VI. 多変量解析

VI-1. 最適化

VI-1-1. 重回帰分析

VI-1-1-1. 疑似逆行列を使った重回帰分析

説明変数が一つしかない時には、次の式のように、目的変数を一つの係数、説明変数、定数、誤差で説明することが出来ます。

$$y = ax + b + e$$

y: 目的変数

x: 説明変数

a: 係数

b: 定数

e: 誤差・説明できない残差

このような形で最適な係数、定数を求めることを単回帰と言います。説明変数が増えると次のような形で目的変数を説明することになります。これを重回帰（線形重回帰）と言います。

$$y = a_0 + a_1x_1 + a_2x_2 + \cdots + a_nx_n + e$$

y: 目的変数

x_i: 説明変数

a_i: 係数

a_0: 定数

e: 誤差・説明出来ない残差

この最適化にはさまざまな考え方、モデル、方法があります。代表的なものが最小二乗法と最尤法です。データが正規分布している場合には、最小法と最尤法の結果は一致します。ここでは、最小二乗法による重回帰分析の疑似逆行列を使った計算する方法を紹介します。

つぎのようなデータセットがあったとします。

$$\mathrm{y}_1, x_{11}, x_{21} \cdots x_{p1}$$

$$\mathrm{y}_2, x_{12}, x_{22} \cdots x_{p2}$$

$$\mathrm{y}_3, x_{11}, x_{23} \cdots x_{p3}$$

$$\vdots$$

$$\mathrm{y}_m, x_{1m}, x_{2m} \cdots x_{pm}$$

式に書くと次のようになります。

$$\mathrm{y}_1 = a_0 + a_1x_{11} + a_2x_{21} + \cdots + a_px_{p1} + e_1$$

$$\mathrm{y}_2 = a_0 + a_1x_{12} + a_2x_{22} + \cdots + a_px_{p2} + e_2$$

$$\mathrm{y}_3 = a_0 + a_1x_{13} + a_2x_{23} + \cdots + a_px_{p3} + e_3$$

$$\vdots$$

$$\mathrm{y}_n = a_0 + a_1x_{1m} + a_2x_{2m} + \cdots + a_px_{pm} + e_m$$

一般的化すると次の式になります。

$$\mathrm{y} = a_0 + a_1 x_1 + a_2 x_2 + \cdots + a_p x_p + e$$

$$a_0\text{: 定数}$$

$$= \begin{pmatrix} 1 & x_1 & \cdots & x_p \end{pmatrix} \begin{pmatrix} a_0 \\ a_1 \\ \vdots \\ a_p \end{pmatrix} + e$$

これを行列で表現します。

$$\boldsymbol{Y} = \begin{pmatrix} y_1 \\ \vdots \\ y_n \end{pmatrix}, \quad \boldsymbol{X_{+1}} = \begin{pmatrix} 1 & x_{11} & \cdots & x_{1p} \\ \vdots & \vdots & \ddots & \vdots \\ 1 & x_{n1} & \cdots & x_{np} \end{pmatrix}, \quad \boldsymbol{A_{+1}} = \begin{pmatrix} a_0 \\ a_1 \\ \vdots \\ a_p \end{pmatrix}, \quad \boldsymbol{E} = \begin{pmatrix} e_1 \\ \vdots \\ e_n \end{pmatrix}$$

$$\boldsymbol{Y} = \boldsymbol{X_{+1}} \boldsymbol{A_{+1}} + \boldsymbol{E}$$

私たちは V-1-4 の最適化と疑似逆行列のところで、疑似逆行列$\left(\boldsymbol{X_{+1}}^{\#}\right)$が次のように $\boldsymbol{A_{+1}}$の最適解を与えることを学びました。

$$\boldsymbol{A_{+1}} = \boldsymbol{X_{+1}}^{\#} \boldsymbol{Y}$$

式 73

直接に逆行列を求める計算もありますし、特異値分解の逆演算として求めることもできます。

$$\boldsymbol{X_{+1}}^{\#} = \left(\boldsymbol{X_{+1}}^{T} \boldsymbol{X_{+1}}\right)^{-\boldsymbol{1}} \boldsymbol{X_{+1}}^{T}$$

$$\boldsymbol{U^T X^T X U} = \boldsymbol{\Sigma} = \begin{pmatrix} \lambda_1 & 0 & \cdots & 0 \\ 0 & \lambda_2 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \lambda_p \end{pmatrix}$$

$$\lambda_1 \geq \lambda_2 \geq \cdots \geq \lambda_p > 0$$

$$\boldsymbol{X_{+1}} \boldsymbol{X_{+1}}^{T} = \begin{pmatrix} 1 & x_{11} & \cdots & x_{1p} \\ \vdots & \vdots & \ddots & \vdots \\ 1 & x_{n1} & \cdots & x_{np} \end{pmatrix} \begin{pmatrix} 1 & \cdots & 1 \\ x_{11} & \cdots & x_{n1} \\ \vdots & \ddots & \vdots \\ x_{1p} & \cdots & x_{np} \end{pmatrix}$$

$$\boldsymbol{V^T} \boldsymbol{X_{+1}} \boldsymbol{X_{+1}}^{T} \boldsymbol{V} = \boldsymbol{\Sigma_V}$$

$$\boldsymbol{\Sigma}^{\#} = \begin{pmatrix} \frac{1}{\gamma_1} & 0 & \cdots & 0 & 0 & \cdots & 0 \\ 0 & \frac{1}{\gamma_2} & \cdots & 0 & 0 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots & \vdots & \ddots & 0 \\ 0 & 0 & \cdots & \frac{1}{\gamma_r} & 0 & \cdots & 0 \\ 0 & 0 & \cdots & 0 & 0 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & 0 & 0 & \cdots & 0 \end{pmatrix}_{n \times p}$$

$$\gamma_i = \lambda_i \geq threshold\ value$$

$$\underline{\boldsymbol{\Sigma}^{\#}}$$

$$X^{\#} = V\Sigma^{\#}U^{T}$$

実際にやってみると、直接求めるのは大変かもしれませんので、特異値分解の逆演算のほうが楽かもしれません。ただし、実際にはコンピュータで計算するので、どちらでも良いようなものです

VI-1-1-2. 重回帰分析の空間幾何学的な意味

ここでは、重回帰分析とはどんなことをしているのかを空間幾何学的に考えます。初めに、多次元空間中の超平面と点の距離について考えます。何故かというと、最小二乗法による回帰分析とは、各データの空間中の点と超平面との距離の総和を最小化することだからです。図 63 に、原点を含む超平面と素の式を示しました。原点の座標は0 : (0 ⋯ 0)です。

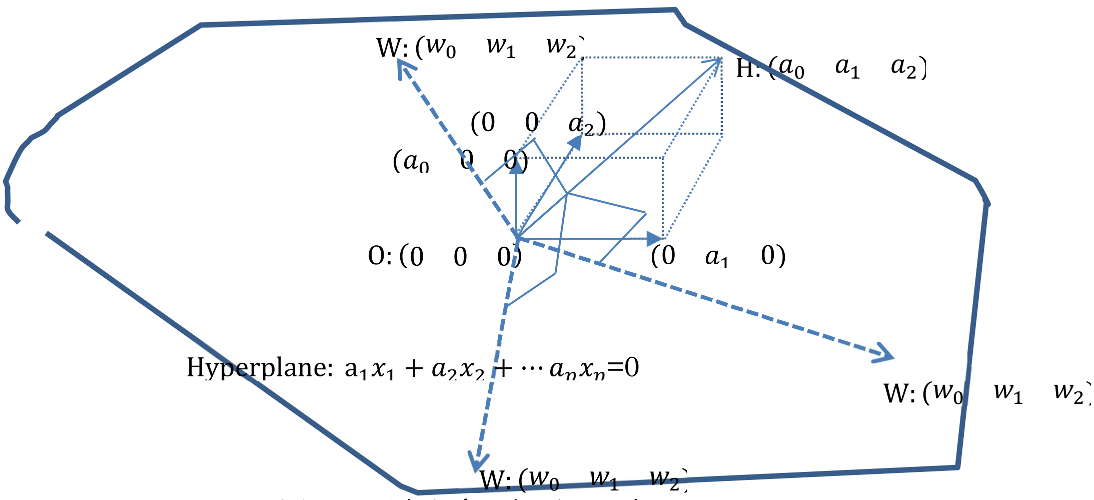

図 63. 原点を含む超平面の式

超平面の傾きは法線ベクトル（平面に直交するベクトル）で表せます。ベクトル$\overrightarrow{\mathrm{OH}}$は超平面の直交する法線です。

$$\overrightarrow{\mathrm{OH}} = (a_0 \quad a_1 \quad \cdots \quad a_p)$$

として、$\overrightarrow{\mathrm{OW}}$を超平面上の任意のベクトルとします。

$$\overrightarrow{\mathrm{OW}} = (w_0 \quad w_1 \quad \cdots \quad w_0)$$

$$\overrightarrow{\mathrm{OH}} \perp \overrightarrow{\mathrm{OW}}$$

直交の条件は内積が 0 になることですから、

$$(w_0 \quad w_1 \quad \cdots \quad w_p)\begin{pmatrix} a_0 \\ a_1 \\ \vdots \\ a_p \end{pmatrix} = 0$$

$$\mathrm{a}_0w_0 + a_1w_1 + \cdots a_pw_p = 0$$

w_iをx_iに置き換えたものが図に示した行平面の式です。簡便化するために、法線ベクトル$\boldsymbol{U}$を単位ベクトルにします。

$$a_0{}^2+a_1{}^2+\cdots+a_p{}^2=1$$

ここで、この原点を含む超平面に平行する平面を一般的な超平面だと考えます。どんな平面でも、平面を平行移動して原点を含む平面を作ることが出来ることは自明でしょう。$\mathrm{X}(x_0 \quad \cdots \quad x_p)$をこの平行した平面上の点だと考えます。二つの平面の距離は一定でこれをtとします。原点を含む平面上の任意の点Wからの法線と並行した平面の交点をXとすると、ベクトル$\overrightarrow{WX}$の長さ$|WX|$がtということです。

