

# データ解析

<http://coconut.sys.eng.shizuoka.ac.jp/data/06/>

静岡大学工学部

安藤和敏

2006.10.16

# 前回やり残したこと

- 分析ツール

# 2-2 1変数を多変数から予測する重 回帰分析

# 重回帰分析のデータ

個体番号	変数 $x$	変数 $u$	変数 $v$	変数 $w$	...	変数 $y$
1	$x_1$	$u_1$	$v_1$	$w_1$	...	$y_1$
2	$x_2$	$u_2$	$v_2$	$w_2$	...	$y_2$
⋮	⋮	⋮	⋮	⋮	⋮	⋮
$i$	$x_i$	$u_i$	$v_i$	$w_i$	...	$y_i$
⋮	⋮	⋮	⋮	⋮	⋮	⋮
$n$	$x_n$	$u_n$	$v_n$	$w_n$	...	$y_n$

# 重回帰分析のデータの例

社員 No	社交 性	勤勉性	企画力	判断力	給与評 価
1	7	6	7	8	10
2	4	5	5	4	4
3	6	8	4	4	8
4	5	5	5	5	8
5	6	6	4	5	6
6	6	5	6	6	7
7	4	4	6	6	8

# 重回帰分析の目的(の一つ)

与えられたデータに「最もよくあてはまる」方程式

回帰方程式

説明変数

$$y = a + bx + cu + dv + ew + \dots$$

を求めること。

目的変数

切片

偏回帰係数

「最もよくあてはまる」方程式ってどういうこと？

# 重回帰分析のデータ (説明変数が2個の場合)

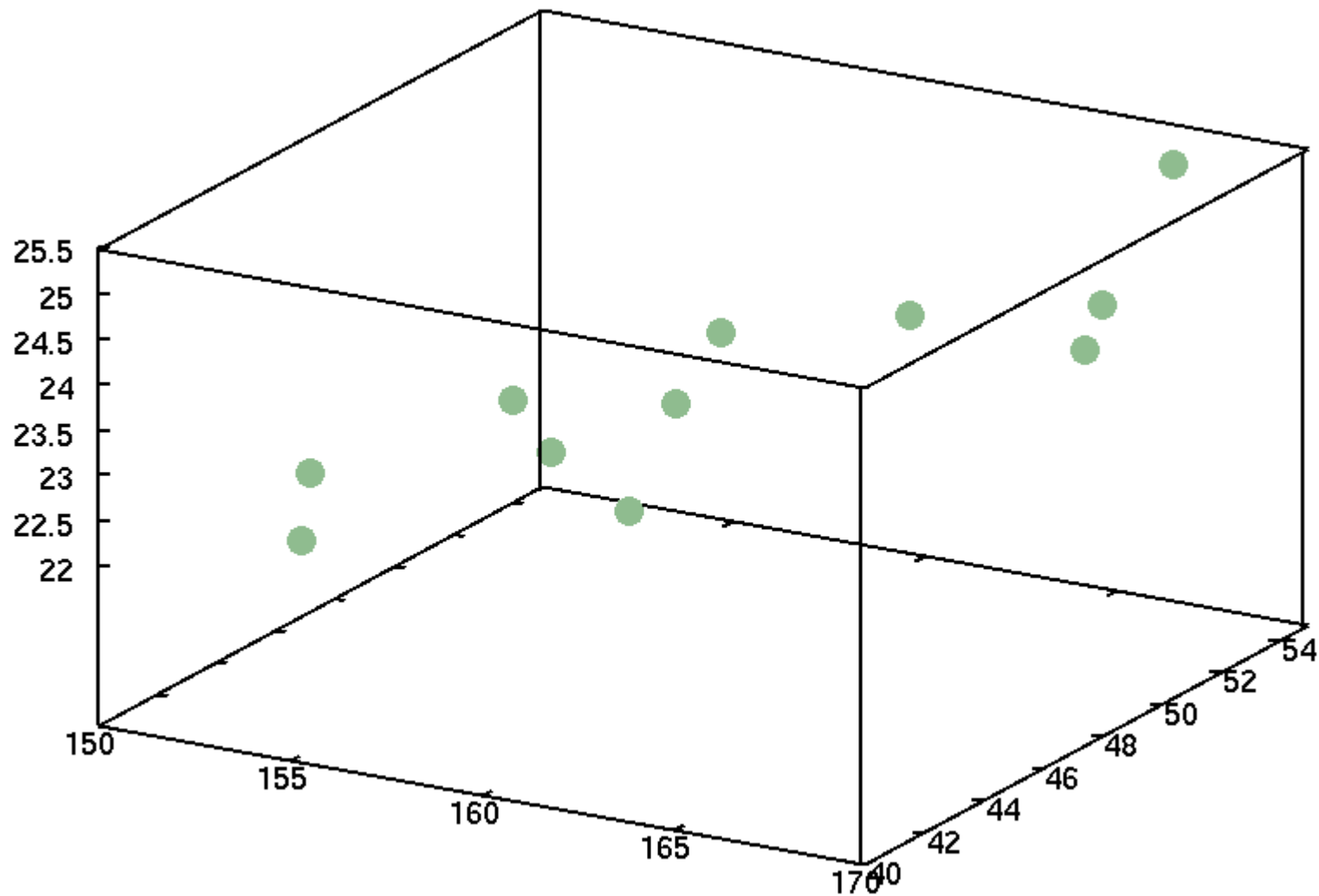
個体番号	変数 $x$	変数 $u$	変数 $y$
1	$x_1$	$u_1$	$y_1$
2	$x_2$	$u_2$	$y_2$
⋮	⋮	⋮	⋮
$i$	$x_i$	$u_i$	$y_i$
⋮	⋮	⋮	⋮
$n$	$x_n$	$u_n$	$y_n$

