_{11}$ | $n_{12}$ | $n_{1\cdot}$ |
| $A_2$ | $n_{21}$ | $n_{22}$ | $n_{2\cdot}$ |
| 計 | $n_{\cdot 1}$ | $n_{\cdot 2}$ | $n$ |

・実際のデータ

表 1.16　一般の独立期待度数表

| $A \diagdown B$ | $B_1$ | $B_2$ | 計 |
|---|---|---|---|
| $A_1$ | $e_{11}$ | $e_{12}$ | $n_{1\cdot}$ |
| $A_2$ | $e_{21}$ | $e_{22}$ | $n_{2\cdot}$ |
| 計 | $n_{\cdot 1}$ | $n_{\cdot 2}$ | $n$ |

・A と B が独立である

(A と B に関係がない)

・$\chi^2$ 値

$$\chi^2 = \frac{(n_{11} - e_{11})^2}{e_{11}} + \frac{(n_{12} - e_{12})^2}{e_{12}} + \frac{(n_{21} - e_{21})^2}{e_{21}} + \frac{(n_{22} - e_{22})^2}{e_{22}}$$

実際のデータがどの程度「独立である」ということから乖離（かいり）しているかを表している．

# 1．9．独立性

・Note

$$\chi^2 = n\frac{(n_{11}n_{22}-n_{12}n_{21})^2}{n_{1\cdot}n_{2\cdot}n_{\cdot 1}n_{\cdot 2}}$$

・この値が**大きければ大きいほど独立でない**傾向がある．
つまり関係があるということがわかる．

・$\chi^2 \leq 3.841$ ⇒ 「独立であるといえないこともない」
あるいは
「関係があるとはいえない」

・$\chi^2 > 3.841$ ⇒ 「独立ではない」あるいは「関係がある」

・$\chi^2 > 6.635$ ⇒ 「関係が強い」

# １．９　独立性（例１．１２の解析結果）

表 **1.14**　例 **1.12** のクロス集計表

| 問題 1 ＼問題 2 | 正 | 誤 | 計 |
|---|---|---|---|
| 正 | 32 | 12 | 44 |
| 誤 | 8 | 28 | 36 |
| 計 | 40 | 40 | 80 |

・実際のデータ

表 **1.17**　例 **1.12** の独立期待度数表

| 問題 1 ＼問題 2 | 正 | 誤 | 計 |
|---|---|---|---|
| 正 | 22 | 22 | 44 |
| 誤 | 18 | 18 | 36 |
| 計 | 40 | 40 | 80 |

・問題 **1** と 問題 **2** が独立である
(問題 **1** と問題 **2** に関係がない)

・$\chi^2$ 値

$$\chi^2 = 80 \times \frac{(32 \times 28 - 12 \times 8)^2}{44 \times 36 \times 40 \times 40} = \frac{2000}{99} \fallingdotseq 20.2 > 6.635$$

・問題 **1** と問題 **2** は関係が強いと判断される.

# １．９．独立性（クラメールの連関係数）

質的データどうしの関係のことを連関ともいう．

・クラメールの連関係数

$$V = \sqrt{\frac{\chi^2}{n}}$$

・**Note :** $0 \leq V \leq 1$ が成り立つ．

・**A** と **B** の関係が強ければ強いほど **1** に近い値をとる．

・**A** と **B** が独立 ⇒ $V$ の値は $0$ に近い値を取る

・**A** と **B** の関係が強い ⇒ $V$ の値は $1$ に近い値を取る

# １．９．独立性 （例　１．12のクラメールの連関係数）

クラメールの連関係数は

$$V = \sqrt{\frac{\frac{2000}{99}}{80}} \fallingdotseq 0.50.$$

いくつ以上だと関係があるというような明確な基準はないが，$\mathbf{0.50}$ という値はそれなりに大きく，問題 $\mathbf{1}$ と問題 $\mathbf{2}$ は関係が強いと言ってよい．

# 1. 10. 回帰直線

**例 1.10** 最高気温とアイスクリームの売り上げ額

| 最高気温 $(x)$ | 26.1 | 20.0 | 23.2 | 25.1 | 25.3 | 27.2 | 28.1 |
|---|---|---|---|---|---|---|---|
| 売り上げ額 $(y)$ | 120 | 55 | 70 | 75 | 70 | 95 | 110 |

