4 相関・回帰

4.3 回帰モデルとモデルの推定

テキスト

芳賀敏郎（2011）医薬品開発のための統計解析

第１部　基礎　改訂版、サイエンティスト社、p.275

第１部　基礎

1. 統計の基礎・・・・・1.1 宝くじの期待値と分散、1.2 サイコロの目の数の期待値と分散
1.3 分散の加法性・中心極限定理・正規分布、1.4 統計的推測、1.5 モデル

2. １組のデータの解析　2.1 データの特徴の記述、2.2 データのグラフ表示と外れ値
2.3 対数変換と対数正規分布、2.4 平均に関する推測（母標準偏差 σ 既知）
2.5 分散に関する推測、2.6 平均に関する推測（母標準偏差 σ 未知）

3. ２組のデータの解析　3.1 データのグラフ化、3.2 平均値の差の t 検定、3.3 分散の違いの検定
3.4 分散が異なる場合の平均値の差の比較
3.5 対応のある場合の平均値の差の t 検定、3.6 検出力と n の決め方
3.7 ノンパラメトリック検定

4. 相関・回帰・・・・・4.1 散布図、4.2 相関係数、**4.3 回帰モデルとモデルの推定**
4.4 誤差を考慮した推定、4.5 回帰分析適用上の諸問題

4.3 回帰モデルとモデルの推定

p.225

テキストの
該当ページ

（1）回帰直線と回帰式
（2）回帰モデル
（3）最小 2 乗法による回帰式の推定
（4）ソルバーによる解法
（5）平方和の分解とσ^2の推定
（6）a, b, S_Rなどの計算式
（7）LINEST 関数による解法

使用するファイル
　Excel ファイル「基本改4.xls」
　サイエンティスト社ホームページからダウンロード
JMP 10.0.2 を使用した結果を表示

★プレゼンテーションの
　スピーカーノートを、
　PDF の注釈に変換してあります

はじめに

p.225

●相関分析と回帰分析

前節：相関係数（相関分析）

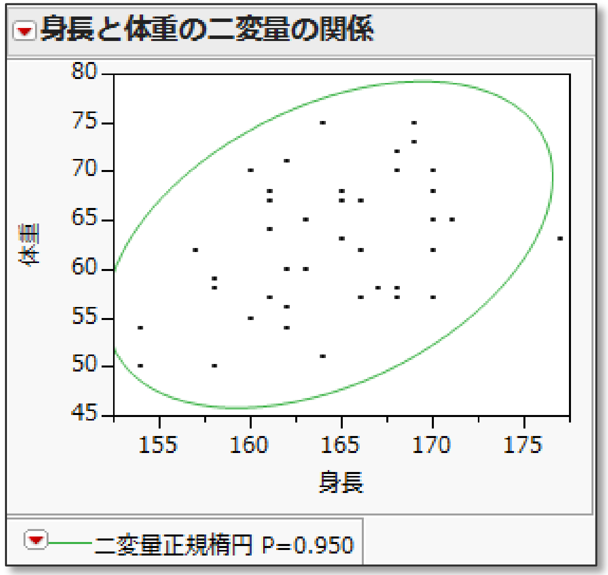
2 つの量的変数の直線的な関係の強さを表す量（直線関係の定量化）

本節：回帰直線と回帰式（単回帰分析）

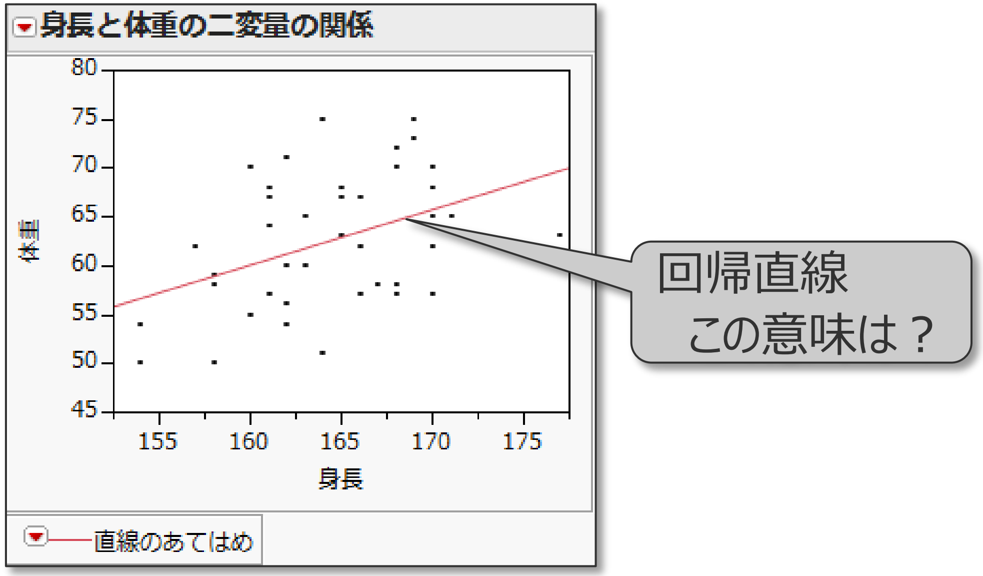
2 つの量的変数の直線関係を表す直線とその式（直線関係の定式化）

一方の変数（目的変数）を他方の変数（説明変数）で予測するためのモデル化

相関分析



単回帰分析



p.225

（1）回帰直線と回帰式

回帰直線と回帰式

p.225

●事例 1

表示4.2.13 分析ツールによる相関係数行列（p.218）

	年齢	身長	体重	血圧(上)	血圧(下)	TC	HDL	中性脂肪	TN	血糖
年齢	1									
身長	-0.246	1								
体重	-0.354	0.415	1							
血圧(上)	0.103	0.118	0.279	1						
血圧(下)	0.070	0.041	0.147	0.765	1					
TC	-0.159	-0.006	0.157	0.400	0.432	1				
HDL	0.025	0.276	-0.154	-0.058	-0.202	-0.259	1			
中性脂肪	-0.162	-0.155	0.132	0.440	0.424	0.515	-0.399	1		
TN	-0.049	0.074	0.267	0.069	-0.063	0.306	-0.150	0.062	1	
血糖	-0.135	-0.135	0.397	0.361	0.058	0.235	-0.088	0.357	0.374	1

$\alpha = 0.05$ で有意

40人の検査項目の相関係数行列

肥満度が高くなると糖尿病にかかり易くなるといわれる

血糖と身長、血糖と体重の相関は強くない

⇒　肥満度の指標（ BMI）と血糖との関係は？　（課題 4.3）

回帰直線と回帰式

p.225

●データ

Excel ファイル「基礎改4.xls」を読込
名前ボックスから「表示4.3.1」（Fig43_01）を選択

表示4.3.1 散布図と回帰式

No.	年齢	身長	体重	BMI	上-下	血糖
1	40	169	75	26.3	48.0	95
2	41	164	75	27.9	34.0	88
3	42	161	68	26.2	60.0	117
4	43	169	73	25.6	28.0	83
5	43	177	63	20.1	48.0	89
6	43	166	57	20.7	36.0	94
7	43	170	65	22.5	46.0	83
・	・	・	・	・	・	・
・	・	・	・	・	・	・
・	・	・	・	・	・	・
38	53	162	56	21.3	40.0	79
39	54	168	58	20.5	34.0	80
40	54	160	70	27.3	68.0	112

BMI （Body Mass Index）
体重(kg)÷身長(m)^2
標準:18.5－25.0、肥満:25以上
（日本肥満学会, 2025）

BMI と血糖との関係は？
↓
Excel で散布図を描画（§4.1、p.197）

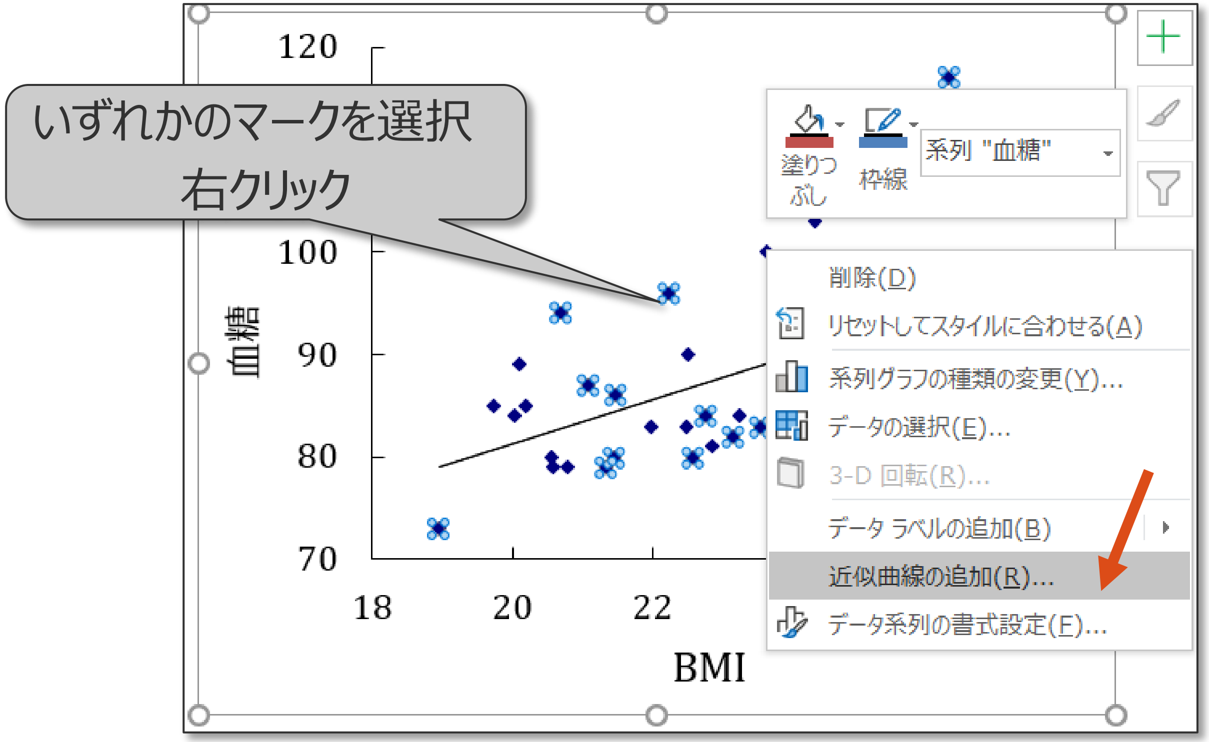
回帰直線と回帰式

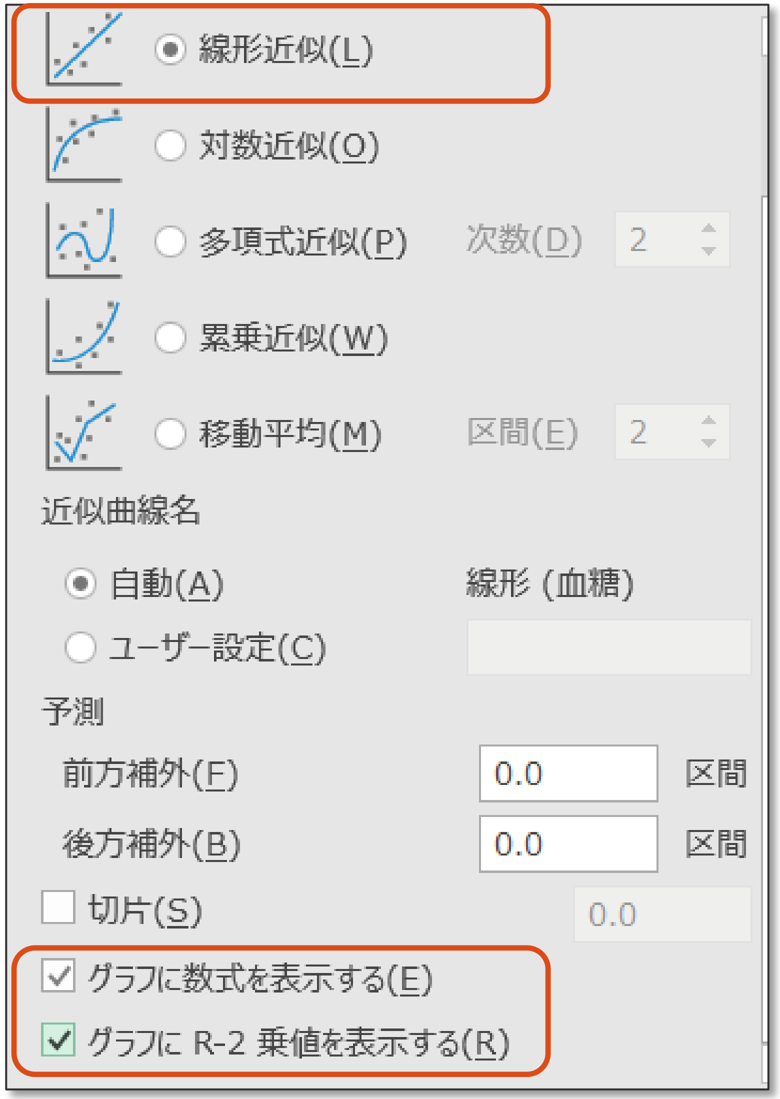
●散布図の近似曲線の追加

マークを選択して右クリック

＞［近似曲線の追加］＞［線形近似］

表示4.3.1 散布図と回帰式





回帰直線と回帰式

p.225

●相関分析→回帰分析

寄与率と相関係数（後述）

$R^2 = 0.2784$

$r = \sqrt{0.2784} = 0.5276$

血糖と身長の相関係数：－0.135

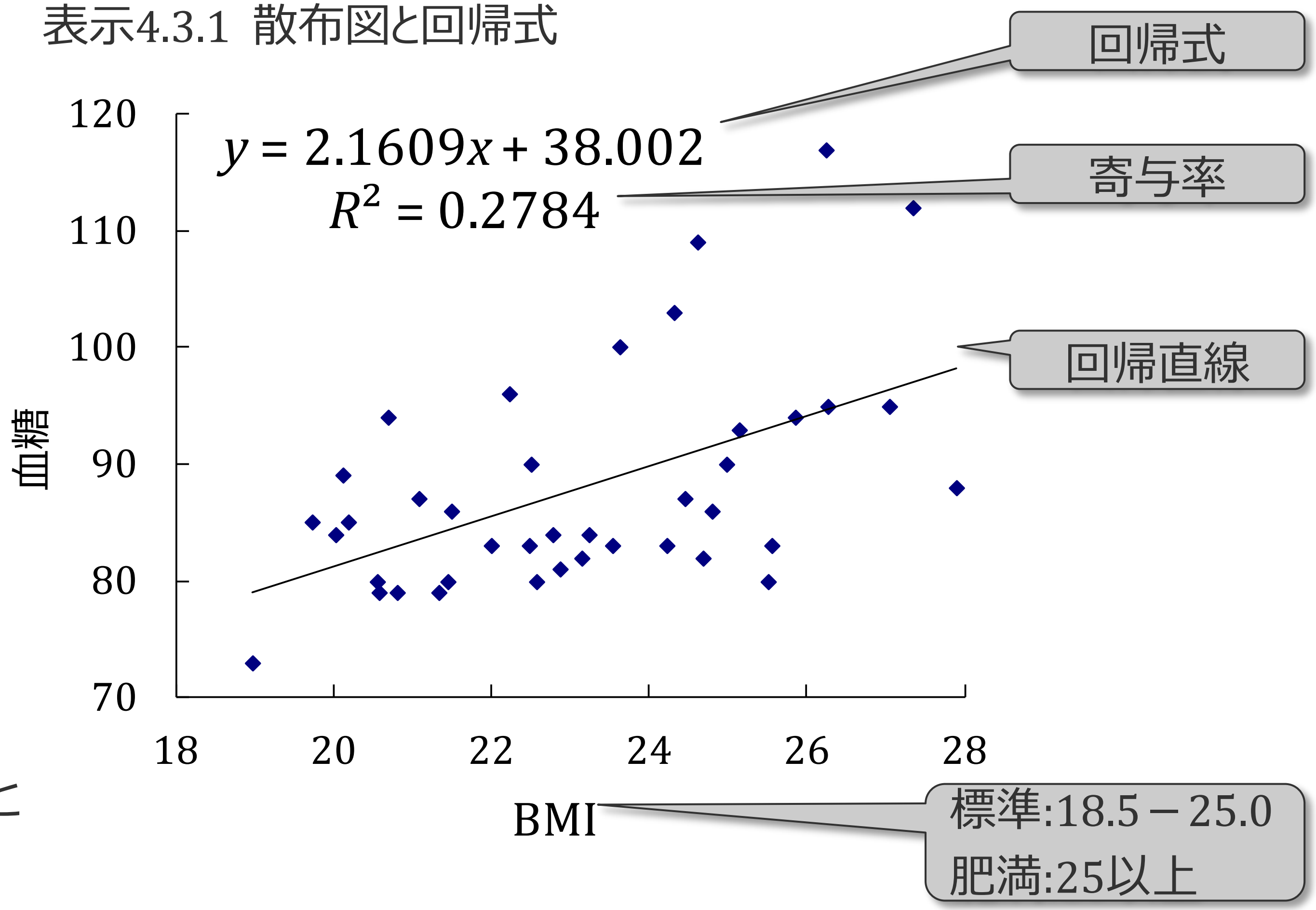
血糖と体重の相関係数：　0.397

散布図と相関係数から、
血糖とBMI の直線関係を確認

↓

回帰直線を用いて、
BMI（x）から血糖（y）が予測できると考えられる

表示4.3.1 散布図と回帰式



回帰直線と回帰式

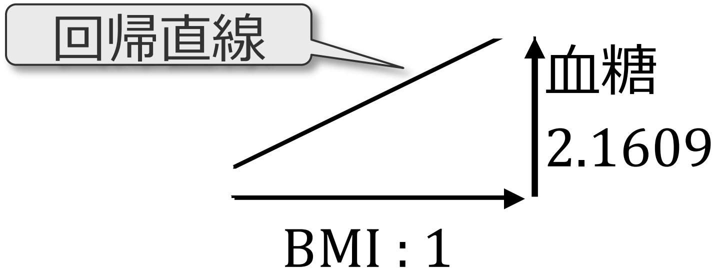
p.226

●回帰式

回帰係数（傾き）

BIM（x）が 1 単位増えると

血糖（y）が平均的に 2.1609 増加



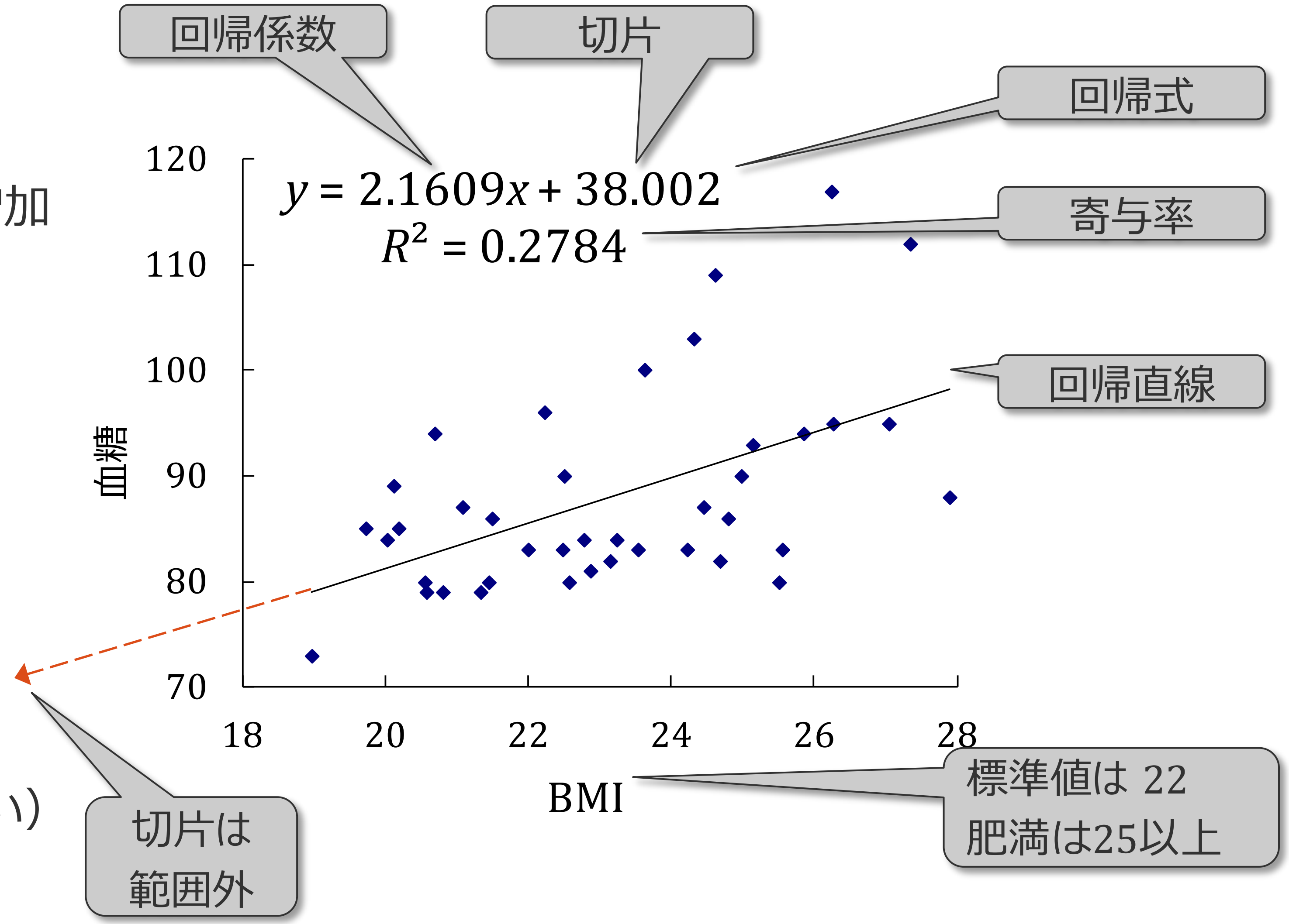
切片

BIM（x）が 0 のときの

血糖（y）の平均値

この場合は意味を持たない

（このようなケースが少なくない）



回帰直線と回帰式

p.226

●回帰式

$y = 2.1609x + 38.002$　（Excel による表示）

$y = 38.002 + 2.1609x$　（一般的に、切片を先に記載）

$y = a + bx$　(4.3.1)　（一般的な形、他には　$y = b_0 + b_1 x$）

y：目的変数、従属変数　予測の対象となる変数（事例 1 では「血糖」）
x：説明変数、独立変数　予測に使う変数（事例 1 では「BMI」）

a：切片、b：回帰係数・・・・・・パラメータ

単回帰分析と重回帰分析

説明変数が１つ　　：単回帰分析　$y = a + bx$
説明変数が複数個：重回帰分析（p.253, p.262）　$y = b_0 + b_1 x_1 + b_2 x_2$