# 重回帰分析のデータ (説明変数が2個の場合)

番号	身長(x)	体重(u)	靴サイズ(y)
1	162	44	24.0
2	165	48	24.5
3	168	53	25.5
4	160	45	22.5
5	158	45	23.0
6	153	43	22.0
7	158	45	23.0
8	168	50	24.0
9	157	52	23.0
10	154	42	23.0
11	170	48	25.0
12	157	45	23.5



# データの3次元プロット



# 説明変数が2個の場合の重回帰分析

与えられたデータに「最もよくあてはまる」平面

回帰方程式  $y = a + bx + cu \dots\dots(1)$

を求めること。

目的変数

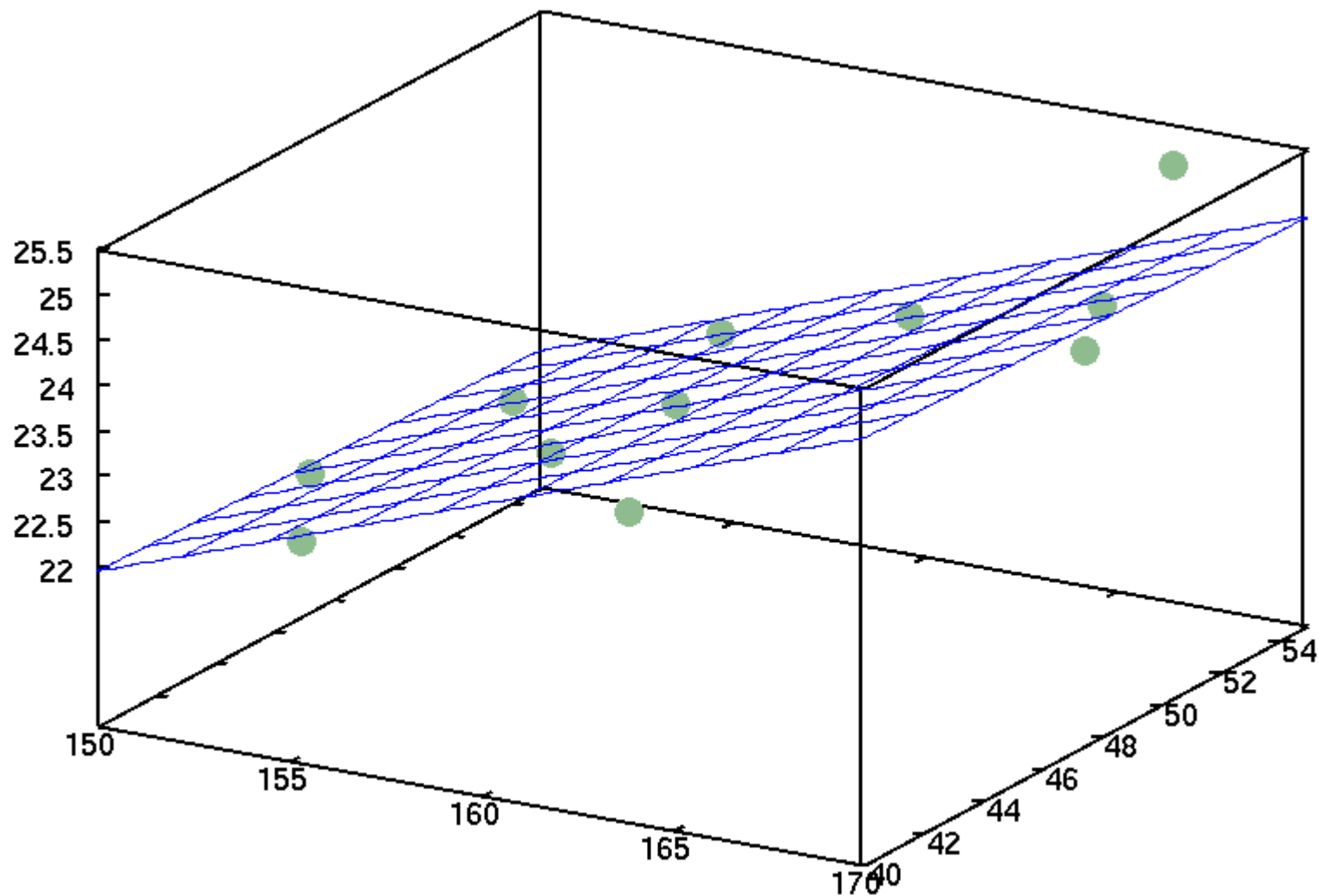
切片

回帰係数

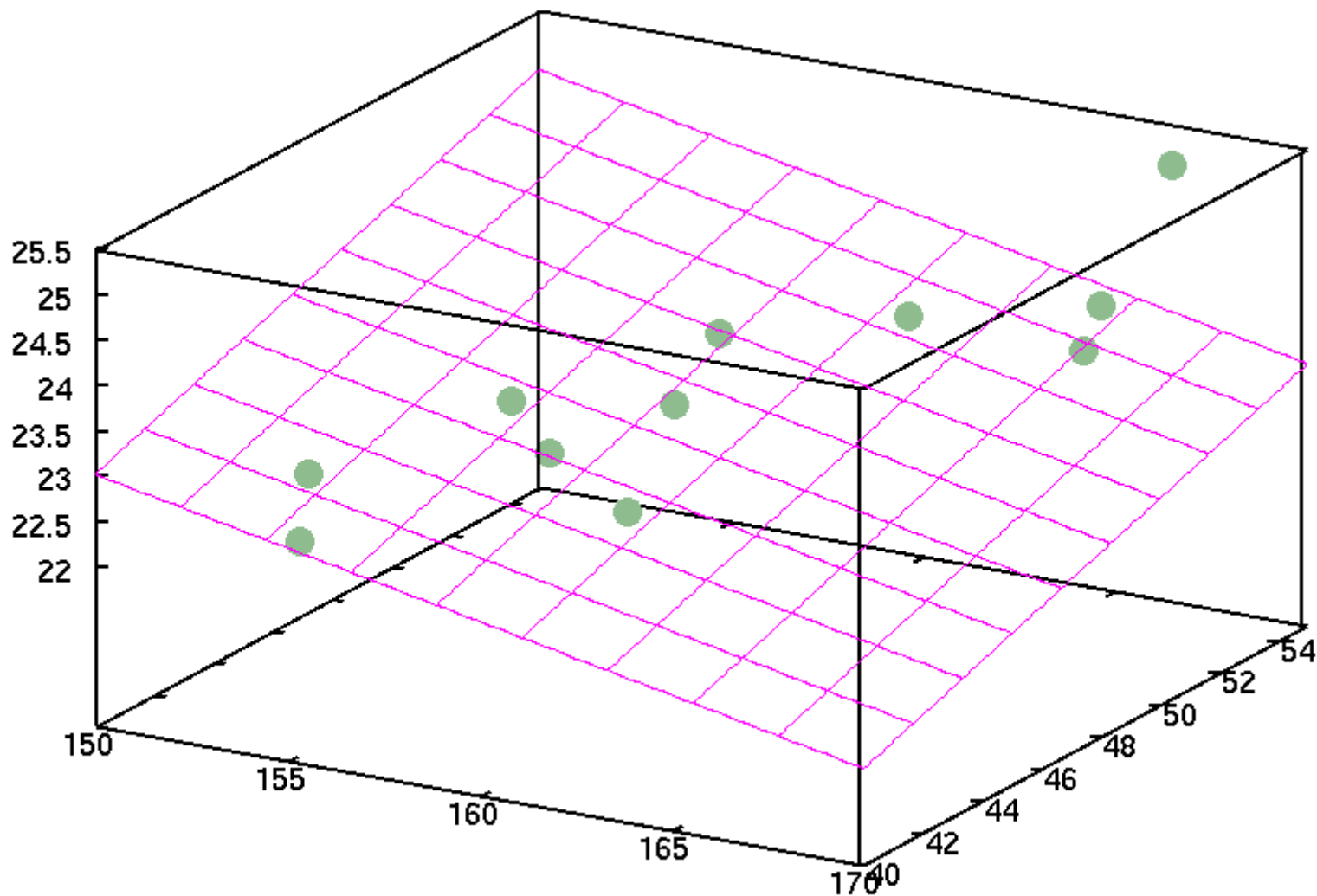
説明変数

「最もよくあてはまる平面」ってどういうこと？

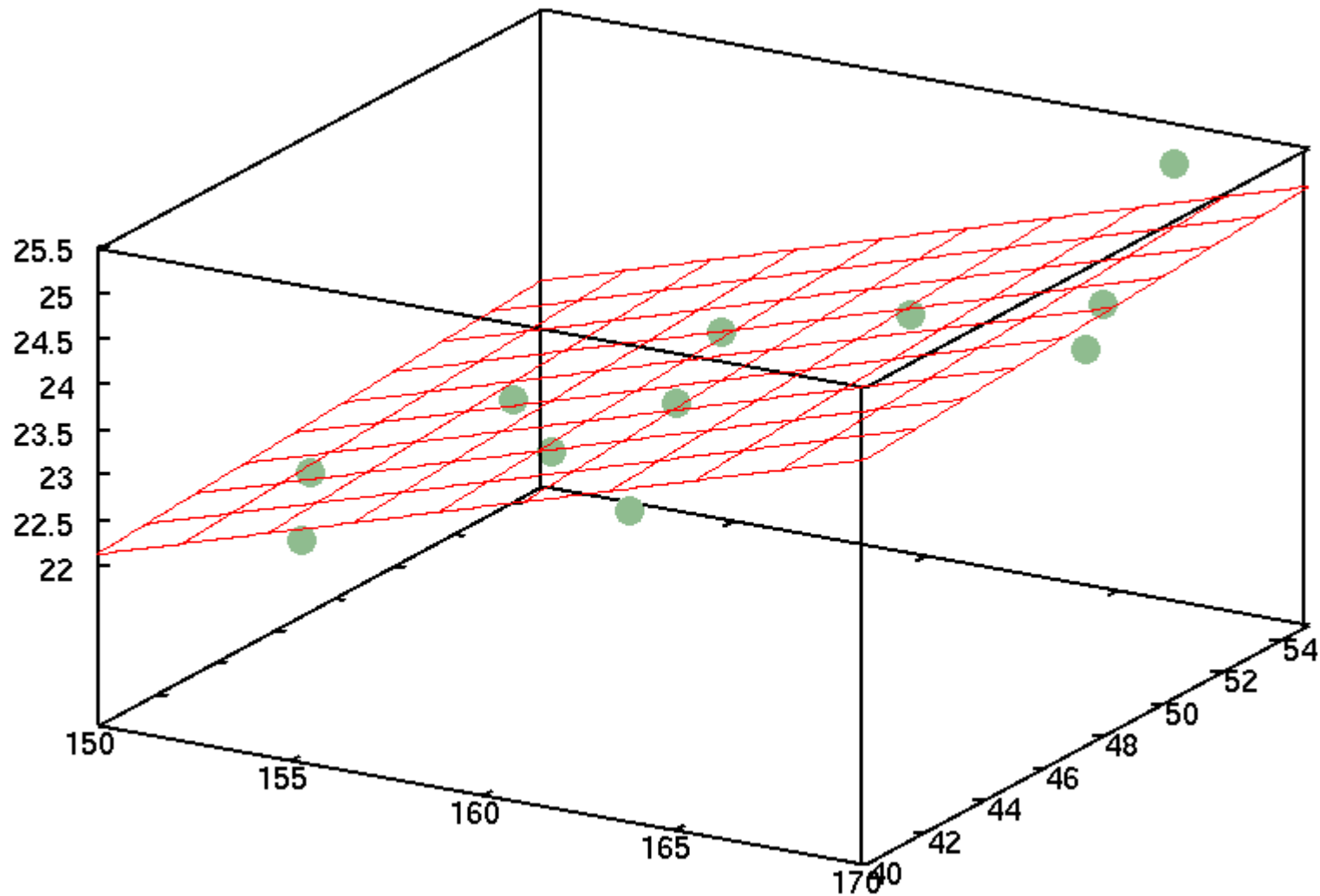
# 平面のあてはめ(1)



