第 III 部　線型モデル

6　多重共線性

ポイント

- 多重共線性
- 独立変数間の相関
- モデルの仮定

6.1　はじめに

- 全ての話には前提があります。
- 前提を満たさないと結果はおかしくなります。
- 前提に関しては、当然満たされるはずですが、満たしていない場合もあります。
- 前提を満たしていな状況がもたらすことを紹介します。

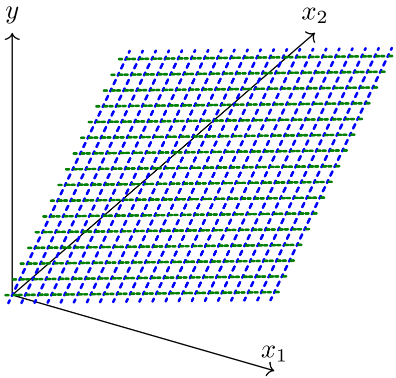
6.2　2 変数の線型モデル

- 変数の組を識別する添え字を i とし、(y_i, x_{i1}, x_{i2}) とあらわす。
- y_i を従属変数とし、x_{i1}, x_{i2} を独立変数とする。
- y, x_1, x_2 は連続量であって、x_1 と x_2 は無相関であるとする。
- 独立変数が複数ある線型モデルを重回帰モデル[1)]とよび、2 変数の重回帰モデルを以下のように定義する。

$$\hat{y}_i = \beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} \tag{1}$$

1）　重回帰モデルと比較するとき、1 変量の線型モデルを単回帰とよぶ。

図 1　$\hat{y}_i = \beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2}$の平面



6.2.1　重回帰モデルにおける誤差の定義

誤差 (e_i) は観測点 (y_i) と平面上の点 $(\hat{y}_i)$ の y 軸上の差として定義する。

$$\begin{aligned} e_i &= y_i - \hat{y}_i & (2) \\ &= y_i - (\beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2}) & (3) \\ &= y_i - \beta_0 - \beta_1 x_{i1} - \beta_2 x_{i2} & (4) \end{aligned}$$

6.2.2　誤差の二乗和

$$\sum_{i=1}^{n} {e_i}^2 = \sum_{i=1}^{n} (y_i - \beta_0 - \beta_1 x_{i1} - \beta_2 x_{i2})^2 \tag{5}$$

6.2.3　偏導関数

誤差の二乗和は $\beta_0, \beta_1, \beta_2$ の関数なので、

$$f(\beta_0, \beta_1, \beta_2) = \sum_{i=1}^{n} (y_i - \beta_0 - \beta_1 x_{i1} - \beta_2 x_{i2})^2 \tag{6}$$

偏微分すると 3 つの偏導関数を得る。

$$\frac{\partial f}{\partial \beta_0} = -2 \sum_{i=1}^{n} (y_i - \beta_0 - \beta_1 x_{i1} - \beta_2 x_{i2}) \tag{7}$$

$$\frac{\partial f}{\partial \beta_1} = -2 \sum_{i=1}^{n} x_{i1} (y_i - \beta_0 - \beta_1 x_{i1} - \beta_2 x_{i2}) \tag{8}$$

$$\frac{\partial f}{\partial \beta_2} = -2 \sum_{i=1}^{n} x_{i2} (y_i - \beta_0 - \beta_1 x_{i1} - \beta_2 x_{i2}) \tag{9}$$

6.2.4　最小二乗解の導出

誤差の二乗和は非負なので下限を持つ。したがって極値が最小値を取る。極値を求めるためには偏導関数をそれぞれ 0 と置いた連立方程式を解けばよい。

$$
\begin{cases}
\displaystyle\sum_{i=1}^{n} (y_i - \beta_0 - \beta_1 x_{i1} - \beta_2 x_{i2}) = 0 \\
\displaystyle\sum_{i=1}^{n} x_{i1} (y_i - \beta_0 - \beta_1 x_{i1} - \beta_2 x_{i2}) = 0 \\
\displaystyle\sum_{i=1}^{n} x_{i2} (y_i - \beta_0 - \beta_1 x_{i1} - \beta_2 x_{i2}) = 0
\end{cases}
\tag{10}
$$

