2015年10月13日

九州大学 海洋システム工学専攻

システム設計特論 （木村）
講義資料

2) 多重回帰

火曜3限（13：00～14：30)
場所：セミナー室

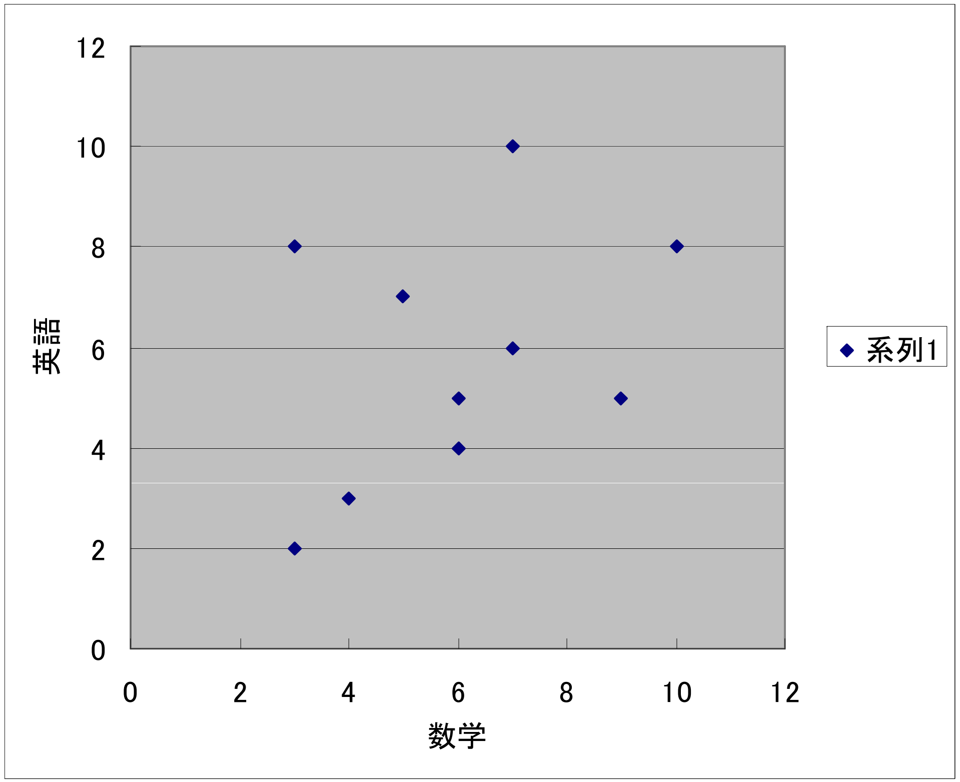
【復習】 回帰分析

「身長」と「体重」
「数学」と「英語」の点数　など　⇒　２変量についての関係を調べる

２変量データの例）

数学 x	英語 y
3	8
6	4
10	8
4	3
7	6
7	10
3	2
9	5
6	5
5	7

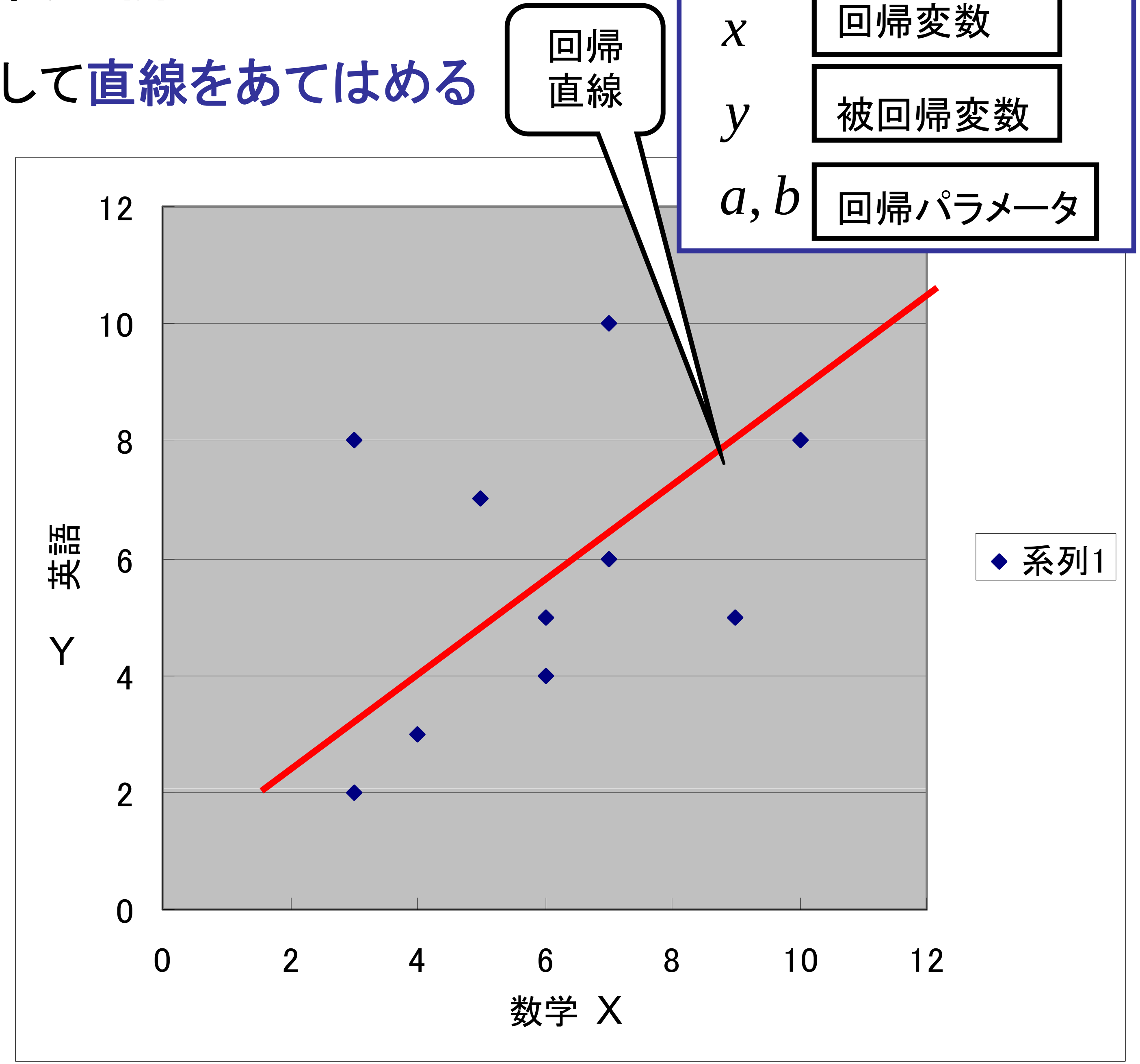
２変量データの表現方法：　散布図



【復習】回帰分析

２変量の関係として**直線をあてはめる**

数学 x	英語 y
3	8
6	4
10	8
4	3
7	6
7	10
3	2
9	5
6	5
5	7



$$y = ax + b$$

x 回帰変数

y 被回帰変数

a, b 回帰パラメータ

【復習】回帰分析

回帰パラメータa,bの求め方

回帰直線
$y = ax + b$ 　yのxに対する回帰直線

n個のデータの組を $(x_1, y_1), (x_2, y_2), \cdots, (x_n, y_n)$ と表す

データの各点から同じ x_i の回帰直線までの距離を d_i とし、

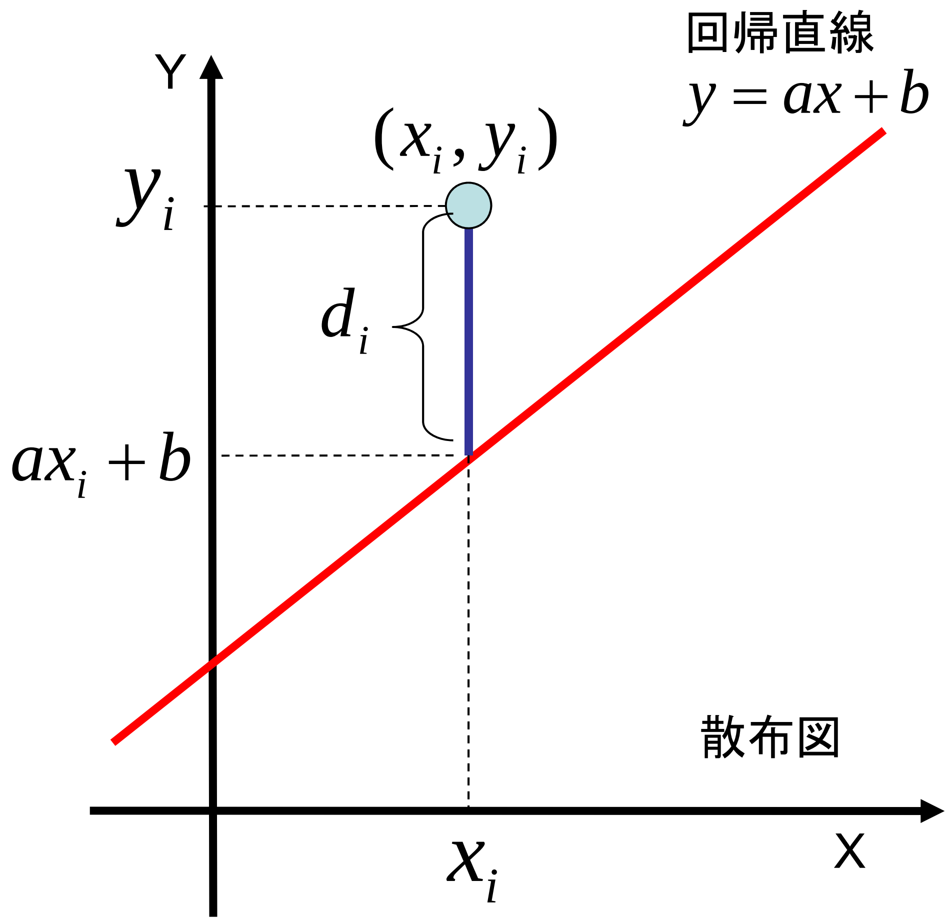
この長さの2乗和Lを最小にするように

回帰パラメータa, bを決める
（**最小2乗法**）

$$L = \sum_{i=1}^{n} d_i^{\,2} = \boxed{\sum_{i=1}^{n} \left(y_i - (ax_i + b)\right)^2}$$

$$\frac{\partial L}{\partial a} = 0, \quad \frac{\partial L}{\partial b} = 0$$

の連立1次方程式を解いてa, bを求める



x の平均 $\bar{x} = \frac{1}{n}\sum_{i=1}^{n} x_i$ y の平均 $\bar{y} = \frac{1}{n}\sum_{i=1}^{n} y_i$ とすると、

$$a = \boxed{\frac{\left(\sum_{i=1}^{n} x_i y_i\right) - n\bar{x}\bar{y}}{\left(\sum_{i=1}^{n} x_i^2\right) - n\bar{x}^2}} \qquad b = \boxed{\bar{y} - a\bar{x}}$$

ここで、

$$S_{xx} = \frac{1}{n}\sum_{i=1}^{n} (x_i - \bar{x})^2$$

x の分散

$$S_{yy} = \frac{1}{n}\sum_{i=1}^{n} (y_i - \bar{y})^2$$

y の分散

$$S_{xy} = \frac{1}{n}\sum_{i=1}^{n} (x_i - \bar{x})(y_i - \bar{y}) = \boxed{\left(\frac{1}{n}\sum_{i=1}^{n} x_i y_i\right) - \bar{x}\,\bar{y}}$$

共分散 Covariance

とおくと、

$$a = \boxed{\frac{S_{xy}}{S_{xx}}}$$

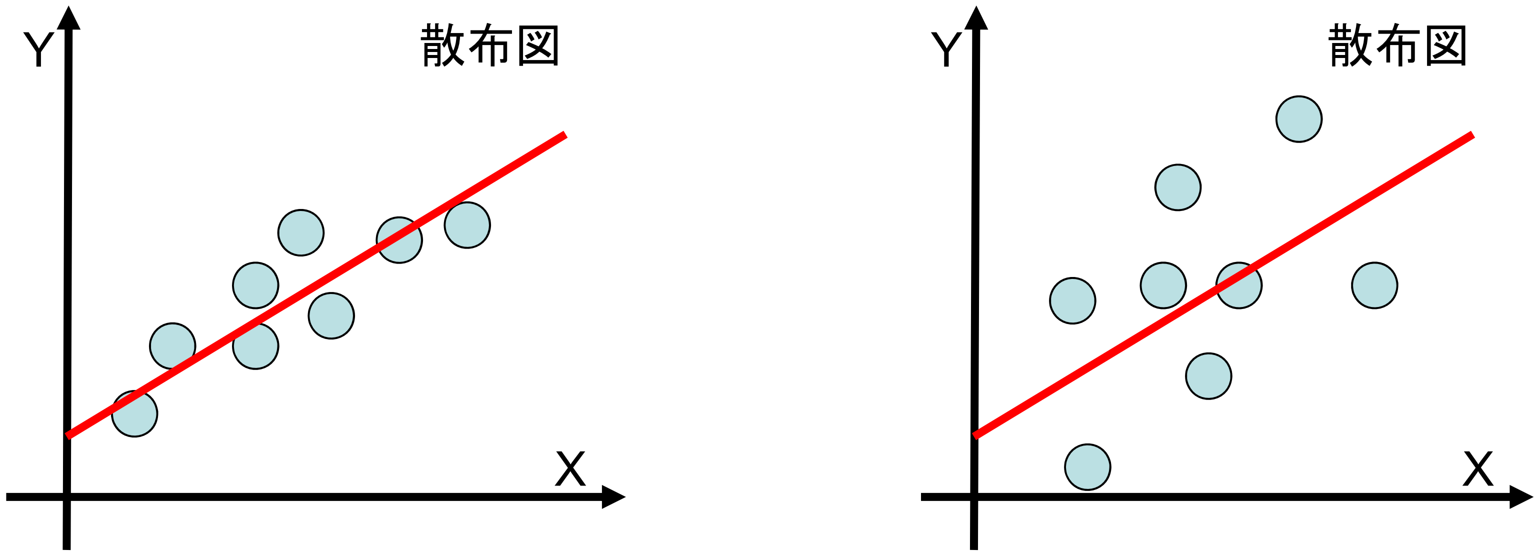
となり、求める直線は

$$y - \bar{y} = \frac{S_{xy}}{S_{xx}}(x - \bar{x})$$

すなわち

点 $(\bar{x}, \bar{y})$ を通り、傾き $\frac{S_{xy}}{S_{xx}}$ の直線である

相関係数　回帰直線のまわりに密集しているデータの度合い



ピアソンの**標本相関係数**

$$r = \frac{S_{xy}}{\sqrt{S_{xx}}\sqrt{S_{yy}}}$$

共分散

x の標準偏差

y の標準偏差

この係数 r は $\boxed{-1 \leq r \leq 1}$ であり、散らばりが少ないとき、|r| は1に近い値をとる。

r > 0　のとき　回帰直線の傾きが正（x, y の間に正の相関）
r < 0　のとき　回帰直線の傾きが負（x, y の間に負の相関）

多重回帰

観測値 y を変数　$x_1, x_2, \cdots x_K$　を用いて以下の式で説明する:

$$y = b_0 + b_1 x_1 + b_2 x_2 + \cdots + b_K x_K + e$$

確率変動・誤差

このとき、n 個の観測値 $(y_1, x_{11}, x_{12}, \cdots x_{1K}), (y_2, x_{21}, x_{22}, \cdots x_{2K}), (y_n, x_{n1}, x_{n2}, \cdots x_{nK})$
によって係数　$b_0, b_1, b_2, \cdots b_K$ の最小2乗推定量を求める。ここで、

目的変数行列

$$\mathbf{y} = \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{bmatrix}$$

説明変数行列

$$\mathbf{X} = \begin{bmatrix} 1 & x_{11} & x_{12} & \cdots x_{1K} \\ 1 & x_{21} & x_{22} & \cdots x_{2K} \\ \vdots & \vdots & \vdots & \ddots \quad \vdots \\ 1 & x_{n1} & x_{n2} & \cdots x_{nK} \end{bmatrix}$$

