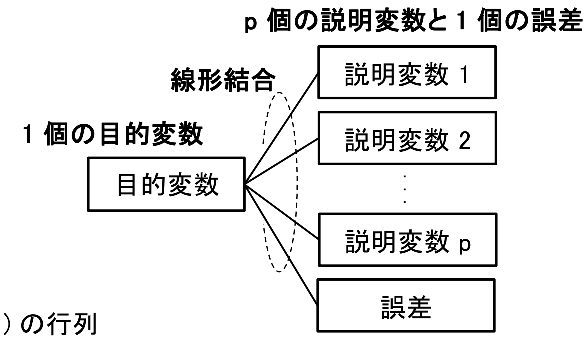
重回帰分析の概要

2023年5月（作成）

倉谷　隆博

1. 重回帰分析の考え方（Multiple regression analysis）

（図1）に示すような、目的変数を複数の説明変数の線形結合（回帰式）で表して、目的変数と説明変数の関係を分析する方法を重回帰分析という。説明変数が1つの場合を単回帰分析、複数の場合を重回帰分析という。



（図1）重回帰分析の構造

重回帰分析で使用する変数を以下にまとめておく。

①説明変数の数：p

②観測データの数：n

③目的変数（既知の確定値）：$\mathbf{y} = [y_1\ \ y_2\ \ \cdots\ \ y_n]^T$　n × 1 のベクトル

④重回帰式を使用して求めた目的変数の推定値：$\hat{\mathbf{y}} = [\hat{y}_i\ \ \hat{y}_2\ \ \cdots\ \hat{y}_n]^T$　n × 1 のベクトル

⑤説明変数（既知の確定値）：$\mathbf{x}_j = [x_{1j}\ \ x_{2j}\ \ \cdots\ \ x_{nj}]^T$　(j = 1, ･･･, p)　n × 1 のベクトル
　但し、x_{ij}：目的変数の値 y_i に対応する説明変数 $\mathbf{x}_j$ の n 個の値

⑥重回帰式に使用する説明変数の値をまとめた行列（計画行列という）：$\mathbf{X} = [\mathbf{e}\ \ \mathbf{x}_1\ \ \mathbf{x}_2\ \ \cdots\ \ \mathbf{x}_p]$　n × (p + 1) の行列
　但し、$\mathbf{e} = [1\ \ 1\ \ \cdots\ \ 1]^T$　n × 1 のベクトル

⑦重回帰式の偏回帰係数（未知の確定値）：$\boldsymbol{\beta} = [\beta_0\ \ \beta_1\ \ \cdots\ \ \beta_p]^T$　(p + 1) × 1 のベクトル

⑧重回帰式の偏回帰係数の推定値（**確率変数**）：$\hat{\boldsymbol{\beta}} = [\hat{\beta}_0\ \ \hat{\beta}_1\ \ \cdots\hat{\beta}_p]^T$　(p + 1) × 1 のベクトル

⑨観測誤差（**確率変数**）：$\mathbf{u} = [u_1\ \ u_2\ \ \cdots\ \ u_n]^T$　n × 1 のベクトル

（仮定）$\mathbf{u}$ は多変量正規分布　$N(\mathbf{0}, \sigma^2 \mathbf{I})$ に従う。 u_i は互いに独立（i.i.d. independently identically distributed）で同一の正規分布 $N(0, \sigma^2)$ に従う。

つまり、 $E[\mathbf{u}] = \mathbf{0}$， $V[\mathbf{u}] = E[(\mathbf{u} - E[\mathbf{u}])(\mathbf{u} - E[\mathbf{u}])^T] = E[(\mathbf{u} - \mathbf{0})(\mathbf{u} - \mathbf{0})^T] = E[\mathbf{u}\,\mathbf{u}^T] = \sigma^2 \mathbf{I}$　（ $\mathbf{I}$ は、n × n の単位行列）

⑩推定残差：$\boldsymbol{\varepsilon} = [\varepsilon_1\ \ \varepsilon_2\ \ \cdots\ \ \varepsilon_n]^T$　n × 1 のベクトル

以下に、重回帰分析で使用する式をまとめておく。 また、これらの式の関係を（図2）に示す。

①重回帰モデル式：　$\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \mathbf{u}$　　･･･　（式1）

目的変数 $\mathbf{y}$ を説明変数 $\mathbf{X}$ の線型結合で表すという数式モデルである。 誤差 $\mathbf{u}$ を加味している。
（式1）をベクトルと行列の要素を使って表すと次のようになる。

$$
\begin{bmatrix} y_1 \\ y_2 \\ \cdots \\ y_n \end{bmatrix}
=
\begin{bmatrix} 1 & x_{11} & x_{12} & & x_{1p} \\ 1 & x_{21} & x_{22} & & x_{2p} \\ & & \cdots & & \\ 1 & x_{n1} & x_{n2} & & x_{np} \end{bmatrix}
\begin{bmatrix} \beta_0 \\ \beta_1 \\ \cdots \\ \beta_p \end{bmatrix}
+
\begin{bmatrix} u_1 \\ u_2 \\ \cdots \\ u_n \end{bmatrix}
=
\begin{bmatrix} \beta_0 + \beta_1 x_{11} + \beta_2 x_{12} + \cdots + \beta_p x_{1p} + u_1 \\ \beta_0 + \beta_1 x_{21} + \beta_2 x_{22} + \cdots + \beta_p x_{2p} + u_2 \\ \cdots \\ \beta_0 + \beta_1 x_{n1} + \beta_2 x_{n2} + \cdots + \beta_p x_{np} + u_n \end{bmatrix}
$$

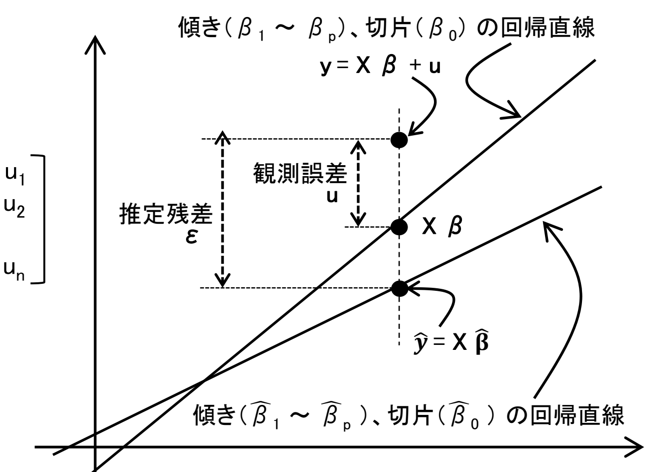
②目的変数の推定式：　$\hat{\mathbf{y}} = \mathbf{X}\hat{\boldsymbol{\beta}}$　　･･･　（式2）

偏回帰係数 $\boldsymbol{\beta}$ の推定値 $\hat{\boldsymbol{\beta}}$ を使って、説明変数 $\mathbf{X}$ の線型結合で目的変数の推定値 $\hat{\mathbf{y}}$ を得る。

③目的変数と目的変数の推定値との関係：　$\mathbf{y} = \hat{\mathbf{y}} + \boldsymbol{\varepsilon}$　　･･･　（式3）

推定にあたっては、推定残差 $\boldsymbol{\varepsilon}$ を伴う。

（式2）は、説明変数 x_{ij} が 1 単位大きくなると目的変数の推定値 $\hat{y}_i$ の値が偏回帰係数 $\hat{\beta}_j$ の値**だけ**大きくなることを示している（偏回帰係数に「**偏**」の文字が使われている所以である）。



（図2）重回帰式の位置付け（イメージ）

2．偏回帰係数(partial regression coefficient)の導出

偏回帰係数の推定値である $\hat{\boldsymbol{\beta}}$ を、p 個の説明変数の観測データ（既知の確定値） $\mathbf{x}_j$ と目的変数の観測データ（既知の確定値） $\mathbf{y}$ から算出する式を導出する。
導出にあたっては、最小二乗法を使う。つまり、目的的変数の推定残差 $\boldsymbol{\varepsilon} = \mathbf{y} - \hat{\mathbf{y}}$ の2乗である $\boldsymbol{\varepsilon}^T\boldsymbol{\varepsilon}$ を最小にする偏回帰係数 $\hat{\boldsymbol{\beta}}$ が最適な推定値であるとする。
最小二乗法を使用して求めた推定値 $\hat{\boldsymbol{\beta}}$ の推定精度が高いことは、$\hat{\boldsymbol{\beta}}$ の分散が最小になることによって確認される（付録6）。
なお、下記の方法とは別に、射影行列を使用した重回帰分析の定式化の方法もあり、これに関しては付録12にまとめておく。

■偏回帰係数の推定値 $\hat{\boldsymbol{\beta}}$ の導出

推定残差 $\varepsilon_i = y_i - \hat{y}_i$ の二乗和である $\sum_{i=1}^{n}(y_i - y_i)^2$ を最小にする $\hat{\boldsymbol{\beta}}$ を求める。（式1）と（式3）から

$$\sum_{i=1}^{n}\varepsilon_i^2 = \sum_{i=1}^{n}(y_i - \hat{y}_i)^2 = (\mathbf{y} - \hat{\mathbf{y}})^T(\mathbf{y} - \hat{\mathbf{y}}) = (\mathbf{y} - \mathbf{X}\hat{\boldsymbol{\beta}})^T(\mathbf{y} - \mathbf{X}\hat{\boldsymbol{\beta}}) = (\mathbf{y}^T - (\mathbf{X}\hat{\boldsymbol{\beta}})^T)(\mathbf{y} - \mathbf{X}\hat{\boldsymbol{\beta}}) = (\mathbf{y}^T - \hat{\boldsymbol{\beta}}^T\mathbf{X}^T)(\mathbf{y} - \mathbf{X}\hat{\boldsymbol{\beta}}) = \mathbf{y}^T\mathbf{y} - \mathbf{y}^T\mathbf{X}\hat{\boldsymbol{\beta}} - \hat{\boldsymbol{\beta}}^T\mathbf{X}^T\mathbf{y} + \hat{\boldsymbol{\beta}}^T\mathbf{X}^T\mathbf{X}\hat{\boldsymbol{\beta}}$$

ここで、$\mathbf{y}^T\mathbf{X}\hat{\boldsymbol{\beta}}$ と $\hat{\boldsymbol{\beta}}^T\mathbf{X}^T\mathbf{y}$ はスカラーで等しい。つまり、$\mathbf{y}^T\mathbf{X}\hat{\boldsymbol{\beta}} = \hat{\boldsymbol{\beta}}^T\mathbf{X}^T\mathbf{y}$ である。

よって、$\sum_{i=1}^{n}\varepsilon_i^2 = \mathbf{y}^T\mathbf{y} - 2\hat{\boldsymbol{\beta}}^T\mathbf{X}^T\mathbf{y} + \hat{\boldsymbol{\beta}}^T\mathbf{X}^T\mathbf{X}\hat{\boldsymbol{\beta}}$

推定残差の二乗和を最小にする $\hat{\boldsymbol{\beta}}$ は、$(\partial/\partial\hat{\boldsymbol{\beta}})(\sum_{i=1}^{n}\varepsilon_i^2) = \mathbf{0}$ を満足する。このことを利用して $\hat{\boldsymbol{\beta}}$ を求める（最小二乗法）。

内積の偏微分（付録3）より、$(\partial/\partial\hat{\boldsymbol{\beta}})(\hat{\boldsymbol{\beta}}^T\mathbf{X}^T\mathbf{y}) = \mathbf{X}^T\mathbf{y}$
また、2次形式の偏微分（付録3）より、$(\partial/\partial\hat{\boldsymbol{\beta}})(\hat{\boldsymbol{\beta}}^T\mathbf{X}^T\mathbf{X}\hat{\boldsymbol{\beta}}) = 2\mathbf{X}^T\mathbf{X}\hat{\boldsymbol{\beta}}$

よって、$(\partial/\partial\hat{\boldsymbol{\beta}})(\sum_{i=1}^{n}\varepsilon_i^2) = (\partial/\partial\hat{\boldsymbol{\beta}})(\mathbf{y}^T\mathbf{y} - 2\hat{\boldsymbol{\beta}}^T\mathbf{X}^T\mathbf{y} + \hat{\boldsymbol{\beta}}^T\mathbf{X}^T\mathbf{X}\hat{\boldsymbol{\beta}}) = (\partial/\partial\hat{\boldsymbol{\beta}})(\mathbf{y}^T\mathbf{y}) - (\partial/\partial\hat{\boldsymbol{\beta}})(2\hat{\boldsymbol{\beta}}^T\mathbf{X}^T\mathbf{y}) + (\partial/\partial\hat{\boldsymbol{\beta}}))(\hat{\boldsymbol{\beta}}^T\mathbf{X}^T\mathbf{X}\hat{\beta})$
$= \mathbf{0} - 2\mathbf{X}^T\mathbf{y} + 2\mathbf{X}^T\mathbf{X}\hat{\boldsymbol{\beta}} = \mathbf{0}$

従って、$\mathbf{X}^T\mathbf{y} = \mathbf{X}^T\mathbf{X}\hat{\boldsymbol{\beta}}$　・・・（式4）

この式を、正規方程式と言う。
正規方程式を満たす $\hat{\boldsymbol{\beta}}$ を求める。これが偏回帰係数 β の推定値になる。

データ数 n は偏回帰係数の数 1 + p より大きいので、n ≧ 1 + p である。説明変数は独立であるとして、rank [X] = p + 1 であり、$\mathbf{X}^T\mathbf{X}$ の逆行列が存在する。
よって、（式4）より、偏回帰係数の推定値 $\hat{\boldsymbol{\beta}}$ が（式5）で求まる。（式5）は、既知の確定値 $\mathbf{x}_i$（$\mathbf{x}_i$ は行列 $\mathbf{X}$ の列ベクトルである）と既知の確定値 $\mathbf{y}$ から、推定残差の二乗和を最小にする偏回帰係数の推定値 $\hat{\boldsymbol{\beta}}$ が求まることを示している。

$\hat{\boldsymbol{\beta}} = (\mathbf{X}^T\mathbf{X})^{-1}\mathbf{X}^T\mathbf{y}$　・・・（式5）

■標準偏回帰係数の推定値 $\hat{\boldsymbol{\beta}}^*$ の導出（計算例は、後述）

標準化した目的変数と説明変数の重回帰式の偏回帰係数を標準偏回帰係数という。標準化によって変数が無次元化されるので、目的変数ごとの偏回帰係数の大小を比較できるようになる。つまり、どの説明変数が目的変数に大きな影響を及ぼしているのかを把握できるようになる。
なお、標準偏回帰係数は、以下のように、偏回帰係数を算出した上で、目的変数と個々の説明変数の標準偏差を加味して求めることもできる。

（式2）と（式3）から、$y_i = \hat{y}_i + \varepsilon_i = \hat{\beta}_0 + \hat{\beta}_1 x_{i1} + \cdots + \hat{\beta}_p x_{ip} + \varepsilon_i$
y_i を標準化して、$y^*_i = (y_i - \bar{y})/\sigma_y \rightarrow y_i = y^*_i\sigma_y + \bar{y}$，　但し、$\bar{y} = (1/n)\sum_{i=1}^{n} y_i$ ，σ_y は目的変数 y の標準偏差
同様に、x_{ij} を標準化して、$x^*_{ij} = (x_{ij} - \bar{x}_j)/\sigma_{x_j} \rightarrow x_{ij} = x^*_{ij}\sigma_{x_j} + \bar{x}_j$　但し、$\bar{x}_j = (1/n)\sum_{i=1}^{n} x_i$ ，σ_{x_j} は説明変数 x_j の標準偏差

これらを上式に代入して、　$y^*_i\sigma_y + \bar{y} = \hat{\beta}_0 + \hat{\beta}_1(x^*_{i1}\sigma_{x_1} + \bar{x}_1) + \cdots + \hat{\beta}_p(x^*_{ip}\sigma_{x_p} + \bar{x}_p) + \varepsilon_i = \hat{\beta}_1 x^*_{i1}\sigma_{x_1} + \cdots + \hat{\beta}_p x^*_{ip}\sigma_{x_p} + \varepsilon_i + \hat{\beta}_0 + \hat{\beta}_1\bar{x}_1 + \cdots + \hat{\beta}_p\bar{x}_p$
後述するように、$\bar{y} = \bar{\hat{y}}$ であるから、$\bar{y} = \bar{\hat{y}} + \bar{\varepsilon} = \bar{\hat{y}}$ より、$\bar{\varepsilon} = 0$　　よって、$\bar{y} = \hat{\beta}_0 + \hat{\beta}_1\bar{x}_1 + \cdots + \hat{\beta}_p\bar{x}_p + \bar{\varepsilon} = \hat{\beta}_0 + \hat{\beta}_1\bar{x}_1 + \cdots + \hat{\beta}_p\bar{x}_p$
$y^*_i\sigma_y = \hat{\beta}_1 x^*_{i1}\sigma_{x_1} + \cdots + \hat{\beta}_p x^*_{ip}\sigma_{x_p} + \varepsilon_i \rightarrow y^*_i = (\sigma_{x_1}/\sigma_y)\hat{\beta}_1 x^*_{i1} + \cdots + (\sigma_{x_p}/\sigma_y)\hat{\beta}_p x^*_{ip} + \varepsilon_i/\sigma_y \rightarrow \hat{y}^*_i = (\sigma_{x_1}/\sigma_y)\hat{\beta}_1 x^*_{i1} + \cdots + (\sigma_{x_p}/\sigma_y)\hat{\beta}_p x^*_{ip}$
従って、標準化偏回帰係数 $\hat{\beta}^*_j$ を偏回帰係数 $\hat{\beta}_j$ を使用して、　$\hat{\beta}^*_j = (\sigma_{x_j}/\sigma_y)\hat{\beta}_j$　・・・（式6）　と置けば、

$\hat{y}^*_i = \hat{\beta}^*_1 x^*_{i1} + \cdots + \hat{\beta}^*_p x^*_{ip}$　と説明変数 x^*_{ij} を使用して、標準化された目的変数の推定値 $\hat{y}^*_i$ を得る重回帰式を設定することができる。

3．決定係数(coefficient of determination)の導出

決定係数は、回帰式が目的変数の観測データにどの程度当てはまっているのかを評価する指標である。

(1)全変動の平方和の分解(ノルム表現に関しては、付録1を参照)

①全変動の平方和を $S_T = \sum_{i=1}^{n}(y_i - \bar{y})^2 = (\mathbf{y} - \bar{y}\,\mathbf{e})^T(\mathbf{y} - \bar{y}\,\mathbf{e}) = \|\mathbf{y} - \bar{y}\mathbf{e}\|^2$ ・・・（式7） ただし、$\bar{y} = (1/n)\sum_{i=1}^{n} y_i$ とおく。

②回帰変動の平方和を $S_R = \sum_{i=1}^{n}(\hat{y}_i - \bar{\hat{y}})^2 = (\hat{\mathbf{y}} - \bar{\hat{y}}\,\mathbf{e})^T(\hat{\mathbf{y}} - \bar{\hat{y}}\,\mathbf{e}) = \|\hat{\mathbf{y}} - \bar{\hat{y}}\,\mathbf{e}\|^2$ ・・・（式8） ただし、$\bar{\hat{y}} = (1/n)\sum_{i=1}^{n}\hat{y}_i$ とおく。

③残差変動の平方和を $S_E = \sum_{i=1}^{n}(y_i - \hat{y}_i)^2 = (\mathbf{y} - \hat{\mathbf{y}})^T(\mathbf{y} - \hat{\mathbf{y}}) = \|\mathbf{y} - \hat{\mathbf{y}}\|^2$ ・・・（式9）とおく。なお、$S_E = \boldsymbol{\varepsilon}^T\boldsymbol{\varepsilon} = \|\boldsymbol{\varepsilon}\|^2$ と表記することもできる。

(式4)より、$X^T(\mathbf{y} - X\hat{\boldsymbol{\beta}}) = \mathbf{0}$ であるので、これを X と $\mathbf{y} - X\hat{\boldsymbol{\beta}}$ との内積であると捉えれば、$(\mathbf{y} - X\hat{\boldsymbol{\beta}})^T X = \mathbf{0}$

ベクトル $\boldsymbol{\varepsilon}$ と行列 X との内積は、$\langle\boldsymbol{\varepsilon}, X\rangle = \boldsymbol{\varepsilon}^T X = (\mathbf{y} - \hat{\mathbf{y}})^T X = (\mathbf{y} - X\hat{\boldsymbol{\beta}})^T X = \mathbf{0}$

よって、ベクトル $\boldsymbol{\varepsilon}$ とベクトル $\hat{\mathbf{y}}$ との内積は、$\langle\boldsymbol{\varepsilon}, \hat{\mathbf{y}}\rangle = \boldsymbol{\varepsilon}^T\hat{\mathbf{y}} = ((\boldsymbol{\varepsilon}^T\hat{\mathbf{y}})^T)^T = (\hat{\mathbf{y}}^T\boldsymbol{\varepsilon})^T = ((X\hat{\boldsymbol{\beta}})^T\boldsymbol{\varepsilon})^T = \boldsymbol{\varepsilon}^T(X\hat{\boldsymbol{\beta}}) = \boldsymbol{\varepsilon}^T X\hat{\boldsymbol{\beta}} = \mathbf{0}$ 同様に、$\langle\hat{\mathbf{y}}, \boldsymbol{\varepsilon}\rangle = \hat{\mathbf{y}}^T\boldsymbol{\varepsilon} = \mathbf{0}$

従って、$\boldsymbol{\varepsilon}$ は X と $\hat{\mathbf{y}}$ それぞれと直交していることが分かる。

つまり、$\boldsymbol{\varepsilon}$ は行列 X を構成する列ベクトル $\mathbf{e}$ と直交しているので、内積 $\langle\boldsymbol{\varepsilon}, \mathbf{e}\rangle = \boldsymbol{\varepsilon}^T\mathbf{e} = \mathbf{e}^T\boldsymbol{\varepsilon} = \sum_{i=1}^{n}\varepsilon_i = 0$ より、残差 $\boldsymbol{\varepsilon}$ の平均は、$(1/n)\sum_{i=1}^{n}\varepsilon_i = 0$ である。

$$
\begin{aligned}
(\mathbf{y} - \bar{y}\,\mathbf{e})^T(\mathbf{y} - \bar{y}\,\mathbf{e}) &= (\hat{\mathbf{y}} + \boldsymbol{\varepsilon} - \bar{y}\,\mathbf{e})^T(\hat{\mathbf{y}} + \boldsymbol{\varepsilon} - \bar{y}\,\mathbf{e}) = (\hat{\mathbf{y}} - \bar{y}\,\mathbf{e})^T(\hat{\mathbf{y}} - \bar{y}\,\mathbf{e}) + (\hat{\mathbf{y}} - \bar{y}\,\mathbf{e})^T\boldsymbol{\varepsilon} + \boldsymbol{\varepsilon}^T(\hat{\mathbf{y}} - \bar{y}\,\mathbf{e}) + \boldsymbol{\varepsilon}^T\boldsymbol{\varepsilon} \\
&= (\hat{\mathbf{y}} - \bar{y}\,\mathbf{e})^T(\hat{\mathbf{y}} - \bar{y}\,\mathbf{e}) + \hat{\mathbf{y}}^T\boldsymbol{\varepsilon} - \bar{y}\,\mathbf{e}^T\boldsymbol{\varepsilon} + \boldsymbol{\varepsilon}^T\hat{\mathbf{y}} - \bar{y}\,\boldsymbol{\varepsilon}^T\mathbf{e} + \boldsymbol{\varepsilon}^T\boldsymbol{\varepsilon} = (\hat{\mathbf{y}} - \bar{y}\,\mathbf{e})^T(\hat{\mathbf{y}} - \bar{y}\,\mathbf{e}) + \boldsymbol{\varepsilon}^T\boldsymbol{\varepsilon}
\end{aligned}
$$

ここで、$\bar{y} - \bar{\hat{y}} = (\mathbf{y}^T\mathbf{e})/n - (\hat{\mathbf{y}}^T\mathbf{e})/n = (\mathbf{y}^T - \hat{\mathbf{y}}^T)\mathbf{e}/n = \boldsymbol{\varepsilon}^T\mathbf{e}/n = 0$ より $\bar{y} = \bar{\hat{y}}$ （目的変数の平均 $\bar{y}$ と目的変数の推定値の平均 $\bar{\hat{y}}$ は等しい）

よって、$(\mathbf{y} - \bar{y}\,\mathbf{e})^T(\mathbf{y} - \bar{y}\,\mathbf{e}) = (\hat{\mathbf{y}} - \bar{\hat{y}}\,\mathbf{e})^T(\hat{\mathbf{y}} - \bar{\hat{y}}\,\mathbf{e}) + \boldsymbol{\varepsilon}^T\boldsymbol{\varepsilon} = (\hat{\mathbf{y}} - \bar{\hat{y}}\,\mathbf{e})^T(\hat{\mathbf{y}} - \bar{\hat{y}}\,\mathbf{e}) + (\mathbf{y} - \hat{\mathbf{y}})^T(\mathbf{y} - \hat{\mathbf{y}})$ つまり、(式10) が成り立つ。 $S_T = S_R + S_E$ ・・・（式10）

(2)決定係数(coefficient of determination)

残差は推定誤差を表している。従って、(式9)の残差変動の平方和が小さいとき、回帰式は目的変数の観測データによく当てはまっていると見なすことができる。

そこで、(式11)で定義される重決定係数 R^2 を、回帰式のあてはまりの良さを表す指標として用いる。$R^2 = 0$ ～ 1 で、1 に近いほど当てはまりがいい。

$R^2 = 1 - S_E / S_T = (S_T - S_E)/S_T = S_R / S_T = \|\hat{\mathbf{y}} - \bar{\hat{y}}\,\mathbf{e}\|^2 / \|\mathbf{y} - \bar{y}\,\mathbf{e}\|^2$ ・・・（式11）

(3)自由度調整済み重決定係数(adjusted coefficient of determination)

説明変数の数が増えるほど重回帰式の当てはまりは良くなる。つまり、重決定係数 R^2 は 1 に近づく。説明変数の数に左右されないように自由度を加味した決定係数を用意する。それが、自由度調整済み重決定係数 R^{*2} である。説明変数の数 p が大きくなると、重決定係数 R^2 は 1 から離れる方向に動く。

$R^{*2} = 1 - (S_E / (n - p - 1)) / (S_T / (n - 1))$ ・・・（式12）

ここで、S_E の自由度は偏回帰係数の数 $p + 1$ が固定化されるので $n - (p + 1)$ である。一方、S_T の自由度は平均 1 つが固定化されるので $n - 1$ である。

(4)決定係数と重相関係数との関係(相関係数は、別途資料「統計的推定の手順」の付録16を参照)

目的変数の観測値ベクトル $\mathbf{y}$ と推定値ベクトル $\hat{\mathbf{y}}$ の重相関係数 ρ は、(式13)で表すことができる。

$\rho = (\mathbf{y} - \bar{y}\,\mathbf{e})^T(\hat{\mathbf{y}} - \bar{\hat{y}}\,\mathbf{e}) / \|\mathbf{y} - \bar{y}\,\mathbf{e}\|\,\|\hat{\mathbf{y}} - \bar{\hat{y}}\,\mathbf{e}\|$ ・・・（式13）

ここで、$\mathbf{y} - \bar{y}\,\mathbf{e} = (\mathbf{y} - \hat{\mathbf{y}}) + (\hat{\mathbf{y}} - \bar{y}\,\mathbf{e}) = \boldsymbol{\varepsilon} + (\hat{\mathbf{y}} - \bar{y}\,\mathbf{e})$ より、

$(\mathbf{y} - \bar{y}\,\mathbf{e})^T(\hat{\mathbf{y}} - \bar{\hat{y}}\,\mathbf{e}) = (\boldsymbol{\varepsilon} + (\hat{\mathbf{y}} - \bar{y}\,\mathbf{e}))^T(\hat{\mathbf{y}} - \bar{\hat{y}}\,\mathbf{e}) = \boldsymbol{\varepsilon}^T(\hat{\mathbf{y}} - \bar{\hat{y}}\,\mathbf{e}) + (\hat{\mathbf{y}} - \bar{y}\,\mathbf{e})^T(\hat{\mathbf{y}} - \bar{\hat{y}}\,\mathbf{e})$

