2008年度 松山大学経営学部 開講科目

経営工学概論

檀 裕也

(dan@cc.matsuyama-u.ac.jp)

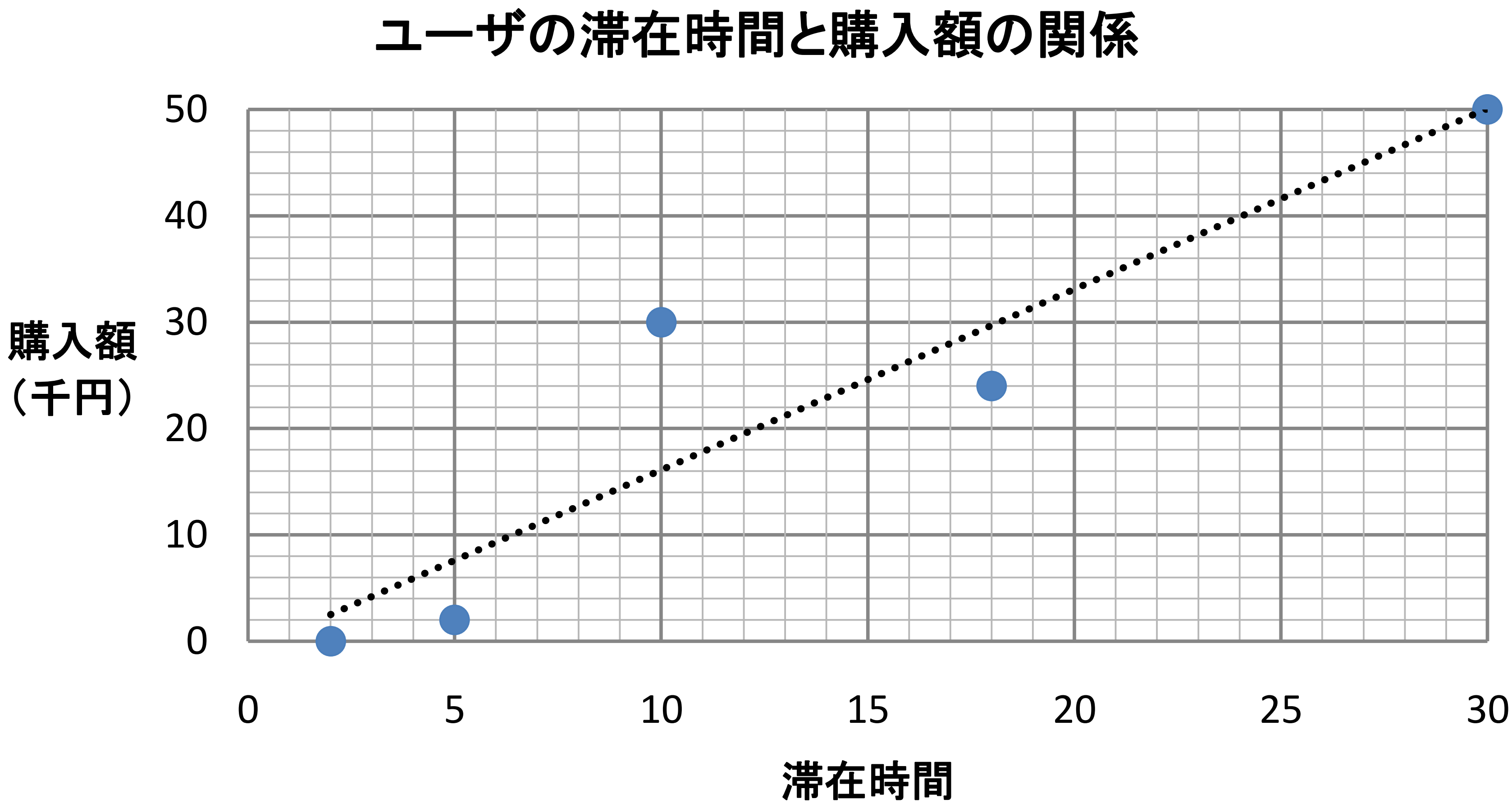
http://www.cc.matsuyama-u.ac.jp/~dan/

前回の課題

- あるネット通販サイトにおけるユーザの滞在時間とそのサイトでの購入額について下の表に示したデータがある。
  - 回帰分析によって、
    滞在時間で購入額を説明せよ。
  - 滞在時間20分の場合の
    ユーザの購入額を予測せよ。