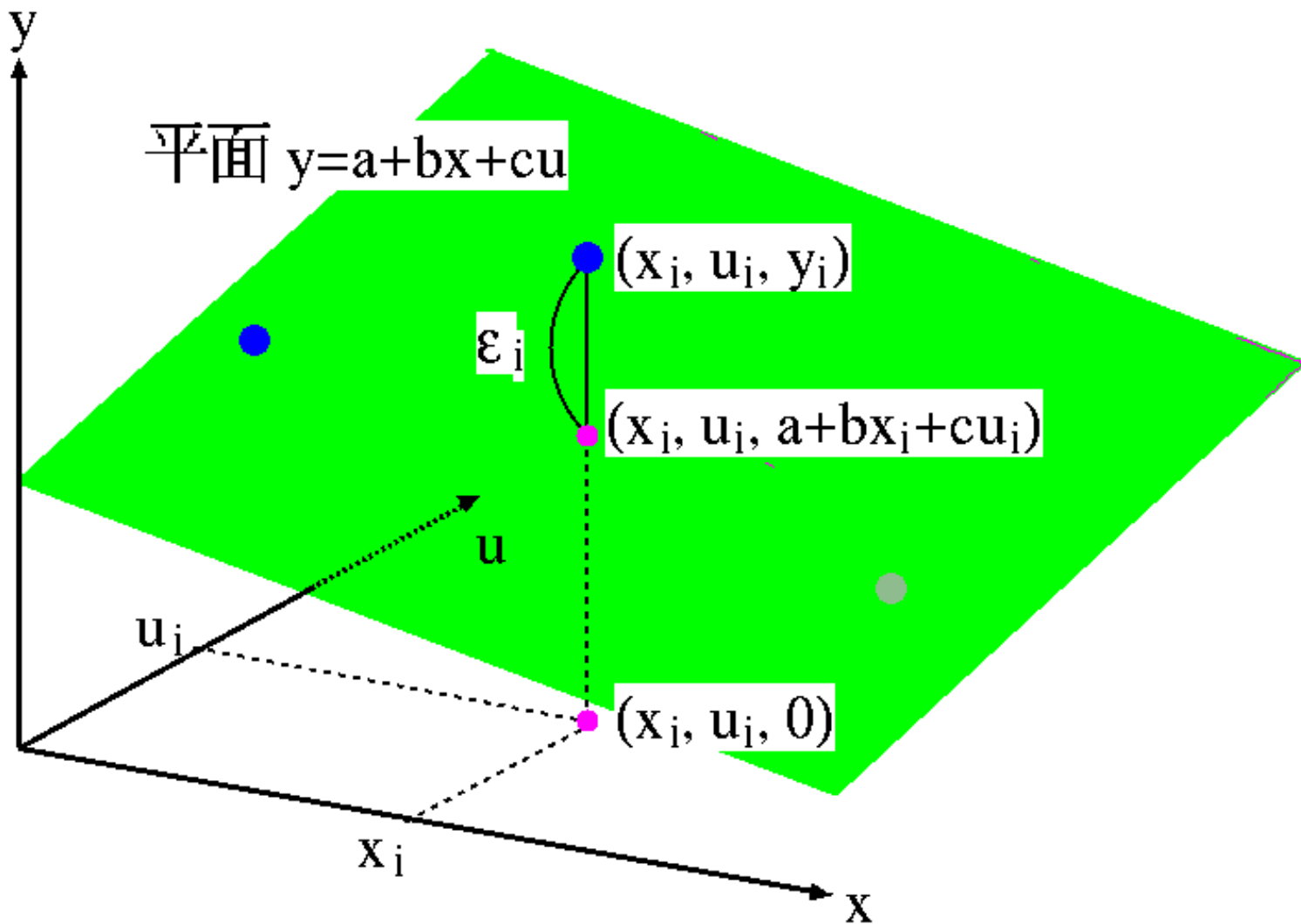
# 平面のあてはめ(2)