回帰係数行列

$$\mathbf{b} = \begin{bmatrix} b_0 \\ b_1 \\ \vdots \\ b_K \end{bmatrix}$$

誤差変数行列

$$\mathbf{e} = \begin{bmatrix} e_1 \\ e_2 \\ \vdots \\ e_n \end{bmatrix}$$

と表すと、

$$\mathbf{y} = \mathbf{X}\mathbf{b} + \mathbf{e}$$

線形表現

誤差変数行列 $\mathbf{e}$ の平方和 $\|\mathbf{e}\|^2$ を最小にする $\mathbf{b}$ を求める

$\rightarrow$　回帰推定（最小2乗法）　回帰モデル

データから回帰モデルを得て何がうれしいか？
回帰モデルによる推定

未知の説明変数(回帰変数)の値が　$(x_{q1}, x_{q2}, \cdots x_{qK})$　で与えられたときの

目的変数(被回帰変数)の値　y_q　をデータから**推定**できる！

それでは、
回帰係数行列　$\mathbf{b} = \begin{bmatrix} b_0 \\ b_1 \\ \vdots \\ b_K \end{bmatrix}$　をデータからどのように求めるか？

データから回帰モデルを得て何がうれしいか？
回帰モデルによる推定

未知の説明変数（回帰変数）の値が $(x_{q1}, x_{q2}, \cdots x_{qK})$ で与えられたときの

目的変数（被回帰変数）の値 y_q をデータから**推定**できる！

$$y_q = b_0 + b_1 x_{q1} + b_2 x_{q2} + \cdots + b_K x_{qK}$$

推定値

誤差eの項はゼロで計算

それでは、
回帰係数行列 $\mathbf{b} = \begin{bmatrix} b_0 \\ b_1 \\ \vdots \\ b_K \end{bmatrix}$ をデータからどのように求めるか？

誤差ベクトル $\mathbf{e}$ の平方和 $\|\mathbf{e}\|^2$ を最小にする $\mathbf{b}$ を最尤推定値 $\hat{\mathbf{b}}$ と表すと、

単純回帰の場合と同様に、回帰係数の各要素で誤差ベクトルの平方和を偏微分し、
これらが全てゼロとした連立方程式を立てて解くことにより、回帰係数ベクトルは
以下の式で計算される：

ただし $\mathbf{X}^{\mathbf{Trans}}$ は $\mathbf{X}$ の転置行列を表す。

$\|\mathbf{e}\|^2$ の最小値 S_e を残差平方和といい、

±σ の範囲内に68.27%のデータが存在

で与えられる。 $\sigma = \sqrt{\dfrac{S_e}{n}}$ より、回帰で推定する場合の精度が分かる

誤差ベクトル $\mathbf{e}$ の平方和 $\|\mathbf{e}\|^2$ を最小にする $\mathbf{b}$ を最尤推定値 $\hat{\mathbf{b}}$ と表すと、

単純回帰の場合と同様に、回帰係数の各要素で誤差ベクトルの平方和を偏微分し、これらがゼロとした連立方程式を立てて解くことにより、回帰係数ベクトルは以下の式で計算される:

$$\hat{\mathbf{b}} = \underbrace{\left(\mathbf{X}^{\mathbf{Trans}}\mathbf{X}\right)^{-1}\mathbf{X}^{\mathbf{Trans}}}_{\text{Xの擬似逆行列 } \mathbf{X}^{+}}\mathbf{y}$$

ただし $\mathbf{X}^{\mathbf{Trans}}$ は $\mathbf{X}$ の転置行列を表す。

pseudo-inverse matrix
ただしXはm行n列、m > n

$\|\mathbf{e}\|^2$ の最小値 S_e を残差平方和といい、

$$S_e = \mathbf{y}^{\mathbf{Trans}}\left\{\mathbf{I} - \mathbf{X}\left(\mathbf{X}^{\mathbf{Trans}}\mathbf{X}\right)^{-1}\mathbf{X}^{\mathbf{Trans}}\right\}\mathbf{y}$$

で与えられる。 ➡ $\sigma = \sqrt{\dfrac{S_e}{n}}$ より、回帰で推定する場合の精度が分かる

±σ の範囲内に68.27%のデータが存在

誤差ベクトル $\mathbf{e}$ の平方和 $\|\mathbf{e}\|^2$ を最小にする $\hat{\mathbf{b}}$ の導出

$\mathbf{y} = \mathbf{X}\mathbf{b} + \mathbf{e}$ より $\|\mathbf{e}\|^2 = \left(\mathbf{X}\mathbf{b} - \mathbf{y}\right)^{Trans}\left(\mathbf{X}\mathbf{b} - \mathbf{y}\right)$

$\mathbf{b}$ の2次関数なので
微分してゼロの $\mathbf{b}$ を求める

$$\frac{\|\mathbf{e}\|^2}{\partial \mathbf{b}} = 2\mathbf{X}^{Trans}\mathbf{X}\mathbf{b} - 2\mathbf{X}^{Trans}\mathbf{y} = 0$$

$\mathbf{b}$ について解く

$$\mathbf{b} = \left(\mathbf{X}^{\mathbf{Trans}}\mathbf{X}\right)^{-1}\mathbf{X}^{\mathbf{Trans}}\mathbf{y}$$

$\hat{\mathbf{b}}$

【復習】 転置行列とは？

$$\begin{bmatrix} 1 & 2 & 3 \\ 4 & 5 & 6 \\ 7 & 8 & 9 \end{bmatrix}^{Trans} = \begin{bmatrix} 1 & 4 & 7 \\ 2 & 5 & 8 \\ 3 & 6 & 9 \end{bmatrix}$$

$$\begin{bmatrix} 1 & 2 & 3 & 4 \\ 5 & 6 & 7 & 8 \\ 9 & 10 & 11 & 12 \end{bmatrix}^{Trans} = \begin{bmatrix} 1 & 5 & 9 \\ 2 & 6 & 10 \\ 3 & 7 & 11 \\ 4 & 8 & 12 \end{bmatrix}$$

$$\mathbf{X}^{Trans}\mathbf{X} = \begin{bmatrix} 1 & 1 & \cdots & 1 \\ x_{11} & x_{21} & \cdots & x_{n1} \\ x_{12} & x_{22} & \cdots & x_{n2} \\ \vdots & \vdots & \ddots & \vdots \\ x_{1k} & x_{2k} & \cdots & x_{nk} \end{bmatrix} \begin{bmatrix} 1 & x_{11} & x_{12} & \cdots x_{1K} \\ 1 & x_{21} & x_{22} & \cdots x_{2K} \\ \vdots & \vdots & \vdots & \ddots \quad \vdots \\ 1 & x_{n1} & x_{n2} & \cdots x_{nK} \end{bmatrix}$$

$$= \begin{bmatrix} n & \sum_{i=1}^{n} x_{i1} & \sum_{i=1}^{n} x_{i2} & \cdots & \sum_{i=1}^{n} x_{ik} \\ \sum_{i=1}^{n} x_{i1} & \sum_{i=1}^{n} x_{i1}{}^{2} & \sum_{i=1}^{n} x_{i1}x_{i2} & \cdots & \sum_{i=1}^{n} x_{i1}x_{ik} \\ \sum_{i=1}^{n} x_{i2} & \sum_{i=1}^{n} x_{i1}x_{i2} & \sum_{i=1}^{n} x_{i2}{}^{2} & \cdots & \sum_{i=1}^{n} x_{i2}x_{ik} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ \sum_{i=1}^{n} x_{ik} & \sum_{i=1}^{n} x_{i1}x_{ik} & \sum_{i=1}^{n} x_{i2}x_{ik} & \cdots & \sum_{i=1}^{n} x_{ik}{}^{2} \end{bmatrix}$$

分散・共分散行列
に関係

多重共線性

観測値 y を変数　$x_1, x_2, \cdots x_K$　を用いて以下の式で説明している：

$$y = b_0 + b_1 x_1 + b_2 x_2 + \cdots + b_K x_K + e$$

確率変動・誤差

もし、例えば変数　$x_1 = x_2$　だったら、 回帰係数ベクトル　$\mathbf{b}$　は一意には定まらない

→　計算が不安定になり、無意味な解が出やすい

一般に、以下の関係式

$$a_0 + a_1 x_1 + a_2 x_2 + \cdots + a_K x_K = 0$$

に近い関係があるとき、データは強い多重共線関係にあるという。
（ただし a は任意の定数）　このような場合、冗長な変数を取り除いてからモデル化

多重共線性の判定：　分散共分散行列の固有値の最大値と最小値の比率が
1000を超える場合、多重共線性ありと判断

補足：分散共分散行列とは？

2変数　x_i, x_j　についての共分散を全ての組合せで
計算して行列として表したもの。
（i=j　の場合、対角要素に相当し、単なる分散になる）

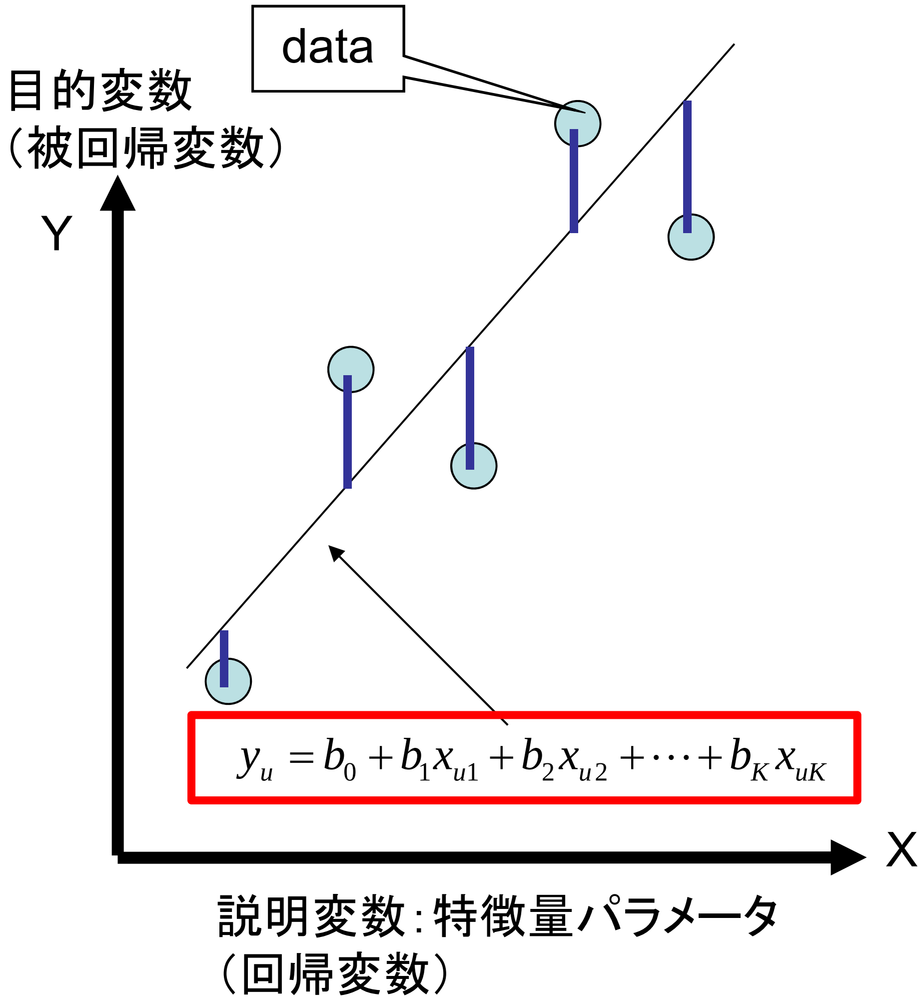
分散・共分散行列

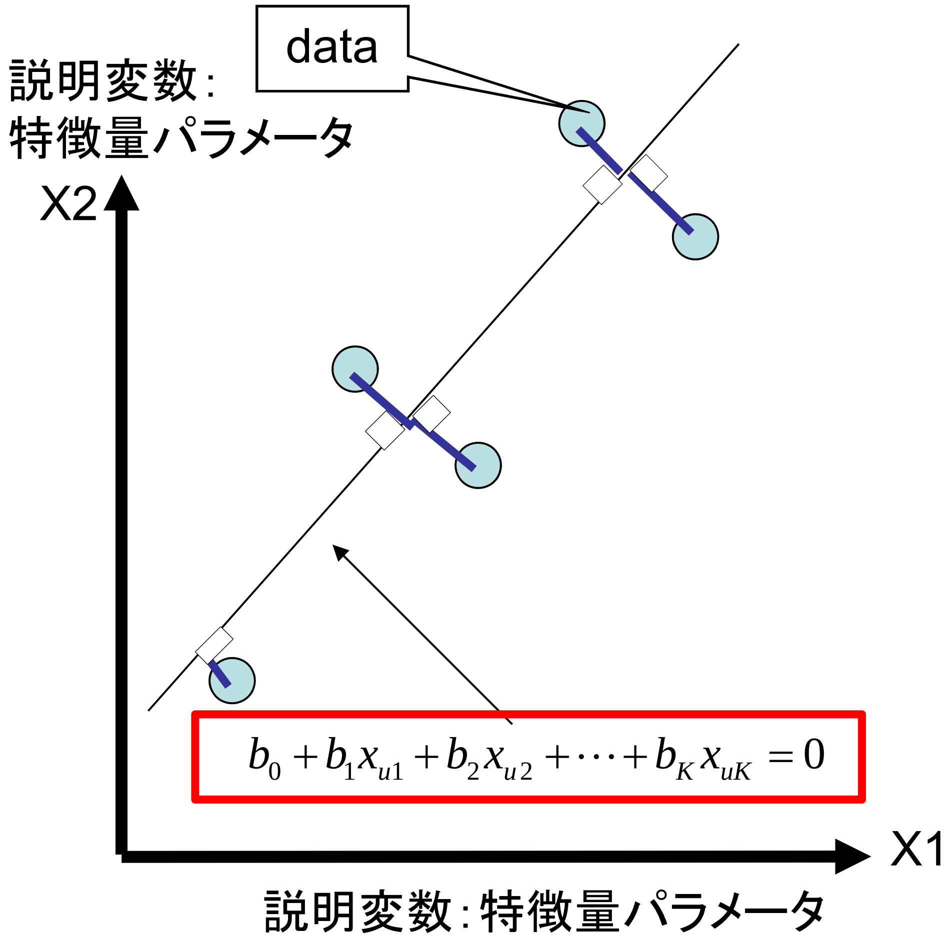
$$
\begin{bmatrix}
\frac{1}{n}\sum_{i=1}^{n}(x_{i1}-\bar{x}_1)^2 & \frac{1}{n}\sum_{i=1}^{n}(x_{i1}-\bar{x}_1)(x_{i2}-\bar{x}_2) & \cdots & \frac{1}{n}\sum_{i=1}^{n}(x_{i1}-\bar{x}_1)(x_{ik}-\bar{x}_k) \\
\frac{1}{n}\sum_{i=1}^{n}(x_{i1}-\bar{x}_1)(x_{i2}-\bar{x}_2) & \frac{1}{n}\sum_{i=1}^{n}(x_{i2}-\bar{x}_2)^2 & \cdots & \frac{1}{n}\sum_{i=1}^{n}(x_{i2}-\bar{x}_2)(x_{ik}-\bar{x}_k) \\
\vdots & \vdots & \ddots & \vdots \\
\frac{1}{n}\sum_{i=1}^{n}(x_{i1}-\bar{x}_1)(x_{ik}-\bar{x}_k) & \frac{1}{n}\sum_{i=1}^{n}(x_{i2}-\bar{x}_2)(x_{ik}-\bar{x}_k) & \cdots & \frac{1}{n}\sum_{i=1}^{n}(x_{ik}-\bar{x}_k)^2
\end{bmatrix}
$$

$$
=\frac{1}{n}
\begin{bmatrix}
\sum_{i=1}^{n} x_{i1}{}^2 & \sum_{i=1}^{n} x_{i2}x_{i1} & \cdots & \sum_{i=1}^{n} x_{ik}x_{i1} \\
\sum_{i=1}^{n} x_{i1}x_{i2} & \sum_{i=1}^{n} x_{i2}{}^2 & \cdots & \sum_{i=1}^{n} x_{ik}x_{i2} \\
\vdots & \vdots & \ddots & \vdots \\
\sum_{i=1}^{n} x_{i1}x_{ik} & \sum_{i=1}^{n} x_{i2}x_{ik} & \cdots & \sum_{i=1}^{n} x_{ik}{}^2
\end{bmatrix}
-
\begin{bmatrix}
\bar{x}_1{}^2 & \bar{x}_2\bar{x}_1 & \cdots & \bar{x}_k\bar{x}_1 \\
\bar{x}_1\bar{x}_2 & \bar{x}_2{}^2 & \cdots & \bar{x}_k\bar{x}_2 \\
\vdots & \vdots & \ddots & \vdots \\
\bar{x}_1\bar{x}_k & \bar{x}_2\bar{x}_k & \cdots & \bar{x}_k{}^2
\end{bmatrix}
$$

