10 ❦ 回帰分析　　2０２４年度後期　基礎数学 B2（火曜４限）

入学年度	学部	学 科		組		番 号			検	フリガナ	
		B		1						氏 名	

2 つの変量 X と Y の間に相関関係があるか，またその強さがどれくらいかを測る指数は**相関係数**と呼ばれる．少し観点を変え，相関関係があると思われる 2 つの変量のうち、一方の変量 X から他方の変量 Y の値がどの程度推測できるかを考えるのが**回帰分析**である．

まず，2 つの変量 X, Y をそれぞれ確率変数と捉え，X を**説明変数**，Y を**目的変数**と呼ぶ．回帰分析とは説明変数と目的変数の関係を回帰式で表し，目的変数が説明変数によってどの程度説明できるかを定量的に分析することである．回帰分析の最も基本的なモデルは回帰式が $Y = a + bX$ という形式の一次式で表せる線形回帰である。

例 1. ある大学の定期試験の結果から 10 人の学生の成績を無作為に抽出してみたところ，数学と物理の点は下の表のようであった．

学生	1	2	3	4	5	6	7	8	9	10
数学 : X(点)	79	63	78	86	65	58	93	83	75	70
物理 : Y(点)	85	70	82	83	75	70	90	77	76	78

まず，仮平均を数学，物理ともに 75 点として，$U = X - 75$，$V = Y - 75$ とおく．すると，

$$
\begin{aligned}
&E(X) = E(U) - 75, & &E(Y) = E(V) - 75,\\
&V(X) = V(U) = E(U^2) - E(U)^2, & &V(Y) = V(V) = E(V^2) - E(V)^2
\end{aligned} \tag{1}
$$

が成り立ち，X と Y の共分散 $\mathrm{Cov}(X, Y)$ についても同様に

$$\mathrm{Cov}(X, Y) = \mathrm{Cov}(U, V) = E(UV) - E(U)E(V) \tag{2}$$

が成り立つ．これらを計算するために，次のような表をまず作る．

	X	Y	U	V	U^2	V^2	UV
1	79	85	4	10	16	100	40
2	63	70	−12	−5	144	25	60
3	78	82	3	7	9	49	63
4	86	83	11	8	121	64	88
5	65	75	−10	0	100	0	0
6	58	70	−17	−5	289	25	85
7	93	90	18	15	324	225	270
8	83	77	8	2	64	4	16
9	75	76	0	1	0	1	0
10	70	78	−5	3	25	9	-15
		和	0	36	1092	502	565
		平均	0	3.6	109.2	50.2	56.5

この結果から，(1)，(2) を用いて

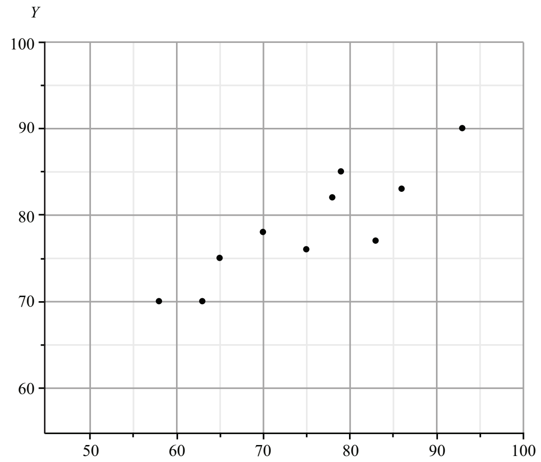
$$V(X) = 109.2, \quad V(Y) = 37.24, \quad \mathrm{Cov}(X, Y) = 56.5$$

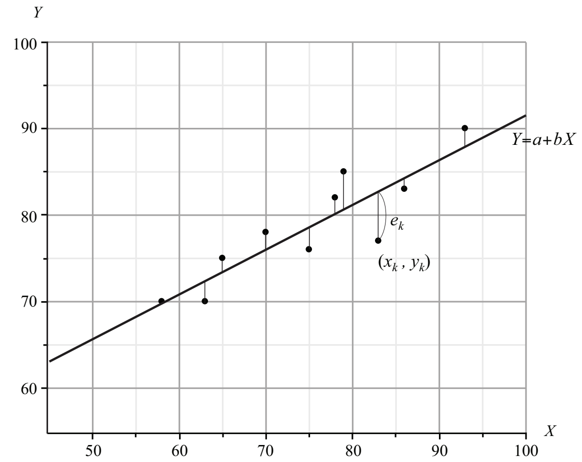
となる．これより，X と Y の間の相関係数 r は

$$r = \frac{\mathrm{Cov}(X, Y)}{\sigma(X)\sigma(Y)} = \frac{\mathrm{Cov}(X, Y)}{\sqrt{V(X)}\sqrt{V(Y)}} = \frac{56.5}{\sqrt{109.2}\sqrt{37.24}} \fallingdotseq 0.886$$

となり，比較的強い正の相関関係があることがわかる．

次に，X と Y の 10 個の標本の値 $(x_1, y_1), (x_2, y_2), \ldots (x_{10}, y_{10})$ をグラフ上の点として表すと，下の左の図のような**散布図**が得られる．点の全体は右上がりの向きに分布している．この場合に，できるだけ客観的に直線を当てはめたい．その時に用いられるのが**最小 2 乗法**である．





