

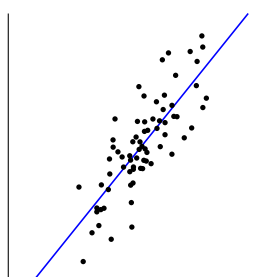
1.5 回帰直線を表す方程式

変量 X の値と変量 Y の値との順序対を要素とする 2 次元のデータについて、相関係数の絶対値が 1 に近いとき、散布図の点は一つ直線つまり 1 次関数のグラフの近くに分布する。つまり、データの各要素の Y の値が X の値の 1 次関数の値に近いことになる。

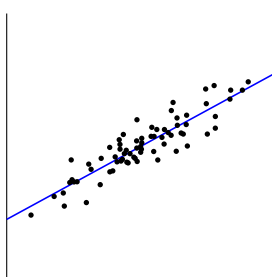
変量 X の値と変量 Y の値との順序対を要素とする 2 次元のデータについて、相関係数の絶対値が 1 に近いとき、散布図の点は一つ直線つまり 1 次関数のグラフの近くに分布する。つまり、データの各要素の Y の値が X の値の 1 次関数の値に近いことになる。その 1 次関数のグラフである直線を Y の X への回帰直線という。

変量 X の値と変量 Y の値との順序対を要素とする 2 次元のデータについて、相関係数の絶対値が 1 に近いとき、散布図の点は一つ直線つまり 1 次関数のグラフの近くに分布する。つまり、データの各要素の Y の値が X の値の 1 次関数の値に近いことになる。その 1 次関数のグラフである直線を Y の X への回帰直線という。

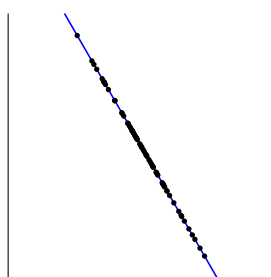
2 次元のデータの散布図に対して回帰直線は例えば次の図の青い直線のようになる。



相関係数 0.8



相関係数 0.9



相関係数 -1

[公式] 変量 X の値と変量 Y の値との順序対を要素とする 2 次元のデータについて, X の平均を $\bar{x}$ とおき, Y の平均を $\bar{y}$ とおき, X の分散を v_X とおき, X と Y との共分散を c_{XY} とおくとき, Y の X への回帰直線は xy 座標平面において方程式 $y = \frac{c_{XY}}{v_X}(x - \bar{x}) + \bar{y}$ で表される.

[公式] 変量 X の値と変量 Y の値との順序対を要素とする 2 次元のデータについて、 X の平均を $\bar{x}$ とおき、 Y の平均を $\bar{y}$ とおき、 X の分散を v_X とおき、 X と Y との共分散を c_{XY} とおくと、 Y の X への回帰直線は xy 座標平面において方程式 $y = \frac{c_{XY}}{v_X}(x - \bar{x}) + \bar{y}$ で表される。

この公式を導く。以下の定理を用いる。

[定理 1.3] 正の自然数 n に対して n 個の数 $x_1, x_2, x_3, \dots, x_n$ を要素とするデータについて、平均を $\bar{x}$ とおくと、分散は

$$v = \frac{1}{n} \sum_{k=1}^n x_k^2 - \bar{x}^2 .$$

[定理 1.4.1] 正の自然数 n に対して、変量 X の値と変量 Y の値との n 個の順序対を要素とする 2 次元のデータ

$$(x_1, y_1), (x_2, y_2), (x_3, y_3), (x_4, y_4), \dots, (x_n, y_n)$$

について、 X の平均を $\bar{x}$ とおき、 Y の平均を $\bar{y}$ とおくと、 X と Y との共分散 c_{XY} は

$$c_{XY} = \frac{1}{n} \sum_{k=1}^n (x_k y_k) - \bar{x} \bar{y} .$$

正の自然数 n に対して、変量 X の値と変量 Y の値との n 個の順序対を要素とする次の 2 次元のデータがあるとする：

$$(x_1, y_1), (x_2, y_2), (x_3, y_3), (x_4, y_4), \dots, (x_n, y_n) .$$

正の自然数 n に対して、変量 X の値と変量 Y の値との n 個の順序対を要素とする次の 2 次元のデータがあるとする：

$$(x_1, y_1), (x_2, y_2), (x_3, y_3), (x_4, y_4), \dots, (x_n, y_n) .$$

$$X \text{ の平均 } \bar{x} \text{ は } \bar{x} = \frac{1}{n} \sum_{k=1}^n x_k . \quad Y \text{ の平均 } \bar{y} \text{ は } \bar{y} = \frac{1}{n} \sum_{k=1}^n y_k .$$

