13章　回帰分析の基礎

2つ以上の変数についての関係を見る.

1つの変数を結果，その他の変数を原因として，因果関係を説明しようとするもの.

厳密な意味での因果関係ではない。

例（因果・相関関係等）

- 勤務年数が長ければ，年間給与は上がる.
- 景気が良くなれば，株価は上がる
- 父親の身長が高ければ，子供の身長も高い.
- 価格が低下すれば需要が増える.
- 自身の兄弟数が多いと，育てる子供の数も多い.
- サッカー人気が上がると，野球人気が落ちる.

例（因果・相関関係等）

- 円が高くなると，輸出不振になる.
- フォアボールが多いと失点が増える.
- 親の年収が高いと，子供の成績もよい.
- 投手の防御率が低いと，勝利数も多い.
- ルックスと性格の関係
- 天候と売り上げ
- 与四球数が多いチームの勝率は低い
- 不景気だとクマのキャラクターの売り上げが上がる.

例（因果・相関関係等）

- B級グルメと地域経済
- 血液型と性格
- 美人と生涯所得の関係
- トヨタ株価と日経平均
- 勉強時間とテスト結果
- 映画の興行収入と作品としての評価
- シュート数と得点

例（因果・相関関係等）

- 東京ディズニーランドとUSJの入場者数の関係.
- 海外旅行者と国内旅行者の数の関係
- 打率と出塁率の関係
- 食事の取り方と体重の関係
- 顔と性格の関係
- ボール支配率と勝率の関係

例（因果・相関関係等）

- 気温とアイスの売り上げ
- ファーストサーブの成功率と勝率
- 親の寿命と子の寿命
- 親の結婚年齢と子の結婚年齢
- 出席率と成績
- 勉強時間と成績の関係
- 煙草の値段と喫煙率

例（因果・相関関係等）

- CDの売り上げと着うたフルのダウンロード数
- 月収とギャンブル収支
- 喫煙者と職業
- 非正規労働者の増減と企業数の増減
- 天候と外出の関係

ここで勉強すること

- 散布図と相関係数
- 最小2乗法と回帰直線
- 決定係数
- 重回帰分析

株価収益率データの標本は？母集団は？

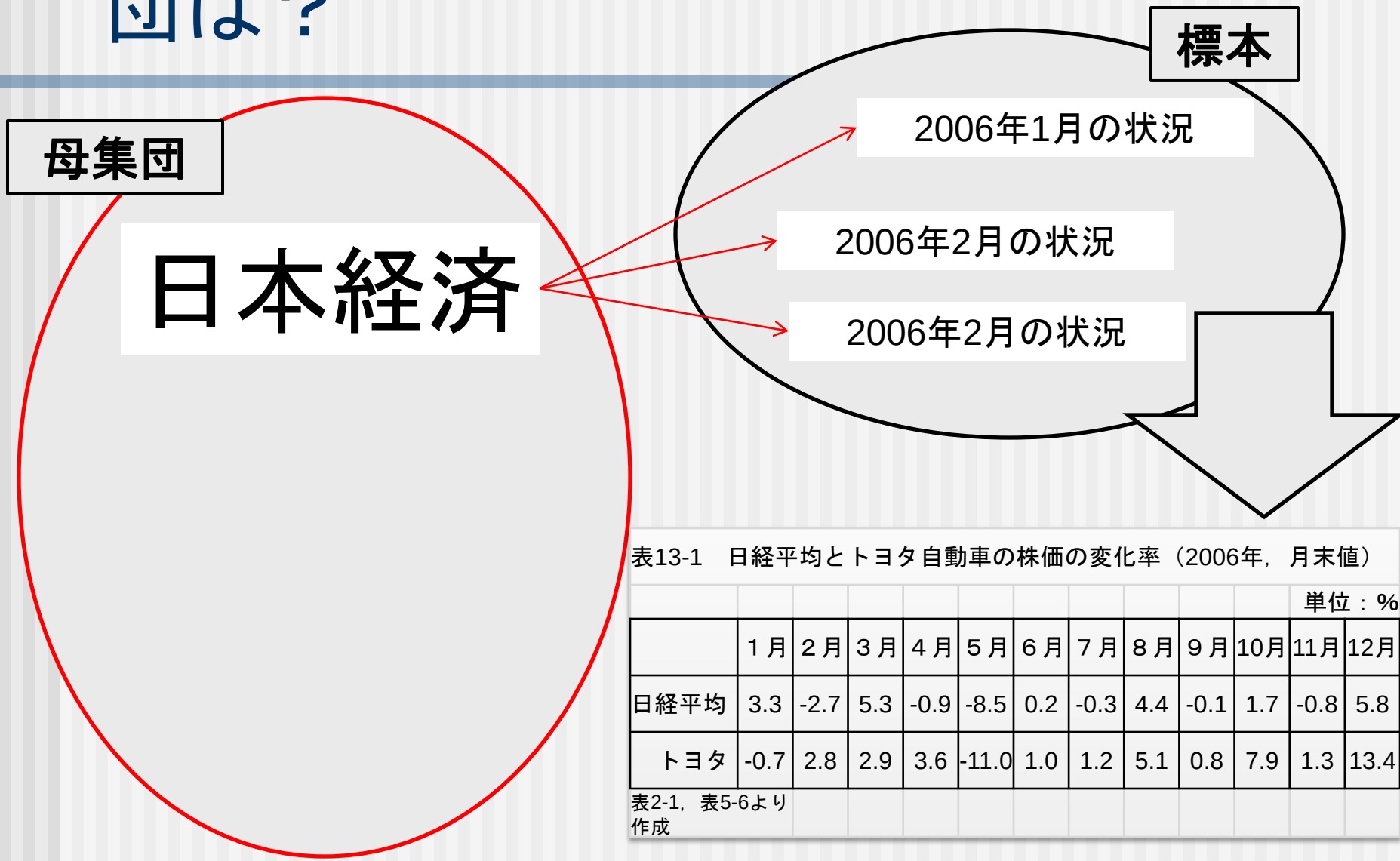


表13-1　日経平均とトヨタ自動車の株価の変化率（2006年，月末値）

単位：%

	1月	2月	3月	4月	5月	6月	7月	8月	9月	10月	11月	12月
日経平均	3.3	-2.7	5.3	-0.9	-8.5	0.2	-0.3	4.4	-0.1	1.7	-0.8	5.8
トヨタ	-0.7	2.8	2.9	3.6	-11.0	1.0	1.2	5.1	0.8	7.9	1.3	13.4

表2-1，表5-6より作成

1．散布図と相関係数

表13-1　日経平均とトヨタ自動車の株価の変化率（2006年，月末値）

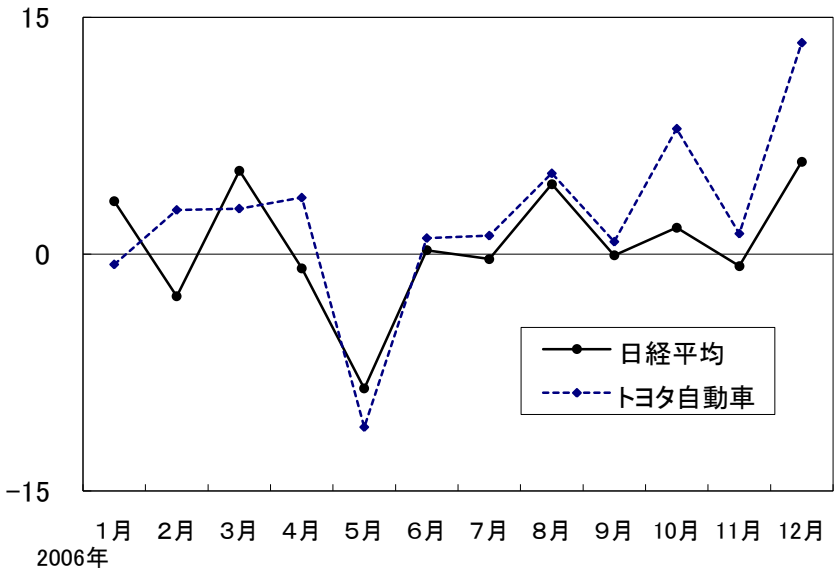
単位：%

	1月	2月	3月	4月	5月	6月	7月	8月	9月	10月	11月	12月
日経平均 x_i	3.3	-2.7	5.3	-0.9	-8.5	0.2	-0.3	4.4	-0.1	1.7	-0.8	5.8
トヨタ y_i	-0.7	2.8	2.9	3.6	-11.0	1.0	1.2	5.1	0.8	7.9	1.3	13.4

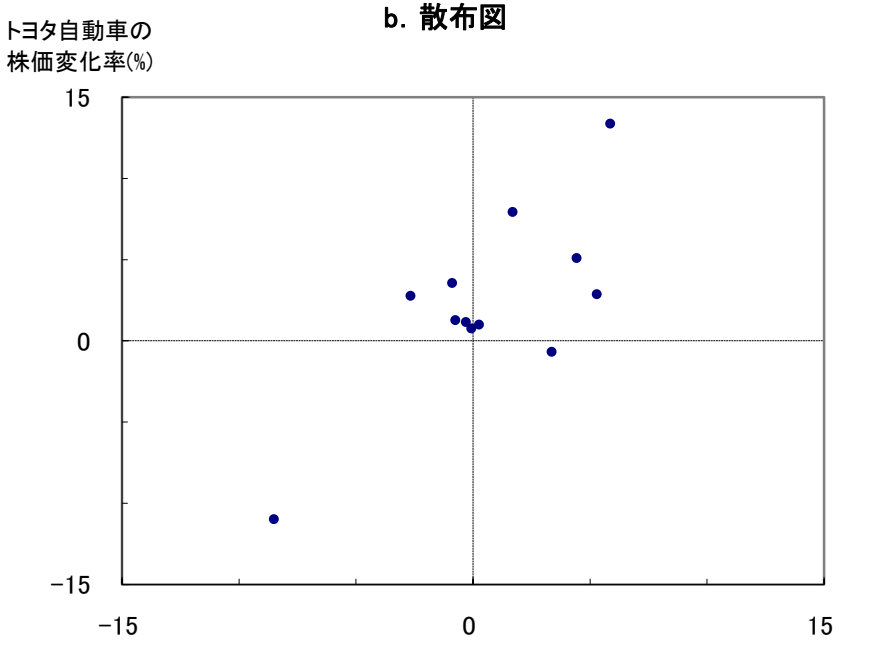
表2-1，表5-6より作成

図13-1　日経平均とトヨタ自動車の株価変化率のグラフ（2006年）

a. 日経平均・トヨタ自動車の月次変化率の推移

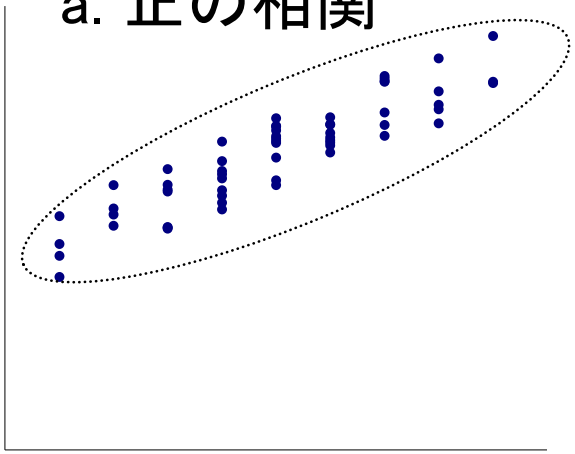


b. 散布図

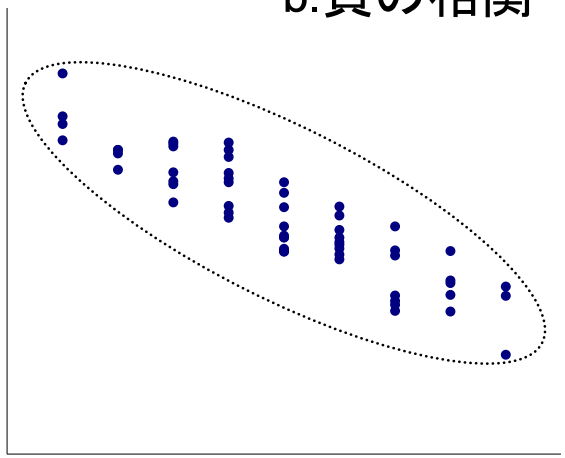


散布図と相関(直観的解釈)

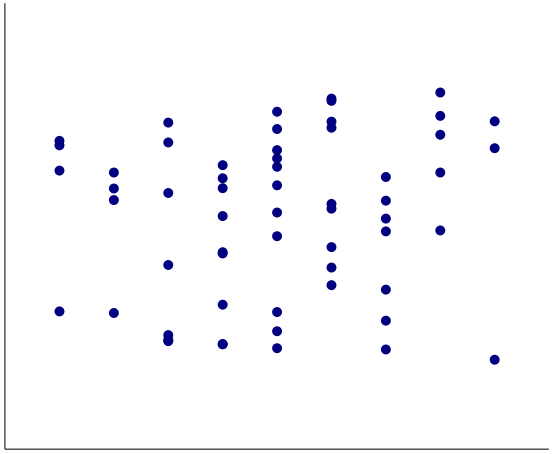
a. 正の相関



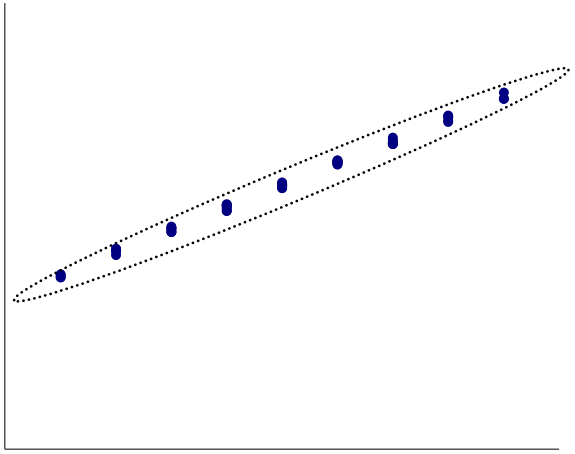
b.負の相関



c. 無相関

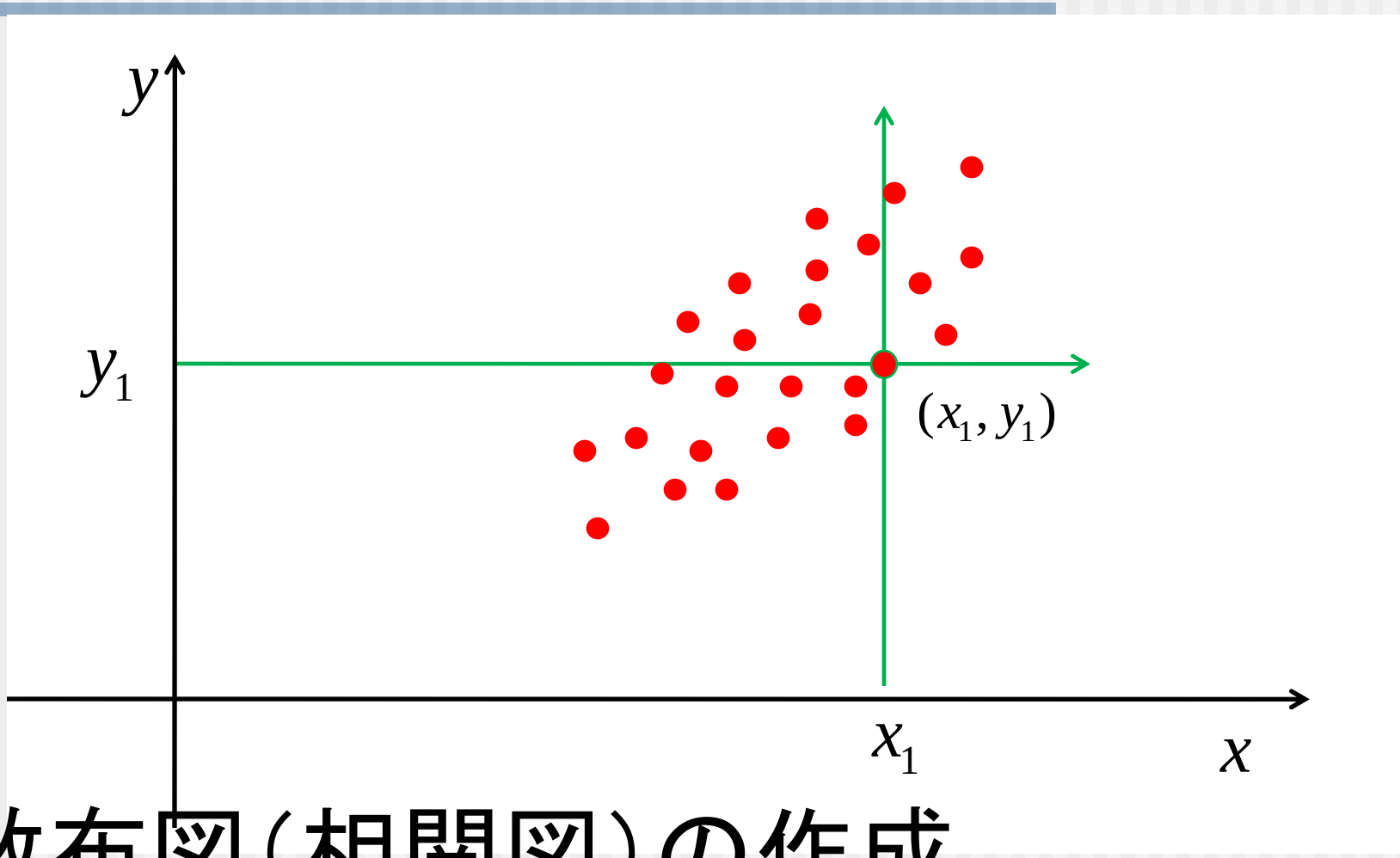


d. 強い正の相関



$(x_i - \bar{x})(y_i - \bar{y})$ の平均である.