回帰直線と回帰式

p.226

●回帰の言葉の起源

国語辞典での「回帰」の説明

「もとの状態に戻ること」　　例：「自然への回帰」、「北回帰線」

回帰分析の生い立ち

父の身長（x）と息子の身長（y）には直線関係がある

傾きは１以下で、息子の平均身長は全集団の平均値に近づく

（父が平均身長より高い180cmの場合、息子の身長は180cmより小さく、平均に近づく）

平均への回帰（regression 、「後戻り」）→「回帰」　　（F. Galton，1822-1911）

（生物学的な現象ではなく、統計学的な現象）

現在の回帰分析

目的変数 y を説明変数 x に回帰させる（ 言葉の起源とは異なった意味に用いている）

例　血糖値を BMI に回帰させる

（§4.6(5) p.260）

p.226

（2）回帰モデル

相関分析のモデル　：２次元正規分布
単回帰分析のモデル：？

回帰モデル

p.226

●母集団の回帰式とサンプルの回帰式

ある被験者から成る母集団を想定、被験者の BMI と血糖 (x, y) には直線関係がある

２変数(x, y) の直線関係　　$\eta = \alpha + \beta x$　（4.3.2）　η ：母平均 イータ

40 組を抽出して回帰式を得た　　$y = a_1 + b_1 x$　（サンプル１）

別の 40 組を抽出して回帰式を得た　　$y = a_2 + b_2 x$　（サンプル２）

別の 80 組を抽出して回帰式を得た　　$y = a_3 + b_3 x$　（サンプル３）

サンプルごとに、回帰式は異なる
データ数を増やせば、回帰式は安定するが、限りがある
限られたサンプルの回帰式から、
母集団の回帰式 (4.3.2) を推定

母集団
BMI と血糖
(x, y)
$\eta = \alpha + \beta x$

サンプル１
n = 40
y= $a_1 + b_1 x$

サンプル２
n = 40
y= $a_2 + b_2 x$

サンプル３
n = 80
y= $a_3 + b_3 x$

回帰モデル

母集団から被験者を集めてきて
BMI x と血糖値 y を調べると
y の母平均 η は x の値によって決まる

母集団
BMI と血糖
(x, y)
$\eta = \alpha + \beta x$

●単回帰モデル

$\eta = \alpha + \beta x$ （η ：母平均、α：母切片、β：母回帰係数）（4.3.2）

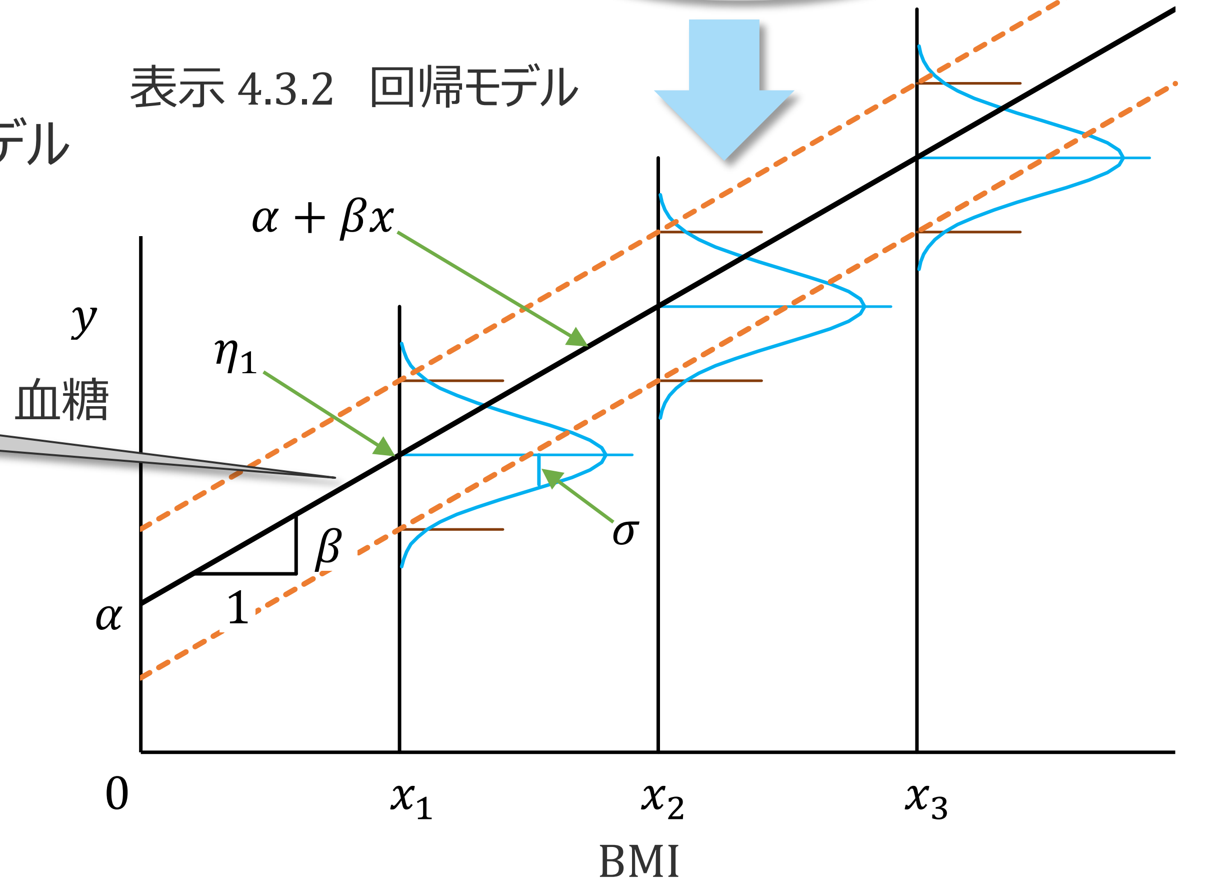
$y = \eta + \varepsilon$ （y ：観測値、ε：誤差）

$y = \eta + \varepsilon = \alpha + \beta x + \varepsilon$ （4.3.3） 単回帰モデル

$\varepsilon \sim \mathrm{N}(0, \sigma^2)$

この直線上に η が存在する

表示 4.3.2 回帰モデル



回帰モデル

p.227

母集団
BMI と血糖
(x, y)
$\eta = \alpha + \beta x$

●単回帰モデル

$\eta = \alpha + \beta x$ （η ：母平均、α：母切片、β：母回帰係数）（4.3.2）

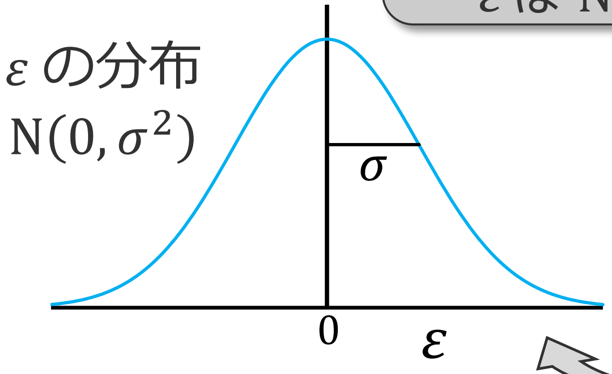
$y = \eta + \varepsilon$ （y ：観測値、ε：誤差）

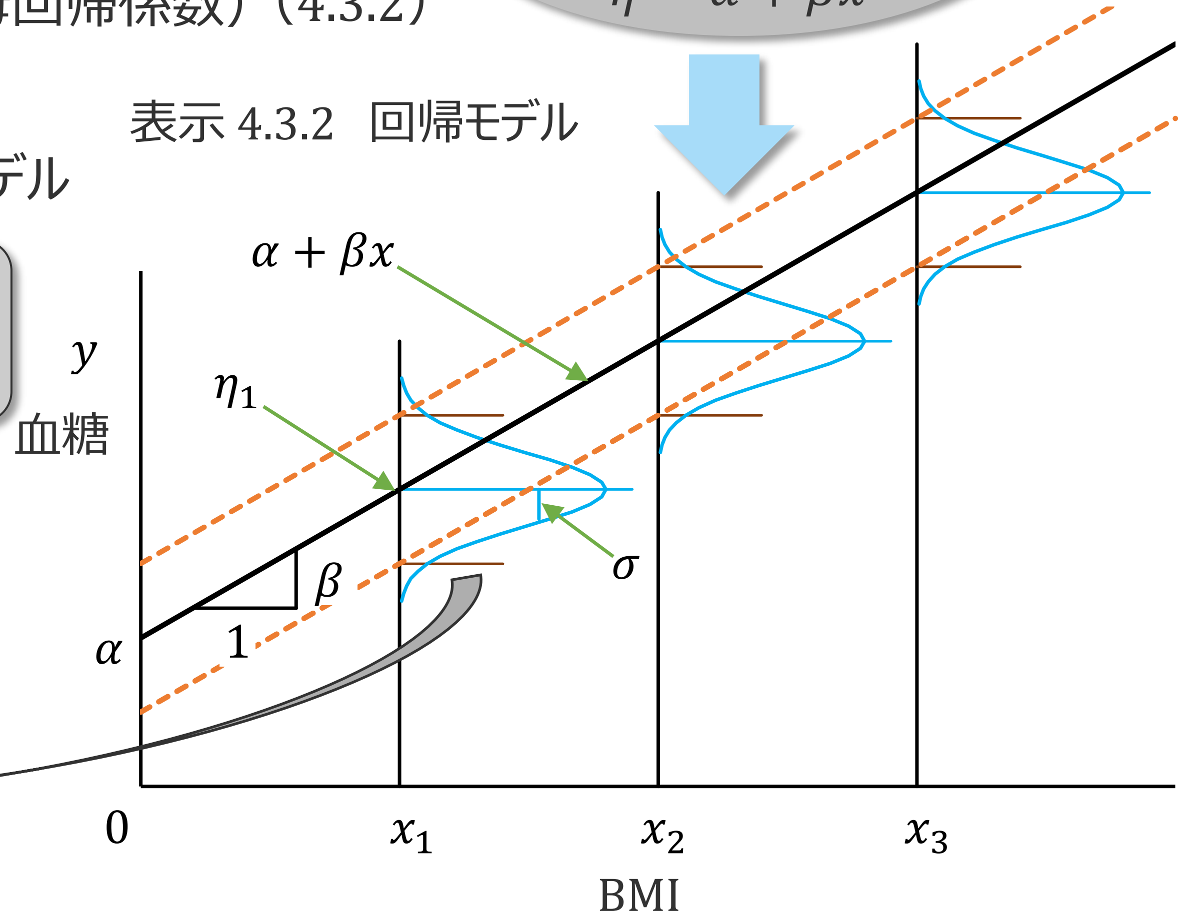
$y = \eta + \varepsilon = \alpha + \beta x + \varepsilon$ （4.3.3） 単回帰モデル

$\varepsilon \sim \mathrm{N}(0, \sigma^2)$

観測値 y は η を中心に変化
その変化を誤差 ε とする
ε は $\mathrm{N}(0, \sigma^2)$ に従う

表示 4.3.2 回帰モデル





回帰モデル

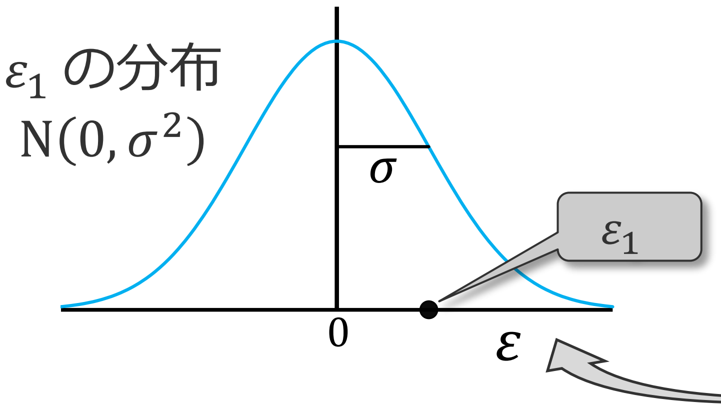
●単回帰モデルの視覚的説明

$\eta = \alpha + \beta x$　　（η ：母平均、α：母切片、β：母回帰係数）（4.3.2）

$y_1 = \eta_1 + \varepsilon_1$　　（y ：観測値、ε：誤差）

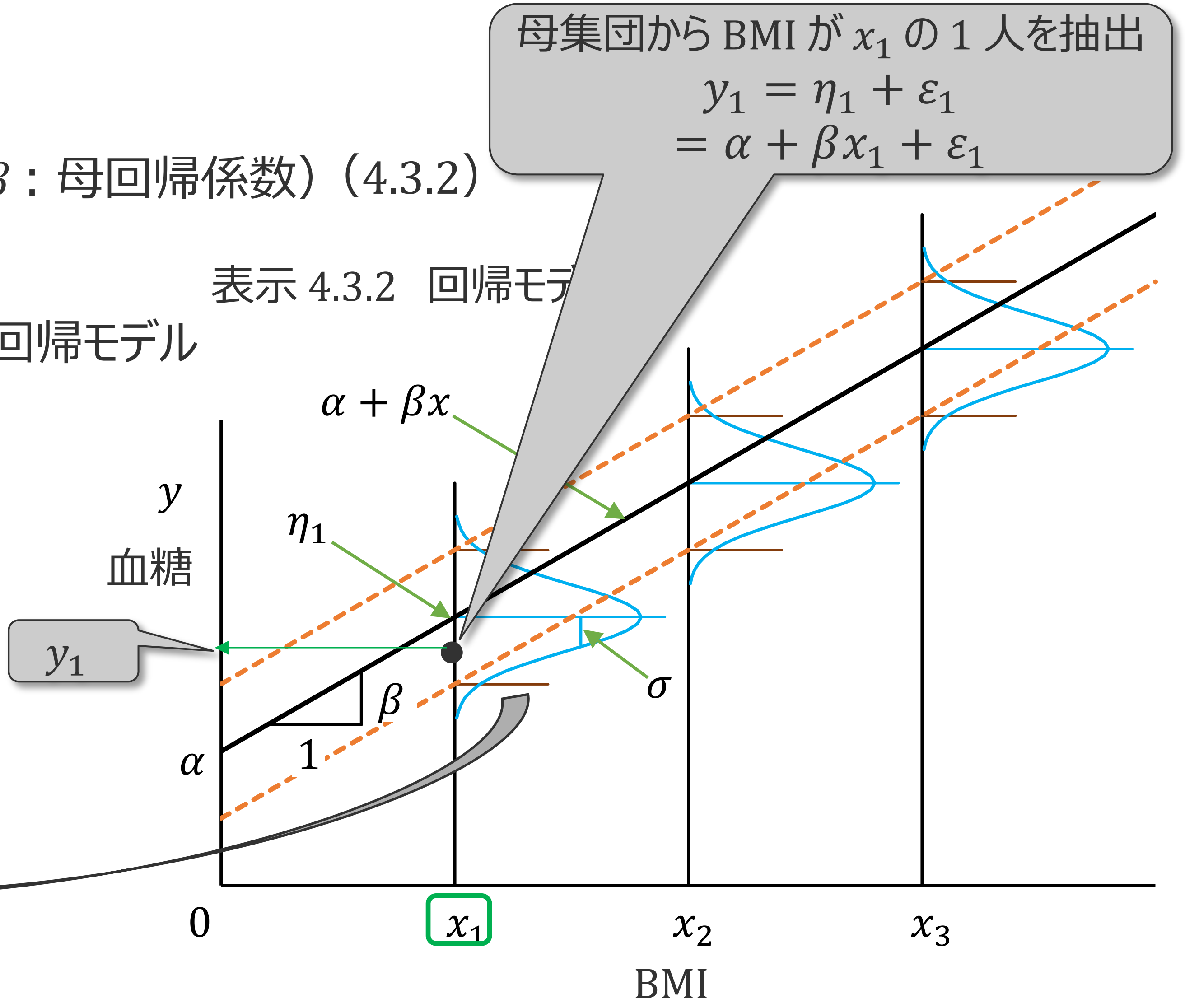
$y_1 = \eta_1 + \varepsilon_1 = \alpha + \beta x_1 + \varepsilon_1$　（4.3.3）　単回帰モデル

$\varepsilon_1 \sim \mathrm{N}(0, \sigma^2)$　誤差



BMI が 特定の値 x_1 の人の血糖 y_1 を測定

$\varepsilon_1 = y_1 - \eta_1$



回帰モデル

p.227

●単回帰モデルの視覚的説明

$\eta = \alpha + \beta x$　　（η ：母平均、α：母切片、β：母回帰係数）（4.3.2）

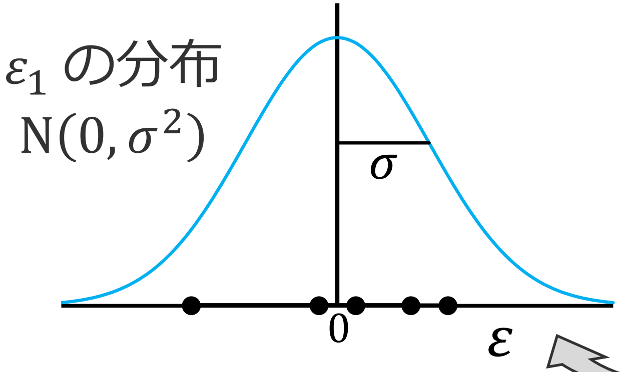
$y_1 = \eta_1 + \varepsilon_1$　　（y ：観測値、ε：誤差）

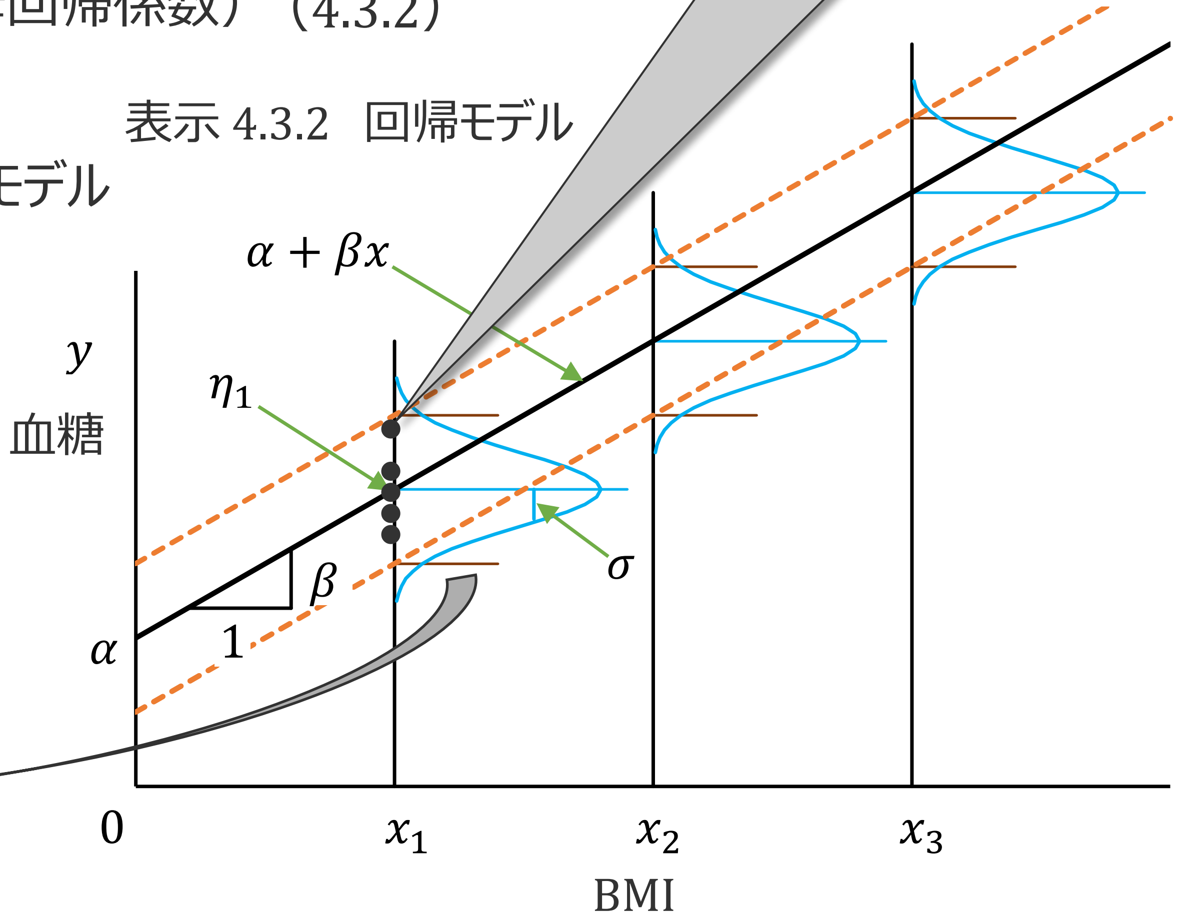
$y_1 = \eta_1 + \varepsilon_1 = \alpha + \beta x_1 + \varepsilon_1$　（4.3.3）　単回帰モデル

