HITAC 基本統計システムにおける重回帰プログラムの統計量

Statistics of the Multiple Regression Program for the Basic Statistical System in HITAC

沢　　　勲　*Isao SAWA*

高　東　敬　*Dongkyong KO*

采　山　寛　幸　*Hiroyuki UNEYAMA*

ABSTRACT

Fiting straight Lines can clearly be seen in the method of ordinary least squares (OLS). Here in this paper, We will indicate some characters of various kinds of statistics, which we have succeeded in summariying from a functional point of view. Namely, there are three assuming conditions for independent variables and disturbance vector with zero means. A first one is the homogeneity and independency of disturbance variances. Secondly, data matrix is non stochastic. A third one is that independent variables are mutually independent. On this point, we insist that we put our operating procedure into vector indication on the basis of these three conditions.

Thus, the important characters of statistics can be described with ease in any different dimension : in particular, favorable characters of estimators ; analysis of variance ; and degrees of freedom of residuals.

1　はじめに

大阪経済法科大学学生のスポーツテストの統計処理において、日立製作所の基本統計システムの重回帰プログラムを適用した。重回帰プログラムでは線形回帰式を用いた。線形回帰式の当てはめには、通常の最小二乗法を適用した。[1~10]

この中で諸統計量の性質を、機能的にまとめることができたので紹介する。この紹介の説明変数と撹乱項に仮定される条件は、３つある。[12] すなわち、第１に撹乱項は期待値０の確率変数であり、分散は一様で互いに独立である。次に、説明変数のデータ行列は固定である。最後に、説明変数は互いに独立である。

これら３つの条件から出発して、操作手順をベクトルで表示することによって、諸性質が簡明に記述される。特に、推定量の望ましい性質、分散分析の構造および残差の自由度などである。これらの統計量の重要な性質を、どの次元にも共通した形で、簡明に記述することがこの論文のテーマである。

2　推定パラメータ[12]

2.1 推　定　式

推定パラメータ $\hat{\beta}$ の推定式 $(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{Y}$ は、最小二乗法の原理「残差 $\mathbf{e}=\mathbf{Y}-\hat{\mathbf{Y}}$ の平方和 $\mathbf{e}'\mathbf{e}$ をパラメータ $\hat{\beta}$ につき最小にする」に基づいて、極値の必要条件式：$\partial(\mathbf{e}'\mathbf{e})/\partial\hat{\beta}=\mathbf{0}$ を $\hat{\beta}$ について解いたものである。実際、式 $\mathbf{e}'\mathbf{e}$ を $\hat{\mathbf{Y}}=\mathbf{X}\hat{\beta}$ を使って展開すると、

$$\begin{aligned}\mathbf{e}'\mathbf{e} &= (\mathbf{Y}-\hat{\mathbf{Y}})'(\mathbf{Y}-\hat{\mathbf{Y}}) = (\mathbf{Y}-\mathbf{X}\hat{\beta})'(\mathbf{Y}-\mathbf{X}\hat{\beta}) \\ &= \mathbf{Y}'\mathbf{Y} - 2\hat{\beta}'\mathbf{X}'\mathbf{Y} + \hat{\beta}'\mathbf{X}'\mathbf{X}\hat{\beta} \quad\cdots\cdots(2.1)\end{aligned}$$

したがって、

$$\frac{\partial}{\partial\hat{\beta}}(\mathbf{e}'\mathbf{e}) = -2\mathbf{X}'\mathbf{Y} + 2\mathbf{X}'\mathbf{X}\hat{\beta} \quad\cdots\cdots(2.2)$$

３つ目の前提条件から $(\mathbf{X}'\mathbf{X})^{-1}$ が存在するから、上式の右辺を $\mathbf{0}$ と置き、$\hat{\beta}$ について解けば、推定式 $(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{Y}$ が得られる。

2.2 期 待 値

$\mathbf{Y}=\mathbf{X}\boldsymbol{\beta}+\mathbf{U}$ を使って上記 $\boldsymbol{\beta}$ の推定式の $\mathbf{Y}$ の部分を書き換え、期待値を取れば、１つ目の前提条件 $\mathrm{E}(\mathbf{U})=\mathbf{0}$, から、

$$\mathrm{E}(\hat{\boldsymbol{\beta}}) = \mathrm{E}((\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{Y}) = \mathrm{E}((\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'(\mathbf{X}\boldsymbol{\beta}+\mathbf{U})) = \boldsymbol{\beta} \quad \cdots\cdots(2.3)$$

したがって、$\hat{\boldsymbol{\beta}}$ は $\mathbf{Y}$ について線型な不偏推定量 (linear unbiased estimator) である。

2.3 分散共分散行列

推定式の $\hat{\boldsymbol{\beta}}=(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{Y}$ と期待値の $\mathrm{E}(\hat{\boldsymbol{\beta}})=\boldsymbol{\beta}$ の結果と $\mathbf{Y}=\mathbf{X}\boldsymbol{\beta}+\mathbf{U}$ から

$$\begin{aligned}\hat{\boldsymbol{\beta}}-\mathrm{E}(\hat{\boldsymbol{\beta}}) &= (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{Y} - \boldsymbol{\beta} = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'(\mathbf{X}\boldsymbol{\beta}+\mathbf{U}) - \boldsymbol{\beta} \\ &= (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{U} \quad \cdots\cdots(2.4)\end{aligned}$$

したがって、１つ目の前提条件〔$\mathrm{E}(\mathbf{U})=0,\ \mathrm{E}(\mathbf{U}\mathbf{U}')=\sigma^2\mathbf{I}_n$〕に注意すると、

$$\begin{aligned}\mathrm{E}((\hat{\boldsymbol{\beta}}-\mathrm{E}(\hat{\boldsymbol{\beta}}))(\hat{\boldsymbol{\beta}}-\mathrm{E}(\hat{\boldsymbol{\beta}}))') &= \mathrm{E}((\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{U})(\mathbf{U}'\mathbf{X}(\mathbf{X}'\mathbf{X})^{-1})) \\ &= \sigma^2(\mathbf{X}'\mathbf{X})^{-1} \quad \cdots\cdots(2.5)\end{aligned}$$

これから第 i 番目の回帰係数 $\hat{\beta}_i$, $i=1,\cdots,k$ の標本分散 $\mathrm{var}(\hat{\beta}_i)$ は、a_{ii} を $(\mathbf{X}\mathbf{X})^{-1}$ の第 i 対角要素として、$\sigma^2 a_{ii}$ となることがわかる。

2.4 var($\hat{\boldsymbol{\beta}}$) の最小性

推定式 (**2.1**節) は、$\boldsymbol{\beta}$ を $\mathbf{Y}$ の一次式で推定することを表示している。これを少し一般化して $\mathbf{C}\boldsymbol{\beta}$ を $\mathbf{b}=\mathbf{A}\mathbf{Y}$ で推定する場合を考えると、下に示すように $\mathrm{var}(\hat{\boldsymbol{\beta}})\leqq\mathrm{var}(\mathbf{b})$ が成立する。したがって、そのような操作範囲では、$\hat{\boldsymbol{\beta}}$ が最小分散を与えることがわかる。いま $\mathbf{C}$ を $r\times k$ ($r\leqq k$) の行列で固定、$\mathbf{A}$ を $r\times n$ の行列、$\mathbf{b}$ を $\mathbf{C}\boldsymbol{\beta}$ の不偏推定量とすれば、

$$\mathrm{E}(\mathbf{b}) = \mathrm{E}(\mathbf{A}\mathbf{Y}) = \mathrm{E}(\mathbf{A}\mathbf{X}\boldsymbol{\beta}+\mathbf{A}\mathbf{U}) = \mathbf{A}\mathbf{X}\boldsymbol{\beta} = \mathbf{C}\boldsymbol{\beta} \Leftrightarrow \mathbf{A}\mathbf{X} = \mathbf{C} \quad \cdots(2.6)$$