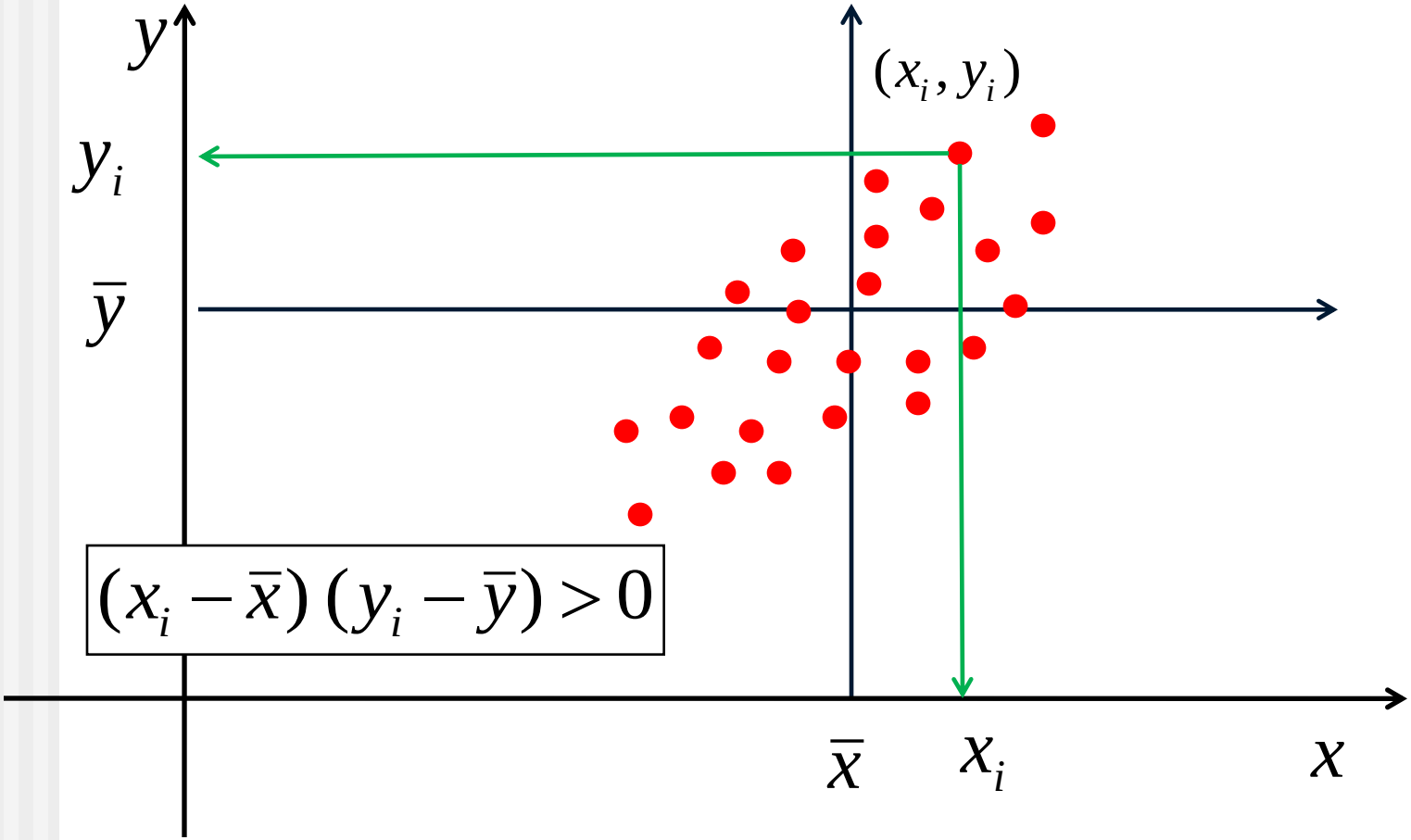
共分散とは何か



散布図(相関図)の作成

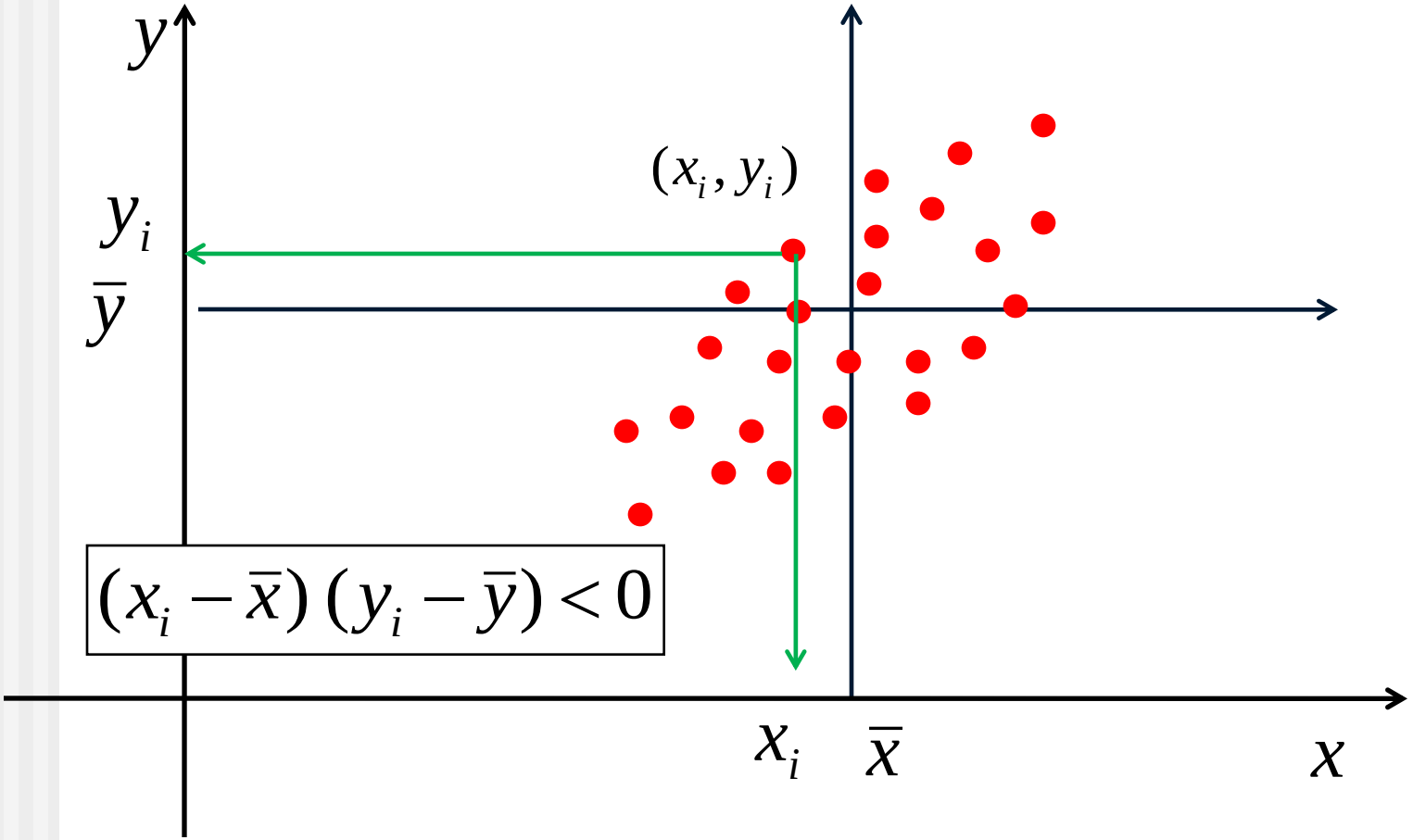
$(x_i - \bar{x})(y_i - \bar{y})$ の平均である.

共分散とは何か（第Ⅰ象限）



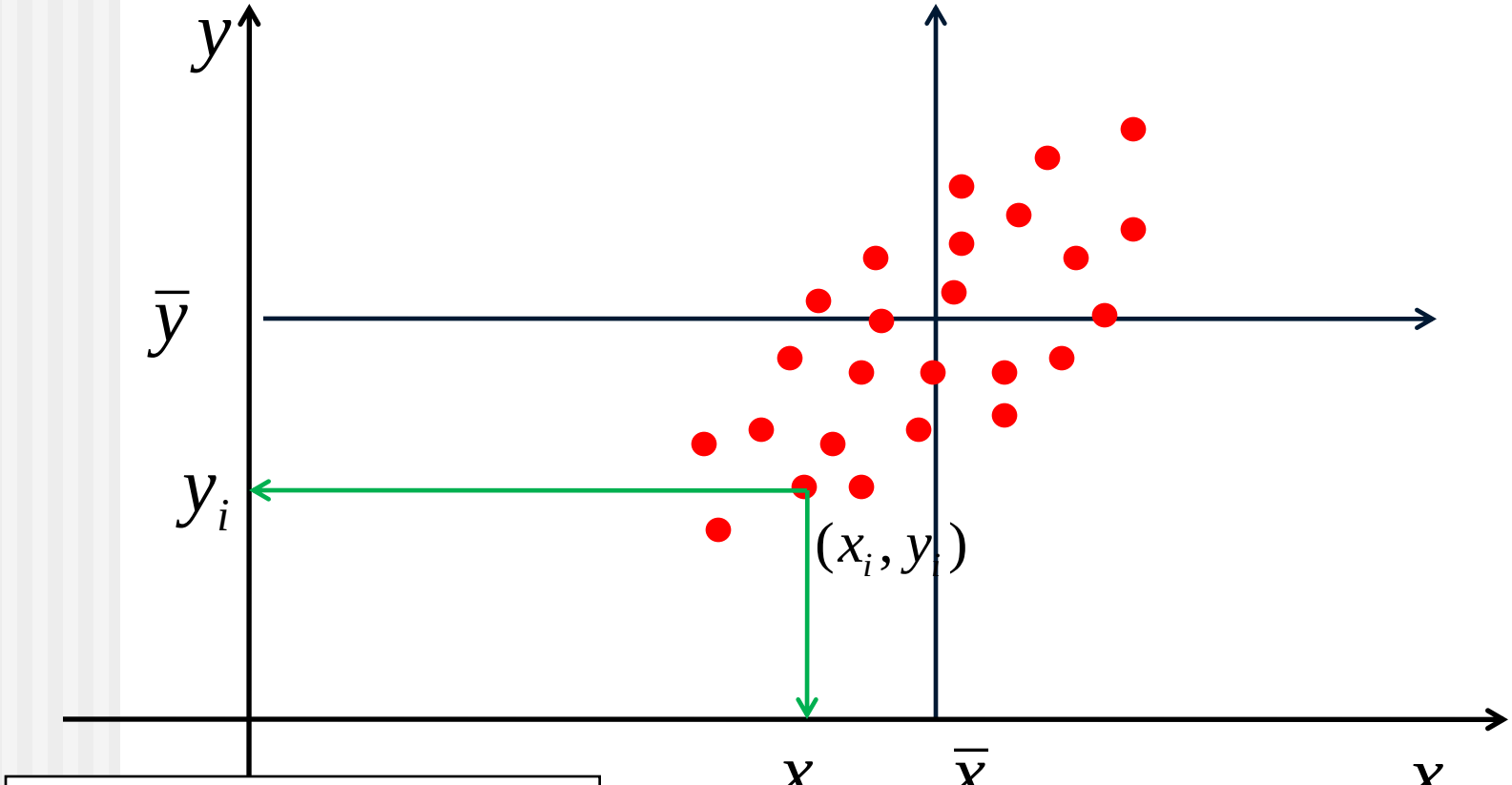
$(x_i - \bar{x})(y_i - \bar{y})$ の平均である.

共分散とは何か（第Ⅱ象限）



$(x_i - \bar{x})(y_i - \bar{y})$ の平均である.

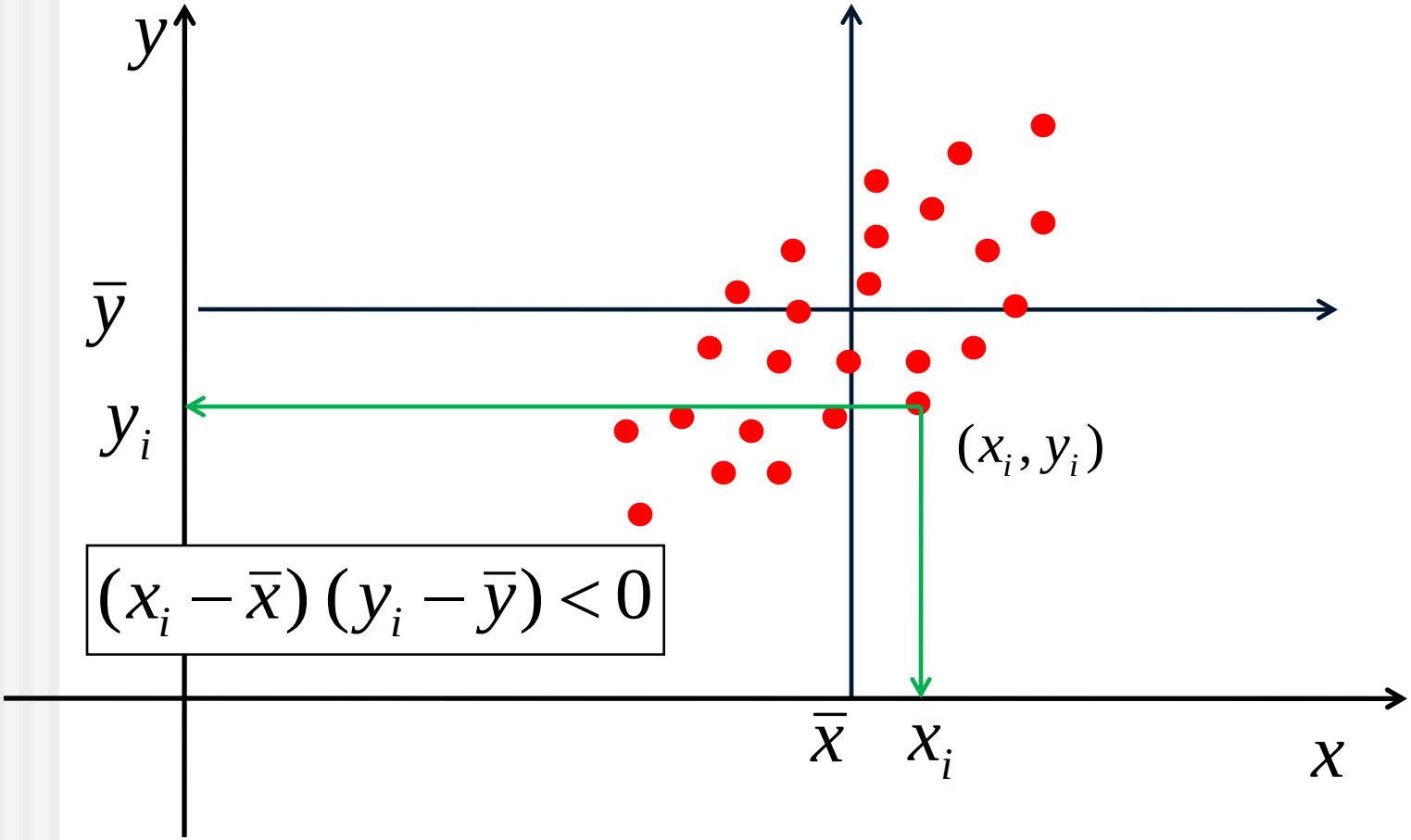
共分散とは何か（第Ⅲ象限）



$$(x_i - \bar{x})\,(y_i - \bar{y}) > 0$$

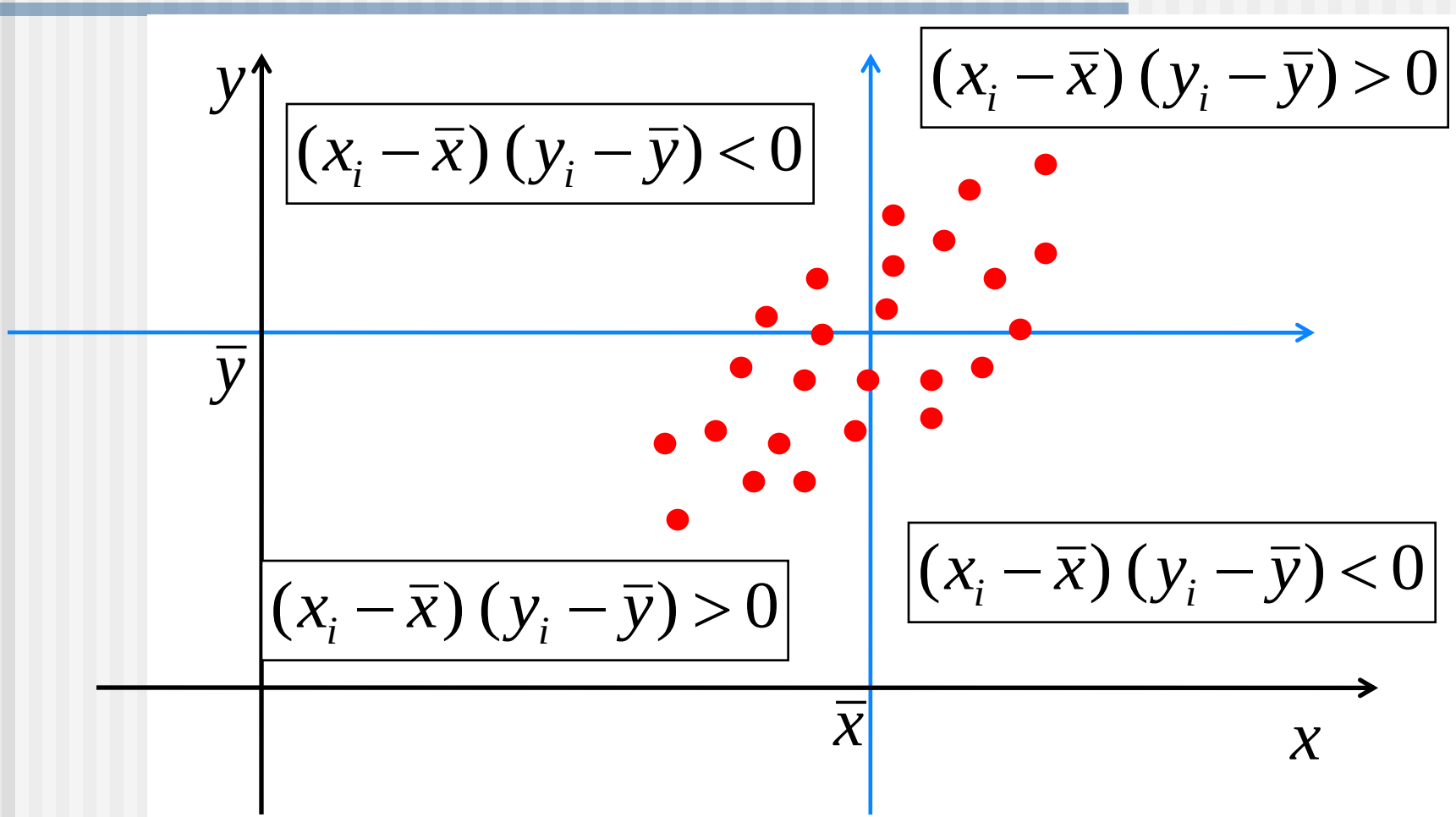
$(x_i - \bar{x})(y_i - \bar{y})$ の平均である.