$\boldsymbol{\varepsilon}$ と $\hat{\mathbf{y}}$，$\boldsymbol{\varepsilon}$ と $\mathbf{e}$ はそれぞれ直交する(上記(1)を参照)ので、$\boldsymbol{\varepsilon}$ と $\hat{\mathbf{y}} - \bar{\hat{y}}\,\mathbf{e}$ は直交する。従って、内積は $\mathbf{0}$ である。つまり、$\langle\boldsymbol{\varepsilon}, \hat{\mathbf{y}} - \bar{\hat{y}}\,\mathbf{e}\rangle = \boldsymbol{\varepsilon}^T(\hat{\mathbf{y}} - \bar{\hat{y}}\,\mathbf{e}) = \mathbf{0}$

従って、$\bar{y} = \bar{\hat{y}}$ の関係()、そして、(付録1)から、

$(\mathbf{y} - \bar{y}\,\mathbf{e})^T(\hat{\mathbf{y}} - \bar{\hat{y}}\,\mathbf{e}) = (\hat{\mathbf{y}} - \bar{y}\,\mathbf{e})^T(\hat{\mathbf{y}} - \bar{\hat{y}}\,\mathbf{e}) = (\hat{\mathbf{y}} - \bar{\hat{y}}\,\mathbf{e})^T(\hat{\mathbf{y}} - \bar{\hat{y}}\,\mathbf{e}) = \|\hat{\mathbf{y}} - \bar{\hat{y}}\,\mathbf{e}\|^2$

$\rho = (\mathbf{y} - \bar{y}\,\mathbf{e})^T(\hat{\mathbf{y}} - \bar{\hat{y}}\,\mathbf{e}) / \|\mathbf{y} - \bar{y}\,\mathbf{e}\|\,\|\hat{\mathbf{y}} - \bar{\hat{y}}\,\mathbf{e}\| = \|\hat{\mathbf{y}} - \bar{\hat{y}}\,\mathbf{e}\|^2 / \|\mathbf{y} - \bar{y}\,\mathbf{e}\|\,\|\hat{\mathbf{y}} - \bar{\hat{y}}\,\mathbf{e}\| = \|\hat{\mathbf{y}} - \bar{\hat{y}}\,\mathbf{e}\| / \|\mathbf{y} - \bar{y}\,\mathbf{e}\|$ ・・・（式14）

従って、(式11)と(式14)を比較して、$R^2 = \rho^2$ ・・・（式15）

4．重回帰式の検定

帰無仮説を全ての偏回帰係数 $\hat\beta_j = 0$（つまり、$\hat{\boldsymbol\beta} = \mathbf{0}$）として、回帰式の有用性を検定するための検定統計量を求める。

（1）$\mathbf{y} - \hat{\mathbf{y}}$ の確率分布を求める。

（式1）$\mathbf{y} = \mathbf{X}\boldsymbol\beta + \mathbf{u}$ で、$\mathbf{u}$ は多変量正規分布 $N(\mathbf{0}, \sigma^2\mathbf{I})$ に従う確率変数ベクトルであり、$\mathbf{y}$ も $\mathbf{y} - \hat{\mathbf{y}}$ も多変量正規分布に従う確率変数ベクトルになる。

① $\mathbf{y} - \hat{\mathbf{y}}$ の期待値 $E[\mathbf{y} - \hat{\mathbf{y}}]$ を求める。

（式1）$\mathbf{y} = \mathbf{X}\boldsymbol\beta + \mathbf{u}$,（式2）$\hat{\mathbf{y}} = \mathbf{X}\hat{\boldsymbol\beta}$,（付式20）$E[\hat{\boldsymbol\beta}] = \boldsymbol\beta$ より、$E[\mathbf{y} - \hat{\mathbf{y}}] = E[\mathbf{y}] - E[\hat{\mathbf{y}}] = E[\mathbf{X}\boldsymbol\beta + \mathbf{u}] - E[\mathbf{X}\hat{\boldsymbol\beta}] = E[\mathbf{X}\boldsymbol\beta] + E[\mathbf{u}] - \mathbf{X}E[\hat{\boldsymbol\beta}] = \mathbf{X}\boldsymbol\beta + \mathbf{0} - \mathbf{X}\boldsymbol\beta = \mathbf{0}$ ・・・（式16）

② $\mathbf{y} - \hat{\mathbf{y}}$ の分散 $V[\mathbf{y} - \hat{\mathbf{y}}]$ を求める。

$\mathbf{H} = \mathbf{X}(\mathbf{X}^T\mathbf{X})^{-1}\mathbf{X}^T$ とおく。 $\mathbf{H}^2 = (\mathbf{X}(\mathbf{X}^T\mathbf{X})^{-1}\mathbf{X}^T)(\mathbf{X}(\mathbf{X}^T\mathbf{X})^{-1}\mathbf{X}^T) = \mathbf{X}(\mathbf{X}^T\mathbf{X})^{-1}\mathbf{I}\,\mathbf{X}^T = \mathbf{X}(\mathbf{X}^T\mathbf{X})^{-1}\mathbf{X}^T = \mathbf{H}$ なので、行列 $\mathbf{H}$ は冪等行列である（付録4）。

また、$\mathbf{H}^T = (\mathbf{X}(\mathbf{X}^T\mathbf{X})^{-1}\mathbf{X}^T)^T = \mathbf{X}(\mathbf{X}^T\mathbf{X})^{-1}\mathbf{X}^T = \mathbf{H}$ が成り立つ。 なお、$\mathbf{I} - \mathbf{H}$ も冪等行列であるので（付録4）、$(\mathbf{I} - \mathbf{H})^2 = \mathbf{I} - \mathbf{H}$, $(\mathbf{I} - \mathbf{H})^T = \mathbf{I} - \mathbf{H}$ も成り立つ。

（式1）$\mathbf{y} = \mathbf{X}\boldsymbol\beta + \mathbf{u}$,（式2）$\hat{\mathbf{y}} = \mathbf{X}\hat{\boldsymbol\beta}$,（付式19）$\hat{\boldsymbol\beta} = \boldsymbol\beta + (\mathbf{X}^T\mathbf{X})^{-1}\mathbf{X}^T\mathbf{u}$ を使って、

$\mathbf{y} - \hat{\mathbf{y}} = (\mathbf{X}\boldsymbol\beta + \mathbf{u}) - \mathbf{X}(\boldsymbol\beta + (\mathbf{X}^T\mathbf{X})^{-1}\mathbf{X}^T\mathbf{u}) = \mathbf{u} - \mathbf{X}(\mathbf{X}^T\mathbf{X})^{-1}\mathbf{X}^T\mathbf{u} = (\mathbf{I} - \mathbf{X}(\mathbf{X}^T\mathbf{X})^{-1}\mathbf{X}^T)\mathbf{u} = (\mathbf{I} - \mathbf{H})\mathbf{u}$ であるので、

$$V[\mathbf{y} - \hat{\mathbf{y}}] = E[(\mathbf{y} - \hat{\mathbf{y}} - E[\mathbf{y} - \hat{\mathbf{y}}])^T(\mathbf{y} - \hat{\mathbf{y}} - E[\mathbf{y} - \hat{\mathbf{y}}])] = E[(\mathbf{y} - \hat{\mathbf{y}} - \mathbf{0})^T(\mathbf{y} - \hat{\mathbf{y}} - \mathbf{0})] = E[(\mathbf{y} - \hat{\mathbf{y}})^T(\mathbf{y} - \hat{\mathbf{y}})] = E[((\mathbf{I} - \mathbf{H})\mathbf{u})^T((\mathbf{I} - \mathbf{H})\mathbf{u})] = E[\mathbf{u}^T(\mathbf{I} - \mathbf{H})^T(\mathbf{I} - \mathbf{H})\mathbf{u}]$$

$$= E[\mathbf{u}^T(\mathbf{I} - \mathbf{H})(\mathbf{I} - \mathbf{H})\mathbf{u}] = E[\mathbf{u}^T(\mathbf{I} - \mathbf{H})^2\mathbf{u}] = E[\mathbf{u}^T(\mathbf{I} - \mathbf{H})\mathbf{u}] = (\mathbf{I} - \mathbf{H})E[\mathbf{u}^T\mathbf{u}] = (\mathbf{I} - \mathbf{H})\,\sigma^2\mathbf{I} = \sigma^2(\mathbf{I} - \mathbf{H})$$

冪等行列 $\mathbf{I} - \mathbf{H}$ は直交行列 $\mathbf{Q}$ を使用して対角化できる（付録4）。なお、$\mathrm{rank}(\mathbf{I} - \mathbf{H}) = \mathrm{rank}(\mathbf{I}) - \mathrm{rank}(\mathbf{H}) = n - (p + 1)$ である。

$$V[\mathbf{y} - \hat{\mathbf{y}}] = \sigma^2(\mathbf{I} - \mathbf{H}) = \sigma^2\mathbf{Q}\,\mathbf{Q}^{-1}(\mathbf{I} - \mathbf{H})\mathbf{Q}\,\mathbf{Q}^{-1} = \sigma^2\mathbf{Q}\begin{bmatrix}\mathbf{I}_{n-p-1} & \mathbf{0}\\ \mathbf{0} & \mathbf{0}\end{bmatrix}\mathbf{Q}^{-1} = \sigma^2\begin{bmatrix}\mathbf{I}_{n-p-1} & \mathbf{0}\\ \mathbf{0} & \mathbf{0}\end{bmatrix}$$ ・・・（式17） 但し、$\mathbf{I}_{n-p-1}$ は、サイズ $n - p - 1$ の単位行列

③（式16）と（式17）から、$y_i - \hat y_i$（$i = 1$ 〜 $n - p - 1$）の確率分布は、正規分布 $N(0, \sigma^2)$ に従うことが分かる。

よって、$\sum_{i=1}^{n-p-1}((y_i - \hat y_i) - 0)^2 / \sigma^2 = \sum_{i=1}^{n-p-1}(y_i - \hat y_i)^2 / \sigma^2$ は、自由度 $n - p - 1$ のカイ二乗分布に従う（別途資料「統計的推定の手順」の付録5を参照）。

従って、サンプル数を n にした、$\sum_{i=1}^{n}(y_i - \hat y_i)^2 / \sigma^2 = S_E / \sigma^2$ も自由度 $n - p - 1$ のカイ二乗分布に従う。

（2）$\hat{\mathbf{y}}$ の確率分布を求める。

$\hat{\mathbf{y}} = \mathbf{X}\hat{\boldsymbol\beta}$ の確率分布を求める。なお、前述した目的変数の平均 $\bar y$ と目的変数の推定値の平均 $\bar{\hat y}$ が等しいことは、以下でも確認できる。

① $\hat{\mathbf{y}}$ の期待値 $E[\hat{\mathbf{y}}]$ を求める。

（付式20）$E[\hat{\boldsymbol\beta}] = \boldsymbol\beta$ であるので、$E[\hat{\mathbf{y}}] = E[\mathbf{X}\hat{\boldsymbol\beta}] = \mathbf{X}E[\hat{\boldsymbol\beta}] = \mathbf{X}\boldsymbol\beta$

一方 $E[\mathbf{u}] = \mathbf{0}$ であるので、$E[\mathbf{y}] = E[\mathbf{X}\boldsymbol\beta + \mathbf{u}] = E[\mathbf{X}\boldsymbol\beta] + E[\mathbf{u}] = \mathbf{X}\boldsymbol\beta + \mathbf{0} = \mathbf{X}\boldsymbol\beta$ 　従って、$\mathbf{X}\boldsymbol\beta = E[\hat{\mathbf{y}}] = E[\mathbf{y}]$ 　つまり、$\bar{\hat y} = \bar y$ ・・・（式18）

② $\hat{\mathbf{y}}$ の分散 $V[\hat{\mathbf{y}}]$ を求める。

$$V[\hat{\mathbf{y}}] = E[(\hat{\mathbf{y}} - E[\hat{\mathbf{y}}])^T(\hat{\mathbf{y}} - E[\hat{\mathbf{y}}])] = E[(\mathbf{X}\hat{\boldsymbol\beta} - \mathbf{X}\boldsymbol\beta)^T(\mathbf{X}\hat{\boldsymbol\beta} - \mathbf{X}\boldsymbol\beta)] = E[(\mathbf{X}(\hat{\boldsymbol\beta} - \boldsymbol\beta))^T(\mathbf{X}(\hat{\boldsymbol\beta} - \boldsymbol\beta))] = E[(\hat{\boldsymbol\beta} - \boldsymbol\beta)^T\mathbf{X}^T\mathbf{X}(\hat{\boldsymbol\beta} - \boldsymbol\beta)]$$

ここで、（付式19）$\hat{\boldsymbol\beta} = \boldsymbol\beta + (\mathbf{X}^T\mathbf{X})^{-1}\mathbf{X}^T\mathbf{u}$ を代入すると、

$$V[\hat{\mathbf{y}}] = E[((\mathbf{X}^T\mathbf{X})^{-1}\mathbf{X}^T\mathbf{u})^T\mathbf{X}^T\mathbf{X}((\mathbf{X}^T\mathbf{X})^{-1}\mathbf{X}^T\mathbf{u})] = E[((\mathbf{X}^T\mathbf{X})^{-1}\mathbf{X}^T\mathbf{u})^T\mathbf{I}\,\mathbf{X}^T\mathbf{u})] = E[\mathbf{u}^T\mathbf{X}(\mathbf{X}^T\mathbf{X})^{-1}\mathbf{X}^T\mathbf{u})] = E[\mathbf{u}^T\mathbf{H}\mathbf{u}] = \mathbf{H}E[\mathbf{u}^T\mathbf{u})] = \mathbf{H}\,\sigma^2\mathbf{I} = \sigma^2\mathbf{H}$$

冪等行列 $\mathbf{H}$ は直交行列 $\mathbf{Q}$ を使用して対角化できる（付録4）。なお、$\mathrm{rank}(\mathbf{H}) = p + 1$ である。

$$V[\hat{\mathbf{y}}] = \sigma^2\mathbf{H} = \sigma^2\mathbf{Q}\,\mathbf{Q}^{-1}\mathbf{H}\mathbf{Q}\,\mathbf{Q}^{-1} = \sigma^2\mathbf{Q}\begin{bmatrix}\mathbf{I}_{p+1} & \mathbf{0}\\ \mathbf{0} & \mathbf{0}\end{bmatrix}\mathbf{Q}^{-1} = \sigma^2\begin{bmatrix}\mathbf{I}_{p+1} & \mathbf{0}\\ \mathbf{0} & \mathbf{0}\end{bmatrix}$$ ・・・（式19）

③（式18）と（式19）から、$\hat y_i$（$i = 1$ 〜 $p + 1$）の確率分布は、正規分布 $N(\bar{\hat y}, \sigma^2)$ に従うことが分かる。

よって、$\sum_{i=1}^{p+1}(\hat y_i - \bar{\hat y})^2 / \sigma^2$ は、自由度 $(p + 1) - 1 = p$ のカイ二乗分布に従う（別途資料「統計的推定の手順」の付録6を参照）。

従って、サンプル数を n にした、$\sum_{i=1}^{n}(\hat y_i - \bar{\hat y})^2 / \sigma^2 = S_R / \sigma^2$ も自由度 p のカイ二乗分布に従う。

（3）以上（1）と（2）の結果を使って、下記の検定統計量 T を設定する。検定統計量 T はカイ二乗分布に従う統計量の比になっており、F 分布に従う（付録15）。

この検定統計量 T は、S_E が S_R に比べて小さな値を取るとき、大きな値をとり、重回帰式が有用と判断される（$\hat{\boldsymbol\beta} = \mathbf{0}$ という帰無仮説は棄却される）。

$$T = \{(S_R / \sigma^2) / p\} / \{(S_E / \sigma^2) / (n - p - 1)\} = (S_R / p) / (S_E / (n - p - 1))$$ ・・・（式20）

5. 個々の偏回帰係数の推定と検定

帰無仮説を個々の偏回帰係数 $\hat{\beta}_j = 0$ として、偏回帰係数 $\hat{\beta}_j$ の有用性を検定するための検定統計量を求める。

（1）j 番目の説明変数 x_j の偏回帰係数の推定値 $\hat{\beta}_j$ が従う確率分布を求める。

前述した通り、ベクトル $\mathbf{y}$ は、多変量正規分布に従う確率変数ベクトルになる。

$\hat{\boldsymbol{\beta}}$ は、（式5） $\hat{\boldsymbol{\beta}} = (X^T X)^{-1} X^T \mathbf{y}$ において行列 X は既知の確定値であるので、ベクトル $\mathbf{y}$ と同じく多変量正規分布に従う確率変数ベクトルになる。

$\hat{\boldsymbol{\beta}}$ の期待値は（付式20）より $E[\hat{\boldsymbol{\beta}}] = \boldsymbol{\beta}$ であり、$\hat{\boldsymbol{\beta}}$ の分散は（付式21）より $V[\hat{\boldsymbol{\beta}}] = \sigma^2 (X^T X)^{-1}$ であり、$\hat{\boldsymbol{\beta}}$ は多変量正規分布 $N(\boldsymbol{\beta}, \sigma^2 (X^T X)^{-1})$ に従う。

ここで、偏回帰係数の推定値 $\hat{\beta}_j$ に注目すると、$E[\hat{\beta}_j] = \beta_j$ ， $V[\hat{\beta}_j] = \sigma^2 a_{jj}$ （但し、a_{jj} は 行列 $(X^T X)^{-1}$ の $(j+1, j+1)$ 番目の要素）である。

従って、$\hat{\beta}_j$ は正規分布 $N(\beta_j, \sigma^2 a_{jj})$ に従うことが分かる。

標準化すれば、$(\hat{\beta}_j - \beta_j) / (\sigma^2 a_{jj})^{1/2}$ ・・・ （式21） は、標準正規分布 $N(0, 1)$ に従う。

（2））誤差 $\mathbf{u}$ の分散 $\sigma^2 \mathbf{I}$ の不変推定量を求める。

$\hat{\beta}_j$ は正規分布 $N(\beta_j, \sigma^2 a_{jj})$ に従う確率変数であるが、母数 σ^2 の値は未知なので、$\hat{\beta}_j$ の統計的推定にあたっては σ^2 を代替するパラメータが必要になる。

それが、下記に示す σ^2 の不変推定量として定義する S^2 である（付録8）。 なお、$(S^2)^{1/2} = S$ ・・・ （式22） これは、観測誤差の分散の標準誤差（付録9）

$S^2 = (\boldsymbol{\varepsilon}^T \boldsymbol{\varepsilon}) / (n-p-1)$ ・・・ （付式30）　　なお、$S^2 = S_E / (n-p-1)$ ・・・ （式23） と表記することもできる。

（3）推定残差の平方和から導かれる確率変数を求める。

推定残差 $\boldsymbol{\varepsilon}^T \boldsymbol{\varepsilon} = (\mathbf{y} - \hat{\mathbf{y}})^T (\mathbf{y} - \hat{\mathbf{y}})$ であることから、（式17）に至る式変形を活用して（式24）が求まる。

$$\boldsymbol{\varepsilon}^T \boldsymbol{\varepsilon} = \mathbf{u}^T (\mathbf{I} - \mathbf{H})\, \mathbf{u} = \mathbf{u}^T Q\, Q^{-1} (\mathbf{I} - \mathbf{H})\, Q\, Q^{-1} \mathbf{u} = \mathbf{u}^T Q \begin{bmatrix} \mathbf{I}_{n-p-1} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} \end{bmatrix} Q^{-1} \mathbf{u} = \mathbf{u}^T \begin{bmatrix} \mathbf{I}_{n-p-1} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} \end{bmatrix} \mathbf{u} = \sum_{i=1}^{n-p-1} u_i^2 \quad \cdots \text{（式24）}$$

u_i は正規分布 $N(0, \sigma^2)$ に従うので、$\sum_{i=1}^{n-p-1} u_i^2 / \sigma^2$ は、自由度 $n-p-1$ のカイ二乗分布に従う（別途資料「統計的推定の手順」の付録5参照）。

従って、 $\boldsymbol{\varepsilon}^T \boldsymbol{\varepsilon} / \sigma^2 = \sum_{i=1}^{n-p-1} u_i^2 / \sigma^2$ ・・・ （式25） は、自由度 $n-p-1$ のカイ二乗分布に従う。

（4）統計的推定と検定に使用する統計量を設定する。

標準正規分布に従う（式21）とカイ二乗分布に従う（式25）を使用し、（付式30）を代入して、

$T = \{ (\hat{\beta}_j - \beta_j) / (\sigma^2 a_{jj})^{1/2} \} / \{ (\boldsymbol{\varepsilon}^T \boldsymbol{\varepsilon} / \sigma^2) / (n-p-1) \}^{1/2} = \{ (\hat{\beta}_j - \beta_j) / (\sigma^2 a_{jj})^{1/2} \} / (S^2 / \sigma^2)^{1/2} = (\hat{\beta}_j - \beta_j) / (S^2 a_{jj})^{1/2}$ ・・・ （式26）

この統計量 T は、自由度 $n-p-1$ の t 分布に従う（付録14）。

σ^2 の不変推定量として定義する S^2 を導入することで、未知の σ^2 を使わない統計量 T を作ることができる。

（5）個々の偏回帰係数の統計的推定

統計的推定の信頼度 = 100 $(1-\alpha)$ % としたとき、（式26）で示す自由度 $n-p-1$ の t 分布に従う統計量の上側 $\alpha/2$ 点は $t_{\alpha/2}(n-p-1)$ 、下側 $\alpha/2$ 点は $t_{1-\alpha/2}(n-p-1)$ で表される。 従って、（式26）を使用して、信頼度 = 100 $(1-\alpha)$ % の信頼区間は次のように表される。

$t_{1-\alpha/2}(n-p-1) \leqq (\hat{\beta}_j - \beta_j) / (S^2 a_{jj})^{1/2} \leqq t_{\alpha/2}(n-p-1)$

この関係を、β_j の信頼区間を表す式に変形すると、（式27）が得られる。

$\hat{\beta}_j - t_{\alpha/2}(n-p-1)\,(S^2 a_{jj})^{1/2} \leqq \beta_j \leqq \hat{\beta}_j + t_{1-\alpha/2}(n-p-1)\,(S^2 a_{jj})^{1/2}$ ・・・ （式27）

（6）個々の偏回帰係数の統計的検定

帰無仮説を $\beta_j = 0$ つまり、偏回帰係数 β_i は有用ではないという帰無仮説を設け、この帰無仮説を棄却できれば意味のある偏回帰係数が得られたとする。

従って、検定統計量は（式26）において $\beta_j = 0$ とした（式28）になる。

$T = \hat{\beta}_j / (S^2 a_{jj})^{1/2}$ ・・・ （式28）

なお、$(S^2 a_{jj})^{1/2}$ ・・・ （式29） を、偏回帰係数の標準誤差という。検定統計量 T は偏回帰係数の推定量を標準誤差で割ったものになっている。

6．重回帰分析の計算例（1/3）

Excel の回帰分析ツールを使用して求めた結果が、前述した（導出した）算式を使用して求められることを確認する。
（例表１）のデータを対象に、Excel の回帰分析ツールを使用して求めた結果が（例表２）である。

（例表１）観測データ

目的変数 y	説明変数 x_1	説明変数 x_2
$y_1 = 5$	$x_{11} = 10$	$x_{12} = 7$
$y_2 = 10$	$x_{21} = 15$	$x_{22} = 15$
$y_3 = 8$	$x_{31} = 20$	$x_{32} = 4$
$y_4 = 3$	$x_{41} = 5$	$x_{42} = 6$

（注）x_1 と x_2 のデータの相関係数は 0.053 であり、
２つの説明変数の間に相関関係はない判断できる。

（例表２）回帰分析結果（Excel）

概要

回帰統計		
重相関 R	0.999209	①
重決定 R2	0.998419	②
補正 R2	0.995257	③
標準誤差	0.214115	④
観測数	4	

分散分析表

	⑤ 自由度	⑥ 変動	⑦ 分散	⑧ 観測された分散比	⑨ 有意 F
回帰	2	28.95415	14.47708	315.7813	0.03976
残差	1	0.045845	0.045845		
合計	3	29			

	係数	標準誤差	t	P-値	下限 95%	上限 95%	下限 95.0%	上限 95.0%
切片	-1.18625	0.325136	-3.64846	0.170307	-5.31749	2.944998	-5.31749	2.944998
X 値 1	0.385673	0.019178	20.10972	0.031631	0.141988	0.629359	0.141988	0.629359
X 値 2	0.358166	0.025628	13.97542	0.045475	0.032528	0.683804	0.032528	0.683804
	⑩	⑪	⑫	⑬	⑭			

■偏回帰係数の算出

（式５）$\hat{\boldsymbol{\beta}} = (X^T X)^{-1} X^T \mathbf{y}$ を使用して、偏回帰係数の推定値 $\hat{\boldsymbol{\beta}}$ を求めて、重回帰式を作成する。

$$X = [\ \mathbf{e}\ \ \mathbf{x}_1\ \ \mathbf{x}_2\ ] = \begin{bmatrix} 1 & x_{11} & x_{12} \\ 1 & x_{21} & x_{22} \\ 1 & x_{31} & x_{32} \\ 1 & x_{41} & x_{42} \end{bmatrix} = \begin{bmatrix} 1 & 10 & 7 \\ 1 & 15 & 15 \\ 1 & 20 & 4 \\ 1 & 5 & 6 \end{bmatrix}, \quad X^T X = \begin{bmatrix} 1 & 1 & 1 & 1 \\ 10 & 15 & 20 & 5 \\ 7 & 15 & 4 & 6 \end{bmatrix} \begin{bmatrix} 1 & 10 & 7 \\ 1 & 15 & 15 \\ 1 & 20 & 4 \\ 1 & 5 & 6 \end{bmatrix} = \begin{bmatrix} 4 & 50 & 32 \\ 50 & 750 & 405 \\ 32 & 405 & 326 \end{bmatrix}$$

$X^T X$ の逆行列 $(X^T X)^{-1}$ は余因子行列を使用して求める（別途資料「現代制御の基礎知識」の付録を参照）。
なお、Excell の逆行列関数 MINVERSE を使用しても、計算サイトを利用しても求めることができる。 https://keisan.casio.jp/exec/system/1278758206

余因子行列は、$\mathrm{adj}\,(X^T X) = \begin{bmatrix} 80475 & -3340 & -3750 \\ -3340 & 280 & -20 \\ -3750 & -20 & 500 \end{bmatrix}$，　余因子行列の行列式は、$|\,\mathrm{adj}\,(X^T X)\,| = \begin{vmatrix} 80475 & -3340 & -3750 \\ -3340 & 280 & -20 \\ -3750 & -20 & 500 \end{vmatrix} = 34900$

よって、$X^T X$ の逆行列 $(X^T X)^{-1}$ は、$(X^T X)^{-1} = (1/\,|\,\mathrm{adj}\,(X^T X)\,|\,)\ \mathrm{adj}\,(X^T X) = (1/3490) \begin{bmatrix} 80475 & -3340 & -3750 \\ -3340 & 280 & -20 \\ -3750 & -20 & 500 \end{bmatrix} = \begin{bmatrix} 2.306 & -0.096 & -0.107 \\ -0.096 & 0.008 & -0.001 \\ -0.107 & -0.001 & 0.014 \end{bmatrix}$

$X^T X$ の逆行列 $(X^T X)^{-1}$ を使って、（式５）より、$\hat{\boldsymbol{\beta}} = (X^T X)^{-1} X^T \mathbf{y} = \begin{bmatrix} 2.306 & -0.096 & -0.107 \\ -0.096 & 0.008 & -0.001 \\ -0.107 & -0.001 & 0.014 \end{bmatrix} \begin{bmatrix} 1 & 1 & 1 & 1 \\ 10 & 15 & 20 & 5 \\ 7 & 15 & 4 & 6 \end{bmatrix} \begin{bmatrix} 5 \\ 10 \\ 8 \\ 3 \end{bmatrix} = \begin{bmatrix} -1.186 \\ 0.386 \\ 0.358 \end{bmatrix}$

従って、（例表１）の重回帰式は（式２）より、 $\hat{y} = X\hat{\boldsymbol{\beta}} = -1.186 + 0.386\,x_1 + 0.358\,x_2$ ・・・（例式１） となる。
求めた偏回帰係数は、（例表２）の ⑩ で確認できる。

