広義の多変量解析のなかで，重回帰分析は最も重要なモデルの1つであり，実用性も高い。このため数多くの良書が出版されている。本講座では，理論の記述に適した行列記述を用いて各種統計量を導くとともに，理解しやすい数値例を示して計算手順を示すことにする。

行列記述を用いることの利点は，重回帰分析の全体的な視点に立つ整理ができることである。行列記述に慣れられていない読者も恐れずに慣れることに努力していただきたい。

2. データ

以下，データは，応答変数 $y$ と $x_1$ から $x_4$ までの4個の説明変数からなる7個の観測データである。

$y$：分娩までの経過時間の自然対数による表示
$x_1$：子宮口開大度
$x_2$：陣痛間欠時間
$x_3$：胎児心拍数
$x_4$：陣痛持続時間

4個の説明変数は観測時点において計測され，応答変数をその時点から分娩までの経過時間を示す。次の7個の時系列データは同一母体からのものである。

このデータに，多重共線性の説明に用いる変数 $x_5$ を追加する。

$x_5$：$x_2+x_4$ ただし，最初のデータのみ，この値に2をさらに加える。

3. 重回帰モデルの定義とパラメータの推定

重回帰モデルは，変数のレベルで表すと，応答変数 $y$ と説明変数を $x_i (i=1,\ldots,p)$ と表わして，

$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \cdots + \beta_p x_p + \varepsilon \quad (3)$$

と表わされる。ここで $\varepsilon$ は誤差である。なお，説明変数は確率変数でも決定論的変数でもよいが，確率変数の場合には，その実現値は正規に測定されるものと仮定する。

これをデータのレベルで表わすと式(4)で説明される。

$$y_i = \beta_0 + \beta_1 x_{1i} + \beta_2 x_{2i} + \cdots + \beta_p x_{pi} + \varepsilon_i \quad (i=1,\ldots,n) \quad (4)$$

ここで，$n$ はサンプル数，$p$ は説明変数の個数を示す。

これを，さらに行列記述すれば式(5)になる。

$$
\begin{bmatrix}
\mathbf{y}_1 \\
\mathbf{y}_2 \\
\vdots \\
\mathbf{y}_n
\end{bmatrix} = 
\begin{bmatrix}
\mathbf{x}_{11} & \mathbf{x}_{12} & \cdots & \mathbf{x}_{1n} \\
\mathbf{x}_{21} & \mathbf{x}_{22} & \cdots & \mathbf{x}_{2n} \\
\vdots & \vdots & \ddots & \vdots \\
\mathbf{x}_{n1} & \mathbf{x}_{n2} & \cdots & \mathbf{x}_{nn}
\end{bmatrix} 
\begin{bmatrix}
\beta_0 \\
\beta_1 \\
\vdots \\
\beta_p
\end{bmatrix} + 
\begin{bmatrix}
\mathbf{e}_1 \\
\mathbf{e}_2 \\
\vdots \\
\mathbf{e}_n
\end{bmatrix} \quad (5)
$$

ただし，これらのモデル中に現われる誤差 $\varepsilon$ について，

しんむら しゅういち 住商コンピュータサービス㈱

1983年9月号
以下の仮定を置く。

i) 不偏性: 期待値は零である。

\[ E(\varepsilon_i) = 0 \]

ii) 等分散性: 分散は一定である。

\[ V(\varepsilon_i) = \sigma^2 \]

iii) 独立性: 誤差が互いに独立である。

\[ \varepsilon_i \perp \varepsilon_j \quad (i \neq j) \]

iv) 正規性: 誤差は正規分布する。

\[ E(\varepsilon_i) = 0, \quad V(\varepsilon_i) = \sigma^2 \]

以上をまとめると, 誤差は平均 0, 分散がの正規分布をすることになる。すなわち, 誤差は正規分布をする。

\[ \varepsilon \sim N(0, \sigma^2) \]

