第 III 部　線型モデル

5　線型モデルの多変数化

ポイント

- 重回帰モデル
- 質的データの分析

5.1　はじめに

- grape は単数形で、grapes は複数形です。
- data は複数形で、datum は単数形です。
- 英語は単数と複数で単語が異なります。
- わざわざ言い換えているのです。
- 特別な意味があります。

5.1.1 総和

定義

n 個の数値があり、x_i を i 番目の値とする。全ての x_i を足し合わせた値を『総和』と呼び、$\sum_{i=1}^{n} x_i$ とあらわす。

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

5.1.2 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。

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

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

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

総和の公式より以下の関係を容易に導くことができる。a, c は定数とする。

$$\sum_{i=1}^{n}(ax_i + c) = a\sum_{i=1}^{n}x_i + nc \tag{5}$$

$$\sum_{i=1}^{n}(x_i - y_i) = \sum_{i=1}^{n}x_i - \sum_{i=1}^{n}y_i \tag{6}$$

$$\sum_{i=1}^{n}(x_i + y_i + z_i) = \sum_{i=1}^{n}x_i + \sum_{i=1}^{n}y_i + \sum_{i=1}^{n}z_i \tag{7}$$

5.1.3　総和と平均

総和を n で割った値を平均と呼び $\bar{x}$ であらわす。

$$\bar{x} = \frac{1}{n}\sum_{i=1}^{n}x_i \tag{8}$$

平均を n 倍した値は総和である。

$$n\bar{x} = \sum_{i=1}^{n}x_i \tag{9}$$

問題 III–5 – 1

次の式を展開しなさい。

$$\frac{1}{n}\sum_{i=1}^{n}\left(u_i-\bar{u}\right)^2$$

問題 III–5 – 2

次の式を展開しなさい。

$$\frac{1}{n}\sum_{i=1}^{n}\left(v_i-\bar{v}\right)\left(w_i-\bar{w}\right)$$

解例 III–5 − 1

$$
\begin{aligned}
\frac{1}{n}\sum_{i=1}^{n}(u_i-\bar{u})^2 &= \frac{1}{n}\sum_{i=1}^{n}\left({u_i}^2-2\bar{u}u_i+\bar{u}^2\right)\\
&= \frac{1}{n}\sum_{i=1}^{n}{u_i}^2-2\bar{u}\frac{1}{n}\sum_{i=1}^{n}u_i+\frac{1}{n}n\bar{u}^2\\
&= \frac{1}{n}\sum_{i=1}^{n}{u_i}^2-2\frac{1}{n}n\bar{u}\bar{u}+\bar{u}^2\\
&= \frac{1}{n}\sum_{i=1}^{n}{u_i}^2-\bar{u}^2
\end{aligned}
$$

解例 III–5 − 2

$$
\begin{aligned}
\frac{1}{n}\sum_{i=1}^{n}(v_i-\bar{v})(w_i-\bar{w}) &= \frac{1}{n}\sum_{i=1}^{n}(v_iw_i-\bar{v}w_i-\bar{w}v_i+\bar{v}\bar{w}) \\
&= \frac{1}{n}\sum_{i=1}^{n}v_iw_i-\bar{v}\frac{1}{n}\sum_{i=1}^{n}w_i-\bar{w}\frac{1}{n}\sum_{i=1}^{n}v_i+\frac{1}{n}n\bar{v}\bar{w} \\
&= \frac{1}{n}\sum_{i=1}^{n}v_iw_i-\bar{v}\bar{w}-\bar{v}\bar{w}+\bar{v}\bar{w} \\
&= \frac{1}{n}\sum_{i=1}^{n}v_iw_i-\bar{v}\bar{w}
\end{aligned}
$$

5.2 分散と共分散

分散並びに共分散の定義を確認する。

$$V_u = \frac{1}{n}\sum_{i=1}^{n}(u_i - \bar{u})^2 \tag{10}$$

$$= \frac{1}{n}\sum_{i=1}^{n}{u_i}^2 - \bar{u}^2 \tag{11}$$

$$Cov_{vw} = \frac{1}{n}\sum_{i=1}^{n}(v_i - \bar{v})(w_i - \bar{w}) \tag{12}$$

$$= \frac{1}{n}\sum_{i=1}^{n}v_i w_i - \bar{v}\bar{w} \tag{13}$$

両辺に n^2 を掛けると、

$$n^2 V_u = n\sum_{i=1}^{n}{u_i}^2 - n\bar{u}n\bar{u} \tag{14}$$

$$= n\sum_{i=1}^{n}{u_i}^2 - \left(\sum_{i=1}^{n}u_i\right)^2 \tag{15}$$

$$n^2 Cov_{vw} = n\sum_{i=1}^{n}v_i w_i - n\bar{v}n\bar{w} \tag{16}$$

$$= n\sum_{i=1}^{n}v_i w_i - \sum_{i=1}^{n}v_i\sum_{i=1}^{n}w_i \tag{17}$$

である。

5.3　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とよび、2 変量の重回帰モデルを以下のように定義する。

$$\hat{y}_i = \beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} \tag{18}$$

