九州大学 工学部地球環境工学科
船舶海洋システム工学コース

海事統計学 （担当：木村）

(2) 相関／回帰

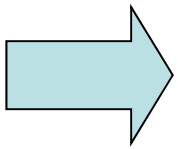
場所： 船2講義室

授業の資料等は

http://sysplan.nams.kyushu-u.ac.jp/gen/index.html

回帰分析・相関分析

「身長」と「体重」
「数学」と「英語」の点数　など

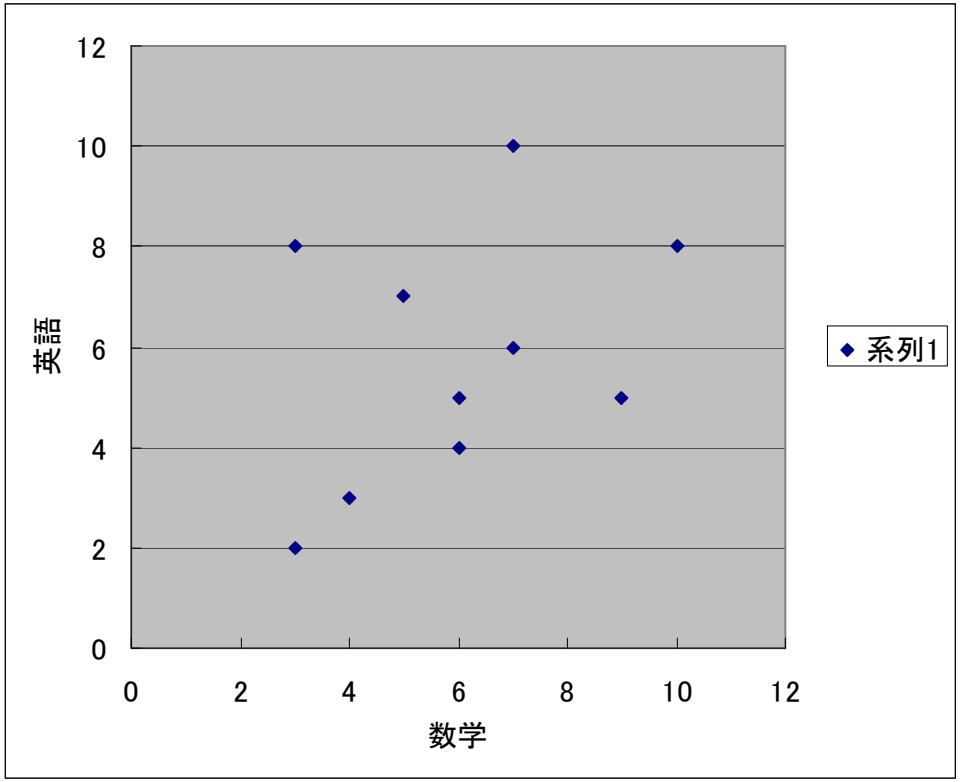


２変量についての関係を調べる

２変量データの例）

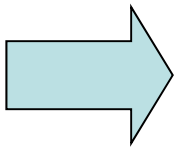
数学 x	英語 y
3	8
6	4
10	8
4	3
7	6
7	10
3	2
9	5
6	5
5	7