行列表記でまとめると,

\[ \begin{pmatrix} 7, 390 \\ 7, 300 \\ 2, 708 \end{pmatrix} = \begin{pmatrix} 1 & 8 & \cdots & 18 \\ 1 & 5 & \cdots & 20 \\ 1 & 5 & \cdots & 10 \end{pmatrix} \begin{pmatrix} \beta_0 \\ \beta_1 \\ \varepsilon_1 \end{pmatrix} \]

次の誤差平方和 (SSE) を最小にする未知数 \( \beta \) の推定値 \( \hat{\beta} \) を求め次の方法を最小二乗法という。

\[ \text{SSE} = \sum (y_i - \hat{y}_i)^2 \]

\[ = \sum (y_i - \beta_0 - \beta_1 x_1 - \cdots - \beta_p x_p)^2 \]

これを \( \hat{\beta}_k (k = 1, \ldots, p) \) で微分すれば,

\[ \frac{\partial}{\partial \hat{\beta}_k} (\text{SSE}) = 2 \sum (y_i - \beta_0 - \beta_1 x_1 - \cdots - \beta_p x_p) (x_{ki}) \]

これを零と置いて得られる \( \hat{\beta} \) の推定方程式の解 \( \hat{\beta}_k \) は, 次の 2 つの方程式が行ることで求まる。ただし, すべてのデータは等しい。

\[ \frac{\partial}{\partial \beta_k} (\text{SSE}) = \sum (y_i - \beta_0 - \beta_1 x_1 - \cdots - \beta_p x_p) (x_{ki}) > 0 \]

以上から, 推定値 \( \hat{\beta} \) は次の正規方程式を解いて求まる。

\[ X'X\hat{\beta} = X'y \quad \text{(正規方程式)} \]

\[ \hat{\beta} = (X'X)^{-1}X'y \quad \text{(解)} \]

【注 4】実際の重回帰分析のアルゴリズムは、行列表記を用いて行う。また, \( X'X \) の関係を求めていくことにより, \( X'y \) の場所に \( \beta \) の推定値が求まる。

【例】正規方程式の解の計算を行う。

\[ \begin{pmatrix} x_1 & x_2 & x_3 \end{pmatrix} \]

\[ \begin{pmatrix} 7 & 46 & 129 & 958 & 268 \\ 46 & 310 & 908 & 6328 & 1654 \\ 958 & 6328 & 17722 & 35400 \end{pmatrix} \]

\[ \begin{pmatrix} x_4 \end{pmatrix} \]

\[ \begin{pmatrix} 268 & 1654 & 4222 & 35400 \end{pmatrix} \]

また, \( X'y \) の逆行列, \( X'y \) の推定値 \( \hat{\beta} \) は次のとおりである。

\[ \begin{pmatrix} x_1 \\ x_2 \\ x_3 \\ x_4 \end{pmatrix} \]

\[ \begin{pmatrix} 56.936 & -1.086 & -0.009 & -0.328 & -0.122 \\ -1.086 & 0.243 & -0.006 & -0.004 & 0.003 \end{pmatrix} \]

\[ \begin{pmatrix} x_1 \\ x_2 \\ x_3 \\ x_4 \end{pmatrix} \]

\[ \begin{pmatrix} -0.009 & -0.006 & 7.2E-4 & 2.6E-4 & 3.9E-5 \\ -0.328 & -0.004 & 2.6E-4 & 0.002 & 5.9E-5 \end{pmatrix} \]

\[ \begin{pmatrix} -0.122 & 0.003 & 3.9E-5 & 5.9E-4 & 4.5E-4 \end{pmatrix} \]

データ行列 \( D \) の各列が, 其の平均値を引き出したものを偏差行列 \( D_d \) とよぶことになる。この時, \( D_d'D_d \) は偏差平方和積和行列になる。\( D \) の各列の平均値を用いて欠測値 \( M \) の要因とするとき, \( D'D \) と \( D_d'D_d \) の関係は次のとおりになる。
Dr/Dã=D−nM'M  ⑴

\[ \begin{align*}
\text{D_dD_d} & = D' - nM'M
\end{align*} \]