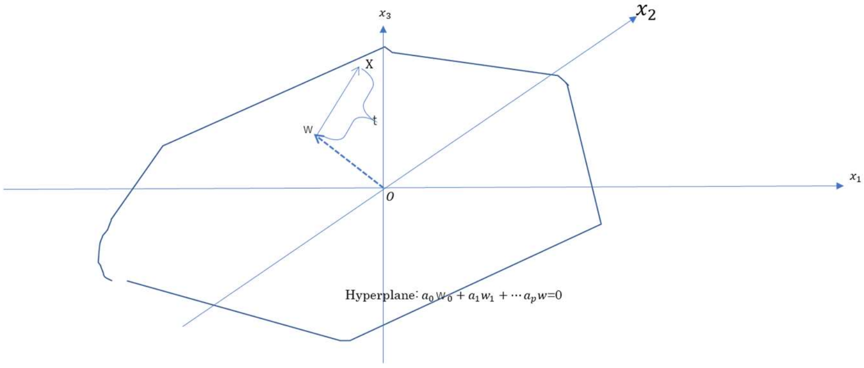

図64. 平面上の点

$\overrightarrow{\mathrm{WX}}$は法線の単位ベクトル$(a_0 \quad \cdots \quad a_p)$と並行してますから、

$$\overrightarrow{\mathrm{WX}}=\overrightarrow{\mathrm{OX}}-\overrightarrow{\mathrm{OW}}=t(a_0 \quad \cdots \quad a_p)$$

$$(x_0-w_0 \quad \cdots \quad x_p-w_p)=t(a_0 \quad \cdots \quad a_p)$$

です。この両辺のベクトルと法線ベクトルは直交しているのでその内積は0です。

$$(x_0-w_0 \quad \cdots \quad x_p-w_p)\begin{pmatrix}a_0\\ \vdots\\ a_p\end{pmatrix}=t(a_0 \quad \cdots \quad a_p)\begin{pmatrix}a_0\\ \vdots\\ a_p\end{pmatrix}$$

$$\mathrm{a}_0x_0+a_1x_1+\cdots a_px_p-(\mathrm{a}_0w_0+a_1w_1+\cdots a_pw_p)=t(a_0{}^2+a_1{}^2+\cdots+a_p{}^2)$$

$$\mathrm{a}_0x_0+a_1x_1+\cdots a_px_p=t$$

$$\because\ \mathrm{a}_0w_0+a_1w_1+\cdots a_pw_p=0$$

$$a_0{}^2+a_1{}^2+\cdots+a_p{}^2=1$$

以上より、平行する超平面の式は以下のように表せます。

$$\mathrm{a}_0x_0+a_1x_1+\cdots a_px_p=t$$

これは $a_0{}^2+a_1{}^2+\cdots+a_p{}^2=1$であれば、原点から超平面の距離はtということです。

$a_0{}^2+a_1{}^2+\cdots+a_p{}^2\neq 1$の場合を含めて、次のように超平面を表します。

$$a_0'x_0+a_1'x_1+\cdots+a_p'x_p=c$$

c: 定数

全ての変数がこの平面にあればすべての点を以下の式で表せます。

$$\boldsymbol{XA'}=\mathrm{c}\begin{pmatrix}1\\ \vdots\\ 1\end{pmatrix}_{n\times 1}$$

そこで

$$a_i = \frac{a_i{}'}{c}$$

とすると

$$\boldsymbol{X} = \begin{pmatrix} x_{10} & \cdots & x_{1p} \\ \vdots & \ddots & \vdots \\ x_{n0} & \cdots & x_{np} \end{pmatrix}_{n \times p}, \quad \boldsymbol{A} = \begin{pmatrix} a_0 \\ \vdots \\ a_p \end{pmatrix}$$

$$a_0 x_0 + a_1 x_1 + \cdots + a_p x_p = 1$$

$$\boldsymbol{XA} = \begin{pmatrix} 1 \\ \vdots \\ 1 \end{pmatrix}_{n \times 1}$$

と書けるはずですが、$\boldsymbol{X}$は正則ではないし、データのすべてが一つの超平面上に乗っていることはないでしょう。

$$\boldsymbol{XA} \neq \begin{pmatrix} 1 \\ \vdots \\ 1 \end{pmatrix}$$

そこで誤差項・誤差行列を考えます。

$$\boldsymbol{XA} - \begin{pmatrix} 1 \\ \vdots \\ 1 \end{pmatrix} = \boldsymbol{E}$$

$$\boldsymbol{E} = \begin{pmatrix} e_1 \\ \vdots \\ e_n \end{pmatrix}$$

e_j: 誤差・残差

この誤差項を最小化する最適な$\boldsymbol{A}$を疑似逆行列$X^{\#}$をつかって次のように求めることが出来ます。

$$\boldsymbol{XA} - \begin{pmatrix} 1 \\ \vdots \\ 1 \end{pmatrix} = \boldsymbol{E}$$

$$\boldsymbol{XA} = \begin{pmatrix} 1 \\ \vdots \\ 1 \end{pmatrix} + \boldsymbol{E}$$

$$\boldsymbol{X}^{\#}\boldsymbol{XA} = \boldsymbol{X}^{\#} \begin{pmatrix} 1 \\ \vdots \\ 1 \end{pmatrix}$$

$$\boldsymbol{A} = \boldsymbol{X}^{\#} \begin{pmatrix} 1 \\ \vdots \\ 1 \end{pmatrix}$$

これが、疑似逆行列を使った重回帰分析の基本的考え方ですが、この解は一般的な重回帰ではありません。重回帰分析とは、目的関数と他の説明変数の関係の表現です。上の結果は、目的変数と説明変数の区別なく、すべての変数間関係を最適に表現しています。目的変数と説明変数を区別しなければならないので、目的変数の係数を 1 にする必要があります。

$$a_0 x_0 + a_1 x_1 + \cdots + a_p x_p = 0$$

そこで、両辺をa_0で割ると、

$$x_0+\frac{a_1}{a_0}x_1+\cdots+\frac{a_p}{a_0}x_p=0$$

$$x_0=-\left(\frac{a_1}{a_0}x_1+\cdots+\frac{a_p}{a_0}x_p\right)$$

つまりx_0を目的変数と考えることが出来ます。そこで$-\frac{a_i}{a_0}$をb_i、x_0を y と書きます。

$$y=b_1x_1+\cdots+b_px_p$$

Y 切片 c をくわえても

$$y=b_1x_1+\cdots+b_px_p+\mathrm{c}$$
$$y-c=b_1x_1+\cdots+b_px_p$$

のように$y-c=\mathrm{y}$とすれば、c を考慮するはありません。

しかし、これはyを説明する式にはなっていません。求めた係数が y を説明するための係数ではなくて、全体の関係を最適に説明するための係数だからです。そのことは、次のように考えるとよくわかります。

$a_0=-1\ x_0=y$とすると、平面の式は次のようになります。

$$y-(a_1x_1+\cdots+a_px_p)=0$$

ここで多次元空間中の点D_j: $\begin{pmatrix}d_{yj} & d_{x1j} & \cdots & d_{xpj}\end{pmatrix}$と超平面の距離は次の式で表せます。

$$d_j=\frac{\left|d_{yj}-(a_1d_{x1j}+\cdots+a_pd_{xpj})\right|}{\sqrt{1^2+{a_1}^2+\cdots+{a_p}^2}}=\frac{|e_j|}{\sqrt{1^2+{a_1}^2+\cdots+{a_p}^2}}$$