6. 重回帰分析の計算例（2/3）

■決定係数の算出

（例表1）の観測データから求めた重回帰式（例式1）の決定係数を求める。（例表1）をもとに、それぞれの変動和である （式8） S_R と（式9） S_E と（式7） S_T を求め、（例表3）を作成する。（例表3）の S_R， S_E， S_T の値は、（例表2）の⑥で確認できる。次いで、（例表3）をもとに、（例表4）の分散分析表を作成する。

（式11） を使って、決定係数を求める。 $R^2 = 1 - S_E / S_T = 1 - 0.046 / 29.000 = 0.998$　（例表2）の②で確認できる。

また、変動の平方和の自由度は、$n = 4$ ， $p = 2$ であるので、$\phi_R = p = 2$, $\phi_E = n - p - 1 = 4 - 2 - 1 = 1$, $\phi_T = \phi_R + \phi_E = 2 + 1 = 3$ （例表2）の⑤で確認できる。

（式12） を使って、自由度調整済重決定係数を求める。 $R^{*2} = 1 - (S_E / \phi_E) / (S_T / \phi_T) = 1 - (0.046 / 1) / (29.000 / 3) = 0.995$　（例表2）の③で確認できる。

（式15） を使って、重相関係数を求める。 $\rho = R = 0.998^{1/2} = 0.999$　（例表2）の①で確認できる。

また、（付式30）を使って、$S^2 = (\boldsymbol{\varepsilon}^T \boldsymbol{\varepsilon}) / (n - p - 1) = S_E / (n - p - 1) = 0.046 / (4 - 2 - 1) = 0.046$

よって、重回帰式の標準誤差 $= S = (S^2)^{1/2} = (0.046)^{1/2} = 0.214$

この重回帰式の標準誤差は、（例表2）の④で確認できる。

（例表3）観測データの推定値と平方和

目的変数 y	説明変数 x_1	説明変数 x_2	推定値 $\hat{y}$	回帰変動	残差変動	全変動
$y_1 = 5$	$x_{11} = 10$	$x_{12} = 7$	$\hat{y}_1 = 5.178$	1.749	0.032	2.25
$y_2 = 10$	$x_{21} = 15$	$x_{22} = 15$	$\hat{y}_2 = 9.971$	12.050	0.001	12.25
$y_3 = 8$	$x_{31} = 20$	$x_{32} = 4$	$\hat{y}_3 = 7.960$	2.131	0.002	2.25
$y_4 = 3$	$x_{41} = 5$	$x_{42} = 6$	$\hat{y}_4 = 2.891$	13.024	0.012	12.25
$\bar{y} = 6.500$			$\bar{\hat{y}} = 6.500$	$S_R = 28.954$	$S_E = 0.046$	$S_T = 29.000$

（例表4）観測データの分散分析表

	自由度	変動和	平均変動和	分散比	p 値
回帰変動	$\phi_R = 2$	$S_R = 28.954$	$S_R / \phi_R = 14.477$	$(S_R / \phi_R) / (S_E / \phi_E) = 315.781$	0.0398
残差変動	$\phi_E = 4 - 2 - 1 = 1$	$S_E = 0.046$	$S_E / \phi_E = 0.046$		
全変動	$\phi_T = \phi_R + \phi_E = 2 + 1 = 3$	$S_T = 29.000$			

■重回帰式の検定

（例表4）の分散分析表を利用して、重回帰式の検定を行う。

回帰の平均変動和 $S_R / \phi_R = 28.854 / 2 = 14.477$, 残差の平均変動和 $S_E / \phi_E = 0.046 / 1 = 0.046$　（例表2）の⑦で確認できる。

従って、（式20）で表す検定統計量 $T = (S_R / \phi_R) / (S_E / \phi_E) = 315.781$　（例表2）の⑧で確認できる。

検定統計量 T は自由度 $(\phi_R , \phi_E) = (2, 1)$ の F 分布に従うので、p 値（有意 F 値）は、Excel を使用して、p 値 = 1 − F.DIST(315.781, 2, 1, true) = 0.0398 ≦0.05

この p 値は、（例表2）の⑨で確認できる。

なお、上側棄却域 5 % に相当する統計量は $F_{0.05}(2, 1) = \text{F.INV}(1 - 0.05, 2, 1) = 199.5 < 315.781$

従って、帰無仮説$(\beta_0 = \beta_1 = \beta_2 = 0)$は棄却され、重回帰式は有用であることが確認される。

6. 重回帰分析の計算例(3/3)

■個々の偏回帰係数の検定と推定(例表5)

(式29)$T = \hat{\beta}_i \,/\, (S^2 a_{jj})^{1/2}$　を使用して、個々の偏回帰係数の検定統計量 T を求める。

$$(X^T X)^{-1} = \begin{bmatrix} \mathbf{2.306} & -0.096 & -0.107 \\ -0.096 & \mathbf{0.008} & -0.001 \\ -0.107 & -0.001 & \mathbf{0.014} \end{bmatrix}$$

であり、a_{jj} は行列 $(X^T X)^{-1}$ の j + 1, j + 1 の要素であるので

$a_{00} = (X^T X)^{-1}{}_{1,1} = 2.306$, $a_{11} = (X^T X)^{-1}{}_{2,2} = 0.008$, $a_{22} = (X^T X)^{-1}{}_{3,1} = 0.014$ である。

また、前述したとおり、$S^2 = 0.046$

(1)偏回帰係数の推定値 β の信頼区間を求める。

(式27) $\hat{\beta}_j - t_{\alpha/2}(n-p-1)\,(S^2 a_{jj})^{1/2} \leqq \beta_j \leqq \hat{\beta}_j + t_{1-\alpha/2}(n-p-1)\,(S^2 a_{jj})^{1/2}$ を使って、偏回帰係数の β の信頼度 95% の推定の信頼区間を求める。

$t_{\alpha/2}(n-p-1) = t_{0.05/2}(4-2-1) = t_{0.025}(1) = \mathrm{T.INV}(1-0.025, 1) = 12.706$

例えば、j = 0 のとき、　$\hat{\beta}_0 - t_{\alpha/2}(n-p-1)\,(S^2 a_{00})^{1/2} = -1.186 - 12.706 * (0.046 * 2.306)^{1/2} = -5.317$

$\hat{\beta}_0 + t_{\alpha/2}(n-p-1)\,(S^2 a_{00})^{1/2} = -1.186 + 12.706 * (0.046 * 2.306)^{1/2} = 2.945$

従って、偏回帰係数の推定値 β_0 の信頼度 95% の信頼区間は、$-5.317 \leqq \beta_0 \leqq 2.945$

β_1, β_2 の信頼区間も同様に求めることができる。 結果を(例表5)に示す。(例表2)の⑭で確認できる。

(2)偏回帰係数の推定値 β の検定を行う。

$\hat{\beta}_0$ の標準誤差は、$(S^2 a_{00})^{1/2} = (0.046 * 2.306)^{1/2} = 0.325$, $(S^2 a_{11})^{1/2} = (0.046 * 0.008)^{1/2} = 0.019$, $(S^2 a_{00})^{1/2} = (0.046 * 0.014)^{1/2} = 0.026$ となる。

(例表2)の⑪で確認できる。

(式29) $T = \hat{\beta}_j \,/\, (S^2 a_{jj})^{1/2}$ を使って、偏回帰係数の推定値 $\hat{\beta}_j$ の検定を行う(帰無仮説は、$\beta_j = 0$ である)。

例えば、j = 0 のとき、$T = \hat{\beta}_0 \,/\, (S^2 a_{00})^{1/2} = -1.186 \,/\, 0.0325 = -3.648$　この統計量は、自由度 $n - p - 1 = 4 - 2 - 1 = 1$ の t 分布に従う。

この統計量に対応する p 値は両側に棄却域を設けるので、2 * T.DIST (− 3.648, 1, true) = 0.170 である。

$\hat{\beta}_1$, $\hat{\beta}_2$ の検定用の統計量も同様に求めることができる。 結果を(例表5)に示す。(例表2)の⑫で確認できる。

検定結果としては、$\hat{\beta}_0$ の p 値 = 0.170 > 0.05 , $\hat{\beta}_1$ の p 値 = 0.032 ≦ 0.05 , $\hat{\beta}_0$ の p 値 = 0.045 ≦ 0.05 となる(p 値は(例表2)の⑬で確認できる)。

$\hat{\beta}_0 = 0$ は棄却できない(受容する)。 $\hat{\beta}_1 = 0$ と $\hat{\beta}_1 = 0$ は棄却して、$\hat{\beta}_1 = 0.386$ と $\hat{\beta}_1 = 0.358$ は採択できるという結果である。

(例表5)個々の偏回帰係数の推定と検定

			推定		検定	
$\hat{\beta}$	$(X^T X)^{-1}$ の対角要素	標準誤差 $(S^2 a_{ii})^{1/2}$	信頼度 95% の信頼区間(下側) $\hat{\beta}_i - t_{\alpha/2}(n-p-1)\,(S^2 a_{ii})^{1/2}$	信頼度 95% の信頼区間(上側) $\hat{\beta}_i + t_{\alpha/2}(n-p-1)\,(S^2 a_{ii})^{1/2}$	検定統計量 T $\hat{\beta}_i / (S^2 a_{ii})^{1/2}$	p 値 (Excel を使用)
$\hat{\beta}_0 = -1.186$	$a_{00} = 2.306$	$(S^2 a_{00})^{1/2} = 0.325$	$-5.317 \leqq \beta_0$	$\beta_0 \leqq 2.945$	$\hat{\beta}_0 / (S^2 a_{00})^{1/2} = -3.648$	2*T.DIST(−3.648, 1, true) = 0.170
$\hat{\beta}_1 = 0.386$	$a_{11} = 0.008$	$(S^2 a_{11})^{1/2} = 0.019$	$0.142 \leqq \beta_0$	$\beta_0 \leqq 0.629$	$\hat{\beta}_1 / (S^2 a_{11})^{1/2} = 20.110$	2*T.DIST(−3.648, 1, true) = 0.032
$\hat{\beta}_2 = 0.358$	$a_{22} = 0.014$	$(S^2 a_{22})^{1/2} = 0.026$	$0.033 \leqq \beta_0$	$\beta_0 \leqq 0.684$	$\hat{\beta}_2 / (S^2 a_{22})^{1/2} = 13.975$	2*T.DIST(−3.648, 1, true) = 0.045

7. 標準偏回帰係数の計算例

標準偏回帰係数の計算方法を２通り示す。

（１）説明変数と目的変数を標準化した後に、偏回帰係数を求める。
こうして求めた偏回帰係数が標準偏回帰係数になる（例表６参照）。
標準化した説明変数 X^* と目的変数 y^* を対象に、（式５） $\hat{\boldsymbol{\beta}}^* = (X^{*T} X^*)^{-1} X^{*T} y^*$ を使用して、標準偏回帰係数の推定値 $\hat{\boldsymbol{\beta}}^*$ を求めて、重回帰式を作成する。

$$X^* = [\ \mathbf{e}\ \ x_1^*\ \ x_2^*\ ] = \begin{bmatrix} 1 & x_{11}^* & x_{12}^* \\ 1 & x_{21}^* & x_{22}^* \\ 1 & x_{31}^* & x_{32}^* \\ 1 & x_{41}^* & x_{42}^* \end{bmatrix} = \begin{bmatrix} 1 & -0.387 & -0.207 \\ 1 & 0.387 & 1.4494 \\ 1 & 1.162 & -0.828 \\ 1 & -1.162 & -0.414 \end{bmatrix},\quad X^{*T} X^* = \begin{bmatrix} 1 & 1 & 1 & 1 \\ -0.387 & 0.387 & 1.162 & -1.162 \\ -0.207 & 1.149 & -0.828 & -0.414 \end{bmatrix} \begin{bmatrix} 1 & -0.387 & -0.207 \\ 1 & 0.387 & 1.4494 \\ 1 & 1.162 & -0.828 \\ 1 & -1.162 & -0.414 \end{bmatrix} = \begin{bmatrix} 4 & 0 & 0 \\ 1 & 3 & 0.16 \\ 0 & 0.16 & 3 \end{bmatrix}$$

$X^{*T} X^*$ の逆行列 $(X^{*T} X^*)^{-1}$ を、Excell の逆行列関数 MINVERSE を使用して求める。

$X^{*T} X^*$ の逆行列 $(X^{*T} X^*)^{-1}$ は、$(X^{*T} X^*)^{-1} = \begin{bmatrix} 0.25 & 0.0 & 0.0 \\ -0.084 & 0.334 & -0.018 \\ -0.004 & -0.018 & 0.334 \end{bmatrix}$

$X^{*T} X^*$ の逆行列 $(X^{*T} X^*)^{-1}$ を使って、（式５）より、$\hat{\boldsymbol{\beta}}^* = (X^{*T} X^*)^{-1} X^{*T} y^* = \begin{bmatrix} 0.25 & 0.0 & 0.0 \\ -0.084 & 0.334 & -0.018 \\ -0.004 & -0.018 & 0.334 \end{bmatrix} \begin{bmatrix} 1 & 1 & 1 & 1 \\ -0.387 & 0.387 & 1.162 & -1.162 \\ -0.207 & 1.149 & -0.828 & -0.414 \end{bmatrix} \begin{bmatrix} -0.482 \\ 1.126 \\ 0.482 \\ -1.126 \end{bmatrix} = \begin{bmatrix} 0 \\ 0.801 \\ 0.556 \end{bmatrix}$

標準偏回帰係数は、$\hat{\beta}^*_1 = 0.801$, $\hat{\beta}^*_2 = 0.556$ となる。
従って、（例表６）の重回帰式は（式２）より、 $\hat{y}^* = X^* \hat{\boldsymbol{\beta}}^* = 0.801\ x^*_1 + 0.556\ x^*_2$ となる。

（例表６）説明変数と目的変数の標準化

目的変数 y	平方和 $(y_i - \bar{y})^2$	標準化した目的変数 y^*	説明変数 x_1	平方和 $(x_{i1} - \bar{x}_1)^2$	標準化した説明変数 x^*_1	説明変数 x_2	平方和 $(x_{i2} - \bar{x}_2)^2$	標準化した説明変数 x^*_2
$y_1 = 5$	$(y_1 - \bar{y})^2 = 2.25$	$y^*_1 = (y_1 - \bar{y})\ /\ \sigma_y$ $= (5 - 6.5)\ /\ 9.667^{1/2}$ $= -0.482$	$x_{11} = 10$	$(x_{11} - \bar{x}_1)^2 = 6.25$	$x^*_{11} = (x_{11} - \bar{x}_1)\ /\ \sigma_{x_1}$ $= (10 - 12.5)\ /\ 41.667^{1/2}$ $= -0.387$	$x_{12} = 7$	$(x_{12} - \bar{x}_2)^2 = 1$	$x^*_{12} = (x_{12} - \bar{x}_2)\ /\ \sigma_{x_2}$ $= (7 - 8)\ /\ 23.333^{1/2}$ $= -0.207$
$y_2 = 10$	$(y_2 - \bar{y})^2 = 12.25$	$y^*_2 = 1.126$	$x_{21} = 15$	$(x_{21} - \bar{x}_1)^2 = 6.25$	$x^*_{21} = 0.387$	$x_{22} = 15$	$(x_{22} - \bar{x}_2)^2 = 49$	$x^*_{22} = 1.449$
$y_3 = 8$	$(y_3 - \bar{y})^2 = 2.25$	$y^*_3 = 0.482$	$x_{31} = 20$	$(x_{31} - \bar{x}_1)^2 = 56.25$	$x^*_{31} = 1.162$	$x_{32} = 4$	$(x_{32} - \bar{x}_2)^2 = 16$	$x^*_{32} = -0.828$
$y_4 = 3$	$(y_4 - \bar{y})^2 = 12.25$	$y^*_4 = -1.126$	$x_{41} = 5$	$(x_{41} - \bar{x}_1)^2 = 56.25$	$x^*_{41} = -1.162$	$x_{42} = 6$	$(x_{42} - \bar{x}_2)^2 = 4$	$x^*_{42} = -0.414$
$\bar{y} = 6.5$	$\sigma_y^2 = \Sigma(y_i - \bar{y})^2$ $/\ (4-1) = 9.667$	$\bar{y}^* = 0.0$ $\sigma_{y^*} = 1.0$	$\bar{x}_1 = 12.5$	$\sigma_{x_1}^2 = \Sigma(x_{i1} - \bar{x}_1)^2$ $/\ (4-1) = 41.667$	$\bar{x}^*_1 = 0.0$ $\sigma_{x^*_1} = 1.0$	$\bar{x}_2 = 8$	$\sigma_{x_2}^2 = \Sigma(x_{i2} - \bar{x}_2)^2$ $/\ (4-1) = 23.333$	$\bar{x}^*_2 = 0.0$ $\sigma_{x^*_2} = 1.0$

（２）偏回帰係数を使って標準偏回帰係数を求める。
（式６） $\hat{\beta}^*_j = (\sigma_{x_j} / \sigma_y)\ \hat{\beta}_j$ を使用して、標準偏回帰係数を求める。
（例式１）より $\hat{\beta}_1 = 0.386$, $\hat{\beta}_2 = 0.358$
また、（付表６）より $\sigma_y = (9.667)^{1/2} = 3.109$, $\sigma_{x_1} = (41.667)^{1/2} = 6.455$, $\sigma_{x_2} = (23.333)^{1/2} = 4.830$ であるので、

$\hat{\beta}^*_1 = (\sigma_{x_1} / \sigma_y)\ \hat{\beta}_1 = (6.455\ /\ 3.109) * 0.386 = 0.801$
$\hat{\beta}^*_2 = (\sigma_{x_2} / \sigma_y)\ \hat{\beta}_2 = (4.830\ /\ 3.109) * 0.358 = 0.556$

この結果は、（１）の結果に等しい。

8．目的変数の信頼区間と予測区間（1/2）

推定した偏回帰係数 $\hat{\beta}$ を使用して偏回帰係数 $\boldsymbol{\beta}$ の信頼区間を求める方法は、「5. 個々の偏回帰係数の推定と検定」で取り上げた。
ここでは、推定した偏回帰係数 $\hat{\beta}$ を使用して、説明変数が特定の値 $\mathbf{x}_a = [1\ \ x_a]$ を取ったときに、(1) 目的変数 $y_a = \mathbf{x}_a\, \boldsymbol{\beta}$ の推定区間（信頼区間という）を求める方法と、(2) 目的変数 $y_{a'} = y_a + u_a$ という観測誤差 u_a を加味した推定区間（予測区間という）を求める方法を整理しておく。

(1) 信頼区間 ： $\hat{y}_a = \mathbf{x}_a\, \hat{\boldsymbol{\beta}}$ の推定区間

説明変数が $\mathbf{x}_a$ の値のとき、目的変数の推定値 $\hat{y}_a$ は（式2）より、$\hat{y}_a = \mathbf{x}_a\, \hat{\boldsymbol{\beta}}$ ・・・（式30） 但し、$\mathbf{x}_a$ は行列 $\mathbf{X}$ の行ベクトル $\mathbf{x}_a = [1\ \ x_{1a}\ \cdots\ x_{na}]$

正規分布に従う $\hat{\boldsymbol{\beta}}$ の線形和である $\hat{y}_a$ も、正規分布の和の再生性により正規分布に従う（別途資料「正規分布の概要」を参照）。

$\hat{y}_a$ の期待値： （付式20） $E[\hat{\boldsymbol{\beta}}] = \boldsymbol{\beta}$ を利用して、$E[\hat{y}_a] = E[\mathbf{x}_a\, \hat{\boldsymbol{\beta}}] = \mathbf{x}_a\, E[\hat{\boldsymbol{\beta}}] = \mathbf{x}_a\, \boldsymbol{\beta} = y_a$ ・・・（式31）

$\hat{y}_a$ の分散： （付式21） $V[\boldsymbol{\beta}\hat{\ }] = \sigma^2 (\mathbf{X}^T \mathbf{X})^{-1}$ と（式30）、（式31）を利用して、

$$V[\hat{y}_a] = V[\mathbf{x}_a \hat{\boldsymbol{\beta}}] = E[(\mathbf{x}_a \hat{\boldsymbol{\beta}} - E[\mathbf{x}_a \hat{\boldsymbol{\beta}}])(\mathbf{x}_a \hat{\boldsymbol{\beta}} - E[\mathbf{x}_a \hat{\boldsymbol{\beta}}])^T] = E[(\mathbf{x}_a \hat{\boldsymbol{\beta}} - \mathbf{x}_a \boldsymbol{\beta})(\mathbf{x}_a \hat{\boldsymbol{\beta}} - \mathbf{x}_a \boldsymbol{\beta})^T] = E[\mathbf{x}_a (\hat{\boldsymbol{\beta}} - \boldsymbol{\beta})(\mathbf{x}_a (\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}))^T] = E[\mathbf{x}_a (\hat{\boldsymbol{\beta}} - \boldsymbol{\beta})(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta})^T \mathbf{x}_a^T]$$
$$= \mathbf{x}_a E[(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta})(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta})^T]\, \mathbf{x}_a^T = \mathbf{x}_a V[\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}]\, \mathbf{x}_a^T = \mathbf{x}_a V[\hat{\boldsymbol{\beta}}]\, \mathbf{x}_a^T = \mathbf{x}_a\, \sigma^2 (\mathbf{X}^T \mathbf{X})^{-1} \mathbf{x}_a^T = \sigma^2 \mathbf{x}_a (\mathbf{X}^T \mathbf{X})^{-1} \mathbf{x}_a^T \quad \cdots \text{（式32）}$$

① 正規分布に従う $\hat{y}_a$ を標準化した $(\hat{y}_a - E[\hat{y}_a]) / (V[\hat{y}_a])^{1/2} = (\mathbf{x}_a \hat{\boldsymbol{\beta}} - y_a) / (\sigma^2 \mathbf{x}_a (\mathbf{X}^T \mathbf{X})^{-1} \mathbf{x}_a^T)^{1/2}$ は、標準正規分布に従う。

② （式25）より、$\boldsymbol{\varepsilon}^T \boldsymbol{\varepsilon} / \sigma^2$ は、自由度 $n-p-1$ のカイ二乗分布に従う。

③ 従って、（付録14）を参考にして①と②から、また、（付式30） $S^2 = \boldsymbol{\varepsilon}^T \boldsymbol{\varepsilon} / (n-p-1)$ を使って、

$$T = ((\mathbf{x}_a \hat{\boldsymbol{\beta}} - y_a) / (\sigma^2 \mathbf{x}_a (\mathbf{X}^T \mathbf{X})^{-1} \mathbf{x}_a^T)^{1/2}) / ((\boldsymbol{\varepsilon}^T \boldsymbol{\varepsilon} / \sigma^2) / (n-p-1))^{1/2} = (\mathbf{x}_a \hat{\boldsymbol{\beta}} - y_a) / (S^2 \mathbf{x}_a (\mathbf{X}^T \mathbf{X})^{-1} \mathbf{x}_a^T)^{1/2} \quad \cdots \text{（式33）}$$

は、自由度 $n-p-1$ の t 分布に従うことが分かる。

この結果を使って、目的変数 $y_a = \mathbf{x}_a\, \boldsymbol{\beta}$ の信頼区間を求める。信頼度 95% の信頼区間は、自由度 $n-p-1$ の t 分布では、上側 2.5 % $= t_{0.025}(n-p-1)$，
下側 2.5 % $= t_{0.975}(n-p-1) = -t_{0.025}(n-p-1)$ で表すことができる（別途資料「統計的推定の手順」を参照）。
従って、（式33）より、 $-t_{0.025}(n-p-1) \leqq (\mathbf{x}_a \hat{\boldsymbol{\beta}} - y_a) / (S^2 \mathbf{x}_a (\mathbf{X}^T \mathbf{X})^{-1} \mathbf{x}_a^T)^{1/2} \leqq t_{0.025}(n-p-1)$

よって、$\mathbf{x}_a\, \boldsymbol{\beta}$ の 95 % の信頼範囲は、$\mathbf{x}_a \hat{\boldsymbol{\beta}} - t_{0.025}(n-p-1)(S^2 \mathbf{x}_a (\mathbf{X}^T \mathbf{X})^{-1} \mathbf{x}_a^T)^{1/2} \leqq y_a \leqq \mathbf{x}_a \hat{\boldsymbol{\beta}} + t_{0.025}(n-p-1)(S^2 \mathbf{x}_a (\mathbf{X}^T \hat{\boldsymbol{\beta}}]\, \mathbf{X})^{-1} \mathbf{x}_a^T)^{1/2}$ ・・・（式34）

(2) 予測区間 ： $\hat{y}_{a'} = \hat{y}_a + u_a$ の推定区間

ここでは、説明変数が $\mathbf{x}_a$ の値のとき、目的変数は $y_{a'}$ であるとしているので、目的変数の推定値 $\hat{y}_{a'}$ は、$\hat{y}_{a'} = \mathbf{x}_a\, \hat{\boldsymbol{\beta}}$ である。

$\hat{y}_{a'}$ の期待値： （式31）を利用して、また、$\mathbf{x}_a$ のとき $y_{a'}$ であるとしているので、$E[\hat{y}_{a'}] = E[\mathbf{x}_a \hat{\boldsymbol{\beta}}] = \mathbf{x}_a E[\hat{\boldsymbol{\beta}}] = \mathbf{x}_a\, \boldsymbol{\beta} = y_{a'}$ ・・・（式35）

$\hat{y}_{a'}$ の分散： （式32）を利用して、$V[\hat{y}_{a'}] = V[\hat{y}_a + u_a] = V[\hat{y}_a] + V[u_a] = \sigma^2 \mathbf{x}_a (\mathbf{X}^T \mathbf{X})^{-1} \mathbf{x}_a^T + \sigma^2 = \sigma^2 (\mathbf{x}_a (\mathbf{X}^T \mathbf{X})^{-1} \mathbf{x}_a^T + 1)$ ・・・（式36）

① 正規分布に従う $\hat{y}_{a'}$ を標準化した $(\hat{y}_{a'} - E[\hat{y}_{a'}]) / (V[\hat{y}_{a'}])^{1/2} = (\mathbf{x}_a \hat{\boldsymbol{\beta}} - y_{a'}) / (\sigma^2 (\mathbf{x}_a (\mathbf{X}^T \mathbf{X})^{-1} \mathbf{x}_a^T + 1))^{1/2}$ は、標準正規分布に従う。

② （式25）より、$\boldsymbol{\varepsilon}^T \boldsymbol{\varepsilon} / \sigma^2$ は、自由度 $n-p-1$ のカイ二乗分布に従う。

③ 従って、（付録14）を参考にして、①と②から、また、（付式30） $S^2 = \boldsymbol{\varepsilon}^T \boldsymbol{\varepsilon} / (n-p-1)$ を使って、

$$T = ((\mathbf{x}_a \hat{\boldsymbol{\beta}} - y_{a'}) / (\sigma^2 (\mathbf{x}_a (\mathbf{X}^T \mathbf{X})^{-1} \mathbf{x}_a^T + 1))^{1/2}) / ((\boldsymbol{\varepsilon}^T \boldsymbol{\varepsilon} / \sigma^2) / (n-p-1))^{1/2} = (\mathbf{x}_a \hat{\boldsymbol{\beta}} - y_{a'}) / (S^2 (\mathbf{x}_a (\mathbf{X}^T \mathbf{X})^{-1} \mathbf{x}_a^T + 1))^{1/2} \quad \cdots \text{（式37）}$$

は、自由度 $n-p-1$ の t 分布に従うことが分かる。

この結果を使って、目的変数 $y_{a'} = y_a + u_a$ の予測区間を求める。
信頼度 95% の予測区間は、自由度 $n-p-1$ の t 分布では、上側 2.5 % $= t_{0.025}(n-p-1)$，下側 2.5 % $= t_{0.975}(n-p-1) = -t_{0.025}(n-p-1)$ で表すことができる
（別途資料「統計的推定の手順」を参照）。
従って、$-t_{0.025}(n-p-1) \leqq (\mathbf{x}_a \hat{\boldsymbol{\beta}} - y_{a'}) / (S^2 (\mathbf{x}_a (\mathbf{X}^T \mathbf{X})^{-1} \mathbf{x}_a^T + 1)^{1/2} \leqq t_{0.025}(n-p-1)$

