

計測工学 第6回

前回 : ▪ 母平均の区間推定 (続き)
 ▪ 相関と回帰



今回 : ▪ 相関係数
 ▪ 回帰分析

6.1 相関係数

・例: 30 人の学生における高校時の評価平均 x と大学 1 年時の評価平均 y の相関を考える.

表 6.1 x - y 一覧表

x	y	x	y	x	y
3.0	2.4	2.9	1.9	3.1	2.8
2.4	2.6	2.7	2.2	3.3	3.2
3.7	3.0	3.7	3.1	2.7	1.8
3.6	3.9	2.7	2.6	3.5	2.7
3.8	3.6	3.3	2.8	2.9	2.1
2.9	3.0	2.8	2.7	2.7	1.7
3.5	3.1	3.1	2.4	2.9	1.7
3.0	2.8	2.8	3.0	3.2	2.3
2.3	2.2	3.0	3.3	3.4	2.6
3.0	2.9	2.2	1.8	2.5	2.7

① 散布図とプロット

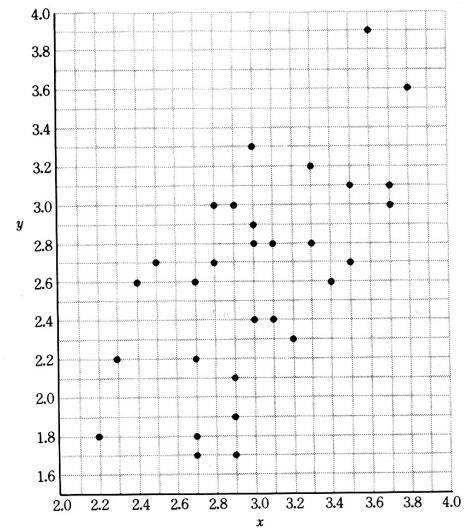


図 6.1 x - y 散布図

② 変数 (x_i, y_i) について,

$$x' = \frac{x_i - \bar{x}}{S_x}, \quad y' = \frac{y_i - \bar{y}}{S_y} \quad \text{による標準化を行い 図プロットする.}$$

$$S_x^2 = S_{xx} = \frac{1}{n} \sum (x_i - \bar{x})^2: x \text{ の標準分散}$$

$$S_x = \sqrt{S_x^2}$$

$$S_y^2 = S_{yy} = \frac{1}{n} \sum (y_i - \bar{y})^2$$

$$S_{xy} = \frac{1}{n} \sum (x_i - \bar{x})(y_i - \bar{y}): x \text{ と } y \text{ の共分散}$$

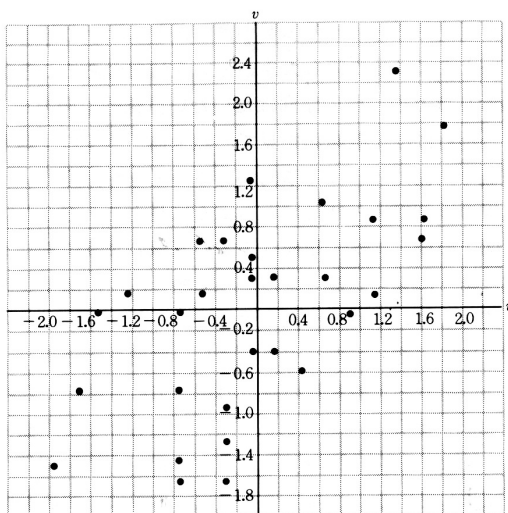
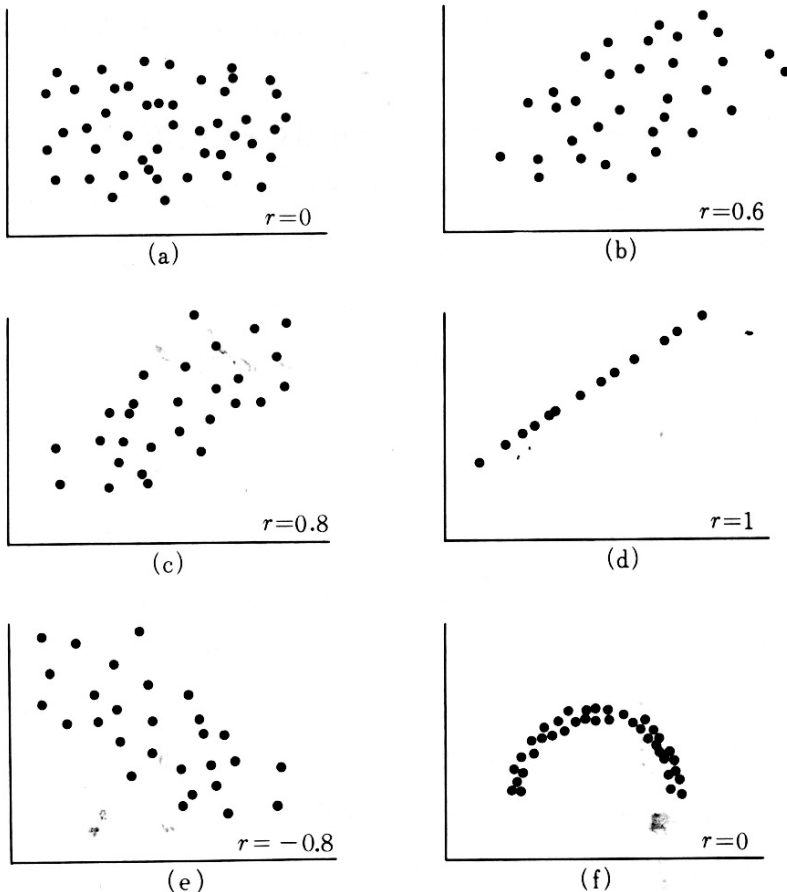