私たちが最適化しようとしているの $\sum_{j=1}^{n}{d_j}^2$ であって $\sum_{j=1}^{n}{e_j}^2$ではありません。. $a_1=a_2=\cdots=a_p=0$ でなければ

$$\sqrt{1^2+{a_1}^2+\cdots+{a_p}^2}\neq 1$$

なので

$$d_j\neq e_j$$

y 軸方向の距離は、点と平面の距離とは明らかに違います。

私たちが考えなければならないのは図 65 に示した。 y 軸方向の距離の総和の最小化です。

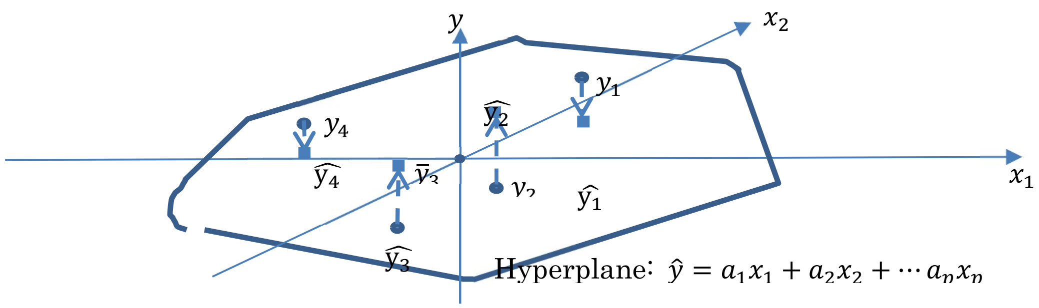

図. 65. データの測点と超平面の y 軸方向の距離

単純化のために、例によってｙ切片を０とします。生のデータではなくて、それぞれの項目の平均値を生データから差し引くということです。もし説明変数の行列Xが正則ならば

$$\boldsymbol{Y} = \boldsymbol{X}\boldsymbol{A}$$

となって次のように関係を表せます。

$$y = a_1 x_1 + a_2 x_2 + \cdots a_p x_p$$

しかし、データは誤差を含んでいて正則ではありません。誤差項を加えて、次のように関係を表します。

$$y = a_1 x_1 + a_2 x_2 + \cdots a_p x_p + e$$

$$\boldsymbol{Y} = \boldsymbol{X}\boldsymbol{A} + \boldsymbol{E}$$

$$\boldsymbol{E} = \begin{pmatrix} e_1 \\ \vdots \\ e_n \end{pmatrix}$$

$$e_i\text{: 誤差・残差}$$

私たちは $\boldsymbol{E}^{\boldsymbol{T}}\boldsymbol{E}$.を最小化しますが、疑似逆行列を使って次のように計算できることはすでに V-3-4 で学んだことです。

$$\boldsymbol{Y} = \boldsymbol{X}\boldsymbol{A}$$

$$\boldsymbol{X}\boldsymbol{A} = \begin{pmatrix} y_1 \\ \vdots \\ y_n \end{pmatrix}$$

$$\boldsymbol{X}^{\#}\boldsymbol{X}\boldsymbol{A} = \boldsymbol{X}^{\#}\begin{pmatrix} y_1 \\ \vdots \\ y_n \end{pmatrix}$$

$$\boldsymbol{A} = \boldsymbol{X}^{\#}\begin{pmatrix} y_1 \\ \vdots \\ y_n \end{pmatrix}$$

切片

上の説明で y 切片を考えないようにするために、あらかじめデータを平均値からの差に作り変えて、後からそれらを加えて切片を作るというやり方です。具体的に手順を示すと、次の通りです。

$$y = d_y - \overline{d_y}$$

$$x_{ij} = d_{ij} - \overline{d_J}$$

$$d\text{: 元データ}$$

$$\overline{d_y}, \overline{d_J}\text{: 平均値}$$

$$\overline{d_J} = \frac{1}{n}\sum_{i=1}^{n} x_{ij}$$

$$\overline{d_y} = \frac{1}{n}\sum_{i=1}^{n} y_i$$

$$y = (x_1 \quad \cdots \quad x_p)\begin{pmatrix} a_1 \\ \vdots \\ a_p \end{pmatrix} = a_1 x_1 + a_2 x_2 + \cdots a_p x_p$$

$$y = a_1 x_1 + a_2 x_2 + \cdots a_p x_p$$

$$d_y - \overline{d_y} = a_1(d_1 - \overline{d_1}) + a_2(d_2 - \overline{d_2}) + \cdots + a_p(d_p - \overline{d_p})$$

と計算して、最後に以下の式で切片を求めます。

$$d_y = a_1 d_1 + a_2 d_2 + \cdots + a_p d_p + \overline{d_y} - (a_1\overline{d_1} + a_2\overline{d_2} + \cdots + a_p\overline{d_p})$$
$$\overline{d_y} - (a_1\overline{d_1} + a_2\overline{d_2} + \cdots + a_p\overline{d_p}) = a_0$$

このやり方の方が面倒だと思えば、変動しない説明変数の列を行列に加えて、直接、疑似逆行列を作って解をもとめるというやり方もあります。

$$\boldsymbol{Y} = \begin{pmatrix} y_1 \\ \vdots \\ y_n \end{pmatrix}, \quad \boldsymbol{X_{+1}} = \begin{pmatrix} 1 & x_{11} & \cdots & x_{1p} \\ \vdots & \vdots & \ddots & \vdots \\ 1 & x_{n1} & \cdots & x_{np} \end{pmatrix}, \quad \boldsymbol{A_{+1}} = \begin{pmatrix} a_0 \\ a_1 \\ \vdots \\ a_p \end{pmatrix}, \quad \boldsymbol{E} = \begin{pmatrix} e_1 \\ \vdots \\ e_n \end{pmatrix}$$

$$\boldsymbol{Y} = \boldsymbol{X_{+1}} \boldsymbol{A_{+1}} + \boldsymbol{E}$$
$$\boldsymbol{A_{+1}} = \boldsymbol{X_{+1}}^{\#} \boldsymbol{Y}$$
$$a_0\text{: constant term}$$

これが最初に示した計算方法です。ずいぶん回りくどい説明をしましたが、何を言いたかったのかというと、回帰というのは、超平面の傾きを最適化しているのだということと、超平面の傾きを表しているのは法線だということです。これは、かなり重要な感覚で、この感覚があると物事を単純化して理解できます。

VI-1-1-3. 微分と連立方程式による説明

疑似逆行列を使うことによって、重回帰分析の手順を単純に公式化できました。しかし、線形代数学を知らない読者にとっては、疑似逆行列が一種のブラックボックスでよくわからないかもしれません。そこで、蛇足ですが、微分方程式を解くという形で、重回帰分析を説明しておきます。

次のデータセットがあったとします。

	説明変数	目的変数
標本番号	0　1　2 ⋯ i ⋯ p	y
1	$a_0 \quad d_{11} \; d_{12} \cdots d_{1i} \cdots d_{1p}$	d_{1y}
2	$a_0 \quad d_{21} \; d_{22} \cdots d_{2i} \cdots d_{2p}$	d_2y
⋮	⋮	⋮
k	$a_0 \quad d_{k1} \; d_{k2} \cdots d_{ki} \cdots dx_{kp}$	d_{ky}
⋮	⋮	⋮
n	$a_0 \quad d_{n1} \; d_{n2} \cdots d_{ni} \cdots d_{np}$	d_{ny}

$$d_{ky} = a_0 + a_1 d_{k1} + a_2 d_{k2} + \cdots + a_i d_{ki} + \cdots + a_p d_{kp} + e_k$$
$$e_k = d_{ky} - (a_0 + a_1 d_{k1} + a_2 d_{k2} + \cdots + a_i d_{ki} + \cdots + a_p d_{kp})$$

$$e_k{}^2 = \left(d_{ky} - (\ a_0 +\ a_1 d_{k1} +\ a_2 d_{k2} + \cdots + a_i d_{ki} + \cdots + a_p d_{kp})\right)^2$$

$$E = \sum_{k=1}^{n} e_k{}^2 = \sum_{k=1}^{n} \left(d_{ky} - (\ a_0 +\ a_1 d_{k1} +\ a_2 d_{k2} + \cdots + a_i d_{ki} + \cdots + a_p d_{kp})\right)^2$$

Eは下に凸の二次関数だから最小値を持つことは明らかです。そこで、E を $a_0, \cdots a_p$で微分して極値を 2 求めます。

$$\frac{dE}{da_i} = 0$$

計算を単純化するため次のようにF_kをおきます。

$$F_k = d_{ky} - (\ a_0 +\ a_1 d_{ki} +\ a_2 d_{k2} + \cdots + a_i d_{kp})$$

$$E = \sum_{k=1}^{n} F_k{}^2$$

$$\frac{dE}{da_i} = \frac{dE}{dF_k} \frac{dF_k}{da_i}$$

$$\frac{dE}{dF} = 2\sum_{k=1}^{n} F_k = 2\sum_{k=1}^{n} \left(d_{ky} - (\ \ a_0 + a_1 d_{k1} +\ a_2 d_{k2} + \cdots + a_i d_{ki} + \cdots + a_p d_{kp})\right)$$

$$\frac{dF_k}{da_i} = 2d_{ki}$$

$$\frac{dF_k}{da_0} = 2n$$

$$\frac{dE}{da_i} = 2\sum_{k=1}^{n} \left(d_{ky} - (\ a_0 +\ a_1 d_{k1} +\ a_2 d_{k2} + \cdots + a_i d_{ki} + \cdots + a_p d_{kp})\right) d_{ki} = 0$$

$$d_{oi} = 1$$

$$\frac{dE}{2da_i} = \sum_{k-1}^{n} d_{ky} d_{ki} - \left(a_0 + \sum_{k=1}^{n} d_{ki}\ \ a_1 \sum_{k-1}^{n} d_{k1} d_{ki} +\ x_2 \sum_{k=1}^{n} d_{k2} d_{ki} + \cdots + x_p \sum_{k=1}^{n} d_{kp} d_{ki}\right) = 0$$

$$a_1 \sum_{k-1}^{n} d_{k1} d_{ki} +\ a_2 \sum_{k=1}^{n} d_{k2} d_{ki} + \cdots + a_p \sum_{k=1}^{n} d_{kp} d_{ki} + a_{p+1} \sum_{k=1}^{n} d_{ki} = \sum_{k-1}^{n} d_{ky} d_{ki}$$