共分散とは何か（第Ⅳ象限）



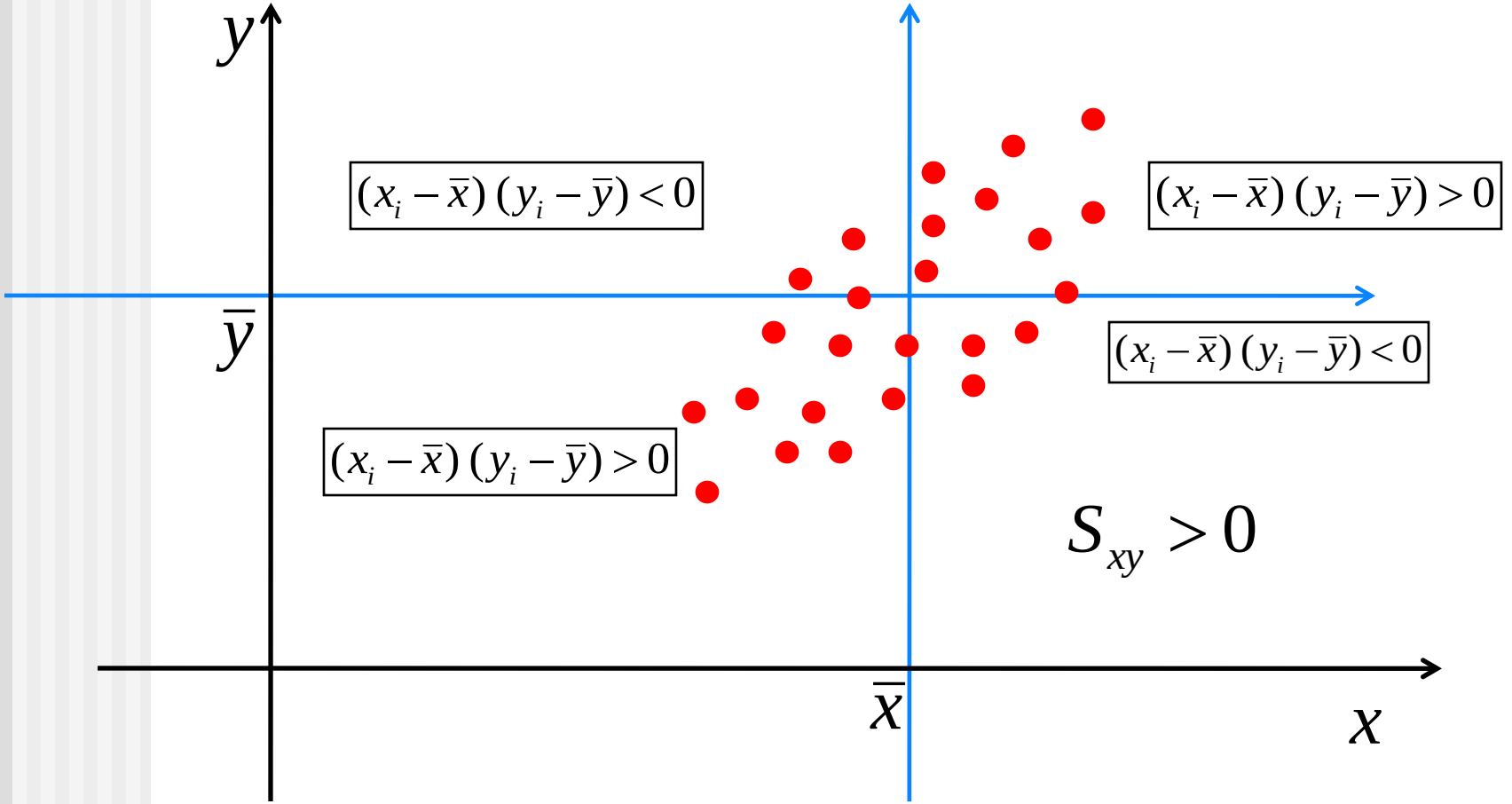
$(x_i - \bar{x})(y_i - \bar{y})$ の平均である.

共分散とは何か（第Ⅰ~Ⅳ象限）



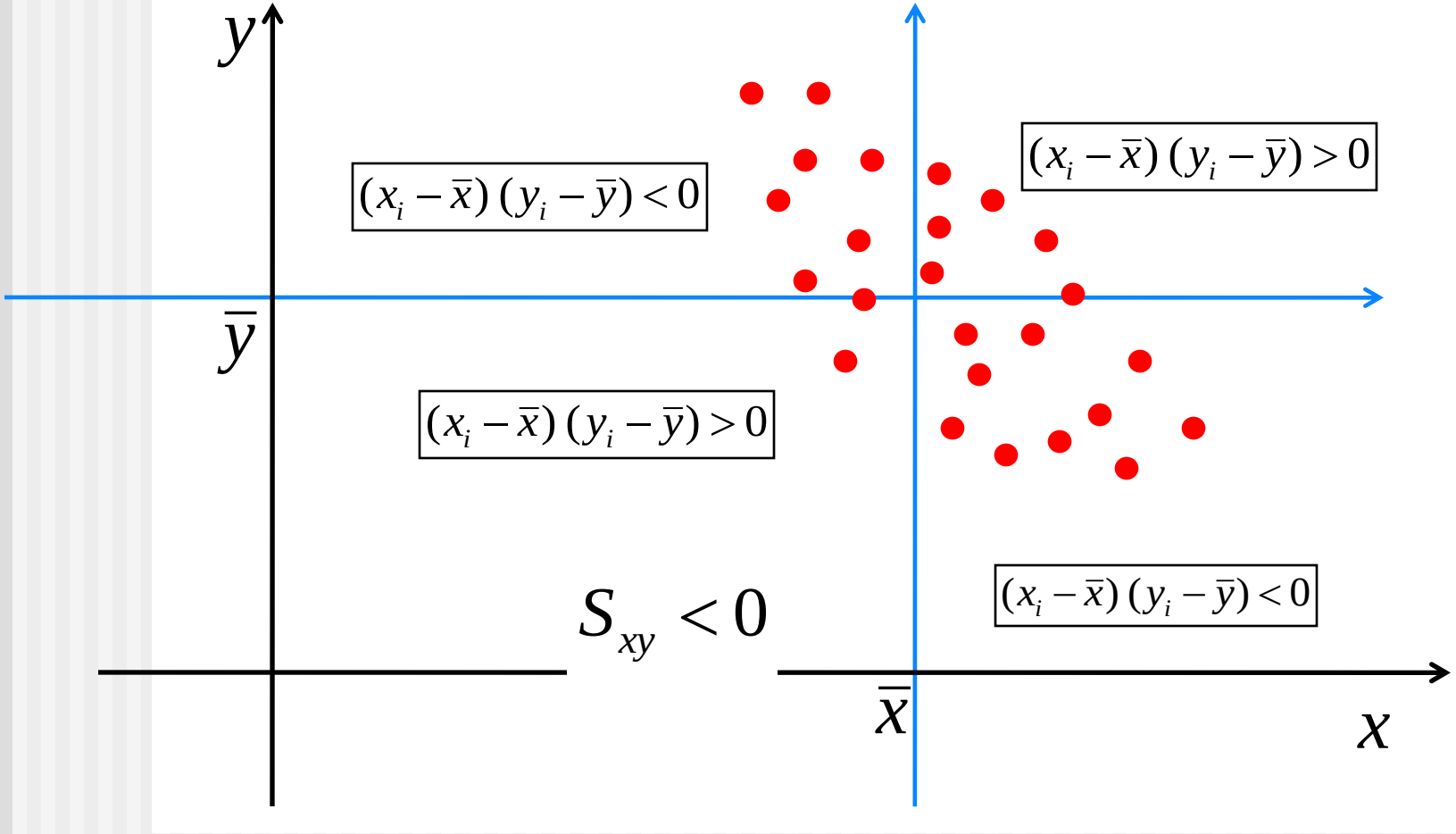
共分散の符号

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



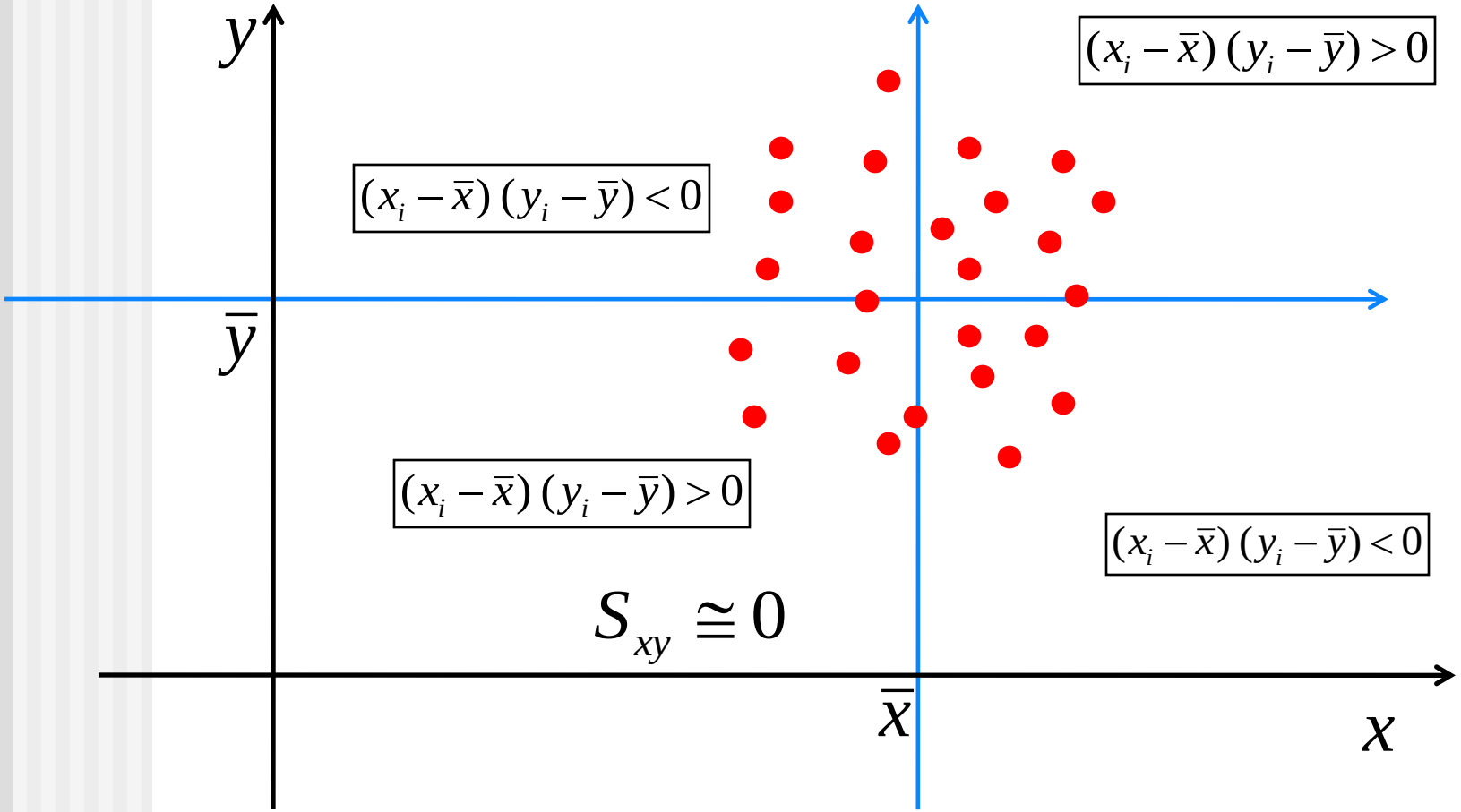
共分散の符号

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



共分散の符号

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



相関係数

$$r_{xy} = \frac{S_{xy}}{S_x S_y} =$$

$$\frac{\frac{1}{n}\sum (x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\frac{1}{n}\sum (x_i - \bar{x})^2}\sqrt{\frac{1}{n}\sum (y_i - \bar{y})^2}}$$

XとYの共分散

Xの標準偏差

Yの標準偏差

相関係数

$$r_{xy} = \frac{S_{xy}}{S_x S_y} = \frac{\frac{1}{n}\sum (x_i - \bar{x})(y_i - \bar{y})}{S_x S_y}$$

$$= \frac{1}{n}\sum \frac{(x_i - \bar{x})}{S_x} \frac{(y_i - \bar{y})}{S_y}$$

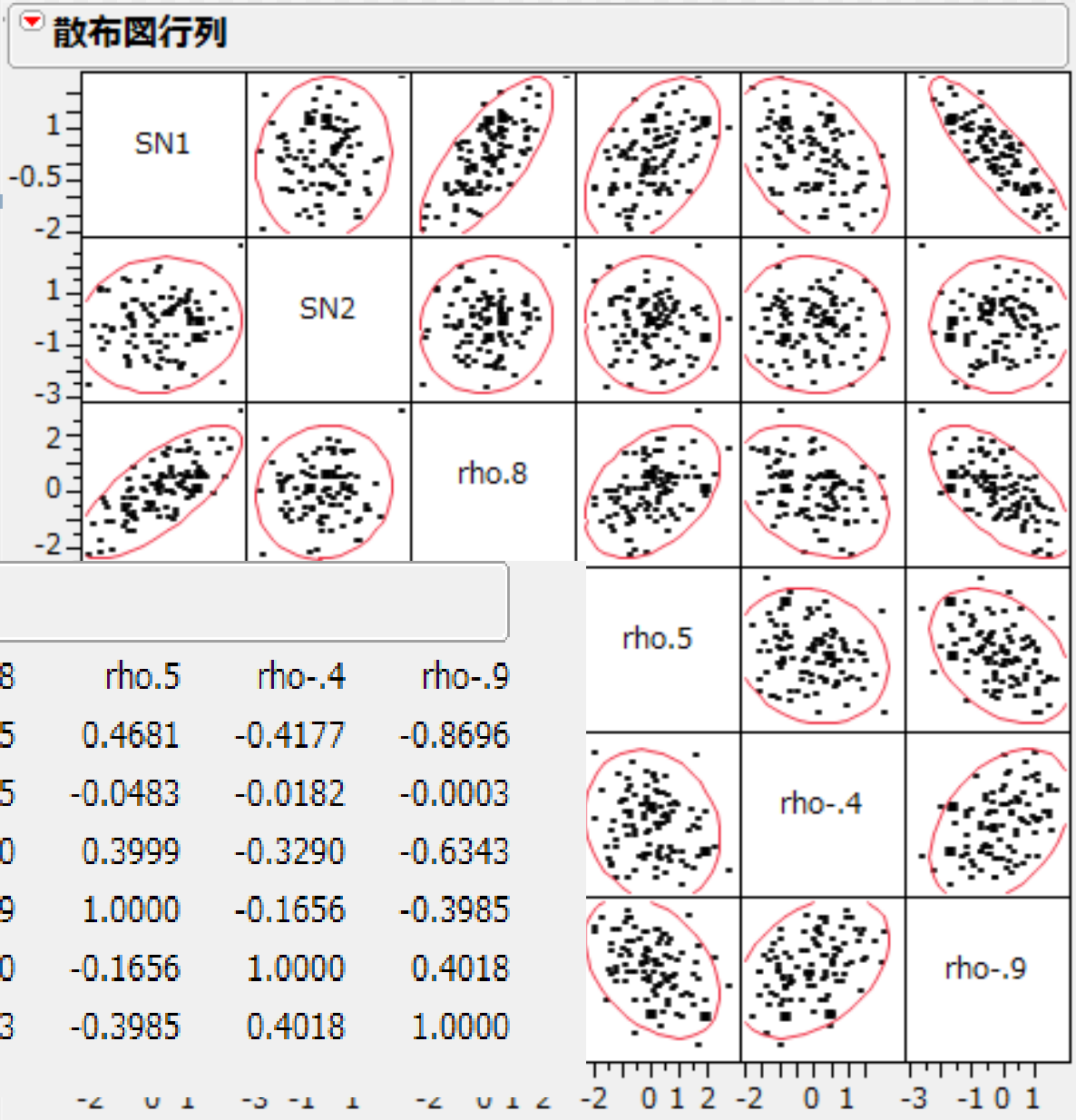
平均０
標準偏差１

平均０
標準偏差１

相関係数の性質

- 最大1，最小−1の値をとる.
- 相関係数の絶対値が1に近い程，相関は強いことが分かる.
- 相関係数の絶対値が1になるのは，データ点が一直線上に位置するときのみである.
- 相関係数は，直線的な関係の強さをはかるもので，曲線的な関係を調べるのには向いていない.