重回帰モデルと対比するとき、1 変量の線型モデルを単回帰モデルと呼ぶ。

図 1　3 次元の観測点

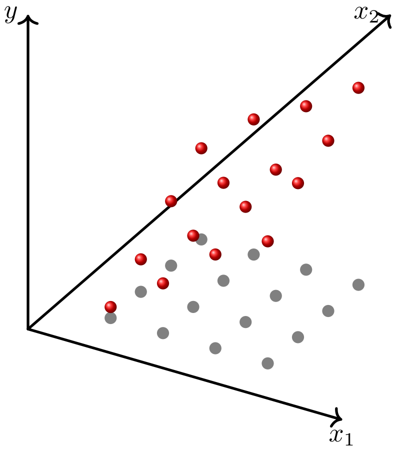


図 2　平面の当てはめ

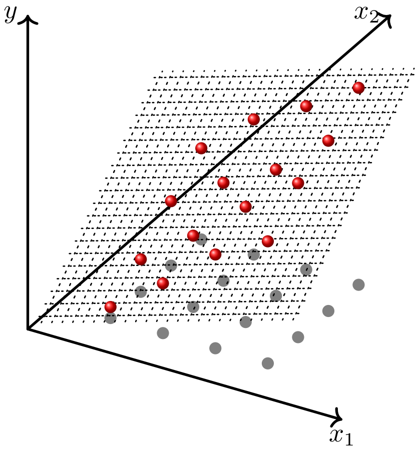


表 1　x_2の違いによる y の差

$x_{i2}=0$	$\hat{y}_i=\beta_0+\beta_1 x_{i1}$
$x_{i2}=1$	$\hat{y}_i=\beta_0+\beta_1 x_{i1}+\beta_2$
$x_{i2}=2$	$\hat{y}_i=\beta_0+\beta_1 x_{i1}+2\beta_2$
$x_{i2}=3$	$\hat{y}_i=\beta_0+\beta_1 x_{i1}+3\beta_2$

図 3　$\boldsymbol{x_2}$ の違いによる $\boldsymbol{\hat{y}_i}$

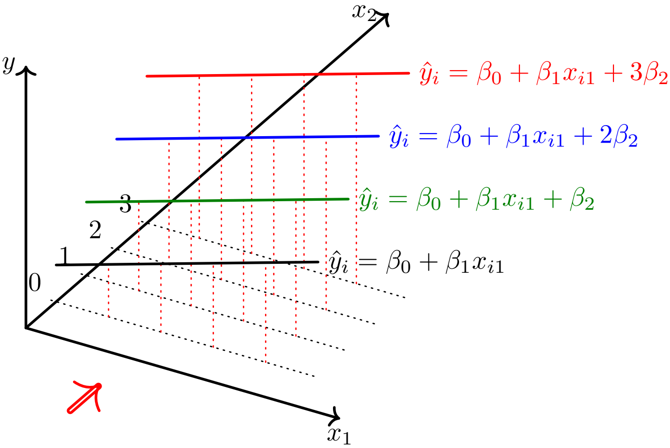
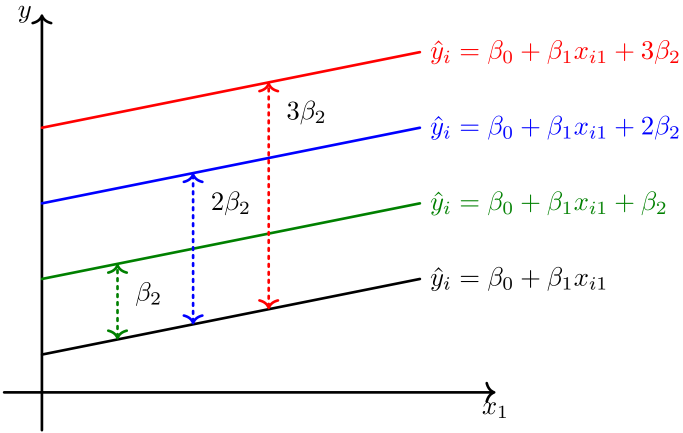


図 4　$x_1 \times y$ 平面上の β_2 の意味



5.4　パラメータの推定手順

1. 誤差の二乗和を求める。
2. 偏導関数を求める。
3. 偏導関数を 0 とする連立方程式を立てる。
4. 連立方程式を解く。

- 得られた解を最小二乗解または最小二乗推定値とよぶ。
- 重回帰モデルは単回帰モデルとまったく同じ手順でパラメータを推定することができる。
- 特に難しい点は無く、煩雑なだけ。

5.4.1 重回帰モデルにおける誤差

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。

$$
\begin{aligned}
e_i &= y_i - \hat{y}_i && (19)\\
&= y_i - (\beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2}) && (20)\\
&= y_i - \beta_0 - \beta_1 x_{i1} - \beta_2 x_{i2} && (21)
\end{aligned}
$$

誤差の二乗和

$$
\sum_{i=1}^{n} {e_i}^2 = \sum_{i=1}^{n} (y_i - \beta_0 - \beta_1 x_{i1} - \beta_2 x_{i2})^2 \tag{22}
$$

5.4.2 偏導関数

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

$$f(\beta_0, \beta_1, \beta_2) = \sum_{i=1}^{n} (y_i - \beta_0 - \beta_1 x_{i1} - \beta_2 x_{i2})^2 \tag{23}$$

である。偏微分すると 3 つの偏導関数を得る。

$$\frac{\partial f}{\partial \beta_0} = -2 \sum_{i=1}^{n} (y_i - \beta_0 - \beta_1 x_{i1} - \beta_2 x_{i2}) \tag{24}$$

$$\frac{\partial f}{\partial \beta_1} = -2 \sum_{i=1}^{n} x_{i1} (y_i - \beta_0 - \beta_1 x_{i1} - \beta_2 x_{i2}) \tag{25}$$

$$\frac{\partial f}{\partial \beta_2} = -2 \sum_{i=1}^{n} x_{i2} (y_i - \beta_0 - \beta_1 x_{i1} - \beta_2 x_{i2}) \tag{26}$$