正の自然数 n に対して、変量 X の値と変量 Y の値との n 個の順序対を要素とする次の 2 次元のデータがあるとする：

$$(x_1, y_1), (x_2, y_2), (x_3, y_3), (x_4, y_4), \dots, (x_n, y_n) .$$

X の平均 $\bar{x}$ は $\bar{x} = \frac{1}{n} \sum_{k=1}^n x_k$. Y の平均 $\bar{y}$ は $\bar{y} = \frac{1}{n} \sum_{k=1}^n y_k$. よって,

$$\sum_{k=1}^n x_k = n\bar{x} , \quad \sum_{k=1}^n y_k = n\bar{y} .$$

正の自然数 n に対して、変量 X の値と変量 Y の値との n 個の順序対を要素とする次の 2 次元のデータがあるとする：

$$(x_1, y_1), (x_2, y_2), (x_3, y_3), (x_4, y_4), \dots, (x_n, y_n) .$$

X の平均 $\bar{x}$ は $\bar{x} = \frac{1}{n} \sum_{k=1}^n x_k$. Y の平均 $\bar{y}$ は $\bar{y} = \frac{1}{n} \sum_{k=1}^n y_k$. よって,

$$\sum_{k=1}^n x_k = n\bar{x} , \quad \sum_{k=1}^n y_k = n\bar{y} .$$

定理 1.3 により, X の分散 v_X は $v_X = \frac{1}{n} \sum_{k=1}^n x_k^2 - \bar{x}^2$, Y の分散 v_Y は

$$v_Y = \frac{1}{n} \sum_{k=1}^n y_k^2 - \bar{y}^2 .$$

正の自然数 n に対して，変量 X の値と変量 Y の値との n 個の順序対を要素とする次の 2 次元のデータがあるとする：

$$(x_1, y_1), (x_2, y_2), (x_3, y_3), (x_4, y_4), \dots, (x_n, y_n) .$$

X の平均 $\bar{x}$ は $\bar{x} = \frac{1}{n} \sum_{k=1}^n x_k$. Y の平均 $\bar{y}$ は $\bar{y} = \frac{1}{n} \sum_{k=1}^n y_k$. よって，

$$\sum_{k=1}^n x_k = n\bar{x} , \quad \sum_{k=1}^n y_k = n\bar{y} .$$

定理 1.3 により， X の分散 v_X は $v_X = \frac{1}{n} \sum_{k=1}^n x_k^2 - \bar{x}^2$, Y の分散 v_Y は

$$v_Y = \frac{1}{n} \sum_{k=1}^n y_k^2 - \bar{y}^2 . \quad \text{よって，}$$

$$\sum_{k=1}^n x_k^2 = n(v_X + \bar{x}^2) , \quad \sum_{k=1}^n y_k^2 = n(v_Y + \bar{y}^2) .$$

正の自然数 n に対して、変量 X の値と変量 Y の値との n 個の順序対を要素とする次の 2 次元のデータがあるとする：

$$(x_1, y_1), (x_2, y_2), (x_3, y_3), (x_4, y_4), \dots, (x_n, y_n) .$$

X の平均 $\bar{x}$ は $\bar{x} = \frac{1}{n} \sum_{k=1}^n x_k$. Y の平均 $\bar{y}$ は $\bar{y} = \frac{1}{n} \sum_{k=1}^n y_k$. よって,

$$\sum_{k=1}^n x_k = n\bar{x} , \quad \sum_{k=1}^n y_k = n\bar{y} .$$

定理 1.3 により, X の分散 v_X は $v_X = \frac{1}{n} \sum_{k=1}^n x_k^2 - \bar{x}^2$, Y の分散 v_Y は

$$v_Y = \frac{1}{n} \sum_{k=1}^n y_k^2 - \bar{y}^2 . \quad \text{よって,}$$

$$\sum_{k=1}^n x_k^2 = n(v_X + \bar{x}^2) , \quad \sum_{k=1}^n y_k^2 = n(v_Y + \bar{y}^2) .$$

定理 1.4.1 により, X と Y との共分散 c_{XY} は $c_{XY} = \frac{1}{n} \sum_{k=1}^n (x_k y_k) - \bar{x} \bar{y}$.

