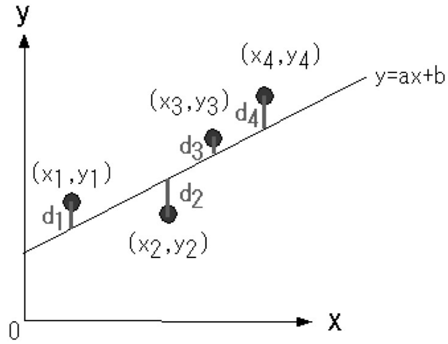
（中部大学の教育支援サイト https://edu.isc.chubu.ac.jp/ を参考にしました。）

問題　４個のデータ　(x_1, y_1)，(x_2, y_2)，(x_3, y_3)，(x_4, y_4)　をもとに
回帰直線の方程式を求めます。

方針　「もっともらしい直線」を「残差の二乗の合計を最小にするする直線」と考えます。



図を参考にしながら　$S=d_1^2+d_2^2+d_3^2+d_4^2$　として S を実際に計算してみます。

$$
\begin{aligned}
S &= \{y_1-(ax_1+b)\}^2+\{y_2-(ax_2+b)\}^2+\{y_3-(ax_3+b)\}^2+\{y_4-(ax_4+b)\}^2 \\
&= y_1^2-2y_1(ax_1+b)+(ax_1+b)^2 \\
&\quad +y_2^2-2y_2(ax_2+b)+(ax_2+b)^2 \\
&\quad +y_3^2-2y_3(ax_3+b)+(ax_3+b)^2 \\
&\quad +y_4^2-2y_4(ax_4+b)+(ax_4+b)^2 \\
&= (y_1^2+y_2^2+y_3^2+y_4^2)+a^2(x_1^2+x_2^2+x_3^2+x_4^2)-2b(y_1+y_2+y_3+y_4) \\
&\quad -2a(x_1y_1+x_2y_2+x_3y_3+x_4y_4)+2ab(x_1+x_2+x_3+x_4)+4b^2
\end{aligned}
$$

データに基づく定数部分を
右の表のように置き換えると
S は次のようになります。

$A=(y_1^2+y_2^2+y_3^2+y_4^2)$
$B=(x_1^2+x_2^2+x_3^2+x_4^2)$
$C=(y_1+y_2+y_3+y_4)$
$D=(x_1y_1+x_2y_2+x_3y_3+x_4y_4)$
$E=(x_1+x_2+x_3+x_4)$

つまり,

$\boxed{S=A+a^2B-2bC-2aD+2abE+4b^2 \quad \cdots (*1)}$　となります。

＊－－－－－＊－－－－－＊－－－－－＊－－－－－＊－－－－－＊

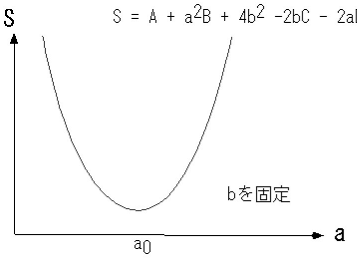
$$\boxed{S = A + a^2B - 2bC - 2aD + 2abE + 4b^2 \quad \cdots (*1)}$$

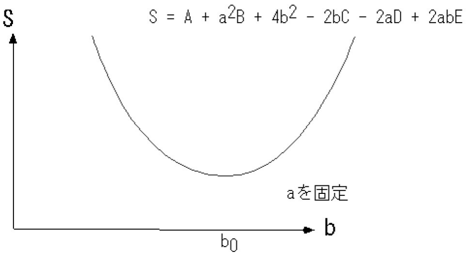
A,B,C,D,E はテータに基づく定数と考えることができます。

そのうえで次のように考えることができます。

- b を固定すると，S は 変数 a に関する２次関数である。
- a を固定すると，S は 変数 b に関する２次関数である。

数学Ⅰでは２次関数の（最小値/最大値を与える）頂点を平方完成によって求めますが,《頂点における接線の傾きは 0 である》ことを用いて S が最小となる点を見つけます。





$(*1)$ を微分（偏微分）します。

$$\frac{\partial S}{\partial a} = -2D + 2aB + 2bE = 0 \quad \cdots ①$$

$$\frac{\partial S}{\partial b} = -2C + 2aE + 8b = 0 \quad \cdots ②$$

①②を連立して a,b を求めます。

b を消去します。

①× 2　　$-4D + 4aB + 4bE = 0$

②× $\frac{1}{2}E$　　$-CE + aE^2 + 4bE = 0$

$$\begin{array}{rr} & -4D + 4aB + 4bE = 0 \\ - & -CE + aE^2 + 4bE = 0 \\ \hline & -4D + CE + (4B - E^2)a = 0 \end{array}$$

$$a = \frac{4D - CE}{4B - E^2}$$

a を消去します。

①× E　　$-2DE + 2aBE + 2bE^2 = 0$

②× B　　$-2BC + 2aBE + 8bB = 0$

$$\begin{array}{rr} & -2DE + 2aBE + 2bE^2 = 0 \\ - & -2BC + 2aBE + 8bB = 0 \\ \hline & -2DE + 2BC + (2E^2 - 8B)b = 0 \end{array}$$

$$b = \frac{DE - BC}{E^2 - 4B} = \frac{BC - DE}{4B - E^2}$$

$a = \dfrac{4D - CE}{4B - E^2}$ を一般的に書くと

$b = \dfrac{DE - BC}{E^2 - 4B} = \dfrac{BC - DE}{4B - E^2}$ は

$$a = \frac{nD - CE}{nB - E^2} \quad \cdots ③$$

$$b = \frac{BC - DE}{nB - E^2} \quad \cdots ④$$

となります。

＊－－－－－＊－－－－－＊－－－－－＊－－－－－＊－－－－－＊

このページでは　【③④が　$a=\dfrac{s_{xy}}{s_x^2}$，　$b=\overline{y}-\dfrac{s_{xy}\overline{x}}{{s_x}^2}=\overline{y}-a\overline{x}$　となることを示します。】