２変量データの表現方法：



回帰分析・相関分析

「身長」と「体重」
「数学」と「英語」の点数　など

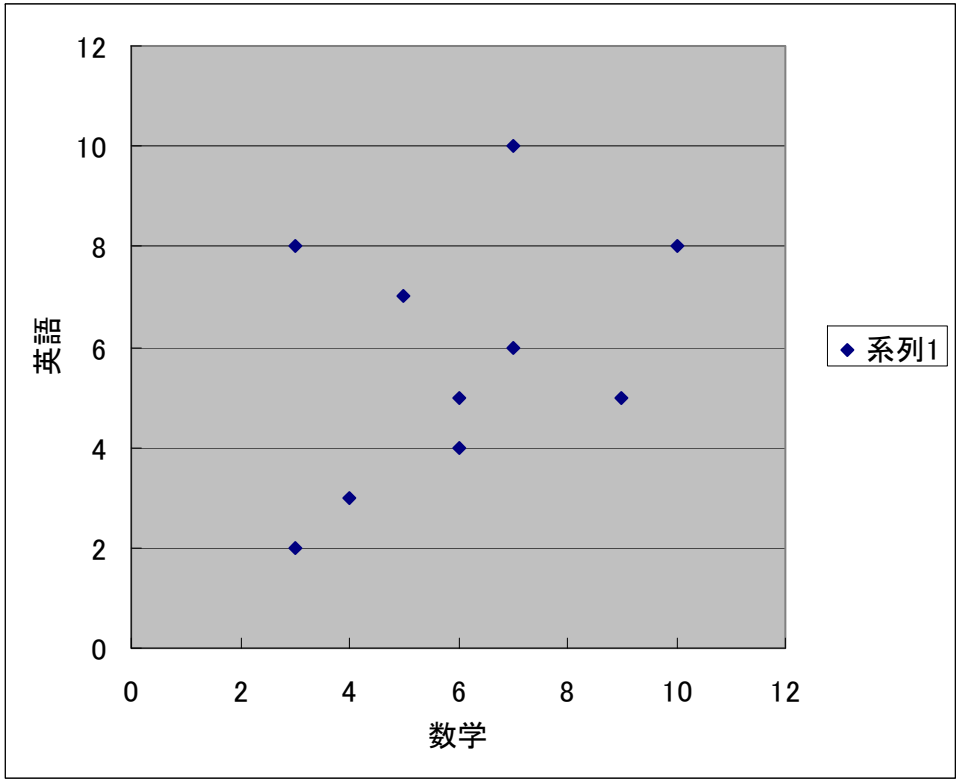


２変量についての関係を調べる

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**



回帰（単純回帰）

２変量の関係として**直線をあてはめる**

$y = ax + b$

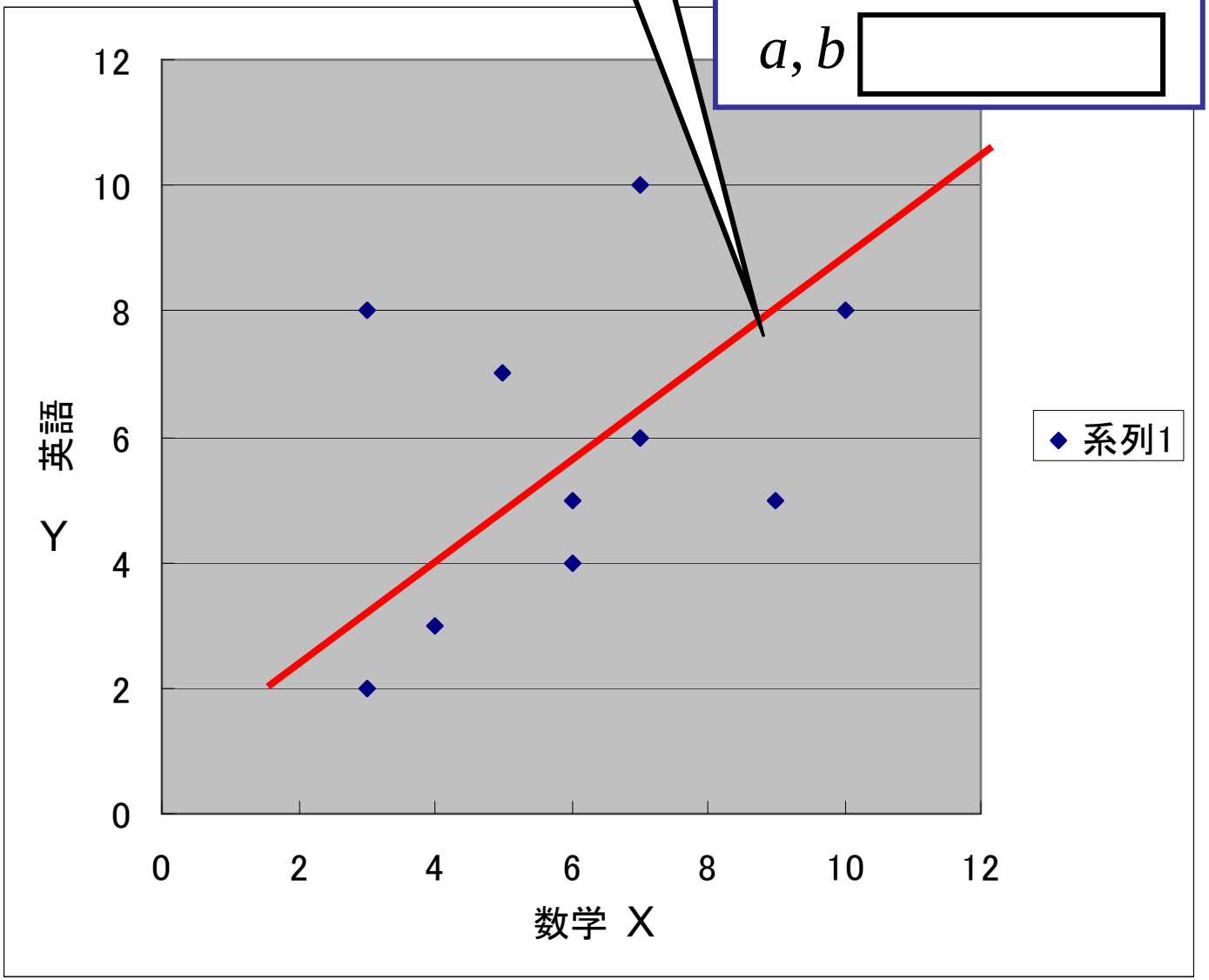
x

y

a, b

回帰直線

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



回帰（単純回帰）

２変量の関係として**直線をあてはめる**

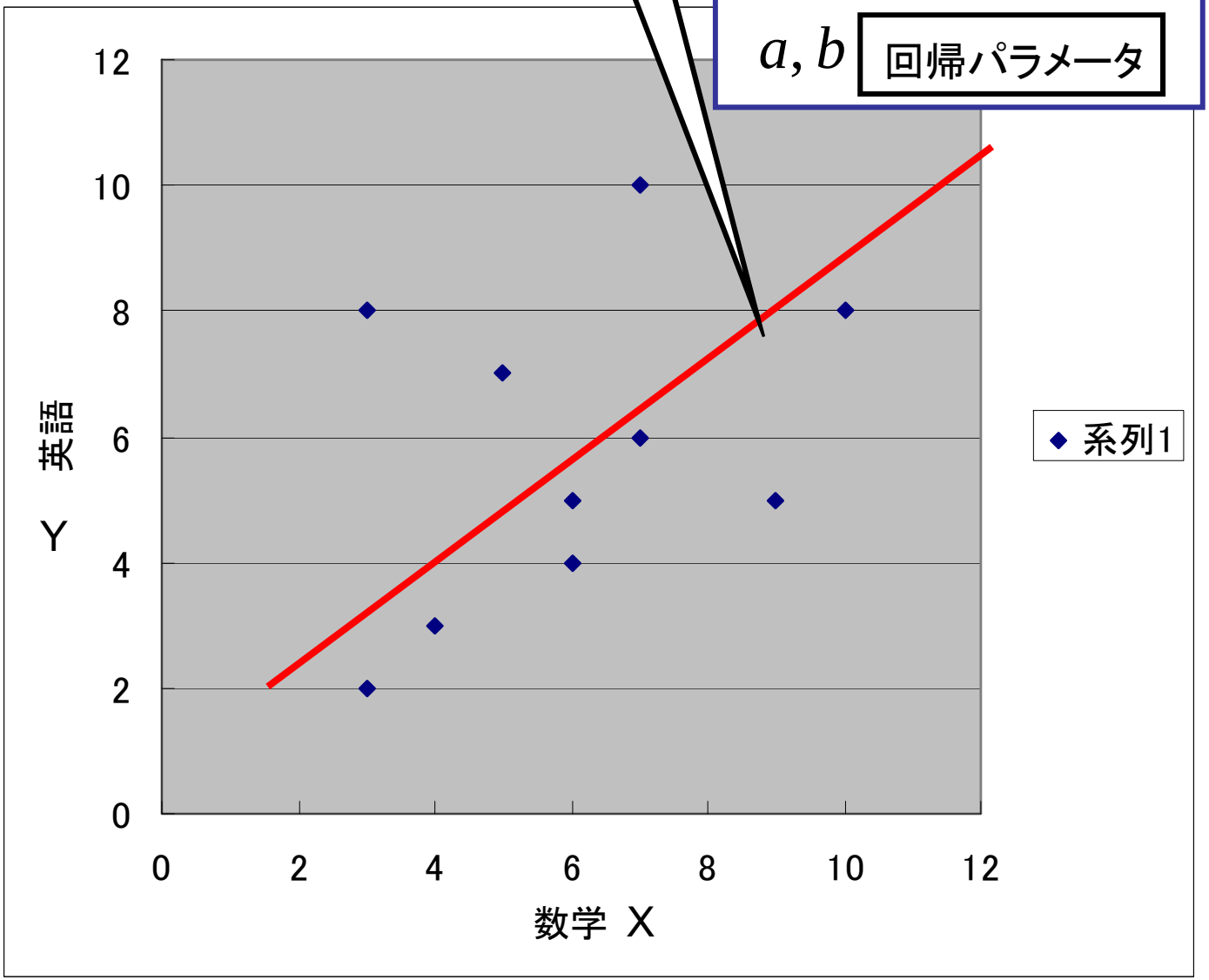
$y = ax + b$

x 回帰変数

y 被回帰変数

a, b 回帰パラメータ

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



回帰（単純回帰）

回帰直線
$y = ax + b$
y の x に対する回帰直線

回帰パラメータa,bの求め方

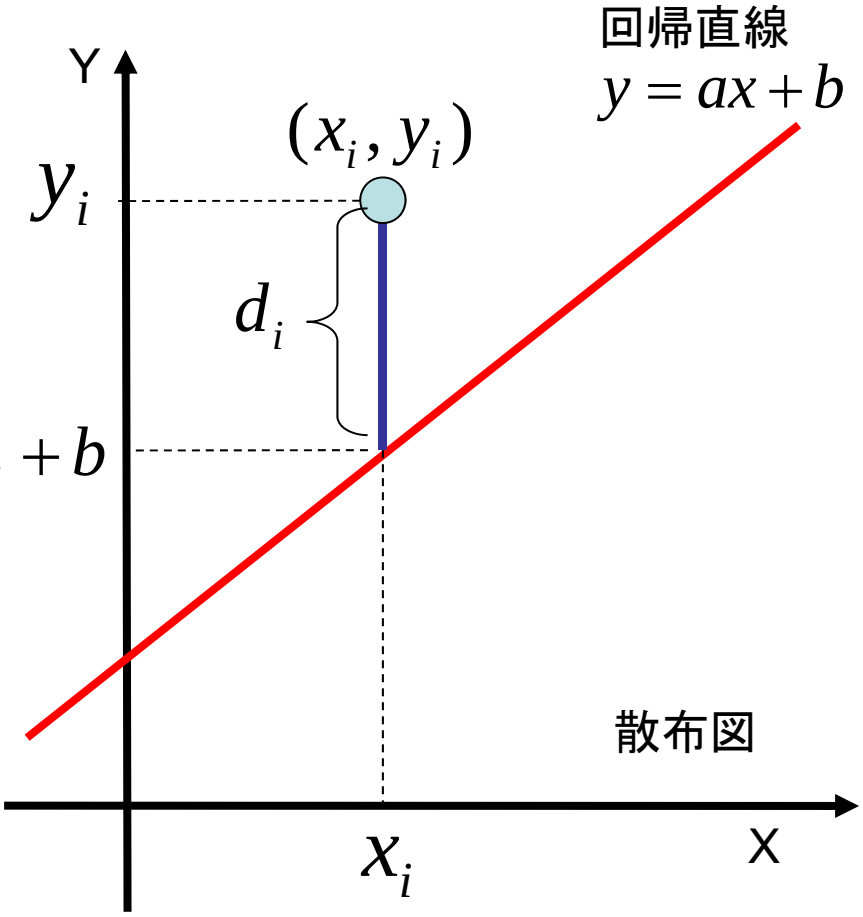
n 個のデータの組を $(x_1, y_1), (x_2, y_2), \cdots, (x_n, y_n)$ と表す