そこで、今度は最後の結果 $\mathbf{A}\mathbf{X}=\mathbf{C}$ を条件として、$\mathrm{var}(\mathbf{b})$ を最小にする $\mathbf{A}$ を探索する：

$$\mathrm{var}(\mathbf{b}) = \mathrm{E}((\mathbf{b}-\mathbf{C}\boldsymbol{\beta})(\mathbf{b}-\mathbf{C}\boldsymbol{\beta})')$$

$$= \mathrm{E}((\mathbf{A}(\mathbf{X}\boldsymbol{\beta}+\mathbf{U})-\mathbf{C}\boldsymbol{\beta})(\mathbf{A}(\mathbf{X}\boldsymbol{\beta}+\mathbf{U})-\mathbf{C}\boldsymbol{\beta})')$$
$$= \mathrm{E}(\mathbf{AUU'A'}) = \sigma^2\mathbf{AA'} \cdots\cdots\cdots(2.7)$$

一方、次の恒等式が成り立つ：

$$\mathbf{AA'} \equiv (\mathbf{C}(\mathbf{X'X})^{-1}\mathbf{X'})(\mathbf{C}(\mathbf{X'X})^{-1}\mathbf{X'})' + (\mathbf{A}-\mathbf{C}(\mathbf{X'X})^{-1}\mathbf{X'})(\mathbf{A}-\mathbf{C}(\mathbf{X'X})^{-1}\mathbf{X'})' \cdots\cdots\cdots(2.8)$$

これから $\mathbf{AA'}$、したがって $\mathrm{var}(\mathbf{b})=\sigma^2\mathbf{AA'}$ の対角要素が最小になるのは、$\mathbf{A}=\mathbf{C}(\mathbf{X'X})^{-1}\mathbf{X'}$ の場合であることがわかる。特に、$\mathbf{C}=\mathbf{I}$ のとき、題記命題を証明している。

上記期待値の不偏性（**2.2**節）と $\mathrm{var}(\hat{\beta})$ の最小性から、$\hat{\beta}$ は最良線型不偏推定量（best linear unbiased estimator (b.l.u.e.)）である ——Gauss-Markov の定理。

2.5 $\hat{\beta}$ は ee′ の最小点[19)]

極値条件から求めた $\hat{\beta}$ が、最小の残差平方和を与えることを確かめる。いま、$\hat{\beta}$, $\mathbf{X'X}$, $\mathbf{X'Y}$, $\mathbf{Y'Y}/2$ を見やすいようにそれぞれ $\mathbf{x}$, $\mathbf{H}$, $\mathbf{b}$, C と書いて、$\mathbf{e'e}/2=\mathrm{f}(\mathbf{x})$ と置くと、(2.1) 式から

$$\mathrm{f}(\mathbf{x}) = \frac{1}{2}\mathbf{x'Hx} - \mathbf{b'x} + \mathrm{C} \cdots\cdots\cdots(2.9)$$

$\hat{\beta}$ に対応する $\frac{\partial}{\partial \mathbf{x}}\mathrm{f}(\mathbf{x})=\mathbf{0}$ の根 $\hat{\mathbf{x}}$ は、$\mathbf{H}\hat{\mathbf{x}}=\mathbf{b}$ を満たすから、少し計算すれば、

$$\mathrm{f}(\mathbf{x}) = \frac{1}{2}(\mathbf{x}-\hat{\mathbf{x}})'\mathbf{H}(\mathbf{x}-\hat{\mathbf{x}}) + \hat{\mathrm{C}},\ \mathbf{x}\epsilon\mathbf{R}^{k},\ \hat{\mathrm{C}} = -\frac{1}{2}\mathbf{b}'\hat{\mathbf{x}} + \mathrm{C} \cdots(2.10)$$

いま $\mathbf{H}$ の固有値と、それに対応する固有ベクトルの列を $(\lambda_i, \mathbf{v}_i)_{i=1,\cdots,k}$, $\mathbf{v}'_i\mathbf{v}_j=\delta_{ij}$ として対角行列 $\Lambda=\mathrm{diag}(\lambda_1\cdots,\lambda_k)$、直交行列＊ $\mathbf{V}=[\mathbf{v}_1,\cdots,\mathbf{v}_k]$ を考えると $\mathbf{HV}=\mathbf{V}\Lambda$, $\mathbf{V'}=\mathbf{V}^{-1}$ が成り立つ。これらにあわせ、変数 $\mathbf{x}$ に変数 $\mathbf{z}=\mathbf{V}'(\mathbf{x}-\hat{\mathbf{x}})$ を施して変形すると、

$$\tilde{\mathrm{f}}(\mathbf{z}) \equiv \mathrm{f}(\mathbf{Vz}+\hat{\mathbf{x}}) = \frac{1}{2}\mathbf{z'V'HVz} + \hat{\mathrm{C}} = \frac{1}{2}\mathbf{z'V'V}\Lambda\mathbf{z} + \hat{\mathrm{C}}$$
$$= \frac{1}{2}\mathbf{z}'\Lambda\mathbf{z} + \hat{\mathrm{C}} \cdots\cdots\cdots(2.11)$$

または、

$$\tilde{f}(\mathbf{z}) = \frac{1}{2}\sum_{i=1}^{\kappa} \lambda_i z_i^2 + \hat{C},\ \mathbf{z}\in\mathbf{R}^{\kappa} \quad \cdots\cdots (2.12)$$

$\mathbf{H}$は正定値（対称）**だから $\forall_i(\lambda_i>0)$***。したがって、$\tilde{f}(\mathbf{z})$ の値域は $[\hat{C},\infty)$ となり、$\mathbf{z}=\mathbf{0}$ が最小値 $\tilde{f}(\mathbf{0})=\hat{C}$ を与える点である。

変数変換 $\mathbf{z}=\mathbf{V}'(\mathbf{x}-\hat{\mathbf{x}})$ の基では $\tilde{f}(\mathbf{z})=f(\mathbf{x})$ だから、$\hat{\mathbf{x}}(\mathbf{z}=\mathbf{V}'(\mathbf{x}-\hat{\mathbf{x}})=\mathbf{0})$ は $f(\mathbf{x})$ の最小値 $\hat{C}$ を与える。$\hat{\mathbf{x}}=\mathbf{H}^{-1}\mathbf{b}$ を元にもどすならば、

$$\hat{\mathbf{x}} = \mathbf{H}^{-1}\mathbf{b} = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{Y} = \hat{\beta} \quad \cdots\cdots (2.13)$$

となる。

* $\mathbf{v}'\ \mathbf{v}=(\delta_{ij})$, i, j=1,…,k

** 3つ目の前提条件から説明変数は互いに独立だから、$\hat{\beta}\neq\mathbf{0}$ のとき $\mathbf{X}\hat{\beta}\neq\mathbf{0}$
したがって、$\hat{\beta}'\mathbf{X}'\mathbf{X}\hat{\beta}>0$.
また、$\mathbf{H}'=(\mathbf{X}'\mathbf{X})'=\mathbf{X}'\mathbf{X}=\mathbf{H}$ だから$\mathbf{H}$は対称。

*** 実対称行列の固有値は実数、さらに正定値ならば正。

3 残　　差[12)]

3.1 残差の和

式(2.2)の右辺を $\hat{\mathbf{Y}}=\mathbf{X}\hat{\beta}$ に注意して変形すれば、次式が得られる。

$$\mathbf{X}'\mathbf{Y}-\mathbf{X}'\mathbf{X}\hat{\beta} = \mathbf{X}'(\mathbf{Y}-\hat{\mathbf{Y}}) = \mathbf{X}'\mathbf{e} \quad \cdots\cdots (3.1)$$

極値条件から $\mathbf{X}'\mathbf{e}=\mathbf{0}$. この等式の1行目（$\mathbf{X}'$ の1行目は $\mathbf{1}'$）は、残差の和が 0（零）であることを示している。