$\varepsilon_1 \sim \mathrm{N}(0, \sigma^2)$　誤差

BMI が x_1 の 5 人を抽出
誤差 ε_1 は正規分布に従う

表示 4.3.2　回帰モデル





BMI が 特定の値 x_1 の人の血糖 y_1 を測定

$\varepsilon_1 = y_1 - \eta_1$

回帰モデル

p.227

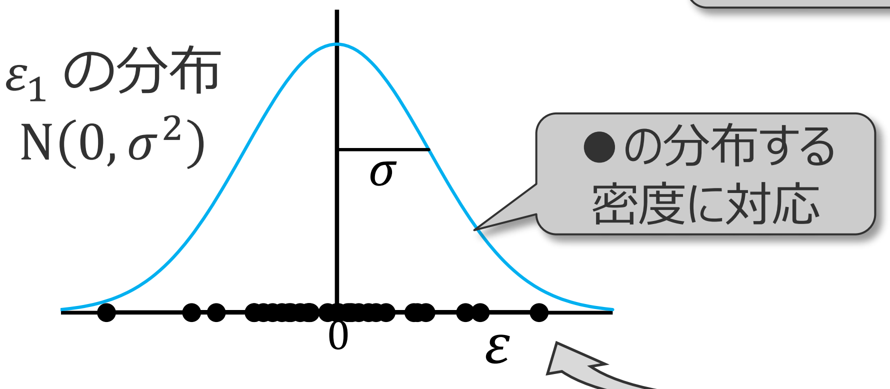
●単回帰モデルの視覚的説明

$\eta = \alpha + \beta x$　（η ：母平均、α：母切片、β：母回帰係数）（4.3.2）

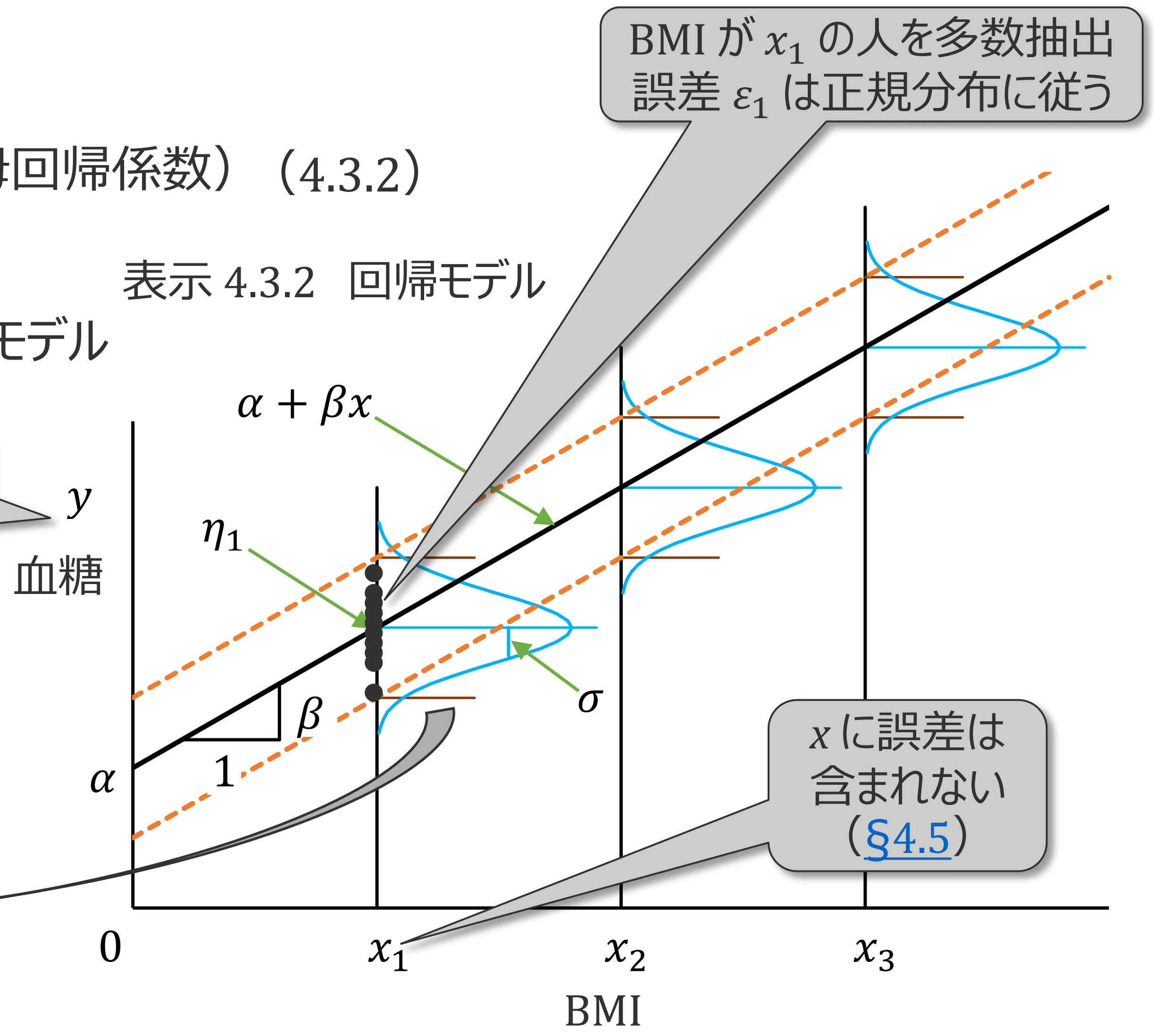
$y_1 = \eta_1 + \varepsilon_1$　（y ：観測値、ε：誤差）

$y_1 = \eta_1 + \varepsilon_1 = \alpha + \beta x_1 + \varepsilon_1$　（4.3.3）　単回帰モデル

$\varepsilon_1 \sim \mathrm{N}(0, \sigma^2)$ 　誤差



表示 4.3.2　回帰モデル



BMI が 特定の値 x_1 の人の血糖 y_1 を測定

$\varepsilon_1 = y_1 - \eta_1$

回帰モデル

p.227

●単回帰モデルの視覚的説明

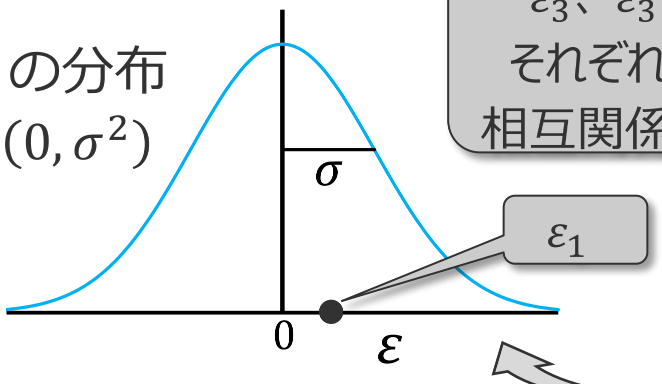
$$y_1 = \eta_1 + \varepsilon_1 = \alpha + \beta x_1 + \varepsilon_1 \quad \varepsilon_1 \sim \mathrm{N}(0, \sigma^2)$$

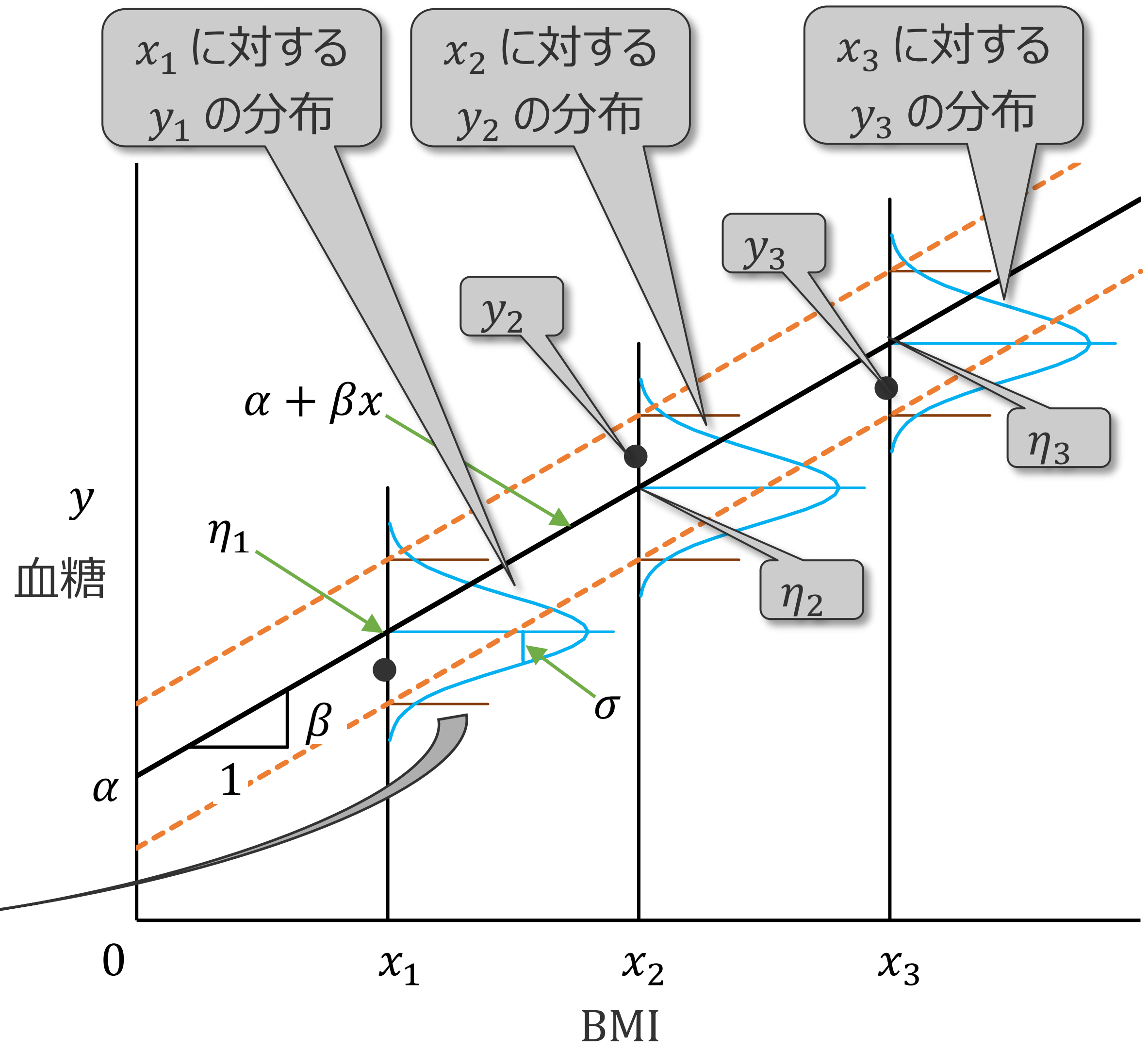
$$y_2 = \eta_2 + \varepsilon_2 = \alpha + \beta x_2 + \varepsilon_2 \quad \varepsilon_2 \sim \mathrm{N}(0, \sigma^2)$$

$$y_3 = \eta_3 + \varepsilon_3 = \alpha + \beta x_3 + \varepsilon_3 \quad \varepsilon_3 \sim \mathrm{N}(0, \sigma^2)$$

(4.3.3)

ε_3、ε_3、ε_3
それぞれ独立
相互関係はない





BMI が 特定の値 x_1 の人の血糖 y_1 を測定

$$\varepsilon_1 = y_1 - \eta_1$$

回帰モデル

p.227

●単回帰モデルの視覚的説明

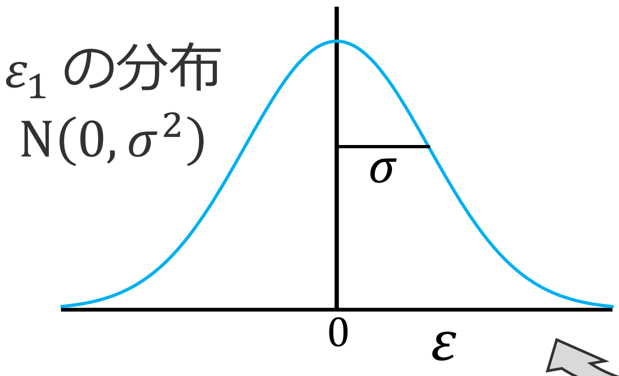
$\eta = \alpha + \beta x$ （η ：母平均） (4.3.2)

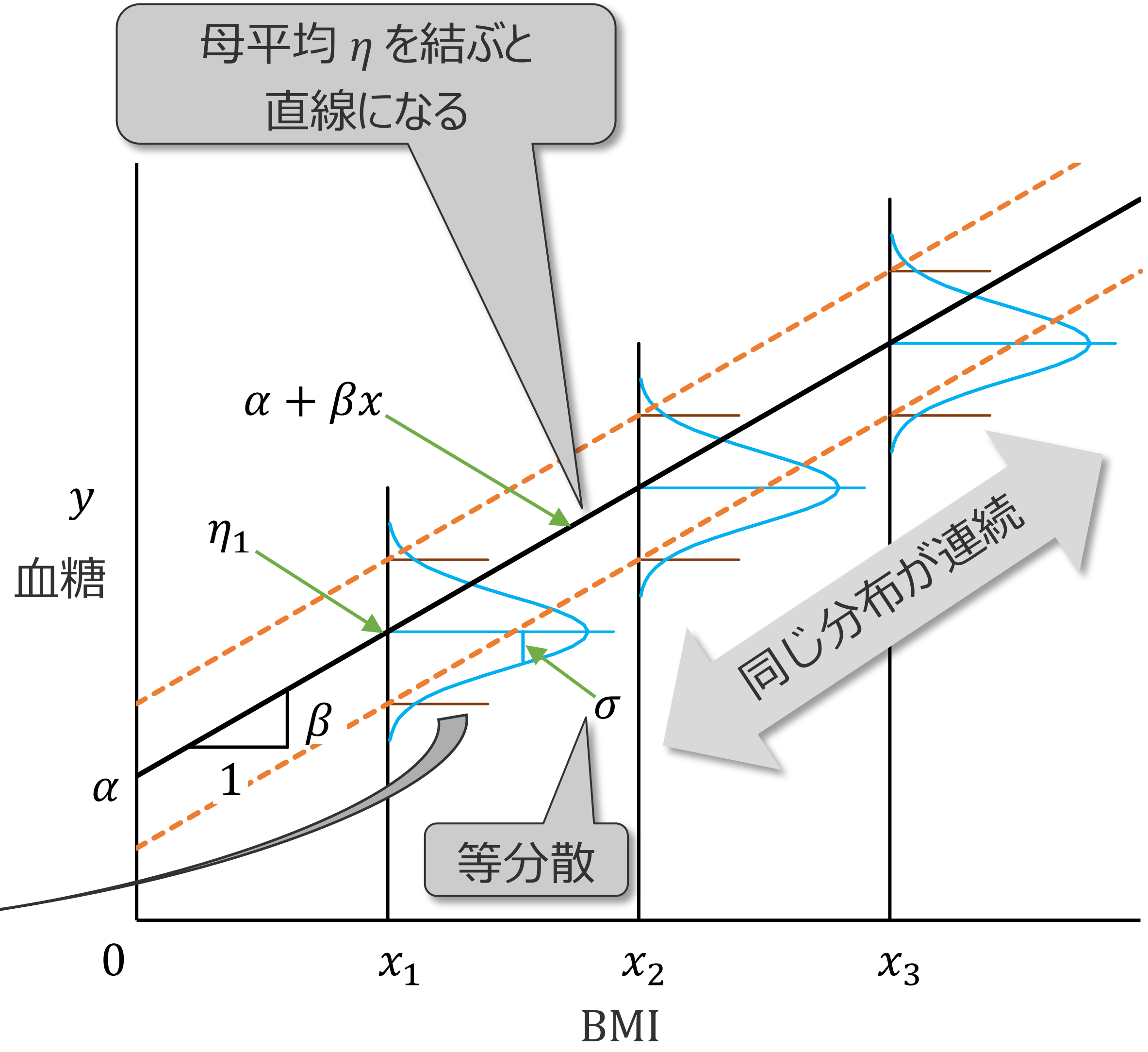
$y = \eta + \varepsilon$ （y ：観測値）

$y = \eta + \varepsilon = \alpha + \beta x + \varepsilon$ (4.3.3)

$\varepsilon \sim \mathrm{N}(0, \sigma^2)$ 誤差

単回帰モデル





BMI が 特定の値 x_1 の人の血糖 y_1 を測定

$\varepsilon_1 = y_1 - \eta_1$

回帰モデル

p.227

●単回帰モデルの視覚的説明

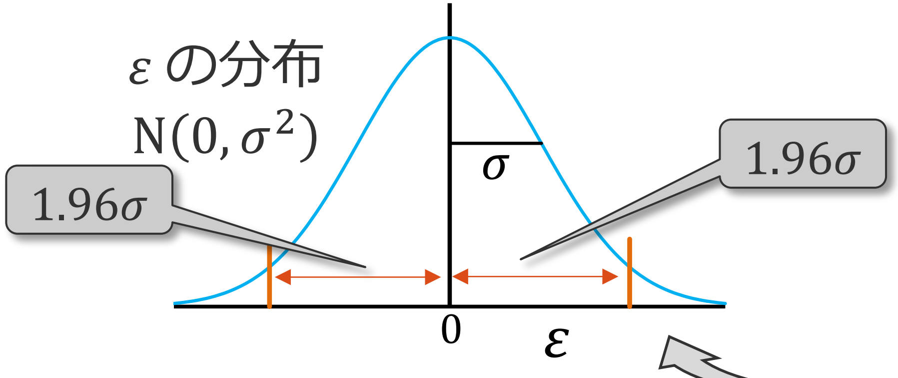
$\eta = \alpha + \beta x$ （η ：母平均） （4.3.2）

$y = \eta + \varepsilon$ （y ：観測値）

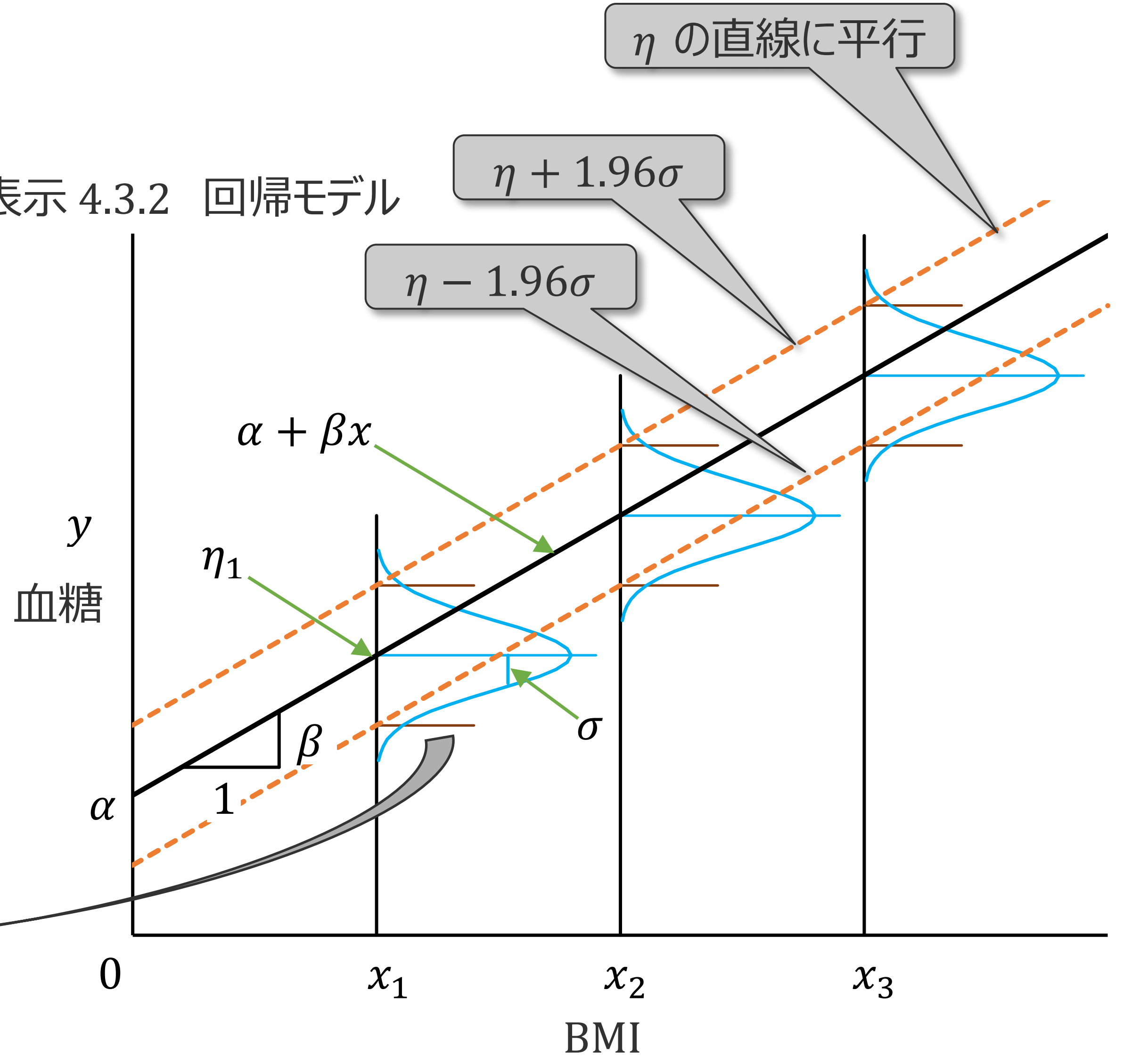
$y = \eta + \varepsilon = \alpha + \beta x + \varepsilon$ （4.3.3）