その前に【③，④を変形するために，式の表現とその意味をまとめておきます。】

分散の定義は「偏差の２乗の平均」つまり　$s_x^2=\dfrac{1}{n}\sum_{i=1}^{n}(x_i-\overline{x})^2$　ですが，

式を変形して「２乗の平均－平均の２乗」　${s_x}^2=\dfrac{1}{n}\sum_{i=1}^{n}x_i^2-(\overline{x})^2$　と書くこともできます。

つまり　$s_x^2=\dfrac{1}{n}\sum_{i=1}^{n}(x_i-\overline{x})^2=\dfrac{1}{n}\sum_{i=1}^{n}x_i^2-(\overline{x})^2=\overline{x^2}-(\overline{x})^2$　…⑤

同様に　$s_y^2=\dfrac{1}{n}\sum_{i=1}^{n}(y_i-\overline{y})^2=\dfrac{1}{n}\sum_{i=1}^{n}y_i^2-(\overline{y})^2=\overline{y^2}-(\overline{y})^2$　…⑥

共分散　$s_{xy}=\dfrac{1}{n}\sum_{i=1}^{n}(x_i-\overline{x})(y_i-\overline{y})=\dfrac{1}{n}\sum_{i=1}^{n}x_iy_i-\overline{x}\cdot\overline{y}=\overline{xy}-\overline{x}\cdot\overline{y}$　…⑦　となります。

相関係数は定義により　$r_{xy}=\dfrac{s_{xy}}{s_xs_y}$　…⑧　これで準備完了です。

【準備ができたところで，③④を⑤⑥⑦⑧を用いて表現することを考えます。】

$$s_x^2=\frac{1}{n}\sum_{i=1}^{n}x_i^2-(\overline{x})^2=\frac{\sum_{i=1}^{n}x_i^2}{n}-\left(\frac{\sum_{i=1}^{n}x_i}{n}\right)^2=\frac{1}{n^2}(nB-E^2)$$ …⑤を変形して

③の分母　$nB-E^2=n^2\cdot s_x^2$　…⑨を得ます。

$$s_{xy}=\frac{1}{n}\sum_{i=1}^{n}x_iy_i-\overline{x}\cdot\overline{y}=\frac{\sum_{i=1}^{n}x_iy_i}{n}-\frac{\sum_{i=1}^{n}x_i}{n}\cdot\frac{\sum_{i=1}^{n}y_i}{n}=\frac{1}{n^2}(nD-CE)$$ …⑦を変形して

③の分子　$nD-CE=n^2\cdot s_{xy}$　…⑩を得ます。

これによって a を簡潔に表すことができました。

$$\boxed{a=\frac{nD-CE}{nB-E^2}=\frac{n^2\cdot s_{xy}}{n^2\cdot s_x^2}=\frac{s_{xy}}{s_x^2}}$$

＊－－－－－＊－－－－－＊－－－－－＊－－－－－＊－－－－－＊

次に　$b=\dfrac{BC-DE}{nB-E^2}$　…④　この式を変形します。

④の分母は③の分母と同じで　$nB-E^2=n^2\cdot s_x^2$

④の分子について右の置き換えを使います。

$$BC-DE=n\cdot\overline{x^2}\times n\cdot\overline{y}-n\cdot\overline{xy}\times n\cdot\overline{x}$$
$$=n^2(\overline{x^2}\cdot\overline{y}-\overline{xy}\cdot\overline{x})$$

⑤　$s_x^2=\overline{x^2}-(\overline{x})^2$　より　$\overline{x^2}=s_x^2+(\overline{x})^2$

⑦　$s_{xy}=\overline{xy}-\overline{x}\cdot\overline{y}$　より　$\overline{xy}=s_{xy}+\overline{x}\cdot\overline{y}$

$$=n^2\{(s_x^2+(\overline{x})^2)\cdot\overline{y}-(s_{xy}+\overline{x}\cdot\overline{y})\cdot\overline{x}\}$$

よって　$b=\dfrac{BC-DE}{nB-E^2}$

$$=\frac{s_x^2\cdot\overline{y}+(\overline{x})^2\cdot\overline{y}-s_{xy}\cdot\overline{x}-(\overline{x})^2\cdot\overline{y}}{s_x^2}$$

$$=\overline{y}-\frac{s_{xy}}{s_x^2}\cdot\overline{x}=\overline{y}-a\cdot\overline{x}$$　示されました。　終

一般の場合

$A=\sum_{i=1}^{n}y_i^2=n\cdot\overline{y^2}$

$B=\sum_{i=1}^{n}x_i^2=n\cdot\overline{x^2}$

$C=\sum_{i=1}^{n}y_i=n\cdot\overline{y}$

$D=\sum_{i=1}^{n}x_iy_i=n\cdot\overline{xy}$

$E=\sum_{i=1}^{n}x_i=n\cdot\overline{x}$

回帰直線の方程式は $y=ax+b$ で

$a=\dfrac{s_{xy}}{s_x^2}$，　$b=\overline{y}-\dfrac{s_{xy}\overline{x}}{{s_x}^2}=\overline{y}-a\overline{x}$

$y-\overline{y}=a(x-\overline{x})$

＊－－－－－＊－－－－－＊－－－－－＊－－－－－＊－－－－－＊