3.2 推定値の算術平均

残差の和から、

$$\begin{aligned} 0 &= \mathbf{1}'\mathbf{e} = \mathbf{1}'(\mathbf{Y}-\hat{\mathbf{Y}}) = \mathbf{1}'\mathbf{Y} - \mathbf{1}'\hat{\mathbf{Y}} \\ &= n\bar{Y} - n\bar{\hat{Y}} = n(\bar{Y}-\bar{\hat{Y}}) \quad \cdots\cdots (3.2) \end{aligned}$$

これから、観測値$\mathbf{Y}$と推定式$\hat{\mathbf{Y}}$の算術平均は等しいことがわかる。

3.3 残差平方和の期待値

まず、$\mathbf{e}=(\mathbf{I}_n-\mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}')\mathbf{U}$ を示すことができる。残差の式 $\mathbf{e}=\mathbf{Y}-\hat{\mathbf{Y}}$ から出発し、$\mathbf{Y}=\mathbf{X}\boldsymbol{\beta}+\mathbf{U}$, $\hat{\mathbf{Y}}=\mathbf{X}\hat{\boldsymbol{\beta}}$, $\hat{\boldsymbol{\beta}}=(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{Y}$ などを使って変形すれば、次の通りである。

$$\begin{aligned}\mathbf{e} &= \mathbf{Y}-\hat{\mathbf{Y}} = \mathbf{Y}-\mathbf{X}\hat{\boldsymbol{\beta}} = (\mathbf{X}\boldsymbol{\beta}+\mathbf{U})-\mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'(\mathbf{X}\boldsymbol{\beta}+\mathbf{U}) \\ &= \mathbf{U}-\mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{U} = (\mathbf{I}_n-\mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}')\mathbf{U} \quad \cdots\cdots(3.3)\end{aligned}$$

ここで、最後の括弧内の行列 $\mathbf{I}_n-\mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}$ は対称、べき等であることが示されるから*、

$$\mathbf{e}'\mathbf{e} = \mathbf{U}'(\mathbf{I}_n-\mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}')\mathbf{U} \quad \cdots\cdots(3.4)$$

両辺の期待値を取って**、

$$\begin{aligned}\mathrm{E}(\mathbf{e}'\mathbf{e}) &= \mathrm{E}(\mathbf{U}'(\mathbf{I}_n-\mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}')\mathbf{U}) \\ &= \sigma^2\mathrm{tr}[\mathbf{I}_n-\mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'] \\ &= \sigma^2\{\mathrm{tr}\ \mathbf{I}_n-\mathrm{tr}[\mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}']\} \\ &= \sigma^2\{\mathrm{n}-\mathrm{tr}[(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{X}]\} \\ &= \sigma^2(\mathrm{n}-\mathrm{k}) \quad \cdots\cdots(3.5)\end{aligned}$$

* $(\mathbf{I}_n-\mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}')'=\mathbf{I}_n-\mathbf{X}((\mathbf{X}'\mathbf{X})^{-1})'\mathbf{X}'=\mathbf{I}_n-\mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'$ だから対称。

また、$(\mathbf{I}_n-\mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}')(\mathbf{I}_n-\mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}')$ を配分則で展開し、簡略すると、

$$\begin{aligned}&\mathbf{I}_n\mathbf{I}_n-\mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{I}_n-\mathbf{I}_n\mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'+\mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}' \\ &=\mathbf{I}_n-\mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\end{aligned}$$

したがって、$\mathbf{I}_n-\mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'$ はべき等である。

** $\mathbf{I}_n-\mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'$ を簡単に $\mathrm{A}=(\mathrm{a}_{ij})$ と書き、1つ目の前提条件を使えば、

$$\begin{aligned}\mathrm{E}(\mathbf{e}'\mathbf{e})&=\mathrm{E}(\mathbf{U}'\mathrm{A}\mathbf{U})=\mathrm{E}(\textstyle\sum_j\sum_i u_i a_{ij} u_j)=\sum_j\sum_i a_{ij}\mathrm{E}(u_i u_j) \\ &=\sigma^2\textstyle\sum_i a_{ii}=\sigma^2\mathrm{tr}\ \mathrm{A}\end{aligned}$$

また、$\mathrm{B}=(\mathrm{b}_{ij})$、$\mathrm{C}=(\mathrm{c}_{in})$ の積 BC、CB が共に意味をもてば、

$$\mathrm{tr}(\mathrm{BC})=\textstyle\sum_i\sum_\alpha b_{i\alpha}c_{\alpha i}=\sum_\alpha\sum_i c_{\alpha i}b_{i\alpha}=\mathrm{tr}(\mathrm{CB})$$

したがって、$\mathbf{e}'\mathbf{e}/(\mathrm{n}-\mathrm{k})$ は σ^2 の不偏推定量である。なお、この値は分散分析表の“回帰からの（残差）平均平方”の値である。

4　種々の性質

4.1　平方和の分割[11)]

全平方和 SS_T、回帰による平方和 SS_D、回帰からの平方和 SS_R は、それぞれ次のように表される。

$$SS_T = (\mathbf{Y}-\bar{Y}\mathbf{1}_n)'(\mathbf{Y}-\bar{Y}\mathbf{1}_n) = \mathbf{Y}'\mathbf{Y} - \bar{Y}\mathbf{1}_n'\mathbf{Y} - \bar{Y}\mathbf{Y}'\mathbf{1}_n + \bar{Y}\bar{Y}\mathbf{1}_n'\mathbf{1}_n$$
$$= \mathbf{Y}'\mathbf{Y} - n(\bar{Y})^2 \quad \cdots\cdots(4.1)$$

$\hat{\mathbf{Y}}=\mathbf{X}\hat{\boldsymbol{\beta}}$, $\hat{\boldsymbol{\beta}}=(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{Y}$, $\bar{\hat{Y}}=\bar{Y}$（**2.1** 節の推定式と **3.2** 節の推定値の算術平均参照）などに注意して、

$$SS_D = (\hat{\mathbf{Y}}-\bar{Y}\mathbf{1}_n)'(\hat{\mathbf{Y}}-\bar{Y}\mathbf{1}_n) = \hat{\mathbf{Y}}'\hat{\mathbf{Y}} - n(\bar{Y})^2 = \hat{\boldsymbol{\beta}}'\mathbf{X}'\mathbf{X}\hat{\boldsymbol{\beta}} - n(\bar{Y})^2$$
$$= \hat{\boldsymbol{\beta}}'\mathbf{X}'\mathbf{Y} - n(\bar{Y})^2 \quad \cdots\cdots(4.2)$$

同様に、$\hat{\mathbf{Y}}=\mathbf{X}\hat{\boldsymbol{\beta}}$, $\hat{\boldsymbol{\beta}}=(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{Y}$. また、$\hat{\boldsymbol{\beta}}'\mathbf{X}'\mathbf{Y}$ はスカラーだから $\hat{\boldsymbol{\beta}}'\mathbf{X}'\mathbf{Y}=\mathbf{Y}'\mathbf{X}\hat{\boldsymbol{\beta}}$ などに注意して、

$$SS_R = (\mathbf{Y}-\hat{\mathbf{Y}})'(\mathbf{Y}-\hat{\mathbf{Y}}) = \mathbf{Y}'\mathbf{Y} - \hat{\boldsymbol{\beta}}'\mathbf{X}'\mathbf{Y} - \mathbf{Y}'\mathbf{X}\hat{\boldsymbol{\beta}} + \hat{\boldsymbol{\beta}}'\mathbf{X}'\mathbf{X}\hat{\boldsymbol{\beta}}$$
$$= \mathbf{Y}'\mathbf{Y} - \hat{\boldsymbol{\beta}}'\mathbf{X}'\mathbf{Y} \quad \cdots\cdots(4.3)$$

したがって、

$$SS_D+SS_R = (\hat{\boldsymbol{\beta}}'\mathbf{X}'\mathbf{Y}-n(\bar{Y})^2)+(\mathbf{Y}'\mathbf{Y}-\hat{\boldsymbol{\beta}}'\mathbf{X}'\mathbf{Y})$$
$$= \mathbf{Y}'\mathbf{Y} - n(\bar{Y})^2 = SS_T \quad \cdots\cdots(4.4)$$

4.2　重相関係数と決定係数[17)]

相関係数が変数を偏差で表しても変わらないこと * から、扱いやすいように $\mathbf{X}$, $\mathbf{Y}$ の代りに、それぞれ平均からの偏差 $\mathbf{x}$, $\mathbf{y}$ を使って、$\sqrt{R^2}=\hat{\rho}(\mathbf{y}, \hat{\mathbf{y}})$ を確かめる。ここで、$\hat{\mathbf{y}}=\mathbf{x}\hat{\boldsymbol{\beta}}$.