よって、y_a の 95 % の信頼区間は、$\mathbf{x}_a \hat{\boldsymbol{\beta}} - t_{0.025}(n-p-1)(S^2 (\mathbf{x}_a (\mathbf{X}^T \mathbf{X})^{-1} \mathbf{x}_a^T + 1)^{1/2} \leqq y_{a'} \leqq \mathbf{x}_a \hat{\boldsymbol{\beta}} + t_{0.025}(n-p-1)(S^2 (\mathbf{x}_a (\mathbf{X}^T \hat{\boldsymbol{\beta}}]\, \mathbf{X})^{-1} \mathbf{x}_a^T + 1)^{1/2}$ ・・・（式38）

8. 目的変数の信頼区間と予測区間(2/2) 計算例

(例表7)に示す観測データの単回帰分析を行って、目的変数の信頼区間と予測区間を求めて図示する。

(例表7)観測データと推定残差

目的変数 y	説明変数 x_1	推定値 $\hat{y}$	推定残差 ε
$y_1 = 90$	$x_{11} = 5$	$\hat{y}_1 = 72.968$	$\varepsilon_1 = 17.032$
$y_2 = 120$	$x_{21} = 10$	$\hat{y}_2 = 122.248$	$\varepsilon_2 = -2.248$
$y_3 = 170$	$x_{31} = 15$	$\hat{y}_3 = 171.527$	$\varepsilon_3 = -1.527$
$y_4 = 40$	$x_{41} = 3$	$\hat{y}_4 = 53.256$	$\varepsilon_4 = -13.256$

(1)偏回帰係数 $\boldsymbol{\beta}$ と σ^2 の不偏推定量 S^2 を求める。

(式5) $\hat{\boldsymbol{\beta}} = (\mathbf{X}^T\mathbf{X})^{-1}\mathbf{X}^T\mathbf{y}$ を使用して、偏回帰係数の推定値 $\hat{\boldsymbol{\beta}}$ を求めて、単回帰式を作成する。

$$\mathbf{X} = [\mathbf{e}\ \ \mathbf{x}] = \begin{bmatrix} 1 & x_{11} \\ 1 & x_{21} \\ 1 & x_{31} \\ 1 & x_{41} \end{bmatrix} = \begin{bmatrix} 1 & 5 \\ 1 & 10 \\ 1 & 15 \\ 1 & 3 \end{bmatrix},\quad \mathbf{X}^T\mathbf{X} = \begin{bmatrix} 1 & 1 & 1 & 1 \\ 5 & 10 & 15 & 3 \end{bmatrix}\begin{bmatrix} 1 & 5 \\ 1 & 10 \\ 1 & 15 \\ 1 & 3 \end{bmatrix} = \begin{bmatrix} 1 & 33 \\ 33 & 359 \end{bmatrix},\quad (\mathbf{X}^T\mathbf{X})^{-1} = \begin{bmatrix} 1.035 & -0.095 \\ -0.095 & 0.012 \end{bmatrix}$$

$$\hat{\boldsymbol{\beta}} = (\mathbf{X}^T\mathbf{X})^{-1}\mathbf{X}^T\mathbf{y} = \begin{bmatrix} 1.035 & -0.095 \\ -0.095 & 0.012 \end{bmatrix}\begin{bmatrix} 1 & 1 & 1 & 1 \\ 5 & 10 & 15 & 3 \end{bmatrix}\begin{bmatrix} 90 \\ 120 \\ 170 \\ 40 \end{bmatrix} = \begin{bmatrix} 0.559 & 0.084 & -0.392 & 0.749 \\ -0.037 & 0.020 & 0.078 & -0.061 \end{bmatrix}\begin{bmatrix} 90 \\ 120 \\ 170 \\ 40 \end{bmatrix} = \begin{bmatrix} 23.689 \\ 9.856 \end{bmatrix}$$

単回帰式 $y_i = 23.689 + 9.856\ x_{i1}$ が求まる。この単回帰式を使用して、(例表7)に示す推定残差 $\boldsymbol{\varepsilon}$ が求まる。

$$\text{(付式30)を使って、}S^2 = (\boldsymbol{\varepsilon}^T\boldsymbol{\varepsilon}) / (n-p-1) = [\,17.032\ \ -2.248\ \ -1.527\ \ -13.256\,]\begin{bmatrix} 17.032 \\ -2.248 \\ -1.527 \\ -13.256 \end{bmatrix} / (4-1-1) = 473.199 / 2 = 236.599$$

また、Excel 関数を使用して、$t_{0.025}(n-p-1) = t_{0.025}(4-1-1) = t_{0.025}(2) = \mathrm{T.INV}\ (1-0.025,\ 2) = 4.303$

(2)例えば、$\mathbf{x}_a = [1\ \ 8]$ のとき($x_a = 8$ のとき)、目的変数 y_a の「信頼区間」を求める。

$$\mathbf{x}_a\hat{\boldsymbol{\beta}} = [1\ \ 8]\begin{bmatrix} 23.689 \\ 9.856 \end{bmatrix} = 102.536,\quad (S^2\,\mathbf{x}_a(\mathbf{X}^T\mathbf{X})^{-1}\mathbf{x}_a^T)^{1/2} = (236.599\ [1\ \ 8]\begin{bmatrix} 1.035 & -0.095 \\ -0.095 & 0.012 \end{bmatrix}\begin{bmatrix} 1 \\ 8 \end{bmatrix})^{1/2} = (236.599 * 0.251)^{1/2} = 7.702$$

従って、(式34)より、信頼区間の下側の値は、$\mathbf{x}_a\hat{\boldsymbol{\beta}} - t_{0.025}(n-p-1)\,(S^2\,\mathbf{x}_a(\mathbf{X}^T\mathbf{X})^{-1}\mathbf{x}_a^T)^{1/2} = 102.536 - 4.303 * 7.702 = 69.397$

信頼区間の上側の値は、$\mathbf{x}_a\hat{\boldsymbol{\beta}} + t_{0.025}(n-p-1)\,(S^2\,\mathbf{x}_a(\mathbf{X}^T\mathbf{X})^{-1}\mathbf{x}_a^T)^{1/2} = 102.536 + 4.303 * 7.702 = 135.675$

(3)例えば、$\mathbf{x}_a = [1\ \ 8]$ のとき($x_a = 8$ のとき)、目的変数 $y_{a'}$ の「予測区間」を求める。

$$\mathbf{x}_a\hat{\boldsymbol{\beta}} = [1\ \ 8]\begin{bmatrix} 23.689 \\ 9.856 \end{bmatrix} = 102.536,\quad (S^2\,(\mathbf{x}_a(\mathbf{X}^T\mathbf{X})^{-1}\mathbf{x}_a^T + 1))^{1/2} = (236.599\ ([1\ \ 8]\begin{bmatrix} 1.035 & -0.095 \\ -0.095 & 0.012 \end{bmatrix}\begin{bmatrix} 1 \\ 8 \end{bmatrix} + 1))^{1/2} = (236.599 * (0.251 + 1))^{1/2} = 17.202$$

従って、(式38)より、

予測区間の下側の値は、

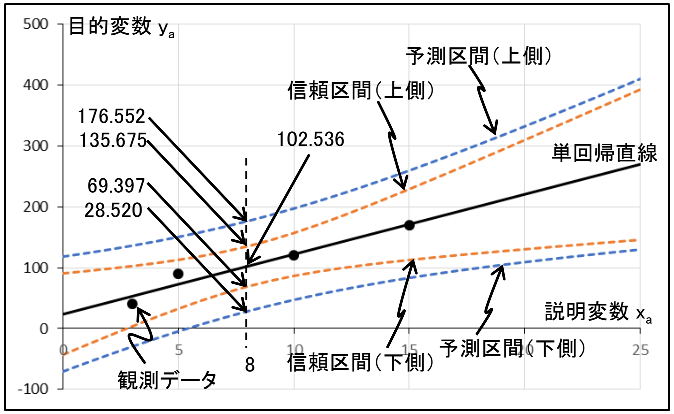
$\mathbf{x}_a\hat{\boldsymbol{\beta}} - t_{0.025}(n-p-1)\,(S^2\,(\mathbf{x}_a(\mathbf{X}^T\mathbf{X})^{-1}\mathbf{x}_a^T + 1))^{1/2} = 102.536 - 4.303 * 17.202 = 28.520$

予測区間の上側の値は、

$\mathbf{x}_a\hat{\boldsymbol{\beta}} + t_{0.025}(n-p\ \ 1)\,(S^2\,(\mathbf{x}_a(\mathbf{X}^T\mathbf{X})^{-1}\mathbf{x}_a^T + 1))^{1/2} = 102.536 + 4.303 * 17.202 = 176.552$

(4)目的変数の信頼区間と予測区間の可視化

単回帰式の説明変数である x_a の値を連続的に変化させることで、
目的変数の信頼区間と予測区間を可視化することができる(例図1)。



(例図1)単回帰式の信頼区間と予測区間の可視化

(例図1)から分かるように、説明変数のデータが存在しないところでは、信頼区間も予測区間も目的変数の区間幅(上下幅)は広がっていく。
回帰式は、データの存在するところ、つまりデータの内挿で使用するのが良く、データの外挿での使用は避けるのが好ましいことが分かる。
例えば、回帰直線を安易に外挿して、縦軸の切片に意味を持たせてはならない。

9. 分散拡大要因(1/3)　相関行列を使った算出方法

VIF_i (Variance Inflation Factor)が持つ意味を示すのが(式39)である。重回帰式(式1)を変形して(式40)で表したとき、説明変数 $\mathbf{x}_j$ の偏回帰係数の推定値 $\hat{\beta}_j$ の分散 $V[\hat{\beta}_j]$ と、説明変数 $\mathbf{x}_j$ だけを使用したときの単回帰式(式42)の偏回帰係数の推定値 $\tilde{\beta}_j$ の分散 $V[\tilde{\beta}_j]$ との比が、説明変数 $\mathbf{x}_j$ の VIF_j になる。VIF_j は特定の説明変数を使った単回帰式を説明変数の数を増やした重回帰式にしたことで、特定の説明変数の偏回帰係数の分散がどれ程大きくなるのか、どれ程推定精度が下がるのかを示している。通常、VIF_j の値が 10 以上で、説明変数 $\mathbf{x}_j$ の数を増やすなかで重回帰式を不安定にする多重共線性(Multicollinearity　略して、マルチコとも言う)が発生していると見なす。多重共線性は説明変数のなかに従属関係がある場合、偏回帰係数を求める(式5)の $(\mathbf{X}^T\mathbf{X})^{-1}$ の逆行列が求まらないことに由来する。相関関係が強い説明変数が存在する場合、この逆行列が不安定になる。これが多重共線性であり、重回帰分析で避けなくてはならない問題である。

$\mathrm{VIF}_j = V[\hat{\beta}_j] \,/\, V[\tilde{\beta}_j]$　・・・(式39)

$\mathbf{y} = \beta_0 + \beta_1 \mathbf{x}_1 + \cdots + \beta_p \mathbf{x}_p + \mathbf{u}$　・・・(式40)　→　推定式　$\hat{\mathbf{y}} = \mathbf{X}\hat{\boldsymbol{\beta}} = \hat{\beta}_0 + \hat{\beta}_1 \mathbf{x}_1 + \cdots + \hat{\beta}_p \mathbf{x}_p$　・・・(式41)

$\mathbf{y} = \beta_0 + \beta_j \mathbf{x}_j + \mathbf{u}$　・・・(式42)　→　推定式　$\tilde{\mathbf{y}} = \mathbf{X}\tilde{\boldsymbol{\beta}} = \tilde{\beta}_0 + \tilde{\beta}_j \mathbf{x}_j$　・・・(式43)

■説明変数の相関行列を使った VIF_i の計算方法　(注:$\sum_{i=1}^{n}$ は、和の範囲を省略して、Σ と表記する)

①標準化した説明変数 $\mathbf{x}^*_j$ と目的変数 $\mathbf{y}^*$ を用意する(別途資料「いろいろな確率分布」の付録を参照)。

$\mathbf{x}^*_j = [x^*_{1j}\ \ x^*_{2j}\ \cdots\ x^*_{nj}]^T$　ただし、$x^*_{ij} = (x_{ij} - \bar{x}_j)/\sigma_{x_j}$,　$\bar{x}_j = 1/n\ \Sigma\ x_{ij}$,　$\sigma_{x_j}{}^2 = 1/n\ \Sigma\ (x_{ij} - \bar{x}_j)^2$　標準化した x^*_{ij} に関して、$E[x^*_{ij}] = \bar{x}^*_j = 0$, $V[x^*_{ij}] = 1$

$\mathbf{y}^* = [y^*_1\ \ y^*_2\ \cdots\ y^*_n]^T$　ただし、$y^*_i = (y_i - \bar{y})/\sigma_y$,　$\bar{y} = 1/n\ \Sigma\ y_i$,　$\sigma_y{}^2 = 1/n\ \Sigma\ (y_i - \bar{y})^2$　標準化した y^*_i に関して、$E[y^*_i] = \bar{y}^* = 0$, $V[y^*_i] = 1$

②さらに、説明変数 $\mathbf{x}^*_j$ と目的変数 $\mathbf{y}^*$ のノルム(付録1)を 1 に正規化する。

$\| \mathbf{x}^*_j \| = (x^*_{1j})^2 + (x^*_{2j})^2 + \cdots + (x^*_{nj})^2 = \Sigma\ (x^*_{ij})^2 = n\ \sigma_{x_j}{}^2$　→　$\| \mathbf{x}^*_j \| = \langle \mathbf{x}^*_j, \mathbf{x}^*_j \rangle = \mathbf{x}^{*T}_j \mathbf{x}^*_j = \Sigma\ (x^*_{ij})^2 = 1$　として扱う

$\| \mathbf{y}^* \| = (y^*_1)^2 + (y^*_2)^2 + \cdots + (y^*_n)^2 = \Sigma\ (y^*_i)^2 = n\ \sigma_y{}^2$ →　$\| \mathbf{y}^* \| = \langle \mathbf{y}^*, \mathbf{y}^* \rangle = \mathbf{y}^{*T}\mathbf{y}^* = \Sigma\ (y^*_i)^2 = 1$　として扱う

③なお、標準化+正規化した説明変数と目的変数を使用して、重回帰式、目的変数の推定式、偏回帰係数の分散を、以下のように表すことができる。

$\mathbf{X}^* = [\mathbf{x}^*_1\ \ \mathbf{x}^*_2\ \cdots\ \mathbf{x}^*_p]$,　$\boldsymbol{\beta}^* = [\beta^*_1\ \ \beta^*_2\ \cdots\ \beta^*_p]$,　$\hat{\boldsymbol{\beta}}^* = [\hat{\beta}^*_1\ \ \hat{\beta}^*_2\ \cdots\ \hat{\beta}^*_p]$　(注、標準化した確率変数を使用するため、定数項の β^*_0 は不要)

$\mathbf{y}^* = \mathbf{X}^*\boldsymbol{\beta}^* + \mathbf{u}^*$　・・・(式44),　$\hat{\mathbf{y}}^* = \mathbf{X}^*\hat{\boldsymbol{\beta}}^*$,　$\mathbf{y}^* = \hat{\mathbf{y}}^* + \boldsymbol{\varepsilon}^*$

$V[\hat{\boldsymbol{\beta}}^*] = \sigma^2 (\mathbf{X}^{*T}\mathbf{X}^*)^{-1}$

相関行列 $\mathbf{R}$ を定義する。

$$\mathbf{R} = \begin{bmatrix} 1 & r_{12} & \cdot & r_{1p} \\ r_{21} & 1 & \cdot & r_{2p} \\ \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot \\ r_{p1} & r_{p2} & \cdot & 1 \end{bmatrix}$$

相関係数に関して、以下を確認することができる。ここでは、x^*_{i1} と x^*_{ij} の相関係数を $r(i1, ij)$ と表す。

$r(i1, ij) = (1/n\ \Sigma\ (x^*_{i1} - \bar{x}^*_1)(x^*_{ij} - \bar{x}^*_j)) \,/\, \{(1/n\ \Sigma\ (x^*_{i1} - \bar{x}^*_1)^2)^{1/2} (1/n\ \Sigma\ (x^*_{ij} - \bar{x}^*_j)^2)^{1/2}\}$

ここで、$\bar{x}^*_1 = 0$, $\bar{x}^*_j = 0$　であるので、

$r(i1, ij) = (1/n\ \Sigma\ x^*_{i1}\, x^*_{ij}) \,/\, \{(1/n\ \Sigma\ x^*_{i1}{}^2)^{1/2} (1/n\ \Sigma\ x^*_{ij}{}^2)^{1/2}\} = (\Sigma\ x^*_{i1}\, x^*_{ij}) \,/\, (\Sigma\ (x^*_{i1})^2\ \Sigma\ (x^*_{ij}{}^2))^{1/2}$

また、$\Sigma\ (x^*_{i1})^2 = 1$,　$\Sigma\ (x^*_{ij})^2 = 1$　であるので、

$r(i1, ij) = \Sigma\ x^*_{i1}\, x^*_{ij}$　なお、$r(ij, ij) = \Sigma\ x^*_{ij}\, x^*_{ij} = \Sigma\ (x^*_{ij})^2 = \| \mathbf{x}^*_j \| = 1$

$\mathbf{X}^{*T}\mathbf{X}^*$ を計算する。

$$\mathbf{X}^{*T}\mathbf{X}^* = \begin{bmatrix} x^*_{11} & x^*_{21} & \cdot & x^*_{n1} \\ x^*_{12} & x^*_{22} & \cdot & x^*_{n2} \\ \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot \\ x^*_{1p} & x^*_{2p} & \cdot & x^*_{np} \end{bmatrix} \begin{bmatrix} x^*_{11} & x^*_{12} & \cdot & x^*_{1p} \\ x^*_{21} & x^*_{22} & \cdot & x^*_{2p} \\ \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot \\ x^*_{n1} & x^*_{n2} & \cdot & x^*_{np} \end{bmatrix} = \begin{bmatrix} \Sigma\ x^*_{i1}{}^2 & \Sigma\ x^*_{i1} x^*_{i2} & \cdot & \Sigma\ x^*_{i1} x^*_{ip} \\ \Sigma\ x^*_{i2} x^*_{i1} & \Sigma\ x^*_{i2}{}^2 & \cdot & \Sigma\ x^*_{i2} x^*_{ip} \\ \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot \\ \Sigma\ x^*_{ip} x^*_{i1} & \Sigma\ x^*_{ip} x^*_{i2} & \cdot & \Sigma\ x^*_{ip}{}^2 \end{bmatrix} = \begin{bmatrix} 1 & r(i1, i2) & \cdot & r(i1, ip) \\ r(i2, i1) & 1 & \cdot & r(i2, ip) \\ \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot \\ r(ip, i1) & r(ip, i2) & \cdot & 1 \end{bmatrix} = \begin{bmatrix} 1 & r_{12} & \cdot & r_{1p} \\ r_{21} & 1 & \cdot & r_{2p} \\ \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot \\ r_{p1} & r_{p2} & \cdot & 1 \end{bmatrix} = \mathbf{R}$$

よって、$V[\hat{\boldsymbol{\beta}}^*] = \sigma^2 (\mathbf{X}^{*T}\mathbf{X}^*)^{-1} = \sigma^2 \mathbf{R}^{-1}$　・・・(式45)　が成り立つ。

そして、説明変数の間に相関関係がない場合は、相関行列 $\mathbf{R}$ の対角要素以外の要素は全て 0 であり、$\mathbf{R} = \mathbf{I}$　→　$\mathbf{R}^{-1} = \mathbf{I}$,　よって、$V[\hat{\boldsymbol{\beta}}^*] = \sigma^2 \mathbf{I}$　・・・(式46)

(式45)と(式46)との比較で、(式45)の $\mathbf{R}^{-1}$ の対角要素は(式46)の対角要素 = 1 に対する比率を表していることが分かる。また(式46)の偏回帰係数の分散は、説明変数間に相関がない場合であるので、(式43)の偏回帰係数の分散と捉えることができる。　従って、$\mathrm{VIF}_j = V[\hat{\beta}_j] \,/\, V[\tilde{\beta}_j] = (\mathbf{R}^{-1})_{jj}$　・・・(付式47)

9. 分散拡大要因(2/3)　決定係数を使った算出方法

ここでは、単回帰式(式43)の説明変数 x_j の決定係数 R_j を使用した VIF_j の計算式(式48)を導出する。

$VIF_j = 1 / (1 - R_j^2)$　・・・(式48)

■決定係数を使った VIF_i の計算方法 (注: $\sum_{i=1}^{n}$ は、和の範囲を省略して、Σ と表記する)

(1) $V[\tilde{\beta}_j]$ の導出

(式42)を対象に、$V[\tilde{\beta}] = \sigma^2 (X^T X)^{-1} = \sigma^2 \left(\begin{bmatrix} 1 & 1 & \cdot & 1 \\ x_{1j} & x_{2j} & \cdot & x_{nj} \end{bmatrix} \begin{bmatrix} 1 & x_{1j} \\ 1 & x_{2j} \\ \cdot & \cdot \\ 1 & x_{nj} \end{bmatrix} \right)^{-1} = \sigma^2 \begin{bmatrix} n & \sum x_{ij} \\ \sum x_{ij} & \sum x_{ij}^2 \end{bmatrix}^{-1} = \sigma^2 / (n \sum x_{ij}^2 - (\sum x_{ij})^2) \begin{bmatrix} \sum x_{ij}^2 & -\sum x_{ij} \\ -\sum x_{ij} & n \end{bmatrix}$

また、$\sum (x_{ij} - \bar{x}_j)^2 = \sum (x_{ij}^2 - 2 x_{ij} \bar{x}_j + \bar{x}_j^2) = \sum x_{ij}^2 - 2 \sum x_{ij} \bar{x}_j + \sum \bar{x}_j^2$

$= \sum x_{ij}^2 - 2 (n \bar{x}_j) \bar{x}_j + n \bar{x}_j^2 = \sum x_{ij}^2 - 2 n \bar{x}_j^2 + n \bar{x}_j^2 = \sum x_{ij}^2 - n \bar{x}_j^2$

よって、$V[\tilde{\beta}_j] = \sigma^2 (X^T X)^{-1}_{jj} = n\ \sigma^2 / (n \sum x_{ij}^2 - (\sum x_{ij})^2)$

$= n\ \sigma^2 / (n \sum x_{ij}^2 - (n \bar{x}_j)^2) = \sigma^2 / (\sum x_{ij}^2 - n \bar{x}_j^2)$

$= \sigma^2 / \sum (x_{ij} - \bar{x}_j)^2$　・・・(式49)

(2) $V[\hat{\beta}_j]$ の導出

説明変数 X_j を目的変数と見なし、他の説明変数を使った重回帰式(式50)を考える。

$x_j = \alpha_0 + \alpha_1 x_1 + \cdots + \alpha_{j-1} x_{j-1} + \alpha_{j+1} x_{j+1} + \cdots + \alpha_p x_p + u$　・・・(式50)

行列 X　　j 列目の列ベクトルを 1 列目に移動する(行列の性質は変わらない)。

$$\begin{bmatrix} 1 & x_{11} & \cdot & x_{1j-1} & x_{1j} & x_{1j+1} & \cdot & x_{1p} \\ \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot \\ 1 & x_{j1} & \cdot & x_{jj-1} & x_{jj} & x_{jj+1} & \cdot & x_{jp} \\ \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot \\ 1 & x_{n1} & \cdot & x_{nj-1} & x_{nj} & x_{nj+1} & \cdot & x_{np} \end{bmatrix} \rightarrow \begin{bmatrix} x_{1j} & 1 & x_{11} & \cdot & x_{1j-1} & x_{1j+1} & \cdot & x_{1p} \\ \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot \\ x_{jj} & 1 & x_{j1} & \cdot & X_{jj-1} & x_{jj+1} & \cdot & x_{jp} \\ \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot \\ x_{nj} & 1 & x_{n1} & \cdot & X_{nj-1} & x_{nj+1} & \cdot & x_{np} \end{bmatrix}$$

ベクトル X_j　　行列 X_{-j}

行列 X の j 列目のベクトルを x_j で表す。
行列 X からベクトル x_j を取り除いた残りの行列を X_{-j} で表す。

(図3)行列の並び替え

(式50)の残差平方和　$\sum (x_{ij} - \hat{x}_j)^2$ は、行列 X の j 番目の行を 1 番目の行に移動させた行列を X_j で、残りの行列を X_{-j} で表すと(図3)、$\hat{\alpha}_j = (X_{-j}^T X_{-j})^{-1} X_{-j}^T X_j$ であるので、

$\sum (x_{ij} - \hat{x}_j)^2 = (X_j - X_{-j} \hat{\alpha}_j)^T (X_j - X_{-j} \hat{\alpha}_j) = (X_j - X_{-j} (X_{-j}^T X_{-j})^{-1} X_{-j}^T X_j)^T (X_j - X_{-j} (X_{-j}^T X_{-j})^{-1} X_{-j}^T X_j)$

$= (X_j^T - X_j^T X_{-j} (X_{-j}^T X_{-j})^{-1} X_{-j}^T) (X_j - X_{-j} (X_{-j}^T X_{-j})^{-1} X_{-j}^T X_j)$

$= X_j^T X_j - X_j^T X_{-j} (X_{-j}^T X_{-j})^{-1} X_{-j}^T X_j - X_j^T X_{-j} (X_{-j}^T X_{-j})^{-1} X_{-j}^T X_j + X_j^T X_{-j} (X_{-j}^T X_{-j})^{-1} X_{-j}^T X_{-j} (X_{-j}^T X_{-j})^{-1} X_{-j}^T X_j$

$= X_j^T X_j - 2 X_j^T X_{-j} (X_{-j}^T X_{-j})^{-1} X_{-j}^T X_j + X_j^T X_{-j} (X_{-j}^T X_{-j})^{-1} X_{-j}^T X_j = X_j^T X_j - X_j^T X_{-j} (X_{-j}^T X_{-j})^{-1} X_{-j}^T X_j$　・・・(式51)

一方、$G = X^T X$ は下記のように表すことができる(図4)。

$G = \begin{bmatrix} F_{j,j} & F_{j,-j} \\ F_{-j,j} & F_{-j,-j} \end{bmatrix}$　但し、$F_{j,j} = X_j^T X_j$, $F_{j,-j} = X_j^T X_{-j}$, $F_{-j,j} = X_{-j}^T X_j$, $F_{-j,-j} = X_{-j}^T X_{-j}$

$$X^T X = \begin{bmatrix} X_j^T \\ X_{-j}^T \end{bmatrix} \begin{bmatrix} X_j & X_{-j} \end{bmatrix} = \begin{bmatrix} X_j^T X_j & X_j^T X_{-j} \\ X_{-j}^T X_j & X_{-j}^T X_{-j} \end{bmatrix}$$

(図4)ブロック行列の演算

ここで、シューア補行列(付録5)を使って、

$G^{-1}_{11} = [F_{j,j} - F_{j,-j} (F_{-j,-j})^{-1} F_{-j,j}]^{-1} = [X_j^T X_j - X_j^T X_{-j} (X_{-j}^T X_{-j})^{-1} X_{-j}^T X_j]^{-1}$

従って、(式51)を参考に、　$G^{-1}_{11} = (\sum (x_{ij} - \hat{x}_j)^2)^{-1}$　　なお、G^{-1}_{11} はスカラーである。

さて、$V[\hat{\beta}_j]$ は、(付式21) $V[\hat{\boldsymbol{\beta}}] = \sigma^2 (X^T X)^{-1}$ の対角要素(分散) $\sigma^2 ((X^T X)^{-1})_{jj}$ である。

また、行列 X の j 番目の列を 1 番目の列に移動した行列で構成されるのが行列 G である。

従って、$V[\hat{\beta}_j] = \sigma^2 ((X^T X)^{-1})_{jj} = \sigma^2 (G^{-1})_{jj} = \sigma^2 G^{-1}_{11} = \sigma^2 (\sum (x_{ij} - \hat{x}_j)^2)^{-1} = \sigma^2 / \sum (x_{ij} - \hat{x}_j)^2$　・・・(式52)

(3) VIF_j の導出

重回帰式(式50)の決定係数は、$R_j^2 = 1 - \sum (x_{ij} - \hat{x}_j)^2 / \sum (x_{ij} - \bar{x}_j)^2$