単回帰モデル

$\varepsilon \sim \mathrm{N}(0, \sigma^2)$ 誤差



表示 4.3.2 回帰モデル



BMI が 特定の値 x_1 の人の血糖 y_1 を測定

$\varepsilon_1 = y_1 - \eta_1$

回帰モデル

p.227

●単回帰モデルの前提

$\eta = \alpha + \beta x$　（η ：母平均）　（4.3.2）

$y = \eta + \varepsilon$　（y ：観測値）

$y = \eta + \varepsilon = \alpha + \beta x + \varepsilon$　（4.3.3） 単回帰モデル

$\varepsilon \sim \mathrm{N}(0, \sigma^2)$

単回帰モデルを想定して解析

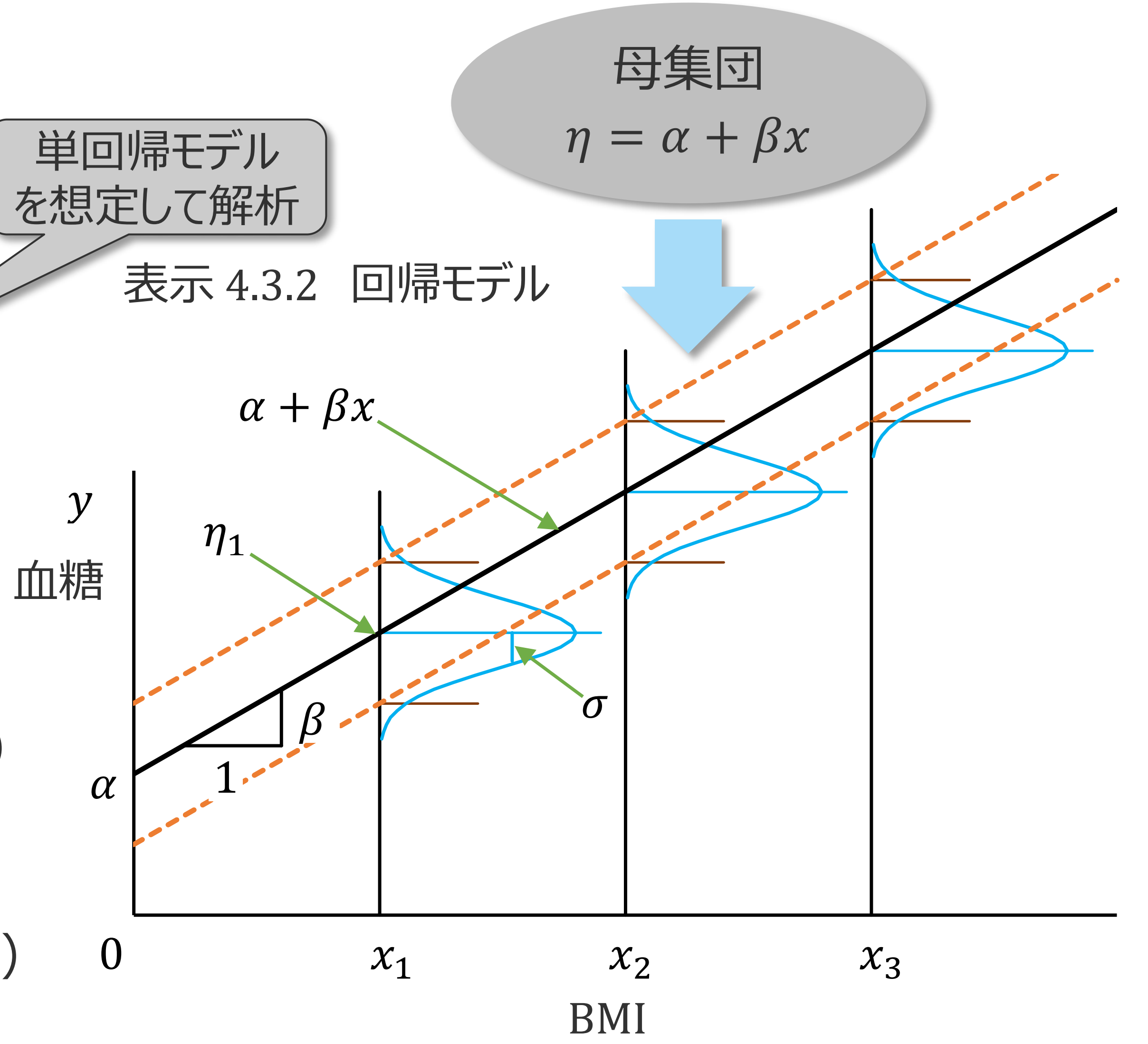
（1） y の方向には誤差があるが 、
x の方向には誤差は含まれない

（2） 誤差は互いに独立で,
各 y について相互に影響関係はない（独立性）

（3） 誤差の期待値は 0 である（不偏性）

（4） 誤差の大きさ σ は、
x の値にかかわらず一定である（等分散性）

（5） 誤差は正規分布にしたがう（正規性）



表示 4.3.2　回帰モデル

p.228

（3）最小２乗法による回帰式の推定

（3） 試行錯誤による回帰式の推定（最小２乗法）
（4）Excel のソルバーによる回帰式の推定（最小２乗法）
（5）平方和の分解と σ^2 の推定（最小２乗法）
（6）数理統計学的な回帰式の推定（最小２乗法）
（7）Excel の LINEST 関数による回帰式の推定（最小２乗法）

テキストの項のタイトルとやや異なる

最小２乗法

p.228

●最小２乗法による平均値の推定

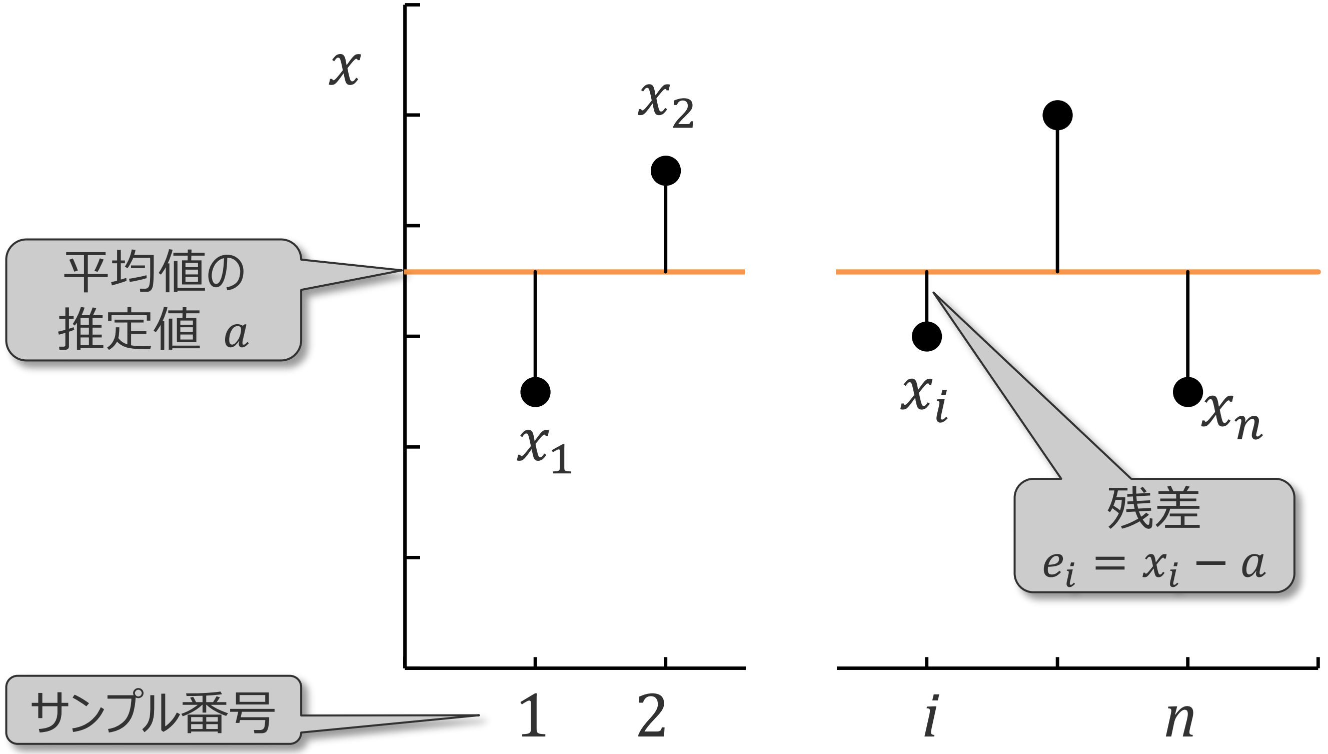
n 個の観測値 x_i から平均値 a を求める（ §2.1 (2)　p.55）
暫定的に a を設定、各観測値 x_i から a までの距離（残差 e_i ）の２乗和（残差平方和 S）を算出
a を移動させて、残差平方和 S が最小になったとき、その a が平均値

$$S = \sum_{I=1}^{n} {e_i}^2$$

$$= \sum_{I=1}^{n} (x_i - a)^2 \Rightarrow \min \qquad (4.3.4)$$

残差

この考え方と同様にして回帰式を推定



最小２乗法

●最小２乗法による回帰式の推定

n 組の観測値（x_i, y_i）から

母切片α、母回帰係数 β の推定値 a, b を求める

回帰式から求められる予測値：$\hat{y}_i$（ワイハット）

$$\hat{y}_i = a + bx_i$$

残差平方和 S を最小にする a, b が推定値

観測値

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

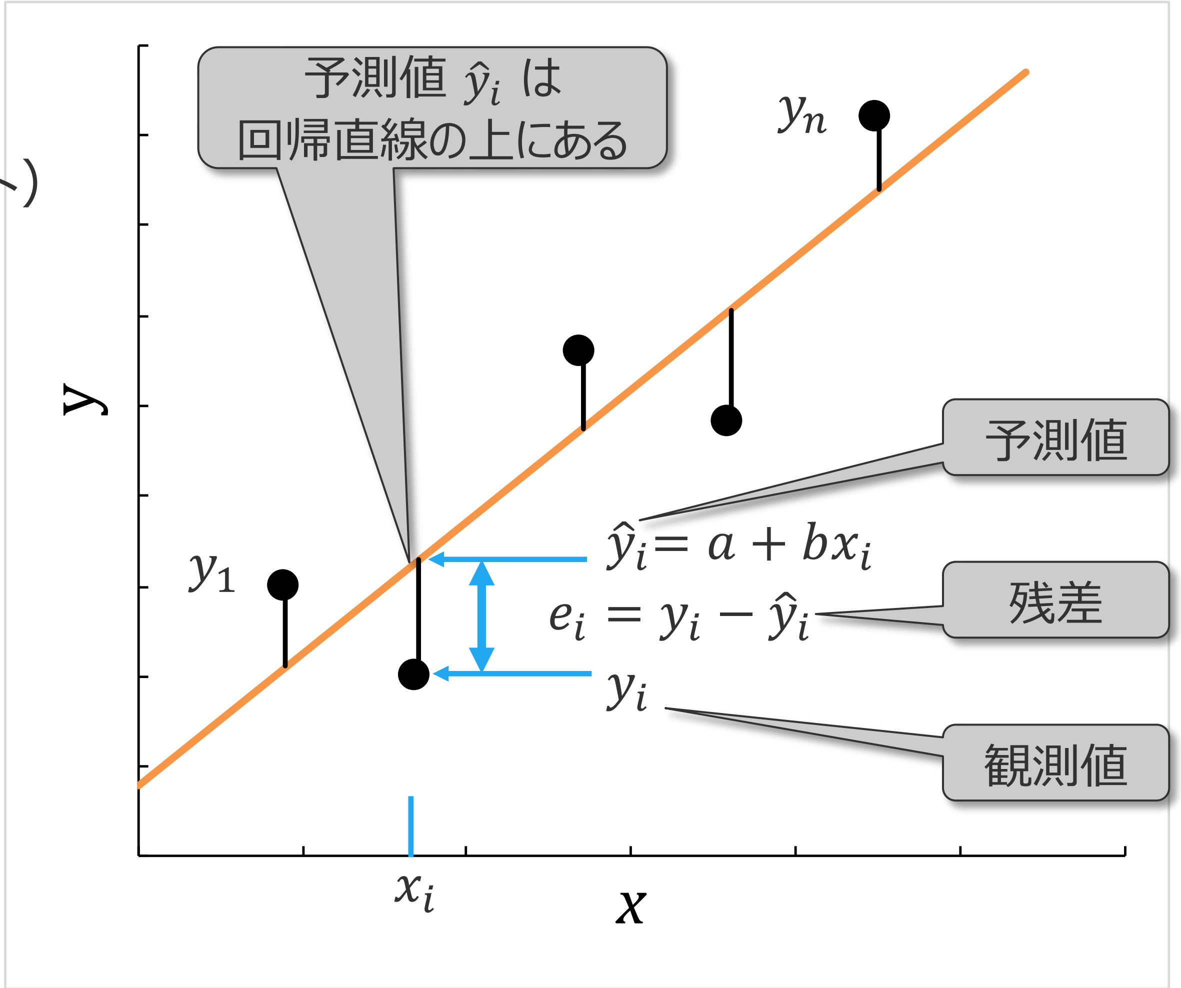
予測値（ワイハット）

$$= \sum_{I=1}^{n} \left(y_i - (a + bx_i)\right)^2$$

$$= \sum_{I=1}^{n} {e_i}^2 \quad \Rightarrow \min \qquad (4.3.5)\ (4.3.6)$$

残差

表示 4.3.5　残差との関係



試行錯誤による回帰式の推定（最小２乗法）

p.229

●事例２

Excel ファイル「基礎改4.xls」、名前ボックスから「表示4.3.3」（Fig43_03）を選択

データと回帰式

６組の２変数データ

最小２乗法で回帰式を当てはめる

（BMI と血糖の事例１から離れる）

求める回帰式

Excel の［近似曲線の追加］で計算

$y = 3.9 + 0.62x$ ・・・求める回帰式

a *b*

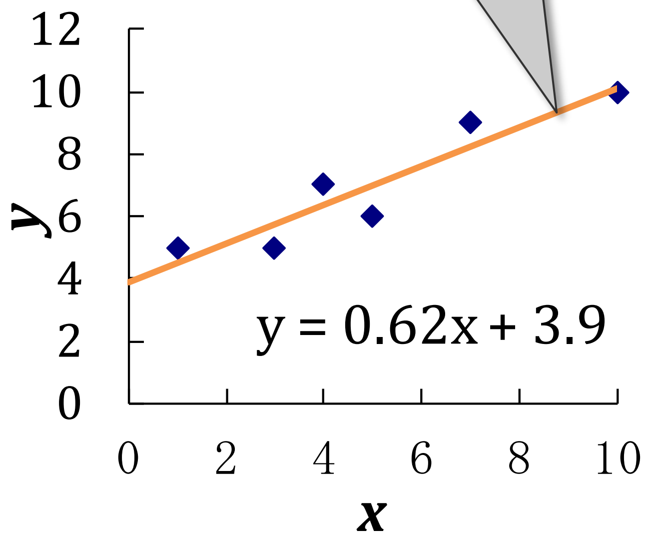
この回帰式をどのようにして求めるか

（様々な方法で説明）

表示4.3.3 データと散布図

i	x	y
1	1	5
2	3	5
3	4	7
4	5	6
5	7	9
6	10	10
平均	5.0	7.0
平方和	50.0	22.0
積和	31.0	

Excelの
［近似曲線の追加］



試行錯誤による回帰式の推定（最小２乗法）

p.229

●Excel の計算式とグラフ

名前ボックスから「表示4.3.4」（Fig43_04）を選択

表示 4.3.4 を利用して、最小２乗法に基づき、試行錯誤で回帰式を求める

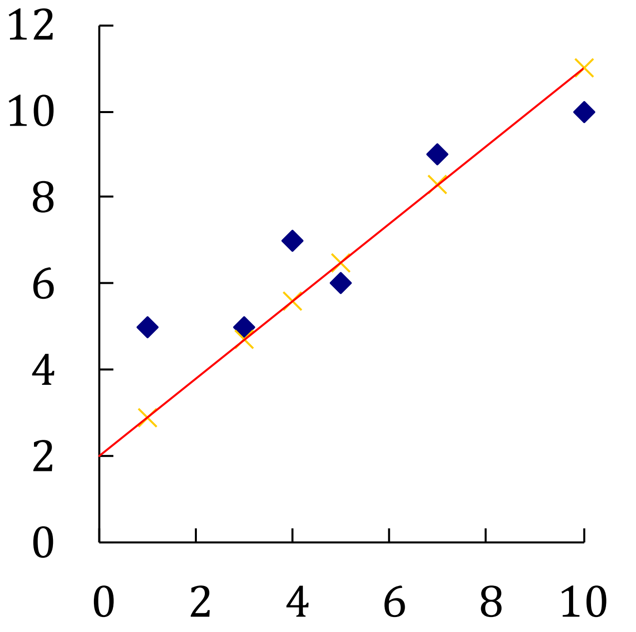
（表示4.3.3 と同じデータ）

表示4.3.4 a, b の値による回帰曲線の変化

	A	B	C	D	E
16	i	x	y	y-hat	e
17	1	1	5	2.9	2.1
18	2	3	5	4.7	0.3
19	3	4	7	5.6	1.4
20	4	5	6	6.5	-0.5
21	5	7	9	8.3	0.7
22	6	10	10	11.0	-1.0

	G	H
17	a	2.0
18	b	0.9
19	S	8.2

表示 4.3.3 と同じ



D17: = H17 + H18 * B17

E17: = C17 − D17

H19: = SUMSQ(E17:E22)

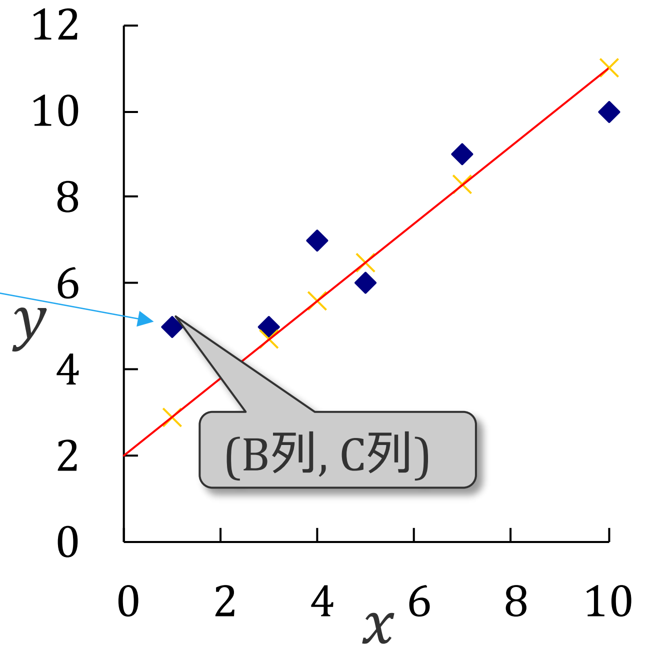
試行錯誤による回帰式の推定（最小２乗法）

p.229

- Excel の計算式とグラフ
  - B 列：x
  - C 列：y
  - D 列：y-hat（予測値、$a + bx_i$）
  - E 列：e（残差、$y_i - \hat{y}_i$）
  - H17:H18：パラメータ a, b
  - H19：残差平方和 S

表示4.3.4　a, b の値による回帰曲線の変化

	A	B	C	D	E	F	G	H	I	J
16	i	x	y	y-hat	e					
17	1	1	5	2.9	2.1		a	2.0		
18	2	3	5	4.7	0.3		b	0.9		
19	3	4	7	5.6	1.4		S	8.2		
20	4	5	6	6.5	-0.5					
21	5	7	9	8.3	0.7					
22	6	10	10	11.0	-1.0					



D17: = \$H\$17 + \$H\$18 * B17

E17: = C17 − D17

H19: = SUMSQ(E17:E22)