関係式が $\mathbf{y}=\mathbf{x}\hat{\boldsymbol{\beta}}+\mathbf{e}$ となっても、以前と同じようにして類似の結果が得られる** :

$$\hat{\boldsymbol{\beta}} = (\mathbf{x}'\mathbf{x})^{-1}\mathbf{x}'\mathbf{y} \quad \cdots\cdots(4.5)$$

$$\hat{\boldsymbol{\beta}}'\mathbf{x}'\mathbf{x}\hat{\boldsymbol{\beta}} = \hat{\boldsymbol{\beta}}'\mathbf{x}'\mathbf{y} \quad \cdots\cdots(4.6)$$

$$\bar{\hat{y}} = \bar{y} = 0 \quad \cdots\cdots (4.7)$$

$$SS_T = \mathbf{y}'\mathbf{y} \quad \cdots\cdots (4.8)$$

$$SS_D = \hat{\beta}'\mathbf{x}'\mathbf{y} \quad \cdots\cdots (4.9)$$

$$SS_R = \mathbf{y}'\mathbf{y} - \hat{\beta}'\mathbf{x}'\mathbf{y} \quad \cdots\cdots (4.10)$$

$$SS_T = SS_D + SS_R \quad \cdots\cdots (4.11)$$

これら一連の結果を使えば、

$$R^2 = \frac{SS_D}{SS_T} = \frac{\hat{\beta}'\mathbf{x}'\mathbf{y}}{\mathbf{y}'\mathbf{y}} \quad \cdots\cdots (4.12)$$

したがって、

$$\sqrt{R^2} = \sqrt{\frac{\hat{\beta}'\mathbf{x}'\mathbf{y}}{\mathbf{y}'\mathbf{y}}} \quad \cdots\cdots (4.13)$$

一方、

$$\rho(\mathbf{y},\hat{\mathbf{y}}) = \frac{\mathbf{y}'\hat{\mathbf{y}}}{\sqrt{\mathbf{y}'\mathbf{y}}\sqrt{\hat{\mathbf{y}}'\hat{\mathbf{y}}}} = \frac{\mathbf{y}'\mathbf{x}\hat{\beta}}{\sqrt{\mathbf{y}'\mathbf{y}}\sqrt{\hat{\beta}'\mathbf{x}'\mathbf{x}\hat{\beta}}}$$

$$= \frac{(\sqrt{\hat{\beta}'\mathbf{x}'\mathbf{y}})^2}{\sqrt{\mathbf{y}'\mathbf{y}}\sqrt{\hat{\beta}'\mathbf{x}'\mathbf{y}}} = \sqrt{\frac{\hat{\beta}'\mathbf{x}'\mathbf{y}}{\mathbf{y}'\mathbf{y}}} \quad \cdots\cdots (4.14)$$

上記 $\sqrt{R^2}$, $\hat{\rho}(\mathbf{y}, \hat{\mathbf{y}})$ の最後の2式を比較すれば、$\sqrt{R^2} = \hat{\rho}(\mathbf{Y}, \hat{\mathbf{Y}})$ が得られる。

* $\mathbf{Z}_1$, $\mathbf{Z}_2$ を期待値 $E(\mathbf{Z}_1)$, $E(\mathbf{Z}_2)$ からの偏差で表したものを $\mathbf{z}_1 = \mathbf{Z}_1 - E(\mathbf{Z}_1)$, $\mathbf{z}_2 = \mathbf{Z}_2 - E(\mathbf{Z}_2)$ と置くと $E(\mathbf{z}_1) = E(\mathbf{z}_2) = \mathbf{0}$。したがって、共分散の式は、

$$\mu(\mathbf{z}_1,\mathbf{z}_2) = E((\mathbf{Z}_1 - E(\mathbf{Z}_1))(\mathbf{Z}_2 - E(\mathbf{Z}_2))) = \mu(\mathbf{Z}_1,\mathbf{Z}_2)$$

また、この等式から $\sigma_i^2 = \mu(\mathbf{z}_i,\mathbf{z}_i) = \mu(\mathbf{Z}_i,\mathbf{Z}_i)$, $i=1, 2$ がわかるから、

$$\rho(\mathbf{z}_1,\mathbf{z}_2) = \frac{\mu(\mathbf{z}_1,\mathbf{z}_2)}{\sigma_1\sigma_2} = \frac{\mu(\mathbf{Z}_1,\mathbf{Z}_2)}{\sigma_1\sigma_2} = \rho(\mathbf{Z}_1,\mathbf{Z}_2)$$

測定値から $\hat{\rho}$ を求める場合は、期待値の代りに算術平均、分散の代りに標本分散（平均からの偏差二乗和の算術平均）を考えればよい。

** $\mathbf{Y} = \mathbf{X}\boldsymbol{\beta} + \mathbf{U} = \mathbf{X}\hat{\boldsymbol{\beta}} + \mathbf{e}$ の平均を取る（各式の左から $\frac{1}{n}\mathbf{1}_n'$ を掛ける）と、$\mathbf{1}_n'\mathbf{e} = \mathbf{0}$ となる（式(3.2)）から、

$$\bar{\mathrm{Y}} = (\bar{\mathrm{X}}_1,\cdots,\bar{\mathrm{X}}_k)\boldsymbol{\beta} + \bar{\mathrm{U}} = (\bar{\mathrm{X}}_1,\cdots,\bar{\mathrm{X}}_k)\hat{\beta}$$

したがって、

$$\begin{aligned}\mathbf{Y} - \bar{\mathrm{Y}}\mathbf{1}_n &= (\mathbf{X}\boldsymbol{\beta}+\mathbf{U}) - \mathbf{1}_n(\bar{\mathrm{X}}_1,\cdots,\bar{\mathrm{X}}_n)\boldsymbol{\beta} - \bar{\mathrm{U}}\mathbf{1}_n \\ &= (\mathbf{X}\hat{\boldsymbol{\beta}}+\mathbf{e}) - \mathbf{1}_n(\bar{\mathrm{X}}_1,\cdots,\bar{\mathrm{X}}_n)\hat{\boldsymbol{\beta}}\end{aligned}$$

$\mathbf{y}$, $\mathbf{x}$ を $\mathbf{Y}$, $\mathbf{X}$ の平均からの偏差として上式の辺々を書き換えると、

$$\mathbf{y} = \mathbf{x}\boldsymbol{\beta} + \mathbf{U} - \bar{\mathrm{U}}\mathbf{1}_n = \mathbf{x}\hat{\boldsymbol{\beta}} + \mathbf{e} \quad \cdots\cdots(*)$$

$\hat{\beta}$ の求め方は以前（**2.1**節）と同じで良いから、

$$\hat{\boldsymbol{\beta}} = (\mathbf{x}'\mathbf{x})^{-1}\mathbf{x}'\mathbf{y}$$

これから、

$$\hat{\boldsymbol{\beta}}'\mathbf{x}'\mathbf{x}\hat{\boldsymbol{\beta}} = \hat{\boldsymbol{\beta}}'\mathbf{x}'\mathbf{x}((\mathbf{x}'\mathbf{x})^{-1}\mathbf{x}'\mathbf{y}) = \boldsymbol{\beta}'\mathbf{x}'\mathbf{y}$$

