

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

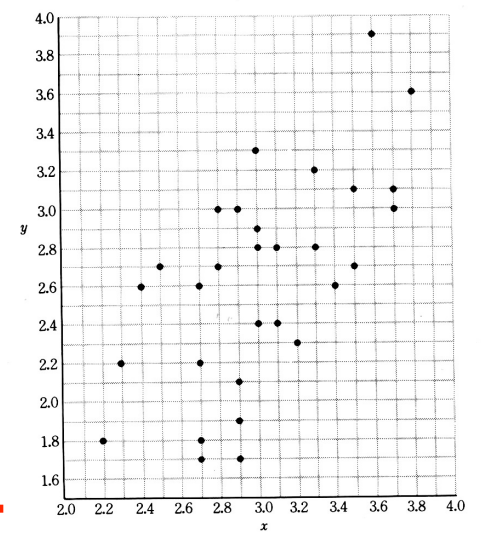
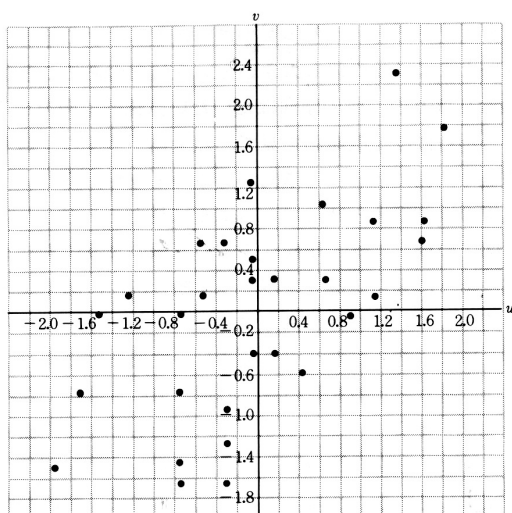


図 6.1 x - y 散布図

②



③

図 6.2 標準化 x - y 散布図

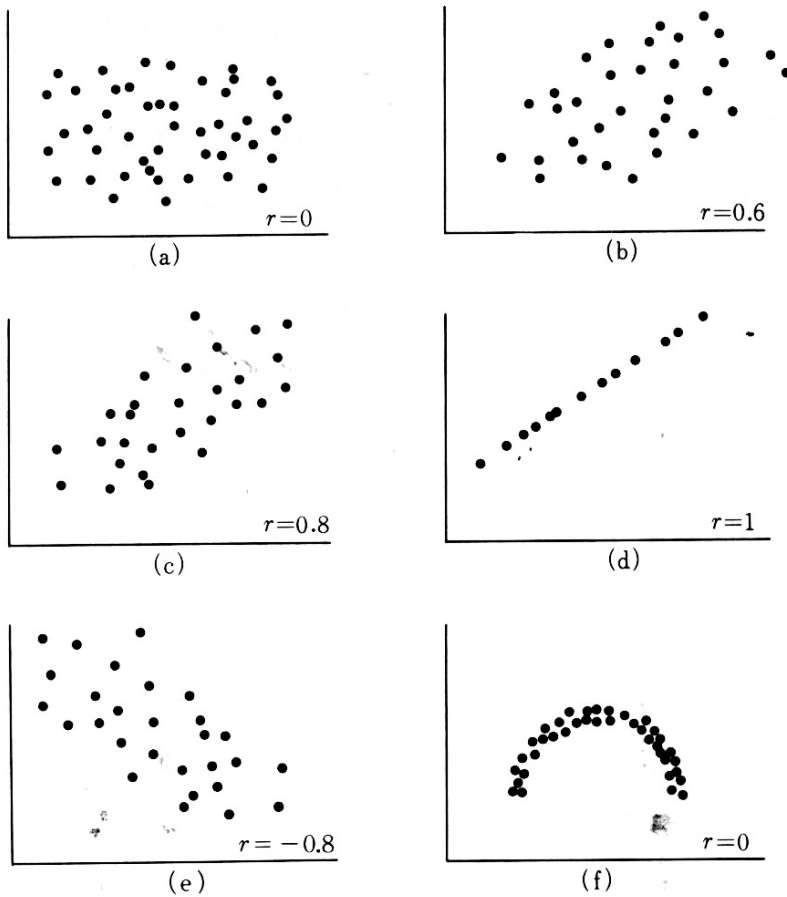


図 6.3 r とデータの傾向の関係

$r = 0$: 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:

・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.

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?

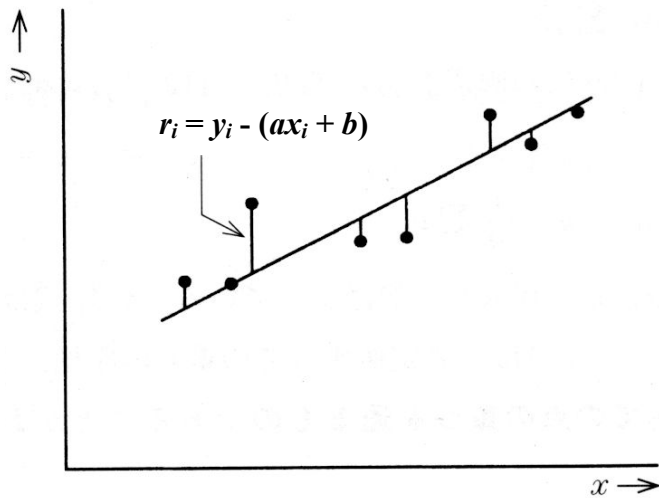


図 6.4 最小 2 乗法による線形回帰分析

- ・ n 個の測定値の対 $(x_1, y_1), (x_2, y_2) \cdots (x_n, y_n)$ 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

→



- ・ 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?

$$\sum_{i=1}^n \{y_i - (ax_i + b)\}^2 = R \quad \text{とおいたとき} \cdots x_i, y_i$$

a, b



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



・ R を変数 a, b で偏微分して 0 と置く・・・最小となる条件！

・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.