時間(分)	購入額(千円)
2	0
5	2
10	30
18	24
30	50

散布図を描く



表計算

x	y	$x_i - \bar{x}$	$y_i - \bar{y}$	$(x_i - \bar{x})(y_i - \bar{y})$	$(x_i - \bar{x})^2$
2	0	-11	-21.2	233.2	121
5	2	-8	-19.2	153.6	64
10	30	-3	8.8	-26.4	9
18	24	5	2.8	14.0	25
30	50	17	28.8	489.6	289
13	21.2			864.0	508

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

$$y - 21.2 = \frac{864}{508}(x - 13)$$

$$S_{xy} = 864$$

$$S_{xx} = 508$$

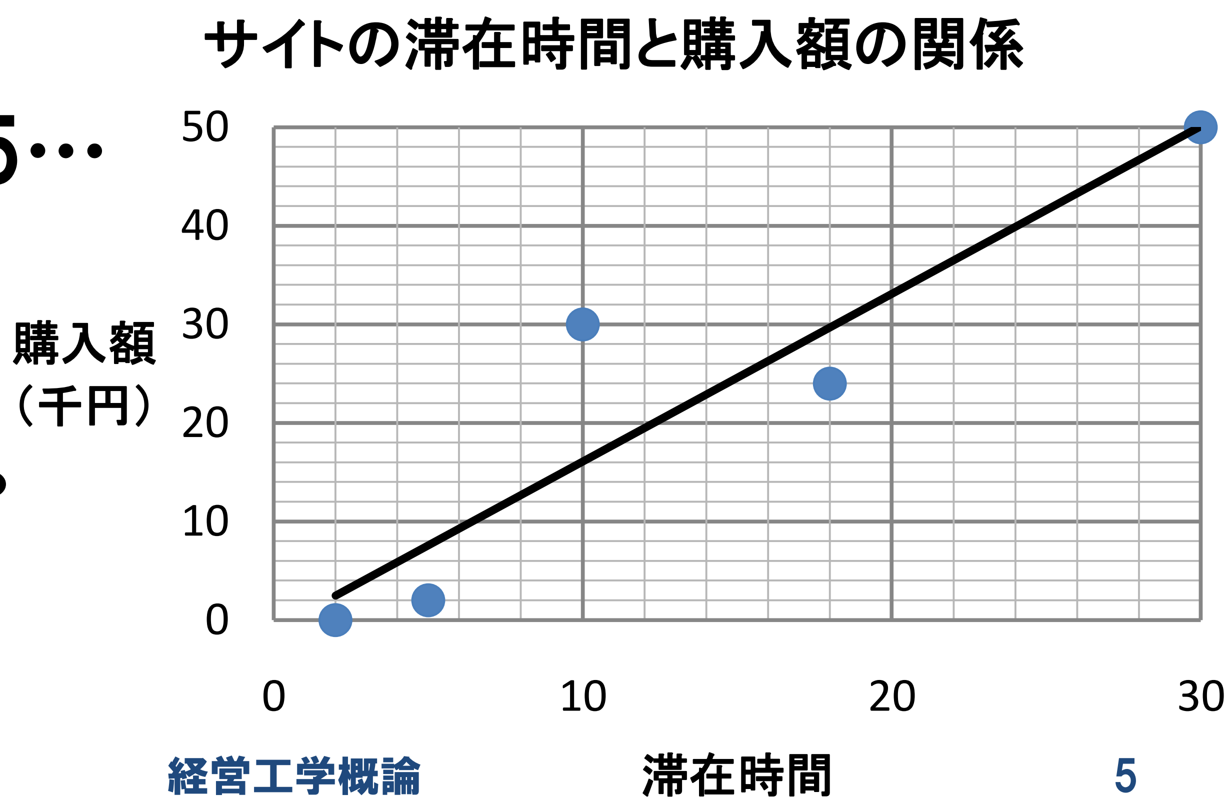
$$y = 1.70x - 0.91$$

解答