これを連立方程式の形に書き換えると

$$na_0 + a_1 \sum_{k-1}^{n} d_{k1} +\ a_2 \sum_{k=1}^{n} d_{k2} + \cdots + a_p \sum_{k=1}^{n} d_{kp} = \sum_{k-1}^{n} d_{ky}$$

$$a_0 \sum_{k=1}^{n} d_{k1} + a_1 + \sum_{k-1}^{n} d_{k1} d_{k1} +\ a_2 \sum_{k=1}^{n} d_{k2} d_{k1} + \cdots + a_p \sum_{k=1}^{n} d_{kp} d_{k1} = \sum_{k-1}^{n} d_{ky} d_{k1}$$

$$+a_0 \sum_{k=1}^{n} d_{k2} + a_1 \sum_{k-1}^{n} d_{k1} d_{k2} +\ a_2 \sum_{k=1}^{n} d_{k2} d_{k2} + \cdots + a_p \sum_{k=1}^{n} d_{kp} d_{k2} = \sum_{k-1}^{n} d_{ky} d_{k2}$$

$$\vdots$$

$$+a_0\sum_{k=1}^{n}d_{kp}+a_1\sum_{k-1}^{n}d_{k1}d_{kp}+\ a_2\sum_{k=1}^{n}d_{k2}d_{kp}+\cdots+a_p\sum_{k=1}^{n}d_{kp}d_{kp}=\sum_{k-1}^{n}d_{ky}d_{kp}$$

この連立方程式を解けば$a_0,\ a_1,\cdots,a_p$の解が得られますが、微分方程式と疑似逆行列の関係を理解するために、この連立方程式を行列の計算に書き換えます。

$$\begin{pmatrix} n & \sum_{k-1}^{n}d_{k1} & \sum_{k=1}^{n}d_{k2} & \cdots & \sum_{k=1}^{n}d_{kp} \\ \sum_{k=1}^{n}d_{k1} & \sum_{k-1}^{n}{d_{k1}}^2 & \sum_{k=1}^{n}d_{k2}d_{k1} & \cdots & \sum_{k=1}^{n}d_{kp}d_{k1} \\ \sum_{k=1}^{n}d_{k2} & \sum_{k=1}^{n}d_{k1}d_{k2} & \sum_{k=1}^{n}{d_{k2}}^2 & \cdots & \sum_{k=1}^{n}d_{kp}d_{k2} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ \sum_{k=1}^{n}d_{kp} & \sum_{k-1}^{n}d_{k1}d_{kp} & \sum_{k=1}^{n}d_{k2}d_{kp} & \cdots & \sum_{k=1}^{n}{d_{kp}}^2 \end{pmatrix}\begin{pmatrix} a_0 \\ a_1 \\ \vdots \\ a_p \end{pmatrix}$$

$$=\begin{pmatrix} 1 & \cdots & 1 \\ d_{11} & \cdots & d_{n1} \\ \vdots & \ddots & \vdots \\ d_{1p} & \cdots & d_{np}1 \end{pmatrix}\begin{pmatrix} d_{1y} \\ \vdots \\ d_{ny} \end{pmatrix}$$

式 i

つぎのように行列を記号で表します。

$$\begin{pmatrix} 1 & d_{11} & \cdots & d_{1p} \\ \vdots & \vdots & \ddots & \vdots \\ 1 & d_{n1} & \cdots & d_{np} \end{pmatrix}_{n\times(p+1)}=\boldsymbol{D_{+1}}$$

$$\begin{pmatrix} a_0 \\ a_1 \\ \vdots \\ a_p \end{pmatrix}=\boldsymbol{A_{+1}}$$

$$\begin{pmatrix} d_{1y} \\ \vdots \\ d_{ny} \end{pmatrix}=\boldsymbol{Y}$$

$$\boldsymbol{D_{+1}}^{\boldsymbol{T}}=\begin{pmatrix} 1 & \cdots & 1 \\ d_{11} & \cdots & d_{n1} \\ \vdots & \ddots & \vdots \\ d_{1p} & \cdots & d_{np}1 \end{pmatrix}$$

$$\boldsymbol{D_{+1}}^{\boldsymbol{T}}\boldsymbol{D_{+1}} = \begin{pmatrix} 1 & \cdots & 1 \\ d_{11} & \cdots & d_{n1} \\ \vdots & \ddots & \vdots \\ d_{1p} & \cdots & d_{np}1 \end{pmatrix}\begin{pmatrix} 1 & d_{11} & \cdots & d_{1p} \\ \vdots & \vdots & \ddots & \vdots \\ 1 & d_{n1} & \cdots & d_{np} \end{pmatrix} = \begin{pmatrix} n & \sum_{k-1}^{n} d_{k1} & \sum_{k=1}^{n} d_{k2} & \cdots & \sum_{k=1}^{n} d_{kp} \\ \sum_{k=1}^{n} d_{k1} & \sum_{k-1}^{n} {d_{k1}}^2 & \sum_{k=1}^{n} d_{k2}d_{k1} & \cdots & \sum_{k=1}^{n} d_{kp}d_{k1} \\ \sum_{k=1}^{n} d_{k2} & \sum_{k=1}^{n} d_{k1}d_{k2} & \sum_{k=1}^{n} {d_{k2}}^2 & \cdots & \sum_{k=1}^{n} d_{kp}d_{k2} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ \sum_{k=1}^{n} d_{kp} & \sum_{k-1}^{n} d_{k1}d_{kp} & \sum_{k=1}^{n} d_{k2}d_{kp} & \cdots & \sum_{k=1}^{n} {d_{kp}}^2 \end{pmatrix}$$

式ｉは次のように表せます。

$$\boldsymbol{D_{+1}}^{\boldsymbol{T}}\boldsymbol{D_{+1}}\boldsymbol{A_{+1}} = \boldsymbol{D_{+1}}^{\boldsymbol{T}}\boldsymbol{Y}$$

両辺に $\left(\boldsymbol{D_{+1}}^{\boldsymbol{T}}\boldsymbol{D_{+1}}\right)^{-1}$をかけます。.

$$\left(\boldsymbol{D_{+1}}^{\boldsymbol{T}}\boldsymbol{D_{+1}}\right)^{-1}\left(\boldsymbol{D_{+1}}^{\boldsymbol{T}}\boldsymbol{D_{+1}}\right)\boldsymbol{A_{+1}} = \left(\boldsymbol{D_{+1}}^{\boldsymbol{T}}\boldsymbol{D_{+1}}\right)^{-1}\boldsymbol{D_{+1}}^{\boldsymbol{T}}\boldsymbol{Y}$$

$$\boldsymbol{A_{+1}} = \left(\boldsymbol{D_{+1}}^{\boldsymbol{T}}\boldsymbol{D_{+1}}\right)^{-1}\boldsymbol{D_{+1}}^{\boldsymbol{T}}\boldsymbol{Y}$$

$$\left(\boldsymbol{D_{+1}}^{\boldsymbol{T}}\boldsymbol{D_{+1}}\right)^{-1}\boldsymbol{D_{+1}}^{\boldsymbol{T}} = \boldsymbol{D_{+1}}^{\#}$$

$\boldsymbol{D_{+1}}^{\#}$: $\boldsymbol{D_{+1}}$の疑似逆行列

これで疑似逆行列による計算と微分方程式による計算が同じだということが確認できましたが、よけいな蛇足ですから、本質的な話に戻ります。これらの計算をもう少し単純化して表現してみます。単純化のために定数項を無視します。

$$\boldsymbol{DA} = \boldsymbol{Y}$$

$$\boldsymbol{A} = \boldsymbol{D}^{\boldsymbol{T}}\boldsymbol{Y}$$

$$(\boldsymbol{D}^{\boldsymbol{T}}\boldsymbol{D})^{-\boldsymbol{1}}\boldsymbol{D}^{\boldsymbol{T}}\boldsymbol{DA} = (\boldsymbol{D}^{\boldsymbol{T}}\boldsymbol{D})^{-\boldsymbol{1}}\boldsymbol{D}^{\boldsymbol{T}}\boldsymbol{Y}$$

$$\boldsymbol{A} = (\boldsymbol{D}^{\boldsymbol{T}}\boldsymbol{D})^{-\boldsymbol{1}}\boldsymbol{D}^{\boldsymbol{T}}\boldsymbol{Y}$$

$$\boldsymbol{S} = \boldsymbol{D}^{\boldsymbol{T}}\boldsymbol{D} = \begin{pmatrix} d_{11} & \cdots & d_{n1} \\ \vdots & \ddots & \vdots \\ d_{1p} & \cdots & d_{np} \end{pmatrix}\begin{pmatrix} d_{11} & \cdots & d_{1p} \\ \vdots & \ddots & \vdots \\ d_{n1} & \cdots & d_{np} \end{pmatrix} = \begin{pmatrix} \sum_{k-1}^{n} {d_{k1}}^2 & \sum_{k=1}^{n} d_{k2}d_{k1} & \cdots & \sum_{k=1}^{n} d_{kp}d_{k1} \\ \sum_{k=1}^{n} d_{k1}d_{k2} & \sum_{k=1}^{n} {d_{k2}}^2 & \cdots & \sum_{k=1}^{n} d_{kp}d_{k2} \\ \vdots & \vdots & \ddots & \vdots \\ \sum_{k-1}^{n} d_{k1}d_{kp} & \sum_{k=1}^{n} d_{k2}d_{kp} & \cdots & \sum_{k=1}^{n} {d_{kp}}^2 \end{pmatrix}$$