(式39)に(式49)と(式52)を代入して、

$VIF_j = V[\hat{\beta}_j] / V[\tilde{\beta}_j] = (\sigma^2 / \sum (x_{ij} - \hat{x}_j)^2) / (\sigma^2 / \sum (x_{ij} - \bar{x}_j)^2) / = \sum (x_{ij} - \bar{x}_j)^2 / \sum (x_{ij} - \hat{x}_j)^2 = 1 / (1 - R_j^2)$　・・・(式53)

9. 分散拡大要因(3/3)　計算例

(例表8)の観測データを対象に、VIF_j を(式47)の方法と(式53)の方法の2通りの方法で求める。
なお、説明変数間の関係(相関係数)を、(例図2)に示しておく。

(例表8)観測データ

目的変数 y	説明変数 x_1	説明変数 x_2	説明変数 x_3
$y_1 = 150$	$x_{11} = 10$	$x_{12} = 13$	$x_{13} = 32$
$y_2 = 180$	$x_{21} = 15$	$x_{22} = 16$	$x_{23} = 45$
$y_3 = 220$	$x_{31} = 20$	$x_{32} = 22$	$x_{33} = 21$
$y_4 = 90$	$x_{41} = 5$	$x_{42} = 8$	$x_{43} = 33$
$y_5 = 150$	$x_{51} = 14$	$x_{52} = 16$	$x_{53} = 28$
$y_6 = 210$	$x_{61} = 19$	$x_{62} = 25$	$x_{63} = 44$
$y_7 = 310$	$x_{71} = 23$	$x_{72} = 30$	$x_{73} = 38$
$y_8 = 170$	$x_{81} = 17$	$x_{82} = 20$	$x_{83} = 34$
$y_9 = 180$	$x_{91} = 12$	$x_{92} = 17$	$x_{93} = 29$
$y_{10} = 140$	$x_{101} = 9$	$x_{102} = 11$	$x_{103} = 21$

(A)説明変数の相関行列を使った計算方法(式47)

(1)説明変数の相関行列 **R** を求める.。
Excel の CORELL 関数を使用して 説明変数の相関行列 **R** を求める。

$$\mathbf{R} = \begin{bmatrix} 1 & 0.9641 & 0.2830 \\ 0.9641 & 1 & 0.3417 \\ 0.2830 & 0.3417 & 1 \end{bmatrix}$$

(2)次いで、**R** の逆行列を求める(Excel の MINVERSE 関数を使用)。

$$\mathbf{R}^{-1} = \begin{bmatrix} 14.6894 & -14.4258 & 0.7719 \\ -14.4258 & 15.2990 & -1.1449 \\ 0.7719 & -1.1449 & 1.1727 \end{bmatrix}$$

(3)逆行列 $\mathbf{R}^{-1}$ の対角要素が、VIF_j になる(一括して求まる)。
従って、$VIF_1 = 14.6894$, $VIF_2 = 15.2990$, $VIF_3 = 1.1727$

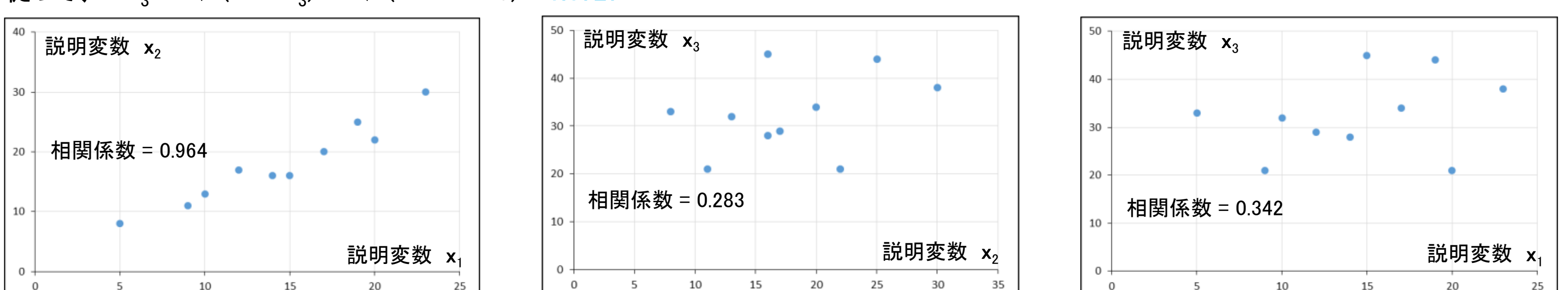
$VIF_1 > 10$, $VIF_2 > 10$ であるので、x_1 と x_2 の両方を説明変数として使用すると、多重共線性が発生すると考えられる。(例図2)から分かるように、x_1 と x_2 の相関関係は強い。
重回帰式を作成する際には、いずれかを説明変数から外すことを考える。

(B)決定係数を使った計算方法(式53)

(1)説明変数 x_1 を目的変数として、残りの x_2 と x_3 を説明変数として重回帰分析する。
Excel の回帰分析の機能を使用して、 この回帰式の決定係数、$R^2_1 = 0.9319$ が求まる。
従って、$VIF_1 = 1 / (1 - R^2_1) = 1 / (1 - 0.9319) = 1.46894$

(2)説明変数 x_2 を目的変数として、残りの x_3 と x_1 を説明変数として重回帰分析する。
Excel の回帰分析の機能を使用して、 この回帰式の決定係数、$R^2_2 = 0.9346$ が求まる。
従って、$VIF_2 = 1 / (1 - R^2_2) = 1 / (1 - 0.9346) = 15.2990$

(3)説明変数 x_3 を目的変数として、残りの x_1 と x_2 を説明変数として重回帰分析する。
Excel の回帰分析の機能を使用して、 この回帰式の決定係数、$R^2_3 = 0.1473$ が求まる。
従って、$VIF_3 = 1 / (1 - R^2_3) = 1 / (1 - 0.1473) = 1.1727$



(例図2)説明変数間の関係

10. ダミー変数(dummy variable)の活用

重回帰分析では、説明変数に量的変数のほか、例えば、資格の有無、経験の有無、曜日、施策の種類などのような質的な変数をダミー変数として使うことができる。

(1)二者択一のダミー変数

どちらか一方を 1 とし、他方を 0 とする。

(例) ・回答ダミー変数 ： Yes → 1， No → 0
・資格ダミー変数 ： 有資格 → 1， 無資格 → 0
・経験ダミー変数 ： 経験済み → 1， 未経験 → 0

(2)複数のなかから選択するダミー変数

(例) 施策A, 施策B, 施策C の3種類の施策があり、どれか1つの施策しか採用できないとき、
・施策Aダミー変数 ： 採用 → 1， 不採用 → 0
・施策Bダミー変数 ： 採用 → 1， 不採用 → 0

なお、施策Cダミー変数は設けない。施策Aダミー変数と施策Bダミー変数が決まれば、施策Cダミー変数は自動で決まる。
施策Aダミー変数 = 1， 施策Bダミー変数 = 0 であれば、施策Cダミー変数 = 0 となる。また、施策Aダミー変数 = 0， 施策Bダミー変数 = 0 であれば、施策Cダミー変数 = 1 となる。つまり、施策Cダミー変数は、施策Aダミー変数と施策Bダミー変数に依存(従属)していることになる。
従属する施策Cダミー変数を設定すれば多重共線性が起こる。よって、複数の選択肢がある場合は、1つ少ない数のダミー変数にする。

(例)説明変数にダミー変数を使うとき

(例表9)観測データ

目的変数 y	説明変数 x_1	説明変数 x_2	説明変数 x_3
230	70	1	0
270	95	0	0
300	100	0	0
240	90	0	1
190	50	1	0
290	90	0	0
250	80	0	1
160	50	1	0
210	70	1	0
170	50	1	0
280	110	0	1
200	80	1	0
180	60	1	0
260	70	0	0
220	60	0	1

説明変数 x_1 は量的変数、三者択一の説明変数をダミー変数 x_2 と x_3 とした(2つ設定)。

①VIF の確認 $VIF_1 = 1.885$， $VIF_2 = 2.648$， $VIF_3 = 1.778$ ＜ 10 であり、多重共線性の心配はない。

②重回帰分析(Excel の回帰分析の機能を活用)

概要

回帰統計		
重相関 R	0.96320563	
重決定 R2	0.92776509	
補正 R2	0.90806466	①
標準誤差	13.5598926	
観測数	15	

(要約)
① 自由度調整済み R2 = 0.908
② 重回帰式のF 検定 の p 値 = 0.000
③ $\hat{\beta}_0$ の t 検定の p 値 = 0.000
④ $\hat{\beta}_1$ の t 検定の p 値 = 0.000
⑤ $\hat{\beta}_2$ の t 検定の p 値 = 0.000
⑥ $\hat{\beta}_3$ の t 検定の p 値 = 0.008

分散分析表

	自由度	変動	分散	測された分散	有意 F	
回帰	3	25977.4224	8659.14082	47.0936449	1.4454E-06	②
残差	11	2022.57755	183.870687			
合計	14	28000				

	係数	標準誤差	t	P-値		下限 95%	上限 95%
切片	170.763332	23.8062389	7.17304959	1.8142E-05	③	118.366154	223.160511
X 値 1	1.29888463	0.25940935	5.00708487	0.00039806	④	0.7279285	1.86984076
X 値 2	-59.123388	11.4201861	-5.177095	0.00030512	⑤	-84.259048	-33.987728
X 値 3	-31.766295	9.9031229	-3.2077048	0.00833953	⑥	-53.562921	-9.9696683

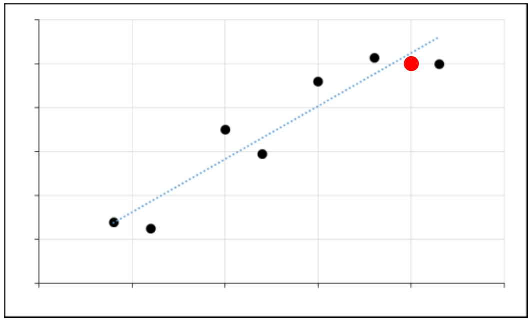
11. 赤池情報量基準を使用した説明変数の選択(1/2)

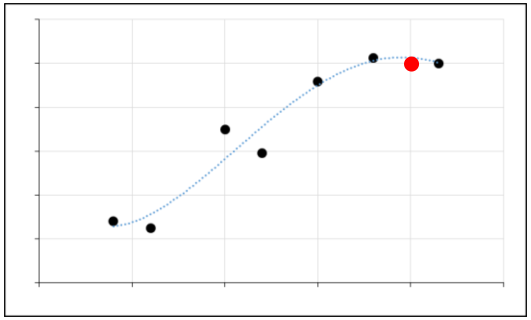
(1)赤池情報量基準(AIC : Akaike's Information Criterion)
AIC は統計モデルの良さを評価するために指標であり、(式54)で表される(尤度に関しては、別途資料「統計的推定の手順」を参照)。
統計モデルごとに AIC の値を計算し、AIC が相対的に(比較して)小さな値をとるほどよい統計モデルであることを示す。

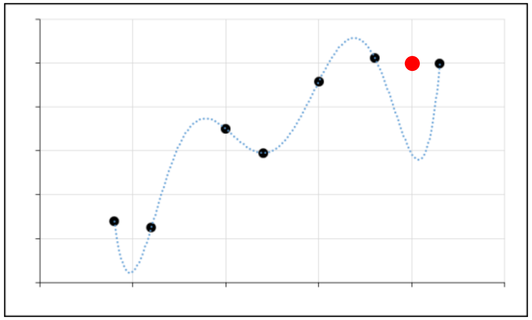
AIC = − 2 ln L + 2 (p + 2)　　ただし、L : 統計モデルの対数最大尤度、 p: パラメータの数(説明変数の数)　　・・・ (式54)

①(式54)の第1項は、観測データへの統計モデルへの適合度合いを表す。 符号はマイナスなので適合度が上がると数値としては小さくなる。
②(式54)の第2項は、パラメータの数を増やすと増加する項として機能する。

観測データの統計モデルへの適合度合いは、パラメータの数や次数を増やすほど良くなる(図5は次数を増やした場合)。
しかし、この場合、統計モデルが複雑になるとともに、観測済みのデータに対しての適合度合いは良くても、新たな観測データに対しての適合度合いに問題が生じる恐れがある(いわゆる過適合、オーバーフィッティングの問題)。従って、パラメータの数や次数をいたずらに増やすことはできない。
そこで、(式54)では、第2項でパラメータの数や次数が増えるのを抑制している。


(a)1次式(直線)の統計モデルを使用した例


(b)3次式の統計モデルを使用した例


(c)6次式の統計モデルを使用した例

(図5)観測データと統計モデルの模式図(Excel のグラフの近似曲線プロットを活用)

統計モデルの次数を上げると観測データへの適合度合いは向上する。この例では、観測データは7つなので、6次の統計モデルを使用すると、統計モデルは観測データと完全に一致する。
しかし。新たな観測データに対しては、この6次の統計モデルの適合度は損なわれることが想像できる。例えば、図中に示すように新たな観測データ ● が得られたとき、このデータは、(a) と (b) では統計モデルに沿っているが、次数の高い統計モデル(c)では大きく逸脱してしまう。このようなことが起こり得る。

(2)重回帰式(重回帰モデル)への AIC の適用
重回帰式における、(式54)の対数最大尤度 ln L は、(付録8)に示す、β の最尤推定値である $\hat{\boldsymbol{\beta}}$ と σ^2 の最尤推定量である $\hat{\sigma}^2 = (1/n)\ \boldsymbol{\varepsilon}^T \boldsymbol{\varepsilon} = S_E/n$ と残差平方和 $S_E = \boldsymbol{\varepsilon}^T \boldsymbol{\varepsilon} = (y - \mathbf{X}\hat{\boldsymbol{\beta}})^T (y - \mathbf{X}\hat{\boldsymbol{\beta}})$ を、(付式24)に代入して得られる。

$$\ln f(\boldsymbol{\beta}, \sigma^2) = -(n/2) \ln \sigma^2 - (1/(2\sigma^2))(y - \mathbf{X}\boldsymbol{\beta})^T (y - \mathbf{X}\boldsymbol{\beta}) - (n/2)\ln(2\pi) \quad \cdots (付式24)$$

$$\begin{aligned}\ln L = \ln f(\hat{\boldsymbol{\beta}}, \hat{\sigma}^2) &= -(n/2)\ln\hat{\sigma}^2 - (1/(2\hat{\sigma}^2))(y - \mathbf{X}\hat{\boldsymbol{\beta}})^T (y - \mathbf{X}\hat{\boldsymbol{\beta}}) - (n/2)\log(2\pi)\\ &= -(n/2)\ln\hat{\sigma}^2 - S_E/(2\hat{\sigma}^2) - (n/2)\ln(2\pi)\\ &= -(n/2)\ln(S_E/n) - S_E/(2S_E/n) - (n/2)\ln(2\pi)\\ &= -(n/2)\ln(S_E/n) - n/2 - (n/2)\ln(2\pi)\\ &= -(n/2)(\ln(2\pi S_E/n) + 1)\end{aligned}$$

これを(式54)に代入して、

$$重回帰式の\ AIC = -2(-(n/2)(\ln(2\pi S_E/n) + 1)) + 2(p + 2) = n(\ln(2\pi S_E/n) + 1) + 2p + 4 \quad \cdots (式55)$$

なお、AIC は絶対値に意味はなく、相対的に大小を比較する指標であるため、使用にあたっては、(式55)の定数項である 4 を割愛しても差し支えない。

11. 赤池情報量基準を使用した説明変数の選択(2/2)

重回帰識の説明変数の選択方法には、知見に基づいて行う方法、自由度調整済みの決定係数を使って行う方法、検定結果を使って行う方法などがあるが、AIC を使用した重回帰式の説明変数の選択方法を例を使って示す。

まず、(1)説明変数全ての組み合せの AIC を求めて、AIC が最小になる説明変数の組み合せで選択する方法を示す。しかし、説明変数の数が増えると(1)の方法では計算回数が増大する。そこで、(2)ステップワイズ法が用意されている。ここでは、各種のステップワイズ法のなかで、変数増減法による説明変数の選択方法を示す。
使用するデータを(例表9)に示す。$VIF_1 = 3.62$, $VIF_2 = 3.90$, $VIF_3 = 2{,}65 < 10$ であり、多重共線性の心配はない。
なお、多重共線性のリスクがあるときには、該当する説明変数を除いてから重回帰分析を行う。

(例表9)観測データ

目的変数	説明変数	説明変数	説明変数
$y_1 = 150$	$x_{11} = 8$	$x_{12} = 15$	$x_{13} = 70$
$y_2 = 180$	$x_{21} = 15$	$x_{22} = 15$	$x_{23} = 45$
$y_3 = 220$	$x_{31} = 20$	$x_{32} = 30$	$x_{33} = 21$
$y_4 = 250$	$x_{41} = 25$	$x_{42} = 35$	$x_{43} = 33$
$y_5 = 100$	$x_{51} = 14$	$x_{52} = 10$	$x_{53} = 80$
$y_6 = 210$	$x_{61} = 19$	$x_{62} = 25$	$x_{63} = 44$
$y_7 = 310$	$x_{71} = 30$	$x_{72} = 40$	$x_{73} = 40$
$y_8 = 170$	$x_{81} = 17$	$x_{82} = 8$	$x_{83} = 60$
$y_9 = 180$	$x_{91} = 12$	$x_{92} = 20$	$x_{93} = 70$
$y_{10} = 140$	$x_{101} = 9$	$x_{102} = 11$	$x_{103} = 75$

(1)説明変数全ての組み合わせの AIC を求めて、AIC が最小になる説明変数の組み合わせを選択する方法
(例表10)に(式55)を使用して求めた、説明変数全ての組み合わせの AIC を示す。この中では、回帰式モデル(e)のAIC の値が最も小さい。よって、説明変数として、x_1 と x_2 を選択することになる。

(2)AIC を使用して、ステップワイズ法(変数増減法)で説明変数の組み合わせを選択する方法

① $\mathbf{y} = \hat{\boldsymbol{\beta}}_0$ の AIC(0) を求める。 AIC(0) = 138.3

② ①のモデルに、説明変数を1つ加える。
$\mathbf{x}_1$ を加えた重回帰式の AIC($\mathbf{x}_1$) = 100.1
$\mathbf{x}_2$ を加えた重回帰式の AIC($\mathbf{x}_2$) = 95.9
$\mathbf{x}_3$ を加えた重回帰式の AIC($\mathbf{x}_3$) = 105.7
なにもしない(説明変数を追加しない) AIC(0) = 138.3
これらのなかから AIC が最小のモデル式を選ぶ → $\mathbf{y} = \hat{\boldsymbol{\beta}}_0 + \hat{\boldsymbol{\beta}}_2\, \mathbf{x}_2$

③ ②のモデルに、説明変数を1つ加える/減らす。
$\mathbf{x}_1$ を加えた重回帰式の AIC($\mathbf{x}_1$, $\mathbf{x}_2$) = 94.7
$\mathbf{x}_3$ を加えた重回帰式の AIC($\mathbf{x}_2$, $\mathbf{x}_3$) = 100.4
なにもしない(説明変数を追加しない) AIC($\mathbf{x}_2$) = 95.9
$\mathbf{x}_2$ を除いた重回帰式の AIC(0) = 138.3
これらのなかから AIC が最小のモデル式を選ぶ → $\mathbf{y} = \hat{\boldsymbol{\beta}}_0 + \hat{\boldsymbol{\beta}}_1\, \mathbf{x}_1 + \hat{\boldsymbol{\beta}}_2\, \mathbf{x}_2$

④ ③のモデルに、説明変数を1つ加える/減らす。
$\mathbf{x}_3$ を加えた重回帰式の AIC($\mathbf{x}_1$, $\mathbf{x}_2$, $\mathbf{x}_3$) = 96.4
なにもしない(説明変数を追加しない) AIC($\mathbf{x}_1$, $\mathbf{x}_2$) = 94.7
$\mathbf{x}_1$ を除いた重回帰式の AIC($\mathbf{x}_2$) = 95.6
$\mathbf{x}_2$ を除いた重回帰式の AIC($\mathbf{x}_1$) = 100.1
これらのなかから AIC が最小のモデル式を選ぶ → $\mathbf{y} = \hat{\boldsymbol{\beta}}_0 + \hat{\boldsymbol{\beta}}_1\, \mathbf{x}_1 + \hat{\boldsymbol{\beta}}_2\, \mathbf{x}_2$
④の結果が③の結果と同じになったので、これが求める重回帰式になる。

(例表10)重回帰モデル式の AIC

重回帰式のモデル	残差平方和	データの数	説明変数の数	AIC	自由度調整済みR2	重回帰式の F 検定の p 値	$\hat{\beta}_0$ の t 検定の p 値	$\hat{\beta}_1$ の t 検定の p 値	$\hat{\beta}_2$ の t 検定の p 値	$\hat{\beta}_{03}$ の t 検定の p 値
(a) $\mathbf{y} = \hat{\boldsymbol{\beta}}_0$	396900	10	0	111.3						
(b) $\mathbf{y} = \hat{\boldsymbol{\beta}}_0 + \hat{\boldsymbol{\beta}}_1 \mathbf{x}_1$	7143.8	10	1	100.1	0.750	0.001	0.045	0.01		
(c) $\mathbf{y} = \hat{\boldsymbol{\beta}}_0 + \hat{\boldsymbol{\beta}}_2 \mathbf{x}_2$	4699.9	10	1	95.9	0.835	0.000	0.001		0.000	
(d) $\mathbf{y} = \hat{\boldsymbol{\beta}}_0 + \hat{\boldsymbol{\beta}}_3 \mathbf{x}_3$	12578.0	10	1	105.7	0.559	0.008	0.000			0.008
(e) $\mathbf{y} = \hat{\boldsymbol{\beta}}_0 + \hat{\boldsymbol{\beta}}_1 \mathbf{x}_1 + \hat{\boldsymbol{\beta}}_2 \mathbf{x}_2$	3410.9	10	2	94.7	0.863	0.000	0.010	0.148	0.028	
(f) $\mathbf{y} = \hat{\boldsymbol{\beta}}_0 + \hat{\boldsymbol{\beta}}_1 \mathbf{x}_1 + \hat{\boldsymbol{\beta}}_3 \mathbf{x}_3$	6045.9	10	2	100.4	0.758	0.003	0.096	0.029		0.297
(g) $\mathbf{y} = \hat{\boldsymbol{\beta}}_0 + \hat{\boldsymbol{\beta}}_2 \mathbf{x}_2 + \hat{\boldsymbol{\beta}}_3 \mathbf{x}_3$	4293.3	10	2	97.0	0.828	0.001	0.052		0.008	0.442
(h) $\mathbf{y} = \hat{\boldsymbol{\beta}}_0 + \hat{\boldsymbol{\beta}}_1 \mathbf{x}_1 + \hat{\boldsymbol{\beta}}_2 \mathbf{x}_2 + \hat{\boldsymbol{\beta}}_3 \mathbf{x}_3$	3311.6	10	3	96.4	0.845	0.002	0.177	0.231	0.068	0.686

12. リッジ回帰（Ridge regression）

説明変数間に従属関係があるとき、（式5） $\hat{\beta} = (X^T X)^{-1} X^T y$ の逆行列 $X^T X^-$ は求まらない。また、従属関係とまでいかないまでも強い相関関係があるとき（多重共線性があるとき）、データによって逆行列 $X^T X^{-1}$ の値が不安定になり偏回帰係数の推定値 $\hat{\beta}$ の値が大きく変動してしまうという問題が起こる。

多重共線性が危惧される場合、偏回帰係数の推定値の変動を抑制して重回帰式を安定化させる手段として正則化という方法が用いられる。リッジ回帰はこの正則化の方法の一つであり、LASSO回帰とともに、（重）回帰分析の過適合を防ぐ目的で、またAI の機械学習における過学習を防ぐ目的で使用される。

（1）リッジ回帰の定式化

偏回帰係数の推定値 $\hat{\beta}$ は、推定残差の二乗和である $\sum_{i=1}^{n} \varepsilon_i^2 = \| y - X\beta \|^2 = (y - X\beta)^T (y - X\beta)$ を最小にする β である。 しかし、多重共線性が存在する場合、ある説明変数のデータを使用して求めた $\hat{\beta}$ の値と新たな説明変数のデータを使用して求めた $\hat{\beta}$ の値が大きく異なることが起こる。そこで、（式56）に示すように、 β の値が大きくならないような制約を設け、その制約の下で残差平方和を最小にする $\hat{\beta}$ を求めるようにする（この工夫を正則化という）。

リッジ回帰は制約に L_2 ノルム（付録1）を使用したものでそのイメージを（図6）に示す。 なお、ラッソ回帰には L_1 ノルムを使用する。

残差平方和 + 制約 = $\| y - X\beta \|^2 + \lambda \| \beta \|^2 = (y - X\beta)^T (y - X\beta) + \lambda \beta^T \beta$ （但し、$\lambda \geqq 0$）・・・ （式56）

（式56）を最小にする β は、 $\partial((y - X\beta)^T (y - X\beta) + \lambda \beta^T \beta) / \partial \beta = 0$ を満足する。

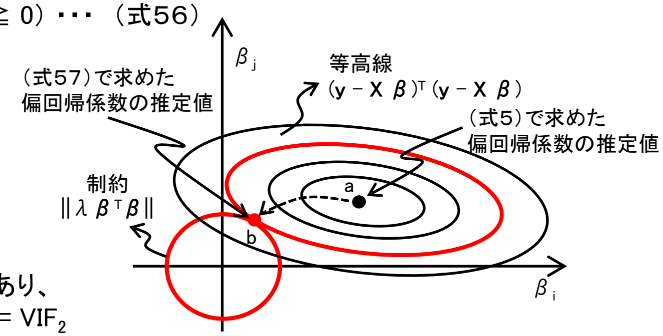
この β を求める手順は（式5）を求める手順と同じである

$$\partial((y - X\beta)^T (y - X\beta) + \lambda \beta^T \beta) / \partial \beta = \partial((y^T - \beta^T X^T)(y - X\beta) + \lambda \beta^T \beta) / \partial \beta$$

$$= \partial(y^T y - y^T X \beta - \beta^T X^T y + \beta^T X^T X \beta + \lambda \beta^T \beta) / \partial \beta$$

$$= - y^T X - X^T y + 2 X^T X \beta + 2 \lambda \beta = - 2 X^T y + 2 (X^T X \beta + \lambda \mathrm{I}) \beta = 0$$

この式を満足する β が $\hat{\beta}$ である。 よって、$\hat{\beta} = (X^T X \beta + \lambda \mathrm{I})^{-1} X^T y$ ・・・ （式57）



制約を加えた上で偏回帰係数を求めることで、
多重共線性で、等高線の底にある不安定な偏回帰係数の値（a）を
等高線の底から離れた安定な値（b）へ移動させることができる。
なお、この図では、β_j に比べて β_i の値を「大幅に」小さくして実現している。

（図6）リッジ回帰の正則化

（2）リッジ回帰を使用した偏回帰係数の算出（例）

（例表11）の説明変数 x_1 と x_2 の間の相関係数は、（A）の場合も（B）の場合もいずれも、ほぼ 1.0 であり、両者の間には従属関係に近い強い相関関係がある。また、分散拡大要因を確認すると、（A）の $VIF_1 = VIF_2 = 1379 > 10$ （B）の $VIF_1 = VIF_2 = 2183 > 10$ であり、多重共線性が発生するリスクが極めて高い。

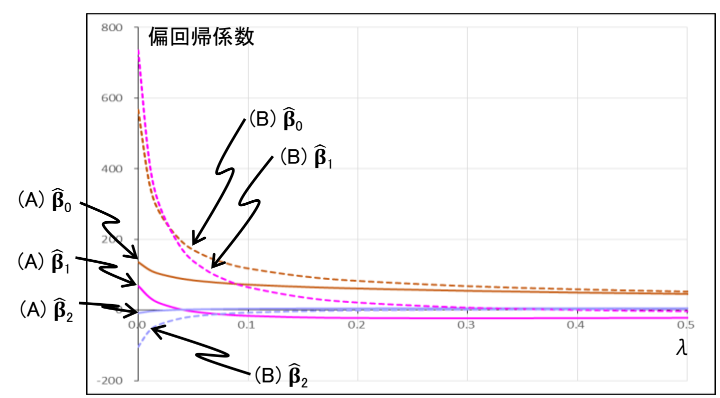
なお、重回帰分析の結果は、（A）、（B）とも重回帰式の F 値、偏回帰係数の p 値はいずれも 0.005 より大きく有意差はないという結果であった。 偏回帰係数の推定値は、以下の通り差が大きく不安定であった。