- 滞在時間x(分)と購入額y(千円)の関係：

$$y = 1.70x - 0.91$$

- x=20のとき
  y=33.105…
  よって、
  33(千円)
  と予測できる。



今回の内容

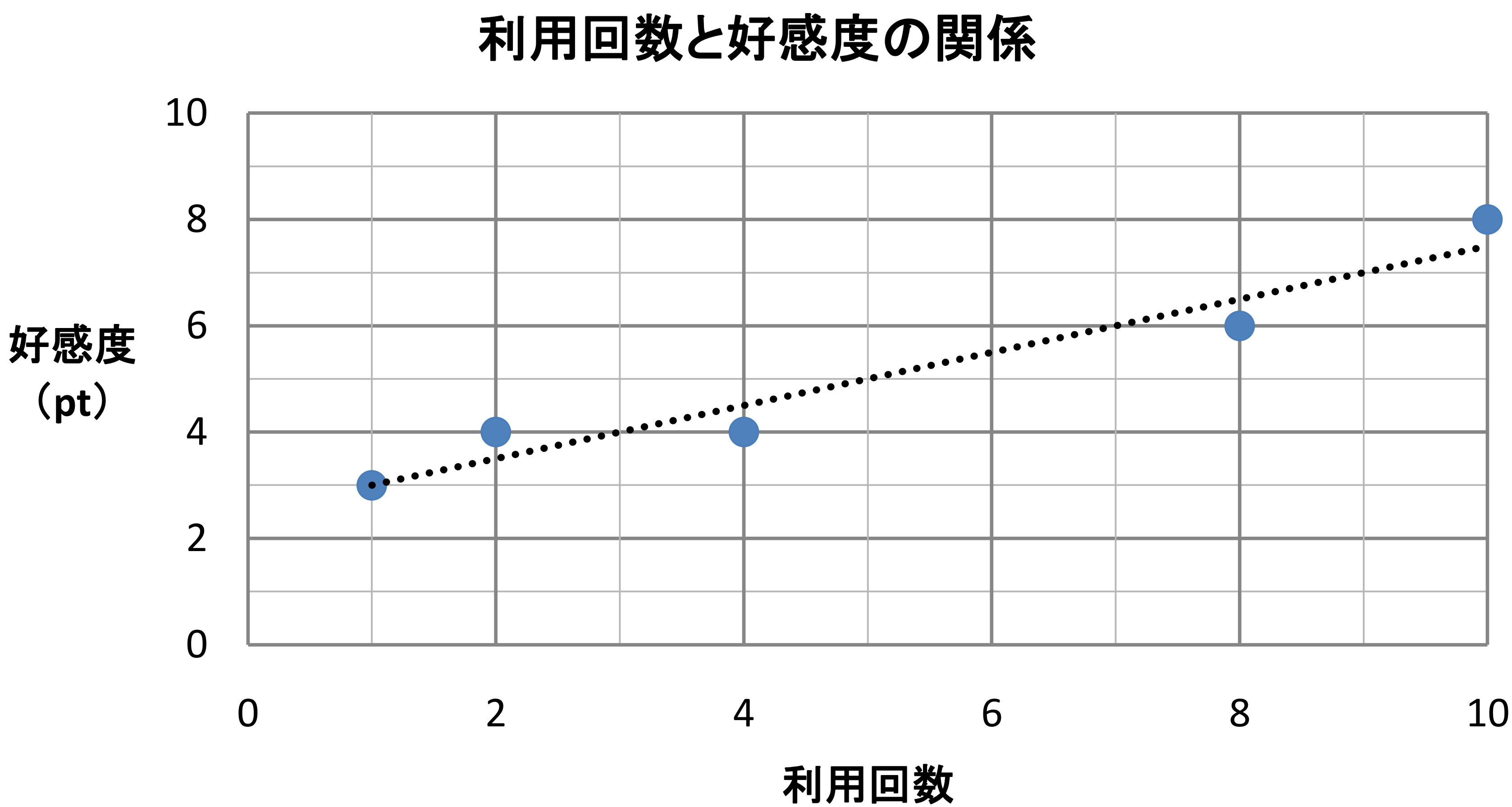
- 回帰分析による予測
  - 前回の補足
  - 最小自乗法

練習

- ある店舗における利用回数と好感度の関係について、下の表に示したデータがある。このとき、回帰分析によって両者の関係を調べよ。

利用回数	好感度（pt）
1	3
2	4
4	4
8	6
10	8

散布図を描く



表を使って計算する

x	y	$x_i - \bar{x}$	$y_i - \bar{y}$	$(x_i - \bar{x})(y_i - \bar{y})$	$(x_i - \bar{x})^2$
1	3	-4	-2	8	16
2	4	-3	-1	3	9
4	4	-1	-1	1	1
8	6	3	1	3	9
10	8	5	3	15	25
5	5			30	60