$\hat{\mathbf{y}}$ は $\mathbf{x}\hat{\boldsymbol{\beta}}$ の別表示だから上の (*) より $\mathbf{y}-\hat{\mathbf{y}}=\mathbf{e}$ これについても、**3.1**節で $\mathbf{Y}$, $\mathbf{X}$ に $\mathbf{y}$, $\mathbf{x}$ を代用すれば、

$$\begin{aligned}&\bar{\hat{\mathrm{y}}}=\bar{\mathrm{y}} \quad \text{また、} \mathrm{n}\bar{\mathrm{y}}=\mathbf{1}_n'\mathbf{y}=\mathbf{1}_n'(\mathbf{Y}-\bar{\mathrm{Y}}\mathbf{1}_n)=\mathbf{1}_n'\mathbf{Y}-\bar{\mathrm{Y}}\mathbf{1}_n'\mathbf{1}_n \\ &\quad =\mathrm{n}\bar{\mathrm{Y}}-\mathrm{n}\bar{\mathrm{Y}}=0\end{aligned}$$

さらに、ここまでの結果を使えば、

$$\begin{aligned}\mathrm{SS_T} &= (\mathbf{y}-\bar{\mathrm{y}}\mathbf{1}_n)'(\mathbf{y}-\bar{\mathrm{y}}\mathbf{1}_n) = \mathbf{y}'\mathbf{y} \\ \mathrm{SS_D} &= (\hat{\mathbf{y}}-\bar{\hat{\mathrm{y}}}\mathbf{1}_n)'(\hat{\mathbf{y}}-\bar{\hat{\mathrm{y}}}\mathbf{1}_n) = \hat{\mathbf{y}}'\hat{\mathbf{y}} = \hat{\boldsymbol{\beta}}'\mathbf{x}'\mathbf{x}\hat{\boldsymbol{\beta}} = \hat{\boldsymbol{\beta}}'\mathbf{x}'\mathbf{y} \\ \mathrm{SS_R} &= (\mathbf{y}-\hat{\mathbf{y}})'(\mathbf{y}-\hat{\mathbf{y}}) = \mathbf{y}'\mathbf{y} - \mathbf{y}'\hat{\mathbf{y}} - \hat{\mathbf{y}}'\mathbf{y} + \hat{\mathbf{y}}'\hat{\mathbf{y}} \\ &= \mathbf{y}'\mathbf{y} - \hat{\boldsymbol{\beta}}'\mathbf{x}'\mathbf{y}\end{aligned}$$

後の２式を辺々加えると、

$$\mathrm{SS_D} + \mathrm{SS_R} = \hat{\boldsymbol{\beta}}'\mathbf{x}'\mathbf{y} + (\mathbf{y}'\mathbf{y}-\hat{\boldsymbol{\beta}}'\mathbf{x}'\mathbf{y}) = \mathbf{y}'\mathbf{y} = \mathrm{SS_T}$$

特に、説明変数が１個で $\mathbf{x}_2$ だけの場合、$\sqrt{\mathrm{R}^2}=|\hat{\rho}(\mathbf{y},\ \mathbf{x}_2)|$ となって重相関係数Rが、$\mathbf{y}$ と $\mathbf{x}_2$ との相関係数に等しくなる。実際 $\hat{\beta}=(\mathbf{x}_2'\mathbf{x}_2)^{-1}\mathbf{x}_2'\mathbf{y}$ で、説明変数が $\mathbf{x}_2$ だけの場合、ベクトルの積和（内積）$\mathbf{x}_2'\mathbf{x}_2$, $\mathbf{x}_2'\mathbf{y}$ がスカラーになることに注意すれば、次のようになる。

$$\begin{aligned}\sqrt{\mathrm{R}^2} &= \sqrt{\frac{\hat{\boldsymbol{\beta}}'\mathbf{x}_2'\mathbf{y}}{\mathbf{y}'\mathbf{y}}} = \sqrt{\frac{((\mathbf{x}_2'\mathbf{x}_2)^{-1}\mathbf{x}_2'\mathbf{y})'(\mathbf{x}_2'\mathbf{y})}{\mathbf{y}'\mathbf{y}}} = \sqrt{\frac{(\mathbf{x}_2'\mathbf{y})^2}{(\mathbf{x}_2'\mathbf{x}_2)(\mathbf{y}'\mathbf{y})}} \\ &= \frac{|\mathbf{x}_2'\mathbf{y}|}{\sqrt{\mathbf{x}_2'\mathbf{x}_2}\sqrt{\mathbf{y}'\mathbf{y}}} = |\hat{\rho}(\mathbf{y},\mathbf{x}_2)| \quad \cdots\cdots(4.15)\end{aligned}$$

（注）一般に決定係数 R^2 と変数間の相関係数との間には、次のような関係が得られる。（**4.3**節参照）

$$R^2 = 1 - \frac{R}{R_{11}}$$

ここで、

$$\mathbf{R} = \begin{bmatrix} \rho(\mathbf{y},\mathbf{y}) & \rho(\mathbf{y},\mathbf{x}_2) \cdots \rho(\mathbf{y},\mathbf{x}_k) \\ \rho(\mathbf{x}_2,\mathbf{y}) & \rho(\mathbf{x}_2,\mathbf{x}_2) \cdots \rho(\mathbf{x}_2,\mathbf{x}_k) \\ \vdots & \vdots \qquad\quad \vdots \\ \rho(\mathbf{x}_k,\mathbf{y}) & \rho(\mathbf{x}_k,\mathbf{x}_2) \cdots \rho(\mathbf{x}_k,\mathbf{x}_k) \end{bmatrix} = \begin{bmatrix} r_{11} & r_{12} \cdots r_{1k} \\ r_{21} & r_{22} \cdots r_{2k} \\ \vdots & \vdots \quad \vdots \\ r_{k1} & r_{k2} \cdots r_{kk} \end{bmatrix}$$

$R = |\mathbf{R}|$、R_{11} は r_{11} の余因子。

これからもすぐ上で述べたのと同じ結果が得られる。説明変数が $\mathbf{x}_2$ だけの場合、

$$\mathbf{R} = \begin{bmatrix} r_{11} & r_{12} \\ r_{21} & r_{22} \end{bmatrix} = \begin{bmatrix} 1 & r_{12} \\ r_{12} & 1 \end{bmatrix}, \quad R = 1 - {r_{12}}^2, \quad R_{11} = 1$$

したがって、$R^2 = 1 - \dfrac{1 - {r_{12}}^2}{1} = {r_{12}}^2$

4.3 重相関係数の別の表現形式[13,17]

X，**Y**を平均からの偏差で表したものを**x**，**y**（**4.2**節**）。**R**，Rは**4.2**節（注）の通り、また**R**の１行と１列を除いた小行列を $\mathbf{R}_{11}$、**R**での r_{ij} の余因子を R_{ij} とする。

証明の骨子は決定係数の定義式 $R^2 = \dfrac{SS_D}{SS_T} = \dfrac{\hat{\boldsymbol{\beta}}'\mathbf{x}'\mathbf{y}}{\mathbf{y}'\mathbf{y}}$ の右辺を相関係数で表すことである。

まず、$\boldsymbol{\beta}$ に関する最小二乗法の要請（**2.1**節）$\mathbf{x}'\mathbf{x}\hat{\boldsymbol{\beta}} = \mathbf{x}'\mathbf{y}$ から $\mathbf{x}'\mathbf{x}$，$\mathbf{x}'\mathbf{y}$ を相関係数で表した形で $\hat{\boldsymbol{\beta}}$ を求める。

$$n\begin{bmatrix} s_2 & & 0 \\ & \ddots & \\ 0 & & s_k \end{bmatrix}\mathbf{R}_{11}\begin{bmatrix} s_2 & & 0 \\ & \ddots & \\ 0 & & s_k \end{bmatrix}\hat{\boldsymbol{\beta}} = n\begin{bmatrix} s_2 & & 0 \\ & \ddots & \\ 0 & & s_k \end{bmatrix} s_1\mathbf{r} \cdots\cdots\cdots\cdots (4.16)$$

