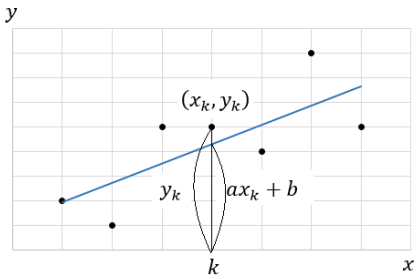
最小二乗法

1．回帰直線

n個のデータ$(x_1, y_1), (x_2, y_2), \cdots, (x_n, y_n)$の散布図に対して，そのデータの状況を直線 $y = ax + b$ で表したい．その一つの方法として**回帰直線**で表すという考え方がある．



その基本的な考え方は，

$$F = \sum_{k=1}^{n} \{y_k - (ax_k + b)\}^2 \tag{1}$$

とし，Fを最小にするa, bを求めることである．この方法を**最小二乗法**と呼んでいる．したがって，回帰直線は最小二乗法によって得られた直線の式である．

2. 回帰直線の正規方程式

具体的に(1)のFを最小にするa, bを求めてみよう．そのためには，Fのa方向の偏微分とb方向の偏微分がどちらも0となるa, bを求めればよい．すなわち，連立方程式

$$\frac{\partial F}{\partial a} = 0, \qquad \frac{\partial F}{\partial b} = 0 \tag{2}$$

を解けばよい．

$$\frac{\partial F}{\partial a} = \sum_{k=1}^{n} \{y_k - (ax_k + b)\}(-x_k) = a\sum_{k=1}^{n} x_k^2 + b\sum_{k=1}^{n} x_k - \sum_{k=1}^{n} x_k y_k$$

$$\frac{\partial F}{\partial b} = \sum_{k=1}^{n} \{y_k - (ax_k + b)\}(-1) = a\sum_{k=1}^{n} x_k + b\sum_{k=1}^{n} 1 - \sum_{k=1}^{n} y_k$$

であるので，よって連立方程式(2)は

$$\begin{pmatrix} \sum_{k=1}^{n} x_k^2 & \sum_{k=1}^{n} x_k \\ \sum_{k=1}^{n} x_k & \sum_{k=1}^{n} 1 \end{pmatrix} \begin{pmatrix} a \\ b \end{pmatrix} = \begin{pmatrix} \sum_{k=1}^{n} x_k y_k \\ \sum_{k=1}^{n} y_k \end{pmatrix} \tag{3}$$

となる．(3)を**正規方程式**と呼ぶ．

問題１．以下の表はある鉱石の密度x $(\mathrm{g/cm^3})$と鉄の含有量$y(\%)$を調べたものである．

x	3.0	3.1	3.2	3.3	3.4	3.5	3.6
y	29	34	35	38	35	39	40

回帰直線a,bを求めるための正規方程式を立てよ．

3. 正規方程式の解法

あるn個のデータ$(x_1,y_1),(x_2,y_2),\cdots,(x_n,y_n)$の散布図の回帰直線を求めるために，以下の正規方程式を得た．

$$\begin{pmatrix}5 & 3\\ 3 & 2\end{pmatrix}\begin{pmatrix}a\\ b\end{pmatrix}=\begin{pmatrix}10\\ 8\end{pmatrix} \tag{4}$$

以下，連立方程式(4)を解く．

$$A=\begin{pmatrix}5 & 3\\ 3 & 2\end{pmatrix},\qquad A^{-1}=\begin{pmatrix}2 & -3\\ -3 & 5\end{pmatrix}$$

より，

$$\begin{pmatrix}a\\ b\end{pmatrix}=\begin{pmatrix}2 & -3\\ -3 & 5\end{pmatrix}\begin{pmatrix}10\\ 8\end{pmatrix}=\begin{pmatrix}-4\\ 10\end{pmatrix}$$

である．したがって，回帰直線の式は

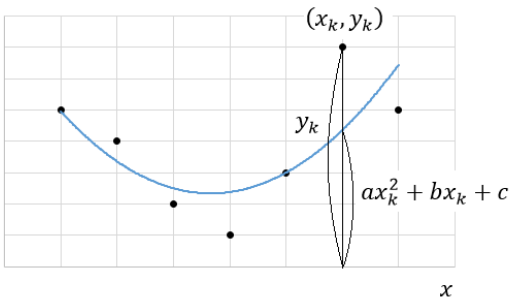
$$y=-4x+10$$

といえる．

問題 2．問題 1 の回帰直線$y=ax+b$の式を求め，それを用いて$x=3.25$のときの含有量yを推定せよ．ただし，a,bはそれぞれ小数点以下第 2 位まで求めよ．

4. ２次回帰曲線

n個のデータ$(x_1,y_1),(x_2,y_2),\cdots,(x_n,y_n)$の散布図に対して，そのデータの状況を最小二乗法を使って 2 次曲線 $y=ax^2+bx+c$ で表したものを**２次回帰曲線**という．