# 平面のあてはめ(3)



# 残差 $\varepsilon_i = y_i - (a + bx_i + cu_i)$



# 残差平方和 $Q$

$$Q = \sum_{i=1}^n \varepsilon_i^2$$
$$= \sum_{i=1}^n \{y_i - (a + bx_i + cu_i)\}^2$$

$Q$ を  $a, b, c$ を変数にもつ3変数関数として見て,  
 $Q(a, b, c)$ を最小にする  $a, b, c$ が, データに「最もよくあてはまる」平面を与えると考える.

このようにして  $a, b, c$ を求める方法を**最小2乗法**と呼ぶ.

どのようにして  $Q(a, b, c)$ を最小にする  $a, b, c$ をもとめるのかを見ていく.

一般に多変数関数の極値(最大値, 最小値)を求めるには, 各変数で偏微分して0と置いた方程式系を解けばよい

$$\left\{ \begin{array}{l} \frac{\partial Q}{\partial a} = \sum_{i=1}^n -2\{y_i - (a + bx_i + cu_i)\} = 0, \\ \frac{\partial Q}{\partial b} = \sum_{i=1}^n -2x_i\{y_i - (a + bx_i + cu_i)\} = 0, \\ \frac{\partial Q}{\partial c} = \sum_{i=1}^n -2u_i\{y_i - (a + bx_i + cu_i)\} = 0 \end{array} \right.$$



# 連立方程式を解く(1)

$$\left\{ \begin{array}{l} \sum_{i=1}^n \{y_i - (a + bx_i + cu_i)\} = 0, \\ \sum_{i=1}^n x_i \{y_i - (a + bx_i + cu_i)\} = 0, \\ \sum_{i=1}^n u_i \{y_i - (a + bx_i + cu_i)\} = 0 \end{array} \right.$$

# 連立方程式を解く(2)

$$\sum_{i=1}^n \{y_i - (a + bx_i + cu_i)\} = 0$$

$$\bar{y} = a + b\bar{x} + c\bar{u}$$

# 連立方程式を解く(3)

$$\begin{aligned} & \sum_{i=1}^n x_i \{y_i - (a + bx_i + cu_i)\} \\ &= \sum_{i=1}^n x_i \{y_i - (\bar{y} - b\bar{x} - c\bar{u} + bx_i + cu_i)\} \\ &= \sum_{i=1}^n x_i \{(y_i - \bar{y}) - b(x_i - \bar{x}) - c(u_i - \bar{u})\} \\ &+ \sum_{i=1}^n \bar{x} \{(y_i - \bar{y}) - b(x_i - \bar{x}) - c(u_i - \bar{u})\} \\ &= \sum_{i=1}^n (x_i - \bar{x}) \{(y_i - \bar{y}) - b(x_i - \bar{x}) - c(u_i - \bar{u})\} \\ &= ns_{xy} - bns_x^2 - cns_{xu} \end{aligned}$$