6.2.5　β_j の推定値

(10) の解を最小二乗解または最小二乗推定値とよび、$\hat{\beta}_j$ であらわす。2 変量の重回帰モデルの最小二乗推定値 $\hat{\beta}_1, \hat{\beta}_2$ は x_1, x_2 の分散、x_1, x_2, y の共分散を使うと、

$$\hat{\beta}_1 = \frac{Cov_{x_1y}V_{x_2} - Cov_{x_2y}Cov_{x_1x_2}}{V_{x_1}V_{x_2} - {Cov_{x_1x_2}}^2} \tag{11}$$

$$\hat{\beta}_2 = \frac{Cov_{x_2y}V_{x_1} - Cov_{x_1y}Cov_{x_1x_2}}{V_{x_1}V_{x_2} - {Cov_{x_1x_2}}^2} \tag{12}$$

であって、$\hat{\beta}_1, \hat{\beta}_2$ を使うと

$$\hat{\beta}_0 = \bar{y} - \hat{\beta}_1\bar{x}_1 - \hat{\beta}_2\bar{x}_2 \tag{13}$$

である。

6.3　独立変数が 3 つの重回帰モデル

独立変数が 3 つある重回帰モデルは、

$$\hat{y}_i = \beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} + \beta_3 x_{i3} \tag{14}$$

である。そして誤差の二乗和は

$$\sum_{i=1}^{n} {e_i}^2 = \sum_{i=1}^{n} (y_i - \hat{y}_i)^2 \tag{15}$$

$$= \sum_{i=1}^{n} (y_i - \beta_0 - \beta_1 x_{i1} - \beta_2 x_{i2} - \beta_3 x_{i3})^2 \tag{16}$$

である。

そして (16) 式の偏導関数を 0 と置いた連立方程式は

$$
\begin{cases}
\displaystyle\sum_{i=1}^{n} (y_i - \beta_0 - \beta_1 x_{i1} - \beta_2 x_{i2} - \beta_3 x_{i3}) = 0 \\
\displaystyle\sum_{i=1}^{n} x_{i1} (y_i - \beta_0 - \beta_1 x_{i1} - \beta_2 x_{i2} - \beta_3 x_{i3}) = 0 \\
\displaystyle\sum_{i=1}^{n} x_{i2} (y_i - \beta_0 - \beta_1 x_{i1} - \beta_2 x_{i2} - \beta_3 x_{i3}) = 0 \\
\displaystyle\sum_{i=1}^{n} x_{i3} (y_i - \beta_0 - \beta_1 x_{i1} - \beta_2 x_{i2} - \beta_3 x_{i3}) = 0
\end{cases}
\tag{17}
$$

である。(17) 式の解が独立変数が 3 つのときの最小二乗推定値である。

和の総和を総和の和であらわし、

$$n\beta_0 + \beta_1 \sum_{i=1}^{n} x_{i1} + \beta_2 \sum_{i=1}^{n} x_{i2} + \beta_3 \sum_{i=1}^{n} x_{i3} = \sum_{i=1}^{n} y_i \tag{18}$$

$$\beta_0 \sum_{i=1}^{n} x_{i1} + \beta_1 \sum_{i=1}^{n} {x_{i1}}^2 + \beta_2 \sum_{i=1}^{n} x_{i1}x_{i2} + \beta_3 \sum_{i=1}^{n} x_{i1}x_{i3} = \sum_{i=1}^{n} x_{i1}y_i \tag{19}$$

$$\beta_0 \sum_{i=1}^{n} x_{i2} + \beta_1 \sum_{i=1}^{n} x_{i1}x_{i2} + \beta_2 \sum_{i=1}^{n} {x_{i2}}^2 + \beta_3 \sum_{i=1}^{n} x_{i2}x_{i3} = \sum_{i=1}^{n} x_{i2}y_i \tag{20}$$

$$\beta_0 \sum_{i=1}^{n} x_{i3} + \beta_1 \sum_{i=1}^{n} x_{i1}x_{i3} + \beta_2 \sum_{i=1}^{n} x_{i2}x_{i3} + \beta_3 \sum_{i=1}^{n} {x_{i3}}^2 = \sum_{i=1}^{n} x_{i3}y_i \tag{21}$$