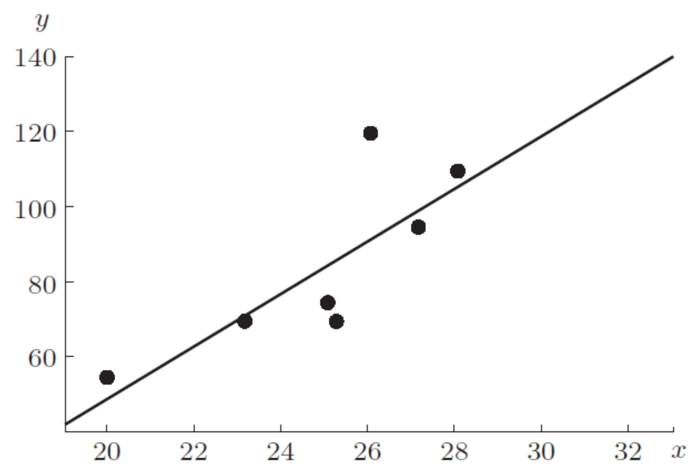



図 1.21　例 1.10 の散布図

およそ図中にあるような直線の関係があるように思える．直線の関係は暑ければ暑いほどアイスクリームの売り上げ額が大きくなることを意味している．

# 1. 10. 回帰直線

・図中で求めようとしている直線の方程式を $y = a + bx$ とする.

・図 1.22 の各点と直線 $y = a + bx$ との縦の線分の長さを考える.

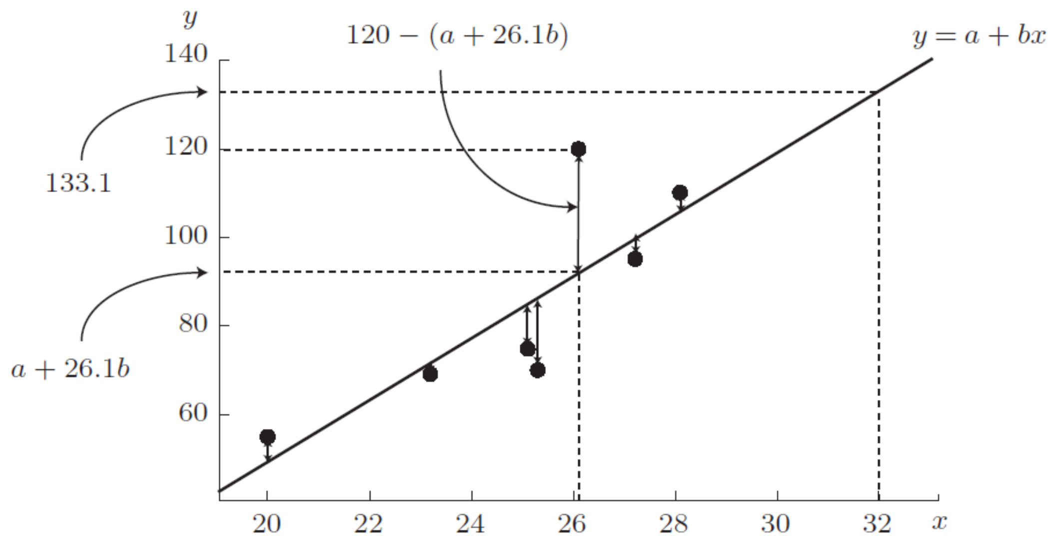



図 **1.22**　例 **1.10** の散布図と回帰直線

・これらの線分の長さの２乗の総和

$$\{120 - (a + 26.1b)\}^2 + \{55 - (a + 20.0b)\}^2 + \cdots + \{110 - (a + 28.1b)\}^2$$

が最小になるような直線 $y = a + bx$ を $y$ の $x$ への回帰直線という.