ここで、次のような記述法を導入します。

$$SS_{ij} = \sum_{k=1}^{n} d_{ki}d_{kj}$$

$$\boldsymbol{S} = \boldsymbol{D}^{\boldsymbol{T}}\boldsymbol{D} = \begin{pmatrix} SS_{11} & SS_{12} & \cdots & SS_{1i} & \cdots & SS_{1n} \\ SS_{21} & SS_{12} & \cdots & SS_{2i} & \cdots & SS_{2n} \\ \vdots & \vdots & \ddots & \vdots & & \vdots \\ SS_{j1} & SS_{j2} & \cdots & SS_{ji} & \cdots & SS_{jn} \\ \vdots & \vdots & & \vdots & \ddots & \vdots \\ SS_{n1} & SS_{n2} & \cdots & SS_{ni} & \cdots & SS_{nn} \end{pmatrix}$$

これを使って、

$$\boldsymbol{D^TY}=\begin{pmatrix} d_{11} & \cdots & d_{n1} \\ \vdots & \ddots & \vdots \\ d_{1p} & \cdots & d_{np} \end{pmatrix}\begin{pmatrix} d_{1y} \\ \vdots \\ d_{ny} \end{pmatrix}=\begin{pmatrix} \sum_{k=1}^{n} d_{k1}d_{ky} \\ \vdots \\ \sum_{k=1}^{n} d_{kp}d_{ky} \end{pmatrix}$$

$$\boldsymbol{D^TDA}=\boldsymbol{D^TY}$$

$$\boldsymbol{SA}=\begin{pmatrix} SS_{1y} \\ \vdots \\ SS_{py} \end{pmatrix}$$

$$\boldsymbol{A}=\boldsymbol{S^{-1}}\begin{pmatrix} SS_{1y} \\ \vdots \\ SS_{py} \end{pmatrix}$$

こうするとコンパクトに解が示せます。著者は簡潔な表現が好みですが、計算過程がわからないので、こういう表現に不満を持つ読者もいるでしょう。そこで、この表現に従った計算の例を示します。

標本番号.	1	2	3	y
1	d_{11}	d_{12}	d_{13}	d_{1y}
2	d_{21}	d_{22}	d_{23}	d_2y
$\vdots$		$\vdots$		$\vdots$
k	d_{k1}	d_{k2}	d_{k3}	d_{ky}
$\vdots$		$\vdots$		$\vdots$
n	d_{n1}	d_{n2}	d_{n3}	d_{ny}

$$\boldsymbol{S}=\boldsymbol{D^TD}=\begin{pmatrix} d_{11} & \cdots & d_{n1} \\ \vdots & \ddots & \vdots \\ d_{1p} & \cdots & d_{np} \end{pmatrix}\begin{pmatrix} d_{11} & \cdots & d_{1p} \\ \vdots & \ddots & \vdots \\ d_{n1} & \cdots & d_{np} \end{pmatrix}=\begin{pmatrix} SS_{11} & SS_{12} & SS_{13} \\ SS_{21} & SS_{22} & SS_{23} \\ SS_{31} & SS_{32} & SS_{33} \end{pmatrix}$$

$$\begin{pmatrix} d_{11} & \cdots & d_{n1} \\ \vdots & \ddots & \vdots \\ d_{1p} & \cdots & d_{np} \end{pmatrix}\begin{pmatrix} d_{1y} \\ \vdots \\ d_{ny} \end{pmatrix}=\begin{pmatrix} SS_{1y} \\ SS_{2y} \\ SS_{3y} \end{pmatrix}$$

Calculation of determinant

$$|\boldsymbol{S}|=\begin{vmatrix} SS_{11} & SS_{12} & SS_{13} \\ SS_{21} & SS_{22} & SS_{23} \\ SS_{31} & SS_{32} & SS_{33} \end{vmatrix}$$

$$=SS_{11}SS_{22}SS_{33}+SS_{12}SS_{23}SS_{31}+SS_{13}SS_{21}SS_{32}-SS_{13}SS_{22}SS_{31}$$
$$-SS_{12}SS_{21}SS_{33}-SS_{11}SS_{23}SS_{32}$$

$$= SS_{11}SS_{22}SS_{33} + SS_{12}SS_{23}SS_{31} + SS_{13}SS_{21}SS_{32} - SS_{11}{SS_{23}}^2 - SS_{22}{SS_{13}}^2 - SS_{33}{SS_{12}}^2$$

余因子行列 $\tilde{\boldsymbol{S}}$

$$\tilde{\boldsymbol{S}} = \begin{pmatrix} \begin{vmatrix} SS_{22} & SS_{23} \\ SS_{32} & SS_{33} \end{vmatrix} & -\begin{vmatrix} SS_{21} & SS_{23} \\ SS_{31} & SS_{33} \end{vmatrix} & \begin{vmatrix} SS_{21} & SS_{22} \\ SS_{31} & SS_{32} \end{vmatrix} \\ -\begin{vmatrix} SS_{12} & SS_{13} \\ SS_{32} & SS_{33} \end{vmatrix} & \begin{vmatrix} SS_{11} & SS_{13} \\ SS_{31} & SS_{33} \end{vmatrix} & -\begin{vmatrix} SS_{11} & SS_{12} \\ SS_{31} & SS_{32} \end{vmatrix} \\ \begin{vmatrix} SS_{12} & SS_{13} \\ SS_{22} & SS_{23} \end{vmatrix} & -\begin{vmatrix} SS_{11} & SS_{13} \\ SS_{21} & SS_{23} \end{vmatrix} & \begin{vmatrix} SS_{11} & SS_{12} \\ SS_{21} & SS_{22} \end{vmatrix} \end{pmatrix}'$$

$\boldsymbol{S}$は対称行列だから

$$\tilde{\boldsymbol{S}} = \tilde{\boldsymbol{S}}^{\boldsymbol{T}}$$

$$\boldsymbol{S}^{-\boldsymbol{1}} = \frac{\tilde{\boldsymbol{S}}}{\begin{vmatrix} SS_{11} & SS_{12} & SS_{13} \\ SS_{21} & SS_{22} & SS_{23} \\ SS_{31} & SS_{32} & SS_{33} \end{vmatrix}}$$

$$S^{-1} = \frac{\begin{pmatrix} \begin{vmatrix} SS_{22} & SS_{23} \\ SS_{32} & SS_{33} \end{vmatrix} & -\begin{vmatrix} SS_{12} & SS_{13} \\ SS_{32} & SS_{33} \end{vmatrix} & \begin{vmatrix} SS_{12} & SS_{13} \\ SS_{22} & SS_{23} \end{vmatrix} \\ -\begin{vmatrix} SS_{21} & SS_{23} \\ SS_{31} & SS_{33} \end{vmatrix} & \begin{vmatrix} SS_{11} & SS_{13} \\ SS_{31} & SS_{33} \end{vmatrix} & -\begin{vmatrix} SS_{11} & SS_{13} \\ SS_{21} & SS_{23} \end{vmatrix} \\ \begin{vmatrix} SS_{21} & SS_{22} \\ SS_{31} & SS_{32} \end{vmatrix} & -\begin{vmatrix} SS_{11} & SS_{12} \\ SS_{31} & SS_{32} \end{vmatrix} & \begin{vmatrix} SS_{11} & SS_{12} \\ SS_{21} & SS_{22} \end{vmatrix} \end{pmatrix}}{\begin{vmatrix} SS_{11} & SS_{12} & SS_{13} \\ SS_{21} & SS_{22} & SS_{23} \\ SS_{31} & SS_{32} & SS_{33} \end{vmatrix}}$$

$$\begin{pmatrix} a_1 \\ a_2 \\ a_3 \end{pmatrix} = \boldsymbol{S}^{-\boldsymbol{1}} \begin{pmatrix} SS_{1y} \\ SS_{2y} \\ SS_{3y} \end{pmatrix}$$