ここで、${s_i}^2$, $i=1,\cdots,k$ は第 i 変数の標本分散、**r** は**x**と**y**との相関ベクトルである。いま、上式の両辺に左から $\dfrac{1}{n}\begin{bmatrix} s_2 & & \\ & \ddots & \\ & & s_k \end{bmatrix}^{-1}$ を掛けると、

$$\mathbf{R}_{11}\begin{bmatrix} s_2 & & 0 \\ & \ddots & \\ 0 & & s_k \end{bmatrix}\hat{\beta} = s_1\mathbf{r} \quad したがって^{*}、\quad \hat{\beta} = \begin{pmatrix} \frac{1}{s_2} & & 0 \\ & \ddots & \\ 0 & & \frac{1}{s_k} \end{pmatrix}\frac{s_1}{R_{11}}\begin{pmatrix} R_{22} & \cdots & R_{k2} \\ \vdots & & \vdots \\ R_{2k} & \cdots & R_{kk} \end{pmatrix}\mathbf{r} \quad \cdots\cdots(4.17)$$

右辺の余因子行列とベクトル $\mathbf{r}$ との積を“異なる行の余因子を用いた展開”の式、

$$r_{11}R_{1i} + (r_{21}R_{2i} + \cdots + r_{k1}R_{ki}) \equiv 0, \quad i = 2, \cdots, k \quad \cdots\cdots(4.18)$$

に注意して書き換えると、

$$\hat{\beta} = \frac{s_1}{R_{11}}\begin{pmatrix} \frac{1}{s_2} & & 0 \\ & \ddots & \\ 0 & & \frac{1}{s_k} \end{pmatrix}\begin{pmatrix} -R_{12} \\ \vdots \\ -R_{1k} \end{pmatrix} \quad \cdots\cdots(4.19)$$

したがって、

$$R^2 = \frac{\hat{\beta}'\mathbf{x}'\mathbf{y}}{\mathbf{y}'\mathbf{y}} = \frac{-s_1}{R_{11}}(R_{12} \cdots R_{1k})\begin{pmatrix} \frac{1}{s_2} & & 0 \\ & \ddots & \\ 0 & & \frac{1}{s_k} \end{pmatrix}\frac{n\begin{pmatrix} s_2 & & 0 \\ & \ddots & \\ 0 & & s_k \end{pmatrix}s_1\mathbf{r}}{n\,{s_1}^2}$$

$$= \frac{-1}{R_{11}}(R_{12} \cdots R_{1k})\mathbf{r} \quad \cdots\cdots(4.20)$$

再び、Rの展開式

$$R = r_{11}R_{11} + (r_{12}R_{12} + \cdots + r_{1k}R_{1k}) \quad \cdots\cdots(4.21)$$

から

$$R^2 = -\frac{1}{R_{11}}(R - R_{11}) = 1 - \frac{R}{R_{11}} \quad \cdots\cdots(4.22)$$

* $\mathbf{R}_{11}^{-1}$ は $\mathbf{R}_{11}$ の余因子行列（r_{ij} をその余因子 R_{ij} で置き換えた行列）を転置し、それを $\mathbf{R}_{11}$ の行列式 R_{11}で割ったものである。

4.4 平面回帰直線の交角[14,18]

(X, Y) が2次元の正規分布 $N(m_1, m_2, {\sigma_1}^2, {\sigma_2}^2, \rho)$ にしたがって分布するとき、

$$y \mapsto E(X|Y=y)\ \text{は直線で}\ x=\mathbf{m_1}+\frac{\sigma_1\rho}{\sigma_2}(y-m_2) \cdots\cdots\cdots(4.23)$$

$$x \mapsto E(Y|X=x)\ \text{は直線で}\ y=m_2+\frac{\sigma_2\rho}{\sigma_1}(y-m_1) \cdots\cdots\cdots(4.24)$$

いま、**Fig. 1** のように前者がｘのｙへの回帰直線、後者がｙのｘへの回帰直線である。

両者の交角を θ とし、

$$\tan\theta_1=\frac{\sigma_2}{\sigma_1\rho} \cdots\cdots\cdots(4.25)$$

$$\tan\theta_2=\frac{\sigma_2\rho}{\sigma_1} \cdots\cdots\cdots(4.26)$$

と置けば、

$$\tan\theta=\tan(\theta_1-\theta_2)=\frac{\tan\theta_1-\tan\theta_2}{1+\tan\theta_1\tan\theta_2}$$

$$=\frac{\sigma_1\sigma_2}{\sigma_1{}^2+\sigma_2{}^2}\cdot\frac{1-\rho^2}{\rho} \cdots\cdots\cdots(4.27)$$

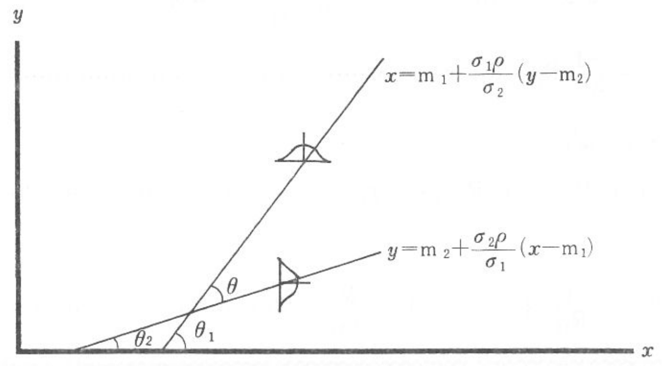

Fig. 1　Location and Angle of Two Simple Linear Regression Lines
平面回帰直線の交角に関する図形

この式から、次のような質的評価をすることができる。

ＸとＹが完全に比例すれば $\rho=1$ したがって、両回帰直線は重なる。

ＸとＹが殆ど独立ならば $\rho\to 0$ したがって、$\theta\to\pm\frac{\pi}{2}$ で両回帰直線は略々直交する。

$0<|\rho|<1$ のとき、これらの中間で、

$$\theta = \tan^{-1}\left(\frac{\sigma_1\sigma_2}{\sigma_1{}^2+\sigma_2{}^2}\cdot\frac{1-\rho^2}{\rho}\right) \quad \cdots\cdots(4.28)$$

4.5 平面回帰直線の式[14]

p 変量正規分布の密度関数 (density function of a p–variate distribution) は、分散共分散行列 Σ を p 次の正定値行列として、[15]

$$f(x_1,x_2,\cdots,x_p)=f(\mathbf{x})=(2\pi)^{-\frac{p}{2}}|\Sigma|^{-\frac{1}{2}}e^{-\frac{1}{2}(\mathbf{x}-\mu)'\Sigma^{-1}(\mathbf{x}-\mu)} \quad \cdots\cdots(4.29)$$

特に、2 変量（X_1, X_2）の場合、

$$\Sigma = E\begin{bmatrix}(x_1-\mu_1)^2 & (x_1-\mu_1)(x_2-\mu_2)\\ (x_2-\mu_2)(x_1-\mu_1) & (x_2-\mu_2)^2\end{bmatrix}=\begin{bmatrix}\sigma_{11} & \sigma_{12}\\ \sigma_{21} & \sigma_{22}\end{bmatrix}=\begin{bmatrix}\sigma_1{}^2 & \sigma_1\sigma_2\rho\\ \sigma_2\sigma_1\rho & \sigma_2{}^2\end{bmatrix} \quad \cdots\cdots(4.30)$$

$$\Sigma^{-1} = \frac{1}{1-\rho^2}\begin{pmatrix}\dfrac{1}{\sigma_1{}^2} & -\dfrac{\rho}{\sigma_1\sigma_2}\\ -\dfrac{\rho}{\sigma_2\sigma_1} & \dfrac{1}{\sigma_2{}^2}\end{pmatrix} \quad \cdots\cdots(4.31)$$