5.5　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。

5.5.1 偏導関数を 0 とおいた連立方程式

$$
\begin{cases}
\displaystyle\sum_{i=1}^{n} (y_i - \beta_0 - \beta_1 x_{i1} - \beta_2 x_{i2}) = 0 & (27) \\
\displaystyle\sum_{i=1}^{n} x_{i1} (y_i - \beta_0 - \beta_1 x_{i1} - \beta_2 x_{i2}) = 0 & (28) \\
\displaystyle\sum_{i=1}^{n} x_{i2} (y_i - \beta_0 - \beta_1 x_{i1} - \beta_2 x_{i2}) = 0 & (29)
\end{cases}
$$

(27) を整理

$$\sum_{i=1}^{n}(y_i - \beta_0 - \beta_1 x_{i1} - \beta_2 x_{i2}) = 0 \tag{27}$$

$$\sum_{i=1}^{n} y_i - n\beta_0 - \beta_1 \sum_{i=1}^{n} x_{i1} - \beta_2 \sum_{i=1}^{n} x_{i2} = 0 \tag{30}$$

$$n\beta_0 + \beta_1 \sum_{i=1}^{n} x_{i1} + \beta_2 \sum_{i=1}^{n} x_{i2} = \sum_{i=1}^{n} y_i \tag{31}$$

同様に整理した (28)(29) と β_0 の打ち消しプラン

(28)(29) を整理したものが (32)(33) である。

$$n\beta_0 + \beta_1 \sum_{i=1}^{n} x_{i1} + \beta_2 \sum_{i=1}^{n} x_{i2} = \sum_{i=1}^{n} y_i \tag{31}$$

$$\beta_0 \sum_{i=1}^{n} x_{i1} + \beta_1 \sum_{i=1}^{n} {x_{i1}}^2 + \beta_2 \sum_{i=1}^{n} x_{i1}x_{i2} = \sum_{i=1}^{n} x_{i1}y_i \tag{32}$$

$$\beta_0 \sum_{i=1}^{n} x_{i2} + \beta_1 \sum_{i=1}^{n} x_{i1}x_{i2} + \beta_2 \sum_{i=1}^{n} {x_{i2}}^2 = \sum_{i=1}^{n} x_{i2}y_i \tag{33}$$

(31) と (32) により β_0 を打ち消し、(31) と (33) により β_0 を打ち消す。

(32) と (31) による β_0 の打ち消し

(32) の両辺に n を掛ける

$$n\beta_0 \sum_{i=1}^{n} x_{i1} + n\beta_1 \sum_{i=1}^{n} {x_{i1}}^2 + n\beta_2 \sum_{i=1}^{n} x_{i1}x_{i2} = n \sum_{i=1}^{n} x_{i1}y_i \tag{34}$$

(31) の両辺に $\sum_{i=1}^{n} x_{i1}$ を掛ける

$$n\beta_0 \sum_{i=1}^{n} x_{i1} + \beta_1 \left(\sum_{i=1}^{n} x_{i1} \right)^2 + \beta_2 \sum_{i=1}^{n} x_{i1} \sum_{i=1}^{n} x_{i2} = \sum_{i=1}^{n} x_{i1} \sum_{i=1}^{n} y_i \tag{35}$$

(34) と (35) の差

$$\beta_1\left(n\sum_{i=1}^{n}{x_{i1}}^2-\left(\sum_{i=1}^{n}x_{i1}\right)^2\right)+\beta_2\left(n\sum_{i=1}^{n}x_{i1}x_{i2}-\sum_{i=1}^{n}x_{i1}\sum_{i=1}^{n}x_{i2}\right)$$
$$=n\sum_{i=1}^{n}x_{i1}y_i-\sum_{i=1}^{n}x_{i1}\sum_{i=1}^{n}y_i \tag{36}$$

ここで、分散, 共分散を使ってあらわすと、(15) (17) より、

$$\beta_1 n^2 V_{x_1}+\beta_2 n^2 Cov_{x_1x_2}=n^2 Cov_{x_1y} \tag{37}$$

$$\beta_1 V_{x_1}+\beta_2 Cov_{x_1x_2}=Cov_{x_1y} \tag{38}$$

を得る。

(33) と (31) による β_0 の打ち消し

(33) の両辺に n を掛ける

$$n\beta_0 \sum_{i=1}^{n} x_{i2} + n\beta_1 \sum_{i=1}^{n} x_{i1}x_{i2} + n\beta_2 \sum_{i=1}^{n} {x_{i2}}^2 = n \sum_{i=1}^{n} x_{i2}y_i \tag{39}$$

(31) の両辺に $\sum_{i=1}^{n} x_{i2}$ を掛ける

$$n\beta_0 \sum_{i=1}^{n} x_{i2} + \beta_1 \sum_{i=1}^{n} x_{i1} \sum_{i=1}^{n} x_{i2} + \beta_2 \left(\sum_{i=1}^{n} x_{i2}\right)^2 = \sum_{i=1}^{n} x_{i2} \sum_{i=1}^{n} y_i \tag{40}$$

(39) と (40) の差

$$\beta_1\left(n\sum_{i=1}^{n}x_{i1}x_{i2}-\sum_{i=1}^{n}x_{i1}\sum_{i=1}^{n}x_{i2}\right)+\beta_2\left(n\sum_{i=1}^{n}{x_{i2}}^2-\left(\sum_{i=1}^{n}x_{i2}\right)^2\right)$$
$$=n\sum_{i=1}^{n}x_{i2}y_i-\sum_{i=1}^{n}x_{i2}\sum_{i=1}^{n}y_i \tag{41}$$