$$\begin{pmatrix} a_1 \\ a_2 \\ a_3 \end{pmatrix} = \frac{1}{\begin{vmatrix} SS_{11} & SS_{12} & SS_{13} \\ SS_{21} & SS_{22} & SS_{23} \\ SS_{31} & SS_{32} & SS_{33} \end{vmatrix}} \begin{pmatrix} \begin{vmatrix} SS_{22} & SS_{23} \\ SS_{32} & SS_{33} \end{vmatrix} & -\begin{vmatrix} SS_{12} & SS_{13} \\ SS_{32} & SS_{33} \end{vmatrix} & \begin{vmatrix} SS_{12} & SS_{13} \\ SS_{22} & SS_{23} \end{vmatrix} \\ -\begin{vmatrix} SS_{21} & SS_{23} \\ SS_{31} & SS_{33} \end{vmatrix} & \begin{vmatrix} SS_{11} & SS_{13} \\ SS_{31} & SS_{33} \end{vmatrix} & -\begin{vmatrix} SS_{11} & SS_{13} \\ SS_{21} & SS_{23} \end{vmatrix} \\ \begin{vmatrix} SS_{21} & SS_{22} \\ SS_{31} & SS_{32} \end{vmatrix} & -\begin{vmatrix} SS_{11} & SS_{12} \\ SS_{31} & SS_{32} \end{vmatrix} & \begin{vmatrix} SS_{11} & SS_{12} \\ SS_{21} & SS_{22} \end{vmatrix} \end{pmatrix} \begin{pmatrix} SS_{1y} \\ SS_{2y} \\ SS_{3y} \end{pmatrix}$$

$$a_1 = \frac{\begin{vmatrix} SS_{22} & SS_{23} \\ SS_{32} & SS_{33} \end{vmatrix} SS_{1y} - \begin{vmatrix} SS_{12} & SS_{13} \\ SS_{32} & SS_{33} \end{vmatrix} SS_{2y} + \begin{vmatrix} SS_{12} & SS_{13} \\ SS_{22} & SS_{23} \end{vmatrix} SS_{3y}}{SS_{11}SS_{22}SS_{33} + SS_{12}SS_{23}SS_{31} + SS_{13}SS_{21}SS_{32} - SS_{11}{SS_{23}}^2 - SS_{22}{SS_{13}}^2 - SS_{33}{SS_{12}}^2}$$

$$a_2 = \frac{-\begin{vmatrix} SS_{21} & SS_{23} \\ SS_{31} & SS_{33} \end{vmatrix} SS_{1y} + \begin{vmatrix} SS_{11} & SS_{13} \\ SS_{31} & SS_{33} \end{vmatrix} SS_{2y} - \begin{vmatrix} SS_{11} & SS_{13} \\ SS_{21} & SS_{23} \end{vmatrix} SS_{3y}}{SS_{11}SS_{22}SS_{33} + SS_{12}SS_{23}SS_{31} + SS_{13}SS_{21}SS_{32} - SS_{11}{SS_{23}}^2 - SS_{22}{SS_{13}}^2 - SS_{33}{SS_{12}}^2}$$

$$a_3 = \frac{\begin{vmatrix} SS_{21} & SS_{22} \\ SS_{31} & SS_{32} \end{vmatrix} SS_{1y} - \begin{vmatrix} SS_{11} & SS_{12} \\ SS_{31} & SS_{32} \end{vmatrix} SS_{2y} + \begin{vmatrix} SS_{11} & SS_{12} \\ SS_{21} & SS_{22} \end{vmatrix} SS_{3y}}{SS_{11}SS_{22}SS_{33} + SS_{12}SS_{23}SS_{31} + SS_{13}SS_{21}SS_{32} - SS_{11}{SS_{23}}^2 - SS_{22}{SS_{13}}^2 - SS_{33}{SS_{12}}^2}$$

これでもまだ式が複雑なので、著者は次のような記述法を提案します。

$$S^{-1}=\frac{\begin{pmatrix}\begin{vmatrix}SS_{22} & SS_{23}\\ SS_{32} & SS_{33}\end{vmatrix} & -\begin{vmatrix}SS_{12} & SS_{13}\\ SS_{32} & SS_{33}\end{vmatrix} & \begin{vmatrix}SS_{12} & SS_{13}\\ SS_{22} & SS_{23}\end{vmatrix}\\ -\begin{vmatrix}SS_{21} & SS_{23}\\ SS_{31} & SS_{33}\end{vmatrix} & \begin{vmatrix}SS_{11} & SS_{13}\\ SS_{31} & SS_{33}\end{vmatrix} & -\begin{vmatrix}SS_{11} & SS_{13}\\ SS_{21} & SS_{23}\end{vmatrix}\\ \begin{vmatrix}SS_{21} & SS_{22}\\ SS_{31} & SS_{32}\end{vmatrix} & -\begin{vmatrix}SS_{11} & SS_{12}\\ SS_{31} & SS_{32}\end{vmatrix} & \begin{vmatrix}SS_{11} & SS_{12}\\ SS_{21} & SS_{22}\end{vmatrix}\end{pmatrix}}{\begin{vmatrix}SS_{11} & SS_{12} & SS_{13}\\ SS_{21} & SS_{22} & SS_{23}\\ SS_{31} & SS_{32} & SS_{33}\end{vmatrix}}$$

$$=\frac{\begin{pmatrix}\begin{vmatrix}SS_{22} & SS_{23}\\ SS_{32} & SS_{33}\end{vmatrix} & -\begin{vmatrix}SS_{12} & SS_{13}\\ SS_{32} & SS_{33}\end{vmatrix} & \begin{vmatrix}SS_{12} & SS_{13}\\ SS_{22} & SS_{23}\end{vmatrix}\\ -\begin{vmatrix}SS_{21} & SS_{23}\\ SS_{31} & SS_{33}\end{vmatrix} & \begin{vmatrix}SS_{11} & SS_{13}\\ SS_{31} & SS_{33}\end{vmatrix} & -\begin{vmatrix}SS_{11} & SS_{13}\\ SS_{21} & SS_{23}\end{vmatrix}\\ \begin{vmatrix}SS_{21} & SS_{22}\\ SS_{31} & SS_{32}\end{vmatrix} & -\begin{vmatrix}SS_{11} & SS_{12}\\ SS_{31} & SS_{32}\end{vmatrix} & \begin{vmatrix}SS_{11} & SS_{12}\\ SS_{21} & SS_{22}\end{vmatrix}\end{pmatrix}}{|\boldsymbol{S}|}$$

$$=\begin{pmatrix}\frac{\begin{vmatrix}SS_{22} & SS_{23}\\ SS_{32} & SS_{33}\end{vmatrix}}{|\boldsymbol{S}|} & -\frac{\begin{vmatrix}SS_{12} & SS_{13}\\ SS_{32} & SS_{33}\end{vmatrix}}{|\boldsymbol{S}|} & \frac{\begin{vmatrix}SS_{12} & SS_{13}\\ SS_{22} & SS_{23}\end{vmatrix}}{|\boldsymbol{S}|}\\ \frac{-\begin{vmatrix}SS_{21} & SS_{23}\\ SS_{31} & SS_{33}\end{vmatrix}}{|\boldsymbol{S}|} & \frac{\begin{vmatrix}SS_{11} & SS_{13}\\ SS_{31} & SS_{33}\end{vmatrix}}{|\boldsymbol{S}|} & -\frac{\begin{vmatrix}SS_{11} & SS_{13}\\ SS_{21} & SS_{23}\end{vmatrix}}{|\boldsymbol{S}|}\\ \frac{\begin{vmatrix}SS_{21} & SS_{22}\\ SS_{31} & SS_{32}\end{vmatrix}}{|\boldsymbol{S}|} & -\frac{\begin{vmatrix}SS_{11} & SS_{12}\\ SS_{31} & SS_{32}\end{vmatrix}}{|\boldsymbol{S}|} & \frac{\begin{vmatrix}SS_{11} & SS_{12}\\ SS_{21} & SS_{22}\end{vmatrix}}{|\boldsymbol{S}|}\end{pmatrix}$$

$$=\begin{pmatrix}SS^{11} & SS^{12} & SS^{13}\\ SS^{21} & SS^{22} & SS^{23}\\ SS^{31} & SS^{32} & SS^{33}\end{pmatrix}$$

$$SS^{ij}=\frac{S\text{の}i,j\text{余因子}}{|\boldsymbol{S}|}$$

$$SS^{ij}=SS^{ji}$$

$$\begin{pmatrix}SS^{11} & \cdots & SS^{1p}\\ \vdots & \ddots & \vdots\\ SS^{p1} & \cdots & SS^{pp}\end{pmatrix}\begin{pmatrix}SS_{11} & \cdots & SS_{1p}\\ \vdots & \ddots & \vdots\\ SS_{p1} & \cdots & SS_{pp}\end{pmatrix}=\boldsymbol{I}$$

これを使うと次のように簡略化出来ます。

$$a_1=SS^{11}SS_{1y}+SS^{12}SS_{2y}+SS^{13}SS_{3y}$$
$$a_2=SS^{21}SS_{1y}+SS^{22}SS_{2y}+SS^{23}SS_{3y}$$
$$a_3=SS^{31}SS_{1y}+SS^{32}SS_{2y}+SS^{33}SS_{3y}$$

VI-1-1-4. 回帰の有意性

行列を使えば$\boldsymbol{A}$の最小二乗解を求める計算は簡単です。私たちが考えなければならないのは回帰係数の有意性です。そのために、各変数によるバリアンスを分離する必要があります