試行錯誤による回帰式の推定（最小２乗法）

p.229

● Excel の計算式とグラフ

B 列：x

C 列：y

D 列：y-hat（予測値、$a + bx_i$）

E 列：e（残差、$y_i - \hat{y}_i$）

H17:H18：パラメータ a, b

H19：残差平方和 S

回帰式から得られる予測値
$\hat{y}$，y−hat（ワイハット）

パラメータ
暫定的に設定

	A	B	C	D	E	F	G	H	I
16	i	x	y	y-hat	e				
17	1	1	5	2.9	2.1		a	2.0	
18	2	3	5	4.7	0.3		b	0.9	
19	3	4	7	5.6	1.4		S	8.2	
20	4	5	6	6.5	-0.5				
21	5	7	9	8.3	0.7				
22	6	10	10	11.0	-1.0				

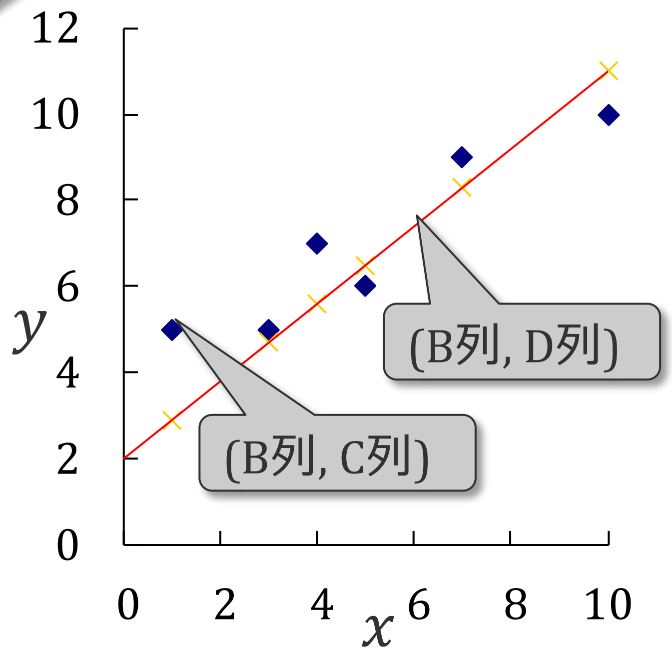
$\hat{y}_i = a + bx_i$

D17: = \$H\$17 + \$H\$18 * B17

E17: = C17 − D17

H19: = SUMSQ(E17:E22)

＄マーク
オートフィルで下方にコピーする時
参照しているセル番地が変わらない



試行錯誤による回帰式の推定（最小２乗法）

p.229

●Excel の計算式とグラフ

B 列：x

C 列：y

D 列：y-hat（予測値、$a + bx_i$）

E 列：e（残差、$y_i - \hat{y}_i$）

H17:H18：パラメータ a, b

H19：残差平方和 S

回帰式から得られる予測値
$\hat{y}$，y-hat（ワイハット）

パラメータ
暫定的に設定

	A	B	C	D	E
16	i	x	y	y-hat	e
17	1	1	5	2.9	2.1
18	2	3	5	4.7	0.3
19	3	4	7	5.6	1.4
20	4	5	6	6.5	-0.5
21	5	7	9	8.3	0.7
22	6	10	10	11.0	-1.0

a	2.0
b	0.9
S	8.2

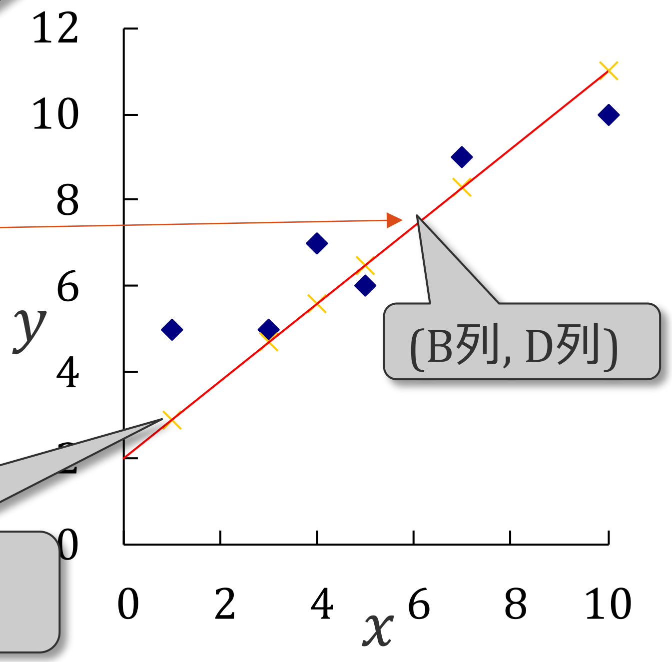
$\hat{y}_i = a + bx_i$

D17: = H17 + H18 * B17

E17: = C17 – D17

H19: =

Excel の「近似曲線の追加」で
表示した直線回帰ではない



試行錯誤による回帰式の推定（最小２乗法）

p.229

●Excel の計算式とグラフ

B 列：x

C 列：y

D 列：y-hat（予測値、$a + bx_i$）

E 列：e（残差、$y_i - \hat{y}_i$）

H17:H18：パラメータ a, b

H19：残差平方和 S

回帰式から得られる予測値 $\hat{y}$，y-hat（ワイハット）

残差

パラメータ

	A	B	C	D	E	F	G	H	I
16	i	x	y	y-hat	e				
17	1	1	5	2.9	2.1		a	2.0	
18	2	3	5	4.7	0.3		b	0.9	
19	3	4	7	5.6	1.4		S	8.2	
20	4	5	6	6.5	-0.5				
21	5	7	9	8.3	0.7				
22	6	10	10	11.0	-1.0				

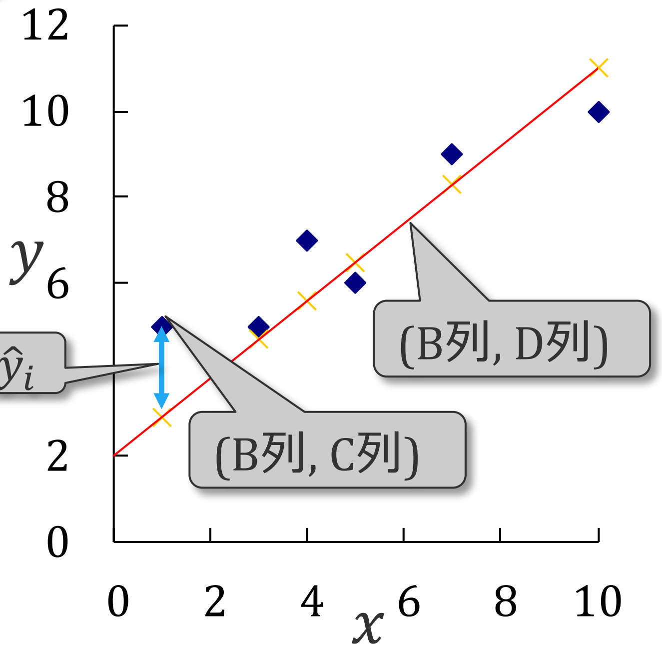
$\hat{y}_i = a + bx_i$

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

D17: = \$H\$17 + \$H\$18 * B17

E17: = C17 − D17

H19: = SUMSQ(E17:E22)



試行錯誤による回帰式の推定（最小２乗法）

p.229

●Excel の計算式とグラフ

B 列：x

C 列：y

D 列：y-hat（予測値、$a + bx_i$）

E 列：e（残差、$y_i - \hat{y}_i$）

H17:H18：パラメータ a, b

H19：残差平方和 S

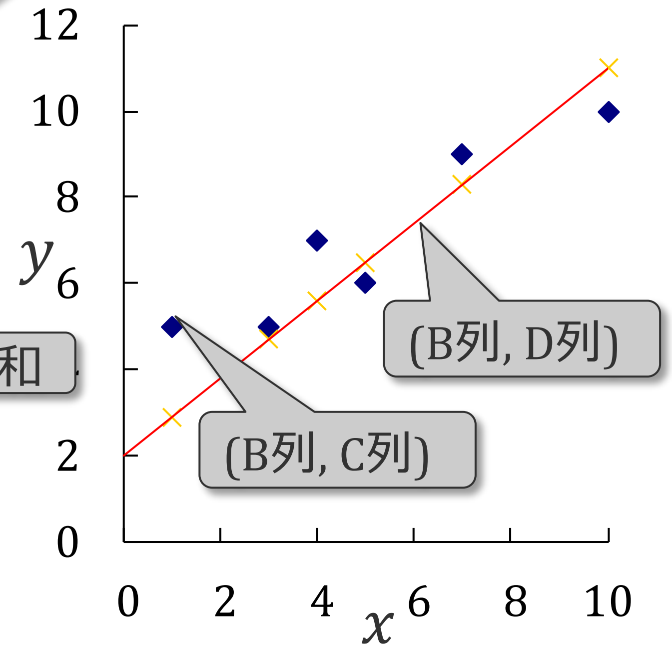
回帰式から得られる予測値 $\hat{y}$，y-hat（ワイハット）

残差

パラメータ

	A	B	C	D	E	F	G	H	I	J
16	i	x	y	y-hat	e					
17	1	1	5	2.9	2.1		a	2.0		
18	2	3	5	4.7	0.3		b	0.9		
19	3	4	7	5.6	1.4		S	8.2		
20	4	5	6	6.5	-0.5					
21	5	7	9	8.3	0.7					
22	6	10	10	11.0	-1.0					

残差平方和



$\hat{y}_i = a + bx_i$

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

$$S = \sum_{i=1}^{n} {e_i}^2$$

D17: = H17 + H18 * B17

E17: = C17 − D17

H19: = SUMSQ(E17:E22)

試行錯誤による回帰式の推定（最小２乗法）

p.229

●Excel の計算式とグラフ

回帰直線は（B列, D列）から描画（Excelの［近似曲線の追加］で表示した回帰直線ではない）

a , b （H17, H18）の値を変えると、y-hat（D 列）が変化、e（E 列）が変化、S（H19）が変化

↓

B列 と C列から描画した回帰直線が変化

a, b を設定して S を求める

$$\hat{y}_i = a + bx_i$$

↓

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

↓

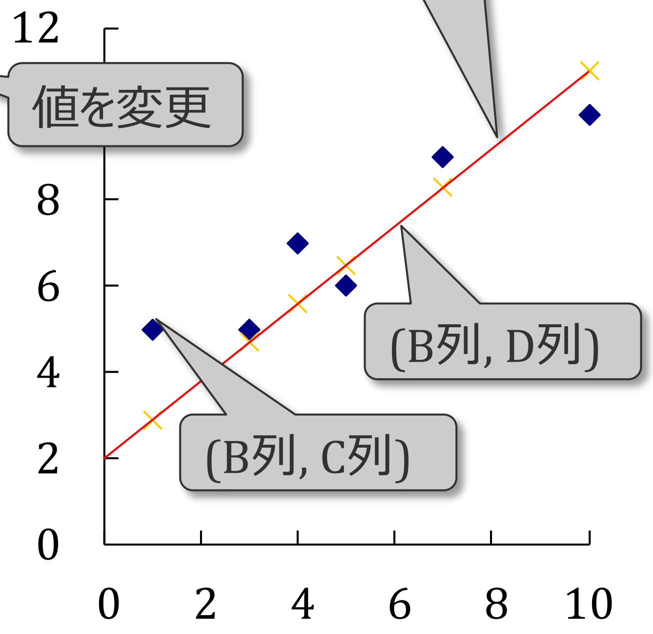
$$S = \sum_{i=1}^{n} {e_i}^2$$

表示4.3.4 a, b の値による回帰曲線の変化

	A	B	C	D	E	F	G	H	I	J
16	i	x	y	y-hat	e					
17	1	1	5	2.9	2.1		a	2.0		
18	2	3	5	4.7	0.3		b	0.9		
19	3	4	7	5.6	1.4		S	8.2		
20	4	5	6	6.5	-0.5					
21	5	7	9	8.3	0.7					
22	6	10	10	11.0	-1.0					

パラメータ

値を変更



D17: = H17 + H18 * B17

E17: = C17 − D17

H19: = SUMSQ(E17:E22)

試行錯誤による回帰式の推定（最小２乗法）

p.229

●試行錯誤による回帰式の推定

a と b のセルの値を変化させて、回帰直線をプロットした観測点にフィットさせる（演習 4.3.1）

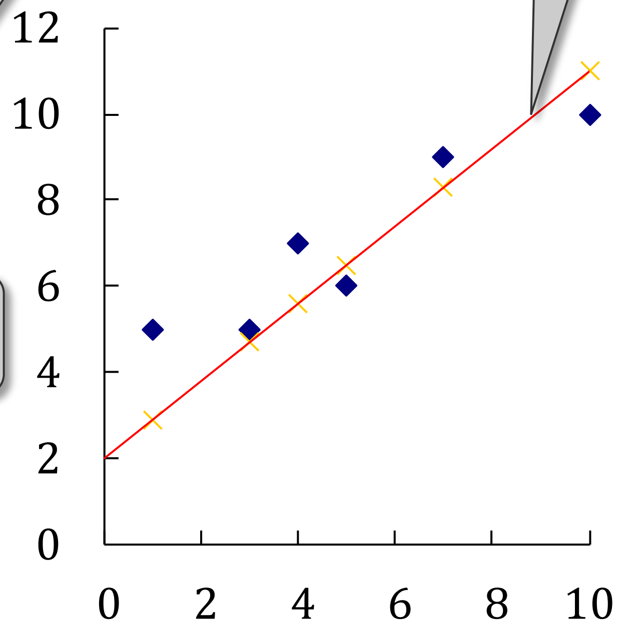
S が最小になるように、a で切片を動かし、b で回帰係数を動かす

Excel には自動的にフィットさせる機能がある → ソルバー

aとbを変化させる

回帰直線が動く

	A	B	C	D	E	F	G	H
16	i	x	y	y-hat	e			
17	1	1	5	2.9	2.1		a	2.0
18	2	3	5	4.7	0.3		b	0.9
19	3	4	7	5.6	1.4		S	8.2
20	4	5	6	6.5	-0.5			
21	5	7	9	8.3	0.7			
22	6	10	10	11.0	-1.0			



Sが変化する

a, b を設定して S を求める

$$\hat{y}_i = a + bx_i$$

↓

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

↓

$$S = \sum_{i=1}^{n} {e_i}^2$$

D17: = \$H\$17 + \$H\$18 * B17

E17: = C17 − D17

H19: = SUMSQ(E17:E22)

（4）ソルバーによる解法

最小２乗法による回帰式の推定

（3） 試行錯誤による回帰式の推定（最小２乗法）
（4）Excel のソルバーによる回帰式の推定（最小２乗法）
（5）平方和の分解と σ^2 の推定（最小２乗法）
（6）数理統計学的な回帰式の推定（最小２乗法）
（7）Excel の LINEST 関数による回帰式の推定（最小２乗法）

テキストの項のタイトルとやや異なる

Excel のソルバーによる回帰式の推定（最小２乗法）

p.230

●Excel の最適化ツール

ゴールシーク

１変数の関数で、y がある値（目標値）になる x を求める機能・・・・・ $y=f(x)$ （4.3.7）

計算式が記録されているセル（y）が指定した値（目標値）になるように、

計算式が参照しているセル（x）の値を調整する（§3.6(4)　p.163）

ソルバー

多変数の関数で、y がある値（目標値）になる複数の x_i を求める機能

$$y=f(x_1,x_2,x_3,\cdots,x_n) \quad (4.3.8)$$

計算式が記録されているセル（y）が「指定した値」「最大」「最小」になるように、

計算式が参照している複数のセル（x_i）の値を調整する（ゴールシークの強化）

今回の事例例：S を最小にする a と b を求める・・・・・・・・・・・ $S=f(a,b)$ （4.3.9）

（前項(3) で試行錯誤により S を最小にする a と b を求めた → この作業を自動的に実施）

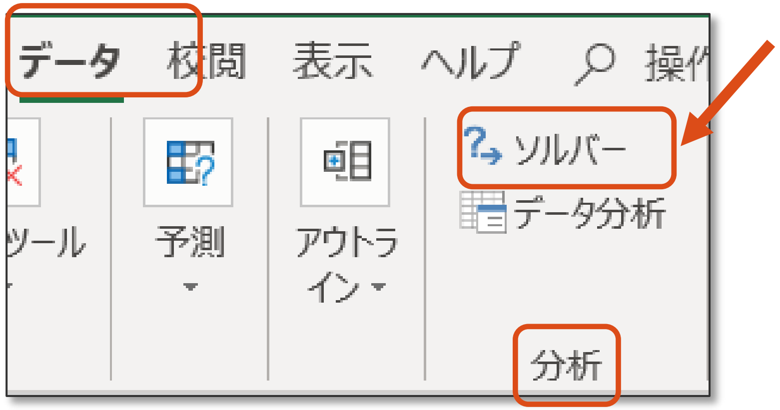
Excel のソルバーによる回帰式の推定（最小２乗法）

p.230

●ソルバーの確認とインストール

ソルバーの選択

［データ］＞［分析］＞［ソルバー］



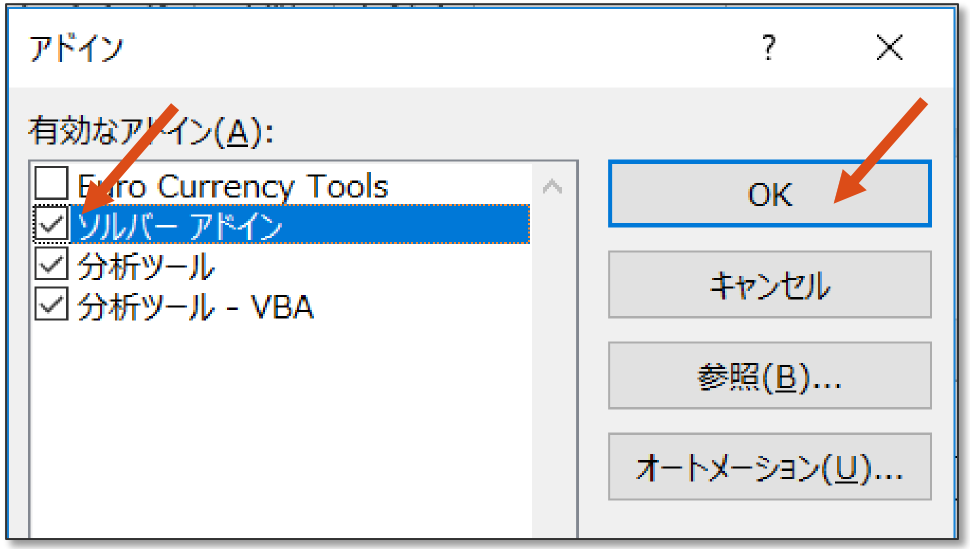
ソルバーのインストール

［ソルバー］が表示されない場合は、インストールが必要

［ファイル］＞［オプション］＞［アドイン］＞

［設定］＞［ソルバーアドイン］にチェック ＞［OK］

（p.235 参照）

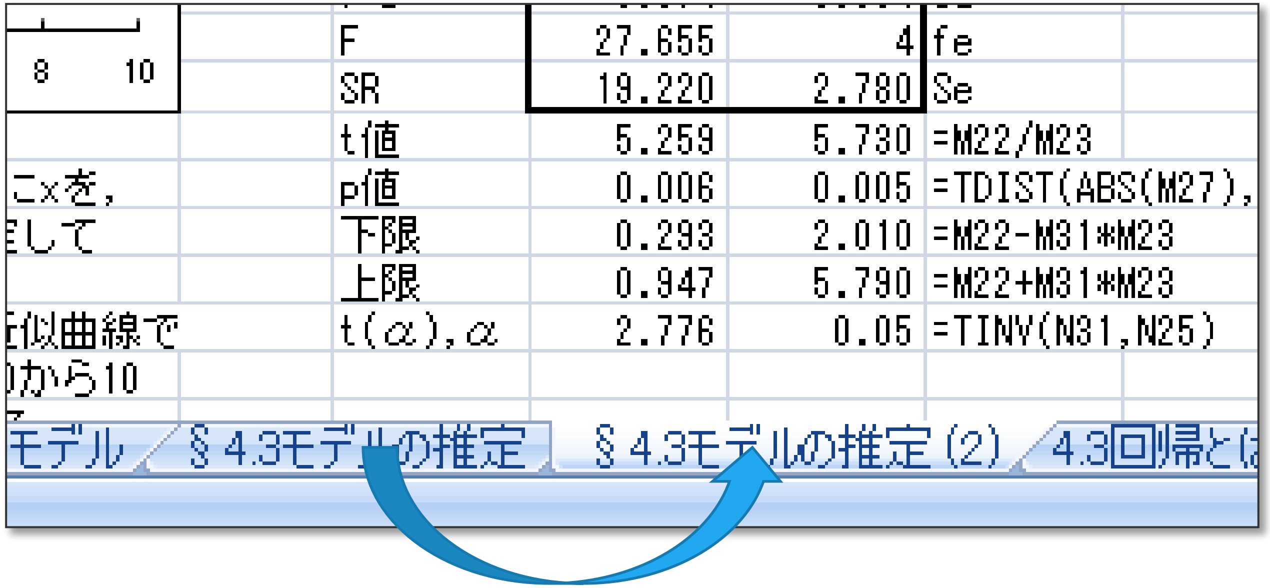


Excel のソルバーによる回帰式の推定（最小２乗法）

p.230

●シートのコピー

始める前に「§4.3モデルの推定」シートをコピー
コピーしたシートで実習する



Ctrlキーを押しながらドラック

Excel のソルバーによる回帰式の推定（最小２乗法）

p.230

変更
a：3.90 → 2.0
b：0.62 → 0.9

●Excel の計算式

表示4.3.6は、表示4.3.4と同じデータと構造

ソルバーで解を求めた結果が表示されている
操作する前に、a とb のセルの数値を変更

a：3.90 → 2.0

b：0.62 → 0.9

（残差平方和 S を最小にする最適な
パラメータではない）

表示4.3.6　ソルバーの解

	L	M	N	O		Q	R	S
3	i	x	y	y-hat	e			
4	1	1	5	4.52	0.48		a	3.900
5	2	3	5	5.76	-0.76		b	0.620
6	3	4	7	6.38	0.62		S	2.78
7	4	5	6	7.00	-1.00			
8	5	7	9	8.24	0.76			
9	6	10	10	10.10	-0.10			
10	平均	5.00	**7.00**	**7.00**	**0.00**			
11	平方和	50.00	**22.00**	**19.22**	**2.78**			
12								
13	相関係数	x	y	y-hat	e			
14	x	1.000	0.935	1.000	**0.00**			
15	y	0.935	1.000	0.935	0.36			
16	y-hat	1.00	0.93	1.00	**0.00**			
17	e	0.00	0.36	0.00	1.00			