相関係数と散布図の関係



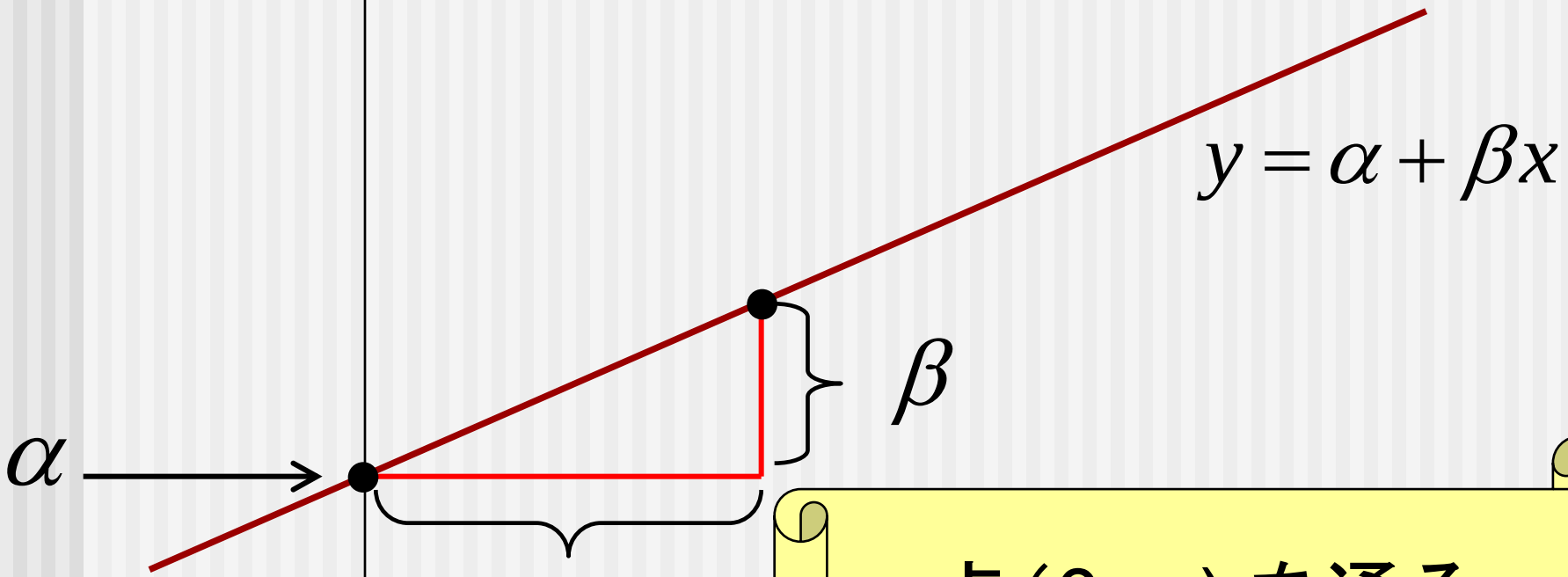
相関

	SN1	SN2	rho.8	rho.5	rho-.4	rho-.9
SN1	1.0000	0.0843	0.7405	0.4681	-0.4177	-0.8696
SN2	0.0843	1.0000	0.0725	-0.0483	-0.0182	-0.0003
rho.8	0.7405	0.0725	1.0000	0.3999	-0.3290	-0.6343
rho.5	0.4681	-0.0483	0.3999	1.0000	-0.1656	-0.3985
rho-.4	-0.4177	-0.0182	-0.3290	-0.1656	1.0000	0.4018
rho-.9	-0.8696	-0.0003	-0.6343	-0.3985	0.4018	1.0000

2　最小2乗法と回帰直線

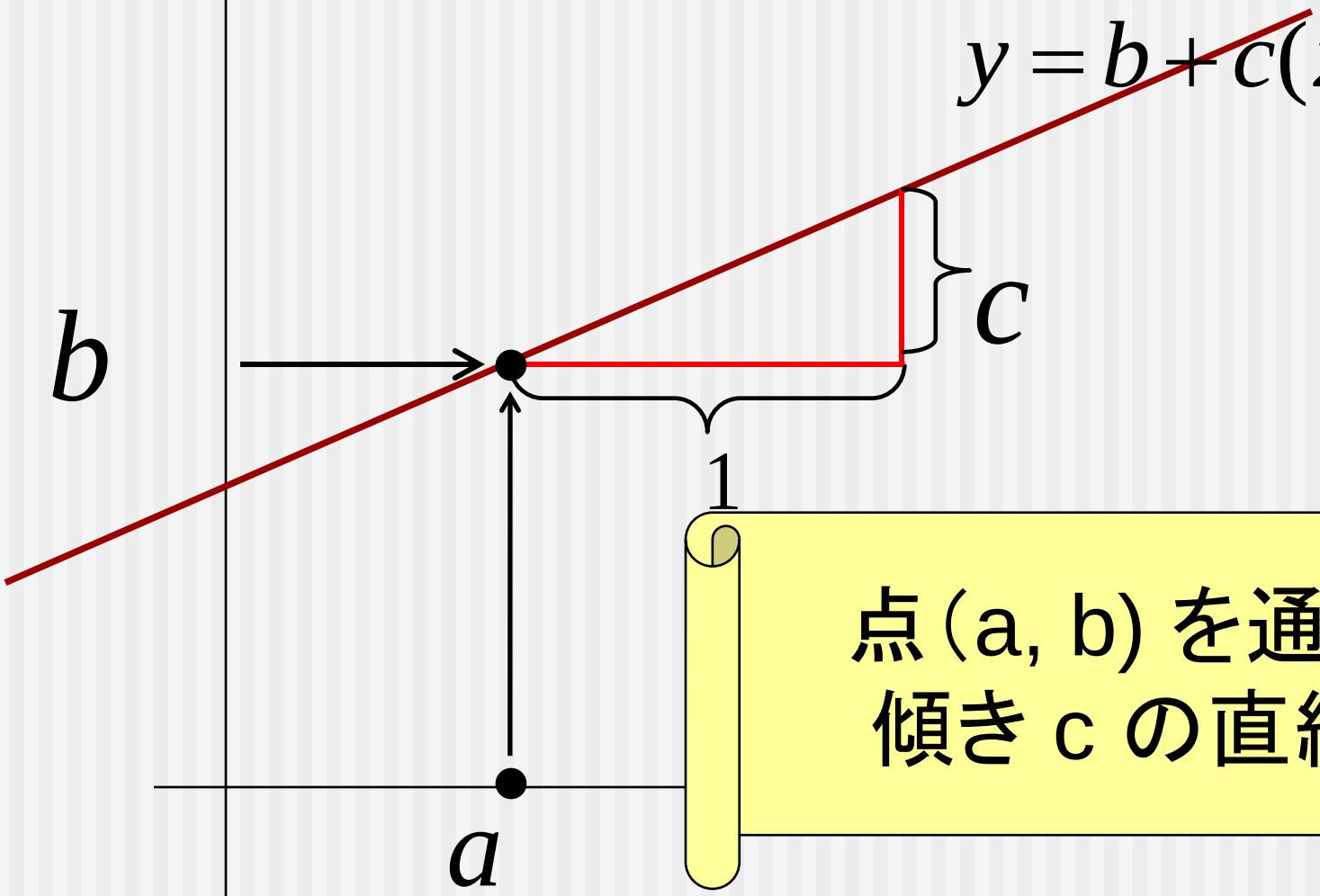
- これまで，2つの変数間の関係の深さについて考えてきた（相関係数）
- 次に，変数に役割を与え，一方の変数を用いて他方の変数を説明することを考える.
- この関係は，必ずしも，因果関係でなくてもよい.