$\mathbf{X}^{Trans}\mathbf{X}$　と見比べよ

参考:「多重回帰」と「平面のあてはめ」の違い





多重回帰
ある変数yとの誤差のみ最小化

（超）平面のあてはめ≒**主成分分析**
全変数間のばらつきに注目

互いに直交するばらつき最大方向　＝　**主成分**

【3次元の点群データに平面をあてはめる計算例】

N個の点群データの座標を $(x_i\ y_i\ z_i)$ ただし $i=1,2,\cdots N$ とする。これらとの2乗距離が最小となる平面 $ax+by+cz+d=0$ の係数 a,b,c,d を求める。

データの各座標軸毎の平均を $\bar{x}=\frac{1}{N}\sum_{i=1}^{N}x_i,\quad \bar{y}=\frac{1}{N}\sum_{i=1}^{N}y_i,\quad \bar{z}==\frac{1}{N}\sum_{i=1}^{N}z_i$ と表すと、

上記の平面はデータの重心 $(\bar{x},\bar{y},\bar{z})$ を通る。

ここで、点 (x_0,y_0,z_0) を通る平面の方程式は

$a(x-x_0)+b(y-y_0)+c(z-z_0)=0$ ただし(a,b,c)は法線ベクトル より

よって $x_0=\bar{x},\quad y_0=\bar{y},\quad z_0=\bar{z}$ として計算すると $d=-a\bar{x}-b\bar{y}-c\bar{z}$ を得る。

次に

点と平面の距離の公式:
平面 $ax+by+cz+d=0$
と点 (x_0,y_0,z_0) との距離は $\frac{|ax_0+by_0+cz_0+d|}{\sqrt{a^2+b^2+c^2}}$

この式の2乗を全データで合計した値を最小化する係数 a,b,c,d を求めれば良いが、分母と分子に係数があり計算困難

平面の方程式は一般に $ax+by+cz+d=0$ と表されるが、任意性がある。そこで

$a^2+b^2+c^2-1=0$ の**等式制約**を加えると上の点と平面の距離の式の分母が1になり
最大化関数が単なる2次式になる。
等式制約の極値問題はラグランジェの未定乗数法で解くのがお約束

データ　$x_i\ y_i\ z_i$　について、

$$\mathbf{X}=\begin{bmatrix} x_1-\bar{x} & y_1-\bar{y} & z_1-\bar{z} \\ x_2-\bar{x} & y_2-\bar{y} & z_2-\bar{z} \\ \vdots & \vdots & \vdots \\ x_n-\bar{x} & y_n-\bar{y} & z_n-\bar{z} \end{bmatrix},\quad \mathbf{a}=\begin{bmatrix} a \\ b \\ c \end{bmatrix},\quad \mathbf{e}=\begin{bmatrix} e_1 \\ e_2 \\ \vdots \\ e_n \end{bmatrix}$$

全データについての
点と平面の距離の2乗

と表すと、　$\|\mathbf{e}\|^2=[\mathbf{Xa}]^T\mathbf{Xa}$

このとき、ラグランジュ関数Lは　$L=[\mathbf{Xa}]^T\mathbf{Xa}+\lambda\left(\|\mathbf{a}\|^2-1\right)$

未定乗数

等式制約
$a^2+b^2+c^2-1=0$

$$\frac{\partial L}{\partial \mathbf{a}} = 2\mathbf{X}^T\mathbf{Xa}-2\lambda\mathbf{a}=0$$

これは固有値方程式の形式

$$\frac{\partial L}{\partial \lambda} = \|\mathbf{a}\|^2-1=0$$

よって行列 $\mathbf{a}=\begin{bmatrix} a \\ b \\ c \end{bmatrix}$ は行列 $\mathbf{X}^T\mathbf{X}$ の固有ベクトルより得る

nで除すると
分散・共分散行列

行列　$\mathbf{X}^T\mathbf{X}$　は半正値対称行列で、
固有値はすべてゼロ以上の実数

3D点群データへの
平面あてはめは、
主成分分析の計算
と等価

3x3行列の固有ベクトルなので答えが3つ出てくるが、
この式は分散・共分散行列を使った主成分分析と同じ。
主成分分析では固有値（＝データの分散）の大きいものに
対応する固有ベクトルを主成分として選択していくことから、
この場合平面の法線ベクトル(a,b,c)としてふさわしいのは
$\mathbf{X}^T\mathbf{X}$ の最小固有値に対応する固有ベクトルである。
（固有ベクトルの2乗ノルムを1にするのは簡単）

主成分分析（PCA）
＝　データの分散が最大になる方向に座標軸を設定し直す

3次元空間上のデータ　$x_i \ y_i \ z_i$　について、

$$\mathbf{X} = \begin{bmatrix} x_1 - \bar{x} & y_1 - \bar{y} & z_1 - \bar{z} \\ x_2 - \bar{x} & y_2 - \bar{y} & z_2 - \bar{z} \\ \vdots & \vdots & \vdots \\ x_n - \bar{x} & y_n - \bar{y} & z_n - \bar{z} \end{bmatrix}$$

と表すと、**分散・共分散行列**　$= \dfrac{1}{n}\mathbf{X}^T\mathbf{X}$

分散・共分散行列　の**固有値**　＝　対応する固有ベクトル方向への**データの分散**

分散・共分散行列の固有値を**大きい順に並べ**、それらに**対応する固有ベクトル**を
順に　**第１主成分、第2主成分、…**と呼び、データがk次元空間であれば主成分はk個

ゼロに近い固有値　＝ 固有値に対応する主成分（固有ベクトル）方向への
データのばらつきが無い　＝　データが縮退

多重共線性の判定へ利用

船舶基本設計における多重回帰の応用

船の使用目的　（タンカー・ばら積み貨物船・自動車運搬船・コンテナ船など）

船の大きさ　　（長さ・幅・喫水深さ）

船の営業速力

ここで多重回帰を利用

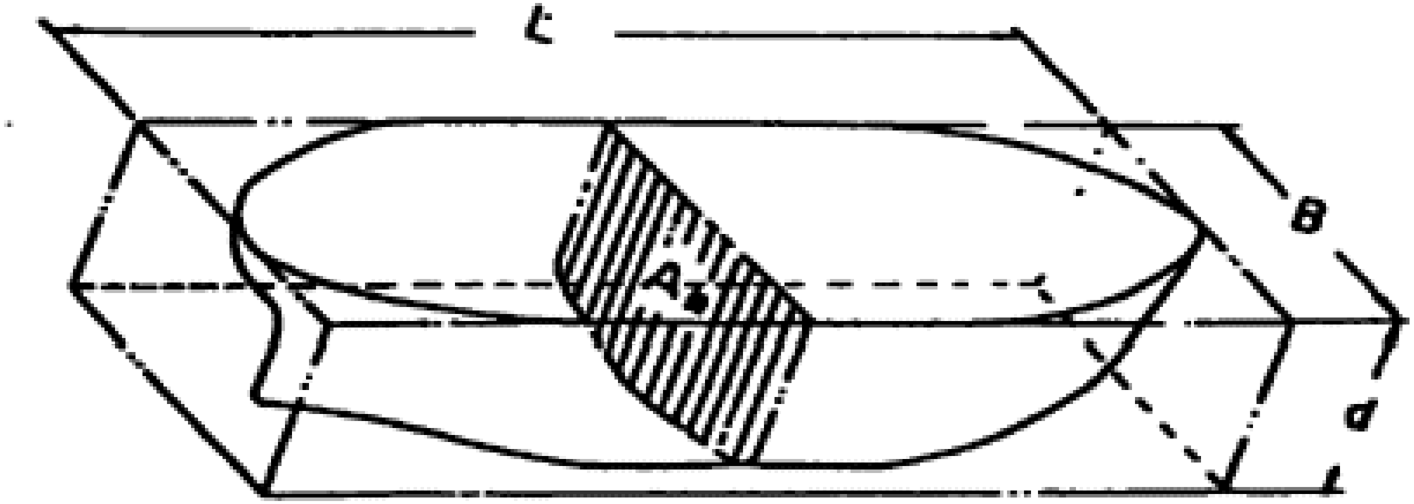
今までに製造したことのある類似船を参考に、おおよその船体形状を決め、流体抵抗を推定　（**推定**　および　**最適化**）

営業速力の違いで最適な船体形状は大きく異なる

・営業速力を出すのに必要な馬力を求め、エンジンを選定、
・貨物室の容積を決定、
・その他もろもろ…

詳細な船体形状を用いて流体シミュレーションを行ったり、
模型実験を行って推定した流体抵抗になることを確認するのはずっと後

（1）方形係数（Block coefficient）、CB この係数は、船体の水線下の
容積のやせている度合を示すもので、船の排水容積と、
これと長さ、幅、喫水の等しい直方体の容積との比で表わされる。



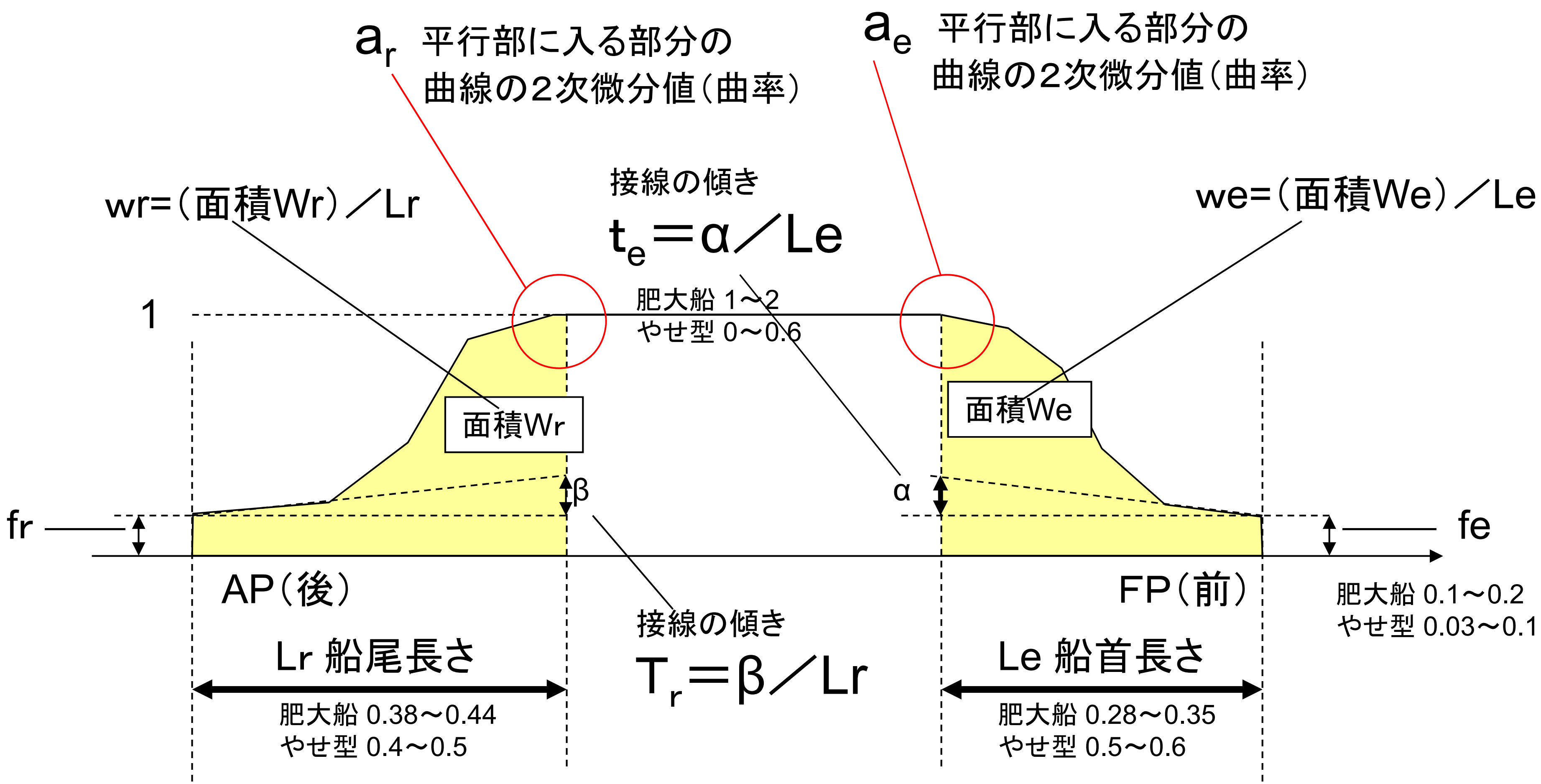
CBの値の小さい船をやせ型船、CBの大きい船を肥大船と言う。
CBの概略値は、
貨物船では0.62～0.84
旅客船では0.50～0.60
漁船では0.55～0.75

（日本財団電子図書館より引用）

船の要目：形状

CPカーブ　船の喫水部を輪切りにした断面積の曲線
最大部の断面積を1とする

パラメータ12個

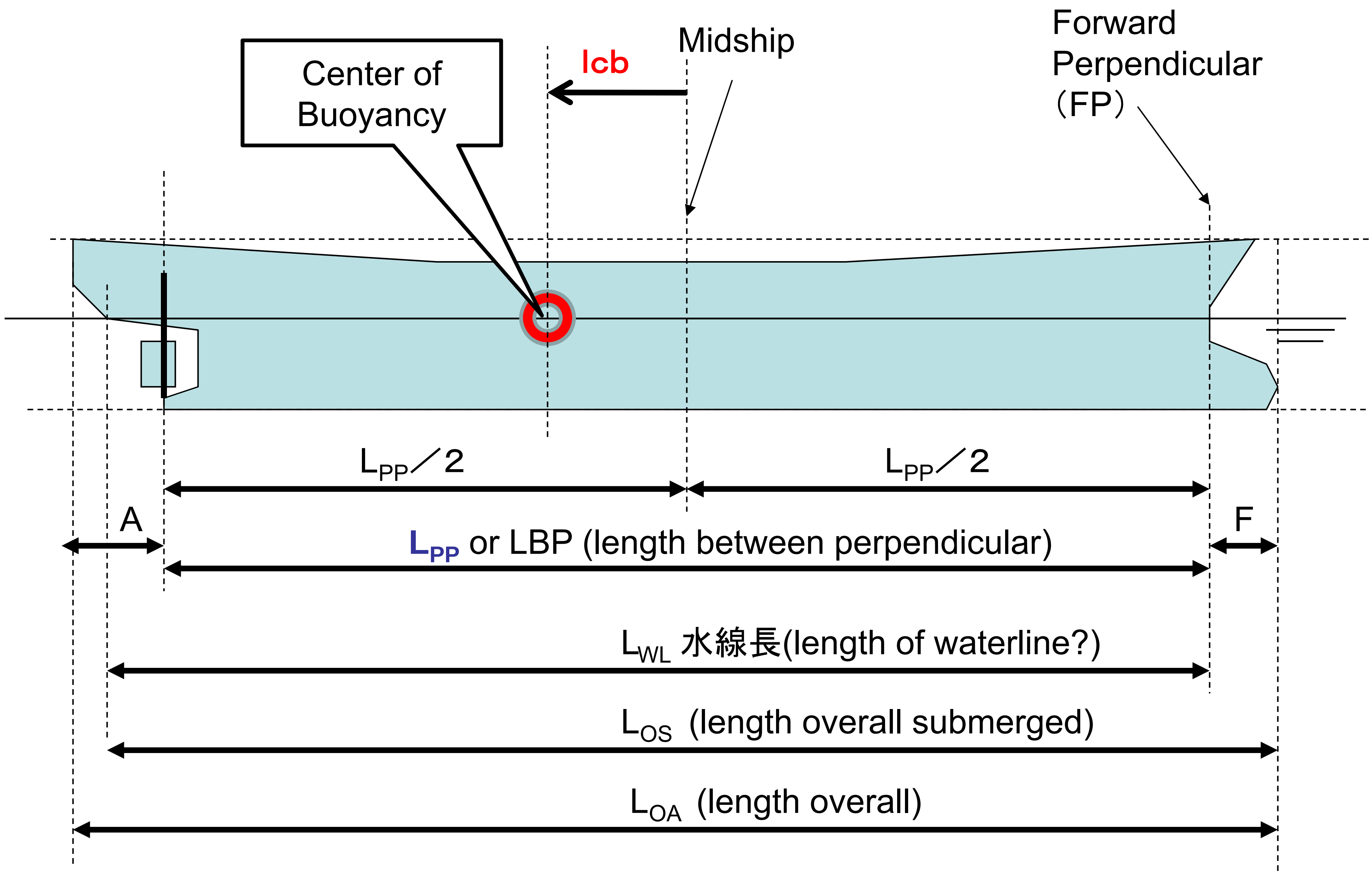


肥大度

$Hr／B = Lr\ (\ 1\ -\ wr\ )\ Lpp／B$　　肥大船 0.65〜0.9　やせ型 0.7〜1.4

$He／B = Le\ (\ 1\ -we\ )\ Lpp\ ／B$　　肥大船 0.3〜0.9　やせ型 0.7〜1.9

Important Parameters of ship hull (9)