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**）

$$L = \sum_{i=1}^{n} d_i^{\,2} =$$



回帰（単純回帰）

回帰直線
$y = ax + b$
y の x に対する回帰直線

回帰パラメータa,bの求め方

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、

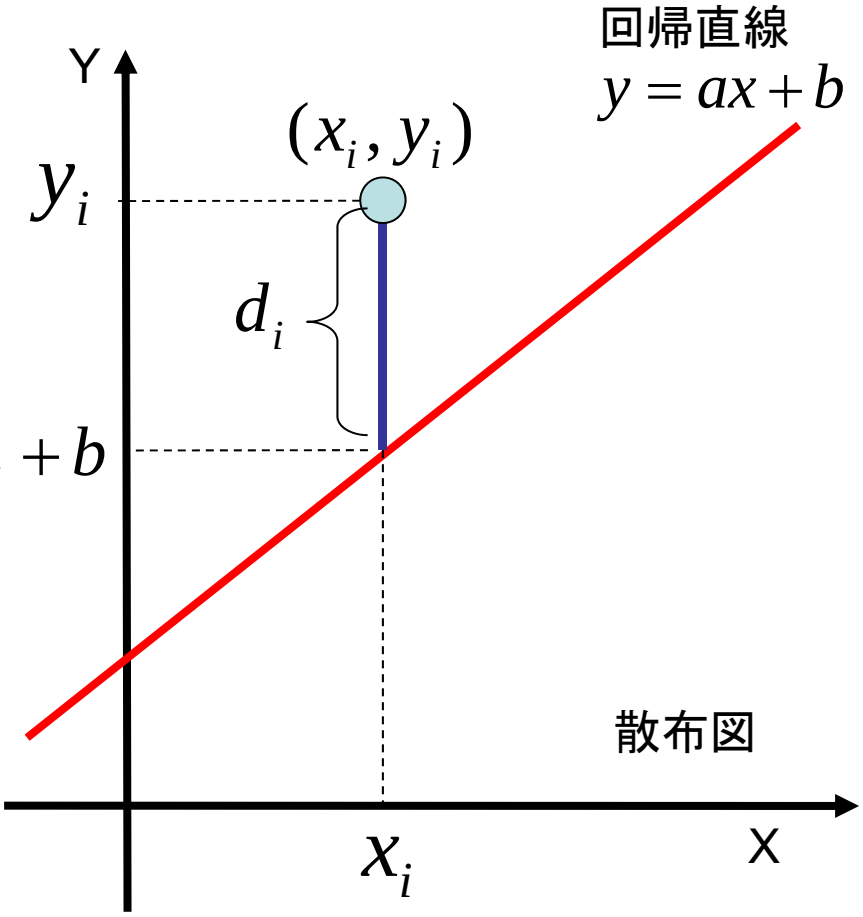
この長さの2乗和 L を最小にするように

回帰パラメータ a, b を決める
（**最小2乗法**）

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

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

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、

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

$$b = \bar{y} - a\bar{x}$$

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

$S_{xy} =$

共分散
Covariance

とおくと、$a = \frac{S_{xy}}{S_{xx}}$

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

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、

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

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

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

共分散 Covariance

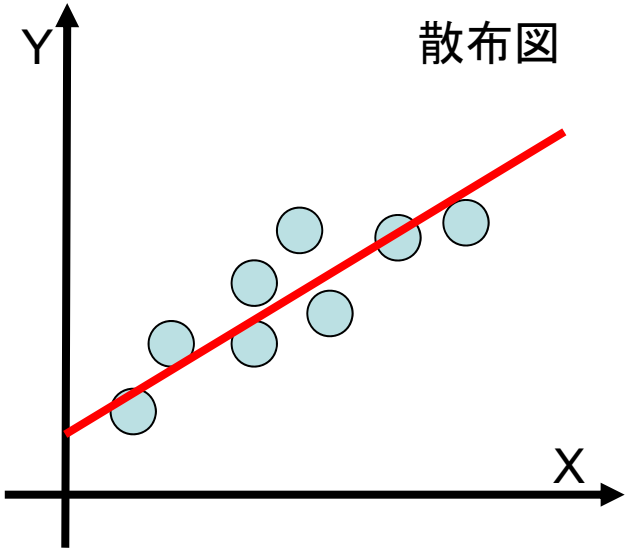
とおくと、 $a = \frac{S_{xy}}{S_{xx}}$

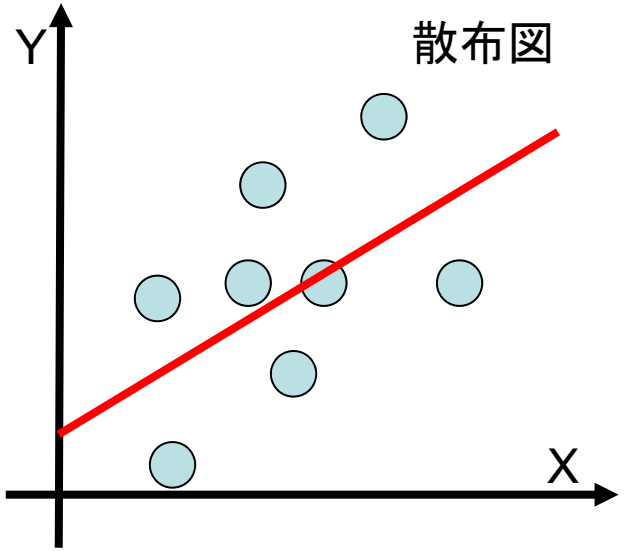
となり、求める直線は $y - \bar{y} = \frac{S_{xy}}{S_{xx}}(x - \bar{x})$　すなわち

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





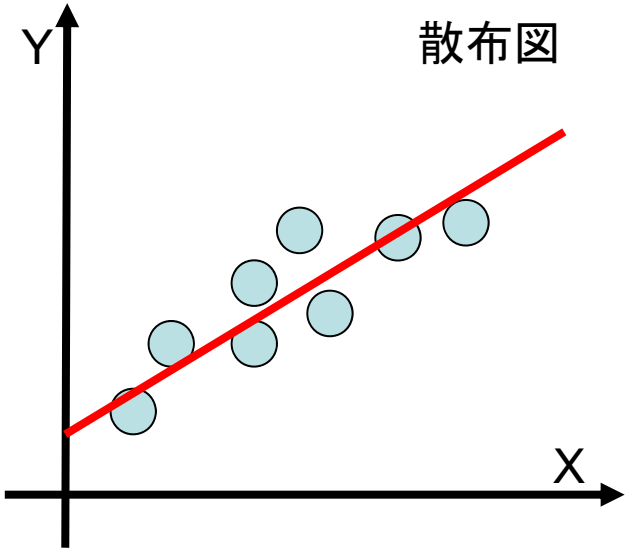
ピアソンの**標本相関係数**

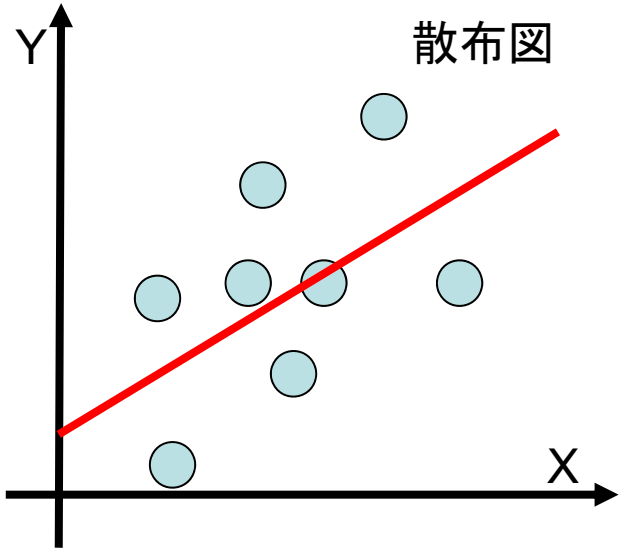
この係数 r は　　であり、散らばりが少ないとき、|r| は1に近い値をとる。

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)

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**

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

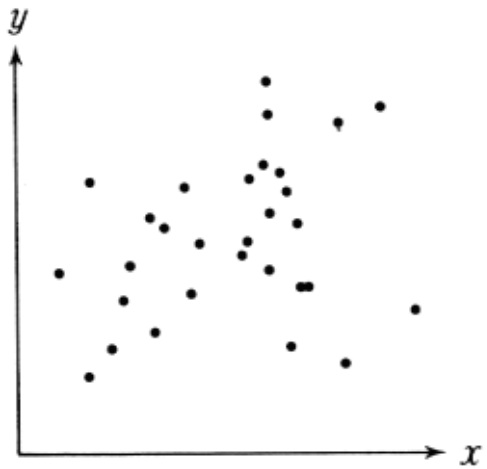
共分散

x の標準偏差

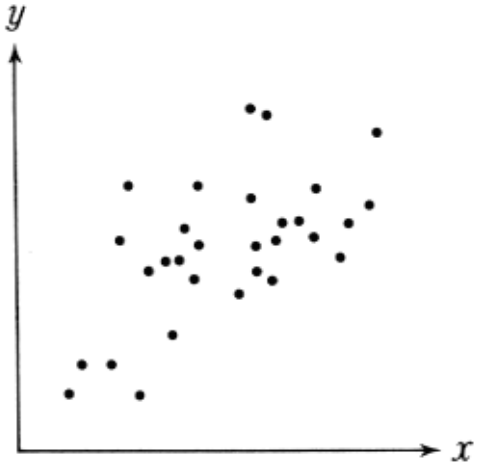
y の標準偏差

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

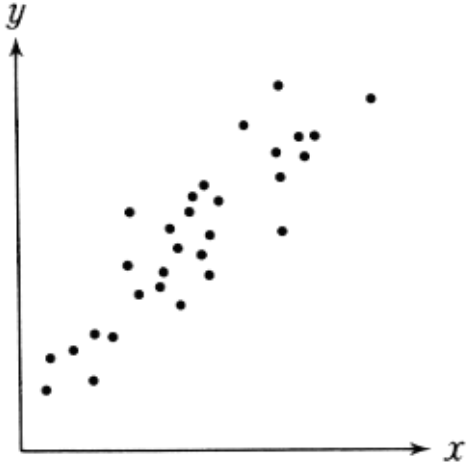
r > 0　のとき　回帰直線の傾きが正　（x, y の間に正の相関）
r < 0　のとき　回帰直線の傾きが負　（x, y の間に負の相関）