目的変数の観測値と回帰による予測値（期待値）の差の式は次のように単純です。

$$y = \hat{y} + e$$

ここでは平均値からの差によって標準化したデータを考えています。

この式の両辺を二乗します。

$$y^2 = (\hat{y} + e)^2 = \hat{y}^2 + 2\hat{y}e + e^2$$

その総和は以下の通りです。

$$\sum_{j=1}^{n} {y_j}^2 == \sum_{j=1}^{n} {\hat{y}_j}^2 + 2\sum_{j=1}^{n} \hat{y}_j e_j + \sum_{j=1}^{n} {e_j}^2$$

右辺の第二項が 0 ならば、次のように平方和を分離できます。

$$\text{目的変数の観測値の平方和}: \mathrm{SS}_y = \sum_{j=1}^{n} {y_j}^2$$

$$\text{目的変数の期待値の平方和}\ \mathrm{SS}_{\hat{y}} = \sum_{j=1}^{n} {\hat{y}_j}^2$$

$$\text{誤差の平方和}\ \mathrm{SS}_e = \sum_{j=1}^{n} {e_j}^2$$

私たちは $\sum_{j=1}^{n} \hat{y}_j e_j$が 0 であることを期待していますが、この式から直接 $\sum_{j=1}^{n} \hat{y}_j e_j = 0$という ことはできませんので証明します。

$\sum_{j=1}^{n} \hat{y}_j e_j = 0$の証明

疑似逆行列を使うと

$$\boldsymbol{A} = \boldsymbol{X}^{\#}\boldsymbol{Y} = (\boldsymbol{X}^{\boldsymbol{T}}\boldsymbol{X})^{-1}\boldsymbol{X}^{\boldsymbol{T}}\boldsymbol{Y}$$

$$\widehat{\boldsymbol{Y}} = \boldsymbol{X}\boldsymbol{A} = \boldsymbol{X}(\boldsymbol{X}^{\boldsymbol{T}}\boldsymbol{X})^{-1}\boldsymbol{X}^{\boldsymbol{T}}\boldsymbol{Y}$$

$$\boldsymbol{Y} = \boldsymbol{I}\,\boldsymbol{Y}$$

$$\boldsymbol{I}: \text{単位行列}$$

$$\boldsymbol{E} = \boldsymbol{Y} - \widehat{\boldsymbol{Y}} = (\boldsymbol{I} - \boldsymbol{X}(\boldsymbol{X}^{\boldsymbol{T}}\boldsymbol{X})^{-1}\boldsymbol{X}^{\boldsymbol{T}})\boldsymbol{Y}$$

$$\boldsymbol{X}^{\boldsymbol{T}}\boldsymbol{E} = \boldsymbol{X}^{\boldsymbol{T}}(\boldsymbol{I} - \boldsymbol{X}(\boldsymbol{X}^{\boldsymbol{T}}\boldsymbol{X})^{-1}\boldsymbol{X}^{\boldsymbol{T}})\boldsymbol{Y} = \boldsymbol{0}$$

$$\because \boldsymbol{X}^{\boldsymbol{T}}(\boldsymbol{I} - \boldsymbol{X}(\boldsymbol{X}^{\boldsymbol{T}}\boldsymbol{X})^{-1}\boldsymbol{X}^{\boldsymbol{T}}) = \boldsymbol{X}^{\boldsymbol{T}} - \boldsymbol{X}^{\boldsymbol{T}}\boldsymbol{X}(\boldsymbol{X}^{\boldsymbol{T}}\boldsymbol{X})^{-1}\boldsymbol{X}^{\boldsymbol{T}} = \boldsymbol{X}^{\boldsymbol{T}} - \boldsymbol{X}^{\boldsymbol{T}} = \boldsymbol{0}$$

$$\boldsymbol{A}^{\boldsymbol{T}}\boldsymbol{X}^{\boldsymbol{T}}\boldsymbol{E} = \boldsymbol{0}$$

$$\boldsymbol{A}^{\boldsymbol{T}}\boldsymbol{X}^{\boldsymbol{T}} = (\boldsymbol{X}\boldsymbol{A})^{\boldsymbol{T}} = \boldsymbol{Y}^{\boldsymbol{T}}$$

$$\boldsymbol{Y}^{\boldsymbol{T}}\boldsymbol{E} = \boldsymbol{0}$$

$$\sum_{j=1}^{n} \hat{y}_j e_j = \boldsymbol{Y}^{\boldsymbol{T}}\boldsymbol{E} = 0$$

証明終わり

以上で分散の分離が出来ました。

$$\mathrm{SS}_y = \mathrm{SS}_{\hat{y}} + \mathrm{SS}_e$$

$$\mathrm{SS}_y=\sum_{j=1}^{n} {y_j}^2,\quad \mathrm{SS}_{\hat{y}}=\sum_{j=1}^{n} \hat{y_j}^2,\quad \mathrm{SS}_e=\sum_{j=1}^{n} {e_j}^2$$

VI-1-1-2. で説明した回帰の幾何学的意味からも、次のことは理解可能だと思います。

$$\mathrm{SS}_y=\mathrm{SS}_{\hat{y}}+\mathrm{SS}_e$$

$$\mathrm{SS}_y=\sum_{j=1}^{n} {y_j}^2,\quad \mathrm{SS}_{\hat{y}}=\sum_{j=1}^{n} \hat{y_j}^2,\quad \mathrm{SS}_e=\sum_{j=1}^{n} {e_j}^2$$

全自由度が $n-1$ で誤差の平方和 SS_e の重度が $n-p-1$ ですから、回帰の自由度は p です。これらを使うと重回帰の分散分析表を表 39 のように作ることが出来ます。

表 39. 重回帰分析の結果の分散分析表

factor	SS	degree od freedom	variance (V)	ratio（F）
total	SS_{yy}	$n-1$	$V_t=\frac{SS_R}{n-1}$	
regression	SS_R	p	$V_R=\frac{SS_R}{p}$	$\mathrm{F}_0=\frac{V_R}{V_e}$
residual	SS_e	$n-p-1$	$V_e=\frac{SS_R}{n-p-1}$	

これによって、$a_1=a_2=\cdots=a_p=0$ を帰無仮説としてＦ検定が出来ます。しかし、個人的には、著者はこの分散分析にあまり意味を感じません。重回帰分析をするときは、もともと説明変数に意味があるだろうと思って回帰しているので、帰無仮説が棄却されて意味がないとは言えないと言われても、あまりうれしくありません。私たちが知りたいのは、回帰式の説明力です。つまり、次の式で示す、全体の分散に占める予測値の分散の割合です。これを決定係数といいます。

$$\mathrm{R}^2=\frac{\mathrm{SS}_{\hat{y}}}{\mathrm{SS}_y}=\frac{\sum_{j=1}^{n} {y_j}^2}{\sum_{j=1}^{n} {y_j}^2}=\frac{\mathrm{SS}_y-\mathrm{SS}_{\hat{e}}}{\mathrm{SS}_y}=1-\frac{\mathrm{SS}_{\hat{e}}}{\mathrm{SS}_y}$$

$$\mathrm{R}=\sqrt{\frac{\mathrm{SS}_{\hat{y}}}{\mathrm{SS}_y}}=\sqrt{\frac{\sum_{j=1}^{n} {y_j}^2}{\sum_{j=1}^{n} {y_j}^2}}$$

R^2: *決定係数*

R: 重回帰係数

重回帰係数は直観的にはわかりやすい回帰の有意性の指標です。しかし、重回帰係数には数学的な弱点があります。重回帰係数は説明変数（因子数）が増加すると増加します。そして、説明変数の数と標本数が一致すると重回帰係数は１になります。たとえば、１０人の人に所持金と左右１０本の指の長さを訊いて、所持金を目的変数として、１０本の指の長さで回帰すると、回帰式が作れて、決定係数は１になります。データの行列が正則になるからです。

そのようなことを避けるために、自由度調整済みの決定係数を使うべきです。自由度調整済決定係数は全分散と予測値の分散の比で次の式で求めることが出来ます。

$$v_e = \frac{\mathrm{SS}_{\hat{e}}}{(n-1-p)}$$

$$v_t = \frac{\mathrm{SS}_y}{(n-1)}$$

$$\mathrm{R^2}_{adj} = 1 - \frac{v_e}{v_t} = 1 - \frac{\mathrm{SS}_{\hat{e}}}{\mathrm{SS}_y}\frac{(n-1)}{(n-1-p)}$$

$\mathrm{R^2}_{adj}$: 自由度調整済決定係数

p: この場合定数項を変数の数に含む。説明変数の自由度は$p-1$

多くの場合、私たちが知りたいのは、ここの説明変数の有意性です。重回帰分析の目的の一つは単純かつ十分な説明です。つまり、不必要な説明変数は取り除きたいのです。単純に考えると、係数の大きさを比較して、小さのものをカットすればよいように思います。しかし、実際には、説明変数は cm mm, 円, kg, トン、個数等々様々な単位が使われています。ミリメータで測られた変数の係数がセンチメータで測られた係数の１０分の１になることはわかります。つまり、直接比較することはできません。しかし、すべての変数の分散が等しければ、直接比較できます。一つの考え方は、すべての変数のデータを標準偏差で割って、標準化しておくことです。これならば直接比較できます。しかし、著者は一般化してこの方法を薦めることはしません。分析方法は分析者が決めるべきです。分析の目的を知っているのは分析者自身だからです。例えば、人々の幸福度と各支出項目の支出額の間で重相関分析をするとします。その場合、ある種の支出項目例えば娯楽費と教養にかかわるような支出はもともと分散が大きく、食費などは分散が少ないのです。そんな場合、絶対金額を使う方が意味があるかもしれません。