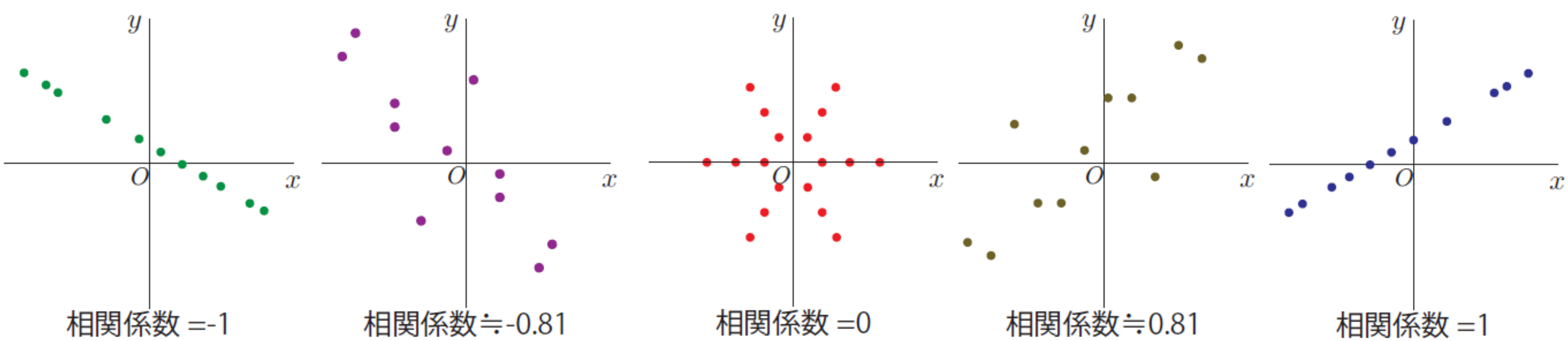
# 公式　1．4

(1) $y$ の $x$ への回帰直線 $y = \hat{a} + \hat{b}x$ の係数 $\hat{a}, \hat{b}$ は

$$\boxed{\hat{a} = (y\ \text{の標本平均}) - \hat{b} \times (x\ \text{の標本平均})}, \quad \boxed{\hat{b} = \frac{x\ \text{と}\ y\ \text{の共分散}}{x\ \text{の標本分散}}}$$

として得られる．（$\hat{a}, \hat{b}$ はそれぞれ“エーハット”，“ビーハット”と読む）．

(2) 相関係数が $1$ または $-1$ に近ければ近いほど回帰直線のデータへのあてはまりが良くなる．



# 公式　１．４

・公式 1.4 (1) の $\hat{a}$, $\hat{b}$ を数式で書くと

$$\hat{a} = \bar{y} - \hat{b}\,\bar{x}, \quad \hat{b} = \frac{c_{xy}}{s_x^2}$$

となる. $\hat{b}$ を回帰係数という.

・このようにして直線を２次元データにあてはめる方法を最小２乗法という.

# 注意

・一般に "$y$ の $x$ への回帰直線" と "$x$ の $y$ への回帰直線" は異なることに注意する.

・２種類のデータがそれぞれ原因と結果と考えられるような場合

「最高気温」が原因であり,「アイスクリームの売り上げ額」が結果

↓

$x$ 軸（横軸）に原因を表すデータ, $y$ 軸（縦軸）に結果を表すデータをとり, $y$ の $x$ への回帰直線を考えるのが一般的である.

・どちらが原因でも結果でもないと考えられる場合

どちらに $x$ 軸, $y$ 軸をとっても差し支えない.

# 例1．10．における回帰直線

**例 1.10** 最高気温とアイスクリームの売り上げ額

| 最高気温 $(x)$ | 26.1 | 20.0 | 23.2 | 25.1 | 25.3 | 27.2 | 28.1 |
|---|---|---|---|---|---|---|---|
| 売り上げ額 $(y)$ | 120 | 55 | 70 | 75 | 70 | 95 | 110 |

- $\bar{x}$，$\bar{y}$，$s_x^2$，$c_{xy}$ を計算する．

$$\bar{x} = \frac{1}{7}(26.1 + 20.0 + 23.2 + 25.1 + 25.3 + 27.2 + 28.1) = 25.0$$

$$\bar{y} = \frac{1}{7}(120 + 55 + 70 + 75 + 70 + 95 + 110) = 85$$

$$\begin{aligned} s_x^2 &= \tfrac{1}{7}\{1.1^2 + (-5.0)^2 + (-1.8)^2 + 0.1^2 + 0.3^2 + 2.2^2 + 3.1^2\} \\ &= \frac{44}{7} \end{aligned}$$

$$\begin{aligned} c_{xy} &= \tfrac{1}{7}\{1.1 \times 35.0 + (-5.0) \times (-30.0) + (-1.8) \times (-15.0) \\ &\quad + 0.1 \times (-10.0) + 0.3 \times (-15.0) + 2.2 \times 10.0 + 3.1 \times 25.0\} \\ &= \frac{309.5}{7} \end{aligned}$$

