

第 III 部 線型モデル

1 直線を当てはめる基準

ポイント

- 最もよく当てはまる直線とは
- 当てはめの基準

- 誤差の定義
- 誤差の二乗和

1.1 はじめに

- 観測点に直線を当てはめてみます。
- もっともよく当てはまるようにあてはめます。
- 「もっともよく」とはどういう意味でしょうか。

問題 III-1-1

以下の式を展開しなさい。

$$(a - b - c)^2$$

解例 III-1-1

$$\begin{aligned}(a-b-c)^2 &= (a-b-c)(a-b-c) \\&= a(a-b-c) - b(a-b-c) - c(a-b-c) \\&= a^2 - ab - ac - ab + b^2 + bc - ac + bc + c^2 \\&= a^2 + b^2 + c^2 - 2ab - 2ac + 2bc\end{aligned}$$

1.2 総和

定義

n 個の数値があり、 x_i を i 番目の値とする。 x_i を全て足し合わせることを『総和』といい $\sum_{i=1}^n x_i$ という記号であらわす。

$$\sum_{i=1}^n x_i = x_1 + x_2 + \cdots + x_n \quad (1)$$

1.2.1 総和の公式

総和の定義より以下の公式を導き出すことができる。

$$\sum_{i=1}^n c = nc \quad ; c \text{ は定数} \quad (2)$$

$$\sum_{i=1}^n ax_i = a \sum_{i=1}^n x_i \quad ; a \text{ は定数} \quad (3)$$

$$\sum_{i=1}^n (x_i + y_i) = \sum_{i=1}^n x_i + \sum_{i=1}^n y_i \quad (4)$$

1.3 2つの変数の関係をあらわす直線

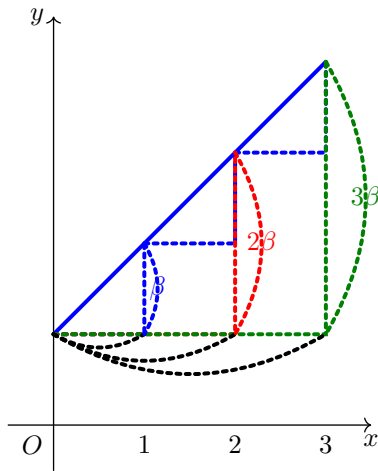
- マーケティングでは、マーケティング変数によって需要に影響を与えることができると考えている。
- 需要に影響を与える変数が操作可能な場合は間接的に需要を操作することが可能となる。
- また需要の予測が困難な場合でも間接的に予測をすることが可能な場合もある。
- これらの例として「広告と売上」「気温と売上」などがあげられる。
- ここでは2つの変数の関係をより分かりやすく表現する方法を考える。

1.3.1 直線をあらわす式

x_i を独立変数、 y_i を従属変数とし、二つの変数の間に線型モデルを仮定する。

$$y_i = \alpha + \beta x_i \quad (5)$$

- β は x_i が 1 単位増加した時の y_i の変化量をあらわす。
- x_i が n 単位増加したら、 y_i は $n\beta$ 変化する。

図 1 線型モデルにおける β の意味

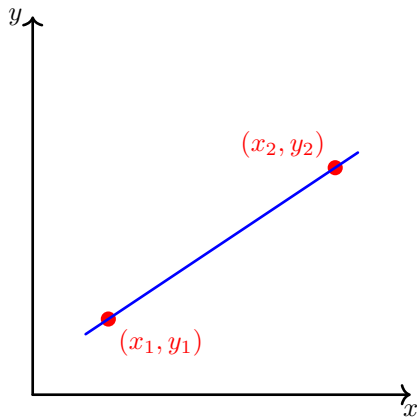
1.3.2 2点を通る直線

観測値が2点であった場合、 (x, y) の関係をもっともよくあらわす直線は2点を通る直線である。

2点を通る直線を求めるには、2組の観測値をそれぞれ x_i, y_i とする連立方程式を解けばよい。

$$\begin{cases} y_1 = \alpha + \beta x_1 \\ y_2 = \alpha + \beta x_2 \end{cases}$$

