

第二章 単回帰¹

劉慶豊²

小樽商科大学

November 11, 2009

¹第二章からの資料は森棟公夫先生著「基礎コース 計量経済学」をもとに作成したものである。

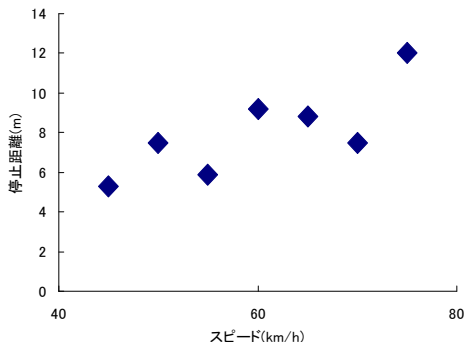
²E-mail:qliu@res.otaru-uc.ac.jp, URL:<http://www.otaru-uc.ac.jp/~qliu/> ◀ ▶ ≡ 🔍 ↺

線形回帰式

表2.1 スピードと停止距離

x	45	50	55	60	65	70	75 (km/h)
y	5.3	7.5	5.9	9.2	8.8	7.5	12 (m)

図2.1 スピードと停止距離



線形回帰式

回帰 確率変数 Y を X に回帰するというのは、 X を持って Y の変動を説明することを意味する。

例 停止距離 (Y) を車の走行スピード (X) に回帰する。車の走行スピード (X) を持って停止距離 (Y) の変動を説明する。

回帰式

$$y_i = \alpha + \beta x_i + u_i \quad i = 1, 2, \dots, n \quad (1)$$

データ $\{(x_1, y_1), (x_2, y_2), \dots, (x_n, y_n)\}$

用語の説明 x_i は説明変数 (独立変数) y_i は被説明変数 (従属変数) u_i は誤差

(1) 式の u_i は誤差

- 誤差項がないなら、 $y_i = \alpha + \beta x_i$ は一本の直線を表す。しかし、各観測値は必ずしも直線に乗らない。そのずれを表すために u_i を足した。
- 誤差項 u_i と説明変数 x_i は互いに独立である（互いに影響を与えない）。

$$E(u_i) = 0, V(u_i) = \sigma^2, i = 1, 2, \dots, n \quad (2)$$

- 個々のズレが互いに影響し合うことはない、 $u_i, i = 1, 2, \dots, n, n$ 個の誤差項は互いに独立である。

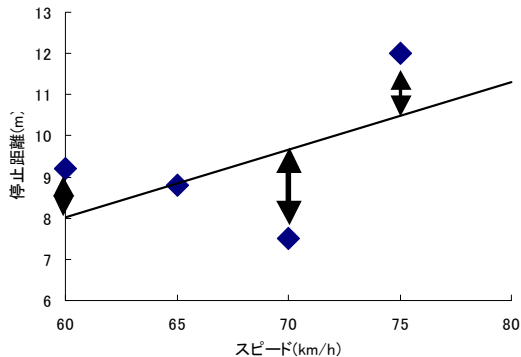
- $y_i = \alpha + \beta x_i$ 中の α と β が未知である。データを用いて未知の α と β を計算する。これを推定という。
- 推定結果は $\hat{\alpha}, \hat{\beta}$ と表記して、推定値と呼ぶ。
- 推定値を利用して $\hat{\alpha} + \hat{\beta}x_i$ で \hat{y}_i を計算する、それを回帰値または予測値と呼ぶ。

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

- 残差は観測値 y_i と予測値 \hat{y}_i の差

$$\hat{u}_i = y_i - \hat{y}_i = y_i - (\hat{\alpha} + \hat{\beta}x_i) \quad i = 1, 2, \dots, n$$

図2.2 回帰直線



一次式と直線

二つの変数 x, y の関係を数式で表すことが出来る。例えば、 $y = \alpha + \beta x$ 。
 $\alpha = 3, \beta = 2$ として、 $y = 3 + 2x$ となる。

$x = 0$ の時、 $y = 3 + 2 \times 0 = 3$, $x = 2$ の時、 $y = 3 + 2 \times 2 = 7$, $x = 5$ のとき $y = 13$ 。