いま，X と Y の間に近似的に $Y = a + bX$ という直線関係が成り立っているとする．（$Y = aX + b$ ではないことに注意．）この関係は近似的な関係なので，$X = x_k$ のときに $Y = y_k$ と $a + bx_k$ との間に誤差

$$e_k = y_k - (a + bx_k)$$

が生じる．これらの誤差をできるだけ小さくするようにするにはいろいろな方法が考えられるが，e_k^2 の平均

$$Q = \frac{1}{10}(e_1^2 + e_2^2 + \cdots + e_{10}^2) = \frac{1}{10}\sum_{k=1}^{10}\bigl(y_k - (a + bx_k)\bigr)^2$$

を最小にするように a，b を定めるのが最小 2 乗法である．この方法によって得られる結果を先に述べると，

$$a = E(Y) - \frac{\mathrm{Cov}(X, Y)}{V(X)}E(X), \quad b = \frac{\mathrm{Cov}(X, Y)}{V(X)} \tag{3}$$

となることが知られている．これは，$Y = a + bX$ が点 $\bigl(E(X), E(Y)\bigr)$ を通り，傾きが上の b の値であるような直線であることを意味している．具体的な数値を計算してみると

$$a = 78.6 - \frac{56.5}{109.2} \times 75.0 \fallingdotseq 39.8, \quad b = \frac{56.5}{109.2} \fallingdotseq 0.52$$

となる．これにより，Y の値を X の値により推定する式 $Y = 39.8 + 0.52X$ が得られる．これより，数学の点数が例えば 80 点だった学生の物理の点数はだいたい 81.4 点だったであろうと推測できる．

$\boxed{1}$ ある自動車が一定の速度で走行中に急ブレーキをかけたとき，停止するまでの距離がどれくらいかを調べたところ，下のようなデータがえられた.

X：速度 (km/h)	42	50	56	64	73	76	80
Y：停止距離 (m)	5.2	7.5	5.9	8.5	8.5	7.8	8.4

X	Y	$U = X - 60$	$V = Y - 7$	U^2	V^2	UV
42	5.2	−18	−1.8	324	3.24	32.4
50	7.5	−10	0.5	100	0.25	−5
56	5.9	−4	−1.1	16	1.21	4.4
64	8.5	4	1.5	16	2.25	6
73	8.5	13	1.5	169	2.25	19.5
76	7.8	16	0.8	256	0.64	12.8
80	8.4	20	1.4	400	1.96	28
	和	21	2.8	1281	11.8	98.1
	平均	3	0.4	183	1.686	14.01

a) X，Y の分散 $V(X)$，$V(Y)$，X と Y の共分散 $\mathrm{Cov}(X,Y)$ をそれぞれ求めよ.

$V(X) = V(U) = E(U^2) - E(U)^2 = 174$

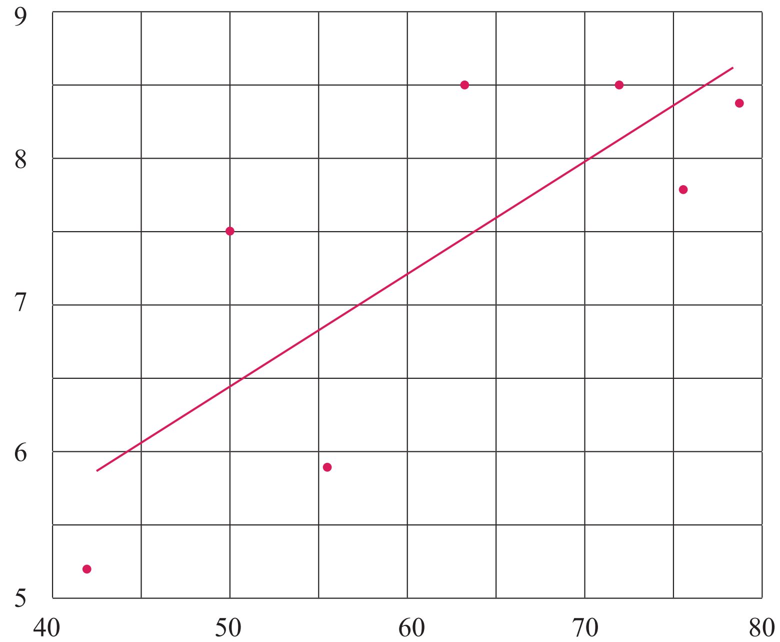
$V(Y) = V(V) = E(V^2) - E(V)^2 = 1.53$

$\mathrm{Cov}(X,Y) = \mathrm{Cov}(U,U) = E(UV) - E(U)E(V) = 12.81$

b) X と Y の間の相関係数を求めよ.

$r = \frac{\mathrm{Cov}(X,Y)}{\sigma(X)\sigma(Y)} = \frac{12.81}{\sqrt{174}\sqrt{1.53}} = 0.784$

c) X と Y の散布図（相関図）を描け.



d) 回帰直線 $Y = a + bX$ を求めて相関図の中に図示せよ.

回帰直線を $Y = a + bX$ としたとき，$b = \frac{\mathrm{Cov}(X,Y)}{V(X)}$, $a = E(Y) - bE(X)$ だから

$$b = 0.074, \quad a = 2.76.$$

したがって，回帰直線は

$$Y = 2.76 + 0.074X$$

e) この自動車が 5 m以内で停止できるようにするには，時速何 km 以下で走行しなければならないか.

$Y = 2.76 + 0.074X < 5$ を C について解くと，$X < 30.41$．したがって，30.41 km 以下で走行しなければならない.

担当：鍬田 政人 (kuwata.04f@g.chuo-u.ac.jp)