同様に、分散, 共分散を使ってあらわすと、(15) (17) より、

$$\beta_1 n^2 Cov_{x_1x_2}+\beta_2 n^2 V_{x_2}=n^2 Cov_{x_2y} \tag{42}$$

$$\beta_1 Cov_{x_1x_2}+\beta_2 V_{x_2}=Cov_{x_2y} \tag{43}$$

を得る。

問題 III–5 – 3

次の連立方程式を解きなさい。ただし、変数は $\beta_1,\ \beta_2$ とし、$V_{x_1}V_{x_2} \neq {Cov_{x_1x_2}}^2$ とする。

$$\begin{cases} \beta_1 V_{x_1} + \beta_2 Cov_{x_1x_2} = Cov_{x_1y} \\ \beta_1 Cov_{x_1x_2} + \beta_2 V_{x_2} = Cov_{x_2y} \end{cases}$$

解例 III–5 – 3

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

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

定理 III–5 – 1

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 の共分散を使うと以下のようにあらわすことができる。ただし、$V_{x_1}V_{x_2} \neq {Cov_{x_1x_2}}^2$ とする。

$$\hat{\beta}_1 = \frac{Cov_{x_1y}V_{x_2} - Cov_{x_2y}Cov_{x_1x_2}}{V_{x_1}V_{x_2} - {Cov_{x_1x_2}}^2} \tag{44}$$

$$\hat{\beta}_2 = \frac{Cov_{x_2y}V_{x_1} - Cov_{x_1y}Cov_{x_1x_2}}{V_{x_1}V_{x_2} - {Cov_{x_1x_2}}^2} \tag{45}$$

問題 III–5 – 4

当てはめるモデルを $y_i = \beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2}$ とし、表の値を使って最小二乗推定値 $\hat{\beta}_1, \hat{\beta}_2$ を求めなさい。

i	x_{i1}	x_{i2}	y_i
1	1	1	12
2	1	2	16
3	1	3	22
4	1	4	27
5	2	1	19
6	2	2	23
7	2	3	28
8	2	4	35

i	x_{i1}	x_{i2}	y_i
9	3	1	27
10	3	2	32
11	3	3	34
12	3	4	41
13	4	1	33
14	4	2	39
15	4	3	42
16	4	4	47

解例 III–5 – 4

解答に際しては (44)(45) を用いても、何らかの統計ソフトを利用しても趣旨は変わりません。

$$
\begin{aligned}
\hat{\beta}_0 &= 0.1875 \\
\hat{\beta}_1 &= 7.025 \\
\hat{\beta}_2 &= 4.825 \\
\hat{y}_i &= 0.1875 + 7.025 \times x_{i1} + 4.825 \times x_{i2}
\end{aligned}
$$

- $\hat{\beta}_0$ は参考として推定値のみを掲載しておく。
- $\hat{\beta}_0$ の最小二乗推定値の求め方に関しては各自に任せる。

5.5.2 独立変数間の相関係数が 0 のとき

独立変数 $x_1,\ x_2$ の相関係数が $\rho_{x_1x_2}=0$ である場合を考える。これは独立変数間は無相関であるとの仮定が完全に満たされたときである。ここで、$V_{x_j}\neq 0$ とする。すると、(44) の分母分子それぞれの第二項は $Cov_{x_1x_2}=0$ なので

$$\hat{\beta}_1=\frac{Cov_{x_1y}V_{x_2}}{V_{x_1}V_{x_2}} \tag{46}$$

$$=\frac{Cov_{x_1y}}{V_{x_1}} \tag{47}$$

である。この値は、単回帰モデル

$$\hat{y}_i=\alpha+\beta x_{i1} \tag{48}$$

の最小二乗推定値 $(\hat{\beta})$ に一致する。$\hat{\beta}_2$ に関しても同様である。

5.6 推定値の大きさの比較の意味

推定値の大きさの比較を考える。説明の便宜上、x_1 の単位をインチ (in) とする。従属変数を y とする 2 変量からなる重回帰モデルを考える。

$$\hat{y}_i = \beta_{01} + \beta_{11} x_{i1} + \beta_{21} x_{i2} \tag{49}$$

x_{i1} を単位を変換しセンチメートル (cm) で測定すると

$$\hat{y}_i = \beta_{02} + \beta_{12} \left(2.54 \times x_{i1}\right) + \beta_{22} x_{i2} \tag{50}$$

$$= \beta_{02} + \left(2.54 \times \beta_{12}\right) x_{i1} + \beta_{22} x_{i2} \tag{51}$$

である。

通常、長さの単位変換は、換算比の積により行われる。より一般化をするために換算比を $\alpha > 0$ とするとモデル式は

$$\hat{y}_i = \beta_{01} + \beta_{11} x_{i1} + \beta_{21} x_{i2} \tag{52}$$

$$\hat{y}_i = \beta_{02} + \alpha\beta_{12} x_{i1} + \beta_{22} x_{i2} \tag{53}$$