Overhang = A／F

問題の定式化

船体形状を表すパラメータ集合

Given：

- L_{PP}
- L／B
- B／d
- Cb
- CPカーブ
- CWカーブ

これらを特徴付けるパラメータ(24個)

$x_1, x_2, \cdots x_K$

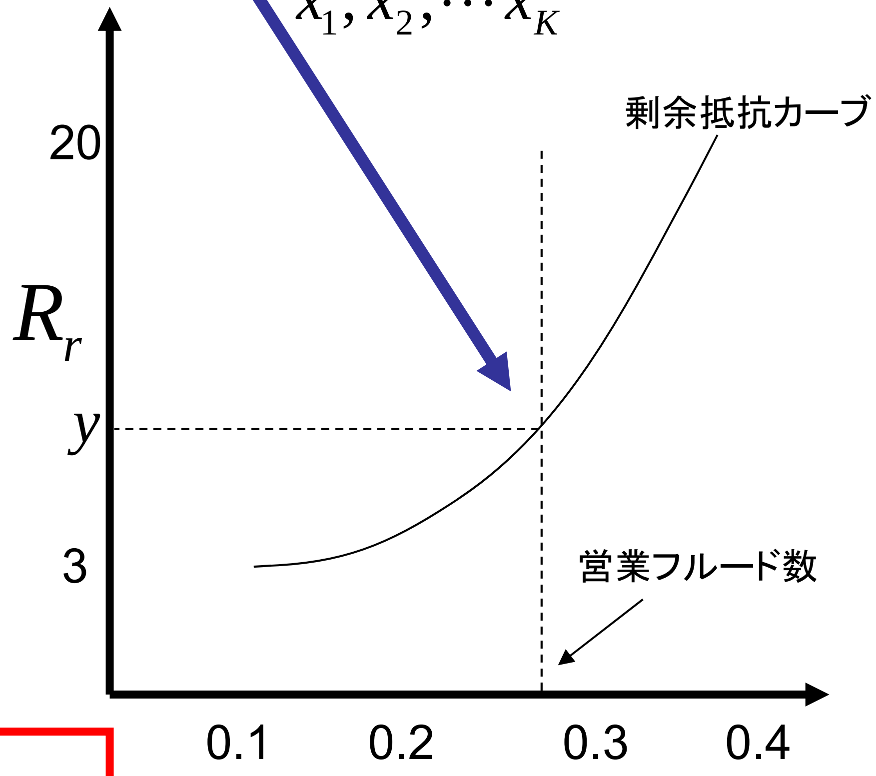
船の抵抗 $R_t = R_f + R_r$

摩擦抵抗　剰余抵抗

これを使う

$= (1 + K)R_f + R_W$

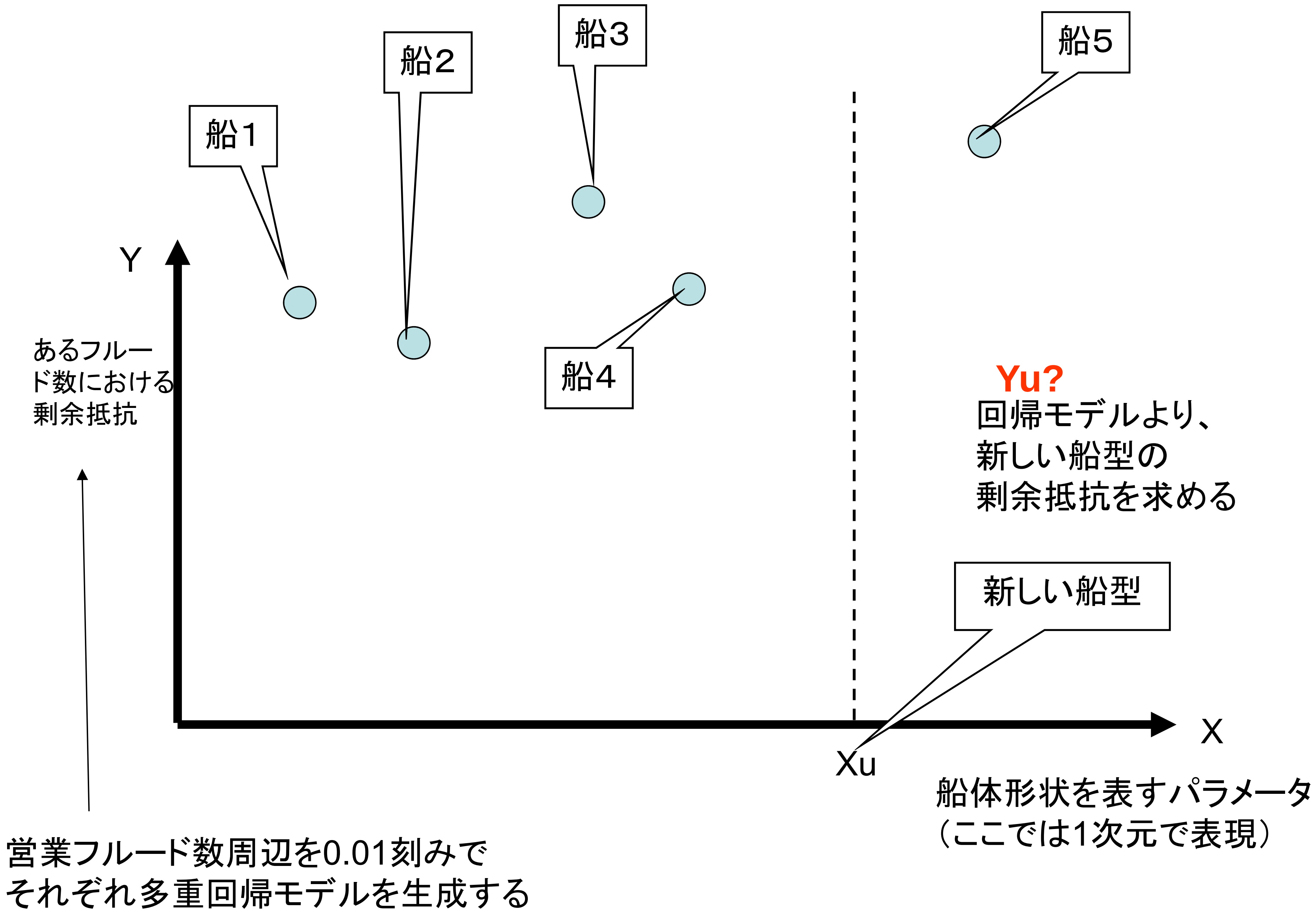
形状影響係数　造波抵抗



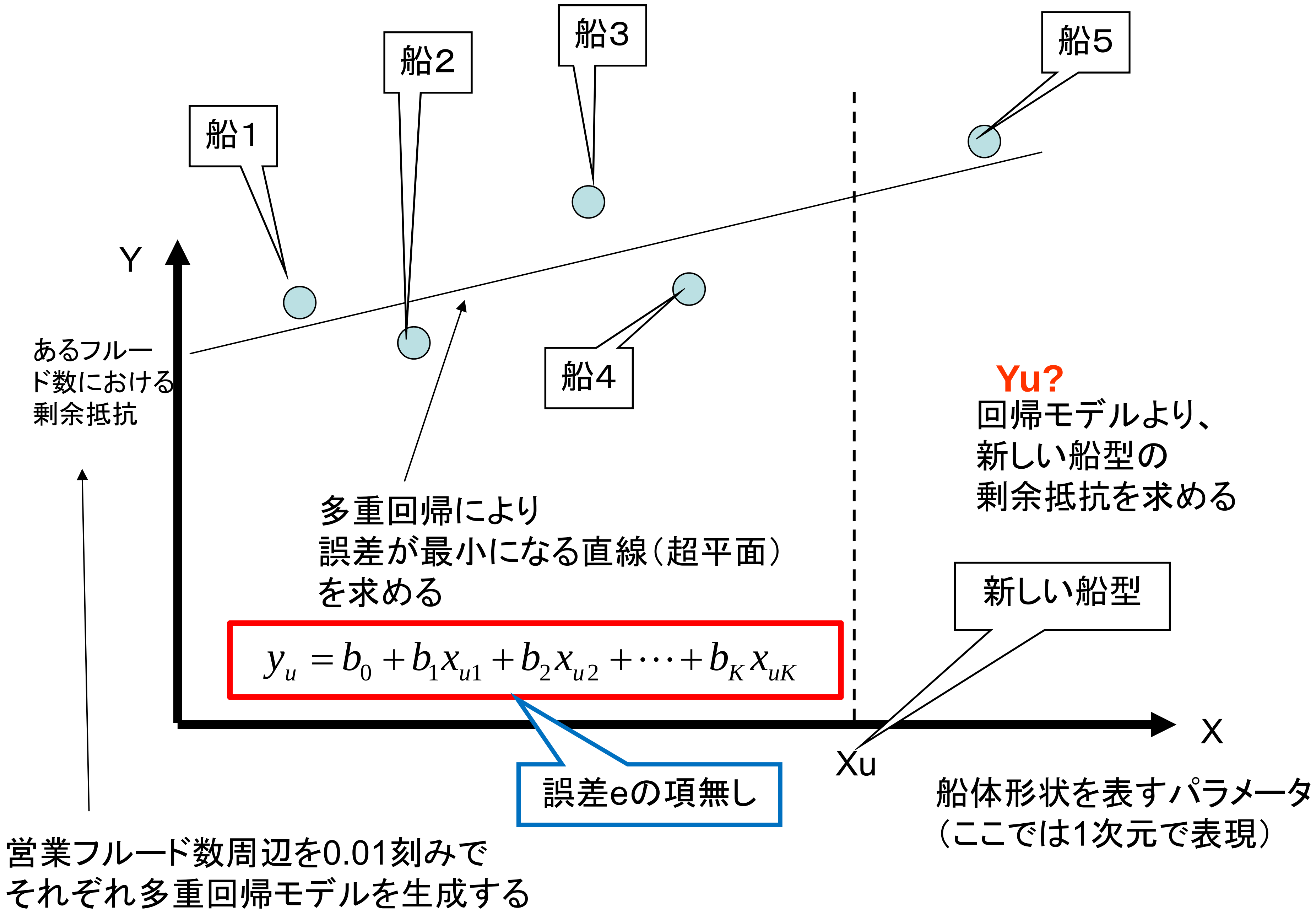
フルード数 $= \dfrac{v}{\sqrt{gL_{w\ell}}}$

60隻分の実験データから、
新しい船型についての剰余抵抗曲線を推定する

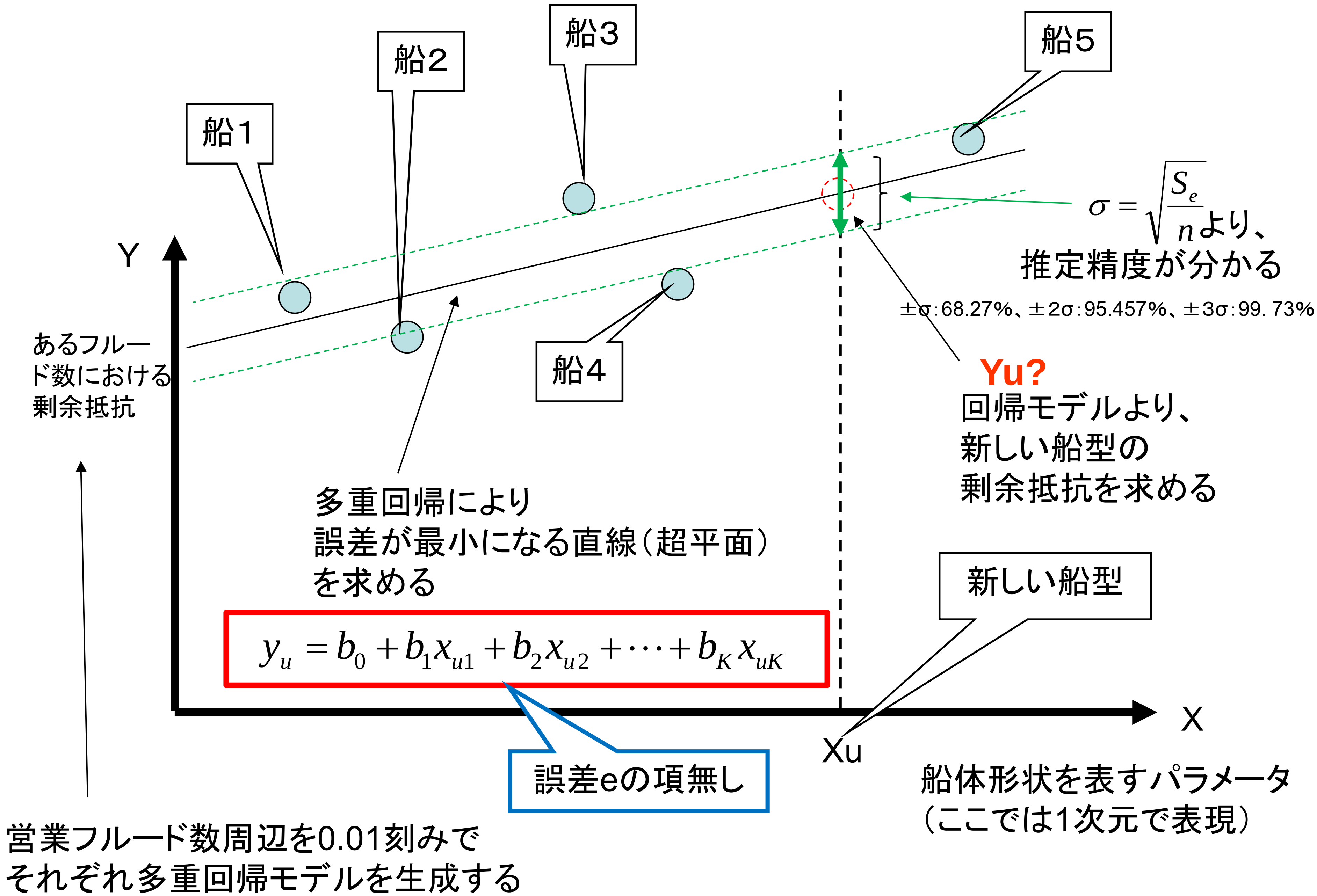
多重回帰を利用した新船型の剰余抵抗値の推定



多重回帰を利用した新船型の剰余抵抗値の推定



多重回帰を利用した新船型の剰余抵抗値の推定



痩せ型船型における推定のシミュレーション

用いた船のデータ（Lppの小さい順57隻分）
n=57

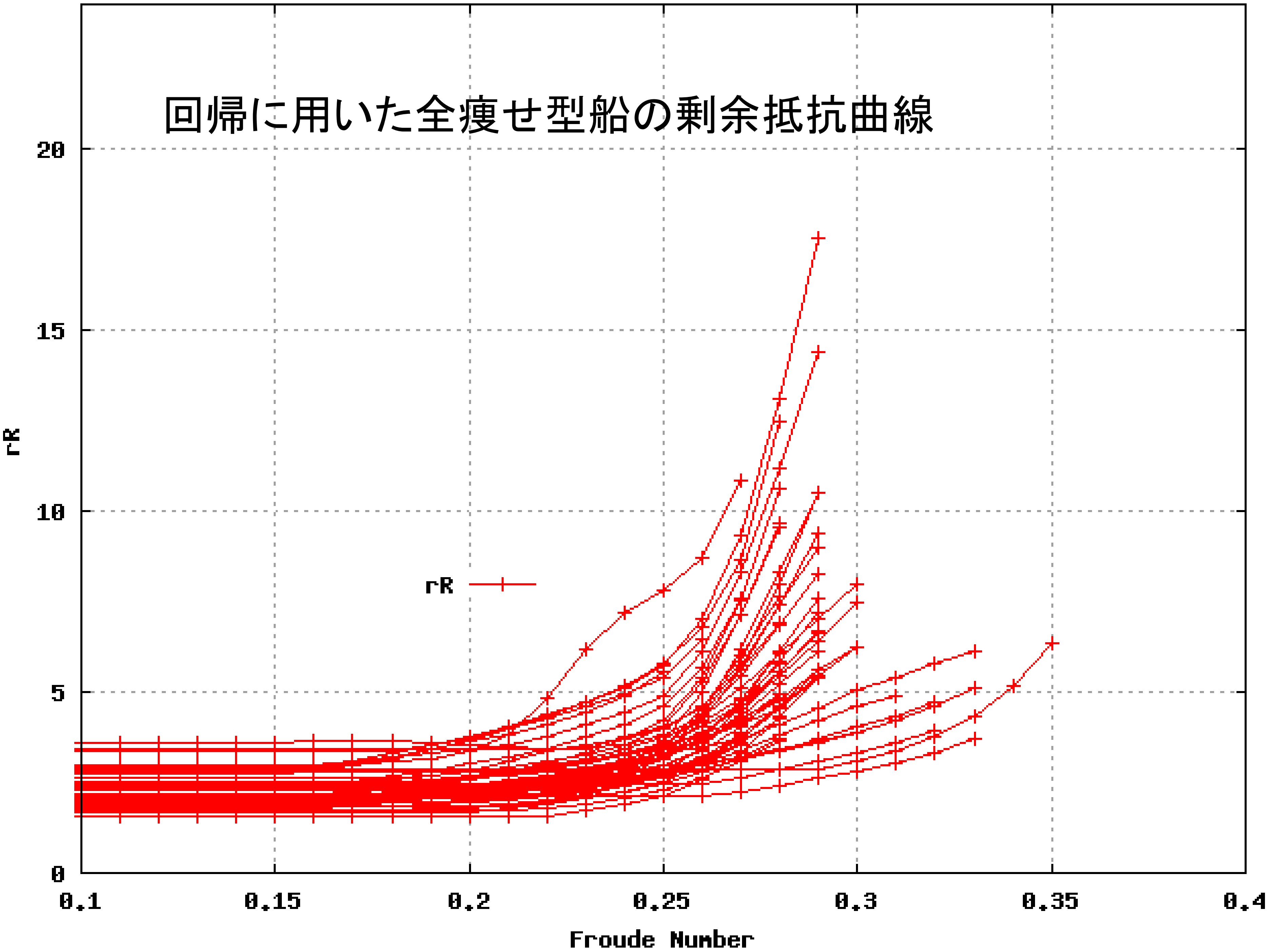
20REF	0281	S5306
0212	42RORO	S3595
0266	0076	47RORO
1200PC	0225	HSS-3
0265	0226	HSS-4
0177	0193	0181
0178	HSS-2	0194
CONT4S	0119	0207
0092	0214	0208
0269	0203	0204
0121RE	S5327S	0268
0260RE	S5502	PMAX
0175	0082	OPMX
0275	HSS-1	
0096RE	0104	
0149RE	P0262	
0179	0278	
0180	0280	
0112	0289	
0113	52PCC	
0221	S3568	
0100RE	S5020	

回帰モデルに用いた
パラメータ19個
K=19

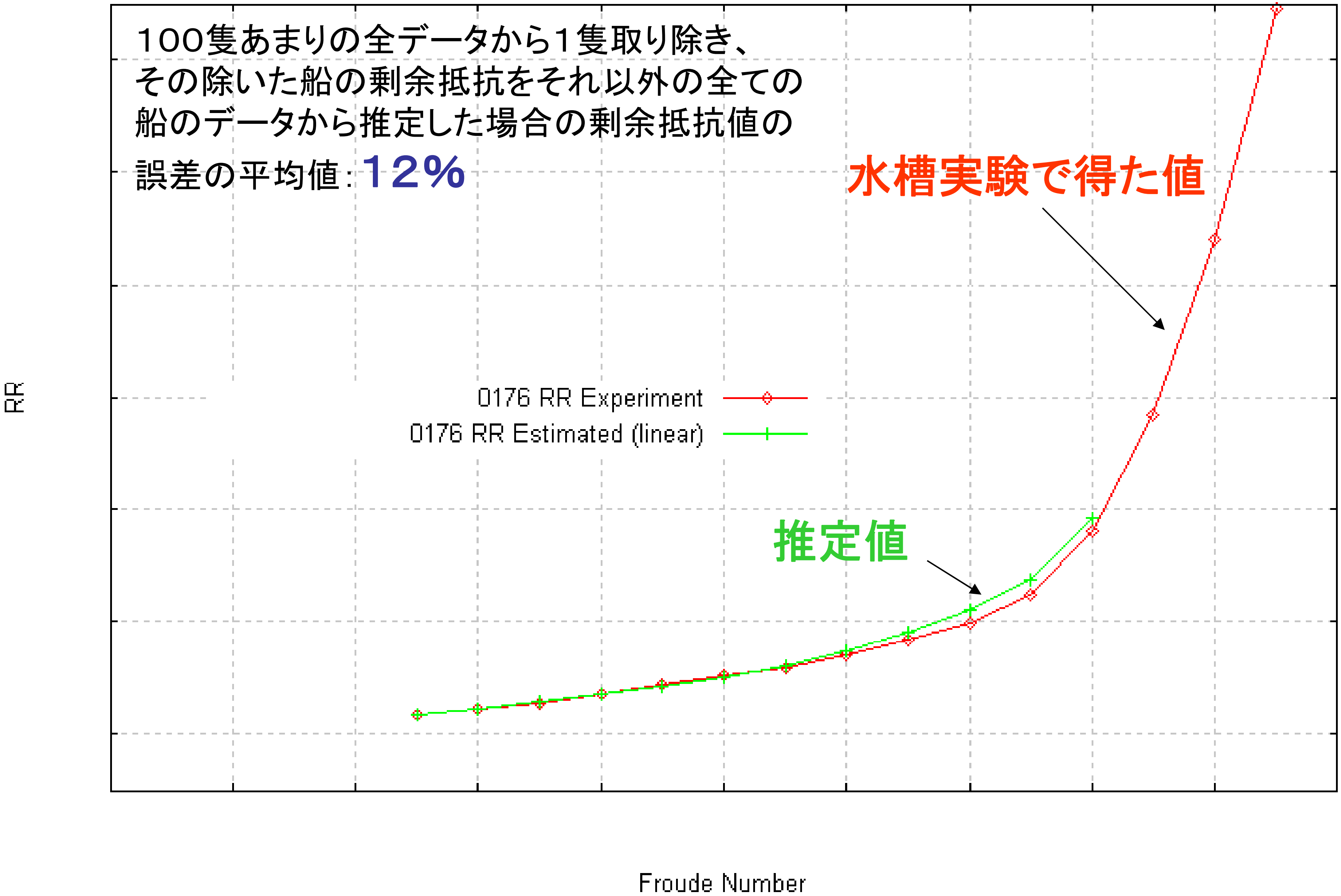
L/B
B/d
Cb
Cm
Cw
lcb
Percent_Lpp
Percent_d
BetaGamma0.3
CPLr
CPfr
CPWr
CPtr
CPar
CPLe
CPfe
CPWe
CPte
CPae

性能を推定する未知船型
0176

ServiceFn	0.234
L/B	6.28319
B/d	2.7561