図2 2点を通る直線



問題 III-1-2

観測点の座標を (x, y) であらわす。直線を $y = \alpha + \beta x$ とする。

2点 $(-3, -3)$, $(2, 7)$ を通る直線を求めよ。

解例 III-1-2

$(-3, -3)$, $(2, 7)$ を代入したそれぞれの式は

$$-3 = \alpha - 3\beta \quad (\text{a})$$

$$7 = \alpha + 2\beta \quad (\text{b})$$

(b) 式から (a) 式を引く

$$\begin{array}{rclcl} 7 & = & \alpha & + & 2\beta \\ -) & -3 & = & \alpha & - & 3\beta \\ \hline 10 & = & & & 5\beta \\ \beta & = & & & 2 \end{array}$$

(b) 式に $\beta = 2$ を代入すると

$$7 = \alpha + 2(2)$$

$$7 = \alpha + 4$$

$$\alpha = 3$$

したがって求める直線の式は

$$y = 3 + 2x$$

問題 III-1-3

以下の連立方程式を解け

$$\begin{cases} 4x - 3y = -9 \\ 3x - 7y = 17 \end{cases}$$

解例 III-1-3

$$4x - 3y = -9 \quad (\text{a})$$

$$3x - 7y = 17 \quad (\text{b})$$

(a) 式を x について解く

$$\begin{aligned} 4x &= 3y - 9 \\ x &= \frac{3y}{4} - \frac{9}{4} \end{aligned} \quad (\text{c})$$

(c) 式を (b) 式に代入

$$\begin{aligned} 3 \left(\frac{3y}{4} - \frac{9}{4} \right) - 7y &= 17 \\ \frac{9y}{4} - \frac{27}{4} - 7y &= 17 \\ \frac{9y}{4} - 7y &= 17 + \frac{27}{4} \\ \frac{9y - 28y}{4} &= \frac{68 + 27}{4} \\ -19y &= 95 \\ y &= -5 \end{aligned} \quad (\text{d})$$

(d) を (a) 式に代入

$$4x - 3(-5) = -9$$

$$4x + 15 = -9$$

$$4x = -9 - 15$$

$$4x = -24$$

$$x = -6$$

従って

$$\begin{cases} x = -6 \\ y = -5 \end{cases}$$

問題 III-1-4

以下の連立方程式を解け。ただし、 x, y を変数とし a_i, b_j を定数とし、 $a_1b_2 \neq a_2b_1$ とする。

$$\begin{cases} a_1x + a_2y = a_3 \\ b_1x + b_2y = b_3 \end{cases}$$

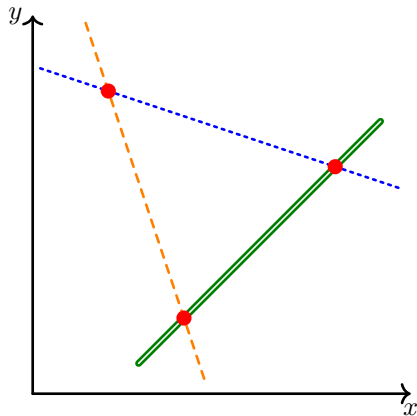
解例 III-1 - 4

$$x = \frac{a_3b_2 - a_2b_3}{a_1b_2 - a_2b_1}$$
$$y = \frac{a_1b_3 - a_3b_1}{a_1b_2 - a_2b_1}$$

1.3.3 当てはまりの基準

- 観測点が2点なら、 x と y の関係をもっともよくあらわす直線は2点を通る直線である。
- 観測点が3点以上では全ての点を通る直線は、特殊な場合を除き引けない。
- 「点を通る」以外の基準が必要である。
- 最もよく当てはまる基準に誤差を利用する。