【例】D/D は式(10)で求めた X'X の1行1列を省いたものに等しくなる。

\[ x_1 \quad x_2 \quad x_3 \quad x_4 \]
\[ 302.286 \quad 847.714 \quad 6295.430 \quad 1761.140 \]
\[ 847.714 \quad 2377.290 \quad 17654.600 \quad 36677.700 \]
\[ 6295.430 \quad 17654.600 \quad 36677.700 \quad 10260.600 \]

よって,

\[ \begin{align*}
\text{D_dD_d} & =
\begin{bmatrix}
7.714 & 60.286 & 32.571 & -107.143 \\
60.286 & 1913.710 & 672.857 & -716.857 \\
32.571 & 672.857 & 782.857 & -1277.710 \\
-107.143 & -716.857 & -1277.710 & 4763.430
\end{bmatrix}
\end{align*} \]

これを自由度 (n−1)で割ったものがデータの分散共分散行列 Vã になる。

\[ \text{Vã}(v_{ij}) = \text{Dã}/(n-1) \quad (12) \]

【例】行列(11')より分散共分散行列は次のとおり。

\[ \begin{align*}
\text{Vã} & =
\begin{bmatrix}
1.000 & 0.496 & 0.419 & -0.559 \\
0.496 & 1.000 & 0.419 & -0.237 \\
0.419 & 0.419 & 1.000 & -0.662 \\
-0.559 & -0.237 & -0.662 & 1.000
\end{bmatrix}
\end{align*} \]

以上のような形で変数にまとめたものを分散分析表(修正前)とよぶ。

<table>
<thead>
<tr>
<th>D.F.</th>
<th>平均平方和</th>
<th>F 値</th>
</tr>
</thead>
<tbody>
<tr>
<td>回帰</td>
<td>p+1</td>
<td>( S_1 / S_2 )</td>
</tr>
<tr>
<td>誤差</td>
<td>n-p-1</td>
<td>( e 'S_3 / e 'S_2 )</td>
</tr>
<tr>
<td>全体</td>
<td>n</td>
<td>( y 'y )</td>
</tr>
</tbody>
</table>

1) 分散分析表の理解を助けるため、以下で射影行列(文献2)を導入する。射影行列 Q は、Q'Q = Q'Q = X(X'X)^{-1}X'y = L(X), rank Q = rank X の性質をもつ。
図 1 射影子の幾何学表現
ただし，D.F.は自由度を示し，行列Xの列数が回帰の，行数から列数を差し引いたものが誤差の自由度を表わす．F値は自由度 \((p+1, n-p-1)\) のF分布にしたがう．

[例] \(y = \beta_0 + \sum_{i=1}^p \beta_i x_i + \epsilon_i\) に対する分散分析表は次のとおり．

<table>
<thead>
<tr>
<th>D.F.</th>
<th>平方和</th>
<th>均方和</th>
<th>F値</th>
</tr>
</thead>
<tbody>
<tr>
<td>回帰</td>
<td>5</td>
<td>264.706</td>
<td>52.941</td>
</tr>
<tr>
<td>誤差</td>
<td>2</td>
<td>2.507</td>
<td>1.254</td>
</tr>
<tr>
<td>全体</td>
<td>7</td>
<td>267.213</td>
<td></td>
</tr>
</tbody>
</table>

このF検定は，次の帰無仮説 \(H_0\) を検定することに等しい．

\[ H_0 : \beta_0 = \beta_1 = \cdots = \beta_p = 0 \] (23)

この検定は現在考えているモデルが \(y = \epsilon\) のモデルと比較して有意か否かの検定であり，当然ずれて有意な情報もたらさない．そこで，すべての回帰モデルのベースとして次の定数項モデルを考えることにする．

\[ y_i = \bar{y} + \epsilon_i (i = 1, \ldots, n) \] (24)