よって

$$\sum_{k=1}^n (x_k y_k) = n(c_{XY} + \bar{x} \bar{y}) .$$

定数 a, b に対して, 変数 x の 1 次関数 $y = ax + b$ のグラフが回帰直線になるようにする. n 個の順序対

$$(x_1, y_1), (x_2, y_2), (x_3, y_3), (x_4, y_4), \dots, (x_n, y_n)$$

において, 自然数を表す変数 $k = 1, 2, 3, \dots, n$ に対して, x_k における 1 次関数の値 $ax_k + b$ が y_k の値に近くなるようにする.

定数 a, b に対して, 変数 x の 1 次関数 $y = ax + b$ のグラフが回帰直線になるようにする. n 個の順序対

$$(x_1, y_1), (x_2, y_2), (x_3, y_3), (x_4, y_4), \dots, (x_n, y_n)$$

において, 自然数を表す変数 $k = 1, 2, 3, \dots, n$ に対して, x_k における 1 次関数の値 $ax_k + b$ が y_k の値に近くなるようにする. そこで, $ax_k + b$ と y_k との差の 2 乗の総和 $\sum_{k=1}^n (ax_k + b - y_k)^2$ の値が最小になるように a 及び b の値を定める.

定数 a, b に対して、変数 x の 1 次関数 $y = ax + b$ のグラフが回帰直線になるようにする. n 個の順序対

$$(x_1, y_1), (x_2, y_2), (x_3, y_3), (x_4, y_4), \dots, (x_n, y_n)$$

において、自然数を表す変数 $k = 1, 2, 3, \dots, n$ に対して、 x_k における 1 次関数の値 $ax_k + b$ が y_k の値に近くなるようにする. そこで、 $ax_k + b$ と y_k との差の 2 乗の総和 $\sum_{k=1}^n (ax_k + b - y_k)^2$ の値が最小になるように a 及び b の値を定める. そのために、 $\sum_{k=1}^n (ax_k + b - y_k)^2$ を、 b について平方完成してから a について平方完成する.

$$\begin{aligned}
& \sum_{k=1}^n (ax_k + b - y_k)^2 \\
&= \sum_{k=1}^n (a^2 x_k^2 + b^2 + y_k^2 + 2ax_k b - 2ax_k y_k - 2by_k) \\
&= a^2 \sum_{k=1}^n x_k^2 + \sum_{k=1}^n b^2 + \sum_{k=1}^n y_k^2 + 2ab \sum_{k=1}^n x_k - 2a \sum_{k=1}^n (x_k y_k) - 2b \sum_{k=1}^n y_k
\end{aligned}$$

$$\begin{aligned}
& \sum_{k=1}^n (ax_k + b - y_k)^2 \\
&= \sum_{k=1}^n (a^2 x_k^2 + b^2 + y_k^2 + 2ax_k b - 2ax_k y_k - 2by_k) \\
&= a^2 \sum_{k=1}^n x_k^2 + \sum_{k=1}^n b^2 + \sum_{k=1}^n y_k^2 + 2ab \sum_{k=1}^n x_k - 2a \sum_{k=1}^n (x_k y_k) - 2b \sum_{k=1}^n y_k \\
&\sum_{k=1}^n x_k^2 = n(v_X + \overline{x}^2) \ , \quad \sum_{k=1}^n y_k^2 = n(v_Y + \overline{y}^2) \ , \quad \sum_{k=1}^n x_k = n\overline{x} \ , \quad \sum_{k=1}^n (x_k y_k) = \\
&n(c_{XY} + \overline{x} \, \overline{y}) \ , \quad \sum_{k=1}^n y_k = n\overline{y} \ .
\end{aligned}$$