# 連立方程式を解く(4)

$$\begin{aligned} & \sum_{i=1}^n u_i \{y_i - (a + bx_i + cu_i)\} \\ &= \sum_{i=1}^n u_i \{y_i - (\bar{y} - b\bar{x} - c\bar{u} + bx_i + cu_i)\} \\ &= \sum_{i=1}^n u_i \{(y_i - \bar{y}) - b(x_i - \bar{x}) - c(u_i - \bar{u})\} \\ &+ \sum_{i=1}^n \bar{u} \{(y_i - \bar{y}) - b(x_i - \bar{x}) - c(u_i - \bar{u})\} \\ &= \sum_{i=1}^n (u_i - \bar{u}) \{(y_i - \bar{y}) - b(x_i - \bar{x}) - c(u_i - \bar{u})\} \\ &= ns_{uy} - bns_{xu} - cns_u^2 \end{aligned}$$

# 連立方程式を解く(5)

$$0 = s_{xy} - bs_x^2 - cs_{xu},$$

$$0 = s_{uy} - bs_{xu} - cs_u^2$$

$$\begin{bmatrix} s_x^2 & s_{xu} \\ s_{xu} & s_u^2 \end{bmatrix} \begin{bmatrix} b \\ c \end{bmatrix} = \begin{bmatrix} s_{xy} \\ s_{uy} \end{bmatrix}$$

# 連立方程式の解

$$\begin{bmatrix} b \\ c \end{bmatrix} = \begin{bmatrix} s_x^2 & s_{xu} \\ s_{xu} & s_u^2 \end{bmatrix}^{-1} \begin{bmatrix} s_{xy} \\ s_{uy} \end{bmatrix},$$

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

# 多重共線性(1)

$$\det \begin{bmatrix} s_x^2 & s_{xu} \\ s_{xu} & s_u^2 \end{bmatrix} = 0 \quad \text{のときは,}$$

のときは, 方程式

$$\begin{bmatrix} s_x^2 & s_{xu} \\ s_{xu} & s_u^2 \end{bmatrix} \begin{bmatrix} b \\ c \end{bmatrix} = \begin{bmatrix} s_{xy} \\ s_{uy} \end{bmatrix}$$

の解は一意に定まらない. なにが起こっているのか?

## 多重共線性(3)

$$\det \begin{bmatrix} s_x^2 & s_{xu} \\ s_{xu} & s_u^2 \end{bmatrix} = 0 \Leftrightarrow s_x^2 s_u^2 - s_{xu}^2 = 0$$

$$\Leftrightarrow r_{xu}^2 = \frac{s_{xu}^2}{s_x^2 s_u^2} = 1$$

であるから、 $x$ と $u$ の間には、 $x = u +$  という関係がある。したがって、 $x$ か $u$ のうちのどちらか一方をモデルから取り除いても $y$ を説明できる。



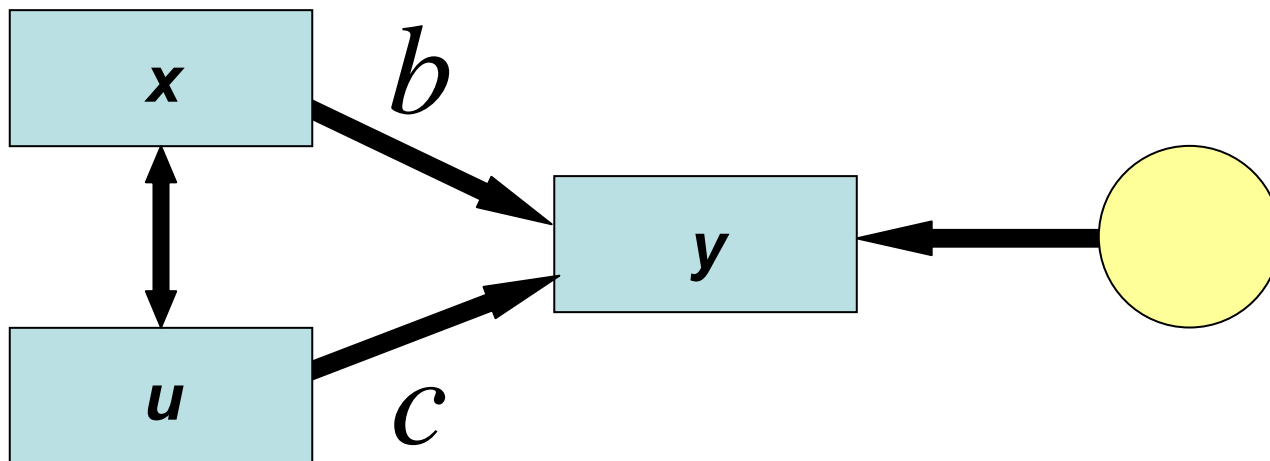
# 多重共線性(3)