このモデルに対する帰無仮説 \(H_0'\) と対立仮説 \(H_1'\) は次のとおり．

\[ H_0' : \beta_0 = \bar{y}, \beta_i = 0 (\text{for } i = 1, \ldots, p) \] (25)

これらの関係を図2に示す．すなわち，分散分析表(22)は回帰平方和として \(\sum \hat{y}_i^2\) を表わすのに対し，モデル(24)をベースにした回帰平方和は，\(\bar{y}\)の偏差平方和の\(\sum \epsilon_i^2\)に差し引く，回帰平方和と全体の平方和から中心効果 \(n\bar{y}^2\)を差し引き，自由度を \(p\) と (n−1) に修正した次の分散分析表を求めたこととなる．

分散分析表(修正済み)

<table>
<thead>
<tr>
<th>D.F.</th>
<th>平方和</th>
<th>均方和</th>
<th>F値</th>
</tr>
</thead>
<tbody>
<tr>
<td>回帰</td>
<td>( \hat{y}'y-n\bar{y}^2 )</td>
<td>( \hat{y}'y-n\bar{y}^2 )</td>
<td>( \hat{y}'y-n\bar{y}^2 )</td>
</tr>
<tr>
<td>誤差</td>
<td>( n-p-1 )</td>
<td>( \epsilon'e )</td>
<td>( \epsilon'e )</td>
</tr>
<tr>
<td>全体</td>
<td>( n-1 )</td>
<td>( y'y-n\bar{y}^2 )</td>
<td></td>
</tr>
</tbody>
</table>

この \(R^2\) 値は，式変形により，平均回帰平方和 \(S_t\) と平均誤差平方和 \(S_e\) の平方和で表わされるので，分散分析表によるF値と，決定係数 \(R^2\) に対する検定は型式が違っても本質的に同じであるので，一方を行えば，他方を行なう必要はない．

5. パラメータの各種統計量

パラメータ \(\beta\) の推定値 \(\hat{\beta}\) の期待値は次式で与えられる．

\[ \hat{\beta} = \hat{\beta}_0, \hat{\beta}_1 = \cdots = \hat{\beta}_p = 0 \] (23)

この \(R^2\) 値が，式変形により，平均回帰平方和 \(S_t\) と平均誤差平方和 \(S_e\) の比で表わされるので，分散分析表によるF値と，決定係数 \(R^2\) に対する検定は型式が違っても本質的に同じであるので，一方を行えば，他方を行なう必要はない．

5. パラメータの各種統計量

パラメータ \(\beta\) の推定値 \(\hat{\beta}\) の期待値は次式で与えられる．

\[ \hat{\beta} = \hat{\beta}_0, \hat{\beta}_1 = \cdots = \hat{\beta}_p = 0 \] (23)
\[ E(\hat{\beta}) = E((XX')^{-1}X'y) = (XX')^{-1}X'E(y) = (XX')^{-1}X'X\beta = \beta \]

\( y \) の分散行列 \( \text{Var}(y) \) は、
\[ \text{Var}(y) = E((y - X\beta)(y - X\beta)') \]
\[ = \text{Var}(\epsilon) = \sigma^2 \]

推定値 \( \hat{\beta} \) の分散行列は、次式になる。
\[ \text{Var}(\hat{\beta}) = \text{Var}((XX')^{-1}X'y) = (XX')^{-1}X'X(XX')^{-1} \sigma^2 = (XX')^{-1} \sigma^2 \]

6. 多重共線性 (multi-collinearity)