$$\begin{aligned}
& \sum_{k=1}^n (ax_k + b - y_k)^2 \\
&= \sum_{k=1}^n (a^2 x_k^2 + b^2 + y_k^2 + 2ax_k b - 2ax_k y_k - 2by_k) \\
&= a^2 \sum_{k=1}^n x_k^2 + \sum_{k=1}^n b^2 + \sum_{k=1}^n y_k^2 + 2ab \sum_{k=1}^n x_k - 2a \sum_{k=1}^n (x_k y_k) - 2b \sum_{k=1}^n y_k \\
&= a^2 n(v_X + \bar{x}^2) + nb^2 + n(v_Y + \bar{y}^2) + 2abn\bar{x} - 2an(c_{XY} + \bar{x}\bar{y}) - 2bn\bar{y} . \\
&\sum_{k=1}^n x_k^2 = n(v_X + \bar{x}^2) \ , \quad \sum_{k=1}^n y_k^2 = n(v_Y + \bar{y}^2) \ , \quad \sum_{k=1}^n x_k = n\bar{x} \ , \quad \sum_{k=1}^n (x_k y_k) = \\
&n(c_{XY} + \bar{x}\bar{y}) \ , \quad \sum_{k=1}^n y_k = n\bar{y} \ .
\end{aligned}$$

$$\begin{aligned}
& \sum_{k=1}^n (ax_k + b - y_k)^2 \\
&= \sum_{k=1}^n (a^2 x_k^2 + b^2 + y_k^2 + 2ax_k b - 2ax_k y_k - 2by_k) \\
&= a^2 \sum_{k=1}^n x_k^2 + \sum_{k=1}^n b^2 + \sum_{k=1}^n y_k^2 + 2ab \sum_{k=1}^n x_k - 2a \sum_{k=1}^n (x_k y_k) - 2b \sum_{k=1}^n y_k \\
&= a^2 n(v_X + \bar{x}^2) + nb^2 + n(v_Y + \bar{y}^2) + 2abn\bar{x} - 2an(c_{XY} + \bar{x}\bar{y}) - 2bn\bar{y} .
\end{aligned}$$

この式を b の 2 次式として平方完成する.

$$\begin{aligned}
& a^2 n(v_X + \bar{x}^2) + nb^2 + n(v_Y + \bar{y}^2) + 2abn\bar{x} - 2an(c_{XY} + \bar{x}\bar{y}) - 2bn\bar{y} \\
&= n\{b^2 + (2a\bar{x} - 2\bar{y})b\} + n\{a^2(v_X + \bar{x}^2) + v_Y + \bar{y}^2 - 2ac_{XY} - 2a\bar{x}\bar{y}\} \\
&= n\{b^2 + 2(a\bar{x} - \bar{y})b + (a\bar{x} - \bar{y})^2\} - n(a^2 \bar{x}^2 - 2a\bar{x}\bar{y} + \bar{y}^2) \\
&\quad + n\{a^2(v_X + \bar{x}^2) + v_Y + \bar{y}^2 - 2ac_{XY} - 2a\bar{x}\bar{y}\} \\
&= n(b + a\bar{x} - \bar{y})^2 + na^2 v_X - 2nac_{XY} + nv_Y .
\end{aligned}$$

$$\begin{aligned}
& \sum_{k=1}^n (ax_k + b - y_k)^2 \\
&= \sum_{k=1}^n (a^2 x_k^2 + b^2 + y_k^2 + 2ax_k b - 2ax_k y_k - 2by_k) \\
&= a^2 \sum_{k=1}^n x_k^2 + \sum_{k=1}^n b^2 + \sum_{k=1}^n y_k^2 + 2ab \sum_{k=1}^n x_k - 2a \sum_{k=1}^n (x_k y_k) - 2b \sum_{k=1}^n y_k \\
&= a^2 n(v_X + \bar{x}^2) + nb^2 + n(v_Y + \bar{y}^2) + 2abn\bar{x} - 2an(c_{XY} + \bar{x}\bar{y}) - 2bn\bar{y} .
\end{aligned}$$

この式を b の 2 次式として平方完成する.