（A） $\hat{\beta}_0 = 135.4$， $\hat{\beta}_1 = 67.9$， $\hat{\beta}_2 = - 8.2$ （B） $\hat{\beta}_0 = 567.0$， $\hat{\beta}_1 = 736.2$， $\hat{\beta}_2 = - 104.8$

このケースにリッジ回帰を使用した結果が（例図3）である（横軸はパラメータ λ）。λ の値を大きくしていくと、（A）の偏回帰係数と（B）の偏回帰係数の値が近づき両者の差が小さくなっていくことが分かる。 $\lambda > 0.3$ 位で「差」は安定する。 リッジ回帰で偏回帰係数の変動を抑えることができる、ということを確認できる。

（例表11）観測データ

1回目のデータ収集（A）			2回目のデータ収集（B）		
y	x_1	x_2	y	x_1	x_2
50.1	4.9	38.2	49.2	4.9	38.2
99.7	9.3	69.8	101.3	9.3	69.7
149.5	2.9	25.2	149.3	2.9	24.7
198.9	7.9	58.6	198.5	7.9	59.0
249.3	11.4	84.2	249.5	11.4	83.0
301.5	6.6	50.4	300.3	6.6	49.9



（例図3）パラメータ λ と偏回帰係数

付録1．ノルム (norm) と内積 (inner product) の関係

ノルムとは、ベクトル空間におけるベクトルの長さの概念である。

（1）ノルムの定義

ベクトル $\mathbf{a} = [a_1 \ \ a_2 \ \cdots \ a_n]^T$ のノルム $\|\mathbf{a}\|$ は、（付式1）で定義される。

$$\|\mathbf{a}\| = (a_1^2 + a_2^2 + \cdots + a_n^2)^{1/2} \qquad \cdots \text{（付式1）}$$

（2）ノルムと内積との関係　（内積に関しては、別途資料「制御工学（現代制御）の基礎知識」の付録1、または、「統計的検定の手順」の付録1を参照）

ベクトル a と a の内積は、ベクトル a のノルムを使って表すことができる。

$$\langle \mathbf{a}, \mathbf{a} \rangle = \mathbf{a}^T \mathbf{a} = \begin{bmatrix} a_1 \\ a_2 \\ \cdot \\ a_n \end{bmatrix}^T \begin{bmatrix} a_1 \\ a_2 \\ \cdot \\ a_n \end{bmatrix} = [a_1 \ \ a_2 \ \cdots \ a_n] \begin{bmatrix} a_1 \\ a_2 \\ \cdot \\ a_n \end{bmatrix} = a_1^2 + a_2^2 + \cdots + a_n^2 = \|\mathbf{a}\|^2 \qquad \cdots \text{（付式2）}$$

従って、ベクトルの和のノルムの二乗は、以下のように展開できる。

$$\|\mathbf{a} + \mathbf{b}\|^2 = \langle \mathbf{a} + \mathbf{b}, \mathbf{a} + \mathbf{b} \rangle = \langle \mathbf{a}, \mathbf{a} \rangle + \langle \mathbf{a}, \mathbf{b} \rangle + \langle \mathbf{b}, \mathbf{a} \rangle + \langle \mathbf{b}, \mathbf{b} \rangle = \|\mathbf{a}\|^2 + 2\langle \mathbf{a}, \mathbf{b} \rangle + \|\mathbf{b}\|^2 = \|\mathbf{a}\|^2 + 2\,\mathbf{a}^T \mathbf{b} + \|\mathbf{b}\|^2 \qquad \cdots \text{（付式3）}$$

（3）L_1 ノルム　と　L_2 ノルム

$(|a_1|^p + |a_2|^p + \cdots + |a_n|^p)^{1/2}$　において、$p = 1$ のとき L_1 ノルム、$p = 2$ のとき L_2 ノルムという。

付録2．行列のトレース（trace）

正方行列に対してトレースは定義される。

（1）トレースの定義

行列 **A** が n × n（n 行 n 列）の正方行列で ij 成分が a_{ij} であるすると、行列 **A** のトレース tr (**A**) は対角要素の和として（付式4）で定義される。

$$\mathrm{tr}\,(\mathbf{A}) = \sum_{i=1}^{n} a_{ii} = a_{11} + a_{22} + \cdots + a_{nn} \qquad \cdots \text{（付式4）}$$

（2）トレースの基本的な性質

行列 **A** , **B** が n × n（n 行 n 列）の正方行列であるとき、

$$\mathrm{tr}\,(\mathbf{A}+\mathbf{B}) = \mathrm{tr}\,(\mathbf{A}) + \mathrm{tr}\,(\mathbf{B}) \qquad \cdots \text{①}$$

$$\mathrm{tr}\,(k\,\mathbf{A}) = k\,\mathrm{tr}\,(\mathbf{A}) \quad \text{ただし、} k \text{ は定数} \qquad \cdots \text{②}$$

$$\mathrm{tr}\,(\mathbf{A}^{T}) = \mathrm{tr}\,(\mathbf{A}) \qquad \cdots \text{③}$$

（3）トレースの対称性

行列 **A** を m × n の行列、行列 **B** を n × m の行列とすると、行列の積 **A B** は m × m の正方行列、行列の積 **B A** は n × n の正方行列になり、（付式5）が成立する。

$$\mathrm{tr}\,(\mathbf{A\,B}) = \sum_{i=1}^{m} (\mathbf{A\,B})_{ii} = \sum_{i=1}^{m}\Big(\sum_{j=1}^{n} \mathbf{A}_{ij}\,\mathbf{B}_{ji}\Big) = \sum_{j=1}^{n}\Big(\sum_{i=1}^{m} \mathbf{B}_{ji}\,\mathbf{A}_{ij}\Big) = \sum_{j=1}^{n} (\mathbf{B\,A})_{jj} = \mathrm{tr}\,(\mathbf{B\,A}) \qquad \cdots \text{（付式5）}$$

（4）2次形式をトレースで表す（2次形式に関しては、別途資料「統計的検定の手順」の付録2を参照）。

$$\text{2次形式}\quad \mathbf{x}^{T}\,\mathbf{A}\,\mathbf{x} = [x_1 \ \ x_2 \ \cdots \ x_n]\begin{bmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ a_{21} & a_{22} & \cdots & a_{2n} \\ & \cdots & & \\ a_{n1} & a_{n2} & \cdots & a_{nn} \end{bmatrix}\begin{bmatrix} x_1 \\ x_2 \\ \cdot \\ x_n \end{bmatrix} = [x_1 \ \ x_2 \ \cdots \ x_n]\begin{bmatrix} a_{11}x_1 + a_{12}x_2 + \cdots + a_{1n}x_n \\ a_{21}x_1 + a_{22}x_2 + \cdots + a_{2n}x_n \\ \cdots \\ a_{n1}x_1 + a_{n2}x_2 + \cdots + a_{nn}x_n \end{bmatrix}$$

$$= (a_{11}x_1 + a_{12}x_2 + \cdots + a_{1n}x_n)\,x_1 + (a_{21}x_1 + a_{22}x_2 + \cdots + a_{2n}x_n)\,x_2 + \cdots + (a_{n1}x_1 + a_{n2}x_2 + \cdots + a_{nn}x_n)\,x_n$$

$$\text{一方、2次形式}\quad \mathbf{A}\,\mathbf{x}\,\mathbf{x}^{T} = \begin{bmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ a_{21} & a_{22} & \cdots & a_{2n} \\ & \cdots & & \\ a_{n1} & a_{n2} & \cdots & a_{nn} \end{bmatrix}\begin{bmatrix} x_1 \\ x_2 \\ \cdot \\ x_n \end{bmatrix}[x_1 \ \ x_2 \ \cdots \ x_n] = \begin{bmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ a_{21} & a_{22} & \cdots & a_{2n} \\ & \cdots & & \\ a_{n1} & a_{n2} & \cdots & a_{nn} \end{bmatrix}\begin{bmatrix} x_1x_1 & x_1x_2 & \cdots & x_1x_n \\ x_2x_1 & x_2x_2 & \cdots & x_2x_n \\ & \cdots & & \\ x_nx_1 & x_nx_2 & \cdots & x_nx_n \end{bmatrix}$$

$$= \begin{bmatrix} (a_{11}x_1 + a_{12}x_2 + \cdots + a_{1n}x_n)\,x_1 & \cdots & & \\ \cdots & (a_{21}x_1 + a_{22}x_2 + \cdots + a_{2n}x_n)\,x_2 & & \\ & & \cdots & \\ & & & (a_{n1}x_1 + a_{n2}x_2 + \cdots + a_{nn}x_n)\,x_n \end{bmatrix}$$

従って、

$$\mathbf{x}^{T}\,\mathbf{A}\,\mathbf{x} = \mathrm{tr}\,(\mathbf{A}\,\mathbf{x}\,\mathbf{x}^{T}) \qquad \cdots \text{（付式6）}$$

付録3．内積の偏微分と二次形式の偏微分

（1）内積の偏微分

$$\partial\langle \mathbf{a}, \mathbf{x}\rangle / \partial \mathbf{x} = \partial\,(\mathbf{a}^T \mathbf{x}) / \partial \mathbf{x} = \partial / \partial \mathbf{x} \left([\,a_1\ \ a_2\ \ \cdots\ \ a_n\,] \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix} \right) = \partial / \partial \mathbf{x}\,(a_1 x_1 + a_2 x_2 + \cdots + a_n x_n) = \begin{bmatrix} a_1 \\ a_2 \\ \vdots \\ a_n \end{bmatrix} = \mathbf{a}$$

$$\partial\langle \mathbf{x}, \mathbf{a}\rangle / \partial \mathbf{x} = \partial\,(\mathbf{x}^T \mathbf{a}) / \partial \mathbf{x} = \partial / \partial \mathbf{x} \left([\,x_1\ \ x_2\ \ \cdots\ \ x_n\,] \begin{bmatrix} a_1 \\ a_2 \\ \vdots \\ a_n \end{bmatrix} \right) = \partial / \partial \mathbf{x}\,(a_1 x_1 + a_2 x_2 + \cdots + a_n x_n) = \begin{bmatrix} a_1 \\ a_2 \\ \vdots \\ a_n \end{bmatrix} = \mathbf{a}$$

よって、$\partial\,(\mathbf{a}^T \mathbf{x}) / \partial \mathbf{x} = \partial\,(\mathbf{x}^T \mathbf{a}) / \partial \mathbf{x} = \mathbf{a}$　　・・・　（付式7）

（2）2次形式の偏微分

x を n × 1 のベクトル、**A** を n × n の正方行列としたとき、**x** と **A x** の内積は、（付式8）で表される（別途資料「統計的検定の手順」の付録1を参照）。

$\langle \mathbf{x}, \mathbf{A}\,\mathbf{x}\rangle = \mathbf{x}^T \mathbf{A}\,\mathbf{x}$　　・・・　（付8）

ここでは、（付式8）の偏微分を求める。

$$\partial\,(\mathbf{x}^T \mathbf{A}\,\mathbf{x}) / \partial \mathbf{x} = \partial / \partial \mathbf{x} \left([\,x_1\ \ x_2\ \ \cdots\ \ x_n\,] \begin{bmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ a_{21} & a_{22} & \cdots & a_{2n} \\ & \cdots & & \\ a_{n1} & a_{n2} & \cdots & a_{nn} \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix} \right) = \partial / \partial \mathbf{x} \left([\,x_1\ \ x_2\ \ \cdots\ \ x_n\,] \begin{bmatrix} a_{11} x_1 + a_{12} x_2 + \cdots + a_{1n} x_n \\ a_{21} x_1 + a_{22} x_2 + \cdots + a_{2n} x_n \\ \cdots \\ a_{n1} x_1 + a_{n2} x_2 + \cdots + a_{nn} x_n \end{bmatrix} \right)$$

$$= \partial / \partial \mathbf{x} \{ (a_{11} x_1^2 + a_{12} x_1 x_2 + \cdots + a_{1n} x_1 x_n) + (a_{21} x_2 x_1 + a_{22} x_2^2 + \cdots + a_{2n} x_2 x_n) + \cdots + (a_{n1} x_n x_1 + a_{n2} x_n x_2 + \cdots + a_{nn} x_n^2) \}$$

$$= \begin{bmatrix} (2\,a_{11} x_1 + a_{12} x_2 + \cdots + a_{1n} x_n) + (a_{21} x_2 + a_{22} x_2 + \cdots + a_{n1} x_n) \\ a_{12} x_1 + (a_{21} x_1 + 2\,a_{22} x_2 + a_{23} x_3 + \cdots + a_{2n} x_n) + (a_{32} x_3 + a_{42} x_4 + \cdots + a_{n2} x_n) \\ \cdots \\ (a_{1n} x_1 + a_{2n} x_2 + \cdots + a_{n-1n} x_{n-1}) + (a_{n1} x_1 + a_{n2} x_2 + \cdots + 2\,a_{nn} x_n) \end{bmatrix}$$

$$= \begin{bmatrix} (a_{11} x_1 + a_{12} x_2 + \cdots + a_{1n} x_n) + (a_{11} x_1 + a_{21} x_2 + \cdots + a_{n1} x_n) \\ (a_{21} x_1 + a_{22} x_2 + \cdots + a_{2n} x_n) + (a_{12} x_1 + a_{22} x_2 + \cdots + a_{n2} x_n) \\ \cdots \\ (a_{n1} x_1 + a_{n2} x_2 + \cdots + a_{nn} x_n) + (a_{1n} x_1 + a_{2n} x_2 + \cdots + a_{nn} x_n) \end{bmatrix} = \begin{bmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ a_{21} & a_{22} & \cdots & a_{2n} \\ & \cdots & & \\ a_{n1} & a_{n2} & \cdots & a_{nn} \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix} + \begin{bmatrix} a_{11} & a_{21} & \cdots & a_{n1} \\ a_{12} & a_{22} & \cdots & a_{n2} \\ & \cdots & & \\ a_{1n} & a_{2n} & \cdots & a_{nn} \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix}$$

よって、 $\partial\,(\mathbf{x}^T \mathbf{A}\,\mathbf{x}) / \partial \mathbf{x} = \mathbf{A}\,\mathbf{x} + \mathbf{A}^T \mathbf{x} = (\mathbf{A} + \mathbf{A}^T)\,\mathbf{x}$　　・・・　（付式9）

ここで、行列 **A** が対称行列のとき、（付式8）は2次形式になり（別途資料「統計的検定の手順」の付録2を参照）、
また、行列 **A** が対称行列のとき $\mathbf{A} = \mathbf{A}^T$ であるので、二次形式の微分は、（付式9）の結果を使用して（付式10）のようになる。

$\partial\,(\mathbf{x}^T \mathbf{A}\,\mathbf{x}) / \partial \mathbf{x} = 2\,\mathbf{A}\,\mathbf{x}$　　・・・　（付式10）

付録4．冪等(べきとう)行列(Idempotent Matrix)

（1）冪等行列の定義

$A^2 = A$ を満たす正方行列 A を、冪等行列という。

（例） $A = \begin{bmatrix} 5 & -2 \\ 10 & -4 \end{bmatrix}$ は、冪等行列である。 何故なら、 $A^2 = \begin{bmatrix} 5 & -2 \\ 10 & -4 \end{bmatrix}\begin{bmatrix} 5 & -2 \\ 10 & -4 \end{bmatrix} = \begin{bmatrix} 5 & -2 \\ 10 & -4 \end{bmatrix} = A$

（2）唯一の正則な（逆行列が存在する）冪等行列は、単位行列である。

行列 A が冪等行列であれば、$A^2 = A$ が成り立つ。 両辺に左から A^{-1} を掛ければ、 $A^{-1} A^2 = A^{-1} A \;\rightarrow\; A = I$ （単位行列）

（3）行列 A が冪等行列であるとき、行列 I − A も冪等行列である。

$(I - A)^2 = (I - A)(I - A) = I - A - A + A^2 = I - A - A + A = I - A$

（4）冪等行列 A が対称行列であるとき、行列 A と行列 I − A との内積は 0 である。

行列 A が対称行列であるので、$A = A^T$

$\langle A, I - A\rangle = A^T (I - A) = A^T - A^T A = A - A A = A - A = 0$

（5）冪等行列の固有値は、0 もしくは 1 である。

行列 A の固有値を λ_i 固有ベクトルを x_i とおくと、固有方程式は、 $A x_i = \lambda_i x_i$ となる。

$\lambda_i x_i = A x_i = A^2 x_i = A A x_i = A \lambda_i x_i = \lambda_i A x_i = \lambda_i \lambda_i x_i = \lambda_i^2 x_i$ より、 $\lambda_i x_i = \lambda_i^2 x_i \;\rightarrow\; \lambda_i (1 - \lambda_i) x_i = 0$

従って、$x_i \neq 0$ より、 $\lambda_i = 0$ もしくは 1

（6）冪等行列 A が対称行列のとき、rank (A) = tr (A) が成立する。

行列の階数 rank (A) とは、行列 A の列（行）ベクトルのうち、一次独立なベクトルの最大個数である。

一般に、対称行列 A の固有方程式を行列で表現すれば、$A X = X \Lambda$ ・・・（付式11） となる。

ここで、1）行列 Λ は、対角要素に行列 A の固有値 λ_i を持つ対角行列である（但し、行列 A が冪等行列のときは、対角要素は 0 もしくは 1 である）。

2）行列 X は、固有値 λ_i に対応した互いに独立な固有ベクトル x_i を列ベクトルとして持つ直交行列（定義、$X^T = X^{-1}$）である。

以下に示すように、一般に対称行列 A は直交行列 X を使用して対角化できる（別途資料「生産技術－制御工学（現代制御）の基礎」の付録を参照）。

（付式11）の両辺に左側から X^{-1} を掛ければ、$X^{-1} A X = X^{-1} X \Lambda \;\rightarrow\; X^{-1} A X = \Lambda$ これが対角化である（直交行列を使った対角化）。

一方、固有方程式の両辺に右側からから X^{-1} を掛ければ、 $A X X^{-1} = X \Lambda X^{-1} \;\rightarrow\; A = X \Lambda X^{-1} \;\rightarrow A = X \Lambda X^T$ となる。これが固有値分解である。

rank (Λ) ≦ rank (X) = n であるので、rank ($X \Lambda X^T$) は、rank (Λ) で決まる。つまり、rank (A) = rank ($X \Lambda X^T$) = rank (Λ) となる。

また、行列 A が冪等行列のとき、行列 Λ は対角要素が 0 もしくは 1 の対角行列で、rank (Λ) は対角行列 Λ の対角要素 = 1 の数 tr (Λ) に等しい。

つまり、rank (A) = tr (Λ) ・・・ ① が成り立つ。

一方、（付録2）トレースの対称性と $X^T = X^{-1}$ の関係を使って、tr (A) = tr ($X \Lambda X^T$) = tr ($\Lambda X^T X$) = tr ($\Lambda X^{-1} X$) = = tr (Λ) ・・・ ② が成り立つ。

従って、式 ① と式 ② より、rank (A) = tr (A) ・・・ （付式12） が成り立つ。

（7）冪等行列の対角化

上述のとおり、冪等行列 A が対称行列のとき、行列 A の対角化は、$X^{-1} A X = \Lambda$ である。

Λ ： 行列 A の固有値（0 もしくは 1）を対角要素に持つ対角行列、 固有値 1 の数はrank (A)

X ： 行列 A の固有値に対応した独立した固有値ベクトルを列ベクトルに持つ行列であり、直交行列である

冪等行列の対角化の式は次のように表すことができる。

$$X^{-1} A X = \Lambda = \begin{bmatrix} I_k & 0 \\ 0 & 0 \end{bmatrix}$$ ・・・ （付式13） ここで、I_k は、rank (I_k) = k = rank (A) の単位行列

付録5. シューア補行列(Schur complement)

(1)定義

行列 A(サイズ s × s)、行列 B(サイズ s × t)、行列 C(サイズ t × s)、行列 D(サイズ s × s)のブロック行列からなる行列 M(サイズ (s + t) × (s + t))を考える。

$$M = \begin{bmatrix} A & B \\ C & D \end{bmatrix} \quad \cdots \text{(付式14)}$$

行列 D が正則、つまり、行行列 D^{-1} が存在するとき、下記の S_D で定義される、サイズ s × s の行列をシューア補行列という。

$$S_D = A - B\,D^{-1}\,B \quad \cdots \text{(付式15)}$$

(2)行列 M を(付16)のように三角行列と対角行列の積で表すことができる。

$$M = \begin{bmatrix} A & B \\ C & D \end{bmatrix} = \begin{bmatrix} \mathrm{I} & B\,D^{-1} \\ 0 & \mathrm{I} \end{bmatrix} \begin{bmatrix} A - B\,D^{-1}\,C & 0 \\ 0 & D \end{bmatrix} \begin{bmatrix} \mathrm{I} & 0 \\ D^{-1}\,C & \mathrm{I} \end{bmatrix} \quad \cdots \text{(付式16)}$$

(確認)
$$\begin{bmatrix} \mathrm{I} & B\,D^{-1} \\ 0 & \mathrm{I} \end{bmatrix} \begin{bmatrix} A - B\,D^{-1}\,C & 0 \\ 0 & D \end{bmatrix} \begin{bmatrix} \mathrm{I} & 0 \\ D^{-1}\,C & \mathrm{I} \end{bmatrix} = \begin{bmatrix} A - B\,D^{-1}\,C & B \\ 0 & D \end{bmatrix} \begin{bmatrix} \mathrm{I} & 0 \\ D^{-1}\,C & \mathrm{I} \end{bmatrix} = \begin{bmatrix} A - B\,D^{-1}\,C + B\,D^{-1}\,C & B \\ D\,D^{-1}\,C & D \end{bmatrix} = \begin{bmatrix} A & B \\ C & D \end{bmatrix}$$

(3)行列 M の逆行列 M^{-1} を(付5D)のように表すことができる。

$$M^{-1} = \begin{bmatrix} A & B \\ C & D \end{bmatrix}^{-1} = \begin{bmatrix} \mathrm{I} & 0 \\ -D^{-1}\,C & \mathrm{I} \end{bmatrix} \begin{bmatrix} (A - B\,D^{-1}\,C)^{-1} & 0 \\ 0 & D^{-1} \end{bmatrix} \begin{bmatrix} \mathrm{I} & -B\,D^{-1} \\ 0 & \mathrm{I} \end{bmatrix} = \begin{bmatrix} (A - B\,D^{-1}\,C)^{-1} & -(A - B\,D^{-1}\,C)^{-1}\,B\,D^{-1} \\ -D^{-1}\,C\,(A - B\,D^{-1}\,C)^{-1} & D^{-1} + D^{-1}\,C\,(A - B\,D^{-1}\,C)^{-1}\,B\,D^{-1} \end{bmatrix} \cdots \text{(付式17)}$$

本文では、この逆行列 M^{-1} のブロック行列 $(A - B\,D^{-1}\,C)^{-1}$ を活用する。

(確認)
$$\begin{bmatrix} \mathrm{I} & 0 \\ -D^{-1}\,C & \mathrm{I} \end{bmatrix} \begin{bmatrix} \mathrm{I} & 0 \\ D^{-1}\,C & \mathrm{I} \end{bmatrix} = \begin{bmatrix} \mathrm{I} & 0 \\ 0 & \mathrm{I} \end{bmatrix} \text{ より、} \begin{bmatrix} \mathrm{I} & 0 \\ D^{-1}\,C & \mathrm{I} \end{bmatrix}^{-1} = \begin{bmatrix} \mathrm{I} & 0 \\ -D^{-1}\,C & \mathrm{I} \end{bmatrix}$$

$$\begin{bmatrix} (A - B\,D^{-1}\,C)^{-1} & 0 \\ 0 & D^{-1} \end{bmatrix} \begin{bmatrix} A - B\,D^{-1}\,C & 0 \\ 0 & D \end{bmatrix} = \begin{bmatrix} \mathrm{I} & 0 \\ 0 & \mathrm{I} \end{bmatrix} \text{ より、} \begin{bmatrix} A - B\,D^{-1}\,C & 0 \\ 0 & D \end{bmatrix}^{-1} = \begin{bmatrix} (A - B\,D^{-1}\,C)^{-1} & 0 \\ 0 & D^{-1} \end{bmatrix}$$

$$\begin{bmatrix} \mathrm{I} & -B\,D^{-1} \\ 0 & \mathrm{I} \end{bmatrix} \begin{bmatrix} \mathrm{I} & B\,D^{-1} \\ 0 & \mathrm{I} \end{bmatrix} = \begin{bmatrix} \mathrm{I} & 0 \\ 0 & \mathrm{I} \end{bmatrix} \text{ より、} \begin{bmatrix} \mathrm{I} & B\,D^{-1} \\ 0 & \mathrm{I} \end{bmatrix}^{-1} = \begin{bmatrix} \mathrm{I} & -B\,D^{-1} \\ 0 & \mathrm{I} \end{bmatrix}$$

従って、(付式16)より、

$$M^{-1} = \begin{bmatrix} A & B \\ C & D \end{bmatrix}^{-1} = \left[\begin{bmatrix} \mathrm{I} & B\,D^{-1} \\ 0 & \mathrm{I} \end{bmatrix} \begin{bmatrix} A - B\,D^{-1}\,C & 0 \\ 0 & D \end{bmatrix} \begin{bmatrix} \mathrm{I} & 0 \\ D^{-1}\,C & \mathrm{I} \end{bmatrix} \right]^{-1} = \begin{bmatrix} \mathrm{I} & 0 \\ D^{-1}\,C & \mathrm{I} \end{bmatrix}^{-1} \begin{bmatrix} A - B\,D^{-1}\,C & 0 \\ 0 & D \end{bmatrix}^{-1} \begin{bmatrix} \mathrm{I} & B\,D^{-1} \\ 0 & \mathrm{I} \end{bmatrix}^{-1}$$

$$= \begin{bmatrix} \mathrm{I} & 0 \\ -D^{-1}\,C & \mathrm{I} \end{bmatrix} \begin{bmatrix} (A - B\,D^{-1}\,C)^{-1} & 0 \\ 0 & D^{-1} \end{bmatrix} \begin{bmatrix} \mathrm{I} & -B\,D^{-1} \\ 0 & \mathrm{I} \end{bmatrix} = \begin{bmatrix} (A - B\,D^{-1}\,C)^{-1} & 0 \\ -D^{-1}\,C\,(A - B\,D^{-1}\,C)^{-1} & D^{-1} \end{bmatrix} \begin{bmatrix} \mathrm{I} & -B\,D^{-1} \\ 0 & \mathrm{I} \end{bmatrix}$$

$$= \begin{bmatrix} (A - B\,D^{-1}\,C)^{-1} & -(A - B\,D^{-1}\,C)^{-1}\,B\,D^{-1} \\ -D^{-1}\,C\,(A - B\,D^{-1}\,C)^{-1} & D^{-1} + D^{-1}\,C\,(A - B\,D^{-1}\,C)^{-1}\,B\,D^{-1} \end{bmatrix}$$

付録6. ガウス･マルコフの定理（Gauss−Markov theorem）

線形回帰式のパラメータを推定するとき、推定残差を最小にするように最小二乗法を用いて求めた推定量は、線形な不偏推定量のなかで分散が最小になる、つまり推定のバラツキが最も小さくなる推定量である、ということを保証するのが、ガウス･マルコフ定理である。
この定理は、パラメータの推定において最小二乗法を使う根拠になる定理である。なお、この推定量のことを最良線形不偏推定量（BLUE：Best Linier Unbiased Estimator）という。