グラフにこの2つの点を書いて繋げば一本の直線になる。

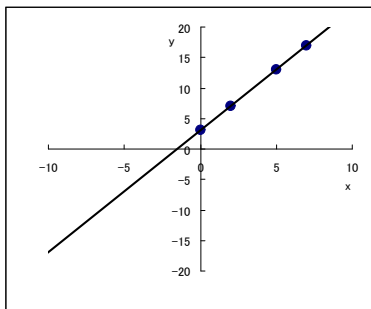
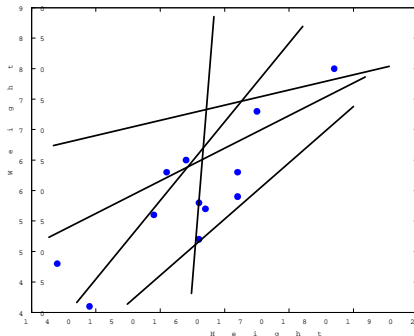


図 1 一次式と直線

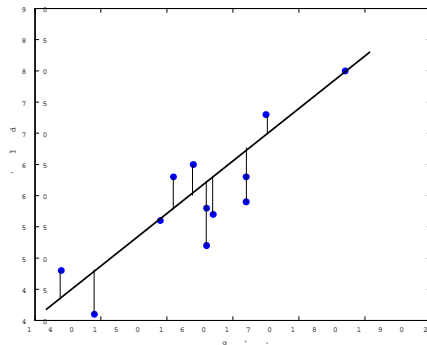
データの散布図に直線を当てはめよう

- 一本の直線を引きたいが、どうやって引けば身長と体重の関係をよく表せるのかを考えよう。
- 何本も直線を引けるが、どれがいいか基準をないと分らない。



最小二乗法の発想

- 観測値（データ）の点から直線への縦の距離の二乗の総和が一番小さくなるように直線、すなわち α と β の推定値を決める。
- 縦の距離そのものを使うこともあるが、その場合では後で出てくる計算が難しくなる。



残差二乗和（残差変動）

$$RSS = \sum_{i=1}^n (y_i - \hat{\alpha} - \hat{\beta}x_i)^2 \quad (3)$$

$$= \sum_{i=1}^n y_i^2 - \hat{\alpha} \sum_{i=1}^n y_i - \hat{\beta} \sum_{i=1}^n x_i y_i. \quad (4)$$

回帰式の推定

最小 2 乗法 (Ordinary Least Squares, OLS) 残差の二乗和を最小にするように $\hat{\alpha}$ と $\hat{\beta}$ を決める方法。

$$\hat{\beta} = \frac{s_{xy}}{s_{xx}} = r_{xy} \frac{s_y}{s_x} \quad (5)$$

$$= \frac{\sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y})}{\sum_{i=1}^n (x_i - \bar{x})^2} \quad (6)$$

$$= \frac{\sum_{i=1}^n x_i y_i - n \bar{x} \bar{y}}{\sum_{i=1}^n x_i^2 - n \bar{x}^2} \quad (7)$$

$$\hat{\alpha} = \bar{y} - \hat{\beta} \bar{x} \quad (8)$$

.

残差分散

誤差項 $u_i = y_i - (\alpha + \beta x_i)$

残差 $\hat{u}_i = y_i - (\hat{\alpha} + \hat{\beta} x_i)$

残差分散

$$s^2 = \frac{\sum_{i=1,n} \hat{u}_i^2}{n-2} \quad (9)$$

表2.2 係数の推定

変数	y	x	x^2	xy	y^2
	5.3	45	2025	238.5	28.09
	7.5	50	2500	375	56.25
	5.9	55	3025	324.5	34.81
	9.2	60	3600	552	84.64
	8.8	65	4225	572	77.44
	7.5	70	4900	525	56.25
	12	75	5625	900	144
和	56.2	420	25900	3487	481.48

$$\begin{aligned}
 \hat{\beta} &= \frac{\sum_{i=1,n} x_i y_i - (\sum_{i=1,n} x_i)(\sum_{i=1,n} y_i) / n}{\sum_{i=1,n} x_i^2 - (\sum_{i=1,n} x_i)^2 / n} \\
 &= \frac{3487 - (420 \times 56.2) / 7}{25900 - (420 \times 420) / 7} \\
 &= 0.1643,
 \end{aligned}$$

$$\begin{aligned}
 \hat{\alpha} &= \left(\frac{1}{n} \sum_{i=1,n} y_i \right) - \hat{\beta} \left(\frac{1}{n} \sum_{i=1,n} x_i \right) \\
 &= \frac{56.2}{7} - 0.1643 \times \frac{420}{7} \\
 &= -1.8294
 \end{aligned}$$