図 6.2 標準化 x - y 散布図

→ ③ x' と y' の積和の平均 $\frac{1}{n} \sum x'_i y'_i$ を「相関係数 r 」と定義する.

$$r = \frac{1}{n} \sum \left\{ \frac{x_i - \bar{x}}{S_x} \cdot \frac{y_i - \bar{y}}{S_y} \right\} = \frac{\frac{1}{n} \sum (x_i - \bar{x})(y_i - \bar{y})}{S_x \cdot S_y}$$

$$= \frac{S_{xy}}{\sqrt{S_x^2} \sqrt{S_y^2}}$$

図 6.3 r とデータの傾向の関係 $r = 0$: 相関無し $r > 0$: 正の相関 $r < 0$: 負の相関 $0 \leq |r| \leq 0.2$: ほとんど相関無し $0.2 < |r| \leq 0.4$: 弱い相関 $0.4 < |r| \leq 0.7$: ある程度の相関 $0.7 < |r| \leq 1.0$: 強い相関

左図(f)より:

本講義の r は 2つの変数間の
線形関係性のみを検討・例題: 学生の入学試験順位 x と 1 年後の順位 y を調べ, そのうち 5 名については $(x, y) \rightarrow$ $(10, 40), (25, 35), (35, 10), (60, 50), (70, 70)$ であった. x と y の相関係数を求めよ.

$$S_x^2 = \frac{1}{n} \sum (x_i - \bar{x})^2 \quad \bar{x} = 40, \bar{y} = 41$$

$$= \frac{1}{5} \{(-30)^2 + (-15)^2 + (-5)^2 + 20^2 + 30^2\} = \underline{490}$$

$$S_y^2 = \underline{384}$$

$$S_{xy} = \frac{1}{n} \sum \{(-30 \times -1) + (-15 \times -6) + (-5 \times -31) + 20 \times 9 + 30 \times 29\}$$

$$= \underline{265}$$

$$\therefore r = \frac{265}{\sqrt{490} \sqrt{384}}$$

$$= 0.61 \dots$$

$$= \underline{0.6}$$

No.	1	2	3	4	5	
x_i	10	25	35	60	70	$\bar{x} = 40$
$x_i - \bar{x}$	-30	-15	-5	+20	+30	
y_i	40	35	10	50	70	$\bar{y} = 41$
$y_i - \bar{y}$	-1	-6	-31	+9	+29	

6.2 回帰分析

●最小2乗法による線形回帰分析

- ・回帰式として $y = ax + b$ を想定する: データから如何に a と b を求めるか?

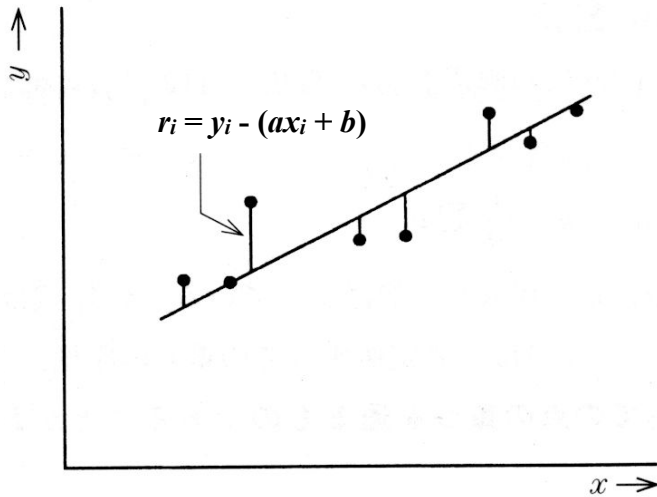


図 6.4 最小2乗法による線形回帰分析

- ・ n 個の測定値の対 $(x_1, y_1), (x_2, y_2) \dots (x_n, y_n)$ を対象とする

- ・ x に誤差はなく, 誤差は全て y に含まれると仮定



- ・ある傾き a と切片 b を持つ直線 $f(x) = ax + b$ と,

i 番目の測定値 y_i との残差 r_i

$$\rightarrow V_i = y_i - f(x_i) = y_i - (ax_i + b)$$



- ・ n 個の測定値に対して, 残差 r_i の2乗和 $\sum_{i=1}^n r_i^2 = \sum_{i=1}^n \{y_i - (ax_i + b)\}^2$ が最小となるのが

最良の回帰直線

- ・問い: なぜ残差 r_i に対して2乗和を取るのか?