Cb	0.7094
Cp	0.7203
Cm	0.9849
Cw	0.8435
Cv	0.84102
lcb	-0.4582


回帰に用いた全痩せ型船の剰余抵抗曲線
20
15
10
5
0
rR
rR
0.1
0.15
0.2
0.25
0.3
0.35
0.4
Froude Number

剰余抵抗(RR)の推定結果



推定精度を上げる…（古典的な）造波抵抗理論式の利用

造波抵抗係数

造波抵抗（kg）

ρ：　流体の密度（kg･s^2／m^4）
S：　浸水表面積（m^2）
v：　　船の速力（m／s）

$$r_W = \frac{R_W}{\frac{1}{2}\rho\, S\, v^2}$$

重力加速度

$$K_0 = \frac{g}{V^2}$$

V：　船の速力（knots）

$$\approx \frac{2C_m{}^2}{\pi}\left(\frac{1}{C_b}\frac{B}{L}\frac{B}{d}\right)^{\frac{2}{3}}\left(1-\exp(-K_0 d)\right)^2\left(B_{00} f_0 f_0 + \cdots + B_{nn} f_n f_n\right)$$

フルード数毎に回帰するので
単なる係数と見なせる

プリズマティックカーブの形状
を表す点列の値に依存

非線形関数による値の変換

データから線形多重回帰

統計解析による実用船型の造波抵抗推定法に関する研究：　多賀野　寛
関西造船協会誌　第147号pp.43—52（昭和48年3月）

造船所からの次なる要求:

いくつかの要目を固定し、他の動かしうる要目について推進抵抗が最小になる値を自動的に探すようにしてほしい
(固定される要目:L,B,D,Cm,**Cb**,%Lpp,および営業フルード数)

問題点:いくつかの要目の値は互いに関連しており勝手に動かせない

例)

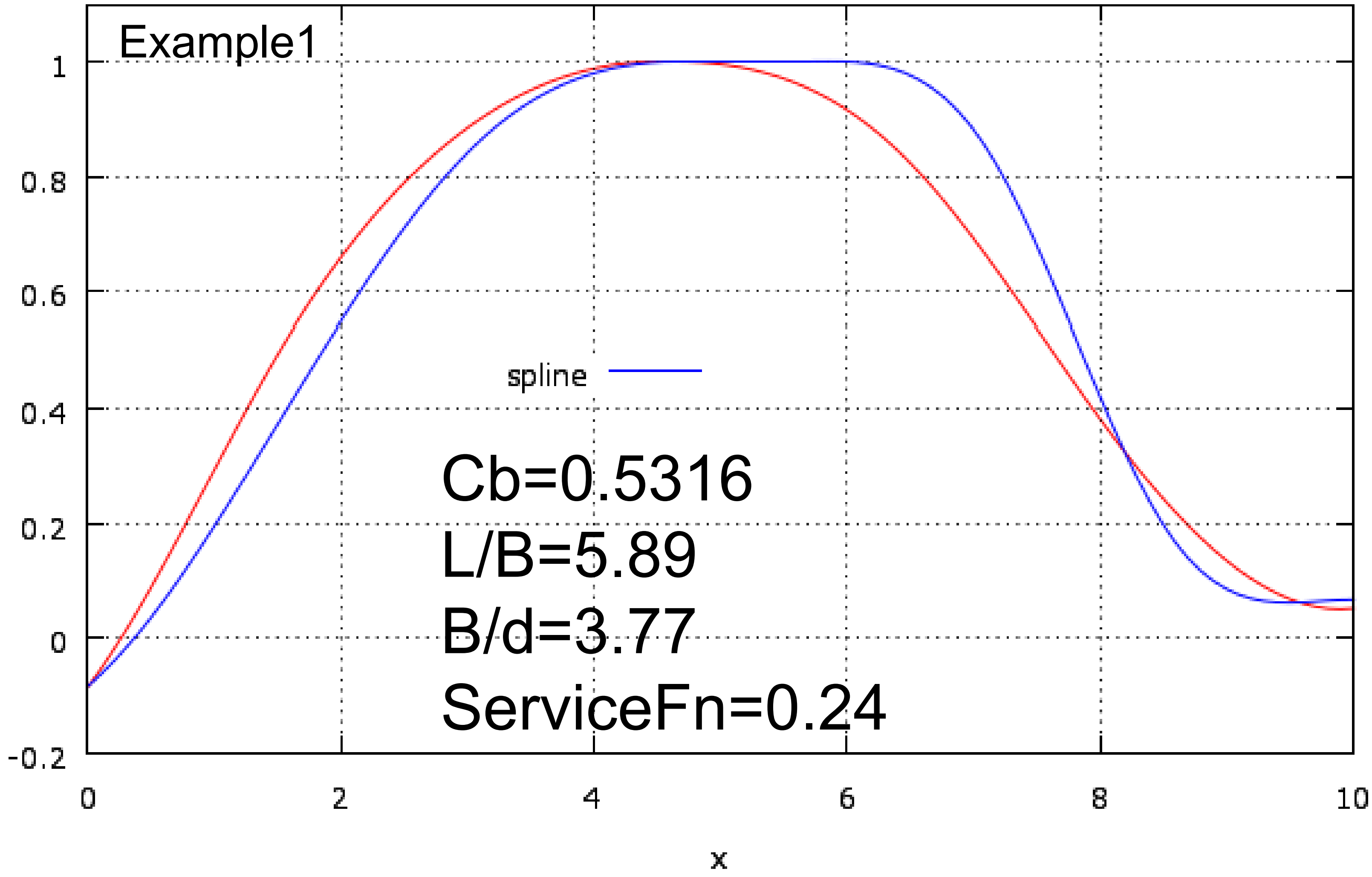
lcb はCpカーブに依存して変化する	--> Cpカーブパラメータから算出
CbもCpカーブによって計算	--> Cpカーブパラメータから算出
Cp	--> Cpカーブパラメータから算出
Cv	--> CpカーブパラメータとCwから算出

また、CPLr + CPLe < 1 という制約

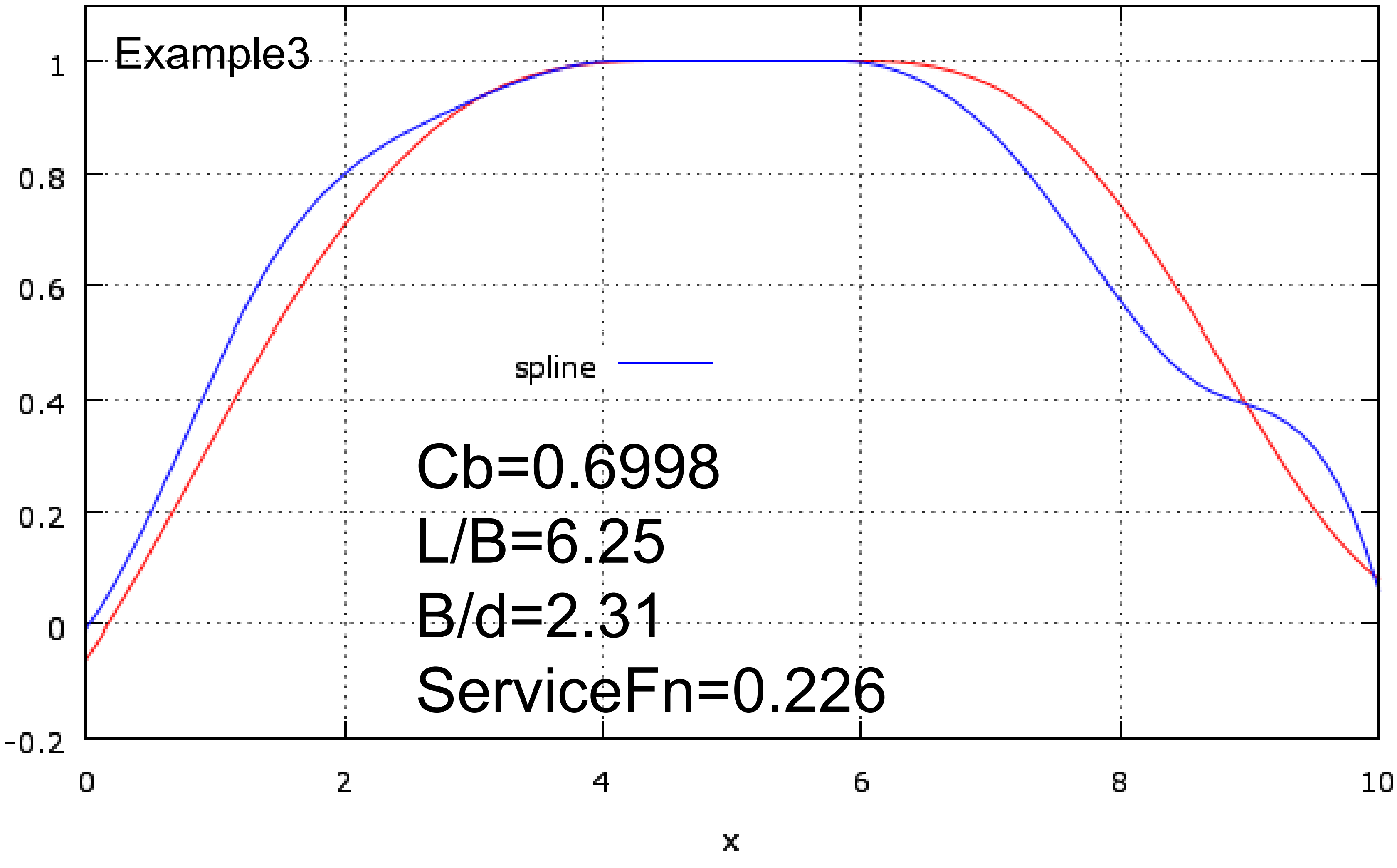
よって、Cb を与えたら、これを満たすようにCpカーブのパラメータを調整しなければならない → 与えたCb と Cpカーブで計算されるCbとの誤差をペナルティとして推進抵抗に上乗せして**非線形最適化**(詳細は次回以降説明)