# 例1．10．における回帰直線

・$s_x^2$，$c_{xy}$ の値から回帰係数を計算する．

$$\hat{b} = \frac{c_{xy}}{s_x^2} = \frac{\frac{309.5}{7}}{\frac{44}{7}} = 7.03 \cdots \fallingdotseq 7.0$$

・$\bar{x}$, $\bar{y}$, $\hat{b}$ の値を使い $\hat{a}$ を計算する．

$$\hat{a} = \bar{y} - \hat{b}\bar{x} \fallingdotseq 85.0 - \frac{309.5}{44} \times 25.0 = -90.85 \cdots \fallingdotseq -90.9$$

・$y$ の $x$ への回帰直線

$$\boxed{y = -90.9 + 7.0x}$$

# 例　1．10における回帰直線

・$y$ の $x$ への回帰直線

$$y = -90.9 + 7.0x$$

たとえば最高気温が $x = 32$ ℃のとき，売り上げ額はおよそ

$$-90.9 + 7.0 \times 32 = 133.1 \quad \text{(千円)}$$

となると予想される．

# まとめ

・クロス集計表

表 1.14　例 1.12 のクロス集計表

| 問題 1 ＼問題 2 | 正 | 誤 | 計 |
|---|---|---|---|
| 正 | 32 | 12 | 44 |
| 誤 | 8 | 28 | 36 |
| 計 | 40 | 40 | 80 |

・独立性, 関係, $\chi^2$ 値, クラメールの連関係数

表 1.15　一般のクロス集計表

| $A \diagdown B$ | $B_1$ | $B_2$ | 計 |
|---|---|---|---|
| $A_1$ | $n_{11}$ | $n_{12}$ | $n_{1\cdot}$ |
| $A_2$ | $n_{21}$ | $n_{22}$ | $n_{2\cdot}$ |
| 計 | $n_{\cdot 1}$ | $n_{\cdot 2}$ | $n$ |

$$\chi^2 = n\frac{(n_{11}n_{22} - n_{12}n_{21})^2}{n_{1\cdot}n_{2\cdot}n_{\cdot 1}n_{\cdot 2}}$$

・$\chi^2 \leq 3.841$ ⇒ 「独立であるといえないこともない」
あるいは
「関係があるとはいえない」

・$\chi^2 > 3.841$ ⇒ 「独立ではない」あるいは「関係がある」

・$\chi^2 > 6.635$ ⇒ 「関係が強い」

# 注意

・関係の方向は２つある.

・最も関係がある場合 (計を固定した場合)

| 問題 1 ＼問題 2 | 正 | 誤 | 計 |
|---|---|---|---|
| 正 | 40 | 4 | 44 |
| 誤 | 0 | 36 | 36 |
| 計 | 40 | 40 | 80 |

・もう一方の関係

| 問題 1 ＼問題 2 | 正 | 誤 | 計 |
|---|---|---|---|
| 正 | 4 | 40 | 44 |
| 誤 | 36 | 0 | 36 |
| 計 | 40 | 40 | 80 |

・最も関係がない場合 (計を固定した場合)

| 問題 1 ＼問題 2 | 正 | 誤 | 計 |
|---|---|---|---|
| 正 | 22 | 22 | 44 |
| 誤 | 18 | 18 | 36 |
| 計 | 40 | 40 | 80 |

# まとめ

・回帰直線

(1) $y$ の $x$ への回帰直線 $y = \hat{a} + \hat{b}x$ の係数 $\hat{a}, \hat{b}$ は

$$\boxed{\hat{a} = (y \text{ の標本平均}) - \hat{b} \times (x \text{ の標本平均})}, \quad \boxed{\hat{b} = \frac{x \text{ と } y \text{ の共分散}}{x \text{ の標本分散}}}$$

として得られる．（$\hat{a}, \hat{b}$ はそれぞれ "エーハット"，"ビーハット" と読む）．

(2) 相関係数が $1$ または $-1$ に近ければ近いほど回帰直線のデータへのあてはまりが良くなる．