$$\det \begin{bmatrix} s_x^2 & s_{xu} \\ s_{xu} & s_u^2 \end{bmatrix} = 0$$

のときも、同様のことが言える。

# 重回帰分析のパス図

$$y = a + bx + cu$$



# 残差平方和の別表現(1)

$$\begin{aligned} Q &= \sum_{i=1}^n \varepsilon_i^2 = \sum_{i=1}^n \{y_i - (a + bx_i + cu_i)\}^2 \\ &= \sum_{i=1}^n \{y_i - \bar{y} + \bar{y} - (a + bx_i + cu_i)\}^2 \\ &= \sum_{i=1}^n \{y_i - \bar{y} + a + b\bar{x} - (a + bx_i + cu_i)\}^2 \\ &= \sum_{i=1}^n \{(y_i - \bar{y}) - b(x_i - \bar{x}) - c(u_i - \bar{u})\}^2 \end{aligned}$$

# 残差平方和の別表現(2)

(つづき)

$$\begin{aligned} &= \sum_{i=1}^n (y_i - \bar{y})^2 + b^2 \sum_{i=1}^n (x_i - \bar{x})^2 + c^2 \sum_{i=1}^n (u_i - \bar{u})^2 \\ &\quad - 2b \sum_{i=1}^n (y_i - \bar{y})(x_i - \bar{x}) - 2c \sum_{i=1}^n (y_i - \bar{y})(u_i - \bar{u}) \\ &\quad + 2bc \sum_{i=1}^n (x_i - \bar{x})(u_i - \bar{u}) \\ &= ns_y^2 + b^2 ns_x^2 + c^2 ns_u^2 - 2nbs_{xy} - 2ncs_{uy} + 2nbc s_{xu} \end{aligned}$$

# 残差平方和の別表現(3)

(つづき)

$$= ns_y^2 + b^2 ns_x^2 + c^2 ns_u^2 - 2nbs_{xy} - 2ncs_{uy} + 2nbc s_{xu}$$

$$= ns_y^2 + n \begin{bmatrix} b & c \end{bmatrix} \begin{bmatrix} s_x^2 & s_{xu} \\ s_{xu} & s_u^2 \end{bmatrix} \begin{bmatrix} b \\ c \end{bmatrix} - 2nbs_{xy} - 2ncs_{uy}$$

$$= ns_y^2 + n \begin{bmatrix} b & c \end{bmatrix} \begin{bmatrix} s_{xy} \\ s_{uy} \end{bmatrix} - 2nbs_{xy} - 2ncs_{uy}$$

$$= ns_y^2 - n \begin{bmatrix} b & c \end{bmatrix} \begin{bmatrix} s_{xy} \\ s_{uy} \end{bmatrix} = ns_y^2 - n \begin{bmatrix} b & c \end{bmatrix} \begin{bmatrix} s_x^2 & s_{xu} \\ s_{xu} & s_u^2 \end{bmatrix} \begin{bmatrix} b \\ c \end{bmatrix}$$

# 残差平方和の別表現(4)

$$s_{\varepsilon}^2 = s_y^2 - [b \quad c] \begin{bmatrix} s_x^2 & s_{xu} \\ s_{xu} & s_u^2 \end{bmatrix} \begin{bmatrix} b \\ c \end{bmatrix}$$

$$= s_y^2 - bs_x^2 - cs_u^2 - 2bcs_{xu}$$

つまり

$$s_y^2 = s_{\varepsilon}^2 + bs_x^2 + cs_u^2 + 2bcs_{xu} \quad \dots\dots (3)$$

# 本日のまとめ

- 説明変数が2個の場合の重回帰分析のモデルを理解した。
- 説明変数が2個の場合の最小2乗法の考え方、及び、回帰方程式の求め方を理解した。
- Excelを用いて重回帰分析を行う方法を理解した。
- 次の式

$$s_y^2 = s_\varepsilon^2 + bs_x^2 + cs_u^2 + 2bcs_{xu} \dots\dots (3)$$

の導出法を理解した。