である。両モデルの定数項と x_{i2} に対応する推定値は等しく

$$\beta_{01} = \beta_{02} \tag{54}$$

$$\beta_{21} = \beta_{22} \tag{55}$$

であって、x_{i1} に対応する推定値には

$$\beta_{11} = \alpha\beta_{12} \tag{56}$$

の関係がある。

5.6.1　最小二乗推定値を使った説明

$\alpha > 0$ を用いて $x_{i1\alpha} = \alpha x_{i1}$ とする。

$x_{i1\alpha}$ の平均 $\bar{x}_{1\alpha}$ は

$$
\begin{aligned}
\bar{x}_{1\alpha} &= \frac{1}{n}\sum_{i=1}^{n} x_{i1\alpha} &(57)\\
&= \frac{1}{n}\sum_{i=1}^{n} \alpha x_{i1} &(58)\\
&= \alpha\frac{1}{n}\sum_{i=1}^{n} x_{i1} &(59)\\
&= \alpha\bar{x}_1 &(60)
\end{aligned}
$$

$x_{1\alpha}$ の分散 $V_{x_{1\alpha}}$ は

$$
\begin{aligned}
V_{x_{1\alpha}} &= \frac{1}{n}\sum_{i=1}^{n}\left(x_{i1\alpha} - \bar{x}_{1\alpha}\right)^2 && (61)\\
&= \frac{1}{n}\sum_{i=1}^{n}\left(\alpha x_{i1} - \alpha\bar{x}_{1}\right)^2 && (62)\\
&= \frac{1}{n}\sum_{i=1}^{n}\left(\alpha x_{i1} - \alpha\bar{x}_{1}\right)\left(\alpha x_{i1} - \alpha\bar{x}_{1}\right) && (63)\\
&= \alpha^2\frac{1}{n}\sum_{i=1}^{n}\left(x_{i1} - \bar{x}_{1}\right)\left(x_{i1} - \bar{x}_{1}\right) && (64)\\
&= \alpha^2\frac{1}{n}\sum_{i=1}^{n}\left(x_{i1} - \bar{x}_{1}\right)^2 = \alpha^2 V_{x_1} && (65)
\end{aligned}
$$

$x_{1\alpha}$ と y の共分散 $Cov_{x_{1\alpha}y}$ は

$$
\begin{aligned}
Cov_{x_{1\alpha}y} &= \frac{1}{n}\sum_{i=1}^{n}\left(x_{i1\alpha} - \bar{x}_{1\alpha}\right)\left(y_i - \bar{y}\right) && (66)\\
&= \frac{1}{n}\sum_{i=1}^{n}\left(\alpha x_{i1} - \alpha\bar{x}_1\right)\left(y_i - \bar{y}\right) && (67)\\
&= \alpha\frac{1}{n}\sum_{i=1}^{n}\left(x_{i1} - \bar{x}_1\right)\left(y_1 - \bar{y}\right) && (68)\\
&= \alpha Cov_{x_1y} && (69)
\end{aligned}
$$

$x_{1\alpha}$ と x_2 の共分散 $Cov_{x_{1\alpha}x_2}$ は

$$
\begin{aligned}
Cov_{x_{1\alpha}x_2} &= \frac{1}{n}\sum_{i=1}^{n}(x_{i1\alpha}-\bar{x}_{1\alpha})(x_{i2}-\bar{x}_2) && (70)\\
&= \frac{1}{n}\sum_{i=1}^{n}(\alpha x_{i1}-\alpha\bar{x}_1)(x_{i2}-\bar{x}_2) && (71)\\
&= \alpha\frac{1}{n}\sum_{i=1}^{n}(x_{i1}-\bar{x}_1)(x_{i2}-\bar{x}_2) && (72)\\
&= \alpha Cov_{x_1x_2} && (73)
\end{aligned}
$$

したがって、単位変換後の最小二乗推定値は

$$\hat{\beta}_{12} = \frac{Cov_{x_{1\alpha}y}V_{x_2} - Cov_{x_2y}Cov_{x_{1\alpha}x_2}}{V_{x_{1\alpha}}V_{x_2} - {Cov_{x_{1\alpha}x_2}}^2} \tag{74}$$

$$= \frac{\alpha Cov_{x_1y}V_{x_2} - Cov_{x_2y}\alpha Cov_{x_1x_2}}{\alpha^2 V_{x_1}V_{x_2} - (\alpha Cov_{x_1x_2})^2} \tag{75}$$

$$= \frac{1}{\alpha}\frac{Cov_{x_1y}V_{x_2} - Cov_{x_2y}Cov_{x_1x_2}}{V_{x_1}V_{x_2} - {Cov_{x_1x_2}}^2} \tag{76}$$

$$= \frac{1}{\alpha}\hat{\beta}_{11} \tag{77}$$

$$\hat{\beta}_{22} = \frac{Cov_{x_2y}V_{x_{1\alpha}} - Cov_{x_{1\alpha}y}Cov_{x_{1\alpha}x_2}}{V_{x_{1\alpha}}V_{x_2} - {Cov_{x_{1\alpha}x_2}}^2} \tag{78}$$

$$= \frac{\cancel{\alpha^2} Cov_{x_2y}V_{x_1} - \cancel{\alpha^2} Cov_{x_1y}Cov_{x_1x_2}}{\cancel{\alpha^2} V_{x_1}V_{x_2} - \cancel{\alpha^2} {Cov_{x_1x_2}}^2} = \hat{\beta}_{21} \tag{79}$$

(31) に $\hat{\beta}_{11}, \hat{\beta}_{21}$ を代入すると

$$
\begin{aligned}
n\beta_{01} + \hat{\beta}_{11} \sum_{i=1}^{n} x_{i1} + \hat{\beta}_{21} \sum_{i=1}^{n} x_{i2} &= \sum_{i=1}^{n} y_i && (80) \\
\beta_{01} + \hat{\beta}_{11} \frac{1}{n} \sum_{i=1}^{n} x_{i1} + \hat{\beta}_{21} \frac{1}{n} \sum_{i=1}^{n} x_{i2} &= \frac{1}{n} \sum_{i=1}^{n} y_i && (81) \\
\beta_{01} + \hat{\beta}_{11} \bar{x}_1 + \hat{\beta}_{21} \bar{x}_2 &= \bar{y} && (82)
\end{aligned}
$$