直線　y＝α＋βx とは？



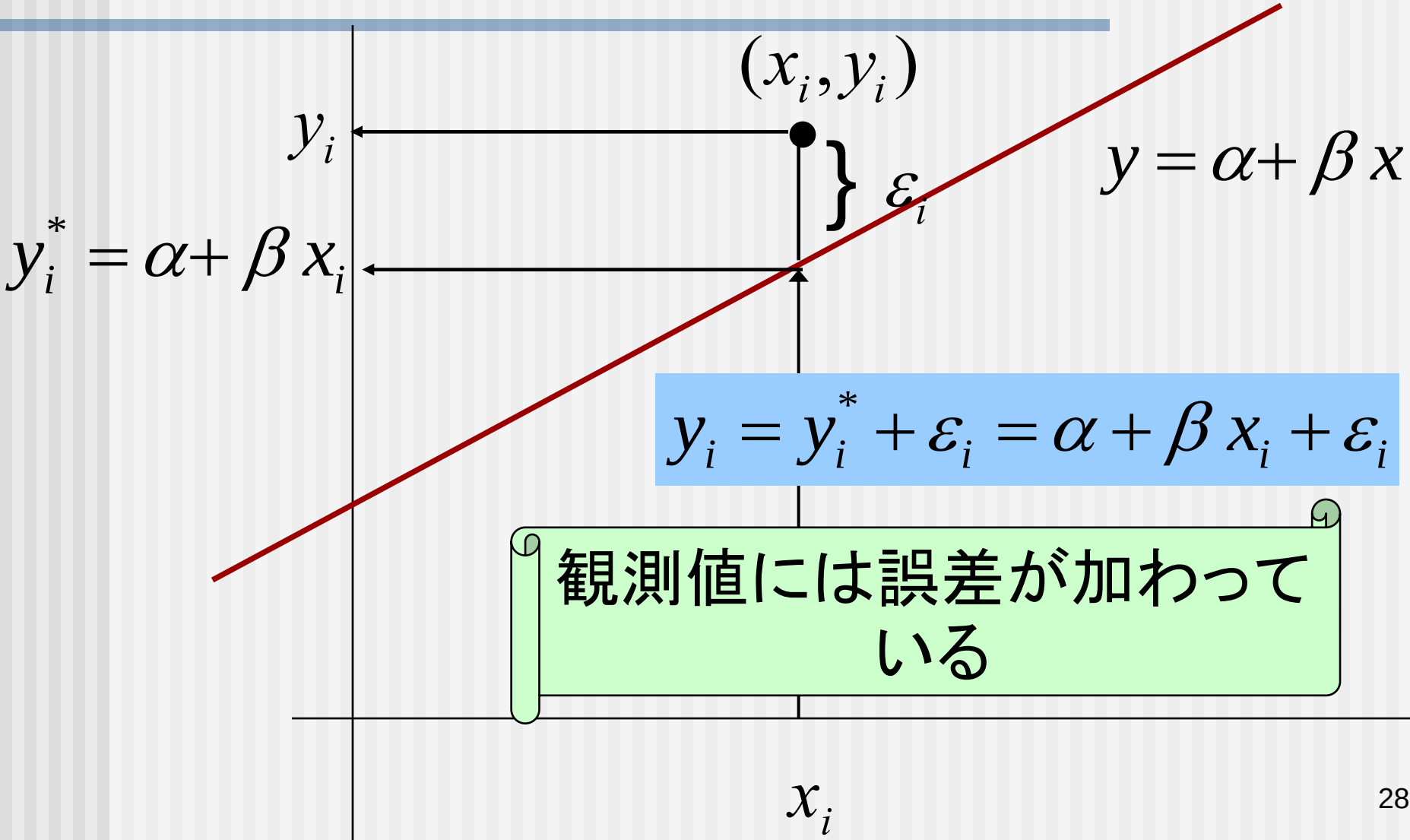
点 (0, α) を通る
傾き β の直線

直線　y＝b＋c(x-a) とは？

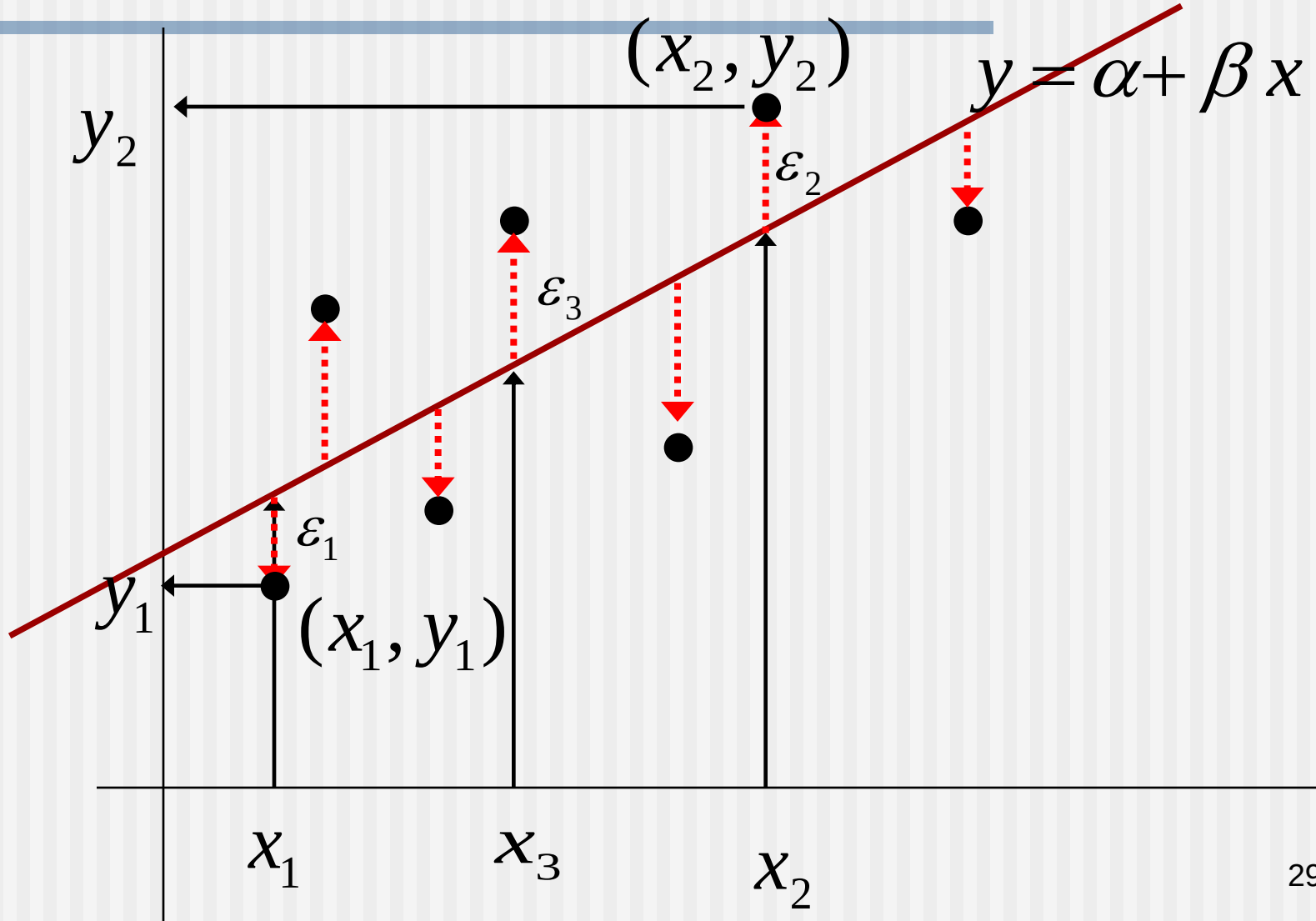


点（a, b) を通る
傾き c の直線

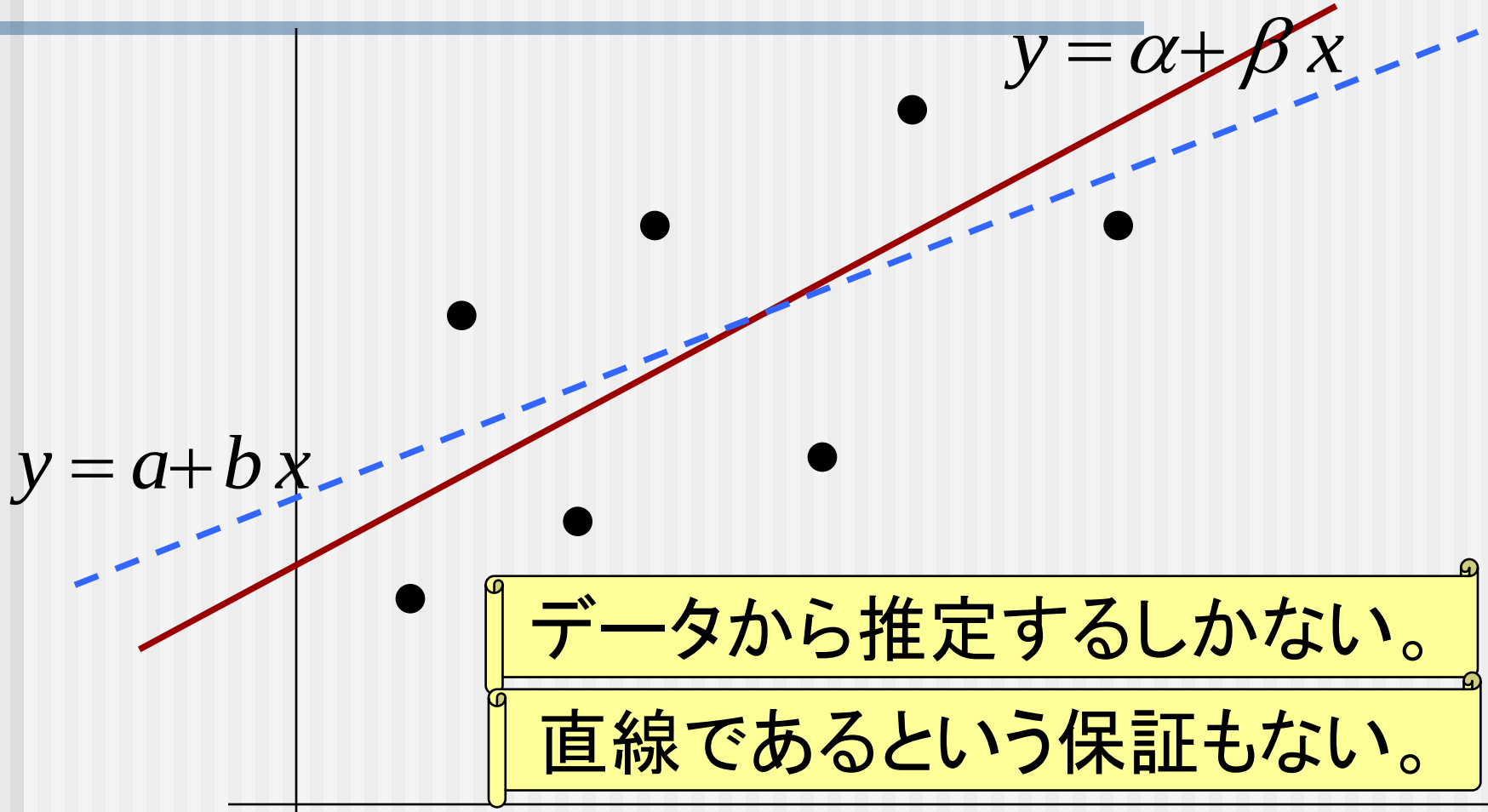
直線　y＝α＋βx を回帰直線と考えるとき



直線　y＝α＋βx を回帰直線と考えるときの観測値の得られ方



回帰直線　y＝α＋βx は未知である



データから推定するしかない。

直線であるという保証もない。

直線 y＝α＋βx の推定法

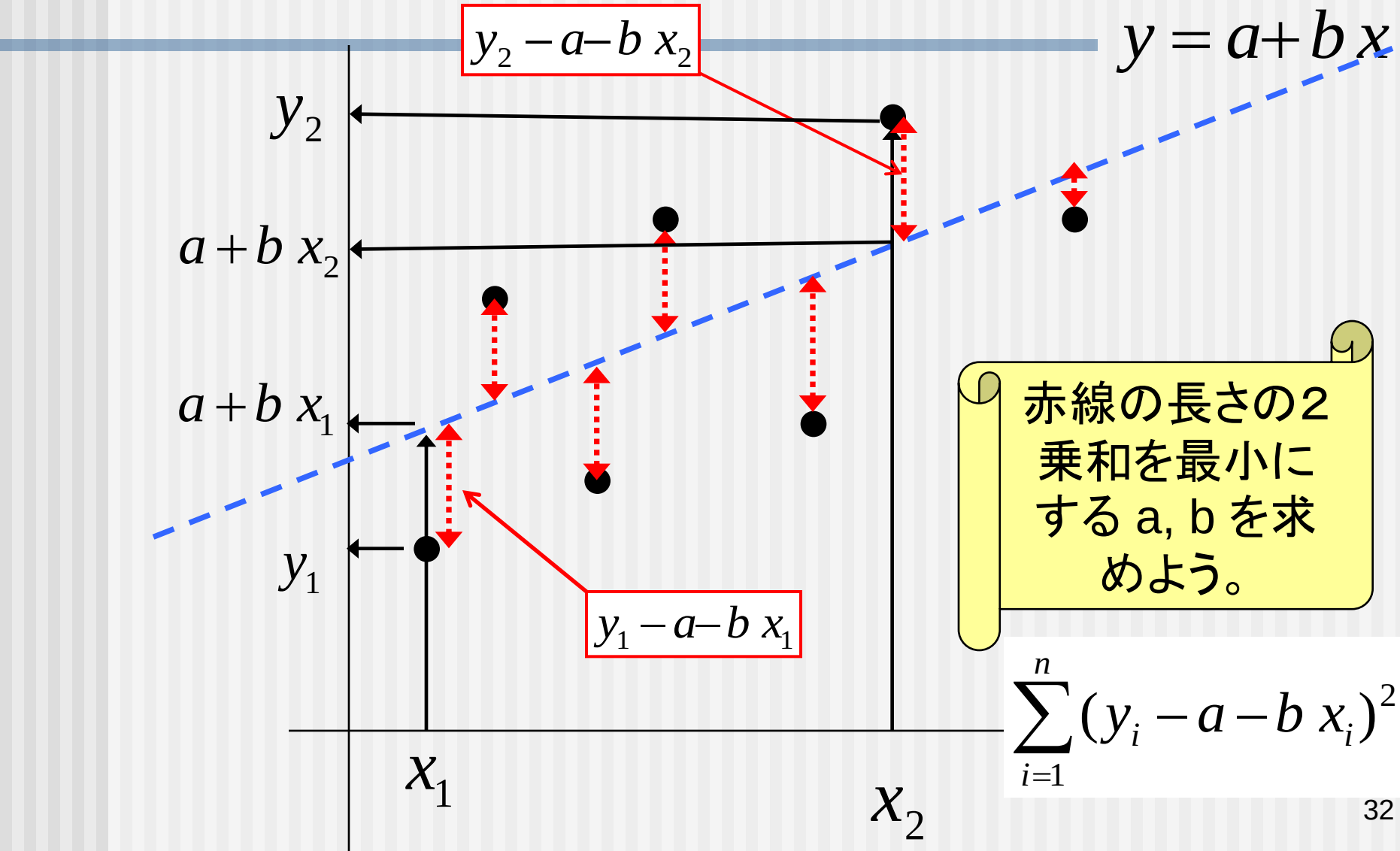
$$(1.5)\frac{1}{n}\sum_{i=1}^{n}\varepsilon_i^2=\frac{1}{n}\sum_{i=1}^{n}(y_i-\alpha-\beta\,x_i)^2$$

- 上の式を最小にするように，αとβを決める.
- 最小2乗法により決めるとも言う.

$$(1.6)SSE=\sum_{i=1}^{n}\varepsilon_i^2=\sum_{i=1}^{n}(y_i-\alpha-\beta\,x_i)^2$$

- 上の Sum of Squared Errors を最少化するとも考えられる.

回帰直線　y＝α＋βx の推定法（図解）



赤線の長さの2乗和を最小にするa, bを求めよう。

$$\sum_{i=1}^{n} (y_i - a - b\,x_i)^2$$

最小2乗推定値の公式
結果を先に示す

$$(1.7)\ b = \frac{\sum_{i=1}^{n}(x_i - \bar{x})(y_i - \bar{y})}{\sum_{i=1}^{n}(x_i - \bar{x})^2} = \frac{s_{xy}}{s_{xx}}$$

$$(1.8)\ a = \bar{y} - b\,\bar{x}$$

ここで,

$$s_{xy} = \frac{1}{n}\sum_{i=1}^{n}(x_i - \bar{x})(y_i - \bar{y}), \quad s_{xx} = \frac{1}{n}\sum_{i=1}^{n}(x_i - \bar{x})^2$$

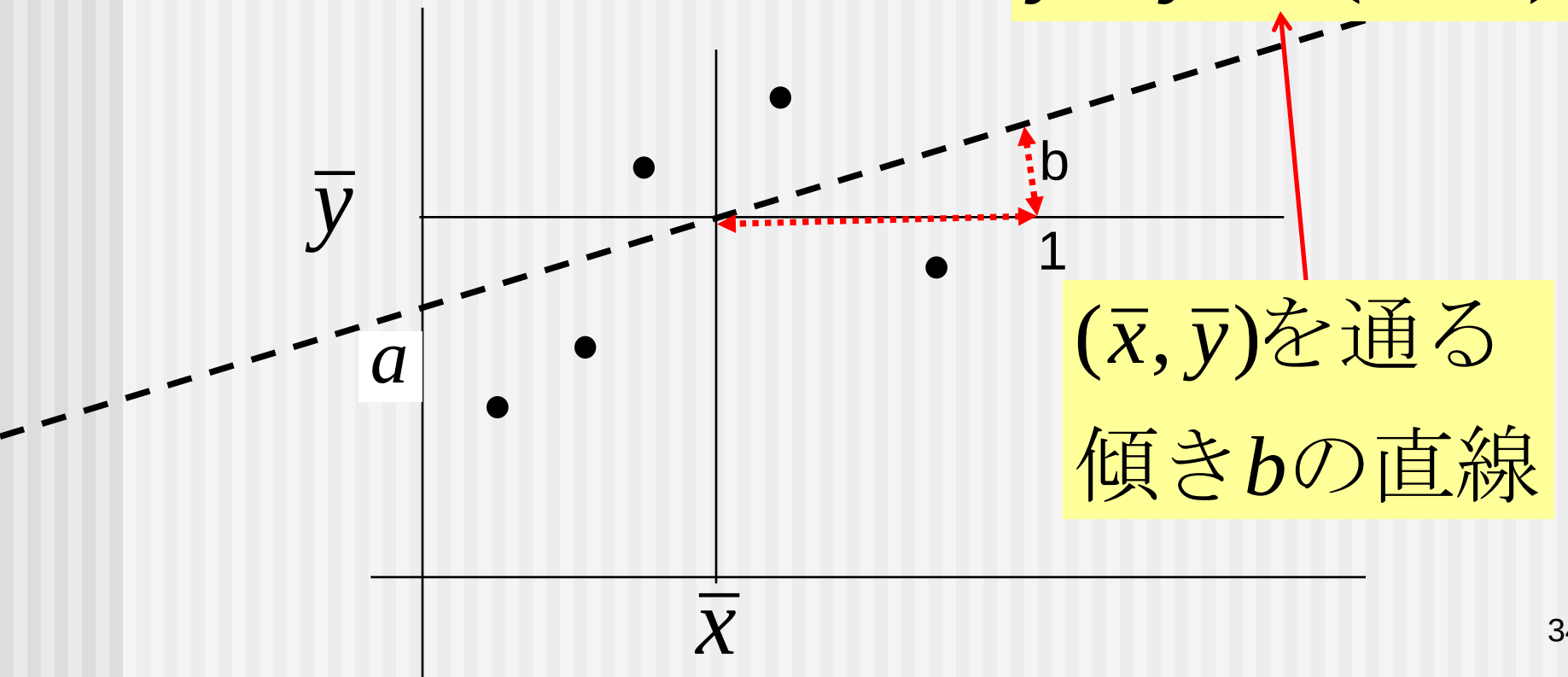
回帰直線とは

$$b = \frac{\sum (y_i - \bar{y})(x_i - \bar{x})}{\sum (x_i - \bar{x})^2} = \frac{s_{xy}}{s_{xx}}$$

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

$y = a + bx$

$y = \bar{y} + b(x - \bar{x})$



$(\bar{x}, \bar{y})$を通る
傾きbの直線

最小2乗推定値の求め方（1）難

$$\sum_{i=1}^{n}(y_i - a - b\,x_i)^2 = \sum(y_i - \bar{y} + \bar{y} - a - b\,x_i + b\bar{x} - b\bar{x})^2$$

$$= \sum\{(y_i - \bar{y}) - b(x_i - \bar{x}) + (\bar{y} - a - b\,\bar{x})\}^2$$

$$= \sum\{\,(y_i - \bar{y})^2 + b^2(x_i - \bar{x})^2 + (\bar{y} - a - b\,\bar{x})^2$$

$$- 2b(x_i - \bar{x})(y_i - \bar{y}) + 2(\bar{y} - a - b\,\bar{x})(y_i - \bar{y})$$

$$- 2b(\bar{y} - a - b\,\bar{x})(x_i - \bar{x})\,\}$$

最小2乗推定値の求め方(2)難

$$\sum\{(y_i-\bar{y})^2+b^2(x_i-\bar{x})^2+(\bar{y}-a-b\bar{x})^2$$
$$-2b(x_i-\bar{x})(y_i-\bar{y})+2b(\bar{y}-a-b\bar{x})(y_i-\bar{y})$$
$$-2(\bar{y}-a-b\bar{x})(x_i-\bar{x})\}$$

$$=\sum(y_i-\bar{y})^2+b^2\sum(x_i-\bar{x})^2+n(\bar{y}-a-b\bar{x})^2$$
$$-2b\sum(y_i-\bar{y})(x_i-\bar{x})-2b(\bar{y}-a-b\bar{x})\sum(x_i-\bar{x})$$
$$+2(\bar{y}-a-b\bar{x})\sum(y_i-\bar{y})$$

0

1．この部分を最小にするようにｂを決める

最小2乗推定値の求め方(3)難

$$\sum_{i=1}^{n}(y_i - a - b\,x_i)^2 = \sum(y_i - \bar{y})^2 + b^2\sum(x_i - \bar{x})^2$$

$$-2b\sum(y_i - \bar{y})(x_i - \bar{x}) + n(\bar{y} - a - b\,\bar{x})^2$$

$$= \sum(x_i - \bar{x})^2\left\{b^2 - 2\frac{\sum(y_i - \bar{y})(x_i - \bar{x})}{\sum(x_i - \bar{x})^2}b\right\}$$

$$+ n(\bar{y} - a - b\,\bar{x})^2 + \sum(y_i - \bar{y})^2$$

$a = \bar{y} - b\,\bar{x}$のとき0になる．

最小2乗推定値の求め方(4)難

$$b^2 - 2\frac{\sum(y_i-\bar{y})(x_i-\bar{x})}{\sum(x_i-\bar{x})^2}b$$

$$=\left(b-\frac{\sum(y_i-\bar{y})(x_i-\bar{x})}{\sum(x_i-\bar{x})^2}\right)^2-\left(\frac{\sum(y_i-\bar{y})(x_i-\bar{x})}{\sum(x_i-\bar{x})^2}\right)^2$$

$$b=\frac{\sum(y_i-\bar{y})(x_i-\bar{x})}{\sum(x_i-\bar{x})^2}$$

のとき最小となる。

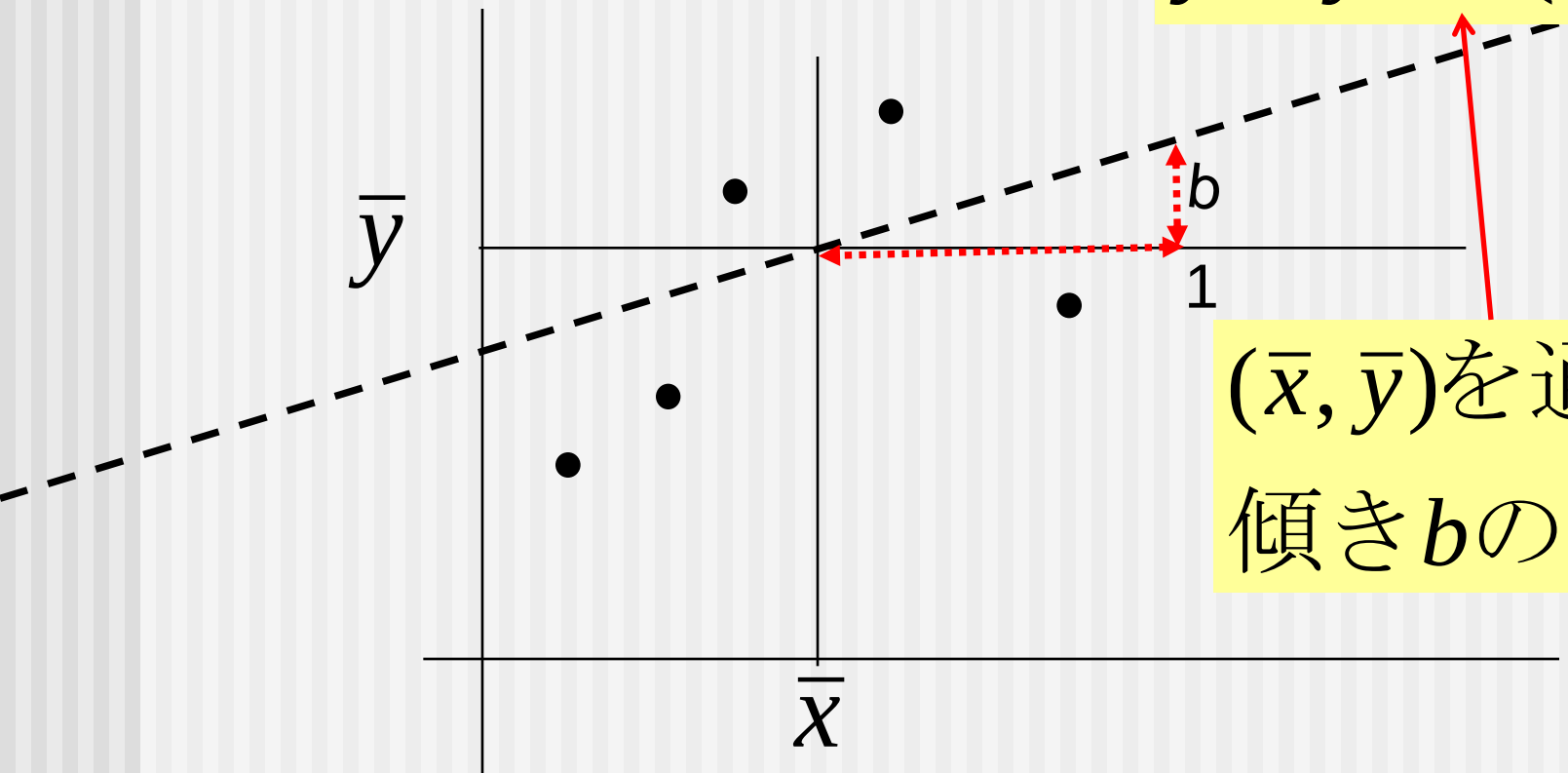
回帰直線とは

$$b = \frac{\sum (y_i - \bar{y})(x_i - \bar{x})}{\sum (x_i - \bar{x})^2} = \frac{s_{xy}}{s_{xx}}$$