(18) を用いて各式から β_0 を打ち消し、x_j の分散を V_j, x_j と x_k の共分散を C_{jk}, x_j と y の共分散を C_{jy} とあらわすと、

$$\left\{ \begin{aligned} \beta_1 V_1 + \beta_2 C_{12} + \beta_3 C_{13} &= C_{1y} \\ \beta_1 C_{12} + \beta_2 V_2 + \beta_3 C_{23} &= C_{2y} \\ \beta_1 C_{13} + \beta_2 C_{23} + \beta_3 V_3 &= C_{3y} \end{aligned} \right. \tag{22}$$

これを解くと最小二乗推定値を得る。

(22) を解いた最小二乗推定値は、

$$\hat{\beta}_1 = \frac{V_2V_3C_{1y} - V_2C_{13}C_{3y} - V_3C_{12}C_{2y} + C_{12}C_{23}C_{3y} + C_{13}C_{23}C_{2y} - {C_{23}}^2C_{1y}}{V_1V_2V_3 - V_1{C_{23}}^2 - V_2{C_{13}}^2 - V_3{C_{12}}^2 + 2C_{12}C_{13}C_{23}} \tag{23}$$

$$\hat{\beta}_2 = \frac{V_1V_3C_{2y} - V_1C_{23}C_{3y} - V_3C_{12}C_{1y} + C_{12}C_{13}C_{3y} - {C_{13}}^2C_{2y} + C_{13}C_{23}C_{1y}}{V_1V_2V_3 - V_1{C_{23}}^2 - V_2{C_{13}}^2 - V_3{C_{12}}^2 + 2C_{12}C_{13}C_{23}} \tag{24}$$

$$\hat{\beta}_3 = \frac{V_1V_2C_{3y} - V_1C_{23}C_{2y} - V_2C_{13}C_{1y} - {C_{12}}^2C_{3y} + C_{12}C_{23}C_{1y} + C_{12}C_{13}C_{2y}}{V_1V_2V_3 - V_1{C_{23}}^2 - V_2{C_{13}}^2 - V_3{C_{12}}^2 + 2C_{12}C_{13}C_{23}} \tag{25}$$

であって、$\hat{\beta}_j$ を使うと、

$$\hat{\beta}_0 = \bar{y} - \hat{\beta}_1\bar{x}_1 - \hat{\beta}_2\bar{x}_2 - \hat{\beta}_3\bar{x}_3 \tag{26}$$

である。

6.4　無相関の仮定が保たれない場合

重回帰モデルは独立変数間に無相関を仮定している。この仮定が保たれない場合を考える。

6.4.1　重複する独立変数

2 変量の重回帰モデルに対して同一の独立変数を 2 回組み込んでみる。

$$y_i = \beta_0 + \beta_1 x_i + \beta_2 x_i \tag{27}$$

$$= \beta_0 + (\beta_1 + \beta_2)\, x_i \tag{28}$$

(27) は単回帰モデルに帰着し、$\hat{\beta}_0$, $\hat{\beta}_1$, $\hat{\beta}_2$ を一意に求めることはできない。

6.4.2 最小二乗推定値を使った説明

$x_{i1} = x_{i2} = x_i$ のとき、(11) は

$$\hat{\beta}_1 = \frac{Cov_{xy}V_x - Cov_{xy}Cov_{xx}}{V_xV_x - {Cov_{xx}}^2} \tag{29}$$