表2.3 回帰値と残差

変数	y	x	回帰値	残差	y ²	回帰値 ²	残差 ²
	5.3	45	5.56	-0.26	28.09	30.96	0.07
	7.5	50	6.39	1.11	56.25	40.78	1.24
	5.9	55	7.21	-1.31	34.81	51.94	1.71
	9.2	60	8.03	1.17	84.64	64.46	1.37
	8.8	65	8.85	-0.05	77.44	78.32	0.00
	7.5	70	9.67	-2.17	56.25	93.54	4.72
	12	75	10.49	1.51	144	110.10	2.27
和	56.2	420	56.2	0	481.48	470.10	11.38
変動和					30.27	18.89	11.38

$$\hat{y}_1 = -1.8294 + 0.1643 \times 45 = 5.564$$

$$\hat{u}_1 = 5.3 - 5.564 = -0.264$$

$$RSS = 481.48 - (-1.8294) \times 56.2 - (0.1643) \times 3487 = 11.38$$

$$s^2 = \frac{11.38}{5} = 2.28$$

総変動 観測値と平均の差の二乗和

$$\begin{aligned} TSS &= \sum_{i=1}^n (y_i - \bar{y})^2 \\ &= \sum_{i=1}^n (y_i - \bar{y}) y_i \\ &= \sum_{i=1}^n y_i^2 - n\bar{y}^2 \end{aligned} \tag{10}$$

決定係数 推定で得られた式の当てはまりの良さを計る尺度。1に近いほど当てはまりが良い。

$$R^2 = 1 - \frac{RSS}{TSS} \quad (11)$$

回帰変動 回帰値と平均の差の二乗和

$$\begin{aligned} ESS &= \sum_{i=1}^n (\hat{y}_i - \bar{y})^2 = \sum_{i=1}^n (\hat{y}_i - \bar{y}) \hat{y}_i \\ &= \sum_{i=1}^n \hat{y}_i^2 - \bar{y} \sum_{i=1}^n \hat{y}_i = \sum_{i=1}^n \hat{y}_i^2 - n\bar{y}^2 \end{aligned}$$

$$RSS = \sum_{i=1}^n y_i^2 - \sum_{i=1}^n \hat{y}_i^2 \quad (12)$$

$$TSS = ESS + RSS \quad (13)$$

$$TSS \geq RSS \quad (14)$$

$$R^2 = \frac{ESS}{TSS} \quad (15)$$

表2.3 回帰値と残差

変数	y	x	回帰値	残差	y ²	回帰値 ²	残差 ²
	5.3	45	5.56	-0.26	28.09	30.96	0.07
	7.5	50	6.39	1.11	56.25	40.78	1.24
	5.9	55	7.21	-1.31	34.81	51.94	1.71
	9.2	60	8.03	1.17	84.64	64.46	1.37
	8.8	65	8.85	-0.05	77.44	78.32	0.00
	7.5	70	9.67	-2.17	56.25	93.54	4.72
	12	75	10.49	1.51	144	110.10	2.27
和	56.2	420	56.2	0	481.48	470.10	11.38
変動和					30.27	18.89	11.38

$$TSS = \sum_{i=1}^n y_i^2 - n\bar{y}^2 = 481.48 - \frac{1}{7}56.2 \times 56.2$$

$$ESS = \sum_{i=1}^n \hat{y}_i^2 - n\bar{y}^2 = 470.10 - \frac{1}{7}56.2 \times 56.2$$

$$R^2 = \frac{ESS}{TSS} = 0.62$$