$$S_{xy} = \sum_{i=1}^{n}(x_i - \bar{x})(y_i - \bar{y}) = 30$$

$$S_{xx} = \sum_{i=1}^{n}(x_i - \bar{x})^2 = 60$$

回帰直線の方程式

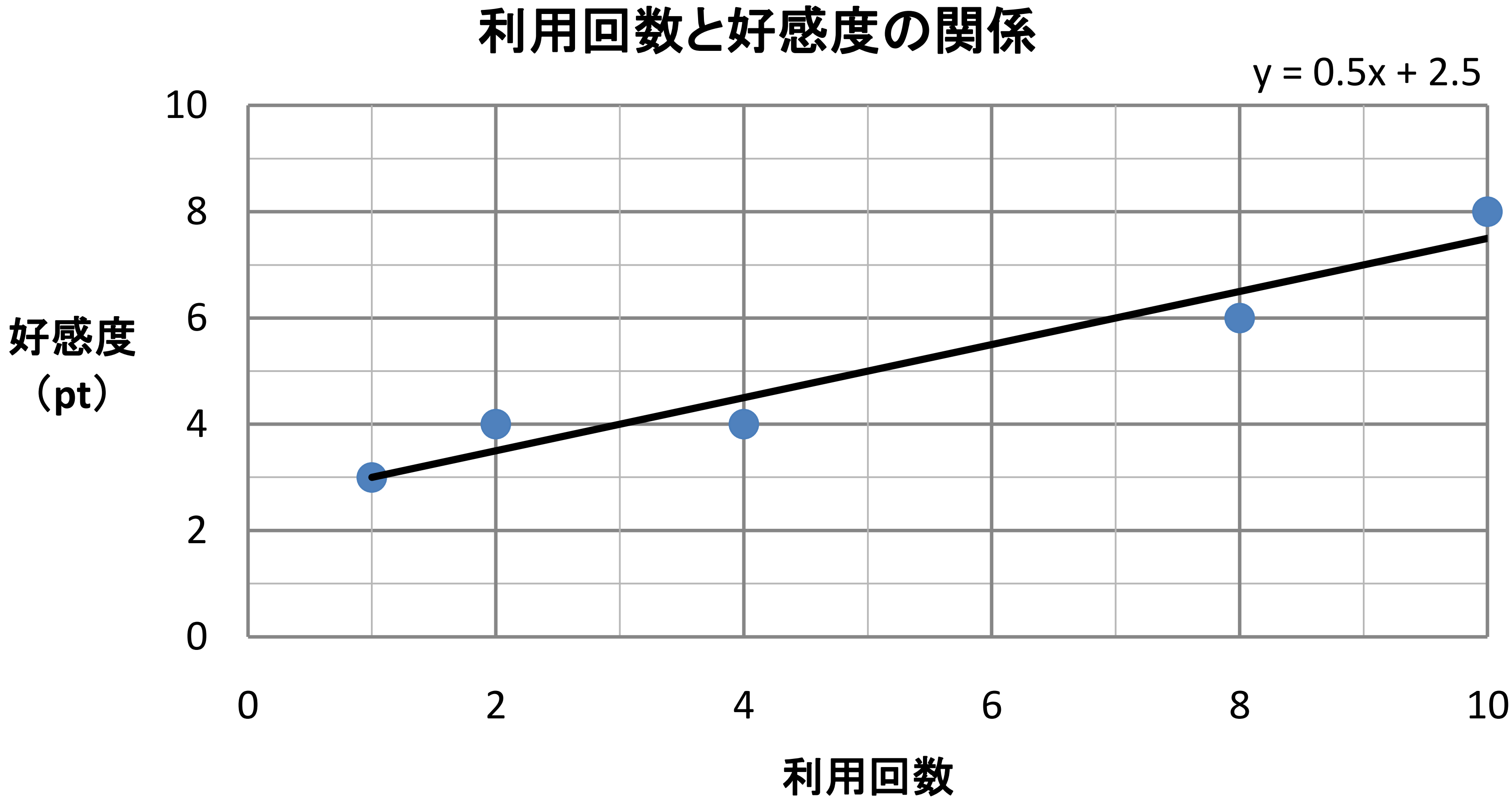
- 計算結果を代入する

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

$$y - 5 = \frac{30}{60}(x - 5)$$

よって $$y = \frac{1}{2}x + \frac{5}{2}$$

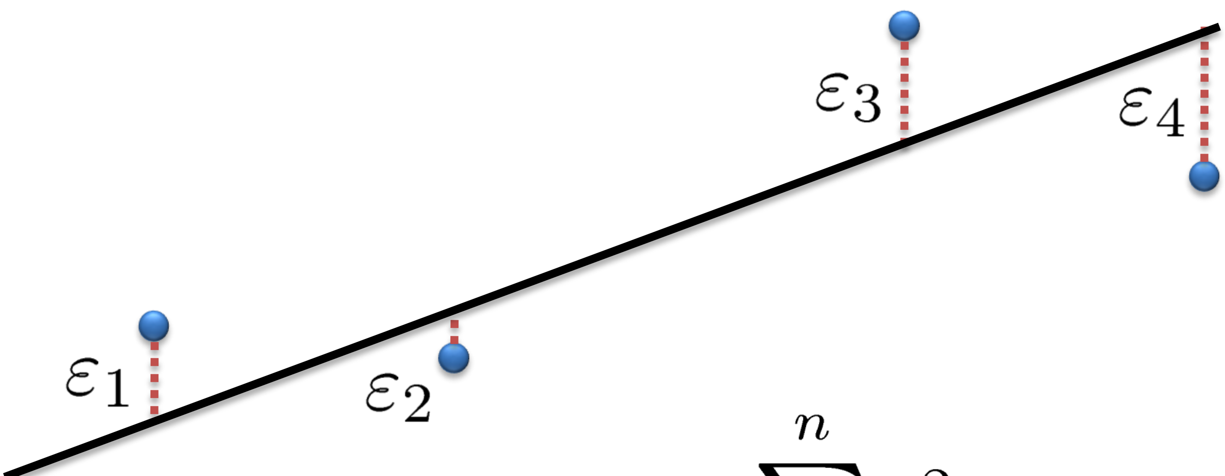
散布図に回帰直線を重ねる



最小自乗法 method of least squares

- 回帰分析において、誤差が最小になるように回帰式の係数を求めること

$$y = ax + b$$



$$\sum_{i=1}^{n} \varepsilon_i^2$$ を最小化する a と b を求めよ！

補題：2次関数の最小値

- 2次関数 $y = ax^2 + bx + c \ (a > 0)$ は

$x = -\dfrac{b}{2a}$ のとき、最小値 $-\dfrac{b^2}{4a} + c$ を取る。

補題の証明

- 平方完成すると、

$$
\begin{aligned}
y &= ax^2 + bx + c \\
&= a\left(x + \frac{b}{2a}\right)^2 - \frac{b^2}{4a} + c \\
&\geqq -\frac{b^2}{4a} + c
\end{aligned}
$$