$$= \frac{Cov_{xy}V_x - Cov_{xy}V_x}{{V_x}^2 - {V_x}^2} \tag{30}$$

となり、分母分子ともに 0 となってしまうため、$\hat{\beta}_1$ を推定することはできない。$\hat{\beta}_2$ も同様である。

6.4.3　過剰なダミー変数

男性なら 1 をとり、それ以外は 0 をとる男性ダミー d_{i1} と女性ならを 1 をとり、それ以外なら 0 をとる女性ダミー d_{i2} を同時にモデルに組み込むことを考える。y_i を従属変数 x_{i1}, d_{i1}, d_{i2} を独立変数として定式化する。ここで、x_{i1} は連続量とする。

$$y_i = \beta_0 + \beta_1 x_{i1} + \beta_2 d_{i1} + \beta_3 d_{i2} \tag{31}$$

男性だった場合

$$d_{i1} = 1 \quad \text{かつ} \quad d_{i2} = 0 \tag{32}$$

であり、女性だった場合は

$$d_{i1} = 0 \quad \text{かつ} \quad d_{i2} = 1 \tag{33}$$

d_{i1} と d_{i2} の間には以下の関係が成り立つ。

$$d_{i2} = 1 - d_{i1} \tag{34}$$

(34) を (31) に代入すると

$$\begin{aligned} y_i &= \beta_0 + \beta_1 x_{i1} + \beta_2 d_{i1} + \beta_3 \left(1 - d_{i1}\right) && (35) \\ &= \beta_0 + \beta_1 x_{i1} + \beta_2 d_{i1} + \beta_3 - \beta_3 d_{i1} && (36) \\ &= \left(\beta_0 + \beta_3\right) + \beta_1 x_{i1} + \left(\beta_2 - \beta_3\right) d_{i1} && (37) \end{aligned}$$

となり、(31) は 2 変数の重回帰モデルに帰着する。したがって $\hat{\beta}_0$, $\hat{\beta}_1$, $\hat{\beta}_2$, $\hat{\beta}_3$ を一意に求めることはできない。

6.5　$\boldsymbol{d_{i1}}$ と $\boldsymbol{d_{i2}}$ の関係

(31) における d_{i1} と d_{i2} の関係を考察する。d_{i1} は男性ダミーであり、d_{i2} は女性ダミーである。

ここで、$\sum_{i=1}^{n} d_{i1} = n_1,\ n > n_1 > 0$ とする。

6.5.1　$\boldsymbol{d_1}$ と $\boldsymbol{d_2}$ の平均

d_{i1} と d_{i2} の平均を求める。

$$\bar{d}_1 = \frac{1}{n}\sum_{i=1}^{n} d_{i1} = \frac{n_1}{n} \tag{38}$$

$$\bar{d}_2 = \frac{1}{n}\sum_{i=1}^{n} d_{i2} = \frac{1}{n}\sum_{i=1}^{n}(1 - d_{i1}) \tag{39}$$

$$= \frac{n - n_1}{n} \tag{40}$$

6.5.2 $\boldsymbol{d_1}$ の分散

d_1 の分散を求める。

$$V_{d_1} = \frac{1}{n}\sum_{i=1}^{n}\left(d_{i1} - \bar{d}_1\right)^2 \tag{41}$$

$$= \frac{1}{n}\sum_{i=1}^{n}{d_{i1}}^2 - {\bar{d}_1}^2 \tag{42}$$

$d_{i1} = 1 \Longrightarrow {d_{i1}}^2 = 1$,

$d_{i1} = 0 \Longrightarrow {d_{i1}}^2 = 0$ なので

$$\sum_{i=1}^{n}{d_{i1}}^2 = \sum_{i=1}^{n} d_{i1} \tag{43}$$

したがって

$$V_{d_1} = \frac{1}{n}n_1 - \left(\frac{n_1}{n}\right)^2 \tag{44}$$

$$= \frac{n_1}{n} - \frac{{n_1}^2}{n^2} \tag{45}$$

$$= \frac{nn_1 - {n_1}^2}{n^2} \tag{46}$$

$$= \frac{n_1(n - n_1)}{n^2} \tag{47}$$

問題 III–6 – 1

d_1, d_2 はともにダミー変数であって $d_{i2} = 1 - d_{i1}$ の関係にある n 組の数とする。$\displaystyle\sum_{i=1}^{n} d_{i1} = n_1,\ n > n_1 > 0$ のとき d_2 の分散を求めなさい。