ある説明変数が他の説明変数の1次結合でほぼ表わされる時、\( \hat{\beta} \) は確定的でなく、多重共線性をもつ。


① 推定値は、データの小さな変化に対して不安定である。
② 推定値は大きな標準誤差をもつ。このため、t 検定が棄却できないことが多い。

多重共線性の検出方法としては、リッジ回帰分析 (文献 [3] pp.201-206)、主成分分析 (文献 [3])、分散拡大要因 (Variance Inflation Factor, VIF) 等がある。これらの方法を以下に解説しよう。

なお、多重共線性が検出された場合、対応としてはパラツキの弱い次元に広く分布するデータを追加するか、多重共線性に係る変数のいくつかをモデルから省くという2つの方法が考えられる。

6.1 分散拡大要因 (VIF)

\( \beta_i \) の VIF は、\( x_i \) を応答変数として残りのすべての説明変数で回帰して得られる多重決定係数 \( R_i^2 \) を用いて次式で表される。
\[ VIF_i = 1/(1 - R_i^2) \]

両相関行列を対比してわかるところは、\( x_3 \) と \( x_4 \) が高度相関をもつことは当然として、\( x_3 \) をモデルに入れたことにより \( x_2 \) と \( x_4 \) の間にも高い相関が認められるようになった。

\( X'X \) を \( X'X \) の \( i \) 番目の対角要素とすれば、\( \hat{\beta_i} \) の標準偏差 stderr(\( \hat{\beta_i} \)) と t 統計量は次式で与えられる。
\[ \text{stderr}(\hat{\beta_i}) = \sqrt{(XX')_{ii}} \]
\[ t = \frac{\hat{\beta_i}}{\text{stderr}(\hat{\beta_i})} \]

[例] 式 (31') と式 (32) から、\( \hat{\beta} \) の標準偏差と t 値は次のとおりになる。

<table>
<thead>
<tr>
<th>( \hat{\beta}_i )</th>
<th>Stderr</th>
<th>t(( \hat{\beta}_i ))</th>
</tr>
</thead>
<tbody>
<tr>
<td>( \hat{\beta}_1 )</td>
<td>8.449</td>
<td>-0.084</td>
</tr>
<tr>
<td>( \hat{\beta}_2 )</td>
<td>0.552</td>
<td>-0.327</td>
</tr>
<tr>
<td>( \hat{\beta}_3 )</td>
<td>0.055</td>
<td>1.778</td>
</tr>
<tr>
<td>( \hat{\beta}_4 )</td>
<td>0.024</td>
<td>-1.175</td>
</tr>
</tbody>
</table>

VIF が10以上の場合は多重共線性が疑われる (文献 [3] pp.201-202)。

[例] 説明変数が \( x_1, x_2, x_3, x_4 \) の4変数の場合、モデル \( x_4 = \hat{\beta}_0 + \hat{\beta}_2 x_2 + \hat{\beta}_3 x_3 + \hat{\beta}_4 x_4 + \epsilon \) の決定係数を \( R_{1,4}^2 \) とすれば、\( x_4 \) の分散拡大要因は VIF_{1,4} = 1/(1 - R_{1,4}^2) になる。同様に \( x_2, x_3, x_4 \) の VIF も計算される。

VIF_{1,4} = 1.875
VIF_{2,4} = 1.385
VIF_{3,4} = 1.863
VIF_{4,123} = 2.156

になる。

多重共線性のない4個の説明変数の組に、\( x_3 \) を追加す
と, 多重共線関係にある $x_2, x_4, x_5$ の分散拡大要因は極端に大きくなる。

以上に示したように多重共線関係にある説明変数の検出は容易に行なえる。しかし、その対応策として、どの変数をどのような基準にもとづいて何個省けばよいのか問題が残る。これを、かりに“多重共線性の解消”問題とよぶが、これは統計的に決めるべき問題ではなく、その問題の専門分野の知識を参考にして決めるべきであろう。

$eta_i$ の各 VIF の値は、$(X'X)^{-1}$ の各 i 番目の対角要素 $(X'X)_{ii}$ の値と比例関係にある。この $(X'X)^{-1}$ は式 (32) からわかるとおり、分散 $s^2$ を $(X'X)_{ii}$ 倍に拡大したものが $eta_i$ の分散になることを示しているので、分散拡大要因とよばれる。