しかし、力づくで最適化すると変な船形になる → 船の性能に関する専門知識が必要
固定する要目を追加するなど

Cpカーブパラメータの最適化



Cpカーブパラメータの最適化



まとめ

観測値 y を変数　$x_1, x_2, \cdots x_K$　を用いて以下の式で説明する:

$$y = b_0 + b_1 x_1 + b_2 x_2 + \cdots + b_K x_K + e$$

【多重回帰】

n 個の観測値

目的変数ベクトル $\mathbf{y} = \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{bmatrix}$　説明変数行列 $\mathbf{X} = \begin{bmatrix} 1 & \boxed{x_{11} \quad x_{12} \quad \cdots \; x_{1K}} \\ 1 & \boxed{x_{21} \quad x_{22} \quad \cdots \; x_{2K}} \\ \vdots & \vdots \quad \vdots \quad \ddots \quad \vdots \\ 1 & \boxed{x_{n1} \quad x_{n2} \quad \cdots \; x_{nK}} \end{bmatrix}$　回帰係数ベクトル $\mathbf{b} = \begin{bmatrix} b_0 \\ b_1 \\ \vdots \\ b_K \end{bmatrix}$　誤差変数ベクトル $\mathbf{e} = \begin{bmatrix} e_1 \\ e_2 \\ \vdots \\ e_n \end{bmatrix}$ と表すと、

$\mathbf{y} = \mathbf{X}\mathbf{b} + \mathbf{e}$　　誤差ベクトル $\mathbf{e}$ の平方和 $\|\mathbf{e}\|^2$ を最小にする $\hat{\mathbf{b}}$ は以下で与えられる:

$$\boxed{\hat{\mathbf{b}} = \left(\mathbf{X}^{\mathrm{Trans}}\mathbf{X}\right)^{-1}\mathbf{X}^{\mathrm{Trans}}\mathbf{y}}$$

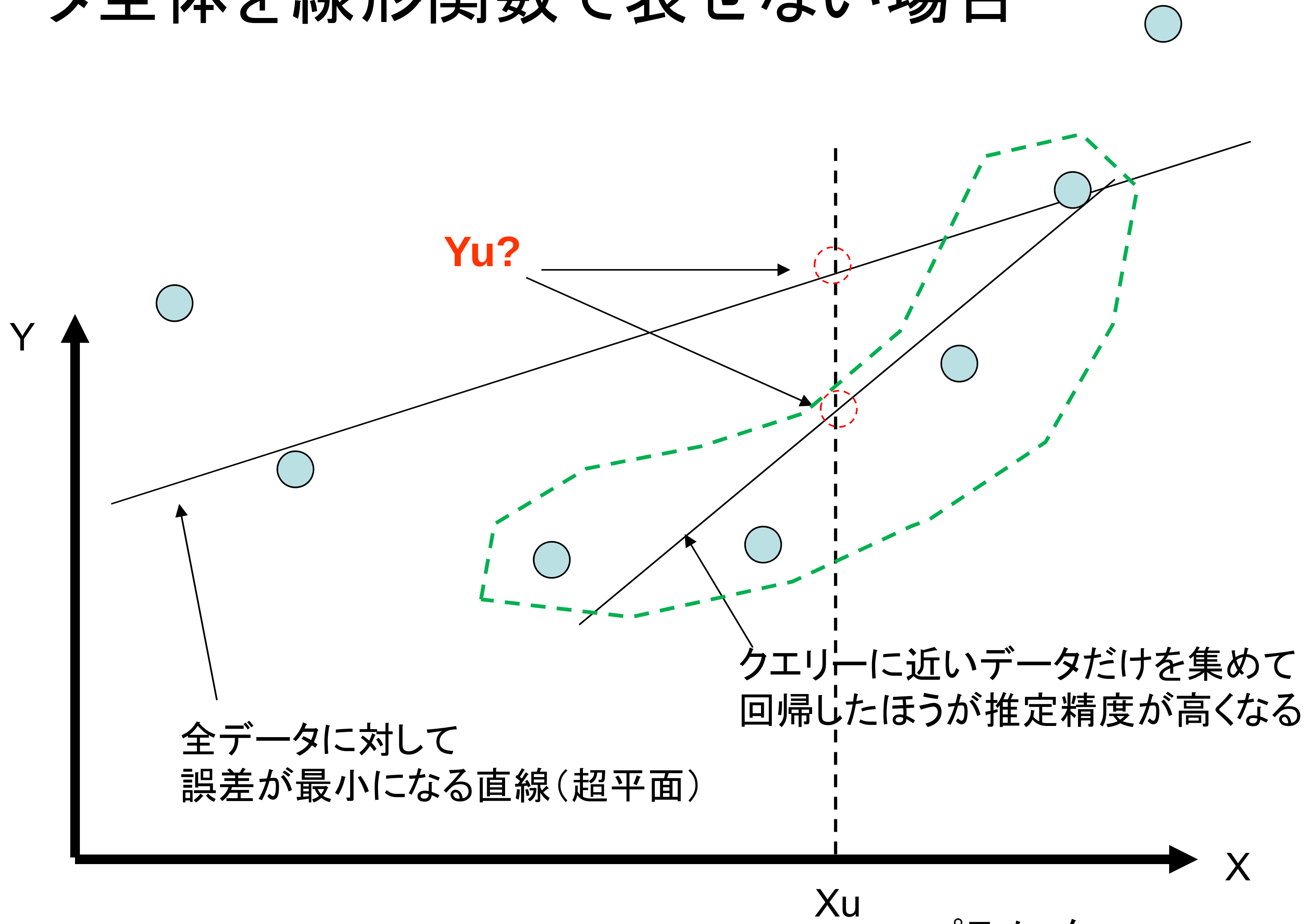
多重共線性に注意して回帰変数を選択

→　**主成分分析**＝分散・共分散行列の固有値・固有ベクトル利用

今回は**性能の推定**への応用を示したが、**工事量**や**物量**の予測などにも使用される

多重回帰はエクセルの[ツール]-[分析ツール]-[データ分析]で使用できる（詳細は検索せよ）

データ全体を線形関数で表せない場合



・データの個数が少ない場合に有効
・近似（補間）される関数が不連続になる

Locally Weighted Regression

・クエリーとデータの距離に応じて重みづけして補間する方法

$$\mathbf{z}_i = w_i \mathbf{x}_i$$

$$\mathbf{z} = \mathbf{w}\mathbf{x}$$

ただし

$$w_i = \sqrt{k\left(d\left(\mathbf{X}_i, \mathbf{q}\right)\right)}$$

カーネル関数

距離関数

i番目のデータの座標

クエリの座標

$$v_i = w_i y_i$$

$$\mathbf{v} = \mathbf{w}\mathbf{y}$$

とおき、クエリの関数値を $\hat{y}(\mathbf{q}) = \mathbf{q}^T \beta\, \mathbf{v}$ と推定するとき、

重み付き誤差関数 $C(\mathbf{q}) = \sum_{i=1}^{n} \left[\left(\mathbf{x}_i^T \beta - y_i \right)^2 K\left(d\left(\mathbf{x}_i, \mathbf{q}\right)\right) \right]$

を最小化するβは

$$\left(\mathbf{z}^T \mathbf{z}\right)\beta = \mathbf{z}^T \mathbf{v}$$

$$\beta = \left(\mathbf{z}^T \mathbf{z}\right)^{-1} \mathbf{z}^T \mathbf{v}$$ で与えられる。

・データの個数が多い場合に有効
・近似（補間）される関数が連続
・距離関数の設定が難しい

ロバスト推定法

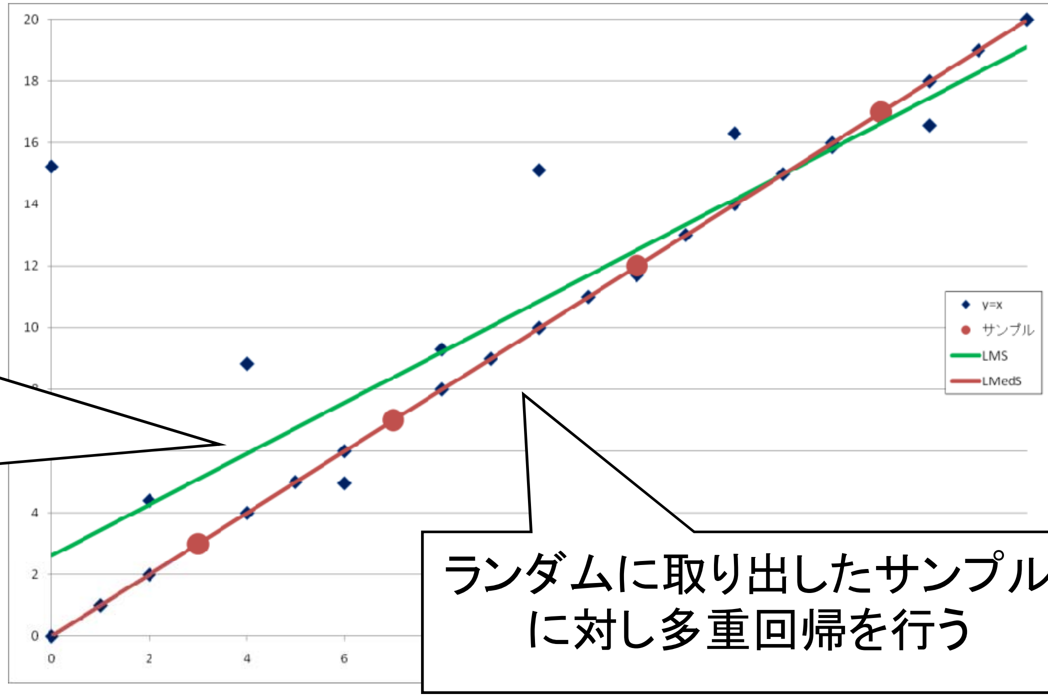
・データに例外値（ノイズ）を多く含む場合、これらを自動的に除去した回帰を行う

最小メジアン法(LMedS：Least Median of Squares)

- ランダムに幾つかのサンプルを抽出し、最小二乗法に当てはめることを繰り返す
- LMedS基準　$LMedS = \min \operatorname{med} \varepsilon_i^2$
- 全データの半分が外れ値でも大きな影響を及ぼさない

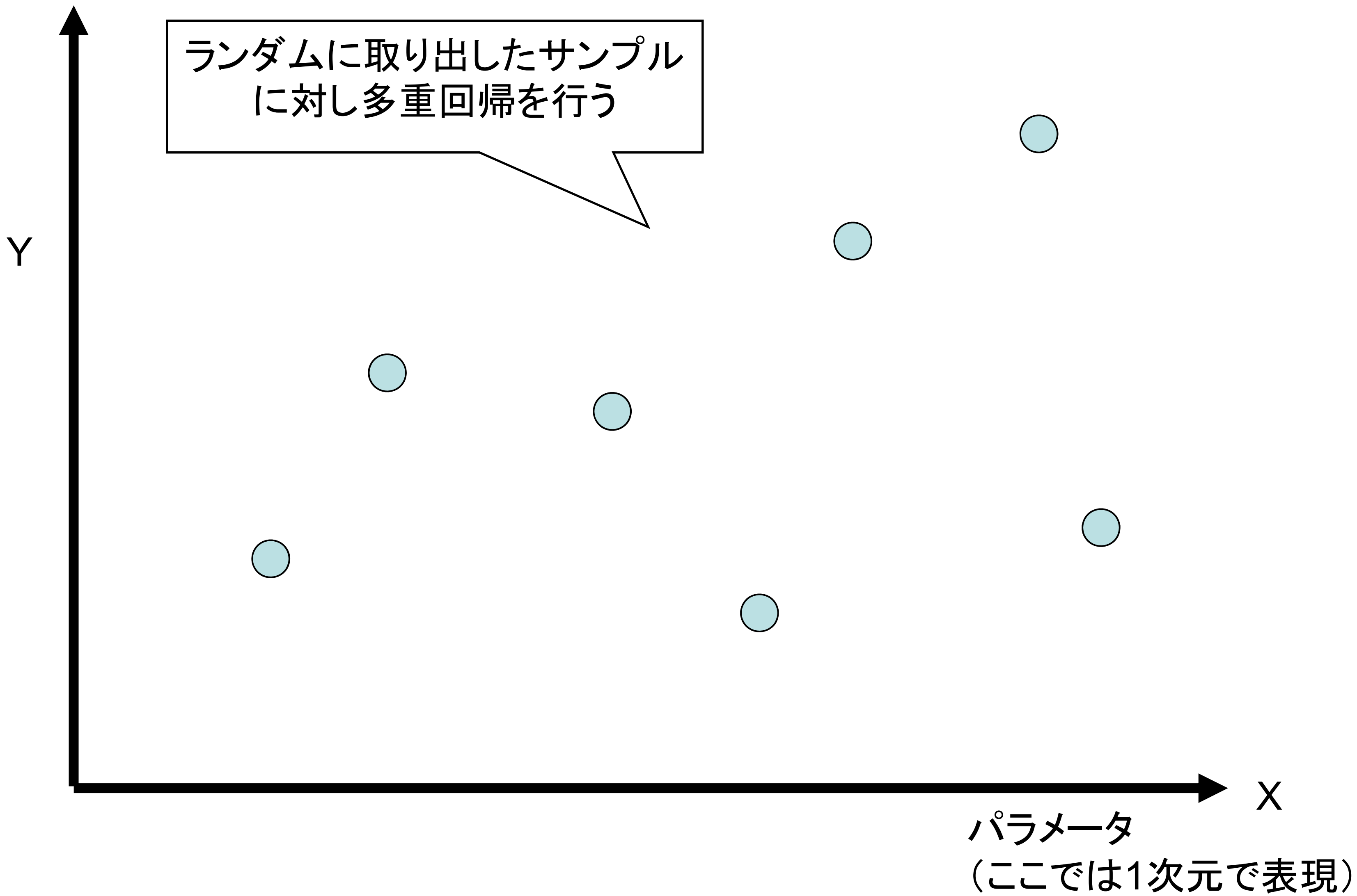
回帰された各直線と全データとの間の誤差の2乗を計算し、昇順に並べ、メディアンになるデータの2乗誤差値を読取る