図3 観測値が3点の場合



1.4 誤差

1.4.1 誤差の定義

当てはめる直線を

$$\hat{y}_i = \alpha + \beta x_i \quad (6)$$

とし、 y 軸上における各点 (y_i) と直線上の点 $(\hat{y}_i)$ の差として誤差 (e_i) を定義する。

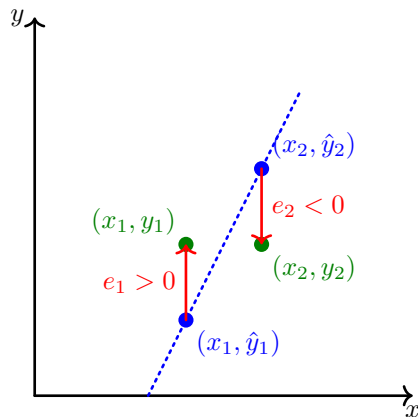
$$e_i = y_i - \hat{y}_i \quad (7)$$

$$= y_i - (\alpha + \beta x_i) \quad (8)$$

$$= y_i - \alpha - \beta x_i \quad (9)$$

定義から明らかなように誤差 (e_i) は符号を持つ。

図4 観測点と直線上の値と誤差



1.4.2 基準の候補

- 誤差の総和
- 誤差の絶対値の総和
- 誤差の二乗の総和

図5 誤差の総和が0の直線

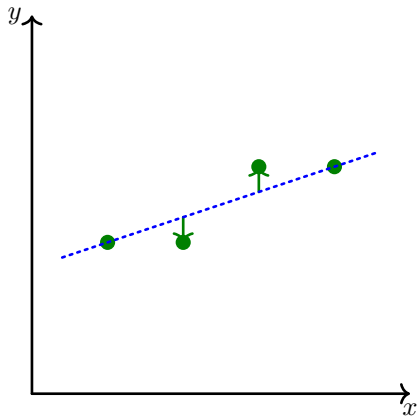
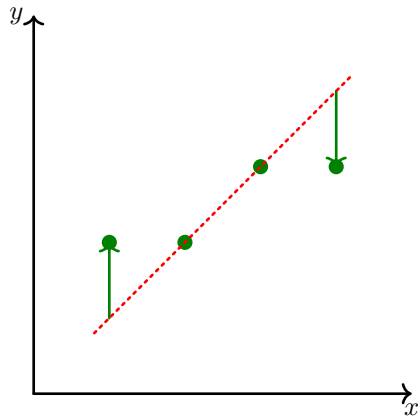


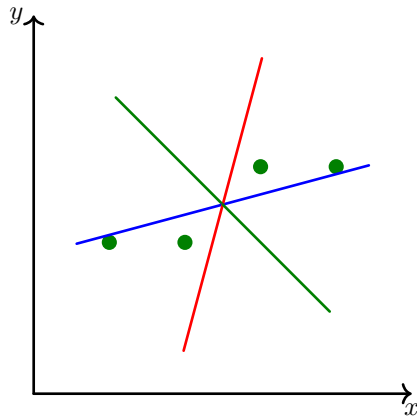
図6 大きな誤差の打ち消しあい



誤差の総和

- 誤差は符号を持っており、正の誤差と負の誤差が互いに打ち消しあってしまう。
- このような問題を『符号の問題』とよぶ。
- 符号の問題により、誤差の総和では直線を当てはめるときの基準とすることはできない。