$$\begin{aligned}
& a^2 n(v_X + \bar{x}^2) + nb^2 + n(v_Y + \bar{y}^2) + 2abn\bar{x} - 2an(c_{XY} + \bar{x}\bar{y}) - 2bn\bar{y} \\
&= n\{b^2 + (2a\bar{x} - 2\bar{y})b\} + n\{a^2(v_X + \bar{x}^2) + v_Y + \bar{y}^2 - 2ac_{XY} - 2a\bar{x}\bar{y}\} \\
&= n\{b^2 + 2(a\bar{x} - \bar{y})b + (a\bar{x} - \bar{y})^2\} - n(a^2 \bar{x}^2 - 2a\bar{x}\bar{y} + \bar{y}^2) \\
&\quad + n\{a^2(v_X + \bar{x}^2) + v_Y + \bar{y}^2 - 2ac_{XY} - 2a\bar{x}\bar{y}\} \\
&= n(b + a\bar{x} - \bar{y})^2 + na^2 v_X - 2nac_{XY} + nv_Y .
\end{aligned}$$

後半の $na^2 v_X - 2nac_{XY} + nv_Y$ を a の 2 次式として平方完成する.

$$\begin{aligned}
na^2 v_X - 2nac_{XY} + nv_Y &= nv_X \left(a^2 - 2\frac{c_{XY}}{v_X} a \right) + nv_Y \\
&= nv_X \left\{ a^2 - 2\frac{c_{XY}}{v_X} a + \left(\frac{c_{XY}}{v_X} \right)^2 \right\} - \frac{nc_{XY}^2}{v_X} + nv_Y \\
&= nv_X \left(a - \frac{c_{XY}}{v_X} \right)^2 - \frac{nc_{XY}^2}{v_X} + nv_Y .
\end{aligned}$$

よって

$$\sum_{k=1}^n (ax_k + b - y_k)^2 = n(b + a\bar{x} - \bar{y})^2 + nv_X \left(a - \frac{c_{XY}}{v_X} \right)^2 - \frac{nc_{XY}^2}{v_X} + nv_Y .$$

$\sum_{k=1}^n (ax_k + b - y_k)^2$ を b について平方完成して a について平方完成すると

$$\sum_{k=1}^n (ax_k + b - y_k)^2 = n(b + a\overline{x} - \overline{y})^2 + nv_X \left(a - \frac{c_{XY}}{v_X} \right)^2 - \frac{nc_{XY}^2}{v_X} + nv_Y \ .$$

$\sum_{k=1}^n (ax_k + b - y_k)^2$ を b について平方完成して a について平方完成すると

$$\sum_{k=1}^n (ax_k + b - y_k)^2 = n(\overline{b + a\bar{x} - \bar{y}})^2 + nv_X \left(a - \frac{c_{XY}}{v_X} \right)^2 - \frac{nc_{XY}^2}{v_X} + nv_Y .$$

この式の値が最小になるのは、 $\overline{b + a\bar{x} - \bar{y}} = 0$ かつ $a - \frac{c_{XY}}{v_X} = 0$ のとき、つ

まり $a = \frac{c_{XY}}{v_X}$ かつ $b = \bar{y} - a\bar{x}$ のときである；

$\sum_{k=1}^n (ax_k + b - y_k)^2$ を b について平方完成して a について平方完成すると

$$\sum_{k=1}^n (ax_k + b - y_k)^2 = n(b + a\bar{x} - \bar{y})^2 + nv_X \left(a - \frac{c_{XY}}{v_X} \right)^2 - \frac{nc_{XY}^2}{v_X} + nv_Y .$$

この式の値が最小になるのは, $b + a\bar{x} - \bar{y} = 0$ かつ $a - \frac{c_{XY}}{v_X} = 0$ のとき, つ

まり $a = \frac{c_{XY}}{v_X}$ かつ $b = \bar{y} - a\bar{x}$ のときである; このとき,

$$ax + b = ax + \bar{y} - a\bar{x} = a(x - \bar{x}) + \bar{y} = \frac{c_{XY}}{v_X}(x - \bar{x}) + \bar{y} .$$

$\sum_{k=1}^n (ax_k + b - y_k)^2$ を b について平方完成して a について平方完成すると

$$\sum_{k=1}^n (ax_k + b - y_k)^2 = n(b + a\bar{x} - \bar{y})^2 + nv_X \left(a - \frac{c_{XY}}{v_X} \right)^2 - \frac{nc_{XY}^2}{v_X} + nv_Y .$$