この値が最小のモデルが最良

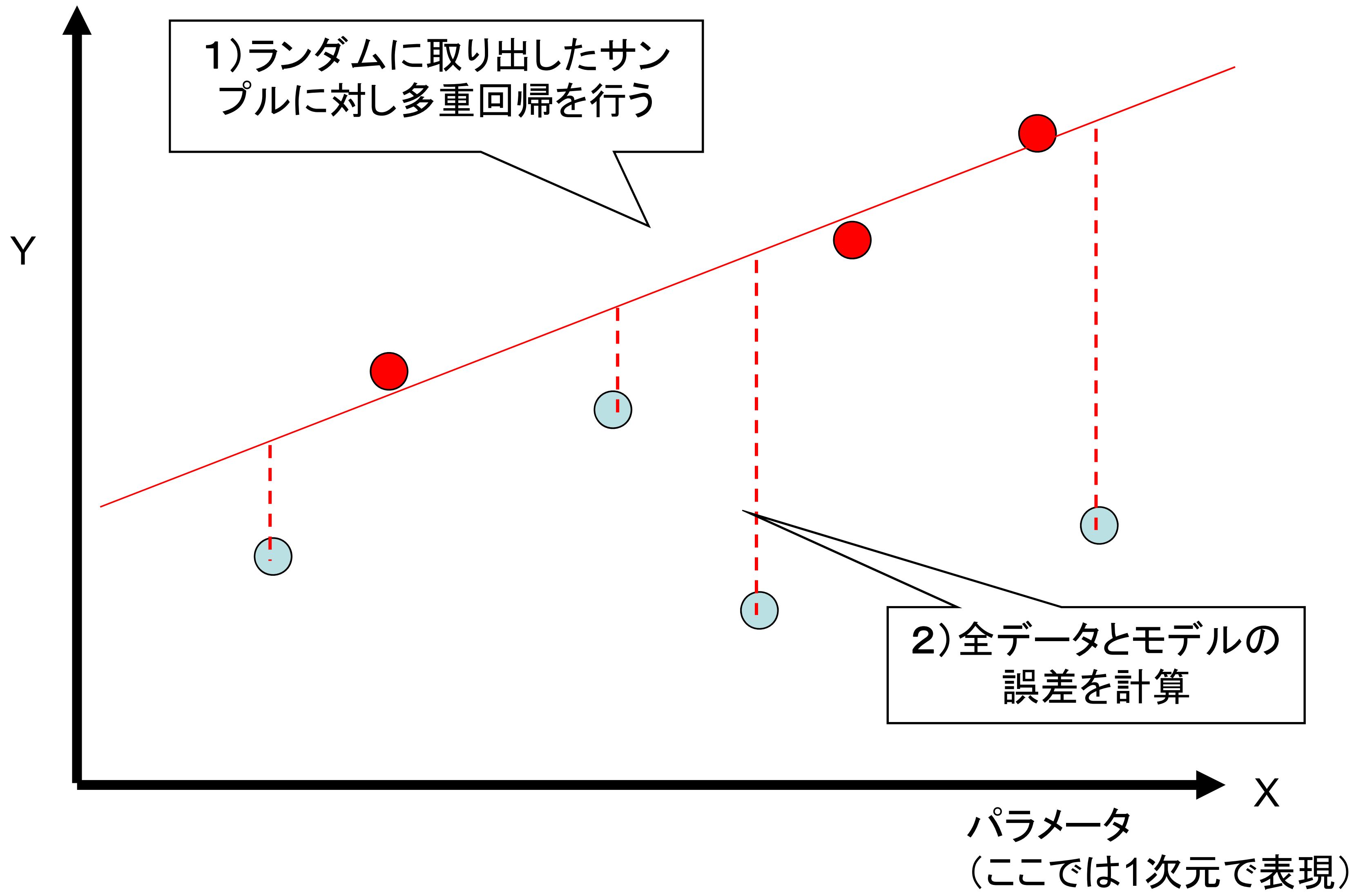


ランダムに取り出したサンプルに対し多重回帰を行う

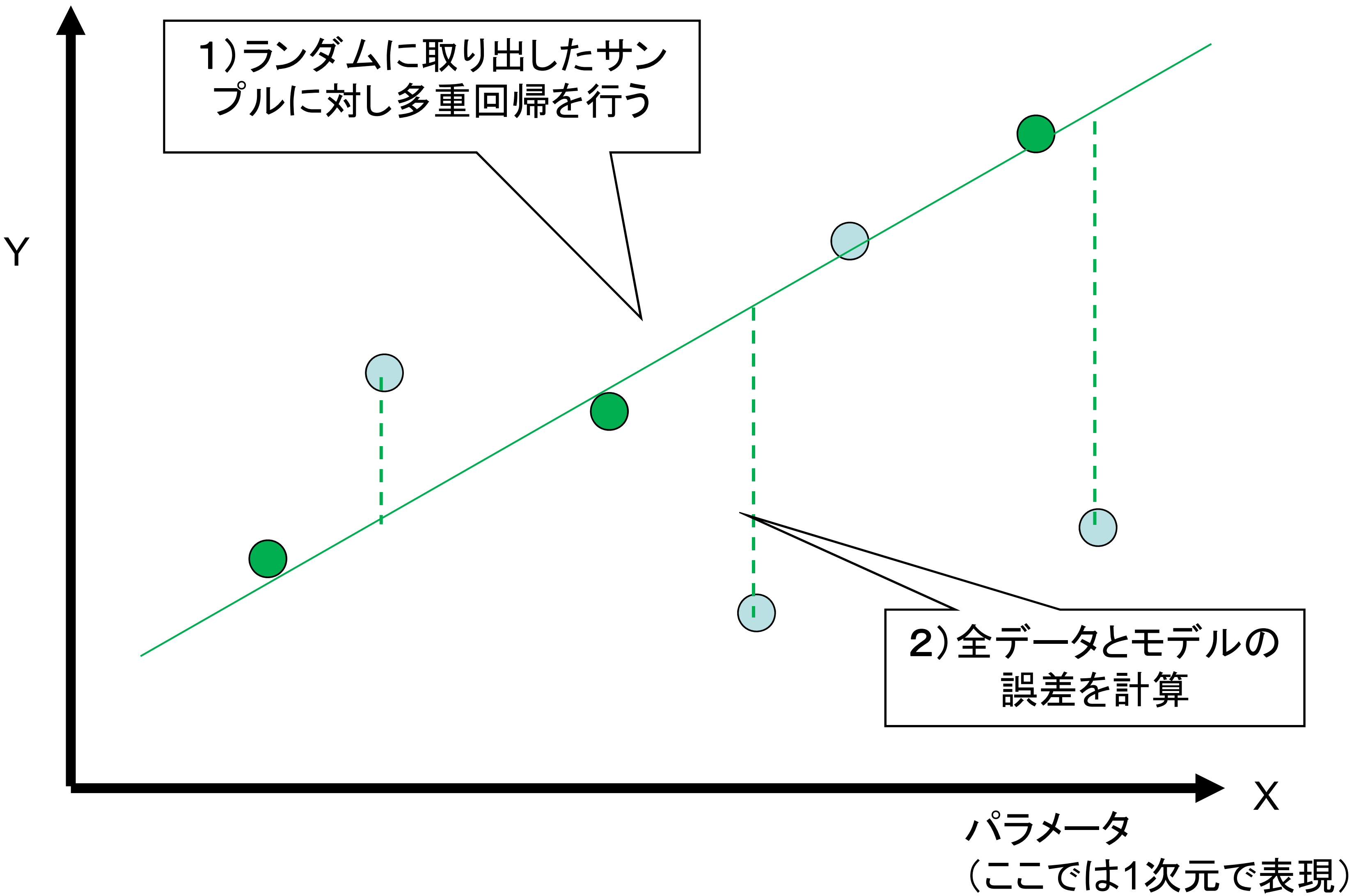
最小メジアン法におけるノイズ処理



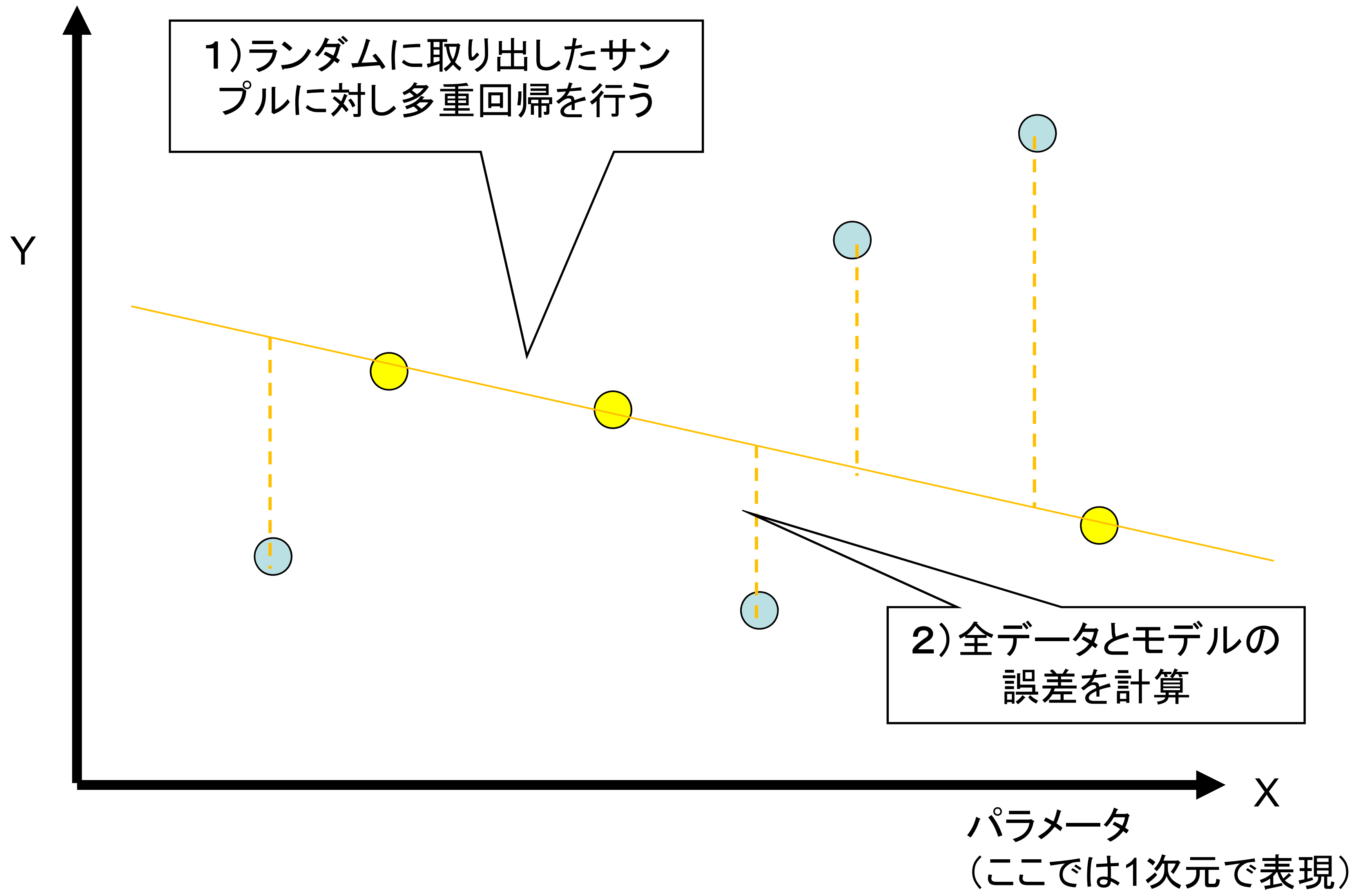
最小メジアン法におけるノイズ処理



最小メジアン法におけるノイズ処理



最小メジアン法におけるノイズ処理



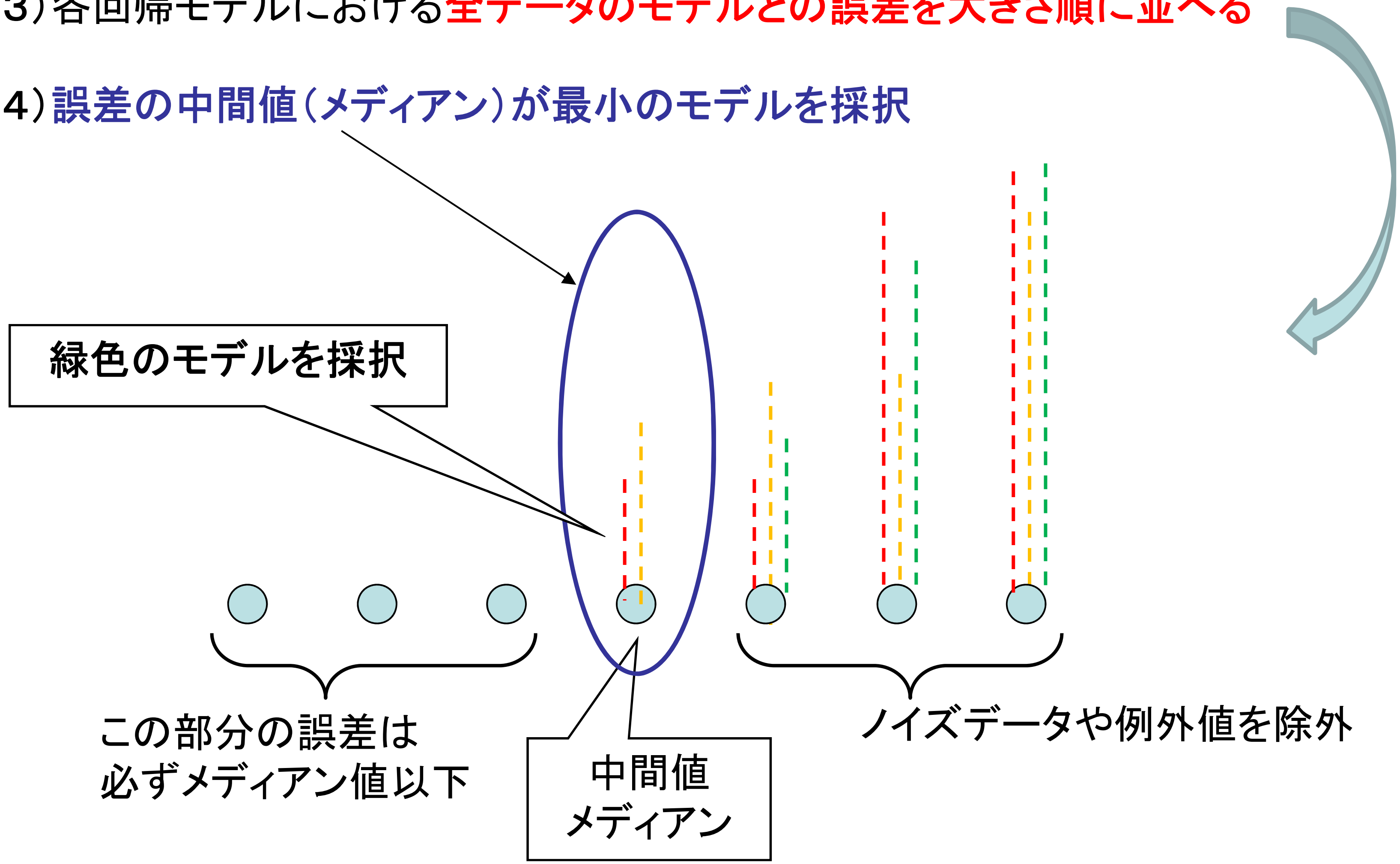
最小メジアン法におけるノイズ処理

1)ランダムに取り出したサンプルに対し多重回帰を行う

2)各回帰モデルにおいて全データとモデルの誤差を計算

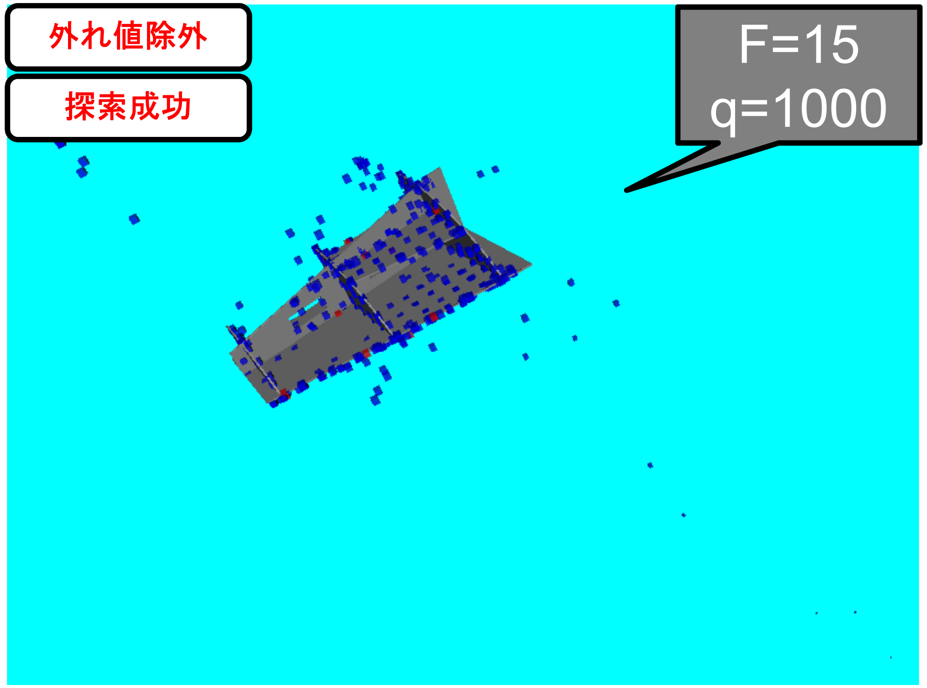
3)各回帰モデルにおける全データのモデルとの誤差を大きさ順に並べる

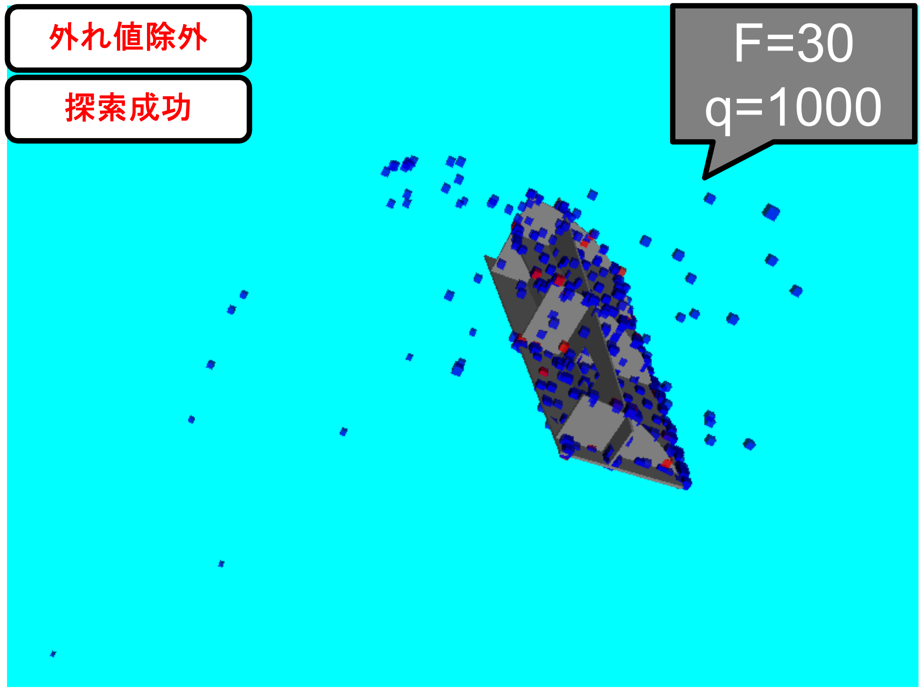
4)誤差の中間値(メディアン)が最小のモデルを採択



外れ値を含む特徴点データと3Dモデルとのマッチング

模型2b(特徴点数:1393)