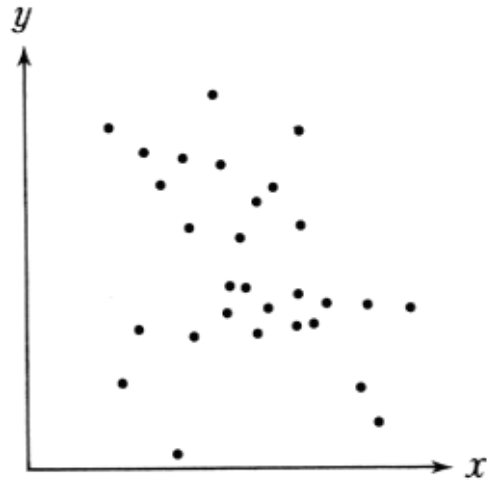
$r=0.3$



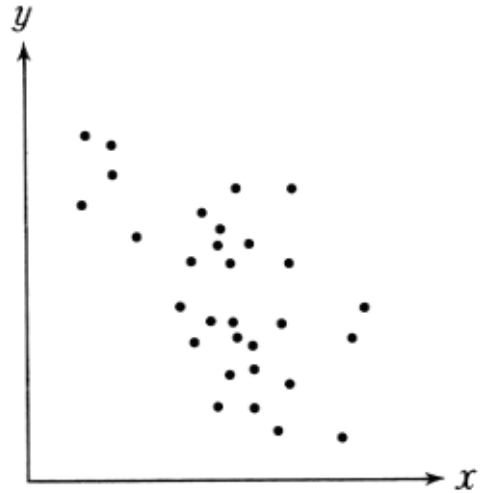
$r=0.6$



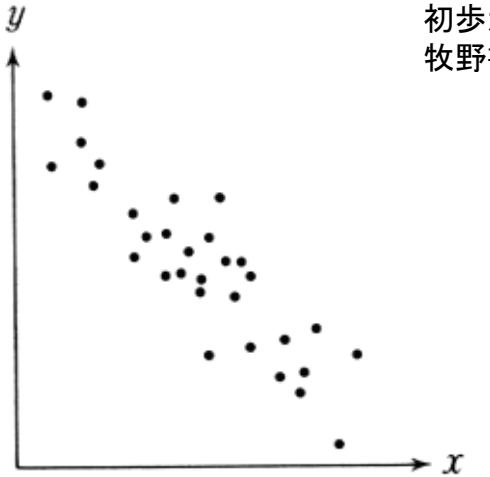
$r=0.9$



$r=-0.3$



$r=-0.6$



$r=-0.9$

【参考文献】
馬場　裕　著
初歩からの統計学
牧野書店

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

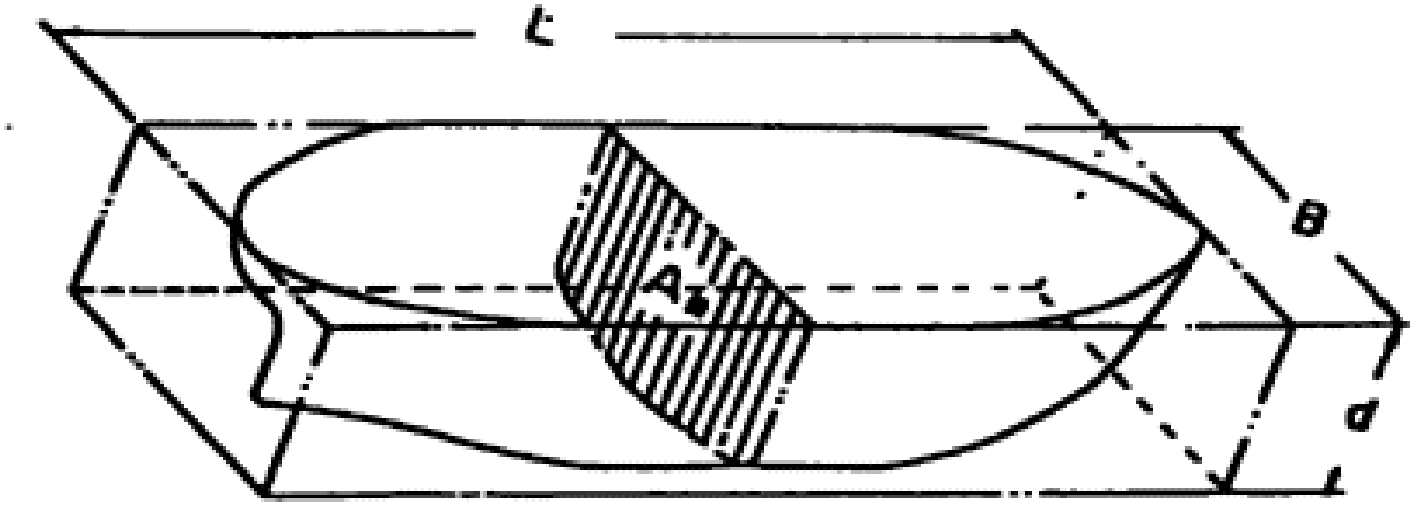
今までに製造したことのある類似船を参考に、おおよその船体形状を決め、
流体抵抗を推定 （**推定**　および　**最適化**）

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、噴水の等しい直方体の容積との比で表わされる。



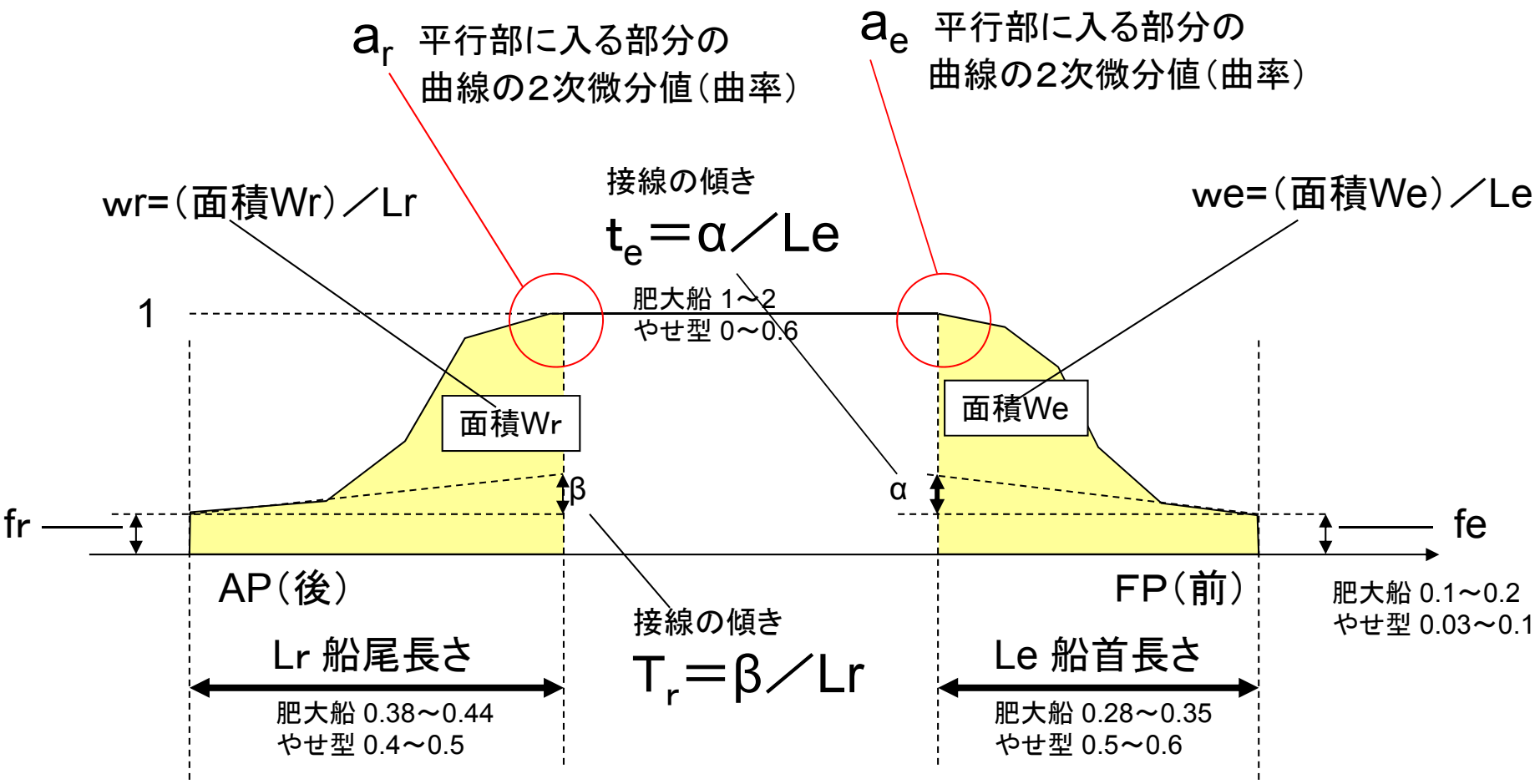
CBの値の小さい船をやせ型船、CBの大きい船を肥大船と言う。
CBの概略値は、
貨物船では0.62～0.84
旅客船では0.50～0.60
漁船では0.55～0.75