各係数の t 検定も変数の選択に使えます。IV-3-1 で説明したように、student の t 検定は確率的に計算された t の臨界値と実測された t 値の比較ですが、この場合、係数の推定値の誤差範囲の中に 0 が含まれる確率です。したがって係数a_iと 0 の距離を標準誤差SE_iで割ったものが t の計算値です。

$$\mathrm{t} = \frac{a_i - 0}{\mathrm{SE}_i}$$

SE_iを求める方法を考えます。

全体の誤差は次の通りです。

$$SE_{total}\sqrt{\frac{v_e}{n}}$$

$$v_e = \frac{\mathrm{SS}_{\hat{e}}}{(n-1-p)}$$

ここで私たちが前の段落で作った記法を使います。

$$\begin{pmatrix} SS^{11} & \cdots & SS^{1p} \\ \vdots & \ddots & \vdots \\ SS^{p1} & \cdots & SS^{pp} \end{pmatrix} \begin{pmatrix} SS_{11} & \cdots & SS_{13} \\ \vdots & \ddots & \vdots \\ SS_{31} & \cdots & SS_{33} \end{pmatrix} = \boldsymbol{I}$$

$\begin{pmatrix} SS^{11} & \cdots & SS^{1p} \\ \vdots & \ddots & \vdots \\ SS^{p1} & \cdots & SS^{pp} \end{pmatrix}$ が全体の誤差を各因子に振り分けていると考えると考えます。

この行列のうち$(SS^{i1} \quad \cdots \quad SS^{ip})$が a_iにかかわる因子ですが、$SS^{i1} \quad \cdots \quad SS^{ip}$は行列式と余因子の比です。また、$SS^{ii}$は$a_i$にかかわる因子の中で$SS_{11}$とは独立の因子で説明変数 i で説明できない部分です。ですから、この比が全体の分散中で、a_i の誤差分散が占めている割合です。

$$SE_i = \sqrt{SS^{ii}\frac{v_e}{n}}$$

$$\mathrm{t} = \frac{a_i}{\sqrt{SS^{ii}\frac{v_e}{n}}}$$

表 40 係数の t 検定の結果の例

変数	係数	標準誤差	t 値	P 値
x_1	a_1	$\sqrt{SS^{11}\frac{v_e}{n}}$	$\frac{a_1}{\sqrt{SS^{11}\frac{v_e}{n}}}$	P_1
x_2	a_2	$\sqrt{SS^{22}\frac{v_e}{n}}$	$\frac{a_2}{\sqrt{SS^{22}\frac{v_e}{n}}}$	P_2
			$\vdots$	
x_p	a_2	$\sqrt{SS^{pp}\frac{v_e}{n}}$	$\frac{a_2}{\sqrt{SS^{pp}\frac{v_e}{n}}}$	P_p

P は統計書またはコンピュータソフトウェアーを参照.

Student の t 検定をつかって、説明変数の選択に重要な情報が得られます。しかし、場合によっては、それだけでは不十分な場合もあります。まず、ある説明変数と他の説明変数が加わった場合の効果（交互作用）についても考えなければならないかもしれません。重回帰分析の目的が数理モデルを作ることであれば、様々な因子の組み合わせの効果も考えなくてはならないでしょう。組み合わせの効果を見つけ出す方法はいくつか考えられます。

交差項を含めて変数の数が限られていれば、総当たりで組み合わせの効果を考えても良いでしょうが、変数の数が増えるとすべての組み合わせを試すことは現実的に難しいでしょう。必要な変数を決定するための方法や説明力の指標はいくつかあります。基本的には説明変数を加えたり差し引いたりします。

変数増加法：最も t 検定のpが小さい変数の単回帰から始めます。単回帰式の回帰式の説明

力の指標を調べます（説明力の指標にはさまざまなものがありますが、ここでは単純に自由度調整済み決定係数を使います。）。次にpの小さい変数を加えて、自由度調整済決定係数が増加すれば、この変数を変数として加え、さらに次にpの小さい変数１つふやし自由度調整済決定係数の変化を調べます。自由度が考慮されているので、変数を加えてもpは必ずしも上昇しません。このような手順で自由度調整済決定係数が増加しなくなるまでこれを繰り返し、自由度調整済み決定係数が増加しなくなるところまでの変数を変数として採用します。、
変数減少法：すべての変数を使って重回帰分析します。次に最も t 検定のpが大きい変数を取り除いて重回帰分析し、自由度調整済決定係数を計算し、決定係数が増加すれば、前の変数を取り除きます。これを繰り返して、決定係数が現象を始めたところで、その時点で残されている変数を重回帰分析の変数とします。
ステップワイズ法：ある数の重回帰分析から始めて、変数を一つたします。足すことによって自由度調整済決定係数等の指標が上昇すれば、その変数を加えます。次に別の変数を取り除きます。自由度調整済み決定係数が増加すればその変数を取り除きます。次に別の変数を加えて同じことを行い。これを繰り返して、最も自由度調整済み決定係数等の指標が高い変数の組み合せを見つけます。

モデルを作る場合、少ない変数の数で高い説明力があることが望ましい。そのような考え方で、赤池の情報基準などいくつかの説明力の効率性の指標があります。それについては、他の教科書あるいはネットの説明などを参照してください。

もう一つの重回帰分析にかかわる数学的な問題は多重共線性（multicollinearity）です。多重共線性とは説明変数同士が相関を持つことです。多重共線性があると、係数を相関を持つ変数同士が引っ張り合ってしまうので、結果が不安定になります。そうなると、モデルを一つに決定できなくなります。多重共線性がある場合には、相関を持つ変数のなかから代表的な説明辺を選ぶか、相関性持つ変数を合成して一つの合成変数を作るなどしてこの問題を解決します。重回帰分析の計算では、分散共分散行列を作りますから、この計算を機械的にやらずに、分散・共分散行列の計算結果をしっかりと確認しましょう。分かりにくければ相関行列を作ってみましょう。重回帰分析の前にこれはやっておくべきです。重回帰分析に先立って主成分分析(PCA)をやっておくというのも良いかもしれません。これで多重共線性がはっきり確認できますし、主成分得点を使って合成変数を作ることもできます。主成分分析は数学的には頑健で確実なのですが、出てきた主成分の意味が個別科学の視点から良く分からないことがあるという、実際上の問題があります。反対に因子分析のように、思わぬ潜在的因子が見つかることもあります。
例えば、100 メートル走の記録と体の各部位の測定値（身長、体重、座高、胸囲、年齢など）をすべての年代の人からデータとして取ったとします。この場合、重回帰分析で短距離走の

能力と身体的特徴の関係を分析することは十分考えられますが、しかし、これらの測定項目はすべて体の大きさと関係しています、したがって強い多重共線性があります。これらのデータを直接、重回帰分析してもあまり意味のある結果は得られないでしょう。これらのデータを主成分分析すると、第一主成分にはおそらく体の大きさが来ます。他の成分はこの第一主成分と直交しています。それらの成分の中で、走力と相関していそうなものを見つけ出せば、それらが走力と関係していると推測できるでしょう。もう一つのやりかたとしては、その分野における知識や経験によってそのような因子を考えるという方法があります。例えばそれらは肥満度、相対的な足の長さ、筋肉量のようなものです。それらは以下のように計算出来ます。

$$肥満度\text{：}\ \mathrm{BMI} = \frac{体重}{(身長)^3}$$

$$相対的足の長さ\text{：}\ 1 - \frac{座高}{身長}$$

$$相対的筋肉量\text{：}\ \frac{胸囲}{肥満度}$$

主成分の二次元プロットによって、私たちは有効な合成変数を見つけることが出来るでしょう。これら合成変数と体の大きさの代表としての身長を使えば、おそらく意味のあるモデルが作れるはずです。

変数の数が少ない時には、変数の選択をコンピュータのソフトウェアーに任せて行うことはあまり勧められません。著者は、官能検査の結果を使って、チーズの物理的性状と味の関係を重回帰分析で分析した修士論文を読んだことがあります。変数選択はステップワイズ法を使っていました。彼女は最も味に関係する物理性状はチーズの硬さだと結論していました。彼女の結論はおそらく正しくないか不十分だと筆者は思います。何故ならば、固さとアミノ酸量はどちらもチーズの発酵時間に関係しています。多重共線性があるのです。官能検査に参加した人はアミノ酸の量や組み合わせの違いを感じ取っていたのだと思います。
この話の含意は、内容を考えずに重回帰することのリスクと盲目的にコンピュータソフトウェアの結果を受け入れることの危険性です。私たちは重回帰分析の結果を自分の知識経験で解釈し、受け入れるべき変数を自分で判断しなければなりません。