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

$y = a + bx$

$y = \bar{y} + b(x - \bar{x})$

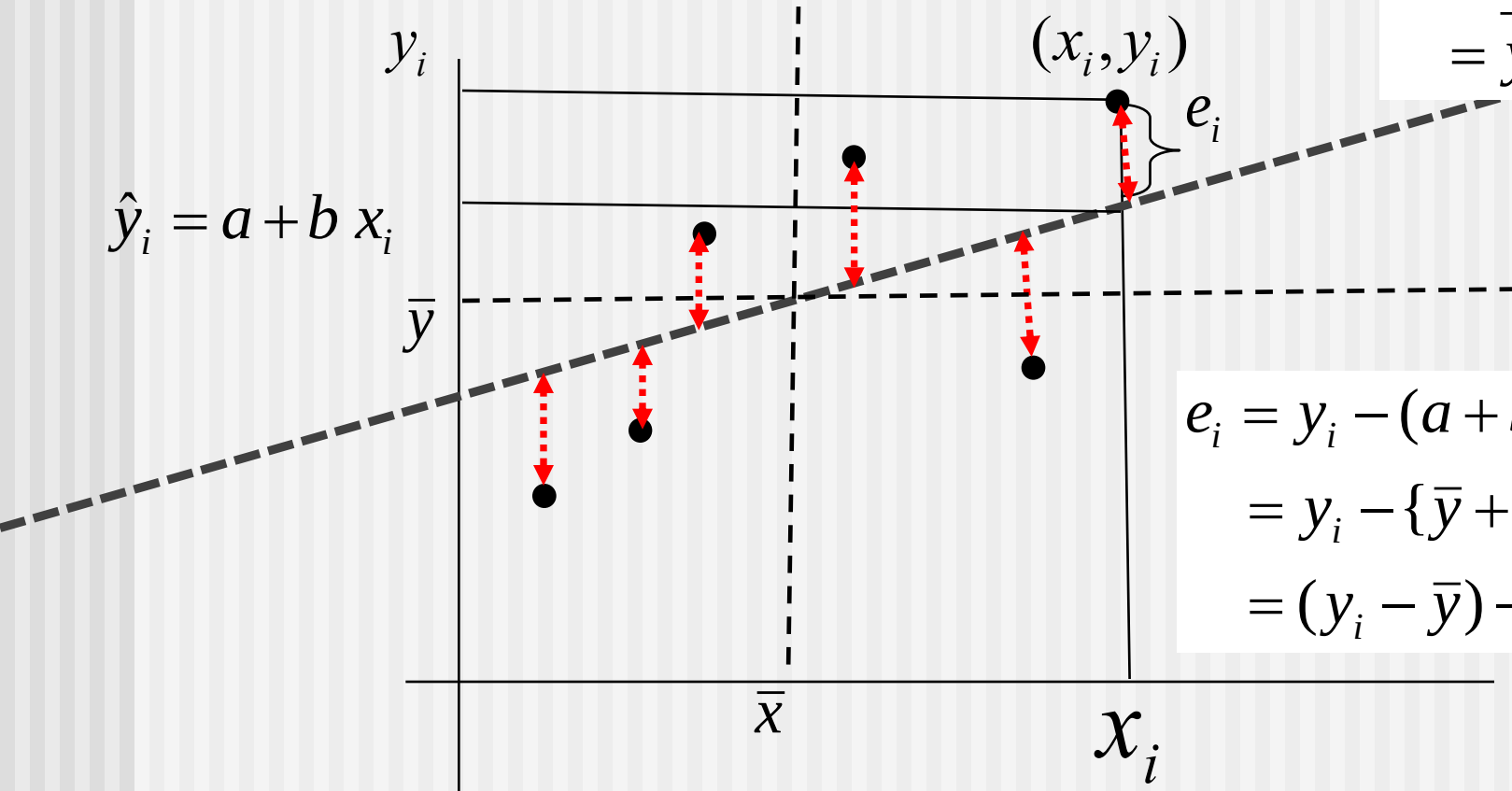


$(\bar{x}, \bar{y})$を通る
傾きbの直線

残差とは？

- 実際の観測値と推定値との差を残差と呼ぶ.

$$e_i = y_i - \hat{y}_i = y_i - (a + b\,x_i)$$

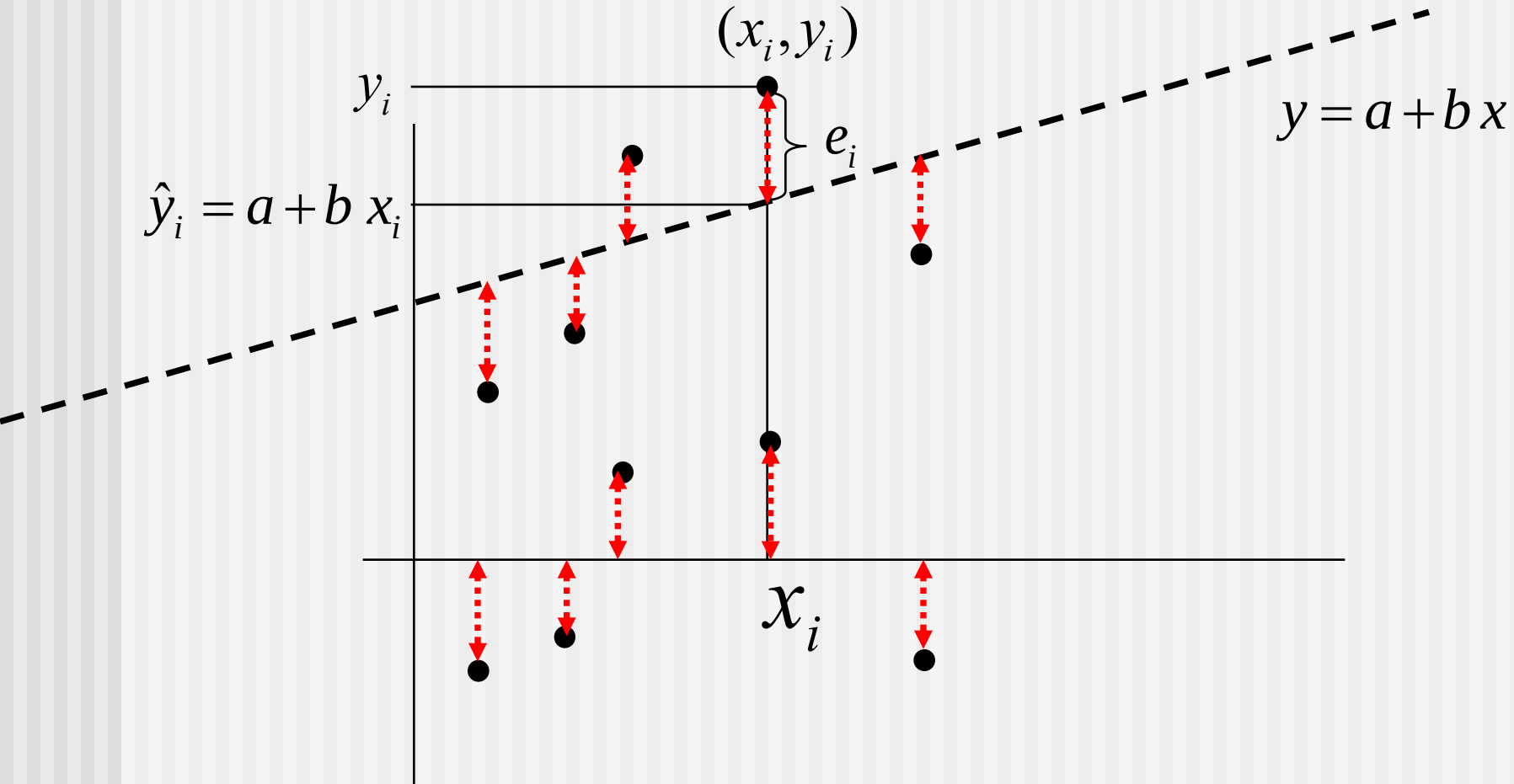


$$y = a + b\,x$$
$$= \bar{y} + b(x - \bar{x})$$

$$e_i = y_i - (a + b\,x_i)$$
$$= y_i - \{\bar{y} + b(x_i - \bar{x})\}$$
$$= (y_i - \bar{y}) - b(x_i - \bar{x})$$

残差プロット

- y 軸に残差をとったものを残差プロットと呼ぶ



残差の和と残差の平方和

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

- 残差の総和は0である.

$$\sum_{i=1}^{n} e_i = \sum_{i=1}^{n} (y_i - a - b\,x_i) = \sum_{i=1}^{n} (y_i - \bar{y}) - b\sum_{i=1}^{n} (x_i - \bar{x}) = 0$$

- 当然のことだが、残差の平均も0である.

$$\bar{e} = 0$$

- 残差の分散は，下のように表される.

$$s_{ee} = \frac{1}{n}\sum_{i=1}^{n} (e_i - \bar{e})^2 = \sum e_i^2 = \frac{1}{n}\sum_{i=1}^{n} (y_i - a - b\,x_i)^2$$

残差分散

$$\hat{y}_i = a + b\,y_i = \bar{y} + b(x_i - \bar{x})$$

$$b = \frac{s_{xy}}{s_{xx}}$$

$$s_{ee} = \frac{1}{n}\sum_{i=1}^{n}(y_i - \hat{y}_i)^2 = \frac{1}{n}\sum_{i=1}^{n}\{y_i - \bar{y} - b(x_i - \bar{x})\}^2$$

$$= \frac{1}{n}\sum(y_i - \bar{y})^2 - 2b\frac{1}{n}\sum(y_i - \bar{y})(x_i - \bar{x}) + b^2\frac{1}{n}\sum(x_i - \bar{x})^2$$

$$= s_{yy} - 2\frac{s_{xy}}{s_{xx}}s_{xy} + \left(\frac{s_{xy}}{s_{xx}}\right)^2 s_{xx} = s_{yy}\left(1 - \frac{s_{xy}^2}{s_{xx}s_{yy}}\right)$$

残差平方和と相関係数の関係

r_{xy}：相関係数

$$s_{ee} = s_{yy}(1 - r_{xy}^2)$$

- 相関係数が1に近いほど，残差平方和は小さくなる．つまり，推定精度が高い．

$R^2 = r_{xy}^2$ を決定係数と呼ぶ．

例題13.2　勤続年数と給与額

	勤続年数（年） x_i	所定内給与額（千円） y_i	$x_i-\bar{x}$	$(x_i-\bar{x})^2$	$y_i-\bar{y}$	$(y_i-\bar{y})^2$	$(x_i-\bar{x})(y_i-\bar{y})$
0年	0.0	217.4	−13.56	183.75	−101.80	10,363.24	1379.96
1 ～2年	1.5	230.3	−12.06	145.34	−88.90	7,903.21	1071.74
3～4年	3.5	246.0	−10.06	101.11	−73.20	5,358.24	736.07
5～9年	7.0	264.1	−6.56	42.98	−55.10	3,036.01	361.21
10～14年	12.0	300.6	−1.56	2.42	−18.60	345.96	28.93
15～19年	17.0	348.3	3.44	11.86	29.10	846.81	100.23
20～24年	22.0	395.4	8.44	71.31	76.20	5,806.44	643.47
25～29年	27.0	426.7	13.44	180.75	107.50	11,556.25	1445.28
30年以上	32.0	444.0	18.44	340.20	124.80	15,575.04	2301.87
合計	122.0	2,872.8	0.00	1,079.72	0.00	60,791.20	8,068.75
平均 $\bar{y}$	13.56	319.20	0.00	119.97	0.00	6,754.58	896.53

$$y = a + b\,x$$
$$= \bar{y} + b(x - \bar{x})$$

$$b = \frac{s_{xy}}{s_{xx}} = \frac{896.53}{119.97} = 7.473$$

$$a = \bar{y} - b\bar{x} = 319.20 - 7.473 \times 13.56$$
$$= 217.9$$

$$y = 217.9 + 7.473\,x$$
$$= 319.20 + 7.473(x - 13.56)$$

例題13.2（続き）

相関係数

$$r_{xy} = \frac{s_{xy}}{\sqrt{s_{xx} s_{yy}}}$$

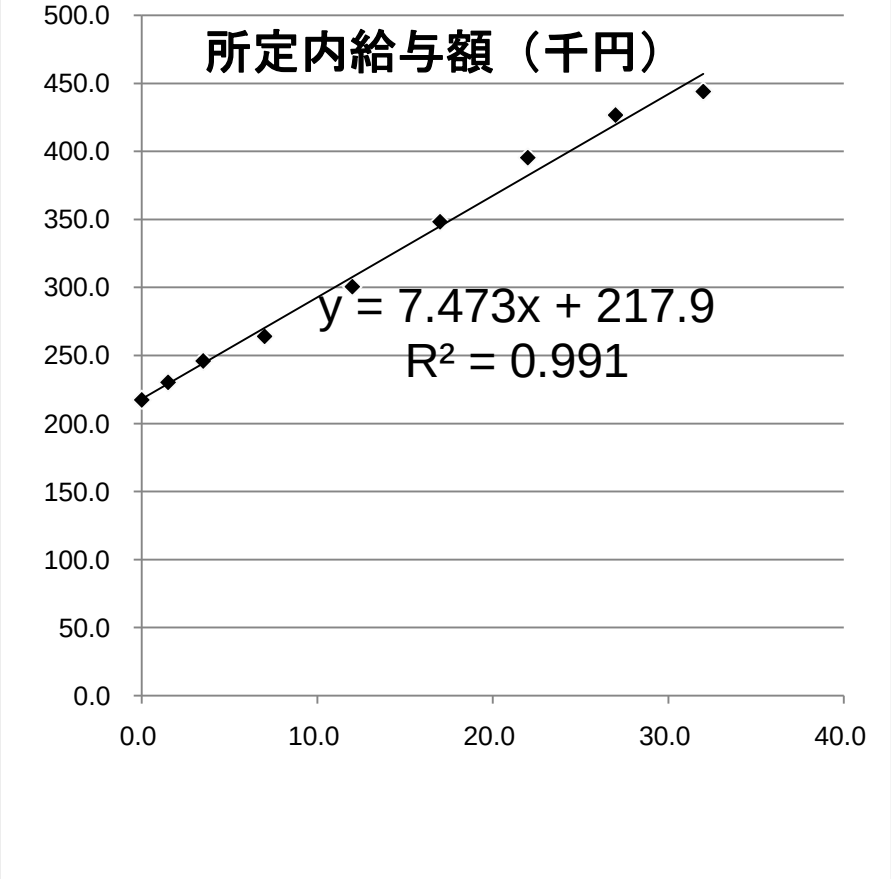
$$= \frac{896.53}{\sqrt{119.97 \times 6754.58}}$$

$$= 0.9959$$

回帰の決定係数

$$R^2 = r_{xy}^2 = (0.9959)^2$$

$$= 0.991$$



例題13.2（続き）

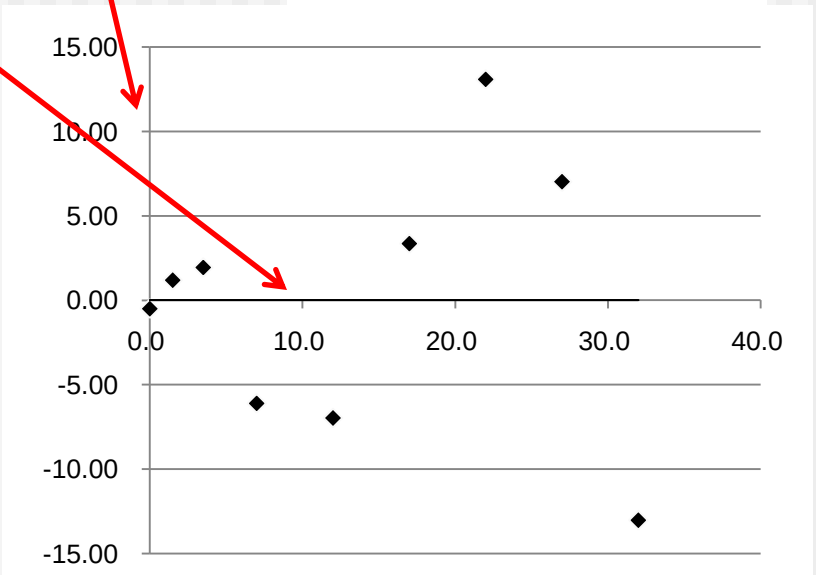
	A	B	C	I	J	K
3		勤続年数（年） x_i	所定内給与額（千円） y_i	推定値 $\hat{y}_i$	残差 $e_i = y_i - \hat{y}_i$	e_i^2
4	0年	0.0	217.4	217.90	−0.50	0.25
5	1 ～2年	1.5	230.3	229.11	1.19	1.42
6	3～4年	3.5	246.0	244.05	1.95	3.78
7	5～9年	7.0	264.1	270.21	−6.11	37.34
8	10～14年	12.0	300.6	307.58	−6.98	48.66
9	15～19年	17.0	348.3	344.94	3.36	11.29
10	20～24年	22.0	395.4	382.31	13.09	171.47
11	25～29年	27.0	426.7	419.67	7.03	49.42
12	30年以上	32.0	444.0	457.04	−13.04	169.91
13	合計	122.0	2,872.8	2,872.80	0.00	493.54
14	平均 $\bar{y}$	13.56	319.20	319.20	0.00	54.84
15	表13-4より作成					