解例 III–6 − 1

$$V_{d_2} = \frac{1}{n}\sum_{i=1}^{n}\left(d_{i2} - \bar{d_2}\right)^2$$
$$= \frac{1}{n}\sum_{i=1}^{n}{d_{i2}}^2 - \bar{d_2}^2$$

$$\bar{d_2} = \frac{n-n_1}{n}$$
$$\sum_{i=1}^{n}{d_{i2}}^2 = \sum_{i=1}^{n}d_{i2}$$
$$\sum_{i=1}^{n}d_{i2} = \sum_{i=1}^{n}(1-d_{i1}) = n - n_1$$

なので

$$V_{d_2} = \frac{1}{n}(n-n_1) - \left(\frac{n-n_1}{n}\right)^2$$
$$= \frac{n_1(n-n_1)}{n^2}$$

6.5.3 $\boldsymbol{d_1}$ と $\boldsymbol{d_2}$ の共分散

d_1 と d_2 の共分散を求める。

$$
\begin{aligned}
Cov_{d_1 d_2} &= \frac{1}{n}\sum_{i=1}^{n}\left(d_{i1}-\bar{d_1}\right)\left(d_{i2}-\bar{d_2}\right) && (48)\\
&= \frac{1}{n}\sum_{i=1}^{n} d_{i1}d_{i2}-\bar{d_1}\bar{d_2} && (49)
\end{aligned}
$$

$d_{i1}d_{i2}=0$ なので第一項は 0 だから

$$
\begin{aligned}
&= -\frac{n_1}{n}\frac{n-n_1}{n} && (50)\\
&= -\frac{n_1(n-n_1)}{n^2} && (51)
\end{aligned}
$$

6.5.4 $\boldsymbol{d_1}$ と $\boldsymbol{d_2}$ の相関係数

d_1 と d_2 の相関係数を求める。

$$\sigma_{d_1} = \sqrt{V_{d_1}} = \sqrt{\frac{n_1(n-n_1)}{n^2}} \tag{52}$$

$$\sigma_{d_2} = \sqrt{V_{d_2}} = \sqrt{\frac{n_1(n-n_1)}{n^2}} \tag{53}$$

$$Cov_{d_1 d_2} = -\frac{n_1(n-n_1)}{n^2} \tag{54}$$

$$\rho_{d_1 d_2} = \frac{Cov_{d_1 d_2}}{\sigma_{d_1}\sigma_{d_2}} \tag{55}$$

$$= \frac{\left(-\frac{n_1(n-n_1)}{n^2}\right)}{\left(\sqrt{\frac{n_1(n-n_1)}{n^2}}\sqrt{\frac{n_1(n-n_1)}{n^2}}\right)} \tag{56}$$

$$= \frac{\left(-\frac{n_1(n-n_1)}{n^2}\right)}{\left(\frac{n_1(n-n_1)}{n^2}\right)} \tag{57}$$

$$= -\frac{n_1(n-n_1)}{n^2}\frac{n^2}{n_1(n-n_1)} \tag{58}$$

$$= -1 \tag{59}$$

6.5.5 $\boldsymbol{x_1}$ と $\boldsymbol{d_1}$ の共分散

x_1 と d_1 の共分散を求める。

$$Cov_{x_1 d_1} = \frac{1}{n}\sum_{i=1}^{n}(x_{i1} - \bar{x_1})\left(d_{i1} - \bar{d_1}\right) \tag{60}$$

$$= \frac{1}{n}\sum_{i=1}^{n} x_{i1} d_{i1} - \bar{x}_1 \bar{d}_1 \tag{61}$$

6.5.6 $\boldsymbol{x_1}$ と $\boldsymbol{d_2}$ の共分散

x_1 と d_2 の共分散を求める。

$$Cov_{x_1 d_2} \tag{62}$$

$$= \frac{1}{n}\sum_{i=1}^{n}(x_{i1} - \bar{x_1})\left(d_{i2} - \bar{d_2}\right) \tag{63}$$

