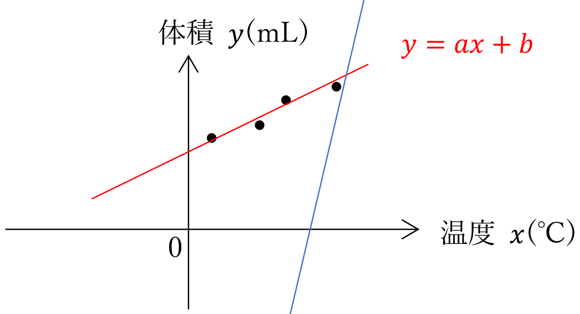
最小自乗法とは

● 測定値から近似曲線（直線）を求める方法である

次のような例で考えてみよう。一定量の気体について，圧力一定で温度を変えたときの体積を測定したところ，下図のような結果が得られた。表中の x_1～x_4 ，y_1～y_1 は測定値である。測定値は誤差を含む。真の値 x ，y に直線的関係 $y = ax + b$ があると思われるとき，測定値から最も確からしい定数 a ，b を求めるのが「最小自乗法」である。



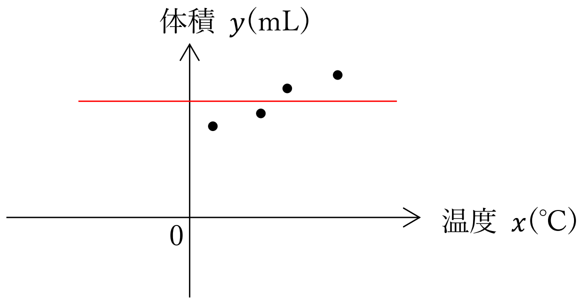
x(℃)	y(mL)
x_1	y_1
x_2	y_2
x_3	y_3
x_4	y_4

$A = x_1 + x_2 + x_3 + x_4$ ，$B = y_1 + y_2 + y_3 + y_4$ ，$C = {x_1}^2 + {x_2}^2 + {x_3}^2 + {x_4}^2$

$D = x_1y_1 + x_2y_2 + x_3y_3 + x_4y_4$ として，$a = \dfrac{4D - AB}{4C - A^2}$ ，$b = \dfrac{BC - AD}{4C - A^2}$ となる。

● なぜこのようになるのか導いてみる

温度 x_1 のとき，真の値 $y = ax_1 + b$ ，測定値 y_1 より，誤差 $y_1 - (ax_1 + b)$ である。この誤差が最小となる直線を求めるのだが，誤差に正負があると，右図のようにとんでもない結果が出てしまうことがある。そこで，誤差を２乗してその和（下式のS ）が最小となる直線を導く。



$S = \{y_1 - (ax_1 + b)\}^2 + \cdots + \{y_4 - (ax_4 + b)\}^2$ ⋯ ⓪

注意してもらいたいのは，x_1 ⋯，y_1 ⋯は測定値なので，定数である。S の値は「変数 a ，b 」により変化すると考える。

上で「定数 a ，b を求める」といっているので，混乱しそうなのだが⋯。

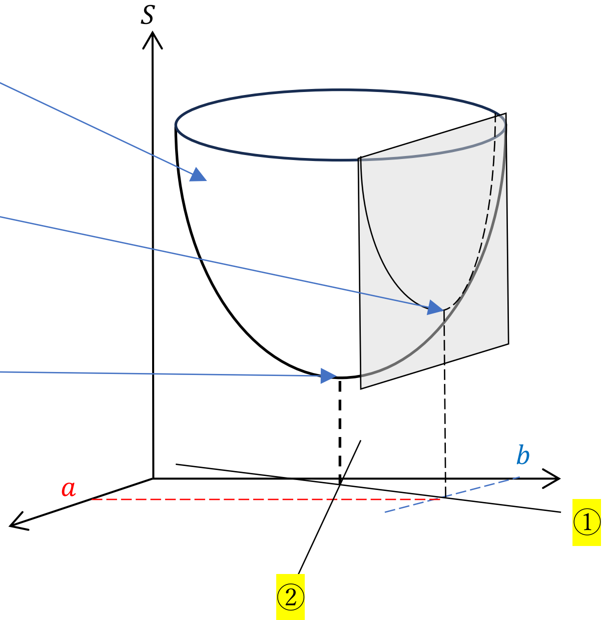
ここでさらに混乱させる表現が出てくるので覚悟してほしい。

S の値は２つの変数 a ，b で決まるのだが，両方変化させると訳がわからなくなる。そこで，一旦 b を定数と考える。すると S =(正の数)× a^2+⋯となり，下に凸の２次曲線となる。したがって， S を a で微分した値が 0 のとき，S が最小値をとる。

$$\frac{\partial S}{\partial a} = 2\{y_1 - (ax_1 + b)\}(-x_1) + \cdots + 2\{y_4 - (ax_4 + b)\}(-x_4) = 0$$

$$\therefore \quad ({x_1}^2 + \cdots + {x_4}^2)a + (x_1 + \cdots + x_4)b - (x_1y_1 + \cdots + x_4y_4) = 0 \qquad \cdots ①$$

ところで，一旦 b を定数と考えたが，b の値が変わると，S の最小値やそのときの a の値も変わってくる。



お椀のような図の側面が式⓪のグラフである。

一旦 b を定数と考え，最低値を求めたのは，灰色の面の切り口（２次曲線）の最小値を求めたことになる。b の値を変えると灰色の切り口が移動し最小値の位置も変わるが，切り口の最小値は常に①の直線の上にある。

S 全体の最小値は，①の直線の上のどこかにあるのだが，このままではわからない。

次に a を定数と考えをると S =(正の数)× b^2+⋯となり，下に凸の２次曲線となる。先ほどと同様に S を b で微分した値が 0 のとき，S が最小値をとる。

$$\frac{\partial S}{\partial b} = 2\{y_1 - (ax_1 + b)\}(-1) + \cdots + 2\{y_4 - (ax_4 + b)\}(-1) = 0$$

$$\therefore \quad (x_1 + \cdots + x_4)a + 4b - (y_1 + \cdots + y_4) = 0 \qquad \cdots \quad ②$$

図のお椀を $S - b$ 面に平行な面で切ったときの切り口（２次曲線）の最小値が，図の②の直線の上にある。

これより，直線①，②の交点の上の点が全体の最小値となる。式①，②を連立させて解くと，はじめに示した結果が得られる。

※　偏微分について

$\dfrac{\partial S}{\partial a}$ や $\dfrac{\partial S}{\partial b}$ を「偏微分」という。S が a，b ２つの変数を持つとき，一方の変数を定数とみて変化を調べる（微分する）ことである。