【例】次の簡単なデータを考える。

$$y = x_1 + x_2$$

$$(X'X)^{-1} = \begin{pmatrix}
3 & 0 & -1 \\
0 & 2 & 9 \\
-1 & 9 & 9
\end{pmatrix}$$

$$a = \begin{pmatrix}
9 \\
7 \\
9
\end{pmatrix}^T$$

モデル $y = a_0 + a_1 x_1 + a_2 x_2 + \epsilon$ に対して、

$$VIF_{x_1} = 1 / (1 - R^2) = 10 / 9$$

$$(X'X)^{-1} = \begin{pmatrix}
3 & 0 & -1 \\
0 & 2 & 9 \\
-1 & 9 & 9
\end{pmatrix}$$

$$a = \begin{pmatrix}
9 \\
7 \\
9
\end{pmatrix}^T$$

モデル $x_1 = b_0 + b_1 x_2 + \epsilon$ に対して、

$$VIF_{x_1} = 1 / (1 - R^2) = 25 / 9$$

$$(X'X)^{-1} = \begin{pmatrix}
3 & 0 & -1 \\
0 & 2 & 9 \\
-1 & 9 & 9
\end{pmatrix}$$

$$b = \begin{pmatrix}
0 \\
1/2
\end{pmatrix}^T$$

8.2 主成分分析の利用

主成分分析は、データが多変量正規分布すなわち確率楕円にしたがうとして、元の変数の作る座標系を座標变换により楕円の軸を新座標系として求める手法である。

各説明変数を、平均 0 (原点移動) と分散 1 (単位系の違い等による影響を除くため) に標準化したデータ行列 $D$ を考える。この行列の列数 (説明変数の数) を $p$, 行数 (データ数) を $n$ とする。ここで $p$ 個の重みベクトル $a = (a_1, a_2, \ldots, a_p)^T$ による次の座標変換を考える。

$$z = Da$$

$$(X'X)^{-1} = \begin{pmatrix}
3 & 0 & -1 \\
0 & 2 & 9 \\
-1 & 9 & 9
\end{pmatrix}$$

$$a = \begin{pmatrix}
9 \\
7 \\
9
\end{pmatrix}^T$$

モデル $x_1 = b_0 + b_1 x_2 + \epsilon$ に対して、

$$R^2 = 0.64$$

$$VIF_{x_1} = 1 / (1 - R^2) = 25 / 9$$

$$(X'X)^{-1} = \begin{pmatrix}
3 & 0 & -1 \\
0 & 2 & 9 \\
-1 & 9 & 9
\end{pmatrix}$$

$$a = \begin{pmatrix}
9 \\
7 \\
9
\end{pmatrix}^T$$

モデル $x_1 = b_0 + b_1 x_2 + \epsilon$ に対して、

$$R^2 = 0.64$$

$$VIF_{x_1} = 1 / (1 - R^2) = 25 / 9$$

$$(X'X)^{-1} = \begin{pmatrix}
3 & 0 & -1 \\
0 & 2 & 9 \\
-1 & 9 & 9
\end{pmatrix}$$

$$a = \begin{pmatrix}
9 \\
7 \\
9
\end{pmatrix}^T$$

モデル $x_1 = b_0 + b_1 x_2 + \epsilon$ に対して、

$$R^2 = 0.64$$

$$VIF_{x_1} = 1 / (1 - R^2) = 25 / 9$$

$$(X'X)^{-1} = \begin{pmatrix}
3 & 0 & -1 \\
0 & 2 & 9 \\
-1 & 9 & 9
\end{pmatrix}$$

$$a = \begin{pmatrix}
9 \\
7 \\
9
\end{pmatrix}^T$$

モデル $x_1 = b_0 + b_1 x_2 + \epsilon$ に対して、

$$R^2 = 0.64$$

$$VIF_{x_1} = 1 / (1 - R^2) = 25 / 9$$
いることから次式で表わされ、さらにデータの相関行列を \( R \) として次式になる。

\[
V_x = \frac{1}{n} \sum_{i=1}^{n} a_i' D_i' D_i a
\]

\[
= a' \left( \frac{1}{n} D_i' D_i \right) a = a' Ra
\]