$$= \frac{1}{n}\sum_{i=1}^{n} x_{i1}d_{i2} - \bar{x}_1\bar{d}_2 \tag{64}$$

$$= \frac{1}{n}\sum_{i=1}^{n} x_{i1}(1 - d_{i1}) - \bar{x}_1(1 - \bar{d}_1) \tag{65}$$

$$= \frac{1}{n}\sum_{i=1}^{n}(x_{i1} - x_{i1}d_{i1}) - (\bar{x}_1 - \bar{x}_1\bar{d}_1) \tag{66}$$

$$= \frac{1}{n}\sum_{i=1}^{n} x_{i1} - \frac{1}{n}\sum_{i=1}^{n} x_{i1}d_{i1} - \bar{x}_1 + \bar{x}_1\bar{d}_1 \tag{67}$$

$$= \bar{x}_1 - \frac{1}{n}\sum_{i=1}^{n} x_{i1}d_{i1} - \bar{x}_1 + \bar{x}_1\bar{d}_1 \tag{68}$$

$$= -\frac{1}{n}\sum_{i=1}^{n} x_{i1}d_{i1} + \bar{x}_1\bar{d}_1 \tag{69}$$

$$= -Cov_{x_1 d_1} \tag{70}$$

これまでの結果をまとめると

$$V_{d_1} = V_{d_2} = -Cov_{d_1 d_2} \tag{71}$$

$$Cov_{x_1 d_1} = -Cov_{x_1 d_2} \tag{72}$$

である。

ここで、$x_{2i} = d_{i1}$, $x_{3i} = d_{i2}$ とし、(23)(24)(25) の表記に合わせ、さらに、それぞれを U, W であらわす。

$$U = V_2 = V_3 = -C_{23} \tag{73}$$

$$W = C_{12} = -C_{13} \tag{74}$$

$U,\ W$ を使い (23) を記述すると、

$$\hat{\beta}_1 = \frac{V_2V_3C_{1y} - V_2C_{13}C_{3y} - V_3C_{12}C_{2y} + C_{12}C_{23}C_{3y} + C_{13}C_{23}C_{2y} - {C_{23}}^2C_{1y}}{V_1V_2V_3 - V_1{C_{23}}^2 - V_2{C_{13}}^2 - V_3{C_{12}}^2 + 2C_{12}C_{13}C_{23}} \tag{23}$$

$$= \frac{U^2C_{1y} + UWC_{3y} - UWC_{2y} - UWC_{3y} + UWC_{2y} - U^2C_{1y}}{V_1U^2 - V_1U^2 - UW^2 - UW^2 + 2UW^2} \tag{75}$$

であり、分母分子ともに 0 となることがわかる。$\hat{\beta}_2,\ \hat{\beta}_3$ も同様である。

男性ダミーと女性ダミーの相関係数は –1 である。このような過剰なダミー変数をモデルに組み込んだ場合、推定値を得ることはできない。

ここでは、2 分類される対象に対して 2 つのダミー変数をモデルに組み込んだ例を示したが、7 つの曜日に対して曜日ダミーを 7 個組み込む場合や、12 カ月の月に対して月ダミーを 12 個組み込む場合も過剰なダミー変数であって、同様に推定値を得ることはできない。

6.6　多重共線性

独立変数を

$$x_{i2} = a + bx_{i1} \tag{76}$$

の関係にある x_{i1},x_{i2} とする。ただし、$b \neq 0$ とする。すると x_{i1} と x_{i2} の相関係数は ± 1 である。この x_{i2} を用いると (1) は

$$y_i = \beta_0 + \beta_1 x_{i1} + \beta_2 (a + bx_{i1}) \tag{77}$$

$$= (\beta_0 + a\beta_2) + (\beta_1 + b\beta_2) x_{i1} \tag{78}$$

となり、線型モデルに帰着する。したがって $\hat{\beta}_0$, $\hat{\beta}_1$, $\hat{\beta}_2$ を一意に求めることはできない。