1）偏回帰係数 β の（一般の）線形不偏推定量 $\tilde{\beta}$ が備えるべき条件を求める。
重回帰式 $\mathbf{y} = \mathbf{X}\,\boldsymbol{\beta} + \mathbf{u}$ ･･･ （式1） の目的変数の値（観測値） $\mathbf{y}$ から推定する偏回帰係数 $\boldsymbol{\beta}$ の不偏推定量を $\tilde{\beta}$ とすると、
$\tilde{\beta} = \mathbf{C}\,\mathbf{y}$ ･･･ ① で表される。式①に（式1）を代入すると、 $\tilde{\beta} = \mathbf{C}\,\mathbf{y} = \mathbf{C}\,(\mathbf{X}\,\boldsymbol{\beta} + \mathbf{u})$
期待値をとると、$E\,[\tilde{\boldsymbol{\beta}}] = E\,[\mathbf{C}\,(\mathbf{X}\,\boldsymbol{\beta} + \mathbf{u})] = \mathbf{C}\,\mathbf{X}\,\boldsymbol{\beta} + E\,[\mathbf{u}] = \mathbf{C}\,\mathbf{X}\,\boldsymbol{\beta} + \mathbf{0} = \mathbf{C}\,\mathbf{X}\,\boldsymbol{\beta}$
$\tilde{\boldsymbol{\beta}}$ が不偏推定量であるためには、$E\,[\tilde{\boldsymbol{\beta}}] = \boldsymbol{\beta}$ でなければならない。 よって、$E\,[\tilde{\boldsymbol{\beta}}] = \mathbf{C}\,\mathbf{X}\,\boldsymbol{\beta} = \boldsymbol{\beta}$ より、$\mathbf{C}\,\mathbf{X} = \mathbf{I}$ （単位行列） という条件が必要になる。
このとき、$\tilde{\boldsymbol{\beta}} = \mathbf{C}\,(\mathbf{X}\,\boldsymbol{\beta} + \mathbf{u}) = \mathbf{C}\,\mathbf{X}\,\boldsymbol{\beta} + \mathbf{C}\,\mathbf{u} = \mathbf{I}\,\boldsymbol{\beta} + \mathbf{C}\,\mathbf{u} = \boldsymbol{\beta} + \mathbf{C}\,\mathbf{u}$ となり、$E\,[\tilde{\boldsymbol{\beta}}] = E\,[\boldsymbol{\beta} + \mathbf{C}\,\mathbf{u}] = E\,[\boldsymbol{\beta}] + \mathbf{C}\,E\,[\mathbf{u}] = \boldsymbol{\beta} + \mathbf{C}\,\mathbf{0} = \boldsymbol{\beta}$ を確認できる。

2）$\tilde{\boldsymbol{\beta}}$ の分散共分散行列を求める。
一般の線形不偏推定量 $\tilde{\boldsymbol{\beta}}$ と推定値 $\hat{\boldsymbol{\beta}}$ の差は、$\tilde{\boldsymbol{\beta}} - \hat{\boldsymbol{\beta}} = \mathbf{C}\,\mathbf{y} - (\mathbf{X}^T\mathbf{X})^{-1}\mathbf{X}^T\mathbf{y} = (\mathbf{C} - (\mathbf{X}^T\mathbf{X})^{-1}\mathbf{X}^T)\,\mathbf{y}$
ここで、$\mathbf{D} = \mathbf{C} - (\mathbf{X}^T\mathbf{X})^{-1}\mathbf{X}^T$ とおくと、$\mathbf{C} = (\mathbf{X}^T\mathbf{X})^{-1}\mathbf{X}^T + \mathbf{D}$ となり、
$\tilde{\boldsymbol{\beta}}$ が不偏推定量である条件 $\mathbf{C}\,\mathbf{X} = \mathbf{I}$ を使うと、$\mathbf{C}\,\mathbf{X} = ((\mathbf{X}^T\mathbf{X})^{-1}\mathbf{X}^T + \mathbf{D})\,\mathbf{X} = (\mathbf{X}^T\mathbf{X})^{-1}\mathbf{X}^T\mathbf{X} + \mathbf{D}\,\mathbf{X} = \mathbf{I} + \mathbf{D}\,\mathbf{X} = \mathbf{I}$ より、$\mathbf{D}\,\mathbf{X} = \mathbf{0}$，$(\mathbf{D}\,\mathbf{X})^T = \mathbf{X}^T\mathbf{D}^T = \mathbf{0}$ が求まる。
これらの関係を使用して、また式変形では、「いろいろな確率分布」の付録48 を参照して、また、（付式21）を利用して、

$$
\begin{aligned}
V\,[\tilde{\boldsymbol{\beta}}] &= E\,[(\tilde{\boldsymbol{\beta}} - E\,[\tilde{\boldsymbol{\beta}}])\,(\tilde{\boldsymbol{\beta}} - E\,[\tilde{\boldsymbol{\beta}}])^T] = E\,[((\boldsymbol{\beta} + \mathbf{C}\,\mathbf{u}) - \boldsymbol{\beta})\,((\boldsymbol{\beta} + \mathbf{C}\,\mathbf{u}) - \boldsymbol{\beta})^T] = E\,[\mathbf{C}\,\mathbf{u}\,(\mathbf{C}\,\mathbf{u})^T] = E\,[\mathbf{C}\,\mathbf{u}\,(\mathbf{u}^T\mathbf{C}^T)] = \mathbf{C}\,E\,[\mathbf{u}\,\mathbf{u}^T]\,\mathbf{C}^T = \mathbf{C}\,\sigma^2\,\mathbf{I}\,\mathbf{C}^T = \sigma^2\,\mathbf{C}\,\mathbf{C}^T \\
&= \sigma^2\,((\mathbf{X}^T\mathbf{X})^{-1}\mathbf{X}^T + \mathbf{D})\,((\mathbf{X}^T\mathbf{X})^{-1}\mathbf{X}^T + \mathbf{D})^T = \sigma^2\,((\mathbf{X}^T\mathbf{X})^{-1}\mathbf{X}^T + \mathbf{D})\,((((\mathbf{X}^T\mathbf{X})^{-1}\mathbf{X}^T)^T + \mathbf{D}^T) = \sigma^2\,((\mathbf{X}^T\mathbf{X})^{-1}\mathbf{X}^T + \mathbf{D})\,(\mathbf{X}\,((\mathbf{X}^T\mathbf{X})^{-1})^T + \mathbf{D}^T) \\
&= \sigma^2\,((\mathbf{X}^T\mathbf{X})^{-1}\mathbf{X}^T + \mathbf{D})\,(\mathbf{X}\,(\mathbf{X}^T\mathbf{X})^{-1} + \mathbf{D}^T) = \sigma^2\,((\mathbf{X}^T\mathbf{X})^{-1}\mathbf{X}^T\mathbf{X}\,(\mathbf{X}^T\mathbf{X})^{-1} + (\mathbf{X}^T\mathbf{X})^{-1}\mathbf{X}^T\mathbf{D}^T + \mathbf{D}\,\mathbf{X}\,(\mathbf{X}^T\mathbf{X})^{-1} + \mathbf{D}\,\mathbf{D}^T) \\
&= \sigma^2\,((\mathbf{I}\,(\mathbf{X}^T\mathbf{X})^{-1} + (\mathbf{X}^T\mathbf{X})^{-1}\mathbf{0} + \mathbf{0}\,(\mathbf{X}^T\mathbf{X})^{-1} + \mathbf{D}\,\mathbf{D}^T) = \sigma^2\,((\mathbf{X}^T\mathbf{X})^{-1} + \mathbf{D}\,\mathbf{D}^T) = \sigma^2\,(\mathbf{X}^T\mathbf{X})^{-1} + \sigma^2\,\mathbf{D}\,\mathbf{D}^T = V\,[\hat{\beta}] + \sigma^2\,\mathbf{D}\,\mathbf{D}^T
\end{aligned}
$$

3）分散共分散行列の対角成分、つまり分散の項に着目する（別途資料「制御工学（現代制御）の基礎知識」の付録1を参照）。

$$
\text{また、}\mathbf{D}\,\mathbf{D}^T = \begin{bmatrix} d_{10} & d_{20} & \cdots & d_{n0} \\ d_{11} & d_{21} & \cdots & d_{n1} \\ & \cdots & & \\ d_{1p} & d_{2p} & \cdots & d_{np} \end{bmatrix} \begin{bmatrix} d_{10} & d_{11} & \cdots & d_{1p} \\ d_{20} & d_{21} & \cdots & d_{2p} \\ & \cdots & & \\ d_{n0} & d_{n1} & \cdots & d_{np} \end{bmatrix} = \begin{bmatrix} d_{10}{}^2 + d_{20}{}^2 + \cdots + d_{n0}{}^2 & & & \cdots \\ & d_{11}{}^2 + d_{21}{}^2 + \cdots + d_{n1}{}^2 & & \\ & & \cdots & \\ \cdots & & & d_{1p}{}^2 + d_{2p}{}^2 + \cdots + d_{np}{}^2 \end{bmatrix}
$$

$$
= \begin{bmatrix} \sum_{i=1}^{n} d_{i0}{}^2 & & \cdots \\ & \sum_{i=1}^{n} d_{i1}{}^2 & \\ & \cdots & \\ \cdots & & \sum_{i=1}^{n} d_{ip}{}^2 \end{bmatrix} \quad \text{より、}\mathbf{D}\,\mathbf{D}^T\text{ の対角要素} = \sum_{i=1}^{n} d_{ik}{}^2
$$

この結果と、分散共分散行列の対角成分は分散 であることから、

$$
V\,[\tilde{\beta}_k] = V\,[\hat{\beta}_k] + \sigma^2 \sum_{i=1}^{n} d_{ik}{}^2
$$

ここで、$\sum_{i=1}^{n} d_{ik}{}^2 \geqq 0$ であるので、 $V\,[\tilde{\beta}_k] \geqq V\,[\hat{\beta}_k]$ ･･･ （付式18） が成り立つ。
この式は、不偏推定量 $\tilde{\boldsymbol{\beta}}$ のなかで最小二乗法に従って求めた不偏推定量 $\hat{\boldsymbol{\beta}}$ の分散が最も小さいことを示している。
つまり、不偏推定量 $\hat{\boldsymbol{\beta}}$ が最良線形不変推定量（BLUE）であることを示している。これが、ガウス･マルコフの定理である。

付録7．偏回帰係数の推定量の確率分布

X は既知の確定値、β は未知の確定値、そして u は多変量標準正規分布 N $(\mathbf{0}, \mathbf{I})$ に従う確率変数である（多変量正規分布に関しては、別途資料「いろいろな確率分布」を参照）。従って、$\hat{\boldsymbol{\beta}}$ は多変量正規分布に従う確率変数になる。 最小二乗法で求めた偏回帰係数の推定量 $\hat{\boldsymbol{\beta}}$ の期待値と分散を求める。

（1）偏回帰係数の推定量 $\hat{\boldsymbol{\beta}}$ の期待値

（式5） $\hat{\boldsymbol{\beta}} = (\mathbf{X}^T \mathbf{X})^{-1} \mathbf{X}^T \mathbf{y}$ に、（式1） $\mathbf{y} = \mathbf{X}\,\beta + \mathbf{u}$ を代入する。なお、E $[\mathbf{u}] = 0$, V $[\mathbf{u}] = \sigma^2 \mathbf{I}$ である。

$$\begin{aligned}\hat{\boldsymbol{\beta}} &= (\mathbf{X}^T \mathbf{X})^{-1} \mathbf{X}^T \mathbf{y} \\ &= (\mathbf{X}^T \mathbf{X})^{-1} \mathbf{X}^T (\mathbf{X}\,\beta + \mathbf{u}) \\ &= (\mathbf{X}^T \mathbf{X})^{-1} \mathbf{X}^T \mathbf{X}\,\beta + (\mathbf{X}^T \mathbf{X})^{-1} \mathbf{X}^T \mathbf{u} \\ &= \beta + (\mathbf{X}^T \mathbf{X})^{-1} \mathbf{X}^T \mathbf{u} \qquad \cdots \quad \text{（付式19）}\end{aligned}$$

$$\begin{aligned}E[\hat{\boldsymbol{\beta}}] &= E[\beta + (\mathbf{X}^T \mathbf{X})^{-1} \mathbf{X}^T \mathbf{u}] \\ &= E[\beta] + E[(\mathbf{X}^T \mathbf{X})^{-1} \mathbf{X}^T \mathbf{u}] \\ &= \beta + (\mathbf{X}^T \mathbf{X})^{-1} \mathbf{X}^T E[\mathbf{u}]\end{aligned}$$

ここで、E $[\mathbf{u}] = \mathbf{0}$ であるので、

$$E[\hat{\boldsymbol{\beta}}] = \beta \qquad \cdots \quad \text{（付式20）}$$

従って、$\hat{\boldsymbol{\beta}}$ は、β の不偏推定量になる（不偏推定量に関しては、別途資料「統計的推定の手順」を参照）。

（2）偏回帰係数の推定量 $\hat{\boldsymbol{\beta}}$ の分散（分散共分散行列）

（付式19）と（付式20）を代入する（行列の式変形には、別途資料「いろいろな確率分布」の付録と「統計的検定の手順」の付録を参照）。

$$\begin{aligned}V[\hat{\boldsymbol{\beta}}] &= E[(\hat{\boldsymbol{\beta}} - E[\hat{\boldsymbol{\beta}}])(\hat{\boldsymbol{\beta}} - E[\hat{\boldsymbol{\beta}}])^T] \\ &= E[(\beta + (\mathbf{X}^T \mathbf{X})^{-1} \mathbf{X}^T \mathbf{u} - \beta)(\beta + (\mathbf{X}^T \mathbf{X})^{-1} \mathbf{X}^T \mathbf{u} - \beta)^T] \\ &= E[(\mathbf{X}^T \mathbf{X})^{-1} \mathbf{X}^T \mathbf{u})((\mathbf{X}^T \mathbf{X})^{-1} \mathbf{X}^T \mathbf{u})^T] \\ &= E[(\mathbf{X}^T \mathbf{X})^{-1} \mathbf{X}^T \mathbf{u})(\mathbf{u}^T \mathbf{X}\,((\mathbf{X}^T \mathbf{X})^{-1})^T)] \\ &= E[(\mathbf{X}^T \mathbf{X})^{-1} \mathbf{X}^T \mathbf{u})(\mathbf{u}^T \mathbf{X}\,(\mathbf{X}^T \mathbf{X})^{-1})] \\ &= (\mathbf{X}^T \mathbf{X})^{-1} \mathbf{X}^T E[\mathbf{u}\,\mathbf{u}^T]\,\mathbf{X}\,(\mathbf{X}^T \mathbf{X})^{-1} \\ &= (\mathbf{X}^T \mathbf{X})^{-1} \mathbf{X}^T \sigma^2 \mathbf{I}\,\mathbf{X}\,(\mathbf{X}^T \mathbf{X})^{-1} \\ &= \sigma^2 (\mathbf{X}^T \mathbf{X})^{-1} \mathbf{X}^T \mathbf{I}\,\mathbf{X}\,(\mathbf{X}^T \mathbf{X})^{-1} \\ &= \sigma^2 (\mathbf{X}^T \mathbf{X})^{-1} \mathbf{X}^T \mathbf{X}\,(\mathbf{X}^T \mathbf{X})^{-1} \\ &= \sigma^2 (\mathbf{X}^T \mathbf{X})^{-1} \qquad \cdots \quad \text{（付式21）}\end{aligned}$$

（3）偏回帰係数の推定量 $\hat{\boldsymbol{\beta}}$ の確率分布

上記で求めた期待値と分散から、 $\hat{\boldsymbol{\beta}}$ は多変量正規分布 N $(\beta,\ \sigma^2 (\mathbf{X}^T \mathbf{X})^{-1})$ 従うことが分かる。

なお、j 番目の説明変数 x_j の偏回帰係数の推定値 $\hat{\beta}_i$ の期待値は E $[\hat{\beta}_j] = \beta_j$, 分散は V $[\hat{\beta}_j] = \sigma^2 a_{jj}$ （但し、a_{jj} は 行列 $\mathbf{X}^T \mathbf{X}$ の $(j+1, j+1)$ 要素）である。

$\hat{\beta}_j$ は正規分布 N $(\beta_j,\ \sigma^2 a_{jj})$ ・・・ （付式22） に従うことが分かる。

ここで、母数 σ^2 の値は未知なので、$\hat{\beta}_j$ の統計的推定と統計的検定を行うにあたっては、σ^2 を代替するパラメータが必要になる。
それが、σ^2 の不変推定量として定義する S^2 である（付録8）。

付録8. 誤差分散の不偏推定量

■偏回帰係数の最尤推定量と観測誤差の分散の最尤推定量

最尤推定量とは尤度(ゆうど)が最大になるときの母数の推定量である。
尤度は標本データの母集団への当てはまり具合のことである(別途資料「統計的推定の手順」を参照)。

誤差 $\mathbf{u}$ が標準正規分布 $N(\mathbf{0}, \sigma^2 \mathbf{I})$ (σ^2 は未知の確定値)に従う確率変数で、X と β も確定値であることから、$\mathbf{y} = X\beta + \mathbf{u}$ の観測値 y_i は互いに独立で正規分布 $N(x_i \beta, \sigma^2)$ に従い($x_i \beta$ はスカラー)、その確率密度関数は(付式23)で表すことができる(別途資料「いろいろな確率分布」を参照)。

$$f(\mathbf{y}) = \prod_{i=1}^{n} (1 / (2\pi\sigma^2)^{1/2})\, e^{-(y_i - x_i \beta)^2 / (2\sigma^2)} \qquad \cdots \text{(付式23)}$$

この確率密度関数を、β と σ^2 がある値を取るとき実現値 $\mathbf{y}$ が得られる確率を求める式であると捉えて、式は同じであるが次のように表現方法を変える。

$$f(\beta, \sigma^2) = \prod_{i=1}^{n} (1 / (2\pi\sigma^2)^{1/2})\, e^{-(y_i - x_i \beta)^2 / (2\sigma^2)}$$

この式は尤度関数である(別途資料「統計的推定の手順」を参照)。そして、計算の簡便さを配慮して、この確率密度関数の対数をとる(対数尤度関数という)。

$$\ln f(\beta, \sigma^2) = \ln \prod_{i=1}^{n} (1 / (2\pi\sigma^2)^{1/2})\, e^{-(y_i - x_i \beta)^2 / (2\sigma^2)} = -(n/2)\ln(2\pi\sigma^2) - (1/(2\sigma^2)) \sum_{i=1}^{n} (y_i - x_i \beta)^2$$
$$= -(n/2)\ln\sigma^2 - (1/(2\sigma^2))(\mathbf{y} - X\beta)^T(\mathbf{y} - X\beta) - (n/2)\ln(2\pi) \qquad \cdots \text{(付式24)}$$

(1)対数尤度関数(尤度関数)を最大する β, つまり、β の最尤推定量を求める。

$$\partial \ln f(\beta, \sigma^2) / \partial\beta = -(1/(2\sigma^2))\, \partial(\mathbf{y} - X\beta)^T(\mathbf{y} - X\beta) / \partial\beta = 0$$

この式の解法は、偏回帰係数の推定値 $\hat{\boldsymbol{\beta}}$ を最小二乗法で求める方法と同じになり、最尤推定量は最小二乗推定量と同じ $\hat{\boldsymbol{\beta}}$ になる。

$\hat{\boldsymbol{\beta}} = (X^T X)^{-1} X^T \mathbf{y}$ $\cdots$ (付式25) が求まる。

(2)対数尤度関数(尤度関数)を最大にする σ^2, つまり、σ^2 の最尤推定量を求める。

$$\partial \ln f(\beta, \sigma^2) / \partial\sigma^2 = -n/(2\sigma^2) + (1/(2\sigma^4)))\,)(\mathbf{y} - X\beta)^T(\mathbf{y} - X\beta) = 0 \quad \rightarrow \quad -n\sigma^2 + (\mathbf{y} - X\beta)^T(\mathbf{y} - X\beta) = 0$$

よって、σ^2 の最尤推定量 $\hat{\sigma}^2 = (1/n)(\mathbf{y} - X\hat{\boldsymbol{\beta}})^T(\mathbf{y} - X\hat{\boldsymbol{\beta}}) = (1/n)\, \boldsymbol{\varepsilon}^T\boldsymbol{\varepsilon}$ $\cdots$ (付式26) が求まる。

■観測誤差の分散(未知の確率変数)の不偏推定量(代替パラメータ)

母数の推定量の期待値が母数と等しいとき、これを満たす推定量を不偏推定量という(別途資料「統計的推定の手順」を参照)。
そこで、$\hat{\sigma}^2$ が不偏推定量であるか否かを確認するために、$\hat{\sigma}^2$ の期待値を調べる(行列 H は冪等行列になる、付録4)。

$H = X(X^T X)^{-1} X^T$ とおくと、$(\mathbf{I} - H)X = X - HX = X - (X(X^T X)^{-1} X^T)X = X - X(X^T X)^{-1}(X^T X) = X - X = \mathbf{0}$ $\cdots$ (付式27) であるので、

$\boldsymbol{\varepsilon} = \mathbf{y} - X\hat{\boldsymbol{\beta}} = \mathbf{y} - X(X^T X)^{-1} X^T \mathbf{y} = (\mathbf{I} - X(X^T X)^{-1} X^T)\mathbf{y} = (\mathbf{I} - H)\mathbf{y} = (\mathbf{I} - H)(X\beta + \mathbf{u}) = (\mathbf{I} - H)X\beta + (\mathbf{I} - H)\mathbf{u} = \mathbf{0}\,\beta + (\mathbf{I} - H)\mathbf{u} = (\mathbf{I} - H)\mathbf{u}$

ここで、$X^T X$ は $(X^T X)^T = X^T (X^T)^T = X^T X$ なので対称行列であり、また、$((X^T X)^{-1})^T = (X^T X)^{-1}$ であるので(別途資料「いろいろな確率分布」の付録48を参照)、

$H^T = (X(X^T X)^{-1} X^T)^T = (X^T)^T ((X^T X)^{-1})^T X^T = X(X^T X)^{-1} X^T = H$

$H^T H = (X(X^T X)^{-1} X^T)^T (X(X^T X)^{-1} X^T) = ((X^T)^T ((X^T X)^{-1})^T X^T)(X(X^T X)^{-1} X^T) = X(X^T X)^{-1} (X^T X)(X^T X)^{-1} X^T = X(X^T X)^{-1}\, \mathbf{I}\, X^T = X(X^T X)^{-1} X^T = H$

$\boldsymbol{\varepsilon}^T\boldsymbol{\varepsilon} = (\mathbf{y} - X\hat{\boldsymbol{\beta}})^T(\mathbf{y} - X\hat{\boldsymbol{\beta}}) = ((\mathbf{I} - H)\mathbf{u})^T(\mathbf{I} - H)\mathbf{u} = \mathbf{u}^T(\mathbf{I} - H)^T(\mathbf{I} - H)\mathbf{u} = \mathbf{u}^T(\mathbf{I} - H^T)(\mathbf{I} - H)\mathbf{u} = \mathbf{u}^T(\mathbf{I} - H - H^T + H^T H)\mathbf{u} = \mathbf{u}^T(\mathbf{I} - H - H + H)\mathbf{u} = \mathbf{u}^T(\mathbf{I} - H)\mathbf{u}$ $\cdots$ (付式28)

また、トレースの対称性(付録2)より、さらに、$X^T X$ はサイズ p + 1 の正方行列なので、

$\mathrm{tr}[H] = \mathrm{tr}[X(X^T X)^{-1} X^T] = \mathrm{tr}[(X^T X)^{-1} X^T X] = \mathrm{tr}[\mathbf{I}_{p+1}]$ ($\mathbf{I}_{p+1}$ は、サイズ p + 1 の単位行列)

従って、$\hat{\sigma}^2$ の期待値 $E[\hat{\sigma}^2]$ は、(付式28)を使って、また、式変形では(付式6)を使って、

$E[\hat{\sigma}^2] = E[(1/n)\, \boldsymbol{\varepsilon}^T\boldsymbol{\varepsilon}] = (1/n)E[\boldsymbol{\varepsilon}^T\boldsymbol{\varepsilon}] = (1/n)E[\mathbf{u}^T(\mathbf{I} - H)\mathbf{u}] = (1/n)E[\mathrm{tr}((\mathbf{I} - H)\mathbf{u}\mathbf{u}^T)] = (1/n)\,\mathrm{tr}((\mathbf{I} - H)E[\mathbf{u}\mathbf{u}^T]) = (1/n)\,\mathrm{tr}((\mathbf{I} - H)\sigma^2\mathbf{I})$
$= (\sigma^2/n)\,\mathrm{tr}(\mathbf{I} - H) = (\sigma^2/n)(\mathrm{tr}(\mathbf{I}) - \mathrm{tr}(H)) = (\sigma^2/n)(\mathrm{tr}(\mathbf{I}_n) - \mathrm{tr}(\mathbf{I}_{p+1})) = (\sigma^2/n)(n - (p+1)) = ((n - p - 1)/n)\,\sigma^2$ $\cdots$ (付式29)

ここで、$S^2 = (1/(n - p - 1))\, \boldsymbol{\varepsilon}^T\boldsymbol{\varepsilon}$ $\cdots$ (付式30) と置くと、(付式26)と(付式29)を使って、

$E[S^2] = E[(1/(n - p - 1))\boldsymbol{\varepsilon}^T\boldsymbol{\varepsilon}] = (1/(n - p - 1))E[\boldsymbol{\varepsilon}^T\boldsymbol{\varepsilon}] = (n/(n - p - 1))E[\hat{\sigma}^2] = (n/(n - p - 1))((n - p - 1)/n)\,\sigma^2 = \sigma^2$

このように、S^2 の期待値が σ^2 になることから、回帰係数の区間推定や検定では S^2 を分散 σ^2 の不偏推定量として定義し、未知パラメータである分散 σ^2 の代替パラメータとして使用されている。

付録9．標準偏差と標準誤差

標準偏差（SD : Standard Deviation）は、記述統計学で使われる統計量であり。標本データ自体のバラツキを表す。
一方、標準誤差（SE : Standard Error）は、推測統計学で使われる統計量であり、推定量のバラツキ、推定の精度を表す。
（付表１）に標準偏差と標準誤差の比較を示す（不偏分散に関しては、別途資料「統計的推測の手順」を参照）。

（付表１）標準偏差と標準誤差の比較

	標準偏差（SD）	標準誤差（SE）
用途	記述統計学 標本データを要約するための統計量の一つで、標本データのバラツキを表す。	推測統計学 標本データを使って母集団の統計量である母数を推測するための統計量の一つで、推測した統計量の分散のバラツキを表す。 つまり、推測した統計量の精度を表す
説明	分散　$\sigma^2 = \sum_{i=1}^{n}(X_i - \overline{X})^2 / n$ X_i ：　標本データ $\overline{X} = \sum_{i=1}^{n} X_i / n$　標本データの平均 標準偏差（SD） = （分散）$^{1/2}$ = $(\sigma^2)^{1/2} = \sigma$	（1）標本平均の標準誤差 ・・・ 別途資料「統計的推定の手順」を参照 標本平均の分散 = σ^2 / n ここで、σ^2 を不偏推定量 $U^2 = \sum_{i=1}^{n}(X_i - \overline{X})^2 / (n-1)$ に置き換える 標本平均の標準誤差 = $(U^2 / n)^{1/2}$ $= \{\sum_{i=1}^{n}(X_i - \overline{X})^2 / (n(n-1))\}^{1/2}$ （2）重回帰式における観測誤差の分散（目的変数の推定値の分散）の標準誤差 ・・・ （式9）と（付式30）を参照 観測誤差の分散 σ^2 の不偏推定量 $S^2 = \sum_{i=1}^{n}(y_i - \hat{y}_i)^2 / (n-p-1)$ 観測誤差の分散の標準誤差 = $(S^2)^{1/2} = \{\sum_{i=1}^{n}(y_i - \hat{y}_i)^2 / (n-p-1)\}^{1/2}$ （3）重回帰式における偏回帰係数（推定値）の分散の標準誤差 ・・・ （式29）を参照 偏回帰係数の分散 $V[\beta_i] = (\sigma^2 a_{ii})^{1/2}$ ここで、σ^2 を不変推定量 $S^2 = \sum_{i=1}^{n}(y_i - \hat{y}_i)^2 / (n-p-1)$ に置き換える 偏回帰係数 β_i の分散の標準誤差 = $(S^2 a_{ii})^{1/2}$ $= \{\sum_{i=1}^{n}(y_i - \hat{y}_i)^2 / (a_{ii}(n-p-1))\}^{1/2}$

付録10. 射影行列(Projection Matrix)

(付式30)で示すような、ベクトル $\mathbf{a}$ を別のベクトル $\mathbf{b}$ へ変換する行列 $\mathbf{P}$ を考える。

$\mathbf{b} = \mathbf{P}\,\mathbf{a}$ ・・・ (付式31)