この式の値が最小になるのは、 $b + a\bar{x} - \bar{y} = 0$ かつ $a - \frac{c_{XY}}{v_X} = 0$ のとき、つまり $a = \frac{c_{XY}}{v_X}$ かつ $b = \bar{y} - a\bar{x}$ のときである；このとき、

$$ax + b = ax + \bar{y} - a\bar{x} = a(x - \bar{x}) + \bar{y} = \frac{c_{XY}}{v_X}(x - \bar{x}) + \bar{y} .$$

Y の X への回帰直線は、 xy 座標平面において方程式 $y = \frac{c_{XY}}{v_X}(x - \bar{x}) + \bar{y}$ で表される。

例 4 人の学生 A 君, B 君, C 君, D 君のある科目の中間試験の点数 X と期

末試験の点数 Y について以下のようになった :

学生	A	B	C	D
中間試験の点数 X	75	67	81	73
期末試験の点数 Y	67	71	83	71

xy 座標平面において, Y の X への回帰直線を表す方程式を求める.

例 4 人の学生 A 君, B 君, C 君, D 君のある科目の中間試験の点数 X と期末試験の点数 Y について以下のようになった:

学生	A	B	C	D
中間試験の点数 X	75	67	81	73
期末試験の点数 Y	67	71	83	71

xy 座標平面において, Y の X への回帰直線を表す方程式を求める. X の平均は $\bar{x} = 74$ であり, X の分散は $v_X = 25$ であり, Y の平均は $\bar{y} = 73$ であり, X と Y との共分散は $c_{XY} = 20$ である.

例 4人の学生A君, B君, C君, D君のある科目の中間試験の点数 X と期末試験の点数 Y について以下のようになった:

学生	A	B	C	D
中間試験の点数 X	75	67	81	73
期末試験の点数 Y	67	71	83	71

xy 座標平面において, Y の X への回帰直線を表す方程式を求める. X の平均は $\bar{x} = 74$ であり, X の分散は $v_X = 25$ であり, Y の平均は $\bar{y} = 73$ であり, X と Y との共分散は $c_{XY} = 20$ である. Y の X への回帰直線を表す方程式は, $y = \frac{c_{XY}}{v_X}(x - \bar{x}) + \bar{y}$ なので,

$$y = \quad (x \quad) \quad ,$$

整理すると $y = \quad$.

例 4 人の学生 A 君, B 君, C 君, D 君のある科目の中間試験の点数 X と期末試験の点数 Y について以下のようになった :

学生	A	B	C	D
中間試験の点数 X	75	67	81	73
期末試験の点数 Y	67	71	83	71

xy 座標平面において, Y の X への回帰直線を表す方程式を求める. X の平均は $\bar{x} = 74$ であり, X の分散は $v_X = 25$ であり, Y の平均は $\bar{y} = 73$ であり, X と Y との共分散は $c_{XY} = 20$ である. Y の X への回帰直線を表す方程式は, $y = \frac{c_{XY}}{v_X}(x - \bar{x}) + \bar{y}$ なので,

$$y = \frac{20}{25}(x - 74) + 73 ,$$

整理すると $y = \frac{4x + 69}{5}$.

終

問1.5 4 人の学生 A 君, B 君, C 君, D 君の身長 (単位は cm) X と体重 (単位は kg) Y について調べると次のようになった.

学生		A	B	C	D
身長 (単位は cm)	X	158	162	174	162
体重 (単位は kg)	Y	61	53	71	55

xy 座標平面において, Y の X への回帰直線を表す方程式を求めよ.

X の平均は 164 であり, X の分散は 36 であり, Y の平均は 60 であり, X と Y との共分散は 32 である. Y の X への回帰直線を表す方程式は, $y = -(x - \quad) + \quad$, 整理すると $y = \frac{x - \quad}{\quad}$.

問1.5 4 人の学生 A 君, B 君, C 君, D 君の身長 (単位は cm) X と体重 (単位は kg) Y について調べると次のようになった.

学生		A	B	C	D
身長 (単位は cm)	X	158	162	174	162
体重 (単位は kg)	Y	61	53	71	55

xy 座標平面において, Y の X への回帰直線を表す方程式を求めよ.

X の平均は 164 であり, X の分散は 36 であり, Y の平均は 60 であり, X と Y との共分散は 32 である. Y の X への回帰直線を表す方程式は,

$y = \frac{32}{36}(x - 164) + 60$, 整理すると $y = \frac{8x - 772}{9}$.

終