1 現［赤チャート数学Ⅰ 例題120］

n 個の2変量データ $(x_i,\ y_i)$ $(i=1,\ 2,\ \cdots\cdots,\ n)$ がある。

$n\geqq 2$ とし，x_i $(i=1,\ 2,\ \cdots\cdots,\ n)$ は互いに異なるとする。x_i と y_i に対して関数

$y=ax+b$ により得られる値 ax_i+b と y_i との差の2乗和を n で割った

$$R=\frac{1}{n}\left[\{y_1-(ax_1+b)\}^2+\{y_2-(ax_2+b)\}^2+\ \cdots\cdots\ +\{y_n-(ax_n+b)\}^2\right]$$

を最小にする a，b を求めることにする。

R を a と b について展開した展開式の各項の係数を，x の平均値 $\overline{x}$ と分散 ${s_x}^2$，y の平均値 $\overline{y}$ と分散 ${s_y}^2$ および x，y の共分散 s_{xy} を用いて表すと

$$R=b^2+2\overline{x}ab+\left(^{ア}\boxed{}\right)a^2-2\overline{y}b-2\left(^{イ}\boxed{}\right)a+{}^{ウ}\boxed{}$$

となる。この式の右辺を，まず b について平方完成し，次に a について平方完成することにより，R を最小にする a，b は $\overline{x}$，$\overline{y}$，${s_x}^2$，s_{xy} を用いて $a={}^{エ}\boxed{}$，$b={}^{オ}\boxed{}$ と求めることができる。

解答　（ア）　${s_x}^2+(\overline{x})^2$　　（イ）　$s_{xy}+\overline{x}\ \overline{y}$　　（ウ）　${s_y}^2+(\overline{y})^2$　　（エ）　$\dfrac{s_{xy}}{{s_x}^2}$

（オ）　$\overline{y}-\dfrac{s_{xy}\overline{x}}{{s_x}^2}$

解説　$R=\dfrac{1}{n}\left\{{y_1}^2-2(ax_1+b)y_1+(ax_1+b)^2\right\}+\left\{{y_2}^2-2(ax_2+b)y_2+(ax_2+b)^2\right\}$

$+\cdots\cdots+\left\{{y_n}^2-2(ax_n+b)y_n+(ax_n+b)^2\right\}$

$=\dfrac{1}{n}\left\{({y_1}^2+{y_2}^2+\cdots\cdots+{y_n}^2)-2a(x_1y_1+x_2y_2+\cdots\cdots+x_ny_n)\right.$

$-2b(y_1+y_2+\cdots\cdots+y_n)+a^2({x_1}^2+{x_2}^2+\cdots\cdots+{x_n}^2)$

$\left.+2ab(x_1+x_2+\cdots\cdots+x_n)+nb^2\right\}$

ここで，$\dfrac{1}{n}(x_1+x_2+\cdots\cdots+x_n)=\overline{x}$，$\dfrac{1}{n}(y_1+y_2+\cdots\cdots+y_n)=\overline{y}$ であり，

${s_x}^2=\overline{x^2}-(\overline{x})^2$，${s_y}^2=\overline{y^2}-(\overline{y})^2$ であるから

$$\frac{1}{n}({x_1}^2+{x_2}^2+\cdots\cdots+{x_n}^2)={s_x}^2+(\overline{x})^2,\quad \frac{1}{n}({y_1}^2+{y_2}^2+\cdots\cdots+{y_n}^2)={s_y}^2+(\overline{y})^2$$

また，$i=1,\ 2,\ \cdots\cdots,\ n$ について

$$(x_i-\overline{x})(y_i-\overline{y})=x_iy_i-\overline{y}x_i-\overline{x}y_i+\overline{x}\,\overline{y}$$

であるから　$x_iy_i=(x_i-\overline{x})(y_i-\overline{y})+\overline{y}x_i+\overline{x}y_i-\overline{x}\,\overline{y}$

これより　$\dfrac{1}{n}(x_1y_1+x_2y_2+\cdots\cdots+x_ny_n)$

$$=\frac{1}{n}\{(x_1-\overline{x})(y_1-\overline{y})+(x_2-\overline{x})(y_2-\overline{y})+\cdots\cdots+(x_n-\overline{x})(y_n-\overline{y})\}$$

$$+\overline{y}(x_1+x_2+\cdots\cdots+x_n)+\overline{x}(y_1+y_2+\cdots\cdots+y_n)-n\overline{x}\,\overline{y}$$

$$=s_{xy}+\overline{y}\,\overline{x}+\overline{x}\,\overline{y}-\overline{x}\,\overline{y}=s_{xy}+\overline{x}\,\overline{y}$$

よって　$R=s_y{}^2+(\overline{y})^2-2a(s_{xy}+\overline{x}\,\overline{y})-2b\overline{y}+a^2\{s_x{}^2+(\overline{x})^2\}+2ab\overline{x}+b^2$

$$=b^2+2\overline{x}ab+\{{}^{ア}s_x{}^2+(\overline{x})^2\}a^2-2\overline{y}b-2({}^{イ}s_{xy}+\overline{x}\,\overline{y})a+{}^{ウ}s_y{}^2+(\overline{y})^2$$

R を変形すると

$$R=b^2+2(a\overline{x}-\overline{y})b+\{s_x{}^2+(\overline{x})^2\}a^2-2(s_{xy}+\overline{x}\,\overline{y})a+s_y{}^2+(\overline{y})^2$$

$$=\{b+(a\overline{x}-\overline{y})\}^2-(a\overline{x}-\overline{y})^2+\{s_x{}^2+(\overline{x})^2\}a^2-2(s_{xy}+\overline{x}\,\overline{y})a+s_y{}^2+(\overline{y})^2$$

$$=\{b+(a\overline{x}-\overline{y})\}^2+s_x{}^2a^2-2s_{xy}a+s_y{}^2$$

$$=\{b+(a\overline{x}-\overline{y})\}^2+s_x{}^2\left(a^2-\frac{2s_{xy}}{s_x{}^2}a\right)+s_y{}^2$$

$$=\{b+(a\overline{x}-\overline{y})\}^2+s_x{}^2\left(a-\frac{s_{xy}}{s_x{}^2}\right)^2-\frac{s_{xy}{}^2}{s_x{}^2}+s_y{}^2$$

$$=\{b+(a\overline{x}-\overline{y})\}^2+s_x{}^2\left(a-\frac{s_{xy}}{s_x{}^2}\right)^2+\frac{s_x{}^2s_y{}^2-s_{xy}{}^2}{s_x{}^2}$$

したがって，R を最小にする a，b は，

$a={}^{エ}\dfrac{s_{xy}}{s_x{}^2}$，$b=\overline{y}-a\overline{x}={}^{オ}\overline{y}-\dfrac{s_{xy}\overline{x}}{s_x{}^2}$　　である。　終