したがって、単位変換前の最小二乗推定値 $\hat{\beta}_{01}$ は、

$$
\hat{\beta}_{01} = \bar{y} - \hat{\beta}_{11}\bar{x}_1 - \hat{\beta}_{21}\bar{x}_2 \tag{83}
$$

である。

ここで x_1 を単位変換後の $\hat{\beta}_{02}$ は、

$$\begin{aligned}\hat{\beta}_{02} &= \bar{y} - \hat{\beta}_{12}\bar{x}_{1\alpha} - \hat{\beta}_{22}\bar{x}_2 &(84)\\ &= \bar{y} - \left(\frac{1}{\alpha}\hat{\beta}_{11}\right)(\alpha\bar{x}_1) - \left(\hat{\beta}_{21}\right)\bar{x}_2 &(85)\\ &= \bar{y} - \hat{\beta}_{11}\bar{x}_1 - \hat{\beta}_{21}\bar{x}_2 &(86)\end{aligned}$$

である。つまり x_1 を α 倍すると、$\hat{\beta}_{12}$ は、$\hat{\beta}_{11}$ の α 分の 1 になるが、$\hat{\beta}_{01}, \hat{\beta}_{21}$ は影響を受けない。このことは、推定値は対応する独立変数の単位によってその大きさを変えるため、推定値の大きさを比較することに意味はないことを示している。ここでは $\alpha > 0$ としたが、$\alpha < 0$ の場合も同様である。上記の議論により直ちに次の定理が導かれる。

定理 III–5 – 2

2 変数重回帰モデルにおいて、独立変数 x_1 を $\alpha \neq 0$ を用いて α 倍すると x_1 に対応する最小二乗推定値 $\hat{\beta}_1$ は $\dfrac{1}{\alpha}$ 倍されるが、x_2 に対応する最小二乗推定値 $\hat{\beta}_2$ 並びに $\hat{\beta}_0$ は影響を受けない。

ここでは 2 変数の重回帰モデルで説明したが、2 変数を超える重回帰モデルにおいても同様である。

問題 III–5 – 5

当てはめるモデルを $y_i = \beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2}$ とし、表の値を使って最小二乗推定値を求めなさい。

i	x_1	x_2	y
1	1	10	12
2	1	20	16
3	1	30	22
4	1	40	27
5	2	10	19
6	2	20	23
7	2	30	28
8	2	40	35

i	x_1	x_2	y
9	3	10	27
10	3	20	32
11	3	30	34
12	3	40	41
13	4	10	33
14	4	20	39
15	4	30	42
16	4	40	47