- なお、等号成立は $x = -\dfrac{b}{2a}$ のときである。

定理：比例関係の場合

- 表のデータについて回帰分析すると、回帰直線の方程式は

$$y = \alpha x$$

である。ただし、

$$\alpha = \frac{\sum_{i=1}^{n} x_i y_i}{\sum_{i=1}^{n} x_i^2}$$

x	y
x_1	y_1
x_2	y_2
:	:
:	:
x_n	y_n

定理の証明

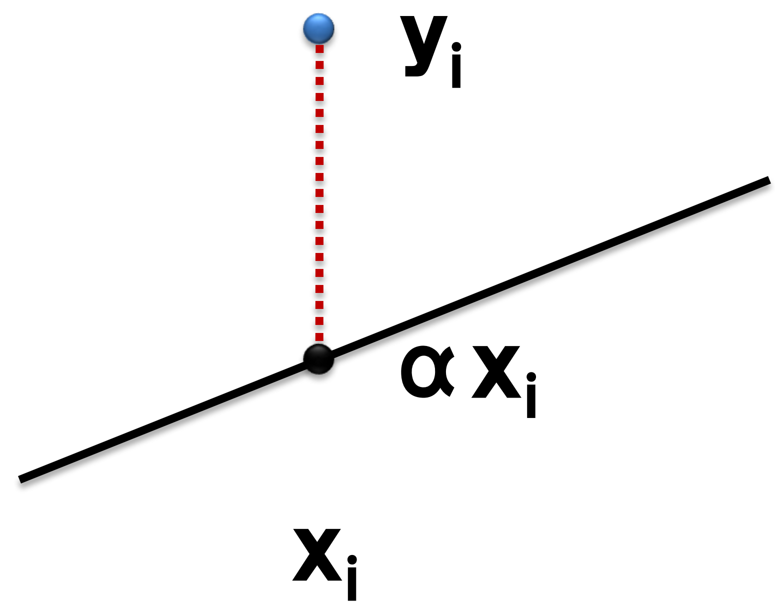
- 回帰直線の方程式を $y = \alpha x$ と仮定すると、データ（x_i、y_i）の回帰式からの誤差 ε_i は

$$\varepsilon_i = y_i - \alpha x_i$$

と表せる。そこで、

$$\sum_{i=1}^{n} \varepsilon_i^2$$

の最小化を目指せばよい。



証明の続き

- 誤差の自乗和をαについて整理すると、

$$\sum_{i=1}^{n}(y_i - \alpha x_i)^2$$
$$= \left(\sum_{i=1}^{n} x_i^2\right)\alpha^2 - 2\left(\sum_{i=1}^{n} x_i y_i\right)\alpha + \sum_{i=1}^{n} y_i^2$$

- 補題から

$\alpha = \sum_{i=1}^{n} x_i y_i / \sum_{i=1}^{n} x_i^2$ **のとき最小となる。**

Σの性質

- 線形性
  - 和の展開

$$\sum_{i=1}^{n}(x_i + y_i) = \sum_{i=1}^{n} x_i + \sum_{i=1}^{n} y_i$$

  - 定数倍

$$\sum_{i=1}^{n} \alpha x_i = \alpha \sum_{i=1}^{n} x_i$$

まとめ

- 回帰分析による予測
  - ２次関数の最小値
  - ガウスの最小自乗法

提出課題

- 最小自乗法について説明せよ。

次回の予定

- 第7回　回帰分析による予測（続き）
  寄与率 R^2
  - 日時：2008年　6月　6日(金）4時限目
  - 場所：845教室