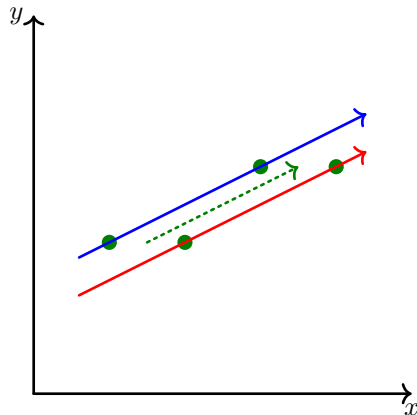
図7 同程度に当てはまる直線



誤差の絶対値の総和

- 符号の問題を解決するために、誤差の絶対値をとることを考える。
- 誤差の絶対値の総和では、同程度に当てはまる直線が無限に引ける。

図8 同程度に当てはまる直線



1.5 誤差の二乗和

- 誤差を二乗することにより符号の問題を解決する。
- 誤差の二乗の総和を『誤差の二乗和』とよぶ。
- 誤差の二乗和を直線を当てはめる基準とし、誤差の二乗和が小さいもの程、優れているとする。
- 計算で、1本だけの直線を引くことができる。
- 大きな誤差を強調し、小さな誤差をより小さく評価する特徴を持つ。
- 誤差の二乗和を最小化する基準を「最小二乗法」という。

問題 III-1-5

データは全部で n 組あり、それぞれの組を (x_i, y_i) とする。直線を

$$\hat{y}_i = \alpha + \beta x_i$$

とし、誤差を

$$e_i = y_i - \hat{y}_i$$

とする。誤差の二乗和を式であらわしなさい。

解例 III-1 - 5

$$\begin{aligned}\sum_{i=1}^n e_i^2 &= \sum_{i=1}^n (y_i - \hat{y}_i)^2 \\ &= \sum_{i=1}^n (y_i - (\alpha + \beta x_i))^2 \\ &= \sum_{i=1}^n (y_i - \alpha - \beta x_i)^2\end{aligned}$$

問題 III-1 - 6

以下の式を展開しなさい。

$$\sum_{i=1}^n (y_i - \alpha - \beta x_i)^2$$

解例 III-1 - 6

$$\begin{aligned}& \sum_{i=1}^n (y_i - \alpha - \beta x_i)^2 \\&= \sum_{i=1}^n (\alpha^2 + 2\alpha\beta x_i + \beta^2 x_i^2 - 2\alpha y_i - 2\beta x_i y_i + y_i^2) \\&= \sum_{i=1}^n \alpha^2 + \sum_{i=1}^n 2\alpha\beta x_i + \sum_{i=1}^n \beta^2 x_i^2 - \sum_{i=1}^n 2\alpha y_i - \sum_{i=1}^n 2\beta x_i y_i + \sum_{i=1}^n y_i^2 \\&= n\alpha^2 + 2\alpha\beta \sum_{i=1}^n x_i + \beta^2 \sum_{i=1}^n x_i^2 - 2\alpha \sum_{i=1}^n y_i - 2\beta \sum_{i=1}^n x_i y_i + \sum_{i=1}^n y_i^2\end{aligned}$$

1.5.1 誤差の二乗和の定義

データが n 組あり、 i 番目の組を (x_i, y_i) とする。当てはめる直線を

$$\hat{y}_i = \alpha + \beta x_i$$

とし、 y 軸上における各点 (y_i) と直線上の点 $(\hat{y}_i)$ の差を誤差 (e_i) とすると、誤差の二乗和は

$$\sum_{i=1}^n e_i^2 = n\alpha^2 + 2\alpha\beta \sum_{i=1}^n x_i + \beta^2 \sum_{i=1}^n x_i^2 - 2\alpha \sum_{i=1}^n y_i - 2\beta \sum_{i=1}^n x_i y_i + \sum_{i=1}^n y_i^2 \quad (10)$$

で定式化される。

1.6 まとめ

- 当てはめているのは直線

$$\hat{y}_i = \alpha + \beta x_i$$

- 誤差の定義は、 y 軸上における各点 (y_i) と直線上の点 $(\hat{y}_i)$ の差

$$e_i = y_i - \hat{y}_i$$

- 直線を当てめる基準は、誤差の二乗和の最小化

$$\sum_{i=1}^n e_i^2 = n\alpha^2 + 2\alpha\beta \sum_{i=1}^n x_i + \beta^2 \sum_{i=1}^n x_i^2 - 2\alpha \sum_{i=1}^n y_i - 2\beta \sum_{i=1}^n x_i y_i + \sum_{i=1}^n y_i^2$$