解例 III–5 – 5

推定に際しては (44)(45)(83) を用いても、何らかの統計ソフトを利用しても趣旨は変わりません。お好きな方法で推定値を得てください。

$$\hat{\beta}_0 = 0.1875$$

$$\hat{\beta}_1 = 7.0250$$

$$\hat{\beta}_2 = 0.4825$$

$$\hat{y}_i = 0.1875 + 7.0250 \times x_{i1} + 0.4825 \times x_{i2}$$

※ 問題 III–5 – 4 と比較すると $\hat{\beta}_2$ が 0.1 倍になっているのが確認できる。

5.7 質的データの分析

5.7.1 ダミー変数

- 0 または 1 の値のみを取る変数をダミー変数と呼ぶ。
- 重回帰モデル (18) において x_2 をダミー変数とする。

$x_{i2} = 0$ の時の (18) は

$$\hat{y}_i = \beta_0 + \beta_1 x_{i1} \tag{87}$$

$x_{i2} = 1$ の時の (18) は

$$\hat{y}_i = \beta_0 + \beta_1 x_{i1} + \beta_2 \tag{88}$$

ダミー変数により重回帰モデル (18) は β_2 だけ $\hat{y}$ を変化させる。

図 5　ダミー変数の意味

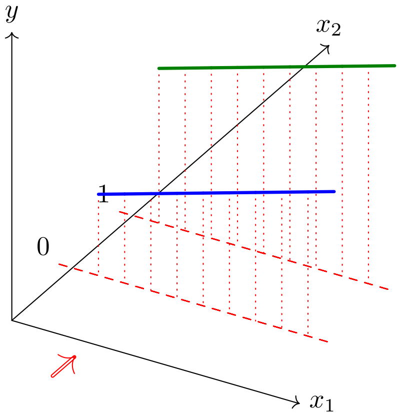


図 6　$x_1 \times y$ 平面上の重回帰式

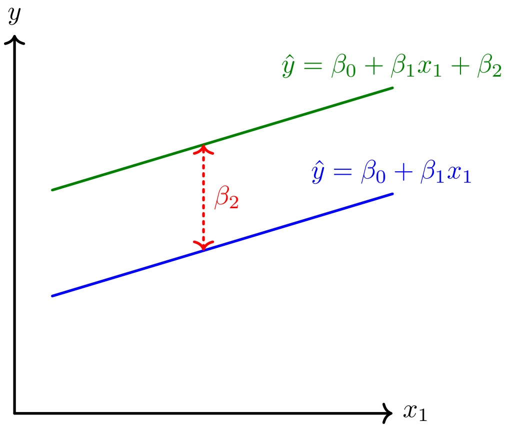


図 7 男性ダミーによる観測点の位置

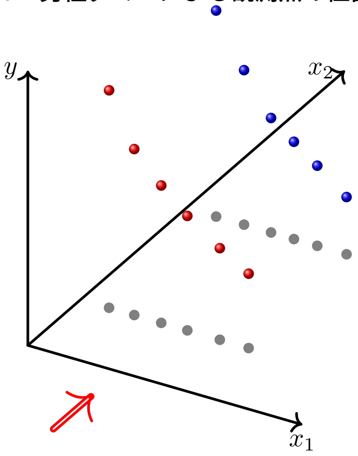
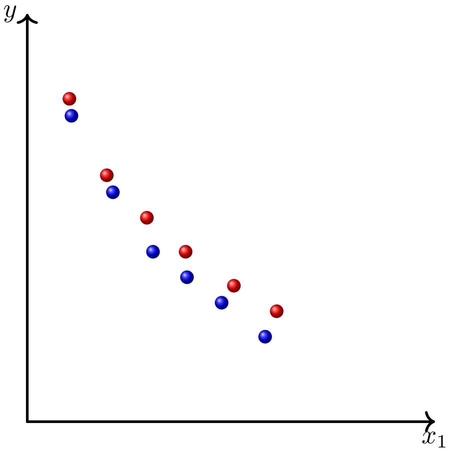


図 8 $x_1 \times y$ 面から見た観測点の位置



問題 III–5 – 6

x_1 は身長、x_2 は男性ダミー、y は 50m 走のタイムである。当てはめるモデルを $y_i = \beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2}$ とし、最小二乗推定値を求めなさい。

i	x_1	x_2	y_i
1	116.5	1	11.6
2	122.6	1	10.7
3	128.5	1	10.0
4	133.5	1	9.7
5	138.6	1	9.4
6	145.0	1	9.0

i	x_1	x_2	y_i
7	116.2	0	11.8
8	121.7	0	10.9
9	127.6	0	10.4
10	133.3	0	10.0
11	140.4	0	9.6
12	146.7	0	9.3

解例 III–5 – 6

推定に際しては (44)(45)(83) を用いても、何らかの統計ソフトを利用しても趣旨は変わりません。お好きな方法で推定値を得てください。

$$\hat{\beta}_0 = 21.14583\cdots$$

$$\hat{\beta}_1 = -0.08254\cdots$$

$$\hat{\beta}_2 = -0.28317\cdots$$

$$\hat{y}_i = 21.14583 - 0.08254 \times x_{i1} - 0.28317 \times x_{i2}$$

問題 III–5 – 7

x_1 は体重、x_2 は男性ダミー、y は 50m 走のタイムである。当てはめるモデルを $y_i = \beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2}$ とし、最小二乗推定値を求めなさい。

i	x_1	x_2	y_i
1	21.3	1	11.6
2	24.2	1	10.7
3	27.7	1	10.0
4	31.1	1	9.7
5	34.4	1	9.4
6	39.2	1	9.0

i	x_1	x_2	y_i
7	21.2	0	11.8
8	23.7	0	10.9
9	27.0	0	10.4
10	30.4	0	10.0
11	35.0	0	9.6
12	40.0	0	9.3

解例 III–5 – 7

推定に際しては (44)(45)(83) を用いても、何らかの統計ソフトを利用しても趣旨は変わりません。お好きな方法で推定値を得てください。

$$\hat{\beta}_0 = 14.16399\cdots$$

$$\hat{\beta}_1 = -0.12963\cdots$$

$$\hat{\beta}_2 = -0.25370\cdots$$

$$\hat{y}_i = 14.16399 - 0.12963 \times x_{i1} - 0.25370 \times x_{i2}$$

5.7.2 データを分けることによる対応

ダミー変数を用いるのではなく、男性と女性にデータを分けそれぞれ分析することも可能である。男性のデータを (x_{i1}, y_{i1})、女性のデータを (x_{i2}, y_{i2}) とする。このデータをそれぞれ二つの単回帰モデルに当てはめることを考える。

$$y_{i1} = \beta_{01} + \beta_{11}x_{i1} \tag{89}$$

$$y_{i2} = \beta_{02} + \beta_{12}x_{i2} \tag{90}$$

それぞれ異なるデータに当てはめているので、通常、

$$\beta_{01} \neq \beta_{02} \tag{91}$$

$$\beta_{11} \neq \beta_{12} \tag{92}$$

である。データを分けて個別に単回帰モデルを当てはめた場合、予測・モデルの当てはまりは、ダミー変数を採用した (18) よりも良好となるが、影響力の解釈という点では難易度が上がる。

問題 III–5 – 8

ある小学校の男子児童から以下のデータを得た。x_{i1} は体重、y_{i1} は 50m 走のタイムである。当てはめるモデルを

$$y_{i1} = \beta_{01} + \beta_{11} x_{i1}$$

とする。$\hat{\beta}_{01}, \hat{\beta}_{11}$ を求めなさい。

i	x_{i1}	y_{i1}
1	21.3	11.6
2	24.2	10.7
3	27.7	10.0
4	31.1	9.7
5	34.4	9.4
6	39.2	9.0

問題 III–5 – 9

ある小学校の女子児童から以下のデータを得た。x_{i2} は体重、y_{i2} は 50m 走のタイムである。当てはめるモデルを

$$y_{i2} = \beta_{02} + \beta_{12} x_{12}$$

とする。$\hat{\beta}_{02}, \hat{\beta}_{12}$ を求めなさい。

i	x_{i2}	y_{i2}
1	21.2	11.8
2	23.7	10.9
3	27.0	10.4
4	30.4	10.0
5	35.0	9.6
6	40.0	9.3

