第 3 回　単回帰モデル①

回帰分析：いくつかの説明要因が興味のある変数 y に与える影響をしらべる分析

1）**単回帰モデル**：説明要因が 1 つだけの回帰モデル

2）**重回帰モデル**：説明要因が 2 つ以上の回帰モデル

[1] 単回帰モデルによる母集団の表現

x が y にどのように影響するかを統計的に推測するとき，「母集団」とは何か？

母集団（知りたい対象の全体）：<u>yとxが発生する仕組み</u>

↓

この仕組みを知れば，実際に y と x がどう発生するかをすべて説明できる。

単回帰モデル：y と x が発生する仕組みは，次の関係で表されると考える。

$y = \alpha + \beta x + u$　・・・ ①

x：**説 明 変 数**（非確率変数）　← 見える

u：**誤　差　項**（確率変数（x で説明できない要因））　← 見えない

α，β：**パラメータ**（y の分布の特徴を表す定数）　← 見えない

y：**従 属 変 数**（①式に従って発生する確率変数）　← 見える

⇒ 単回帰モデルで表される母集団の特徴：α，β の値と u の発生方法

⇒ これらがわかれば，y と x の発生する仕組み（y と x の関係）がわかる。

（例）賃金方程式：$wage = \alpha + \beta\, educ + u$　・・・ ②

wage：時給　educ：教育年数　u：能力など（観察できず，時給に影響する変数）

（ ②式の仮定 ） wage に影響する重要な要因で，観察できるのは educ だけである。

（係数βの意味） 教育年数が 1 年ふえると，時給は β だけ上がる。

[2] 最小二乗推定量による回帰モデルの推定

（問題設定）モデル①で表される母集団からランダム標本をとり，n 組の(x_i, y_i)を観察したとする。このとき，(x_i, y_i)について次の関係が成り立つ。

$$y_i = \alpha + \beta x_i + u_i \quad (i = 1, \cdots, n) \quad \cdot\cdot\cdot \text{③}$$

このとき，α と β の値を推定したい。

（問題の図形的意味）

実際に観察した n 組の(x_i, y_i)があるとき，

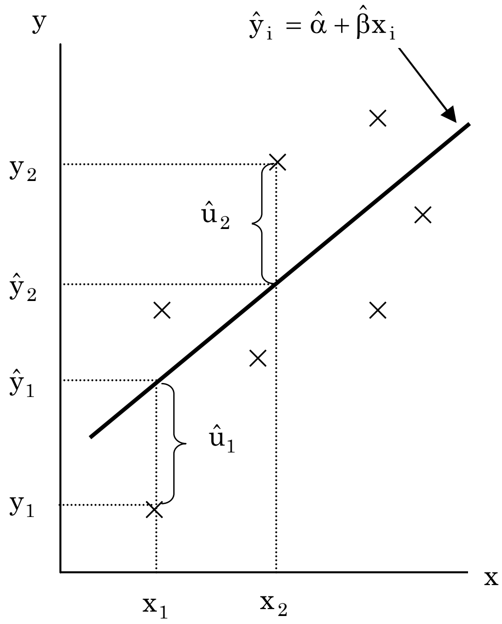
$\boxed{\alpha \text{ と } \beta \text{ の値 } \hat{\alpha} \text{ と } \hat{\beta} \text{ を求める}}$

＝ n 組の（x_i, y_i）の関係を最もうまく表す

直線 $\underline{\hat{y}_i = \hat{\alpha} + \hat{\beta} x_i}$ を求める。

（フィット）

⇒ できるだけ**残差** $\hat{u}_i = y_i - \hat{\alpha} - \hat{\beta} x_i$ が小さくなるような $\hat{\alpha}$ と $\hat{\beta}$ を求める



×印は観察された(x_i, y_i)

最小二乗法： 残差平方和 $\sum_{i=1}^{n} \hat{u}_i^2 = \sum_{i=1}^{n} (y_i - \hat{\alpha} - \hat{\beta} x_i)^2$ を $\hat{\alpha}$ と $\hat{\beta}$ について最小化するように $\hat{\alpha}$ と $\hat{\beta}$ を推定する方法。

$\hat{\alpha}$ と $\hat{\beta}$ を求める最小化の条件： $\dfrac{\partial \sum_{i=1}^{n} \hat{u}_i^2}{\partial \hat{\alpha}} = 0$, $\dfrac{\partial \sum_{i=1}^{n} \hat{u}_i^2}{\partial \hat{\beta}} = 0$

（復習①：和の微分） $\dfrac{\partial \sum_{i=1}^{n} f_i(x)}{\partial x} = \dfrac{\partial \{ f_1(x) + \cdots + f_n(x) \}}{\partial x} = \sum_{i=1}^{n} \dfrac{\partial f_i(x)}{\partial x}$

（復習②：合成関数の微分） $y = f(z)$, $z = g(x)$ ⇒ $\dfrac{\partial f(g(x))}{\partial x} = \dfrac{\partial f(z)}{\partial z} \dfrac{\partial g(x)}{\partial x}$

$\hat{\alpha}$ の条件： $\dfrac{\partial \sum_{i=1}^{n} \hat{u}_i^2}{\partial \hat{\alpha}} = 0 \;\Rightarrow$

$\hat{\beta}$ の条件： $\dfrac{\partial \sum_{i=1}^{n} \hat{u}_i^2}{\partial \hat{\beta}} = 0 \;\Rightarrow$

これらを $\hat{\alpha}$ と $\hat{\beta}$ について解くと，次の公式を得る（ただし， $\overline{x} = \dfrac{1}{n}\sum_{i=1}^{n} x_i$ ， $\overline{y} = \dfrac{1}{n}\sum_{i=1}^{n} y_i$ ）。

$$\hat{\alpha} = \overline{y} - \hat{\beta}\overline{x}$$

$$\hat{\beta} = \frac{S_{xy}}{S_{xx}}, \qquad S_{xx} = \sum_{i=1}^{n} (x_i - \overline{x})^2, \quad S_{xy} = \sum_{i=1}^{n} (x_i - \overline{x})(y_i - \overline{y})$$

※ 二つの条件から $\hat{\alpha}$ と $\hat{\beta}$ の公式を導く計算

（例）次のような 5 組の(x_i, y_i)が観察されたとき，$\hat{\alpha}$，$\hat{\beta}$，残差 $\hat{u}_i$ を計算する。

x_i	2	2	3	4	4
y_i	6	8	9	10	12