（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

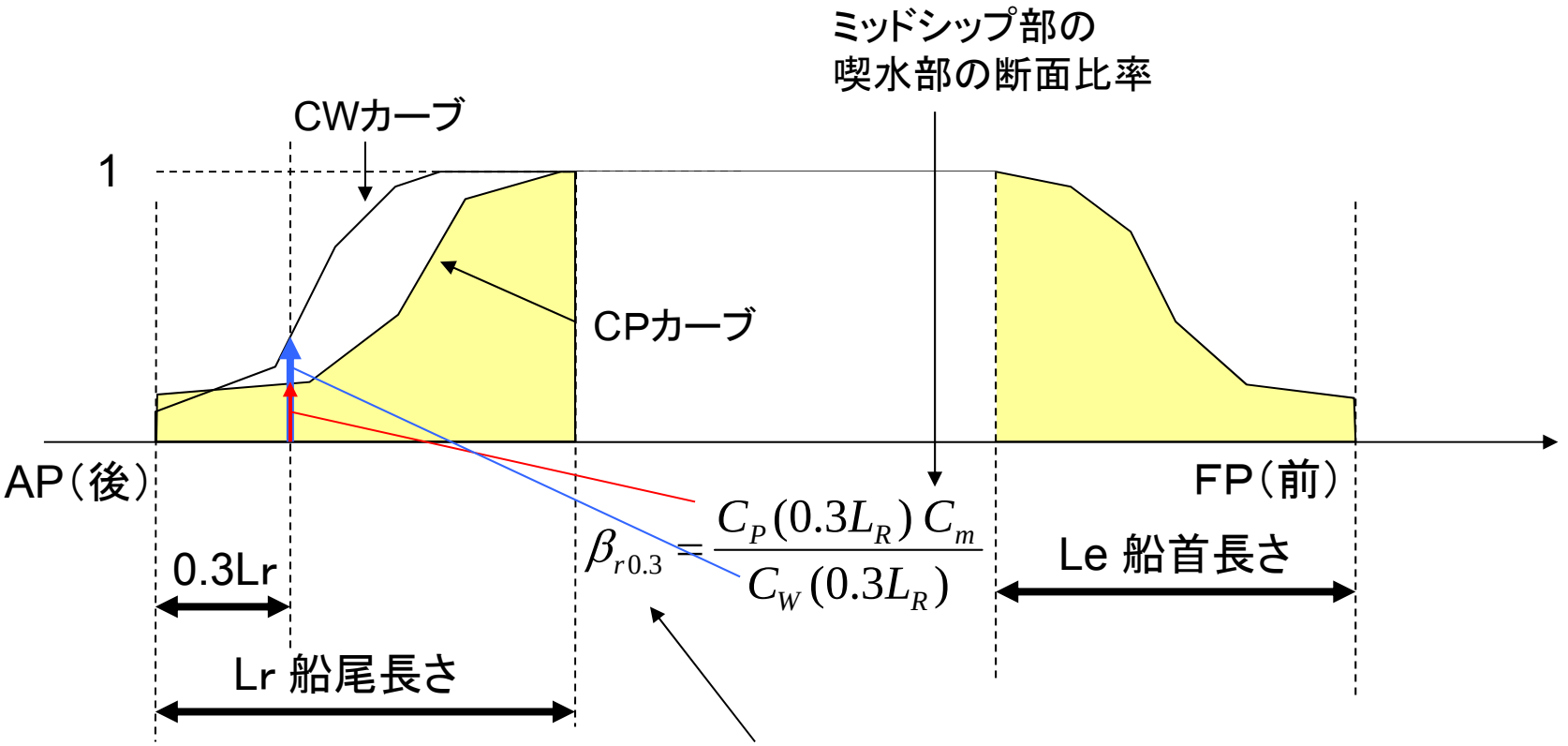
$$Hr/B = Lr\,(1 - wr)\,Lpp/B$$

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

船の要目：形状　CPカーブとCWカーブの両方に関連



Cp、CWカーブ自体は5次式で表現し、データ点をフィッティング

Percent_Lpp　バルバスバウ長さ
Percent_d　　バルバスバウ深さ

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）

$x_1, x_2, \cdots x_K$

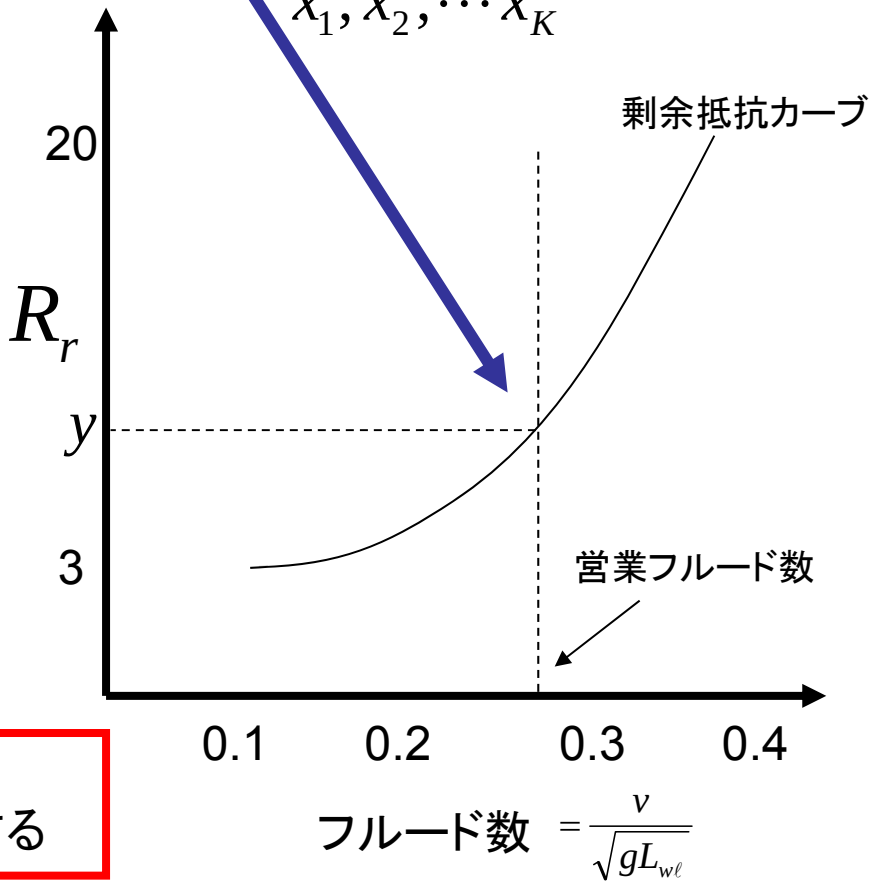
船の抵抗 $R_t = R_f + R_r$

摩擦抵抗　剰余抵抗

これを使う

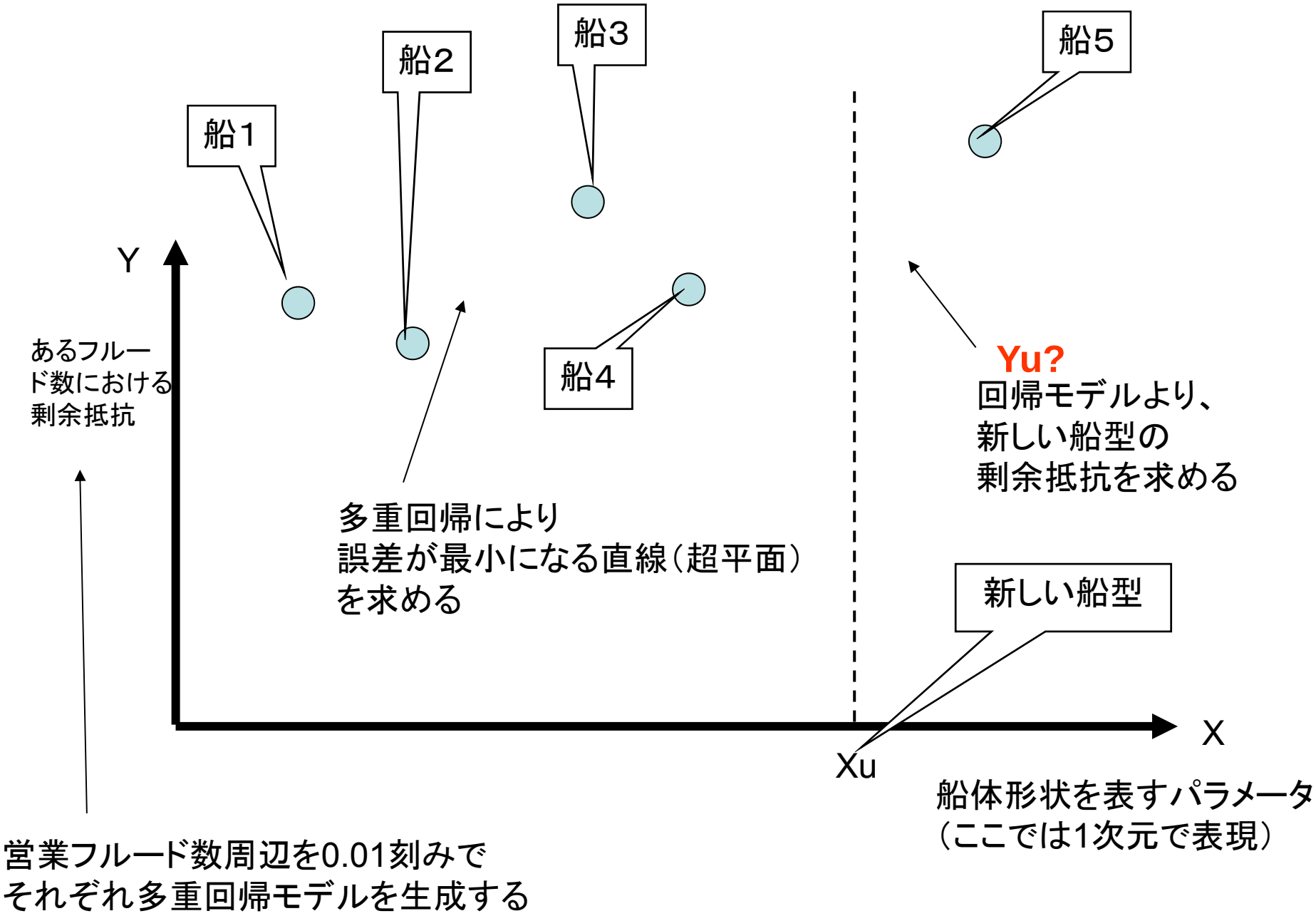
$= (1+K)R_f + R_W$

形状影響係数　造波抵抗

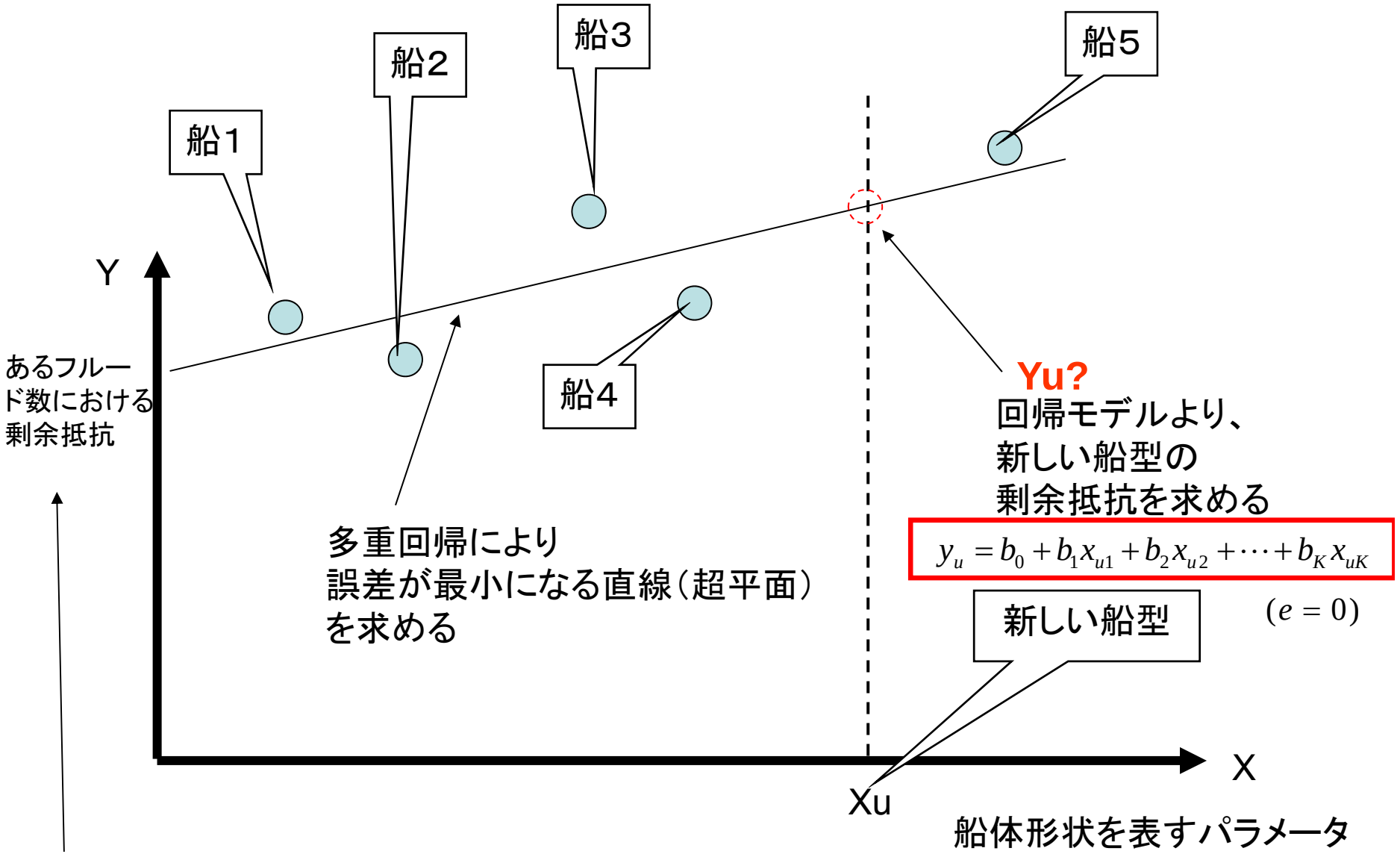


60隻分の実験データから、
新しい船型についての剰余抵抗曲線を推定する

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



痩せ型船型における推定のシミュレーション

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	