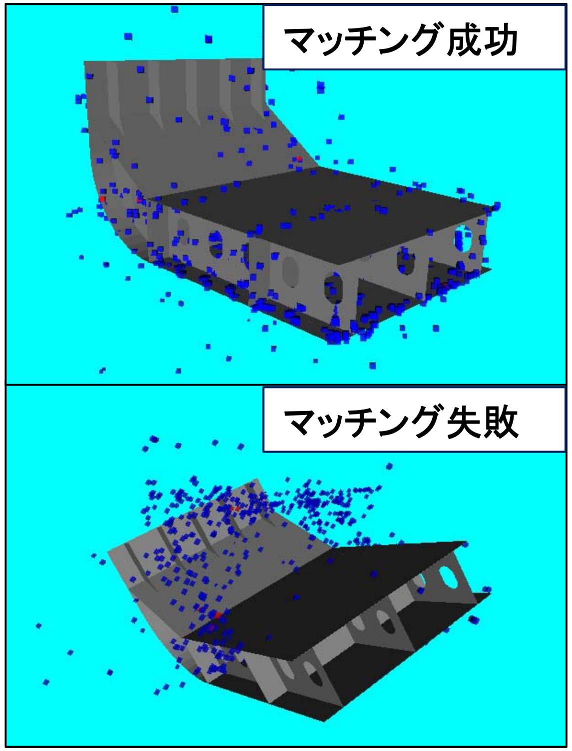


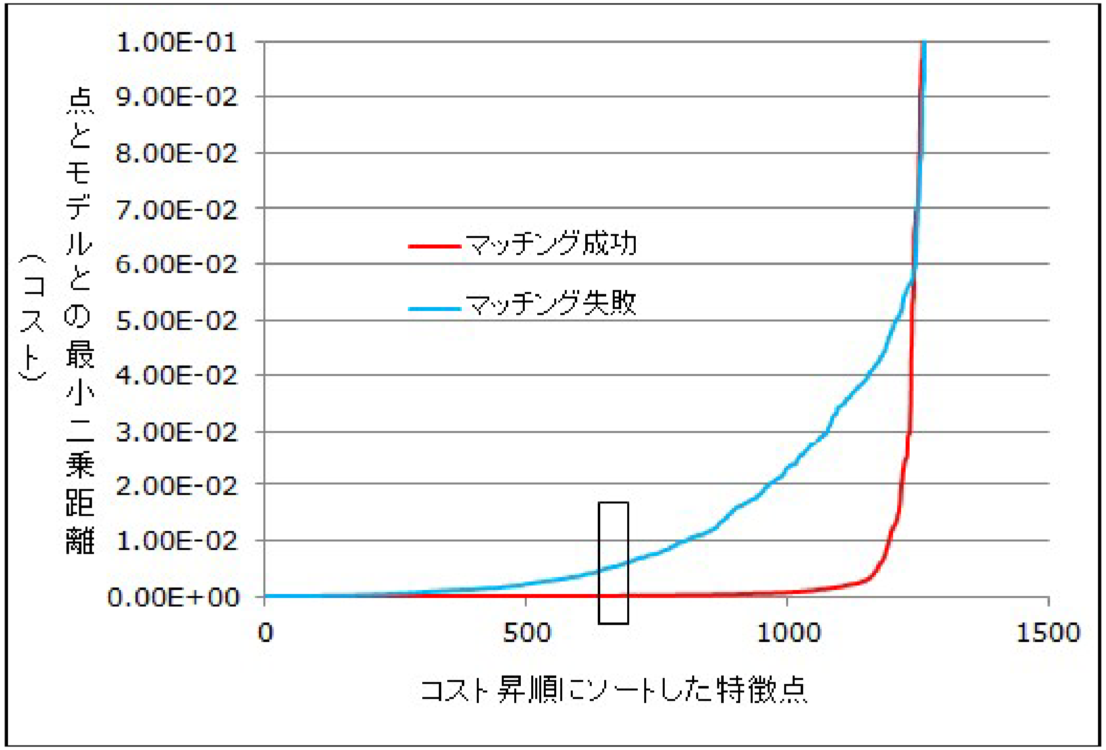


結果

- 外れ値を良く除外
- 全データ数に対し外れ値の割合が少ない

外れ値を含む特徴点データと3Dモデルとのマッチング





最小メジアン法を用いたマッチング

1.無作為にF個の点を選び、マッチング処理を実行。

2.全ての点のコストを昇順にソート。

3.1と2を一定回数繰り返しメジアン（中間値）が最小なものを選択。

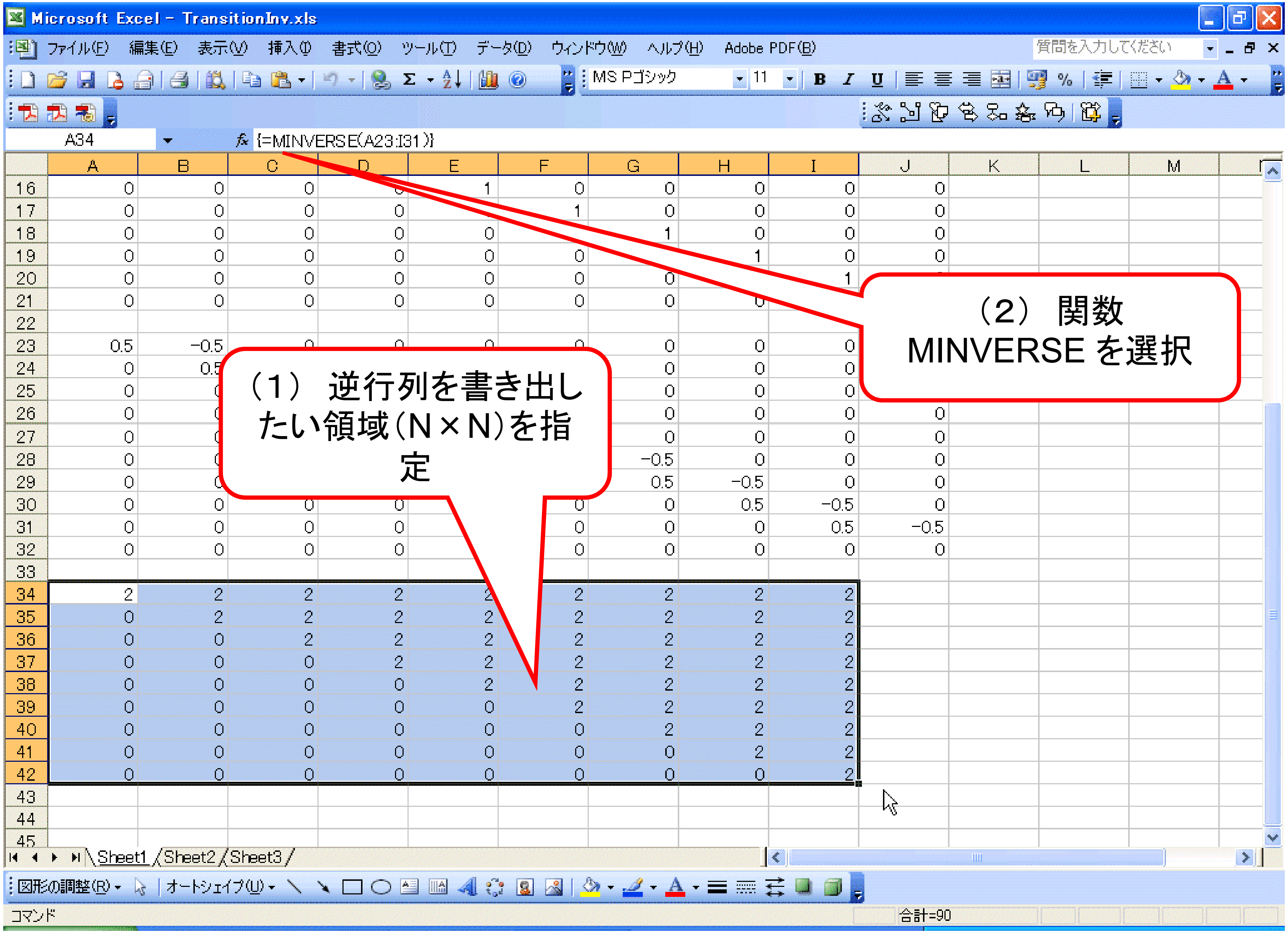
【レポート課題】

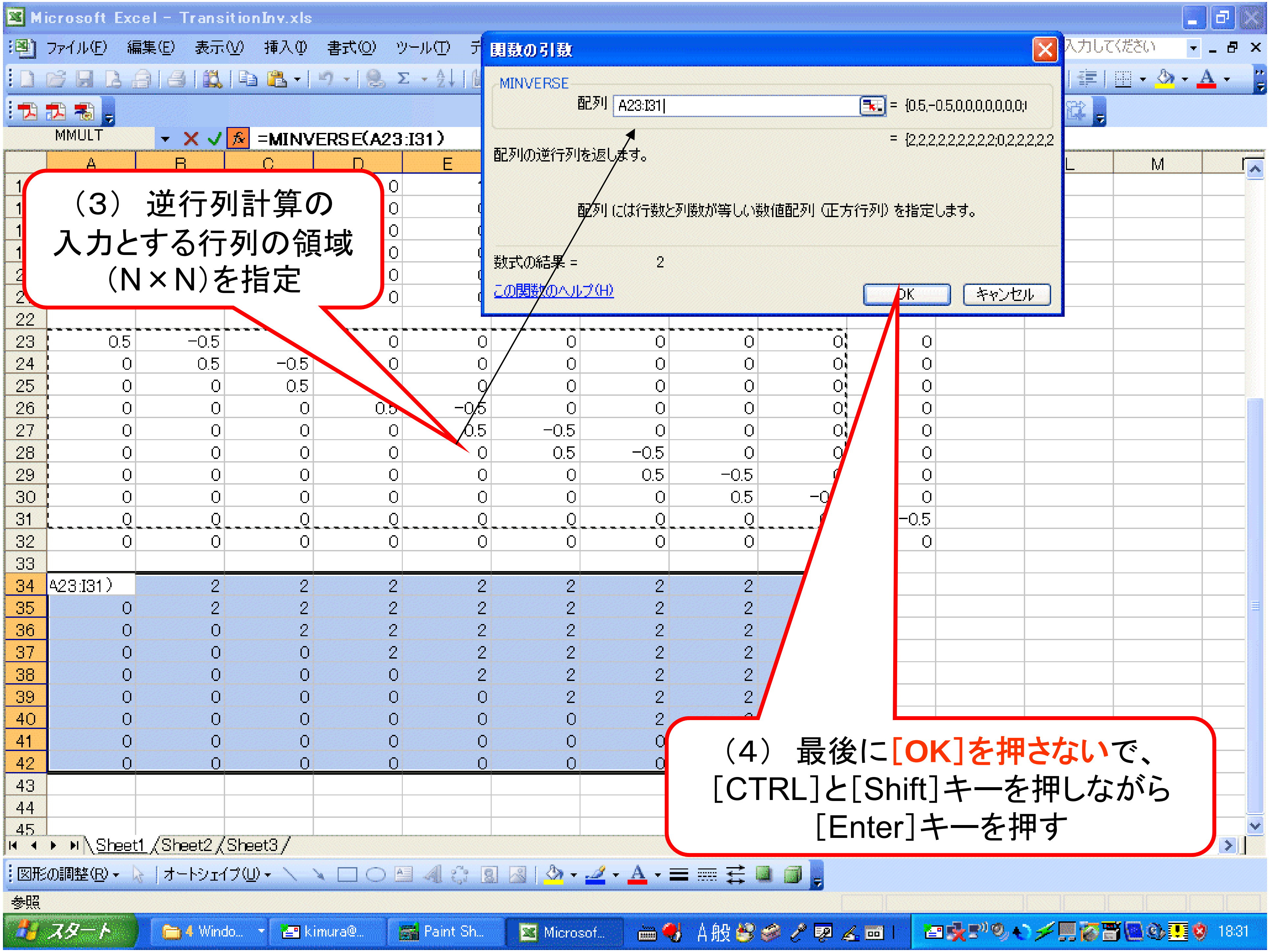
自動車の燃費について以下のようなデータがある。
別紙に指示するように各自の担当車種のデータ以外のデータを用いて多重回帰を行い，
その回帰結果から担当車種の燃費yを推定せよ。このとき推定値と実データとの誤差を評価せよ。
提出期限：10月20日講義開始前まで

車名	x1	x2	x3	X4	x5	x6	y
クラウン	1.360	4.778	2.4251	125	17.5	8.8	8.7
マークII	1.245	4.100	2.4082	125	17.5	8.8	9.5
カムリ	1.070	3.214	2.3575	120	17.6	8.7	10.6
ソアラ	1.235	4.100	2.3052	125	17.5	8.8	9.2
セドリック	1.420	4.625	2.4251	130	17.5	9.5	8.9
ローレル	1.175	3.889	2.3660	125	17.0	9.1	9.2
スカイライン	1.175	4.111	2.3198	125	17.0	9.1	9.2
レパード	1.220	3.900	2.2899	125	17.0	9.1	9.4
カペラ	1.030	3.450	2.3829	120	17.0	8.6	10.2
ギャラン∑	1.180	3.665	2.3645	110	16.7	8.5	10.6
ルーチェ	1.150	3.909	2.3829	120	17.0	8.6	9.8

x1:車体重量(1000kg)，　x2：減速比，　x3：幅×高さ(m^2)，　x4：最大出力(ps)，
x5：最大トルク(kgm)，　x6：圧縮比，　y：10モード走行(km/ℓ)

【参考】 Excel2003で逆行列を求める方法




Microsoft Excel - TransitionInv.xls
関数の引数
MINVERSE
配列 A23:I31
=MINVERSE(A23:I31)
配列の逆行列を返します。
配列 には行数と列数が等しい数値配列 (正方行列) を指定します。
数式の結果 = 2
この関数のヘルプ(H)
OK
キャンセル
（3） 逆行列計算の
入力とする行列の領域
（N×N）を指定
（4） 最後に[OK]を押さないで、
[CTRL]と[Shift]キーを押しながら
[Enter]キーを押す