重回帰式・補遺

7.3　重回帰式

被説明変数を 2 個以上の説明変数で表した式のことを重回帰式という。説明変数が 2 個の場合の重回帰式は回帰平面を作る。重回帰式を $\hat{z} = a + bx + cy$ とし，残差平方和を $S(a, b, c)$ とおく。

$$S(a,b,c) = \sum_{i=1}^{n}(z_i - \hat{z}_i)^2 = \sum_{i=1}^{n}(z_i - a - bx_i - cy_i)^2$$

正規方程式　残差平方和 $S(a, b, c)$ を偏微分し，最小二乗解を求める。

$$\begin{aligned} S_a(a,b,c) &= -2\sum(z_i - a - bx_i - cy_i) = 0 \\ S_b(a,b,c) &= -2\sum x_i(z_i - a - bx_i - cy_i) = 0 \\ S_c(a,b,c) &= -2\sum y_i(z_i - a - bx_i - cy_i) = 0 \end{aligned}$$

回帰定数 a，偏回帰係数 b，c は次の正規方程式を満たす。

$$\begin{aligned} &\sum(z_i - a - bx_i - cy_i) = 0 \\ &\sum x_i(z_i - a - bx_i - cy_i) = 0 \\ &\sum y_i(z_i - a - bx_i - cy_i) = 0 \end{aligned}$$

重回帰式の導出　正規方程式をデータの大きさ n で割り，平均値の式で表す。

$$\bar{z} - a - b\bar{x} - c\bar{y} = 0 \tag{i}$$
$$\overline{xz} - a\bar{x} - b\overline{x^2} - c\overline{xy} = 0 \tag{ii}$$
$$\overline{yz} - a\bar{y} - b\overline{xy} - c\overline{y^2} = 0 \tag{iii}$$

(ii) から (i) $\times\ \bar{x}$ を引き，(iii) から (i) $\times\ \bar{y}$ を引いて，文字 a を消去する。

$$\begin{aligned} \overline{xz} - \bar{x}\bar{z} - b(\overline{x^2} - \bar{x}^2) - c(\overline{xy} - \bar{x}\bar{y}) = 0 \\ \overline{yz} - \bar{y}\bar{z} - b(\overline{xy} - \bar{x}\bar{y}) - c(\overline{y^2} - \bar{y}^2) = 0 \end{aligned}$$

分散や共分散を置き換えて連立方程式を解くと，次の結果が得られる。

$$a = \bar{z} - b\bar{x} - c\bar{y}, \qquad b = \frac{s_y^2 s_{xz} - s_{xy}s_{yz}}{s_x^2 s_y^2 - {s_{xy}}^2}, \qquad c = \frac{s_x^2 s_{yz} - s_{xy}s_{xz}}{s_x^2 s_y^2 - {s_{xy}}^2} \tag{4}$$

決定係数と重相関係数　観測値 y と回帰値 $\hat{y}$ の相関係数 $r_{y\hat{y}}$ のことを重相関係数という。決定係数は重相関係数の 2 乗で表すことができる。

$$\begin{aligned} r_{y\hat{y}} &= \frac{\sum(y_i - \bar{y})(\hat{y}_i - \bar{y})}{\sqrt{\sum(y_i - \bar{y})^2}\sqrt{\sum(\hat{y}_i - \bar{y})^2}} = \frac{\sum\{(\hat{y}_i - \bar{y}) + (y_i - \hat{y}_i)\}(\hat{y}_i - \bar{y})}{\sqrt{\mathrm{TSS}}\sqrt{\mathrm{ESS}}} \\ &= \frac{\sum(\hat{y}_i - \bar{y})^2 + \sum(\hat{y}_i - \bar{y})(y_i - \hat{y}_i)}{\sqrt{\mathrm{TSS}}\sqrt{\mathrm{ESS}}} = \frac{\mathrm{ESS} + \sum(\hat{y}_i - \bar{y})(y_i - \hat{y}_i)}{\sqrt{\mathrm{TSS}}\sqrt{\mathrm{ESS}}} \end{aligned}$$

正規方程式から $\sum(\hat{y}_i-\bar{y})(y_i-\hat{y}_i)=0$ なので，$r_{y\hat{y}}=\sqrt{\mathrm{ESS}/\mathrm{TSS}}$，${r_{y\hat{y}}}^2=R^2$ が成り立つ。決定係数は次のいずれかの方法で求めることができる。

$$R^2=\frac{\mathrm{ESS}}{\mathrm{TSS}}=\frac{\sum(\hat{y}_i-\bar{y})^2}{\sum(y_i-\bar{y})^2},\qquad R^2=1-\frac{\mathrm{RSS}}{\mathrm{TSS}}=1-\frac{\sum(y_i-\hat{y}_i)^2}{\sum(y_i-\bar{y})^2},$$
$$R^2={r_{y\hat{y}}}^2=\frac{\{\sum(y_i-\bar{y})(\hat{y}_i-\bar{y})\}^2}{\sum(y_i-\bar{y})^2\sum(\hat{y}_i-\bar{y})^2} \tag{5}$$

特に単回帰（説明変数が 1 個）の場合は，決定係数は x と y の相関係数の 2 乗に等しい。

$$R^2={r_{xy}}^2=\frac{\{\sum(x_i-\bar{x})(y_i-\bar{y})\}^2}{\sum(x_i-\bar{x})^2\sum(y_i-\bar{y})^2}$$

本稿の参考文献

統計学入門（基礎統計学）
東京大学教養学部統計学教室（編）東京大学出版会 978-4-13-042065-5

統計学
久保川 達也 （著） 東京大学出版会 978-4-13-062921-8

共立講座２１世紀の数学 第２版 １４ 統計
竹村 彰通 （著） 共立出版 978-4-320-01851-8

経済・経営系のための統計入門 （事例でわかる統計シリーズ）
景山 三平 （監修） 実教出版 978-4-407-33711-2

基本統計学 第４版
宮川 公男 （著） 有斐閣 978-4-641-16455-0

www5e.biglobe.ne.jp/~emm386/statistics/