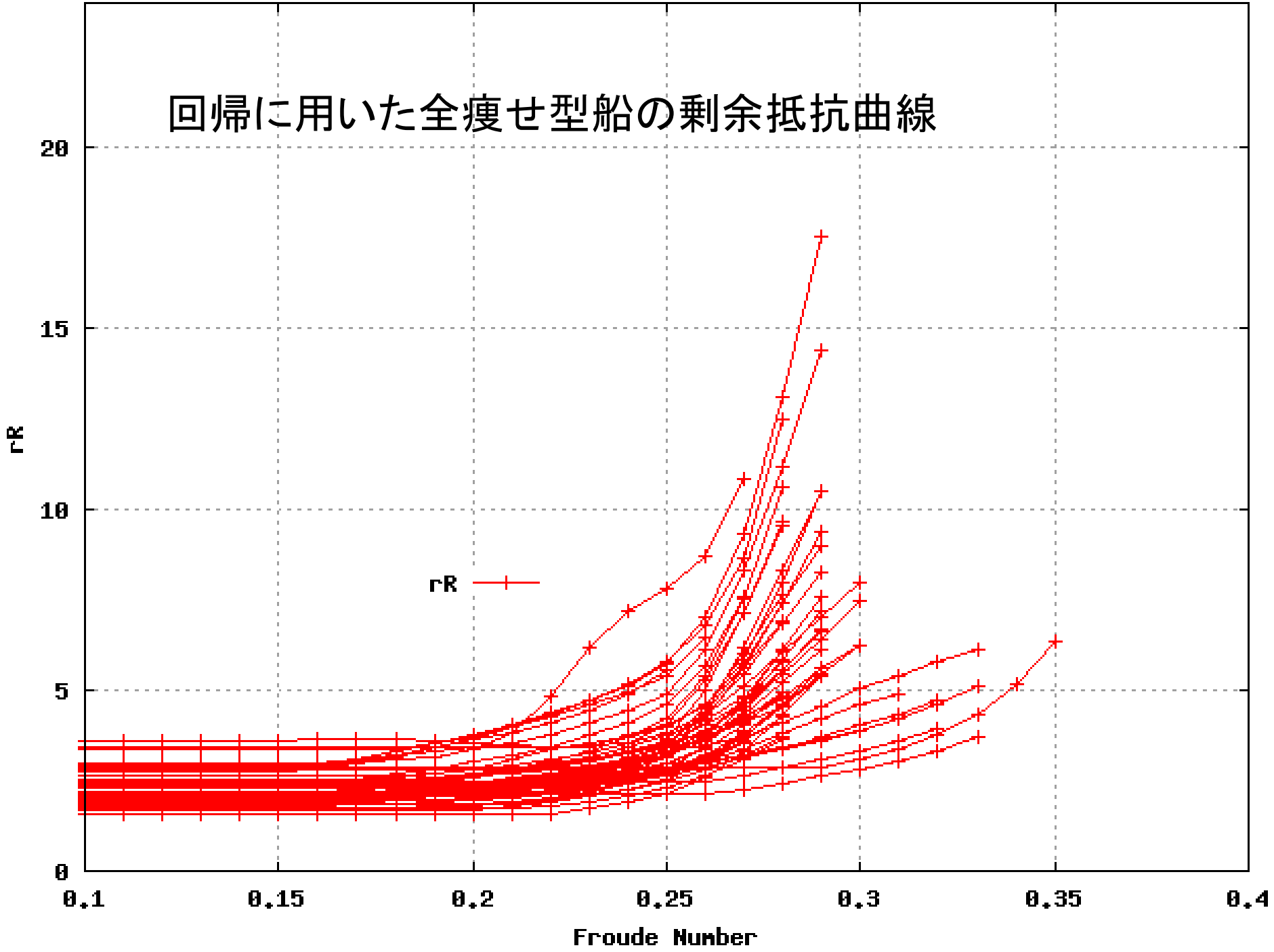
用いた船のデータ（Lppの小さい順57隻分）
n=57

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

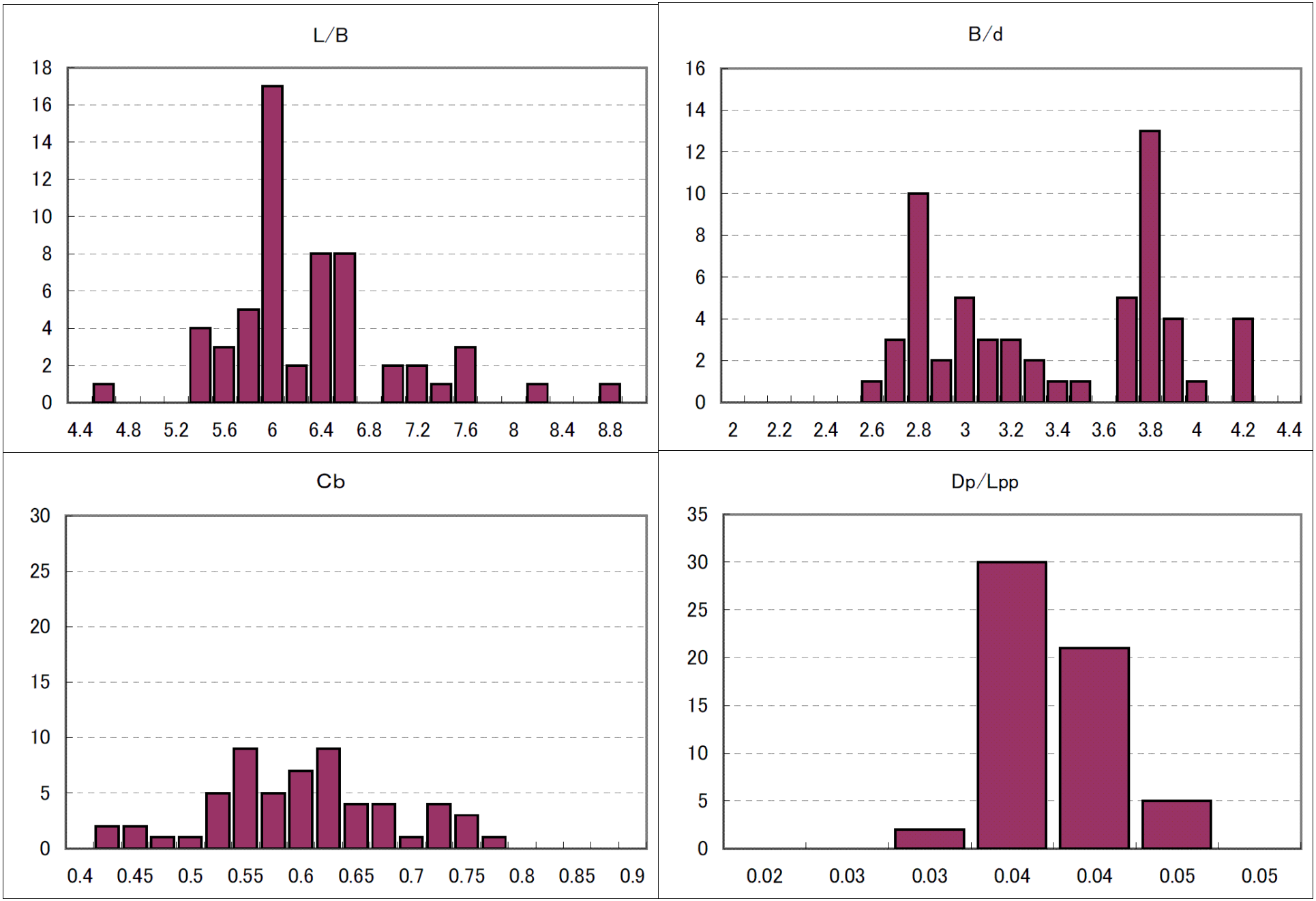
回帰モデルに用いた
パラメータ19個
K=19

性能を推定する未知船型
0176

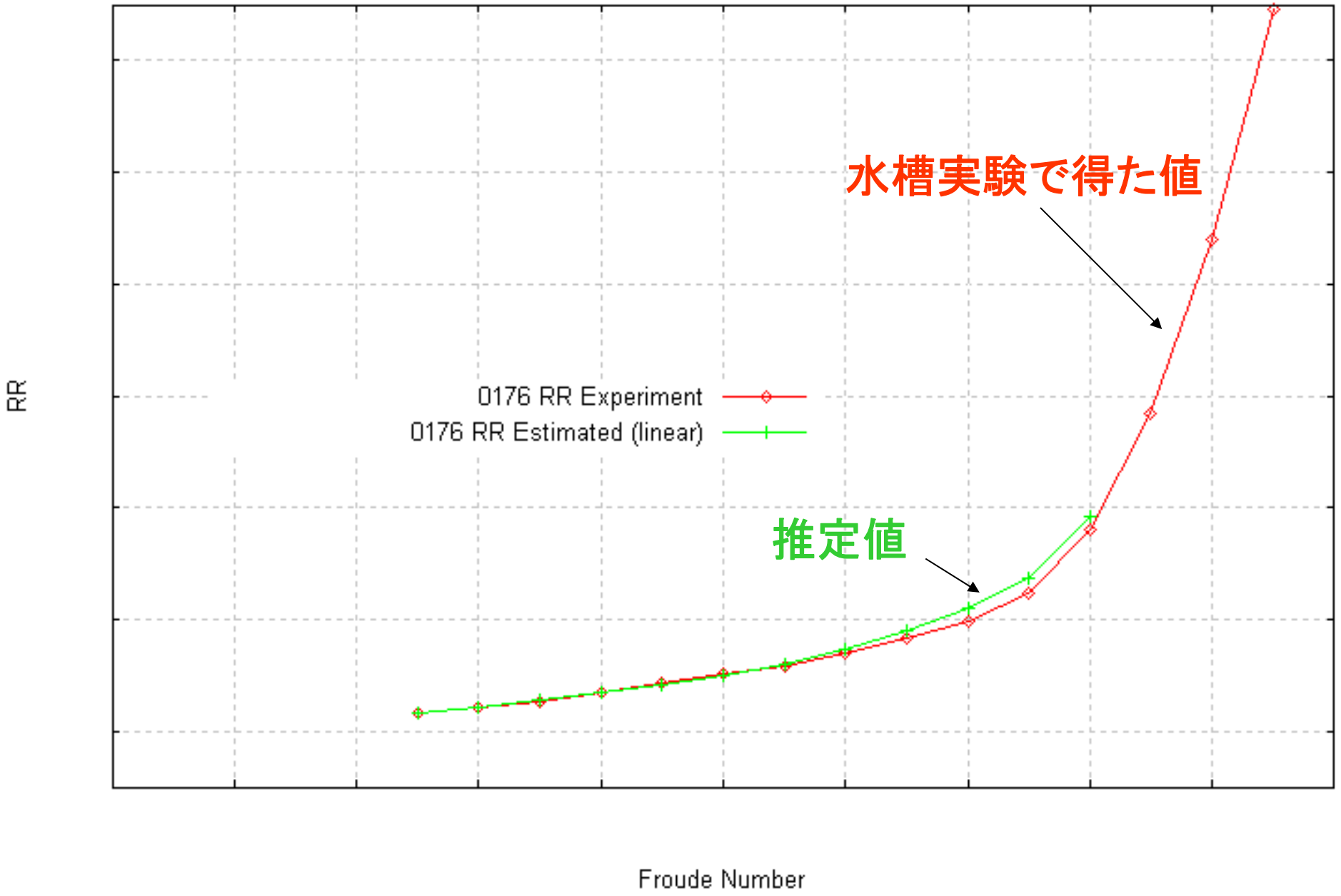
ServiceFn	0.234
L/B	6.28319
B/d	2.7561
Cb	0.7094
Cp	0.7203
Cm	0.9849
Cw	0.8435
Cv	0.84102
lcb	-0.4582


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

回帰に用いたデータの各変数のヒストグラム



剰余抵抗(RR)の推定結果



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

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

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、

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

線形表現

誤差ベクトル $\mathbf{e}$ の平方和 $\|\mathbf{e}\|^2$ を最小にする $\mathbf{b}$ を求める

→ 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:

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

で与えられる。

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:

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

Xの擬似逆行列 $\mathbf{X}^{+}$

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。 → $\sqrt{\frac{S_e}{n}}$ 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 $\hat{\mathbf{b}}$ の導出

2次関数の極値問題

$\mathbf{y} = \mathbf{Xb} + \mathbf{e}$ より $\|\mathbf{e}\|^2 = (\mathbf{Xb} - \mathbf{y})^{Trans}(\mathbf{Xb} - \mathbf{y})$ ← $\mathbf{b}$ の2次関数なので微分してゼロの $\mathbf{b}$ を求める

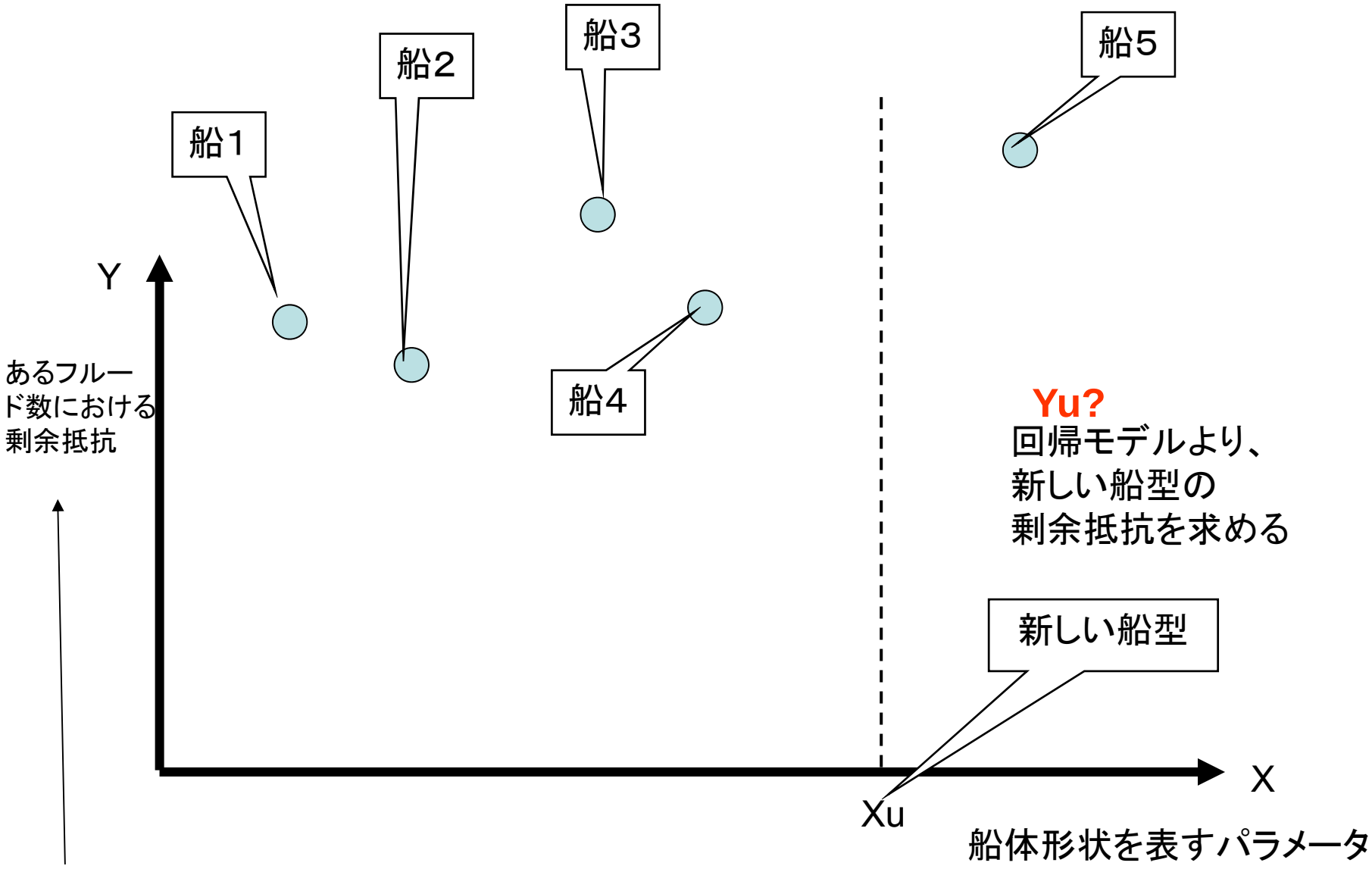
↓

$$\frac{\|\mathbf{e}\|^2}{\partial \mathbf{b}} = 2\mathbf{X}^{Trans}\mathbf{Xb} - 2\mathbf{X}^{Trans}\mathbf{y} = 0$$ ← $\mathbf{b}$ について解く