解例 III–5 − 8

$$\hat{\beta}_{01} = 14.127903\cdots$$

$$\hat{\beta}_{11} = -0.136973\cdots$$

$$\hat{y}_{i1} = 14.127903 - 0.136973 \times x_{i1}$$

解例 III–5 − 9

$$\hat{\beta}_{02} = 13.973983\cdots$$

$$\hat{\beta}_{12} = -0.123203\cdots$$

$$\hat{y}_{i2} = 13.973983 - 0.123203 \times x_{i2}$$

5.7.3 解釈方法の違い

解例 III–5 – 7 を例とすると、男性は女性に比べ 50m 走のタイムにおいて $-0.253\,70$ 秒差があると解釈できる。つまり、体重のもたらす 50m 走のタイムへの影響は等しく、同じ体重の場合のタイムの差は性別の差であることを仮定している。

解例 III–5 – 8, 解例 III–5 – 9 を例とすると、体重 1kg 増加したときの 50m 走のタイムの変化量は、男性の場合 $-0.136\,973$ 秒、女性の場合 $-0.123\,203$ 秒であると解釈できる。この場合、体重のもたらす 50m 走のタイムへの影響は性別ごとに異なることを仮定している。

これらの違いは、仮説の立て方の違いである。

5.8　季節変動の把握

- 周期的な変化を季節変動と呼ぶ。
- ダミー変数により季節変動の把握が可能。
- 季節変動分析用のダミー変数の例
    - 特定の曜日なら 1 を取りそれ以外の曜日であれば 0 を取る曜日ダミー
    - 特定の月なら 1 を取りそれ以外の月であれば 0 を取る月ダミー

5.8.1 ダミー変数を使った曜日の差の把握

d_{i1} を火曜日なら 1 それ以外の曜日であれば 0 をとる火曜日ダミーとする。同様に d_{i2} を水曜日ダミー、d_{i3} を木曜日ダミー、d_{i4} を金曜日ダミー、d_{i5} を土曜日ダミー、d_{i6} を日曜日ダミーとする。x_i, y_i を連続量として、重回帰モデルを以下のように定義する。

$$\hat{y}_i = \beta_0 + \beta_1 x_i + \beta_{d1} d_{i1} + \beta_{d2} d_{i2} + \beta_{d3} d_{i3} + \beta_{d4} d_{i4} + \beta_{d5} d_{i5} + \beta_{d6} d_{i6} \tag{93}$$

5.8.2　月曜日の時のモデル式

月曜日の場合、全ての曜日ダミーは 0 であり、ダミー変数に対応するパラメータは係数が 0 なのでモデルから除かれる。

$$\begin{aligned}\hat{y}_i &= \beta_0 + \beta_1 x_i + \beta_{d1}(0) + \beta_{d2}(0) + \beta_{d3}(0) + \beta_{d4}(0) + \beta_{d5}(0) + \beta_{d6}(0) && (94)\\ &= \beta_0 + \beta_1 x_i && (95)\end{aligned}$$

5.8.3 火曜日の時のモデル式

火曜日の場合は火曜ダミーのみが 1 をとり、それ以外は 0 を取る。したがって、β_{11} だけがモデルに残り (95) を規準として β_{11} だけ $\hat{y}_i$ を変化させる。他の曜日も同様である。

$$
\begin{aligned}
\hat{y}_i &= \beta_0 + \beta_1 x_i + \beta_{d1}(1) + \beta_{d2}(0) + \beta_{d3}(0) + \beta_{d4}(0) + \beta_{d5}(0) + \beta_{d6}(0) && (96) \\
&= \beta_0 + \beta_1 x_i + \beta_{d1} && (97)
\end{aligned}
$$

5.8.4 曜日によるモデルの違い

ダミー変数によりそれぞれの曜日と月曜日の差を把握できる。

$$\text{月曜日} \quad \hat{y}_i = \beta_0 + \beta_1 x_i \tag{95}$$

$$\text{火曜日} \quad \hat{y}_i = \beta_0 + \beta_1 x_i + \beta_{d1} \tag{97}$$

$$\text{水曜日} \quad \hat{y}_i = \beta_0 + \beta_1 x_i + \beta_{d2} \tag{98}$$

$$\text{木曜日} \quad \hat{y}_i = \beta_0 + \beta_1 x_i + \beta_{d3} \tag{99}$$

$$\text{金曜日} \quad \hat{y}_i = \beta_0 + \beta_1 x_i + \beta_{d4} \tag{100}$$

$$\text{土曜日} \quad \hat{y}_i = \beta_0 + \beta_1 x_i + \beta_{d5} \tag{101}$$

$$\text{日曜日} \quad \hat{y}_i = \beta_0 + \beta_1 x_i + \beta_{d6} \tag{102}$$

5.9 まとめ

- 1 変量の線型モデルを単回帰モデルと呼び、多変量の線型モデルを重回帰モデルと呼ぶ。
- 重回帰モデルもパラメータの推定方法は単回帰モデルと同じである。
- 独立変数間に無相関を仮定している。
- 完全に仮定が満たされた場合、重回帰モデルの最小二乗推定値は、単回帰モデルのそれに一致する。
- 最小二乗推定値は独立変数の単位の影響を受けるため大小比較に意味はない。
- ダミー変数を採用すると質的な違いを分析可能となる。
- 季節変動の把握も可能となる。