Excel のソルバーによる回帰式の推定（最小２乗法）

p.230

● Excel の計算式

表示4.3.6は、表示4.3.4と同じデータと構造

ソルバーの機能

a、b の値を暫定的に設定して S を求める

$$\hat{y}_i = a + bx_i = 2.0 + 0.9x_i$$

↓

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

↓

$$S = \sum_{i=1}^{n} {e_i}^2 = 8.20$$

S を最小にする
a、b を求める
（ソルバー）

表示4.3.6　ソルバーの解

	L	M	N	O	P	Q	R	S
3	i	x	y	y-hat	e			
4	1	1	5	2.90	2.10		a	2.000
5	2	3	5	4.70	0.30		b	0.900
6	3	4	7	5.60	1.40		S	8.20
7	4	5	6	6.50	-0.50			
8	5	7	9	8.30	0.70			
9	6	10	10	11.00	-1.00			
10	平均	5.00	**7.00**	**6.50**	**0.50**			
11	平方和	50.00	**22.00**	**40.50**	**6.70**			
12								
13	相関係数	x	y	y-hat	e			
14	x	1.000	0.935	1.000	**-0.76**			
15	y	0.935	1.000	0.935	-0.49			
16	y-hat	1.00	0.93	1.00	**-0.76**			
17	e	-0.76	-0.49	-0.76	1.00			

変数セル

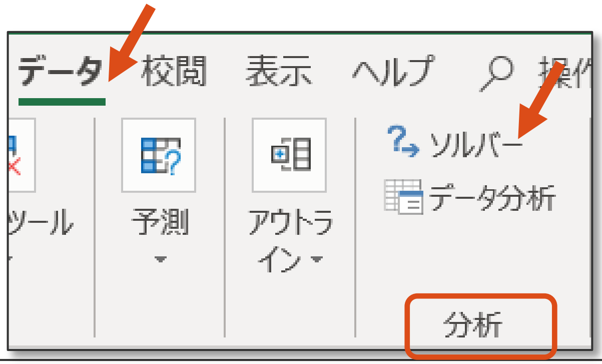
目的のセル

Excel のソルバーによる回帰式の推定（最小２乗法）

p.231

● ソルバーの操作

［データ］＞［分析］グループ＞［ソルバー］

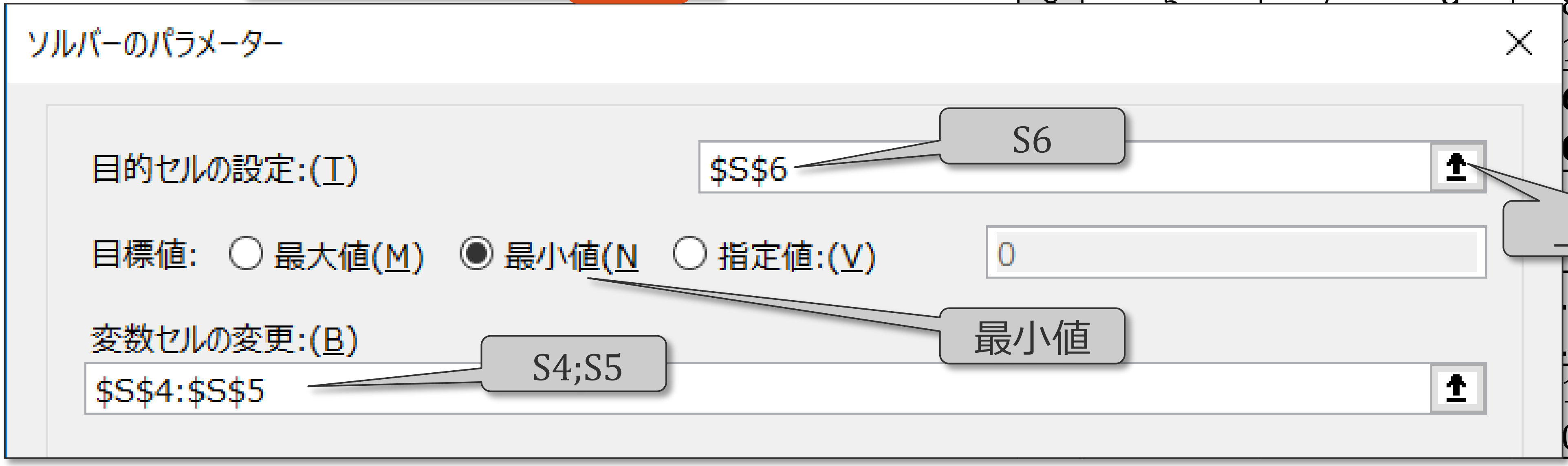


表示4.3.6　ソルバーの解

	L	M	N	O	P	Q	R	S
3	i	x	y	y-hat	e			
4	1	1	5	2.90	2.10		a	2.000
5	2	3	5	4.70	0.30		b	0.900
6	3	4	7	5.60	1.40		S	8.20
7	4	5	6	6.50	-0.50			
8	5	7	9	8.30	0.70			
				1.00	-1.00			
				6.50	**0.50**			
				0.50	**6.70**			
				.000	**-0.76**			
				.935	-0.49			
				1.00	**-0.76**			
				0.76	1.00			

変数セル

目的のセル



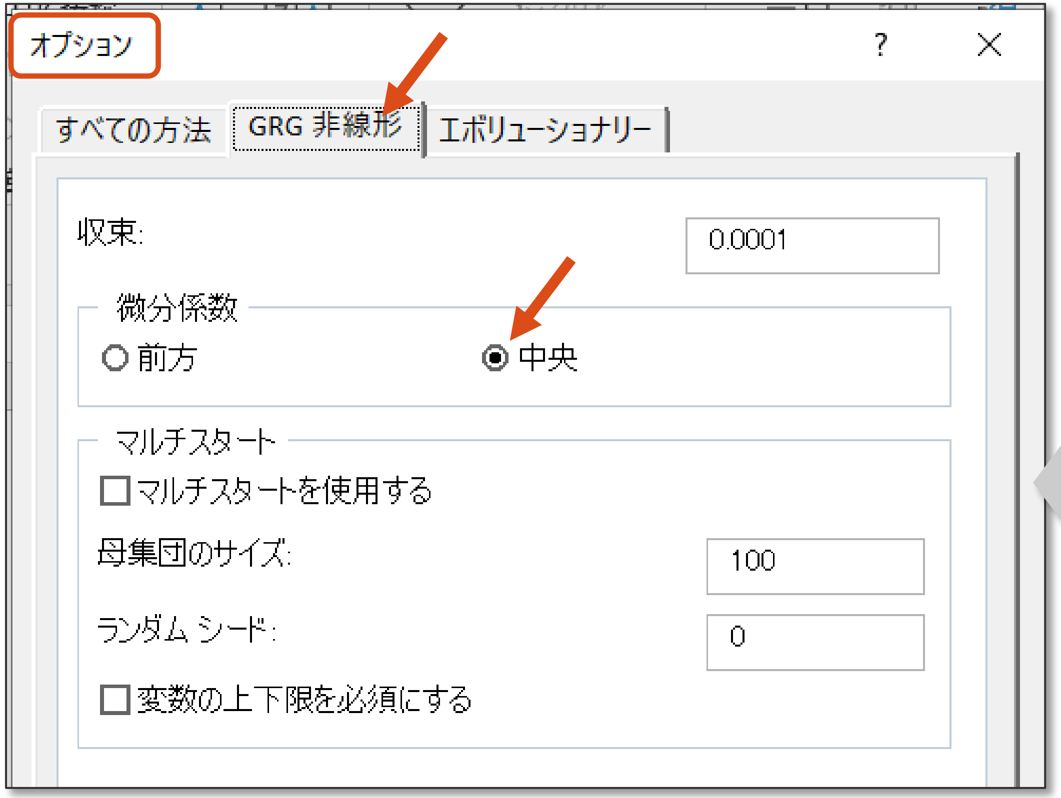
Excel のソルバーによる回帰式の推定（最小２乗法）

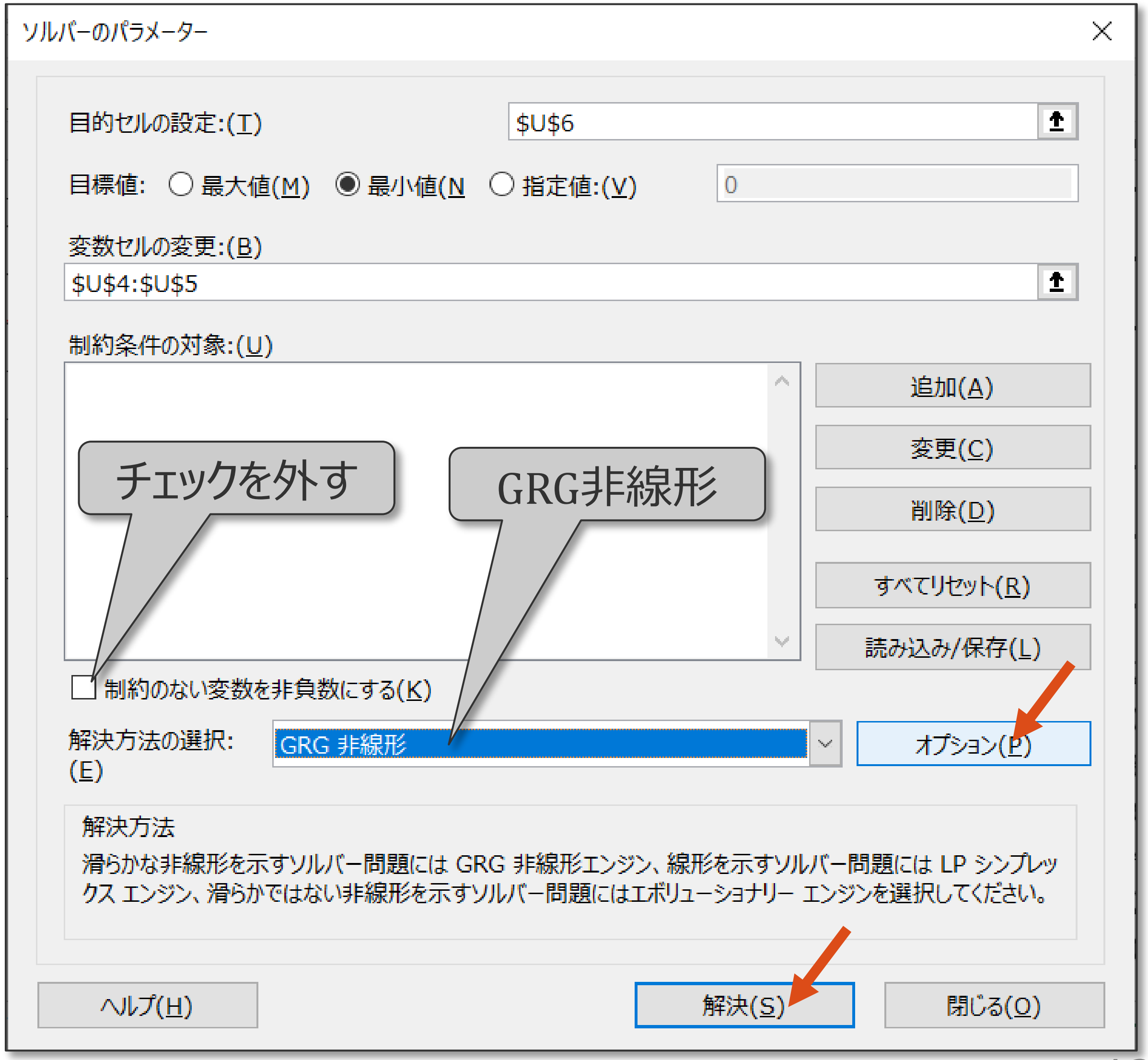
p.231

●ソルバーの操作

［データ］＞［分析］グループ＞［ソルバー］

オプション



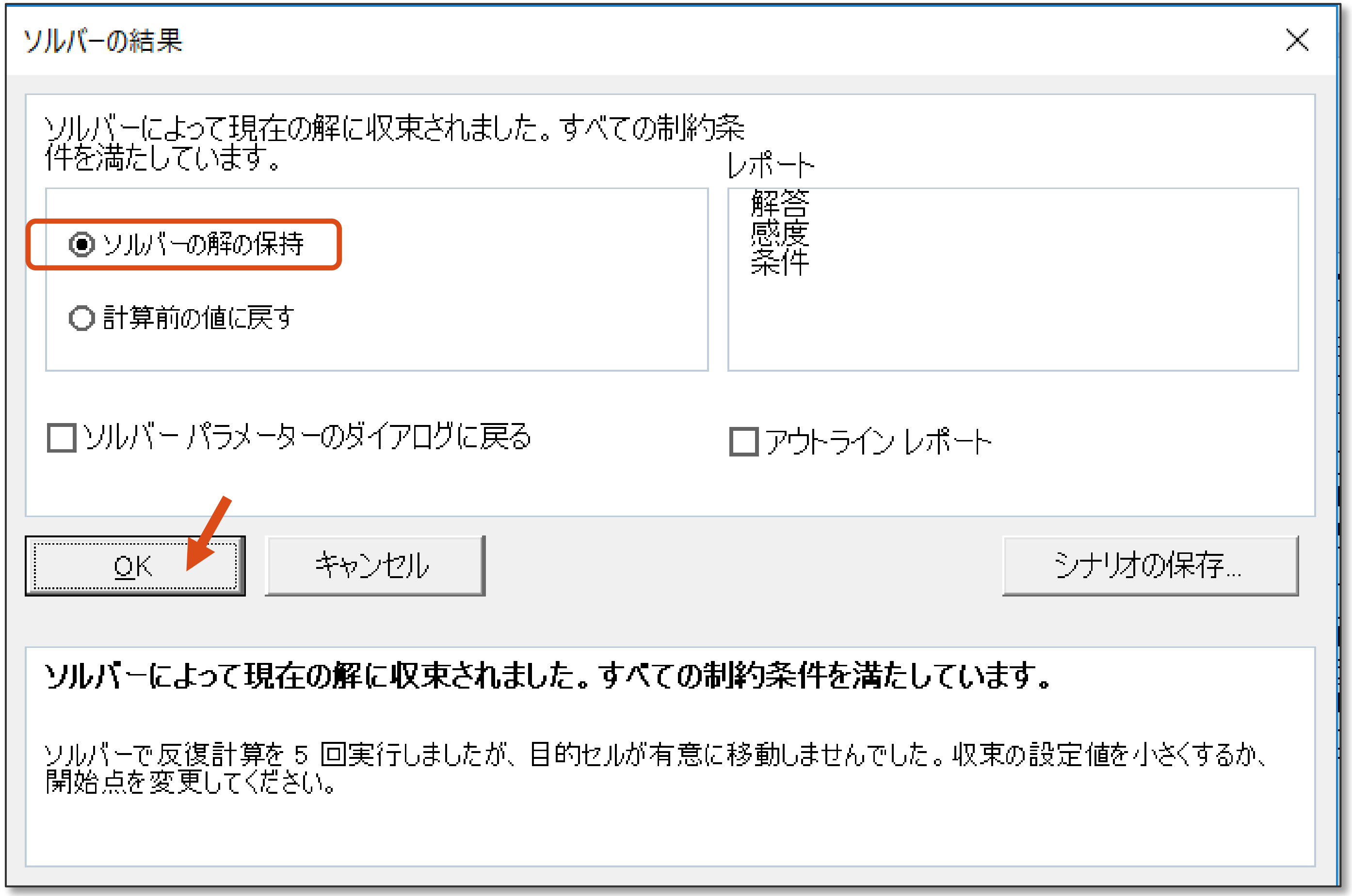


Excel のソルバーによる回帰式の推定（最小２乗法）

p.231

- ソルバーの操作

  ［データ］＞［分析］グループ＞［ソルバー］



Excel のソルバーによる回帰式の推定（最小２乗法）

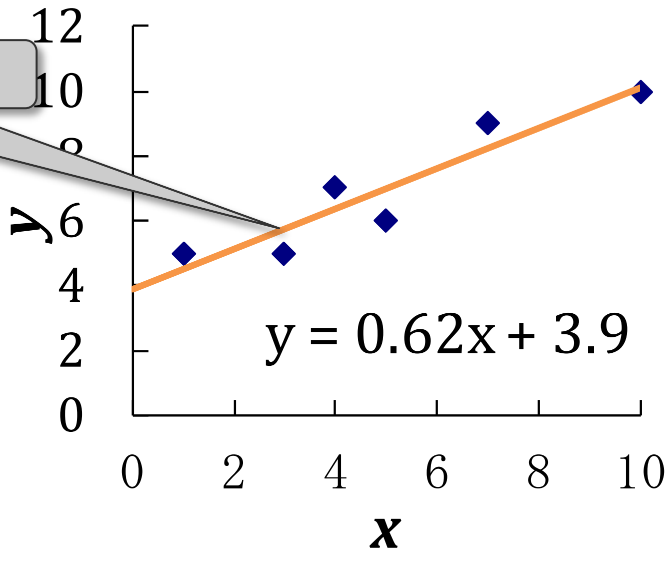
p.231

●ソルバーの解

残差平方和 S が最小になる a と b が求められ、
目的セルと変更セルに上書きされる
これを反映して $\hat{y}_i$ が再計算される

表示4.3.3 データと散布図

最もフィットした回帰直線



表示4.3.6 ソルバーの解

	L	M	N	O	P	R	S
3	i	x	y	y-hat	e		
4	1	1	5	4.52	0.48	a	3.900
5	2	3	5	5.76	-0.76	b	0.620
6	3	4	7	6.38	0.62	S	2.78
7	4	5	6	7.00	-1.00		
8	5	7	9	8.24	0.76		
9	6	10	10	10.10	-0.10		
10	平均	5.00	**7.00**	**7.00**	**0.00**		
11	平方和	50.00	**22.00**	**19.22**	**2.78**		
12							
13	相関係数	x	y	y-hat	e		
14	x	1.000	0.935	1.000	**0.00**		
15	y	0.935	1.000	0.935	0.36		
16	y-hat	1.00	0.93	1.00	**0.00**		
17	e	0.00	0.36	0.00	1.00		

変更セル
目的セル
の上書き

新しい
パラメータ
で再計算

Excel のソルバーによる回帰式の推定（最小２乗法）

p.231

●ソルバーの解

実務において、単回帰分析でソルバーを使うことはない（最小２乗法の仕組みをイメージ化）
テキストでは、以下の分野でソルバーの機能を頻繁に使用
本節で使い方を十分に理解する

単回帰分析（本節）
- 回帰直線のあてはめ

非線形回帰分析（第３部）
- 指数曲線のあてはめ
- ロジスティック曲線、ゴンペルツ曲線、プロビット曲線のあてはめ
- Emax モデル、Michaelis-Menten モデルのあてはめ
- 薬物動態モデルのあてはめ

ロジスティック回帰分析（第３部）

Excel のソルバーによる回帰式の推定（最小２乗法）

p.231

●残差の制約条件

(1)残差の平均が 0（残差の合計が 0）

観測値と予測値の平均は一致、合計も一致

$$\bar{y}_{\cdot} = \bar{\hat{y}}_{\cdot}, \qquad \sum y_i = \sum \hat{y}_i \qquad (4.3.10)$$

したがって、残差の平均は 0、残差の合計も 0

$$\bar{e}_{\cdot} = \bar{y}_{\cdot} - \bar{\hat{y}}_{\cdot} = 0$$

$$\sum e_i = \sum y_i - \sum \hat{y}_i = 0 \qquad (4.3.11)$$

(2) 残差 e は説明変数 x と無相関

残差 e は予測値 $\hat{y}$ と無相関

$$r(e, x) = 0 \qquad r(e, \hat{y}) = 0 \qquad (4.3.12)$$

ソルバーでパラメータの解を求めた状態

表示4.3.6　ソルバーの解

	L	M	N	O	P	Q	R	S
3	i	x	y	y-hat	e			
4	1	1	5	4.52	0.48		a	3.900
5	2	3	5	5.76	-0.76		b	0.620
6	3	4	7	6.38	0.62		S	2.78
7	4	5	6	7.00	-1.00			
8	5	7	9	8.24	0.76			
9	6	10	10	10.10	-0.10			
10	平均	5.00	**7.00**	**7.00**	**0.00**			
11	平方和	50.00	**22.00**	**19.22**	**2.78**			
12								
13	相関係数	x	y	y-hat	e			
14	x	1.000	0.935	1.000	**0.00**			
15	y	0.935	1.000	0.935	0.36			
16	y-hat	1.00	0.93	1.00	**0.00**			
17	e	0.00	0.36	0.00	1.00			