6.6.1　最小二乗推定値を用いた説明

$x_{i2} = a + bx_{i1}$ だから

$$\bar{x}_2 = a + b\bar{x}_1, \quad V_{x_2} = b^2 V_{x_1}, \quad Cov_{x_1 x_2} = bV_{x_1} \qquad \text{であって} \tag{79}$$

$$Cov_{x_2 y} = bCov_{x_1 y} \qquad \text{なので} \tag{80}$$

β_1 の推定値 (11) は

$$\hat{\beta}_1 = \frac{b^2 V_{x_1} Cov_{x_1 y} - b^2 V_{x_1} Cov_{x_1 y}}{b^2 {V_{x_1}}^2 - (bV_{x_1})^2} \tag{81}$$

となり、分母分子ともに 0 となってしまうので $\hat{\beta}_1$ を求めることができない。$\hat{\beta}_2$ に関しても同様である。

6.6.2　多重共線性

独立変数間の相関が強い場合、確率的に (81) に近づいてしまい、以下のような現象が起こる。この性質を多重共線性という。

- 少数のデータの追加・削除により推定値が大きく変化する。
- 異なるデータに適用すると推定値が大きく異なる。
- 推定値の符号が先験的な知識と逆になる。
- 決定係数は高いが、推定値が非有意となる。
- 推定値を求めることができない。

問題 III–6 – 2

ある小学校の児童から以下のデータを得た。x_1 は身長、x_2 は体重、y は 50m 走のタイムである。表の値を使い $\hat{\beta}_0, \hat{\beta}_1, \hat{\beta}_2$ を求めなさい。

i	x_1	x_2	y
1	116.5	21.3	11.6
2	122.6	24.2	10.7
3	128.5	27.7	10.0
4	133.5	31.1	9.7
5	138.6	34.4	9.4
6	145.0	39.2	9.0

i	x_1	x_2	y
7	116.2	21.2	11.8
8	121.7	23.7	10.9
9	127.6	27.0	10.4
10	133.3	30.4	10.0
11	140.4	35.0	9.6
12	146.7	40.0	9.3

解例 III–6 – 2

高機能のアプリでは推定できないと思います。素朴なアプリを利用してください。

$$\hat{\beta}_0 = 33.89156\cdots$$

$$\hat{\beta}_1 = -0.23734\cdots$$

$$\hat{\beta}_2 = 0.24908\cdots$$

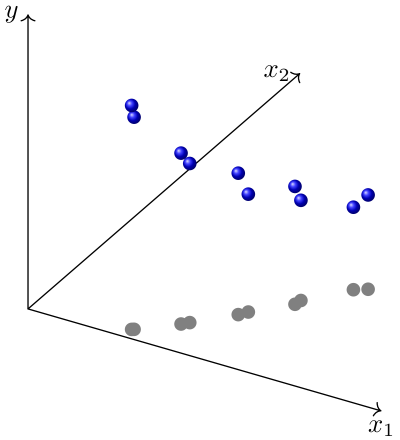
$$\hat{y} = 33.89156 - 0.23734 \times x_1 + 0.24908 \times x_2$$

単回帰による推定値

$$\hat{y} = 20.98577 - 0.08240 \times x_1$$

$$\hat{y} = 14.04194 - 0.12979 \times x_2$$

図 2　相関の強い独立変数の具体例



6.6.3 多重共線性への対応

- 多重共線性は独立変数間に強い相関がある場合に発生する。
- 多重共線性への対処は相関の強い独立変数のうちの一つをモデルから除けばよい。
- 問題は、「どの変数を除くか」である。
- マーケティングでは、より興味を持っている変数を残す場合が多い。

6.7 まとめ

- 独立変数間の相関が強い場合は多重共線性が発生する。
- 極端な例では独立変数間の相関係数が ± 1 の場合、推定値を得ることができない。
- マーケティング変数は相関が高い場合が多いので注意が必要である。
- マーケティングではどの変数を残すのかをかなり恣意的に決定する。
- どの変数を使って分析するかは恣意的であるが、その結果は客観的である。