すなわち，最小二乗法の考え方は，

$$F=\sum_{k=1}^{n}\{y_k-(ax_k^2+bx_k+c)\}^2 \tag{5}$$

とし，Fを最小にするa,b,cを求めることである．そのためには，連立方程式

$$\frac{\partial F}{\partial a}=0,\qquad \frac{\partial F}{\partial b}=0,\qquad \frac{\partial F}{\partial c}=0 \tag{6}$$

を解けばよい．

$$\frac{\partial F}{\partial a}=\sum_{k=1}^{n}\{y_k-(ax_k^2+bx_k+c)\}(-x_k^2)=a\sum_{k=1}^{n}x_k^4+b\sum_{k=1}^{n}x_k^3+c\sum_{k=1}^{n}x_k^2-\sum_{k=1}^{n}x_k^2y_k$$
$$\frac{\partial F}{\partial b}=\sum_{k=1}^{n}\{y_k-(ax_k^2+bx_k+c)\}(-x_k)=a\sum_{k=1}^{n}x_k^3+b\sum_{k=1}^{n}x_k^2+c\sum_{k=1}^{n}x_k-\sum_{k=1}^{n}x_ky_k$$
$$\frac{\partial F}{\partial c}=\sum_{k=1}^{n}\{y_k-(ax_k^2+bx_k+c)\}(-1)=a\sum_{k=1}^{n}x_k^2+b\sum_{k=1}^{n}x_k+c\sum_{k=1}^{n}1-\sum_{k=1}^{n}y_k$$

であるので，よって連立方程式(6)は

$$\begin{pmatrix}\sum_{k=1}^{n}x_k^4 & \sum_{k=1}^{n}x_k^3 & \sum_{k=1}^{n}x_k^2\\ \sum_{k=1}^{n}x_k^3 & \sum_{k=1}^{n}x_k^2 & \sum_{k=1}^{n}x_k\\ \sum_{k=1}^{n}x_k^2 & \sum_{k=1}^{n}x_k & \sum_{k=1}^{n}1\end{pmatrix}\begin{pmatrix}a\\ b\\ c\end{pmatrix}=\begin{pmatrix}\sum_{k=1}^{n}x_k^2y_k\\ \sum_{k=1}^{n}x_ky_k\\ \sum_{k=1}^{n}y_k\end{pmatrix} \tag{7}$$

となる．(7)が２次回帰曲線を求めるための正規方程式である．

例えば，以下の７個のデータ$(x_1,y_1),\cdots,(x_7,y_7)$ の２次回帰曲線を求めてみよう．

								計
x_k	1	2	3	4	5	6	7	28
y_k	5	4	2	1	3	7	5	27
x_k^4	1	16	81	256	625	1296	2401	4676
x_k^3	1	8	27	64	125	216	343	784
x_k^2	1	4	9	16	25	36	49	140
$x_k^2y_k$	5	16	18	16	75	252	245	627
x_ky_k	5	8	6	4	15	42	35	115

正規方程式は

$$\begin{pmatrix}4676 & 784 & 140\\ 784 & 140 & 28\\ 140 & 28 & 7\end{pmatrix}\begin{pmatrix}a\\ b\\ c\end{pmatrix}=\begin{pmatrix}627\\ 115\\ 27\end{pmatrix}$$

であり，

$$\begin{pmatrix}4676 & 784 & 140\\ 784 & 140 & 28\\ 140 & 28 & 7\end{pmatrix}^{-1}=\frac{1}{84}\begin{pmatrix}1 & -8 & 12\\ -8 & 67 & -108\\ 12 & -108 & 204\end{pmatrix}$$

である．よって，

$$\begin{pmatrix} a \\ b \\ c \end{pmatrix} = \frac{1}{84}\begin{pmatrix} 1 & -8 & 12 \\ -8 & 67 & -108 \\ 12 & -108 & 204 \end{pmatrix}\begin{pmatrix} 627 \\ 115 \\ 27 \end{pmatrix} = \frac{1}{84}\begin{pmatrix} 31 \\ -227 \\ 51 \end{pmatrix}$$

である．したがって，２次回帰曲線は

$$y = \frac{1}{81}x^2 - \frac{227}{81}x + \frac{51}{81}$$

となる．

問題 3．以下の 7 個のデータ$(x_1, y_1), \cdots, (x_7, y_7)$ の２次回帰曲線を求めよ．

x	1	2	3	4	5	6	7
y	2	6	11	12	10	5	3

回帰直線a, bを求めるための正規方程式を立てよ．

解答

問題 1．$\begin{pmatrix} 76.51 & 23.1 \\ 23.1 & 7 \end{pmatrix}\begin{pmatrix} a \\ b \end{pmatrix} = \begin{pmatrix} 829.3 \\ 250 \end{pmatrix}$　問題 2．$y = 15.36x - 14.96$　問題 3．$y = -\frac{43}{42}x^2 + \frac{172}{21}x - \frac{37}{7}$