$\bar{e}_{\cdot} = 0$

無相関

無相関

Excel のソルバーによる回帰式の推定（最小２乗法）

p.232

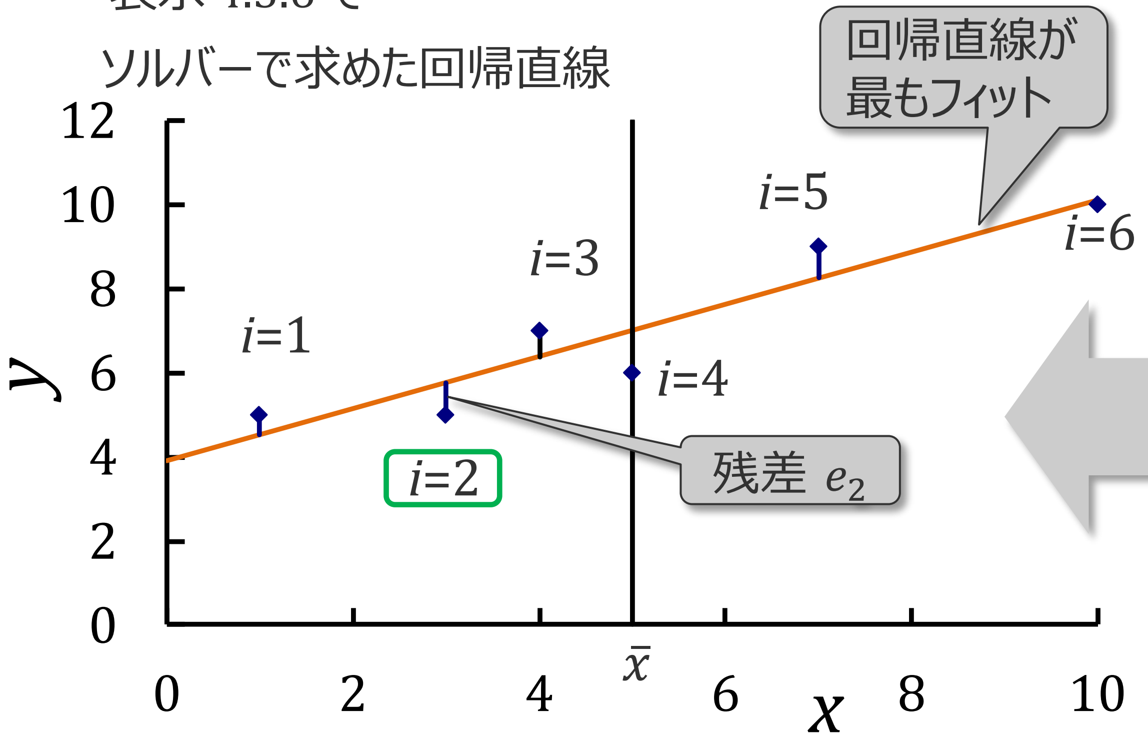
●残差の制約条件

残差の平均が 0 で、残差 e と説明変数 x とは無相関になることを、視覚的・直感的に確認

（演習4.3.2）

表示 4.3.6 で

ソルバーで求めた回帰直線



表示4.3.6　ソルバーの解

ソルバーでパラメータを求めた状態

	L	M	N	O	P
3	i	x	y	y-hat	e
4	1	1	5	4.52	0.48
5	2	3	5	5.76	-0.76
6	3	4	7	6.38	0.62
7	4	5	6	7.00	-1.00
8	5	7	9	8.24	0.76
9	6	10	10	10.10	-0.10
10	平均	5.00	**7.00**	**7.00**	**0.00**
11	平方和	50.00	**22.00**	**19.22**	**2.78**
12					
13	相関係数	x	y	y-hat	e
14	x	1.000	0.935	1.000	**0.00**
15	y	0.935	1.000	0.935	0.36
16	y-hat	1.00	0.93	1.00	**0.00**
17	e	0.00	0.36	0.00	1.00

R	S
a	3.900
b	0.620
S	2.78

$\bar{e}_. = 0$

無相関

無相関

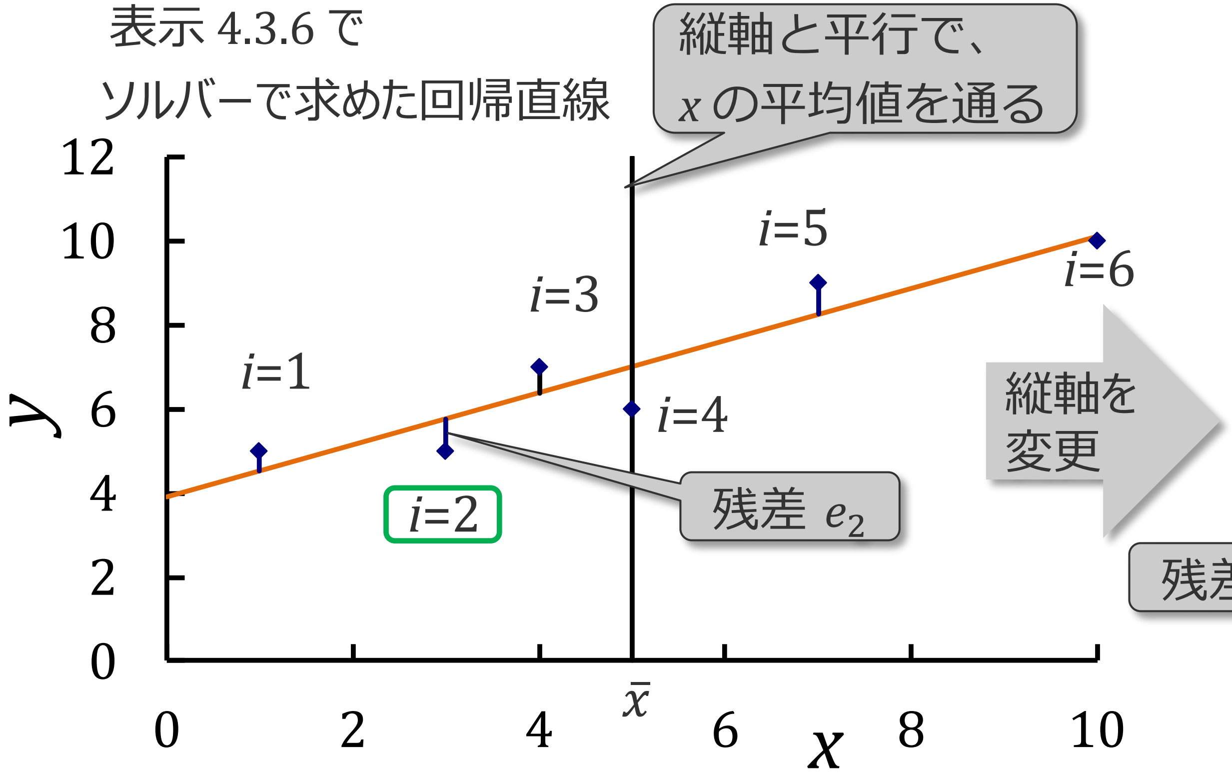
Excel のソルバーによる回帰式の推定（最小２乗法）

p.232

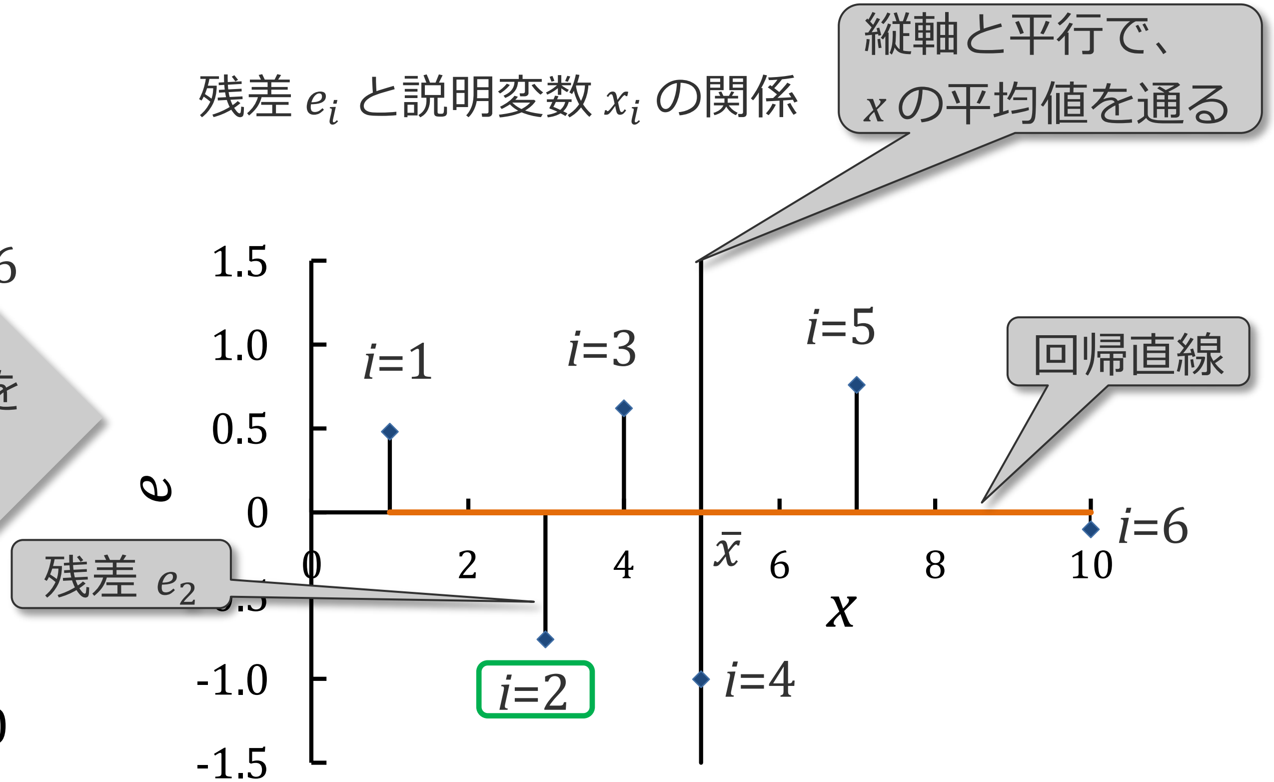
●残差の制約条件

残差 e_i の合計は０、したがって、平均も０ であることを視覚的に確認（演習4.3.2　p.270）

表示 4.3.6 で
ソルバーで求めた回帰直線



残差 e_i と説明変数 x_i の関係



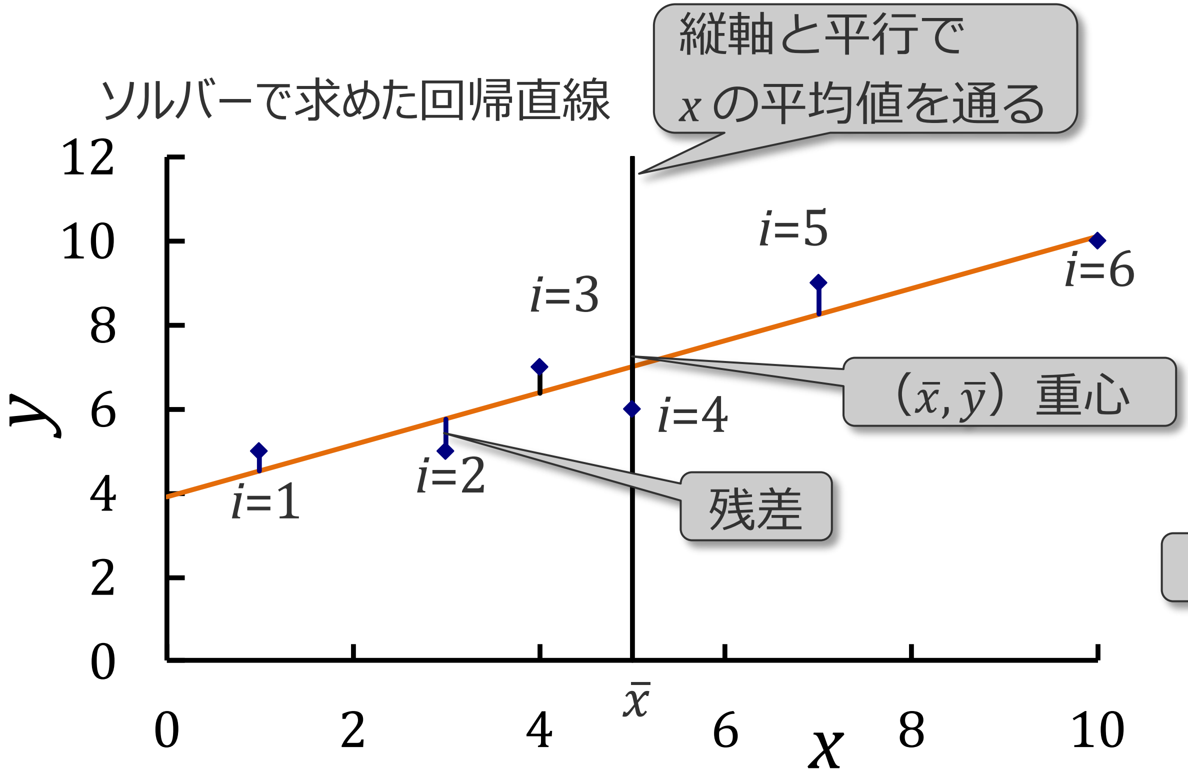
Excel のソルバーによる回帰式の推定（最小２乗法）

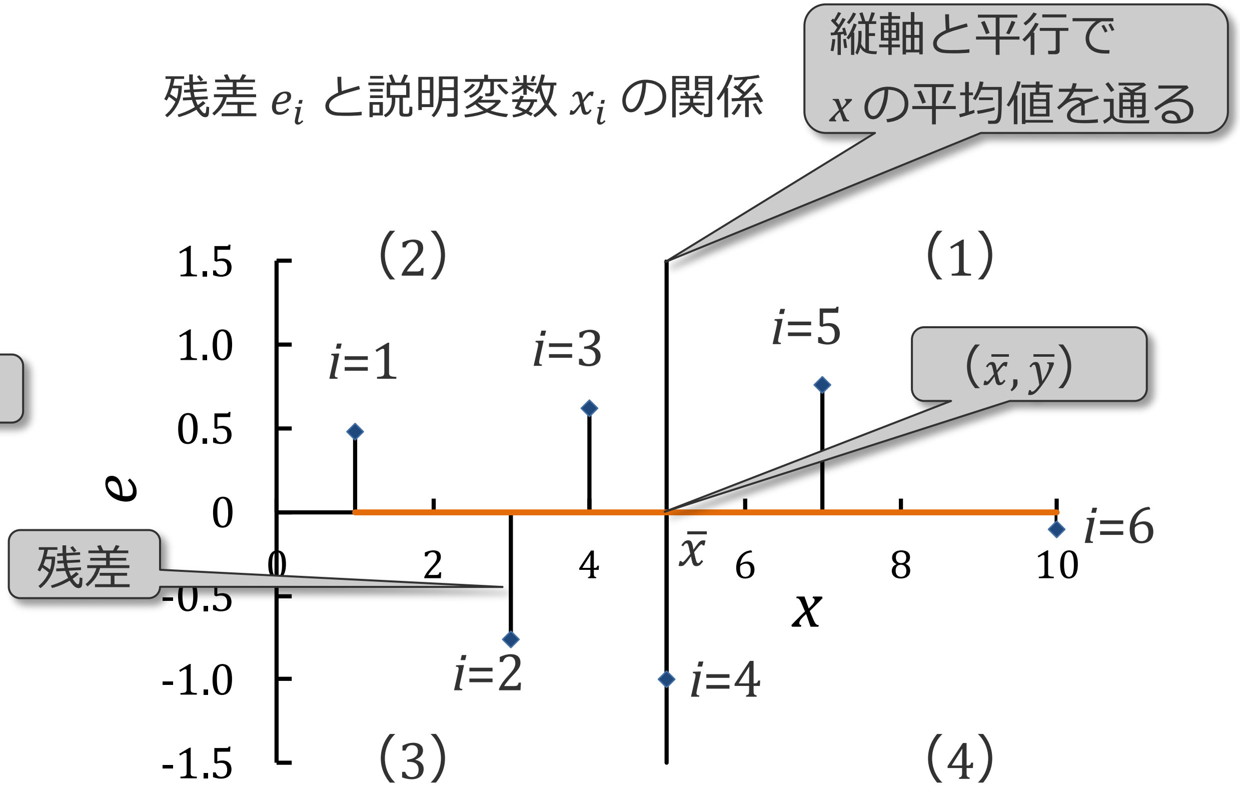
p.232

●残差の制約条件

回帰直線は必ず重心 $(\bar{x}, \bar{y})$ を通る　$\bar{y} = a + b\bar{x}$

重心 $(\bar{x}, \bar{y})$ を通る直線であれば、いずれの直線も残差の合計は 0





Excel のソルバーによる回帰式の推定（最小２乗法）

p.232

●残差の制約条件

表示4.3.6

13	相関係数	x	y	y-hat	e
14	x	1.000	0.935	1.000	**-0.85**
15	y	0.935	1.000	0.935	-0.61
16	y-hat	1.00	0.93	1.00	**-0.85**
17	e	-0.85	-0.61	-0.85	1.00

負の相関

負の相関

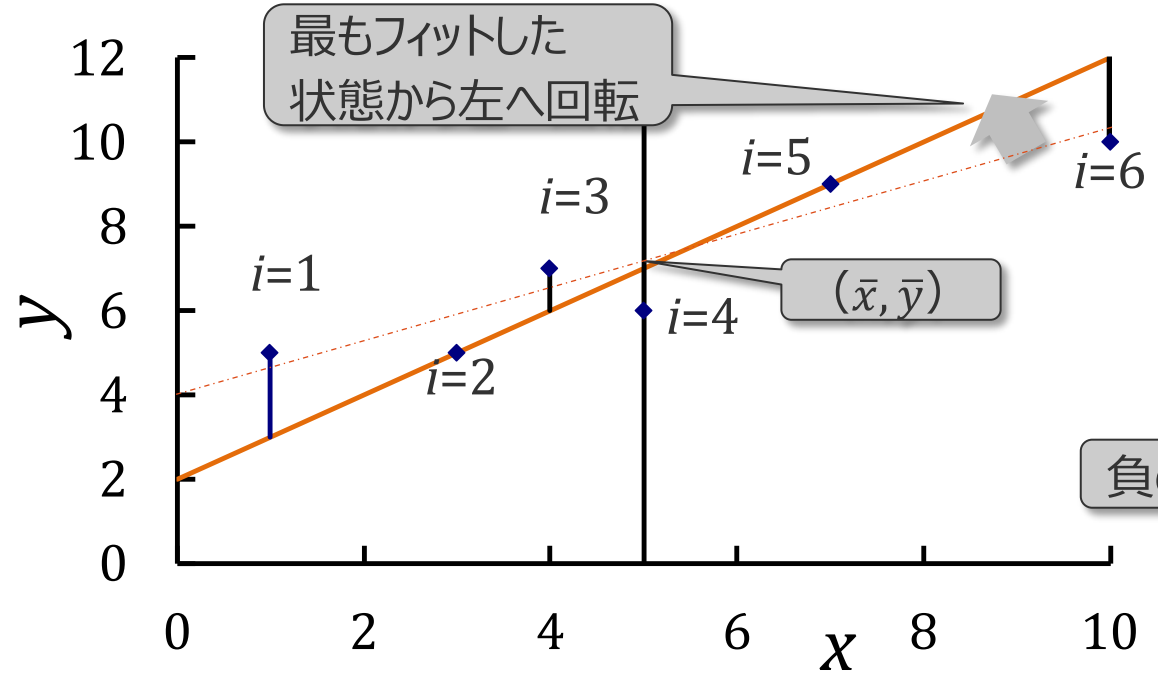
	L	M	N	O	P
3	i	x	y	y-hat	e
4	1	1	5	3.00	2.00
5	2	3	5	5.00	0.00
6	3	4	7	6.00	1.00
7	4	5	6	7.00	-1.00
8	5	7	9	9.00	0.00
9	6	10	10	12.00	-2.00
10	平均	5.00	**7.00**	**7.00**	**0.00**
11	平方和	50.00	**22.00**	**50.00**	**10.00**