(39)

ここで、\( a' a = 1 \) の条件で \( V_x \) を最大にするものを考え、条件つき極値問題なので、ラグランジェの未定乗数を \( A \) として、次の式を最大にする \( a \) を求めればよい。

\[
\varphi = a' Ra - \lambda (a' a - 1)
\]

\[
\frac{\partial \varphi}{\partial a} = 2Ra - \lambda (2a) = 0
\]

式(41)は、相関行列 \( R \) の固有値問題になる。

\[
(R - \lambda E) a = 0
\]

(42)

ただし、ここで \( E \) は単位行列、\( \lambda \) は固有値、\( a \) は固有ベクトルである。

一方、\( Ra = \lambda a \) の両辺の左側に、\( a' \) を乗じれば、

\[
V_x = a' Ra = a' a = \lambda
\]

(43)

となり、固有値 \( \lambda \) は座標 \( a \) でのデータの分散を与える。

相関行列 \( R \) の階数が \( p \) なら、\( p \) 次の固有値 \( \lambda_i \) と固有ベクトル \( a_i \) が求まる。

固有値の大きさ順に並べて \( \lambda_1, \ldots, \lambda_p \) とする。対応する固有ベクトル \( a_1, \ldots, a_p \) は、第 \( 1 \) 主成分軸、……、第 \( p \) 主成分軸とよばれる新座標系の係数を与える。このようにして求めた \( p \) 個の新座標系で、元のデータ \( D_i \) は新座標 \( (D_i a_1, \ldots, D_i a_p) \) に変換される。

もし \( \lambda_p = 0 \) ならば、第 \( p \) 主成分軸上のデータ \( D_i a_p(i = 1, \ldots, n) \) の分散が相対的に零になり、\( D_i a_p \) は一定値とみなせる。元の変数の期待値は零に規準化されており、これらの変数の変数の変数 \( D_i a_p \) の期待値も零になる。すなわち、元の \( i \) 番目の変数を \( x_i \) とすれば、

\[
a_i = a_{i1} + a_{i2} x_2 + \cdots + a_{ip} x_p 
\]

\[a_{ip} \neq 0\]

という関係式が成り立つ。この式が変数 \( x_1, \ldots, x_p \) の間の多重共線性を表わす。したがって、\( a_i \) を零とみなせば特定の変数間の強い多重共線性を検出できる（文献3）p.179）。

変数 \( x_1 \) の作成過程から次式(46)が期待される。

\[
x_2 + x_3 - x_4 = 0
\]

(46)

しかし、実際には式(45)になったのは、データ数が少ないため最初のデータに加えられたバイアスの影響と、データが多変量正規分布から離散しているためと考えられる。

### 参考文献

1) N. ドレイパー他：応用回帰分析, 森北出版, 1968
2) 石井吉郎：実験計画法の基礎, サイエンス社, 1972
3) S. チャタジー他：回帰分析の実際, 新曜社, 1971
4) J. ジョンストン：計量経済学の方法, 東洋経済新報社, 1975
5) 小林龍一：相関・回帰分析法入門, 日科技連, 1972
6) SAS ユーザーズガイド, SAS Inc., 1982
8) 新村秀一：多重共線関係の解消とその影響, 1983年度OS学会春季研究発表会, 156/157
11) 竹内啓：現象と行動のなかの統計数理, 新曜社, 1972
12) 坂元慶行, 石黒真木夫, 北川源四郎：情報量統計学, 共立出版社, 1983