残差の評価に対する符号の影響を排除するため.

- ・ $\sum_{i=1}^n \{y_i - (ax_i + b)\}^2 = R$ とおいたとき... x_i, y_i : 定数 (データとして与えられる)
- a, b : 変数 (回帰式の定数として決定可能な値)



R が最小となるような a と b を決定する.

$$\begin{aligned}
 R &= \sum_{i=1}^n (y_i - ax_i - b)^2 = \sum_{i=1}^n (y_i^2 + a^2 x_i^2 + b^2 - 2by_i - 2ax_i y_i + 2abx_i) \\
 &= \sum_{i=1}^n y_i^2 + a^2 \sum_{i=1}^n x_i^2 + nb^2 - 2b \sum_{i=1}^n y_i - 2a \sum_{i=1}^n x_i y_i + 2ab \sum_{i=1}^n x_i \\
 &= A + a^2 B + nb^2 - 2bC - 2aD + 2abE
 \end{aligned}$$

・ R を変数 a, b で偏微分して 0 と置く... 最小となる条件！

$$\frac{\partial R}{\partial a} = 2aB - 2D + 2bE = 0 \quad \left. \vphantom{\frac{\partial R}{\partial a}} \right\} \rightarrow a = \frac{nD - CE}{nb - E^2}, \quad b = \frac{BC - DE}{nb - E^2}$$

$$\frac{\partial R}{\partial b} = 2nb - 2C + 2aE = 0$$

$$a = \frac{n \sum x_i y_i - \sum x_i \sum y_i}{n \sum x_i^2 - (\sum x_i)^2}, \quad b = \frac{\sum x_i^2 \sum y_i - \sum x_i y_i \sum x_i}{n \sum x_i^2 - (\sum x_i)^2}$$

・例題：P.2 の例題[学生の入学試験順位 x と 1 年後の順位 y , $\bar{x} = 40$, $\bar{y} = 41$]について最小

二乗法による線形回帰分析を行え.

$$\bar{x} = 40, \quad \bar{y} = 41, \quad \sum x_i = n \cdot \bar{x} = 200, \quad \sum y_i = n \cdot \bar{y} = 205$$

$$\sum x_i^2 = (10^2 + 25^2 + 35^2 + 60^2 + 70^2) = 10450$$

$$\sum y_i^2 = (40^2 + 35^2 + 10^2 + 50^2 + 70^2) = 10325$$

$$\sum x_i y_i = (10 \times 40 + 25 \times 35 + 35 \times 10 + 60 \times 50 + 70 \times 70) = 9525$$

$$\begin{aligned}
 \therefore a &= \frac{5 \cdot 9525 - 200 \cdot 205}{5 \cdot 10450 - (200)^2} \\
 &= 0.5408 \dots \\
 &= 0.541
 \end{aligned}$$

$$\begin{aligned}
 b &= \frac{10450 \cdot 205 - 9525 \cdot 200}{5 \cdot 10450 - (200)^2} \\
 &= 19.36 \dots \\
 &= 19.4
 \end{aligned}$$

$$y = 0.541x + 19.4$$

6.4 第5回小テスト解答

- Q.1 正規分布する母集団 (σ^2 未知) から以下の標本を取出したとき, 母平均 μ の信頼区間を 68.3% の信頼係数で求めよ. 計算過程も明示すること. 標本: 12.9, 13.3, 13.1, 13.0, 12.7

A.1 標本平均 $\bar{x} = \frac{1}{n} \sum_{i=1}^n x_i = 13.0$, 標本数 $n = 5$

不偏分散 $u^2 = \frac{n}{n-1} s^2 = \frac{1}{n-1} \sum_{i=1}^n (x_i - \bar{x})^2 = 0.05$

t 分布表より $t_{n-1}(z\%) = 1.141$

$t_{n-1}(Z\%) \sqrt{\frac{u^2}{n}} = 0.1141 = 0.1$

$\therefore 12.9 \leq \mu \leq 13.1$

自由度 ($N-1$)	標本数 (N)	$t_{N-1}(z\%), z(\%)$: 信頼度			
		68.3%	90%	95%	99%
		$P(\text{有意水準}), \text{両側検定}$			
		31.7%	10%	5%	1%
1	2	1.837	6.314	12.706	63.657
2	3	1.321	2.920	4.303	9.925
3	4	1.197	2.353	3.182	5.841
4	5	1.141	2.132	2.776	4.604
5	6	1.110	2.015	2.571	4.032
6	7	1.090	1.943	2.447	3.707
7	8	1.077	1.895	2.365	3.500
8	9	1.066	1.860	2.306	3.355
9	10	1.059	1.833	2.262	3.250