(付式31)で、行列 $\mathbf{P}$ が冪等行列(付録4)であるとき、変換行列 $\mathbf{P}$ を射影行列という。さらに 行列 $\mathbf{P}$ が対称行列のとき、行列 $\mathbf{P}$ は直交射影行列であるという。

1)射影行列

(付図1)に示すように、行列 $\mathbf{P}$ によって点 A が点 B に移るとき、直線 AB 上の全ての点が点 B へ移るのであれば、あたかも点 A の射影が点 B であると見なせる(行列 $\mathbf{P}$ が射影行列と呼ばれる所以である)。行列 $\mathbf{P}$ が冪等行列であるとき、「点 A が点 B へ移るとき、直線 AB の全ての点が点 B へ移る」ことを確認する。

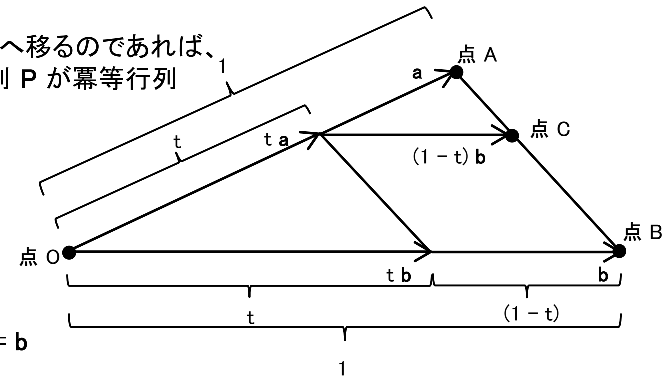
(確認)

冪等行列 $\mathbf{P}$ を使用して、ベクトル $\mathbf{a}$ をベクトル $\mathbf{b}$ に移すことができるということを式に表せば、$\mathbf{b} = \mathbf{P}\,\mathbf{a}$ である。一方、直線 AB 上の全ての点を点 B へ移すことができるとは、直線 AB 上の任意の点 C を点 B へ移すことができるということであり、t を任意の実数(比率)としたとき、(付図1)に示すように点 C は $t\,\mathbf{a} + (1-t)\,\mathbf{b}$ で表されるので、

$\mathbf{P}\,(t\,\mathbf{a} + (1-t)\,\mathbf{b}) = \mathbf{b}$ ・・・ (付式32) が成り立つことを示せばいい。

$\mathbf{P}\,(t\,\mathbf{a} + (1-t)\,\mathbf{b}) = t\,\mathbf{P}\,\mathbf{a} + (1-t)\,\mathbf{P}\,\mathbf{b} = t\,\mathbf{b} + (1-t)\,\mathbf{P}^2\,\mathbf{a} = t\,\mathbf{b} + (1-t)\,\mathbf{P}\,\mathbf{a} = t\,\mathbf{b} + (1-t)\,\mathbf{b} = \mathbf{b}$

よって、(付式31)が成り立つ。



(付図1)射影行列

2)直交射影行列

3次元での直交射影行列の機能を(付図2)に示す。

ベクトル $\mathbf{v}$ を 平面 $\mathbf{a}_1 - \mathbf{a}_2$ に対して行列 $\mathbf{P}$ を使用して垂直に射影したのが、ベクトル $\mathbf{w}$ である。 $\mathbf{w} = \mathbf{P}\,\mathbf{v}$ ・・・ (付式33)

また、ベクトル $\mathbf{w}$ は $\mathbf{A}$ が張る空間(ここでは、平面 $\mathbf{a}_1 - \mathbf{a}_2$)のベクトルであり、実数 s と t を使用して表される。

$$\mathbf{w} = s\,\mathbf{a}_1 + t\,\mathbf{a}_2 = \mathbf{A}\,\mathbf{x},\ \ \text{但し、}\mathbf{A} = [\mathbf{a}_1\ \ \mathbf{a}_2],\ \ \mathbf{x} = \begin{bmatrix} s \\ t \end{bmatrix} \quad \cdots\cdots \ (\text{付式34})$$

ベクトル $\mathbf{v} - \mathbf{w} = \mathbf{v} - \mathbf{A}\,\mathbf{x}$ は平面 $\mathbf{a}_1 - \mathbf{a}_2$ と直交するので、ベクトル $\mathbf{v} - \mathbf{A}\,\mathbf{x}$ と、それぞれのベクトル $\mathbf{a}_1, \mathbf{a}_2$ との内積は 0 になる(別途資料「統計的検定の手順」の付録1を参照)。

$\langle \mathbf{a}_1, \mathbf{v} - \mathbf{A}\,\mathbf{x} \rangle = \mathbf{a}_1^T\,(\mathbf{v} - \mathbf{A}\,\mathbf{x}) = 0$, $\langle \mathbf{a}_2, \mathbf{v} - \mathbf{A}\,\mathbf{x} \rangle = \mathbf{a}_2^T\,(\mathbf{v} - \mathbf{A}\,\mathbf{x}) = 0$ まとめて、$\mathbf{A}^T\,(\mathbf{v} - \mathbf{A}\,\mathbf{x}) = 0$

$\mathbf{A}^T\,\mathbf{v} = \mathbf{A}^T\,\mathbf{A}\,\mathbf{x}$ (この式を正規方程式と言う) ・・・ (付式35)

ベクトル $\mathbf{a}_1$ と $\mathbf{a}_2$ が独立であるとすれば、$\mathbf{A}^T\,\mathbf{A}$ の逆行列は存在するので、(付式35)より、$\mathbf{x} = (\mathbf{A}^T\,\mathbf{A})^{-1}\,\mathbf{A}^T\,\mathbf{v}$

一方、$\mathbf{A}\,\mathbf{x} = \mathbf{w} = \mathbf{P}\,\mathbf{v}$ であるので、$\mathbf{A}\,(\mathbf{A}^T\,\mathbf{A})^{-1}\,\mathbf{A}^T\,\mathbf{v} = \mathbf{P}\,\mathbf{v}$ よって、$\mathbf{P} = \mathbf{A}\,(\mathbf{A}^T\,\mathbf{A})^{-1}\,\mathbf{A}^T$ ・・・ (付式36)

(付式36)が直交射影行列である。n 次元に拡張したときの直交射影行列も(付式36)で表される。

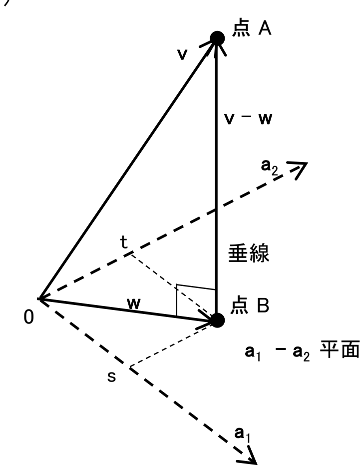
ここで、(付式36)で表される行列 $\mathbf{P}$ が直交射影行列であることを確認しておく。つまり、行列 $\mathbf{P}$ が冪等性を持つ行列であり、かつ対称行列であることを確認しておく。

①冪等性を持つ行列($\mathbf{P}^2 = \mathbf{P}$)であることの確認

$\mathbf{P}^2 = (\mathbf{A}\,(\mathbf{A}^T\,\mathbf{A})^{-1}\,\mathbf{A}^T)\,(\mathbf{A}\,(\mathbf{A}^T\,\mathbf{A})^{-1}\,\mathbf{A}^T) = \mathbf{A}\,(\mathbf{A}^T\,\mathbf{A})^{-1}\,(\mathbf{A}^T\,\mathbf{A})\,(\mathbf{A}^T\,\mathbf{A})^{-1}\,\mathbf{A}^T = \mathbf{A}\,\mathbf{I}\,(\mathbf{A}^T\,\mathbf{A})^{-1}\,\mathbf{A}^T = \mathbf{A}\,(\mathbf{A}^T\,\mathbf{A})^{-1}\,\mathbf{A}^T = \mathbf{P}$

②対称行列($\mathbf{P}^T = \mathbf{P}$)であることの確認(式変形は、別途資料「いろいろな確率分布」の付録48を参照)

$\mathbf{P}^T = (\mathbf{A}\,(\mathbf{A}^T\,\mathbf{A})^{-1}\,\mathbf{A}^T)^T = ((\mathbf{A}^T\,\mathbf{A})^{-1}\,\mathbf{A}^T))^T\,\mathbf{A}^T = (\mathbf{A}^T)^T\,((\mathbf{A}^T\,\mathbf{A})^{-1})^T\,\mathbf{A}^T = \mathbf{A}\,(\mathbf{A}^T\,\mathbf{A})^{-1}\,\mathbf{A}^T = \mathbf{P}$



(付図2)直交射影行列

また、行列 $\mathbf{P}$ が直交射影行列であるとき、以下の通り、行列 $\mathbf{I} - \mathbf{P}$ も直交射影行列になる。

①冪等性を持つ行列であることの確認 $(\mathbf{I} - \mathbf{P})^2 = (\mathbf{I} - \mathbf{P})\,(\mathbf{I} - \mathbf{P}) = \mathbf{I} - \mathbf{P} - \mathbf{P} - \mathbf{P}^2 = \mathbf{I} - \mathbf{P} - \mathbf{P} + \mathbf{P} = \mathbf{I} - \mathbf{P}$

②対称行列であることの確認 $(\mathbf{I} - \mathbf{P})^T = \mathbf{I} - \mathbf{P}^T = \mathbf{I} - \mathbf{P}$

付録11．直交射影行列(Orthogonal projection matrix)の例

点 A をベクトル $\mathbf{a}_1$ 上の点 B に直交射影する場合を考える（付図3）。

付録10の（付式36）を使用して、直交射影行列を求める。

$$\mathbf{P} = \mathbf{A}\,(\mathbf{A}^T \mathbf{A})^{-1}\,\mathbf{A}^T \qquad \cdots \quad (\text{付式}36)$$

ここで、点Aの射影先である点Bの空間を表す行列（ここでは、ベクトル）は、$\mathbf{A} = [\ \mathbf{a}_1\ ]$ である。
ベクトル $\mathbf{a}_1$ は点 O と点 G を通るベクトルであるので、点 O と点 G を結ぶベクトルを代わりに使用すれば、
行列（ここでは、ベクトル） $\mathbf{A}$ は、点 G の座標を使用して下記のように表すことができる。

$$\mathbf{A} = \begin{bmatrix} \cos^2\theta \\ \cos\theta\ \sin\theta \end{bmatrix}$$

よって、（付式36）は、

$$\mathbf{P} = \mathbf{A}\,(\mathbf{A}^T \mathbf{A})^{-1}\,\mathbf{A}^T = \begin{bmatrix} \cos^2\theta \\ \cos\theta\ \sin\theta \end{bmatrix} \left(\begin{bmatrix} \cos^2\theta \\ \cos\theta\ \sin\theta \end{bmatrix}^T \begin{bmatrix} \cos^2\theta \\ \cos\theta\ \sin\theta \end{bmatrix} \right)^{-1} \begin{bmatrix} \cos^2\theta \\ \cos\theta\ \sin\theta \end{bmatrix}^T$$

$$= \begin{bmatrix} \cos^2\theta \\ \cos\theta\ \sin\theta \end{bmatrix} \left([\cos^2\theta \quad \cos\theta\ \sin\theta] \begin{bmatrix} \cos^2\theta \\ \cos\theta\ \sin\theta \end{bmatrix} \right)^{-1} [\cos^2\theta \quad \cos\theta\ \sin\theta]$$

$$= \begin{bmatrix} \cos^2\theta \\ \cos\theta\ \sin\theta \end{bmatrix} [\cos^4\theta + \cos^2\theta\ \sin^2\theta]^{-1}\ [\cos^2\theta \quad \cos\theta\ \sin\theta]$$

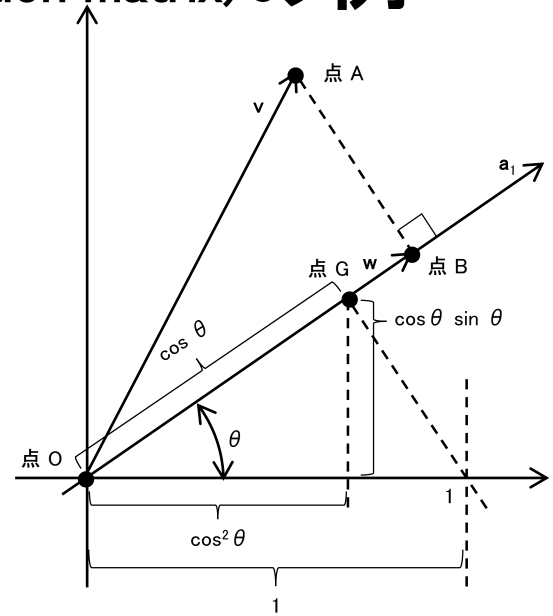
$$= (\cos^2\theta\,(\cos^2\theta + \sin^2\theta))^{-1} \begin{bmatrix} \cos^2\theta \\ \cos\theta\ \sin\theta \end{bmatrix} [\cos^2\theta \quad \cos\theta\ \sin\theta] = (1/\cos^2\theta) \begin{bmatrix} \cos^4\theta & \cos^3\theta\ \sin\theta \\ \cos^3\theta\ \sin\theta & \cos^2\theta\,\sin^2\theta \end{bmatrix} = \begin{bmatrix} \cos^2\theta & \cos\theta\ \sin\theta \\ \cos\theta\ \sin\theta & \sin^2\theta \end{bmatrix}$$

この結果を使用して、例えば、$\theta = \pi / 4$ のときには、

$$\mathbf{P} = \begin{bmatrix} \cos^2(\pi/4) & \cos(\pi/4)\sin(\pi/4) \\ \cos(\pi/4)\sin(\pi/4) & \sin^2(\pi/4) \end{bmatrix} = \begin{bmatrix} (1/\sqrt{2})^2 & (1/\sqrt{2})^2 \\ (1/\sqrt{2})^2 & (1/\sqrt{2})^2 \end{bmatrix} = \begin{bmatrix} 1/2 & 1/2 \\ 1/2 & 1/2 \end{bmatrix}$$

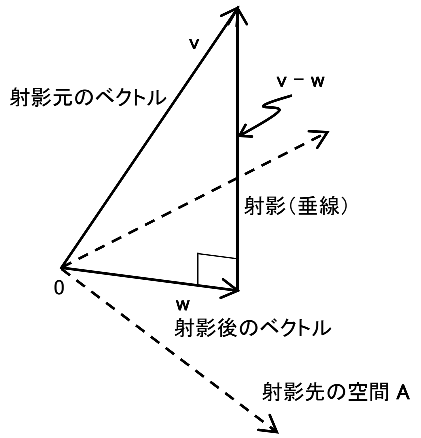
よって、点 A の座標が (1, 2) のときのベクトル $\mathbf{a}_1$ への直交射影である点 B の座標は、 $\mathbf{w} = \mathbf{P}\,\mathbf{v} = \begin{bmatrix} 1/2 & 1/2 \\ 1/2 & 1/2 \end{bmatrix} \begin{bmatrix} 1 \\ 2 \end{bmatrix} = \begin{bmatrix} 3/2 \\ 3/2 \end{bmatrix}$ より、

点 B の座標は、(3/2, 3/2) であることが分かる。



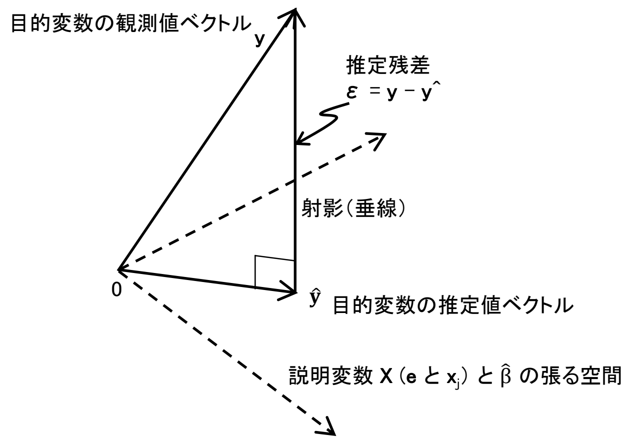
（付図3）直交射影行列の例

付録12. 射影行列を使った偏回帰係数の導出



射影行列 $P = A\,(A^T A)^{-1} A^T$

（付図4）直交射影行列



射影行列 $H = X\,(X^T X)^{-1} X^T$

（付図5）重回帰式の構造

重回帰式の偏回帰係数の推定値 $\hat{\boldsymbol{\beta}}$ は、直交射影行列（付録10）の考え方を使って求めることもできる。

（付図4）は（付録10）の直交射影の（付図2）を書き換えたものである。（付図5）は（付図4）を参考にして重回帰式の構造に合わせて作成したものである。
すなわち、目的変数の観測値ベクトル y を説明変数の空間に直交射影したのが目的変数の推定値ベクトル $\hat{y}$ に相当する。
直交射影は説明変数の空間に垂線を下す射影であるので、このとき推定残差（垂線）ベクトル $\boldsymbol{\varepsilon}$ の大きさは最小になる。つまり、このときの $\hat{\boldsymbol{\beta}}$ が最小二乗法を使って求めた推定値になる。

（付図4）に倣って作成した（付図5）に示す直交射影行列 H を使用すると、（付式37）が推定値ベクトル $\hat{\mathbf{y}}$ を表す式になる。 この式は、（式2）である。

$\hat{\mathbf{y}} = \mathbf{H}\,\mathbf{y} = \mathbf{X}\,(\mathbf{X}^T \mathbf{X})^{-1} \mathbf{X}^T \mathbf{y} = \mathbf{X}\,\hat{\boldsymbol{\beta}}$ ・・・ （付式37）

この（付式37）から、X （つまり、X の列ベクトル x_i）と目的変数 y の観測データから偏回帰係数の推定値 $\hat{\boldsymbol{\beta}}$ が求まる。 （式38）は、（式5）である。

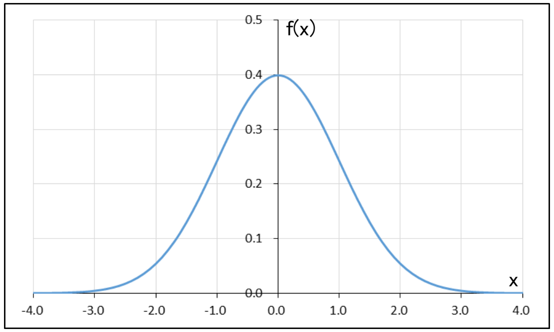
$\hat{\boldsymbol{\beta}} = (\mathbf{X}^T\mathbf{X})^{-1} \mathbf{X}^T \mathbf{y}$ ・・・ （式38）

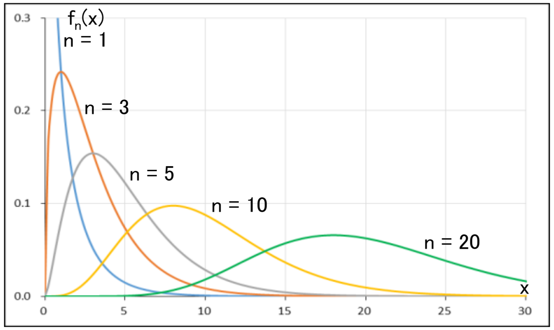
付録13. カイ二乗分布の要約　（別途資料「いろいろな確率分布」参照）

■概要

互いに独立な確率変数 $X_1, X_2, \cdots, X_n$ が標準正規分布（$\mu = 0, \sigma^2 = 1$）に従うとき、
確率変数 $X_1^2 + X_2^2 + \cdots X_n^2$ が従う確率分布を、自由度 n （n は和の数に相当）のカイ二乗分布という。

確率変数 X_i が、標準正規分布に従う　ならば、→　$\sum_{i=1}^{n} X_i^2$ は、自由度 n のカイ二乗分布に従う





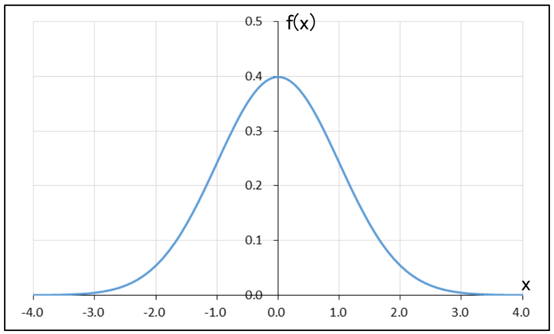
（付図7）カイ二乗分布の導出

付録14. t 分布の要約　（別途資料「いろいろな確率分布」参照）

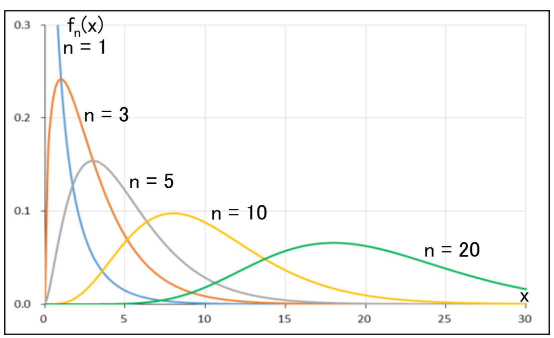
■概要

t 分布は、標準正規分布を自由度を加味したカイ二乗分布でスケーリングした確率分布である。

確率変数 X が、標準正規分布に従う

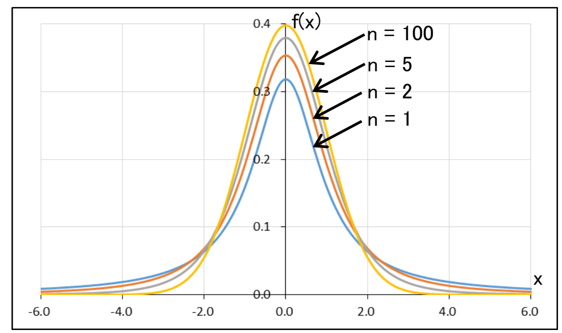


確率変数 Y が、自由度 n のカイ二乗分布に従う



ならば、

X / (Y / n)$^{1/2}$ は、自由度 n の t 分布に従う



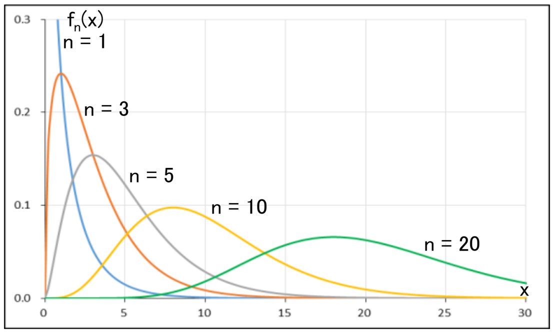
（付図8）t 分布の導出

付録15. F 分布の要約 （別途資料「いろいろな確率分布」参照）

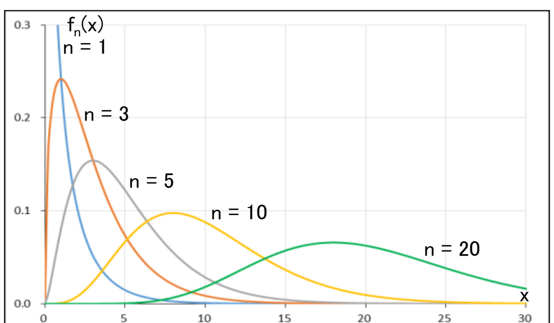
■概要

自由度 m のカイ二乗分布に従う確率変数と自由度 n のカイ二乗分布に従う確率変数が互いに独立であるとき、
自由度を加味した確率変数の比が従う確率分布を、パラメータとして自由度 (m, n) の F 分布という。

確率変数 X が、自由度 m のカイ二乗分布に従う

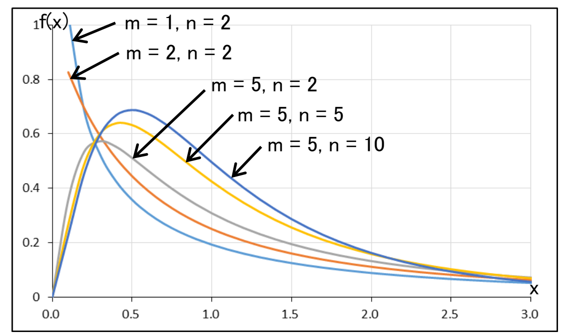


確率変数 Y が、自由度 n のカイ二乗分布に従う



ならば、

(X / m) / (Y / n) は、自由度 (m, n) の F 分布に従う



（付図9）F 分布の導出

参考にした資料

石井俊全：「意味がわかる統計学」、 ベレ出版
涌井良幸、涌井貞美：「中学数学でわかる統計の授業」、 日本実業出版
涌井良幸、涌井貞美：「統計学の図鑑」、 技術評論社
阿部真人：「データ分析に必須の知識・考え方　統計学入門」、 ソシム
西内啓：「統計学が最強の学問である【実践編】」、 ダイヤモンド社
高橋敬子：「線形代数から始める多変量解析」、 プレアデス出版
数理アラカルト：「行列のトレースの性質」、 行列のトレース - 公式/性質-　(証明付) - 理数アラカルト - (risalc.info)
数学の景色、「直交行列の定義と性質１０個とその証明」、直交行列の定義と性質10個とその証明 | 数学の景色 (mathlandscape.com)
高校数学の美しい物語：「二次形式の微分」、二次形式の意味，微分，標準形など | 高校数学の美しい物語 (manabitimes.jp)
Hatena Blog：「シューア補行列」、シューア補行列の定義とその背景，逆行列補題との関係をまとめる - エンジニアを目指す浪人のブログ (hatenablog.com)
村澤友康（甲南大学）：「多重正規分布の線型変換」、 us-ln12.pdf (fc2.com)
高校数学の美しい物語：「射影行列のイメージと楽しい公式」、射影行列のイメージと楽しい公式 | 高校数学の美しい物語 (manabitimes.jp)
Python 数値計算：「正射影ベクトルと射影行列」、正射影ベクトルと射影行列 (atelierkobato.com)
統計学講義用：「行列表記の統計学入門」、 統計学講義用 – YouTube　 行列表記の統計学入門 第5回 ガウスマルコフ定理 - YouTube
高校数学の美しい物語：「正規方程式の導出と計算例」、正規方程式の導出と計算例 | 高校数学の美しい物語 (manabitimes.jp)
データサイエンスLab.：「自由度調整済み決定係数」、説明変数を増やすと必ずR2が大きくなる。変数選択したいなら自由度調整済決定係数の出番です！ - YouTube
ウサギさんの統計学サロン：「重回帰分析・線型回帰・回帰係数の最小二乗推定量と検定」、【統計学】重回帰分析・線形回帰　回帰係数の計算と検定 (multivariate-statistics.com)
有意に無意味な話：「教師あり学習法、線型回帰の性質」、 教師あり学習手法 | 有意に無意味な話 (starpentagon.net)
古谷知之（慶応大学）：「統計基礎」、 統計基礎_19_10 (keio.ac.jp)
麻生良文（慶応大学）：「回帰分析（重回帰）」、回帰分析（重回帰） (keio.ac.jp)
いちばんやさしい、医療統計：「標準偏差と標準誤差の違い」、 標準偏差と標準誤差の違いをわかりやすく！計算式やエラーバーでの使い分けは？｜いちばんやさしい、医療統計 (best-biostatistics.com)
いちばんやさしい、医療統計：「ダミー変数」、 ダミー変数とは？変換する方法や解釈の仕方を具体例で解説！｜いちばんやさしい、医療統計 (best-biostatistics.com)
福井正康（福山平成大学）：「重回帰分析（標準偏回帰係数）」、Microsoft Word - mstattext02.docx (heisei-u.ac.jp)
Qiita：「決定係数からVIF（分散拡大要因）を算出」、 VIF(Variance inflation factor; 分散拡大係数)は重回帰モデルの偏回帰係数の分散に現れるという話 – Qiita
Stat704 Data Analysis I (University of South Carolina)：「Multicollinearity and Variance Inflation Factors」、 vif_704[233].pdf
大森裕浩（東京大学）：「VIFの解釈」、 chapter4_vif.pdf (u-tokyo.ac.jp)
唐沢好男（電気通信大学）：「赤池情報量基準（AIC）を学ぶ」、TR_YK_048_AIC.pdf (uec.ac.jp)
予備校のノリで学ぶ「大学の数学・物理」：「L1/L2正則化の意味【機械学習】」、L1/L2正則化の意味【機械学習】 - YouTube