$$s_{ee} = s_{yy}(1 - R^2)$$

$$= 6754.58 \times (1 - 0.991)$$

$$= 54.84$$

残差プロット



3　決定係数

r_{xy}：相関係数

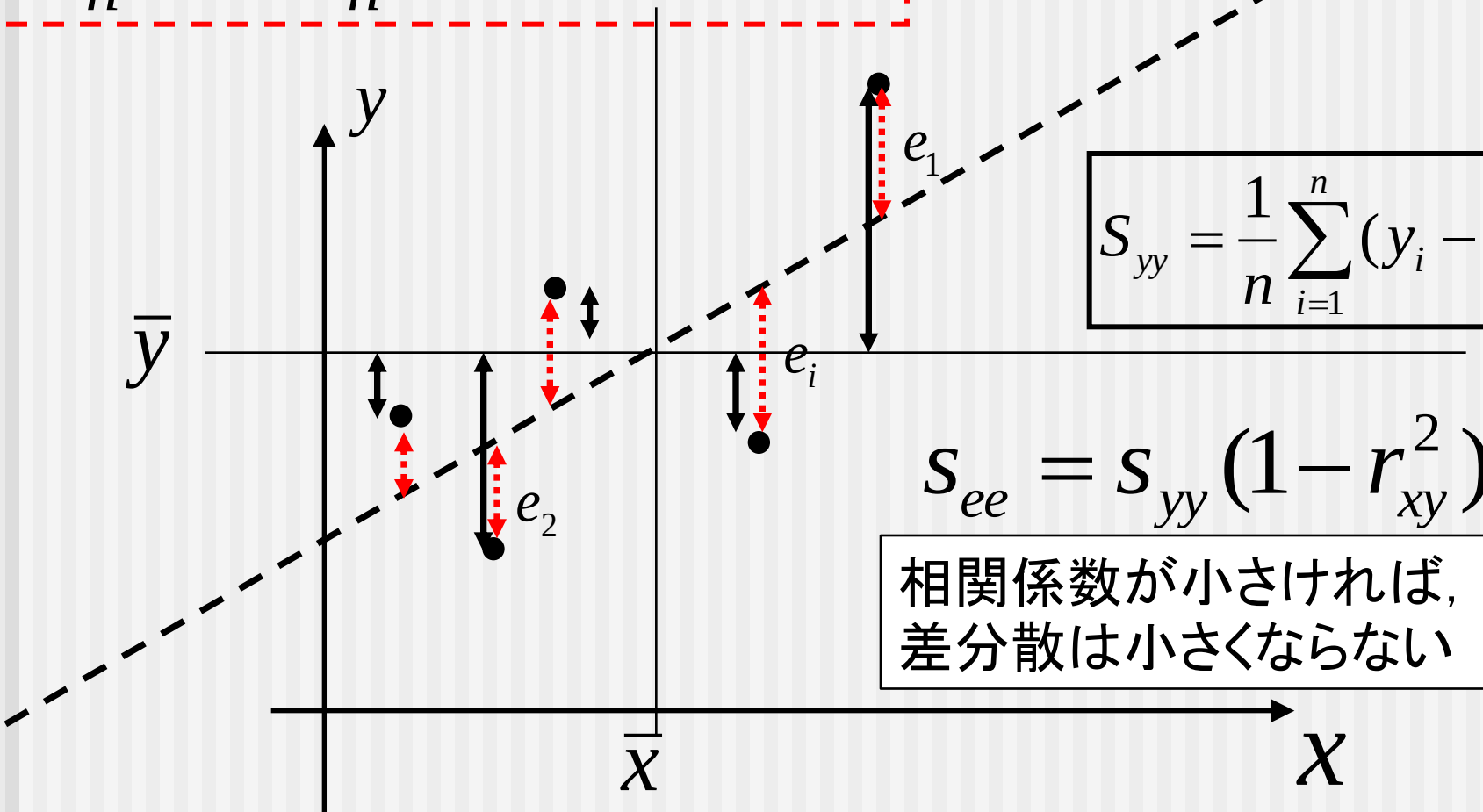
$R^2(=r_{xy}^2)$：回帰の決定係数という

- 決定係数は相関係数を2乗したものであるが，その他にもさまざまな解釈ができる．

決定係数の意味（小さな相関）

図を書いてみる

$$s_{ee} = \frac{1}{n}\sum e_i^2 = \frac{1}{n}\sum \{y_i - (a + b\,x_i)\}^2$$



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

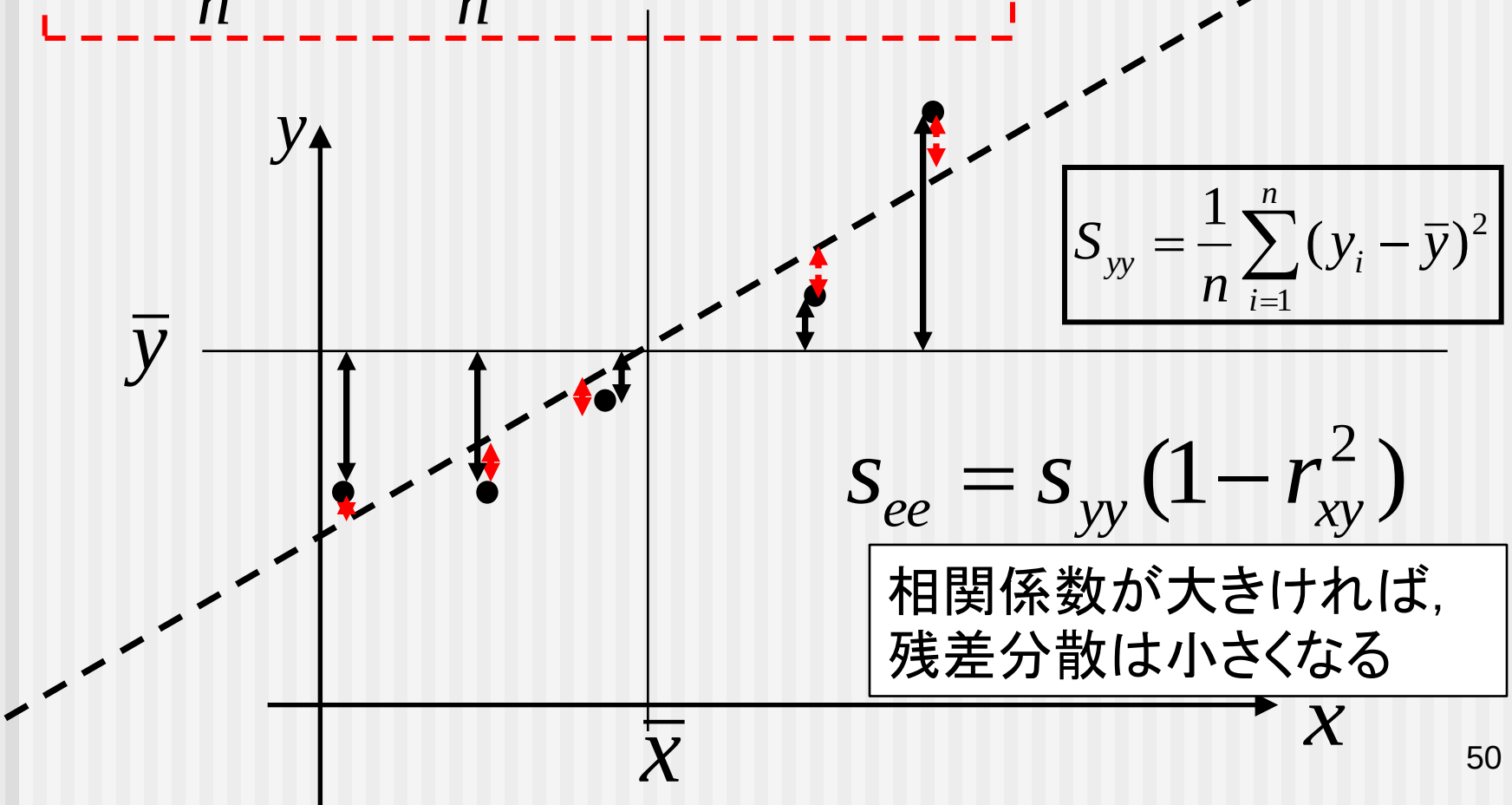
$$s_{ee} = s_{yy}(1 - r_{xy}^2)$$

相関係数が小さければ，残差分散は小さくならない

決定係数の意味（大きな相関）

図を書いてみる

$$s_{ee} = \frac{1}{n}\sum e_i^2 = \frac{1}{n}\sum \{y_i - (a + b\,x_i)\}^2$$



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

$$s_{ee} = s_{yy}(1 - r_{xy}^2)$$

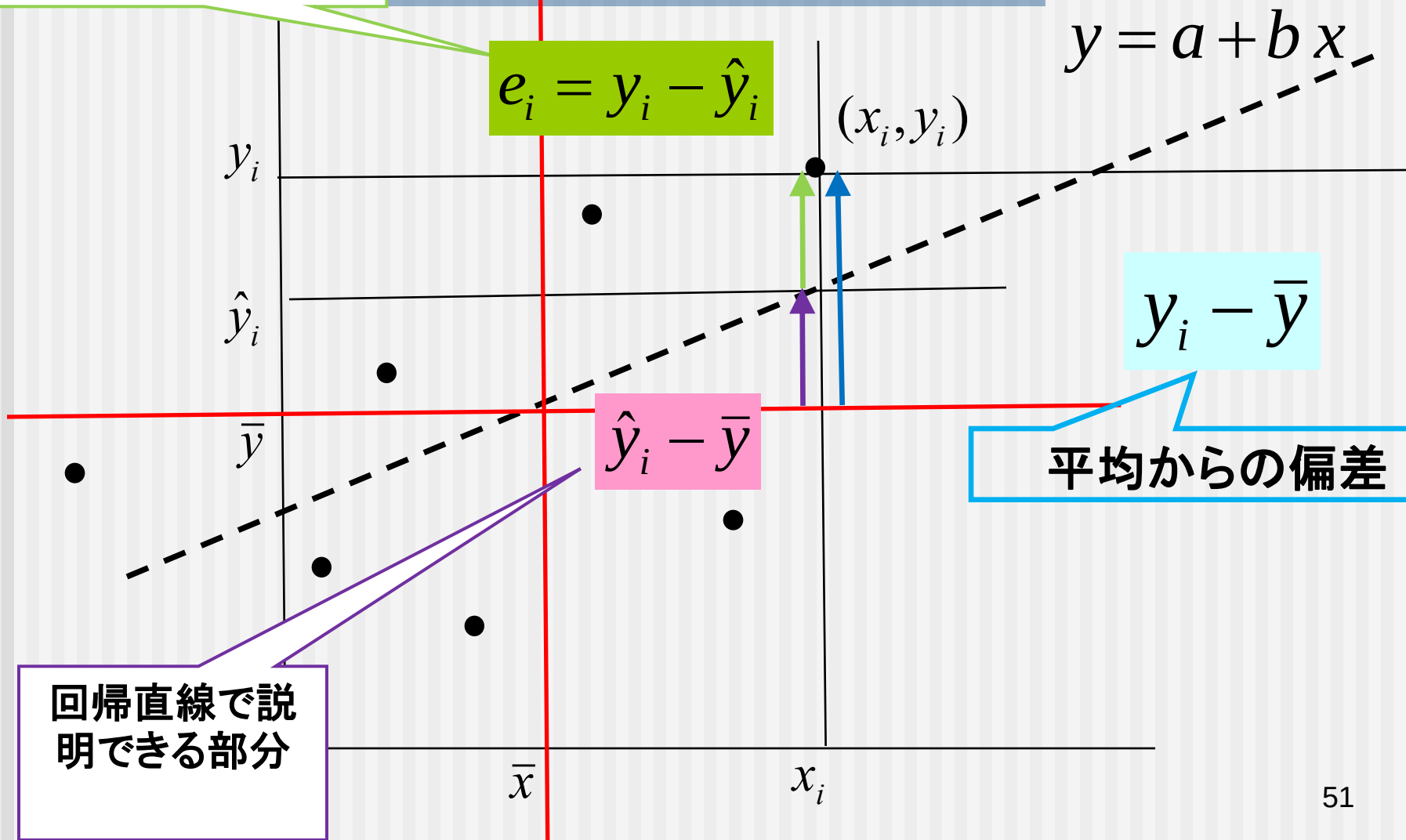
相関係数が大きければ，
残差分散は小さくなる

y の変動の分解と決定係数

$$y_i - \bar{y} = (y_i - \hat{y}_i) + (\hat{y}_i - \bar{y})$$

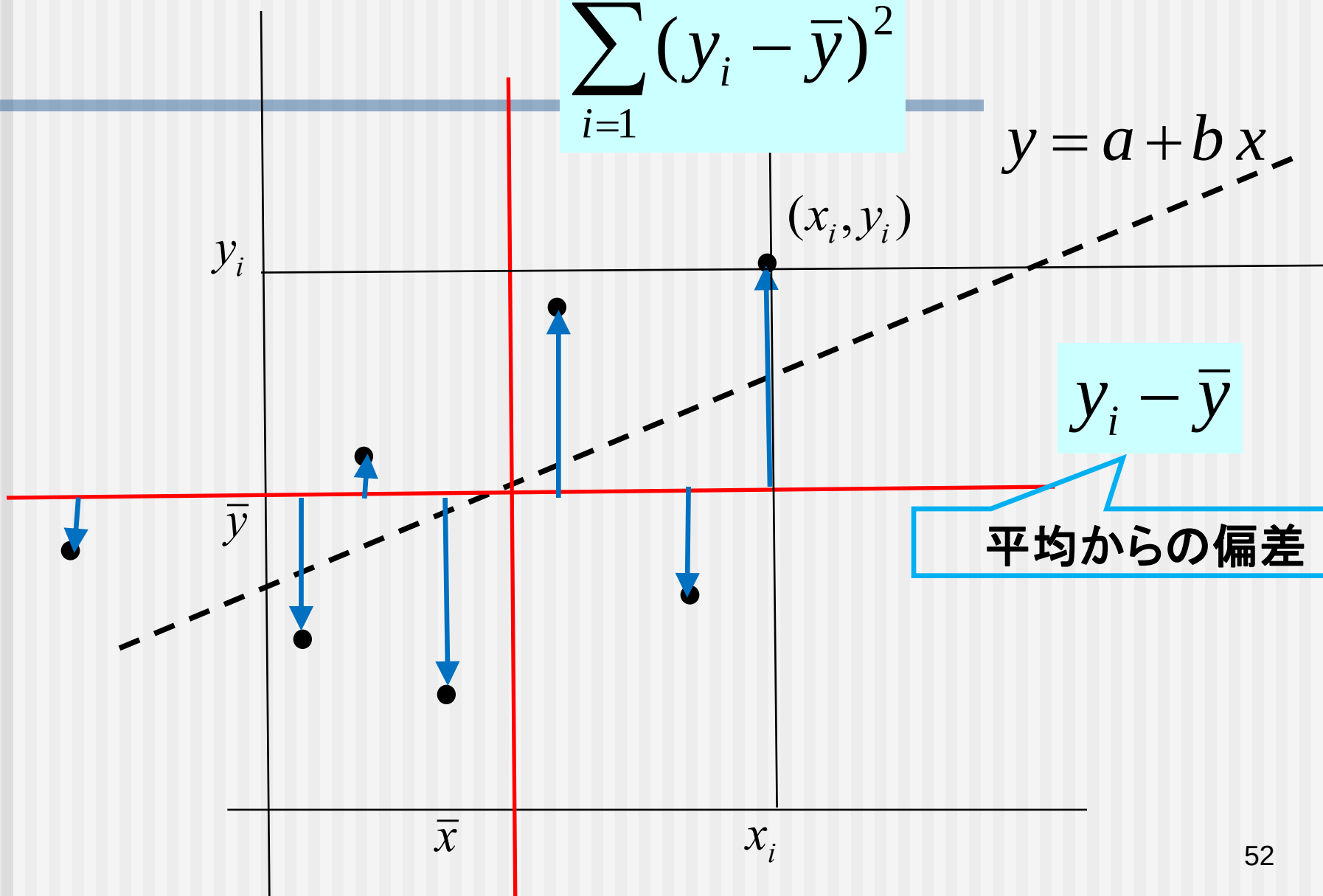
残差： 回帰直線では説明しきれない部分

$$e_i = y_i - \hat{y}_i$$



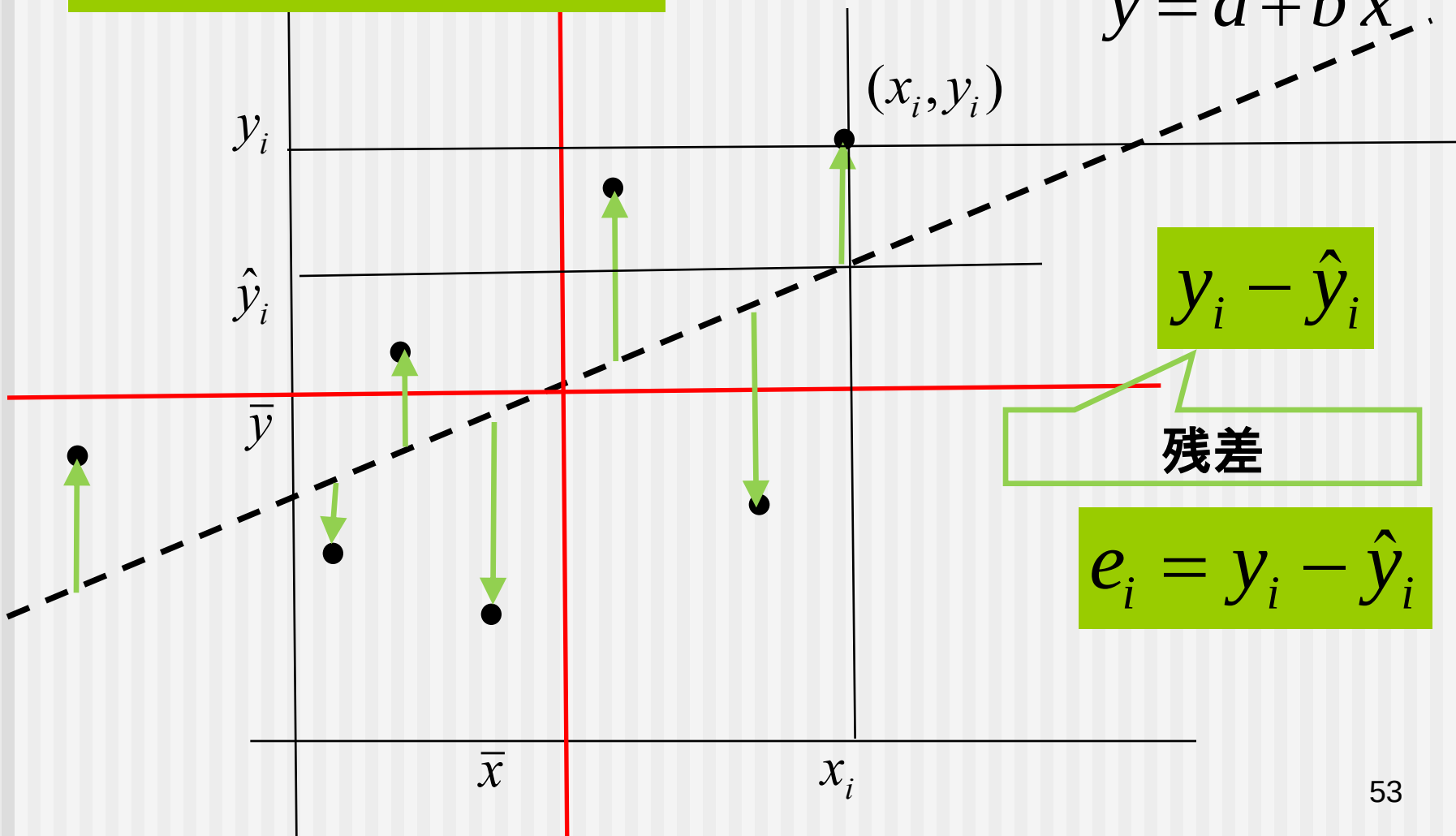
y の全変動(平均からの変動)