したがって、

$$f(x_1,\ x_2) = \frac{1}{2\pi\sigma_1\sigma_2\sqrt{1-\rho^2}}\exp\left\{-\frac{1}{2(1-\rho^2)}\times\left[\frac{(x_1-\mu_1)^2}{\sigma_1{}^2}-2\rho\frac{(x_1-\mu_1)(x_2-\mu_2)}{\sigma_1\sigma_2}+\frac{(x_2-\mu_2)^2}{\sigma_2{}^2}\right]\right\} \quad \cdots\cdots(4.32)$$

X_2 が値 x_2 を取る場合、X_1 の条件付密度 (conditional density of X_1 given x_2) および X_1 が値 x_1 を取る場合の X_2 の条件付密度は、それぞれ

$$g(x_1|X_2=x_2) = \frac{f(x_1,\ x_2)}{h(x_2)} \quad \cdots\cdots(4.33)$$

$$h(x_2|X_1=x_1) = \frac{f(x_1,\ x_2)}{g(x_1)} \quad \cdots\cdots(4.34)$$

ここで、$g(x_1)$ は X_1 の周辺密度 (marginal density of X_1)、$h(x_2)$ は X_2 の周辺密度である。$f(x_1,\ x_2)$ を x_2 につき、全積分区間にわたって積分すれば

$g(x_1)$、x_1 につき積分すれば $h(x_2)$ が得られる：

$$g(x_1)=\frac{1}{\sqrt{2\pi}\sigma_1}e^{-\frac{(x_1-\mu_1)^2}{2\sigma_1{}^2}} \quad\cdots\cdots(4.35)$$

$$h(x_2)=\frac{1}{\sqrt{2\pi}\sigma_2}e^{-\frac{(x_2-\mu_2)^2}{2\sigma_2{}^2}} \quad\cdots\cdots(4.36)$$

したがって、

$$g(x_1|X_2=x_2)=\frac{1}{\sigma_1\sqrt{2\pi(1-\rho^2)}}\times\exp\left\{-\frac{1}{2\sigma_1{}^2(1-\rho^2)}\left((x_1-\mu_1)-\frac{\rho\sigma_1}{\sigma_2}(x_2-\mu_2)\right)^2\right\} \quad\cdots\cdots(4.37)$$

$$h(x_2|X_1=x_1)=\frac{1}{\sigma_2\sqrt{2\pi(1-\rho^2)}}\times\exp\left\{-\frac{1}{2\sigma_2{}^2(1-\rho^2)}\left((x_2-\mu_2)-\frac{\rho\sigma_2}{\sigma_1}(x_1-\mu_1)\right)^2\right\} \quad\cdots\cdots(4.38)$$

これらは、それぞれ

$$N\left(\mu_1+\frac{\rho\sigma_1}{\sigma_2}(x_2-\mu_2),\ \sigma_1{}^2(1-\rho^2)\right),$$

$$N\left(\mu_2+\frac{\rho\sigma_2}{\sigma_1}(x_1-\mu_1),\ \sigma_2{}^2(1-\rho^2)\right)$$

となるから回帰直線は、

$$x_2 \mapsto E(X_1|X_2=x_2) \text{ すなわち } x_2 \mapsto \mu_1+\frac{\rho\sigma_1}{\sigma_2}(x_2-\mu_2) \quad\cdots\cdots(4.39)$$

$$x_1 \mapsto E(X_2|X_1=x_1) \text{ すなわち } x_1 \mapsto \mu_2+\frac{\rho\sigma_2}{\sigma_1}(x_1-\mu_1) \quad\cdots\cdots(4.40)$$

4.6 生データと偏差データ[12]

生データと平均値からの偏差データでは、これらを処理する式の形に微妙な違いが生じる。したがって、これらを対比して次表にまとめた。

生データの場合

$$\mathbf{Y}=\begin{bmatrix} Y_1 \\ Y_2 \\ \vdots \\ \vdots \\ Y_n \end{bmatrix} \quad \mathbf{X}=\begin{bmatrix} 1 & X_{21} & \cdots\cdots & X_{k1} \\ 1 & X_{22} & \cdots\cdots & X_{k2} \\ \vdots & \vdots & & \vdots \\ \vdots & \vdots & & \vdots \\ 1 & X_{2n} & \cdots\cdots & X_{kn} \end{bmatrix} \quad \boldsymbol{\beta}=\begin{bmatrix} \beta_1 \\ \beta_2 \\ \vdots \\ \vdots \\ \beta_k \end{bmatrix}$$

$$\hat{\beta}=\begin{bmatrix} \hat{\beta}_1 \\ \hat{\beta}_2 \\ \vdots \\ \vdots \\ \beta_k \end{bmatrix} \quad \mathbf{u}=\begin{bmatrix} u_1 \\ u_2 \\ \vdots \\ \vdots \\ u_n \end{bmatrix} \quad \mathbf{e}=\begin{bmatrix} e_1 \\ e_2 \\ \vdots \\ \vdots \\ e_n \end{bmatrix}$$

$$\mathbf{Y}=\mathbf{X}\boldsymbol{\beta}+\mathbf{U}=\mathbf{X}\hat{\beta}+\mathbf{e}$$

$\boldsymbol{\beta}$ の最小二乗推定量は、

$$\hat{\beta}=(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{Y}$$

で、その分散-共分散行列は、

$$\mathrm{var}(\hat{\beta})=\sigma^2(\mathbf{X}'\mathbf{X})^{-1}$$

と表される。これは $\boldsymbol{\beta}$ の最良線型不偏推定量となっている。

σ^2 の不偏推定量は、

$$\hat{s}^2=\mathbf{e}'\mathbf{e}/(n-k)$$

で与えられる。また、重相関係数は

$$R^2_{1\cdot 23\cdots k}=\frac{\hat{\beta}'\mathbf{X}'\mathbf{Y}-\frac{1}{n}(\sum Y)^2}{\mathbf{Y}'\mathbf{Y}-\frac{1}{n}(\sum Y)^2}$$

仮説：$\beta_2=\cdots\cdots=\beta_k=0$ を検定するには、*

$$F=\frac{R^2/(k-1)}{(1-R^2)/(n-k)}$$

が自由度（$k-1$, $n-k$）のF分布にしたがうことを利用すればよい。また、

$$\beta_i=0 \qquad i=1,\cdots\cdots,k$$

偏差で表現した場合

$$\mathbf{Y}=\begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ \vdots \\ y_n \end{bmatrix} \quad \mathbf{X}=\begin{bmatrix} x_{21} & \cdots\cdots & x_{k1} \\ x_{22} & \cdots\cdots & x_{k2} \\ \vdots & & \vdots \\ \vdots & & \vdots \\ x_{2n} & \cdots\cdots & x_{kn} \end{bmatrix} \quad \boldsymbol{\beta}=\begin{bmatrix} \beta_2 \\ \beta_3 \\ \vdots \\ \vdots \\ \beta_k \end{bmatrix}$$

$$\hat{\beta}=\begin{bmatrix} \hat{\beta}_2 \\ \hat{\beta}_3 \\ \vdots \\ \vdots \\ \hat{\beta}_k \end{bmatrix} \quad \mathbf{u}=\begin{bmatrix} u_1 \\ u_2 \\ \vdots \\ \vdots \\ u_n \end{bmatrix} \quad \bar{\mathbf{u}}=\begin{bmatrix} \bar{u} \\ \bar{u} \\ \vdots \\ \vdots \\ \bar{u} \end{bmatrix} \quad \mathbf{e}=\begin{bmatrix} e_1 \\ e_2 \\ \vdots \\ \vdots \\ e_n \end{bmatrix}$$

$$\mathbf{Y}=\mathbf{X}\boldsymbol{\beta}+\mathbf{u}-\bar{\mathbf{u}}=\mathbf{X}\hat{\beta}+\mathbf{e}$$

最小二乗推定量は、

$$\hat{\beta}=(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{Y}$$

また、

$$\hat{\beta}_1=\bar{Y}-\hat{\beta}_2\bar{X}_2-\cdots\cdots-\hat{\beta}_k\bar{X}_k$$