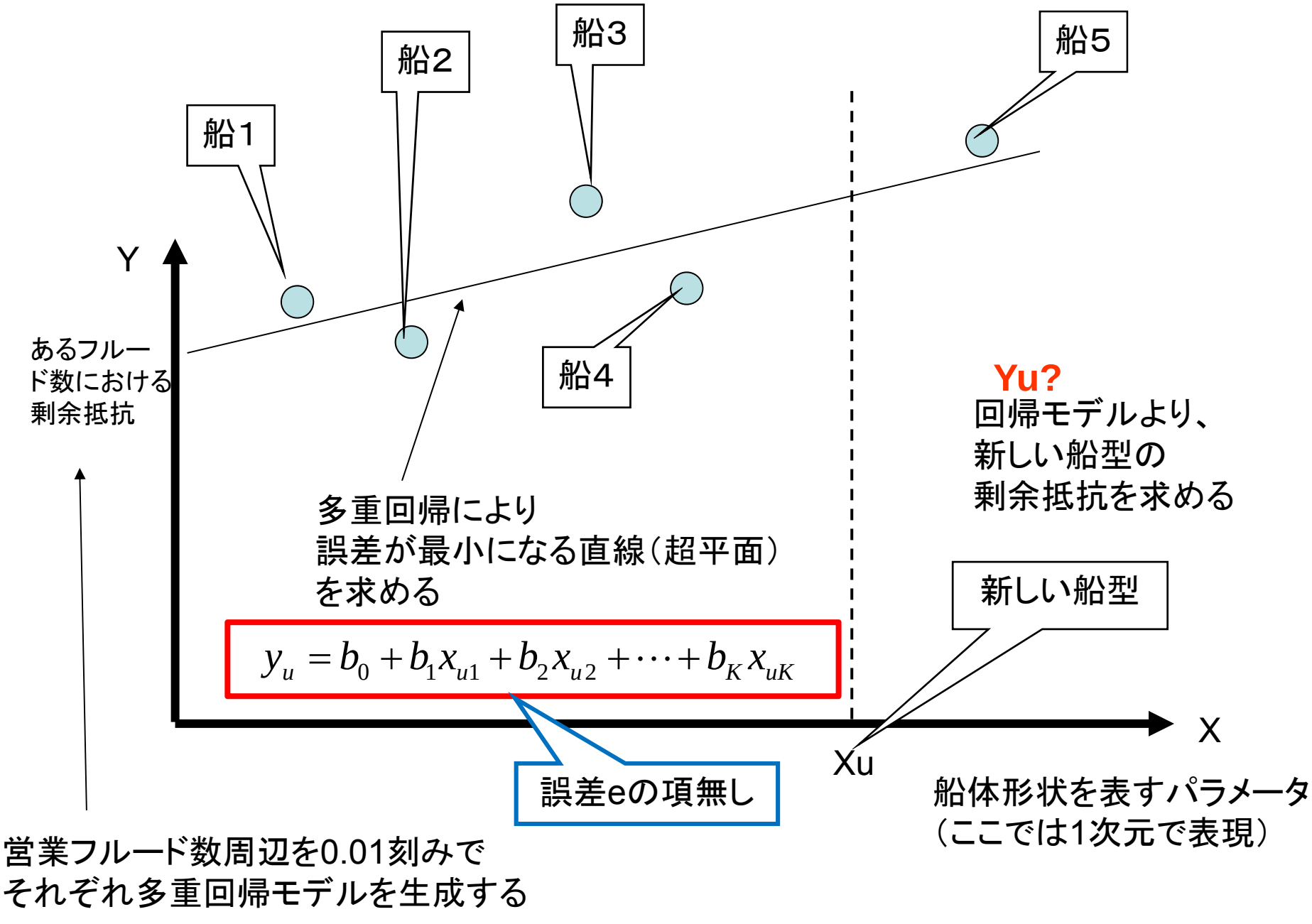
↓

$\hat{\mathbf{b}}$ ← $\mathbf{b} = (\mathbf{X}^{\mathbf{Trans}}\mathbf{X})^{-1}\mathbf{X}^{\mathbf{Trans}}\mathbf{y}$

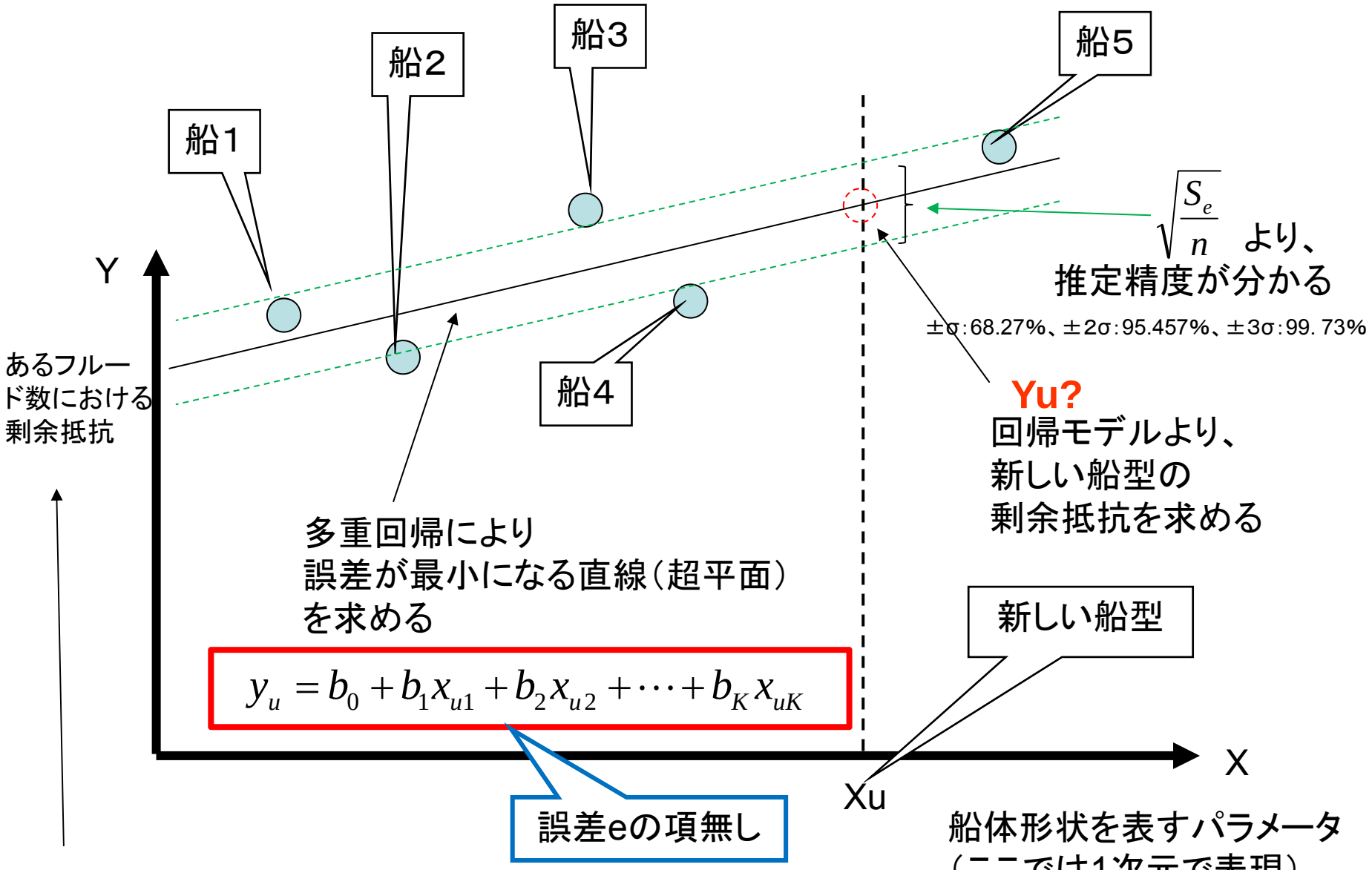
多重回帰を利用した新船型の剰余抵抗値の推定



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



【復習】 転置行列とは？

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

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

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

【復習】 転置行列とは？

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

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

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

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

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

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

→　回帰係数ベクトルを求める逆行列計算が不安定になり、無意味な解が出やすい

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

→ 回帰係数ベクトルを求める逆行列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

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

→ 回帰係数ベクトルを求める逆行列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

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

（超）平面の方程式:
データが超平面上にほぼ存在

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

分散共分散行列の固有値 ＝ 主成分分析

データを互いに直交する成分方向の分散を用いて表現

最大固有値の固有ベクトル ＝ 第一主成分 ＝ データの分散が最大の方向
最大固有値 ＝ 第一主成分方向の分散

以下、2番目に大きい固有値…最小の固有値まで全て同様

まとめ

【単純回帰】

回帰直線
$y = ax + b$

y の x に対する回帰直線

回帰パラメータ a, b を決める
（**最小2乗法**）

ピアソンの**標本相関係数**

【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:

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

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

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:

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

Xの擬似逆行列

多重共線性に注意して回帰変数を選択