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



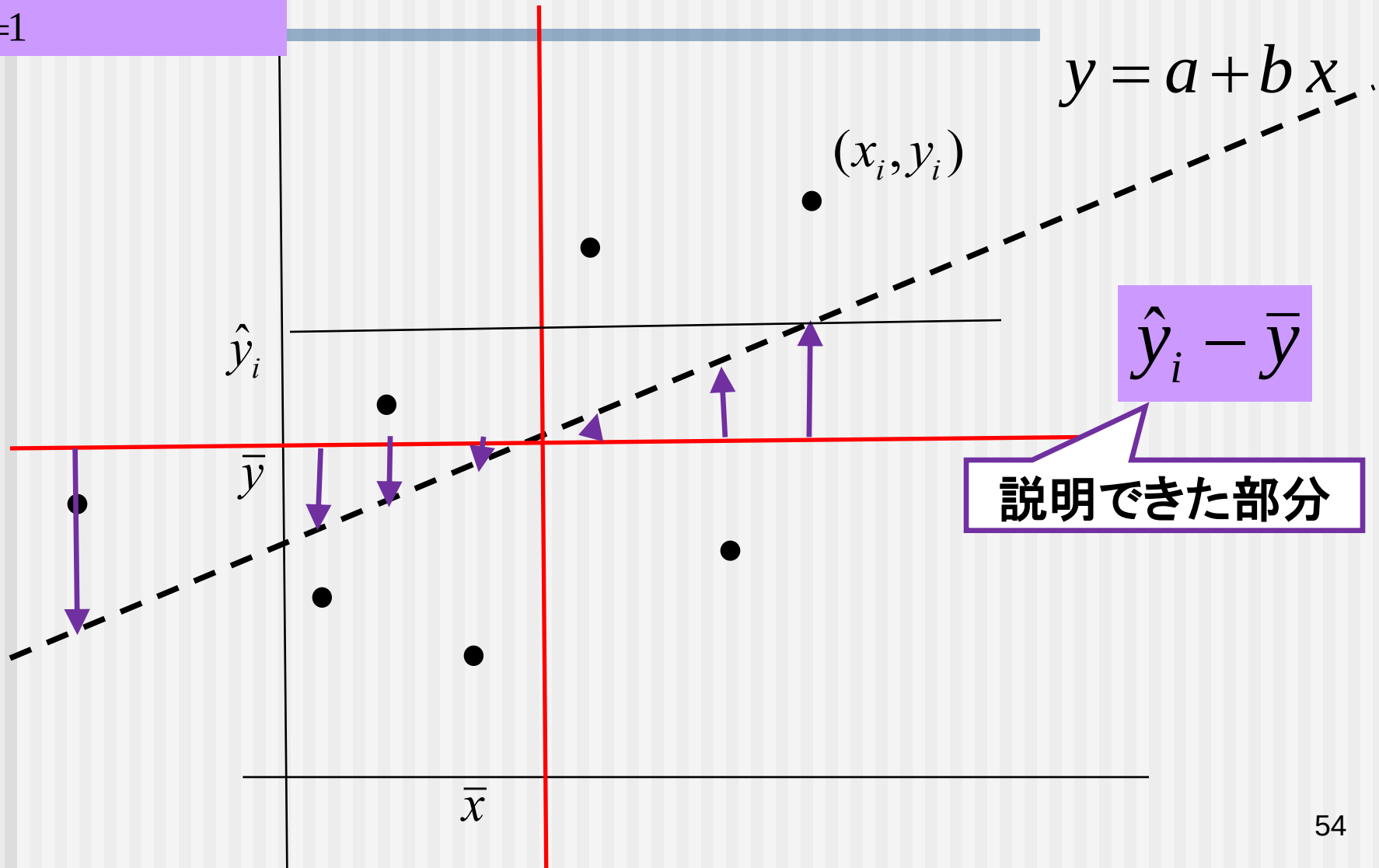
回帰で説明されない変動（残差）

$$\sum_{i=1}^{n}(y_i-\hat{y}_i)^2=\sum_{i=1}^{n}e_i^2$$



回帰で説明される変動

$$\sum_{i=1}^{n}(\hat{y}_i-\bar{y})^2$$



決定係数とは（2）難

$$\sum_{i=1}^{n}(y_i-\bar{y})^2=\sum\{(y_i-\hat{y}_i)+(\hat{y}_i-\bar{y})\}^2$$

$$=\sum(y_i-\hat{y}_i)^2+\sum(\hat{y}_i-\bar{y})^2+2\sum(y_i-\hat{y}_i)(\hat{y}_i-\bar{y})$$

0

なぜなら、

$$\hat{y}_i=\bar{y}+b(x_i-\bar{x})$$

$$b=\frac{s_{xy}}{s_{xx}}=\frac{\sum(x_i-\bar{x})(y_i-\bar{y})}{\sum(x_i-\bar{x})^2}$$

$$\sum(y_i-\hat{y}_i)(\hat{y}_i-\bar{y})=\sum\{y_i-\bar{y}-b(x_i-\bar{x})\}\{b(x_i-\bar{x})\}$$

$$=b\sum(y_i-\bar{y})(x_i-\bar{x})-b^2\sum(x_i-\bar{x})^2=0$$

全変動

回帰からの変動

回帰による変動

決定係数とは(3)　難

$$\sum_{i=1}^{n}(y_i-\bar{y})^2=\sum(y_i-\hat{y}_i)^2+\sum(\hat{y}_i-\bar{y})^2$$

$$1=\frac{\sum(y_i-\hat{y}_i)^2}{\sum_{i=1}^{n}(y_i-\bar{y})^2}+\frac{\sum(\hat{y}_i-\bar{y})^2}{\sum_{i=1}^{n}(y_i-\bar{y})^2}$$

決定係数：R^2

全変動のうち、**回帰による変動**の占める割合

$$\frac{\sum(\hat{y}_i-\bar{y})^2}{\sum_{i=1}^{n}(y_i-\bar{y})^2}=\frac{\sum\{b(x_i-\bar{x})\}^2}{\sum_{i=1}^{n}(y_i-\bar{y})^2}=\frac{b^2\frac{1}{n}\sum(x_i-\bar{x})^2}{\frac{1}{n}\sum_{i=1}^{n}(y_i-\bar{y})^2}=\frac{s_{xy}^2}{s_{xx}\,s_{yy}}$$

決定係数とは（4）難

この表現が後に重要になる

$$R^2 = \frac{\sum_{i=1}^{n}(\hat{y}_i - \bar{y})^2}{\sum_{i=1}^{n}(y_i - \bar{y})^2} = 1 - \frac{\sum_{i=1}^{n}(y_i - \hat{y}_i)^2}{\sum_{i=1}^{n}(y_i - \bar{y})^2} = 1 - \frac{\sum_{i=1}^{n} e_i^2}{\sum_{i=1}^{n}(y_i - \bar{y})^2}$$

$$R^2 = \frac{\sum_{i=1}^{n}(\hat{y}_i - \bar{y})^2}{\sum_{i=1}^{n}(y_i - \bar{y})^2} = \frac{s_{xy}^2}{s_{xx}\, s_{yy}}$$

相関係数の2乗

決定係数のもう一つの意味

回帰モデルの説明力を示すもの.

計算すれば，この等式が成り立つことが分かる.

$$\sum (y_i - \bar{y})^2 = \sum (y_i - \hat{y}_i)^2 + \sum (\hat{y}_i - \bar{y})^2$$

両辺を $\sum (y_i - \bar{y})^2$ で割ると，

$$\frac{\sum (y_i - \bar{y})^2}{\sum (y_i - \bar{y})^2} = \frac{\sum (y_i - \hat{y}_i)^2}{\sum (y_i - \bar{y})^2} + \frac{\sum (\hat{y}_i - \bar{y})^2}{\sum (y_i - \bar{y})^2}$$

$$= 1$$

この部分を R^2 と呼ぶ.

決定係数の意味（さらに）

$$\frac{\sum (y_i - \bar{y})^2}{\sum (y_i - \bar{y})^2} = \frac{\sum (y_i - \hat{y}_i)^2}{\sum (y_i - \bar{y})^2} + \frac{\sum (\hat{y}_i - \bar{y})^2}{\sum (y_i - \bar{y})^2} \quad R^2$$

$$=1$$

$$R^2 = 1 - \frac{\sum (y_i - \hat{y}_i)^2}{\sum (y_i - \bar{y})^2} = 1 - \frac{\sum e_i^{\,2}}{\sum (y_i - \bar{y})^2}$$

決定係数のまとめ

決定係数は，全変動のうち回帰で説明できる割合である．

$$R^2 = r_{xy}^2$$

$$s_{ee} = s_{yy}(1 - r_{xy}^2)$$

$$R^2 = \frac{\sum(\hat{y}_i - \bar{y})^2}{\sum(y_i - \bar{y})^2}$$

決定係数は，相関係数の２乗である．

$$R^2 = 1 - \frac{\sum e_i^2}{\sum(y_i - \bar{y})^2}$$

決定係数は，全変動のから回帰で説明できなかった部分を除いた割合である．

4　重回帰モデル

- 家計の消費水準を，可処分所得と消費者物価により説明する.
- 一人当り賃金上昇率を，消費者物価上昇率と失業率により説明する.
- 株価水準の変動を内外金利水準や鉱工業生産指数，為替レート等の変動や，金融的変数の変動で説明したりする.
- 説明変数が複数あるということは、思わぬ問題を引き起こす。詳しくは、計量経済学で。

重回帰モデル（数式，推定法）

- データが得られるメカニズムは以下の式で表される.

$$(3.1)\ y_i = \alpha + \beta_1 x_{1i} + \beta_2 x_{2i} + \square + \beta_K x_{Ki} + \varepsilon_i$$

- 推定値は最小2乗法，つまり以下の式を最小にするものとして得られる.

$$(3.2)\sum_{i=1}^{n}(y_i - \alpha - \beta_1 x_{1i} - \beta_2 x_{2i} - \cdots - \beta_K x_{Ki})^2$$

推定値・残差

パラメータの推定値を次のように表そう.

$$a, b_1, b_2, b_3, \square, b_K$$

すると各観測の推定値は,

$$(3.4)\ \hat{y}_i = a + b_1 x_{1i} + b_2 x_{2i} + \square + b_K x_{Ki}$$

残差は,

$$(3.5)\ e_i = y_i - \hat{y}_i, \quad \sum_{i=1}^{n} e_i = 0$$

回帰モデルの選択

- ある会社の株価を予想したい.
  - 株価は会社の成長性，安定性，収益性などの要素で決まると言われる.
  - 成長性の指標として，昨年の売上高成長率を採用するのか，それとも，５年間の平均を採用するのか.
  - あるいは，経常利益を考えるのか？
  - それとも，両方を採用するか？

回帰モデルの候補は数えきれない

- 説明変数として何を採用するのか？
- 説明変数をいくつ採用するのか？
- 競合する回帰モデルの優劣を示す数値が必要になる.
- その一つが，修正決定係数である.

修正決定係数とは？

修正前の決定係数

$$R^2 = 1 - \frac{\frac{1}{n}\sum e_i^2}{\frac{1}{n}\sum (y_i - \bar{y})^2}$$

- 決定係数はモデルの選択に使えない。
- 説明変数の数を増やせば，決定係数は必ず増加ずる.
- 説明変数が多いと有利。
- その理由は、下のような2つのモデルを考えてみる。

$$\begin{cases} y_i = \alpha + \beta_1 x_{1i} \cdots\cdots\cdots\cdots\cdots\cdots (1) \\ y_i = \alpha + \beta_1 x_{1i} + \beta_2 x_{2i} \cdots\cdots (2) \end{cases}$$

説明変数が多いと・・・・

$$\sum e_i^2 = \min_{a,b_1} \sum (y_i - a - b_1 x_{1i})^2 \cdots\cdots (1')$$

$$\sum e_i^2 = \min_{a,b_1,b_2} \sum (y_i - a - b_1 x_{1i} - b_2 x_{2i})^2 \cdots (2')$$

■このとき、(1’)と(2’)とではどちらが小さいか、考えてみよう。

■(2’) が小さくなる。

■なぜなら、(2’) では　b_2=0 と固定したとき、(1’) と同じになるので、この制約を外せば(1’)よりも小さくなることが期待できる。

修正決定係数の定義:単回帰のとき

$$y_i = \alpha + \beta_1 x_{1i} \cdots\cdots\cdots\cdots\cdots\cdots (1)$$

$$R^2 = 1 - \frac{\frac{1}{n}\sum e_i^2}{\frac{1}{n}\sum (y_i - \bar{y})^2}$$

$$\sum_{i=1}^{n} (y_i - \bar{y})^2 = \min_a \sum (y_i - a)^2$$

$$\sum e_i^2 = \min_{a,b_1} \sum (y_i - a - b_1 x_{1i})^2$$

$$\bar{R}^2 = 1 - \frac{\sum (y_i - \hat{y}_i)^2 / (n-2)}{\sum_{i=1}^{n} (y_i - \bar{y})^2 / (n-1)} = 1 - \frac{\sum e_i^2 / (n-2)}{\sum_{i=1}^{n} (y_i - \bar{y})^2 / (n-1)}$$

説明変数が一つ増えると,

$$y_i = \alpha + \beta_1 x_{1i} + \beta_2 x_{2i} \cdots\cdots (2)$$

$$R^2 = 1 - \frac{\frac{1}{n}\sum e_i^2}{\frac{1}{n}\sum (y_i - \bar{y})^2}$$

$$\sum_{i=1}^{n} (y_i - \bar{y})^2 = \min_a \sum (y_i - a)^2$$

$$\sum e_i^2 = \min_{a,b_1,b_2} \sum (y_i - a - b_1 x_{1i} - b_2 x_{2i})^2$$

$$\bar{R}^2 = 1 - \frac{\sum (y_i - \hat{y}_i)^2 / (n-3)}{\sum_{i=1}^{n} (y_i - \bar{y})^2 / (n-1)} = 1 - \frac{\sum e_i^2 / (n-3)}{\sum_{i=1}^{n} (y_i - \bar{y})^2 / (n-1)}$$

一般の場合

説明変数がK個のときには，下の修正決定係数を用いる.

$$\sum_{i=1}^{n}(y_i-\bar{y})^2=\min_a\sum(y_i-a)^2$$

$$\sum e_i^2=\min_{a,b_1,b_2,\cdots,b_K}\sum(y_i-a-b_1x_{1i}-b_2x_{2i}-\cdots-b_Kx_{Ki})^2$$

$$\bar{R}^2=1-\frac{\sum(y_i-\hat{y}_i)^2/(n-K-1)}{\sum_{i=1}^{n}(y_i-\bar{y})^2/(n-1)}=1-\frac{\sum e_i^2/(n-K-1)}{\sum_{i=1}^{n}(y_i-\bar{y})^2/(n-1)}$$

決定係数：回帰による変動の割合

決定係数は，説明変数が1つの場合と同様に，以下のように表される．

$$R^2 = \frac{\sum_{i=1}^{n}(\hat{y}_i - \bar{y})^2}{\sum_{i=1}^{n}(y_i - \bar{y})^2} = 1 - \frac{\sum_{i=1}^{n}(y_i - \hat{y}_i)^2}{\sum_{i=1}^{n}(y_i - \bar{y})^2} = 1 - \frac{\sum e^2}{\sum_{i=1}^{n}(y_i - \bar{y})^2}$$

$$= 1 - \frac{\sum e^2 / n}{\sum_{i=1}^{n}(y_i - \bar{y})^2 / n} = 1 - \frac{s_{ee}}{s_{yy}}$$