で与えられる。これらは最良線型不偏推定量、$\hat{\beta}$ の分散-共分散行列は、

$$\mathrm{var}(\hat{\beta})=\sigma^2(\mathbf{X}'\mathbf{X})^{-1}$$

σ^2 の不偏推定量は、

$$\hat{s}^2=\mathbf{e}'\mathbf{e}/(n-k)$$

また、重相相関係数は、

$$R^2_{1\cdot 23\cdots k}=\frac{\hat{\beta}'\mathbf{X}'\mathbf{Y}}{\mathbf{Y}'\mathbf{Y}}$$

仮説：$\beta_2=\cdots\cdots=\beta_k=0$ の検定は、*

$$F=\frac{R^2/(k-1)}{(1-R^2)/(n-k)}$$

が自由度（$k-1$, $n-k$）のF分布をすることを用いる。また、

$$\beta_1=0 \qquad i=2,3,\cdots\cdots,k$$

を単独に検定する場合は、 $t=\dfrac{\hat{\beta}_i-\beta_i}{\hat{s}\sqrt{a_{ii}}}$ が自由度 n−k の t 分布をすることを用いる。ここで a_{ii} は $(\mathbf{X'X})^{-1}$ の第 i 番目の主対角要素を表す。	を単独に検定する場合は、 $t=\dfrac{\hat{\beta}_i-\beta_i}{\hat{s}\sqrt{a_{ii}}}$ が自由度 n−k の t 分布にしたがうことを用いる。ここで a_{ii} は $(\mathbf{X'X})^{-1}$ で X_i に対応する主対角要素を表す。（この場合、a_{ii} が第（i−1）番目の主対角要素となっていることに注意。たとえば、最初の主対角要素が X_2 に対応している。）
β_i（i=1,2,……,k）に対する 100(1−ε)% 信頼区間は、 $\hat{\beta}_i \pm t_{\varepsilon/2}\hat{s}\sqrt{a_{ii}}$ で与えられる。	β_i（i=2,3,……,k）に対する 100(1−ε)% 信頼区間は、 $\hat{\beta}_i \pm t_{\varepsilon/2}\hat{s}\sqrt{a_{ii}}$ となる。
仮説：$\mathbf{c}'\boldsymbol{\beta}=r_0$（ただし、$\mathbf{c}$ は k×1 の列ベクトルとする）の検定は、 $t=\dfrac{\mathbf{c}'\hat{\boldsymbol{\beta}}-r_0}{\hat{s}\sqrt{\mathbf{c}'(\mathbf{X'X})^{-1}\mathbf{c}}}$ が自由度 n−k の t 分布にしたがうことを利用して行うことができる。	仮説：$\mathbf{c}'\boldsymbol{\beta}=r_0$（$\mathbf{c}$ は k×1 の列ベクトル）を検定するには、 $t=\dfrac{\mathbf{c}'\hat{\boldsymbol{\beta}}-r_0}{\hat{s}\sqrt{\mathbf{c}'(\mathbf{X'X})^{-1}\mathbf{c}}}$ が自由度 n−k の t 分布にしたがうことを用いればよい。
$\mathbf{c}'\boldsymbol{\beta}$ に対する 100(1−ε)% 信頼区間は、 $\mathbf{c}'\hat{\boldsymbol{\beta}} \pm t_{\varepsilon/2}\hat{s}\sqrt{\mathbf{c}'(\mathbf{X'X})^{-1}\mathbf{c}}$ となる。	$\mathbf{c}'\boldsymbol{\beta}$ に対する 100(1−ε)% 信頼区間は、 $\mathbf{c}'\boldsymbol{\beta} \pm t_{\varepsilon/2}\hat{s}\sqrt{\mathbf{c}'(\mathbf{X'X})^{-1}\mathbf{c}}$ となる。

＊以降攪乱項 $\mathbf{U}$ が $N(\mathbf{0},\ \sigma^2\mathbf{1}_n)$ に従うことを仮定する。

5　おわりに

この論文は、大阪経済法科大学学生のスポーツテストにおいて、統計処理を行ったプログラムのモデルを作成する手段を取り扱った。プログラムは、株式会社日立製作所の基本統計システム（BASIS）の重回帰プログラムである。ここでは、BASIS の重回帰プログラムを適用する際に発生したいろいろな問題点を解明する目的がある。問題点の解明に端を発した内容は、適用されるパッケージに依存するものではない。

ここで取り上げた諸性質を認識することが、説明変数のレビューに必要である。さらに、モデルの改善にとって有意義であることを提案している。

参 考 文 献

(1) 勝　英雄、沢　　勲「1987年度　大阪経済法科大学学生のスポーツテスト・データ（体格・体力）に関する統計的分析」大阪経済法科大学論集 **41** P.21～84 (1990)
(2) 沢　　勲、大森敏行「1986年度　大阪経済法科大学学生のスポーツテスト・データ（体格・体力）に関する統計的分析」大阪経済法科大学論集 **42** P.5～70 (1990)
(3) 沢　　勲、森下泰行「1985年度　大阪経済法科大学学生のスポーツテスト・データ（体格・体力）に関する統計的分析」大阪経済法科大学論集 **43** P.29～89 (1990)
(4) 沢　　勲、高垣英夫「1984年度　大阪経済法科大学学生のスポーツテスト・データ（体格・体力）に関する統計的分析」大阪経済法科大学論集 **44** P.35～91 (1991)
(5) 沢　　勲、中澄孝司「1983年度　大阪経済法科大学学生における体格診断と体力診断テストに関する統計的分析」大阪経済法科大学論集 **46** P.13～68 (1991)
(6) 沢　　勲、中澄孝司、高垣英夫「大阪経済法科大学学生（1983～1987年）の立位体前屈と踏台昇降運動の体力統計」大阪経済法科大学総合科学研究所年報 **10** P.23～58 (1991)
(7) 沢　　勲、森下泰行、大森敏行「大阪経済法科大学学生（1983～1987年）の垂直跳と背筋力の体力統計」大阪経済法科大学論集 **45** P.41～88 (1991)
(8) 大森敏行、勝　英雄、沢　　勲「コンピュータ処理による体格診断テスト」大阪経済法科大学情報科学センターニュース **7** P.4～6 (1990)
(9) 高垣英夫、沢　　勲「コンピュータ処理による体力診断テストの握力と伏臥上体そらし」大阪経済法科大学情報科学センターニュース **8** P.3～7 (1990)
(10) 中澄孝司、沢　　勲「コンピュータ処理による体力診断テストの立位体前屈と踏台昇降運動」大阪経済法科大学情報科学センターニュース **8** P.7～11 (1990)
(11) N.R. Draper, H. Smith「Applied Regression Analysis」John Wiley & Sons (1981)
(12) J. Johnston「Econometric Methods」McGraw-Hill (1963)（竹内　啓訳「計量経済学の方法」東洋経済新報社　1974）
(13) J. Johnston「Econometric Methods」McGraw-Hill (1972)（竹内　啓他共訳「計量経済学の方法　上」東洋経済新報社　1983）
(14) 河田敬義、丸山文行「数理統計」裳華房 (1977)
(15) T.W. Anderson「An Introduction to Multivariate Statistical Analysis」John Wiley & Sons (1958)
(16) 釆山寛幸「BASIS B02 出力結果に因る関係式に関する事後スクリーニング法」Q.A資料 (1990)

⒄ 釆山寛幸「重相関係数を単相関係数で表わすこと」Q. A資料（1990）
⒅ 釆山寛幸「平面回帰直線の交角の評価」Q. A資料（1990）
⒆ O. Axelsson, V.A. Barker「Finite Element Solution of Boundary Value Problems」Academic Press（1984）
⒇ 沢勲・釆山寛幸・安光烈「HITAC 基本統計システムにおける重回帰プログラムとデータスクリーニング」大阪経済法科大学論集 **46** P. 1～12（1991）