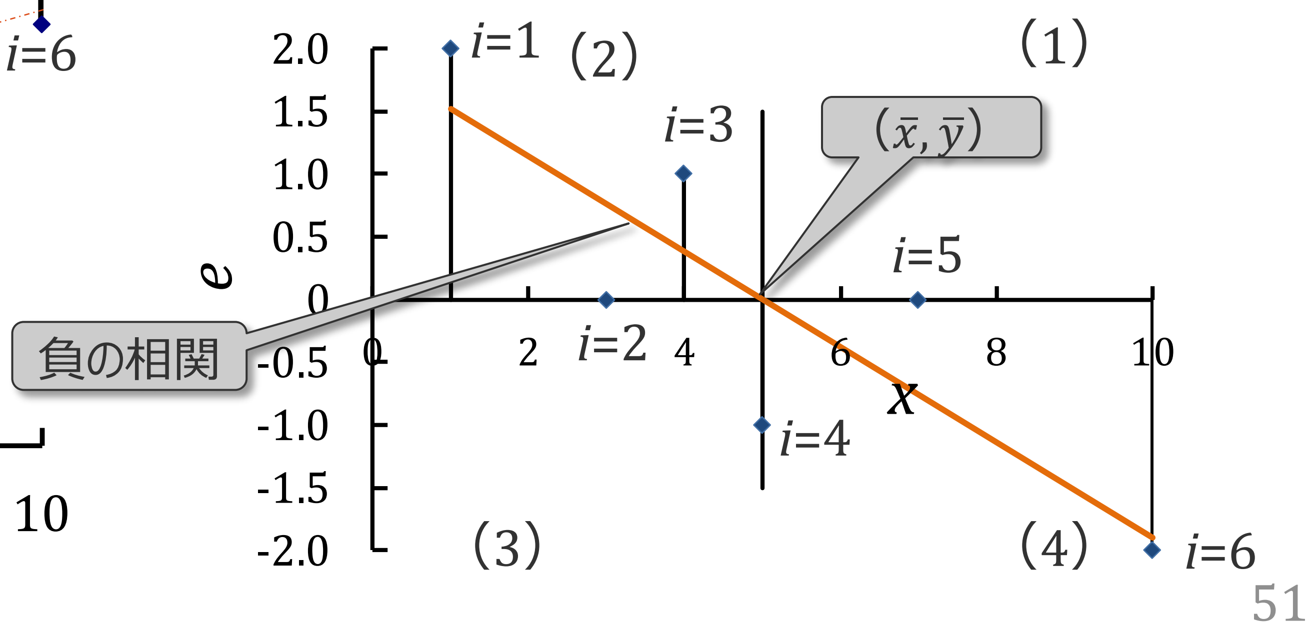
Q R S

a	2.000
b	1.000
S	10.00

a	3.900
b	0.620
S	2.78

残差の平均（合計）は 0





Excel のソルバーによる回帰式の推定（最小２乗法）

p.232

●残差の制約条件

表示4.3.6

13	相関係数	x	y	y-hat	e
14	x	1.000	0.935	1.000	**0.68**
15	y	0.935	1.000	0.935	0.90
16	y-hat	1.00	0.93	1.00	**0.68**
17	e	0.68	0.90	0.68	1.00

正の相関

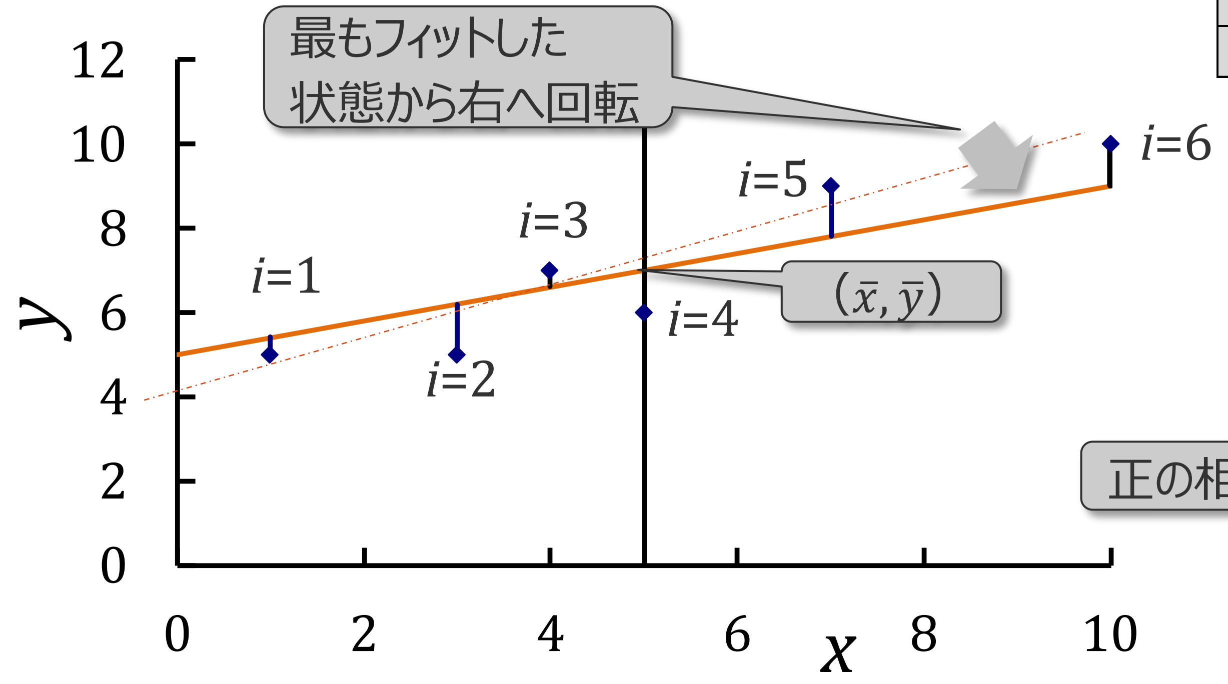
正の相関

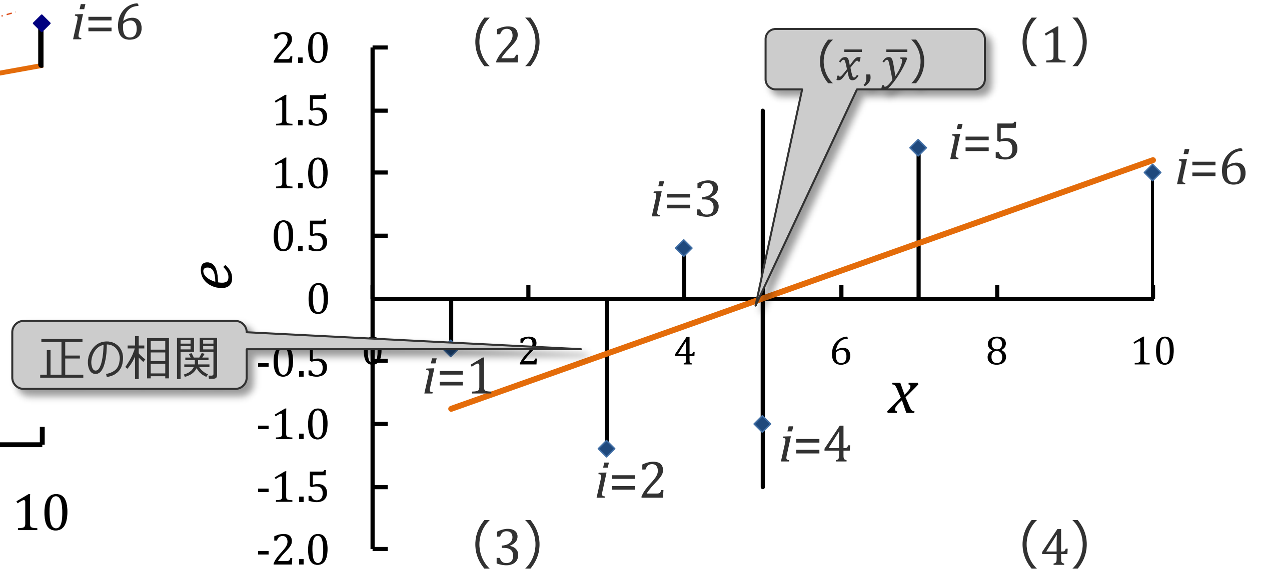
	L	M	N	O	P
3	i	x	y	y-hat	e
4	1	1	5	5.40	-0.40
5	2	3	5	6.20	-1.20
6	3	4	7	6.60	0.40
7	4	5	6	7.00	-1.00
8	5	7	9	7.80	1.20
9	6	10	10	9.00	1.00
10	平均	5.00	**7.00**	**7.00**	**0.00**
11	平方和	50.00	**22.00**	**8.00**	**5.20**

a	5.000
b	0.400
S	5.20

a	3.900
b	0.620
S	2.78

残差の平均（合計）は 0





Excel のソルバーによる回帰式の推定（最小２乗法）

p.232

●残差の制約条件

表示4.3.6

13	相関係数	x	y	y-hat	e
14	x	1.000	0.935	1.000	**0.00**
15	y	0.935	1.000	0.935	0.36
16	y-hat	1.00	0.93	1.00	**0.00**
17	e	0.00	0.36	0.00	1.00

制約条件２

無相関

無相関

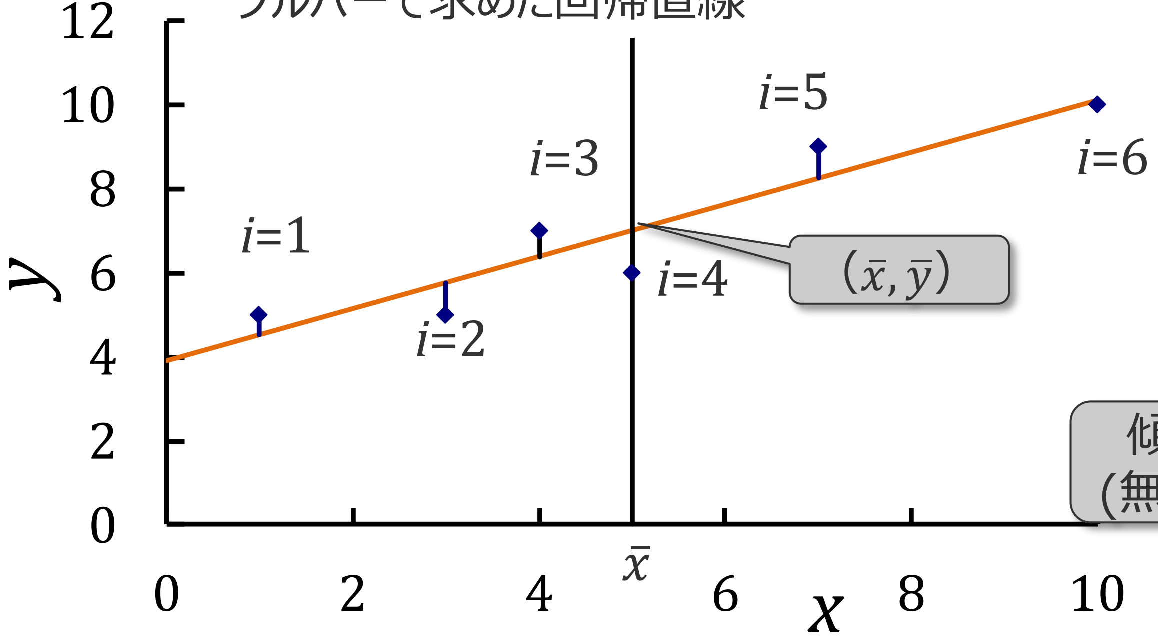
	L	M	N	O	P	Q	R	S
3	i	x	y	y-hat	e			
4	1	1	5	4.52	0.48		a	3.900
5	2	3	5	5.76	-0.76		b	0.620
6	3	4	7	6.38	0.62		S	2.78
7	4	5	6	7.00	-1.00			
8	5	7	9	8.24	0.76			
9	6	10	10	10.10	-0.10			
10	平均	5.00	**7.00**	**7.00**	**0.00**			
11	平方和	50.00	**22.00**	**19.22**	**2.78**			

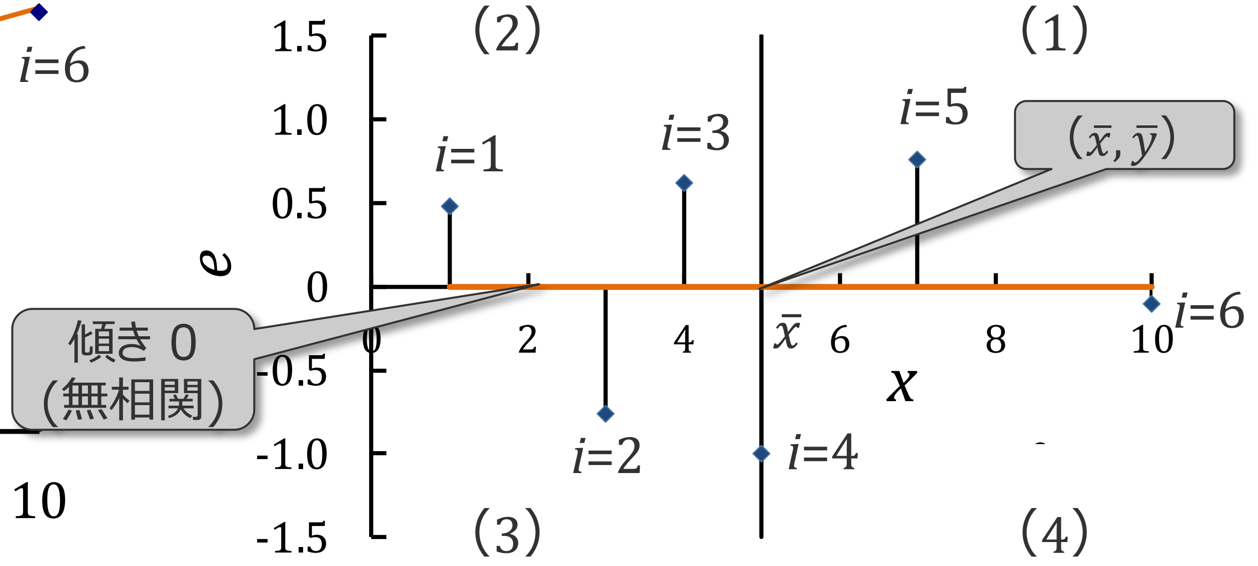
最小値

制約条件１

残差の平均
（合計）は０

ソルバーで求めた回帰直線





Excel のソルバーによる回帰式の推定（最小２乗法）

p.232

●残差の制約条件

最小２乗法で回帰直線をフィットさせるための
残差の「2 つの制約」

(1) 残差の平均は 0（残差の合計は 0）

(2) 残差 e と説明変数 x は無相関

（残差 e と予測値 $\hat{y}$ は無相関）

$\hat{y}_i = a + bx_i$

→ 数学的証明（§4.6(1)(2) p.255）

平方和の分解

次の項で、平方和の分解を取り上げる

表示4.3.6 ソルバーの解

	L	M	N	O	P	Q	R	S
3	i	x	y	y-hat	e			
4	1	1	5	4.52	0.48		a	3.900
5	2	3	5	5.76	-0.76		b	0.620
6	3	4	7	6.38	0.62		S	2.78
7	4	5	6	7.00	-1.00			
8	5	7	9	8.24	0.76			
9	6	10	10	10.10	-0.10			
10	平均	5.00	**7.00**	**7.00**	**0.00**			
11	平方和	50.00	**22.00**	**19.22**	**2.78**			
12								
13	相関係数	x	y	y-hat	e			
14	x	1.000	0.935	1.000	**0.00**			
15	y	0.935	1.000	0.935	0.36			
16	y-hat	1.00	0.93	1.00	**0.00**			
17	e	0.00	0.36	0.00	1.00			

残差の平均 0

無相関

無相関

p.232

（5）平方和の分解と σ^2 の推定

最小２乗法による回帰式の推定

（3） 試行錯誤による回帰式の推定（最小２乗法）
（4）ソルバーによる回帰式の推定（最小２乗法）
（5）平方和の分解と σ^2 の推定（最小２乗法）
（6）数理統計学的な回帰式の推定（最小２乗法）
（7）LINEST 関数による回帰式の推定（最小２乗法）

テキストの項のタイトルとやや異なる

平方和の分解

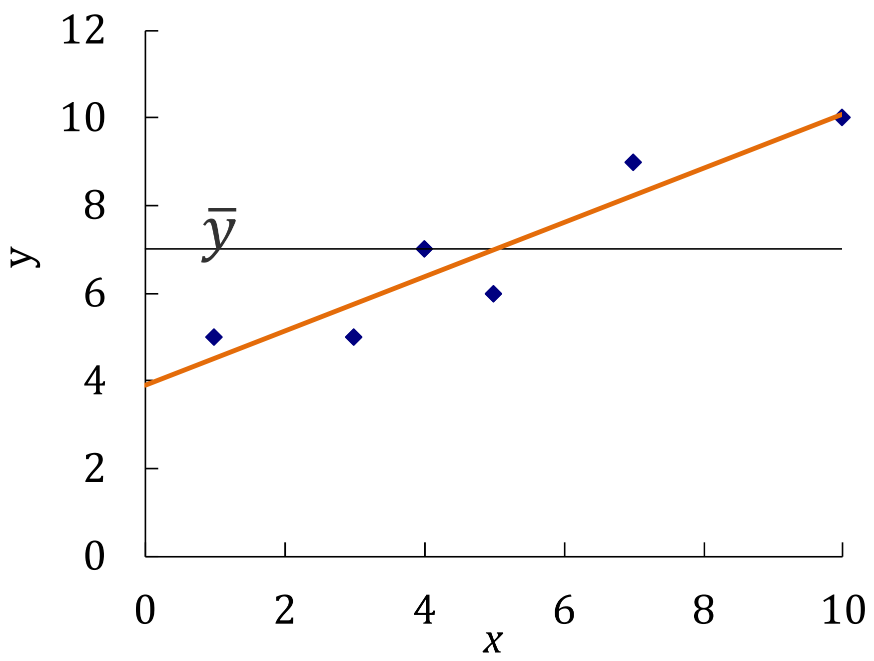
p.232

● 3 つの平方和

総平方和 = 回帰平方和 + 残差平方和

$$S_T = S_R + S_e \qquad (4.3.13)$$

$$22.00 = 19.22 + 2.78$$



表示4.3.6　ソルバーの解（一部）

	L	M	N	O	P
3	i	x	y	y-hat	e
4	1	1	5	4.52	0.48
5	2	3	5	5.76	-0.76
6	3	4	7	6.38	0.62
7	4	5	6	7.00	-1.00
8	5	7	9	8.24	0.76
9	6	10	10	10.10	-0.10
10	平均	5.00	**7.00**	**7.00**	**0.00**
11	平方和	50.00	**22.00**	**19.22**	**2.78**

総平方和 Total

回帰平方和 Regression

残差平方和 Error

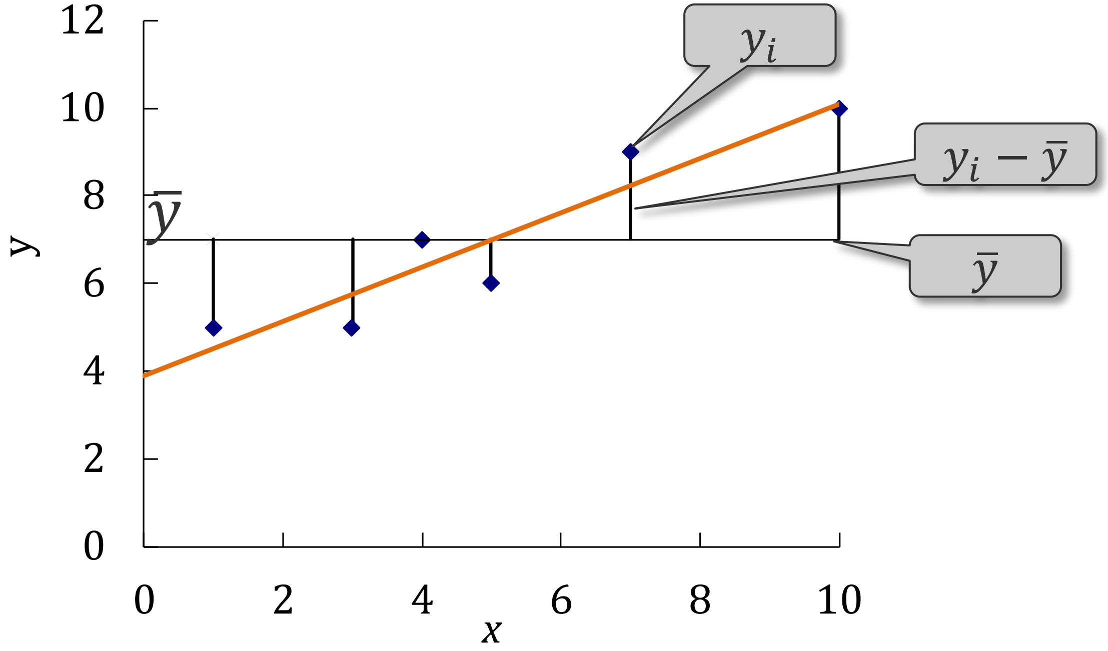
平方和の分解

●総平方和とその自由度

総平方和 = 回帰平方和 + 残差平方和

S_T = S_R + S_e

22.00 = 19.22 + 2.78



表示4.3.6　ソルバーの解（一部）

	L	M	N	O	P
3	i	x	y	y-hat	e
4	1	1	5	4.52	0.48
5	2	3	5	5.76	-0.76
6	3	4	7	6.38	0.62
7	4	5	6	7.00	-1.00
8	5	7	9	8.24	0.76
9	6	10	10	10.10	-0.10
10	平均	5.00	**7.00**	**7.00**	**0.00**
11	平方和	50.00	**22.00**	**19.22**	**2.78**

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

自由度：$\nu_T = n - 1 = 6 - 1 = 5$　（§2.1）

平方和の分解

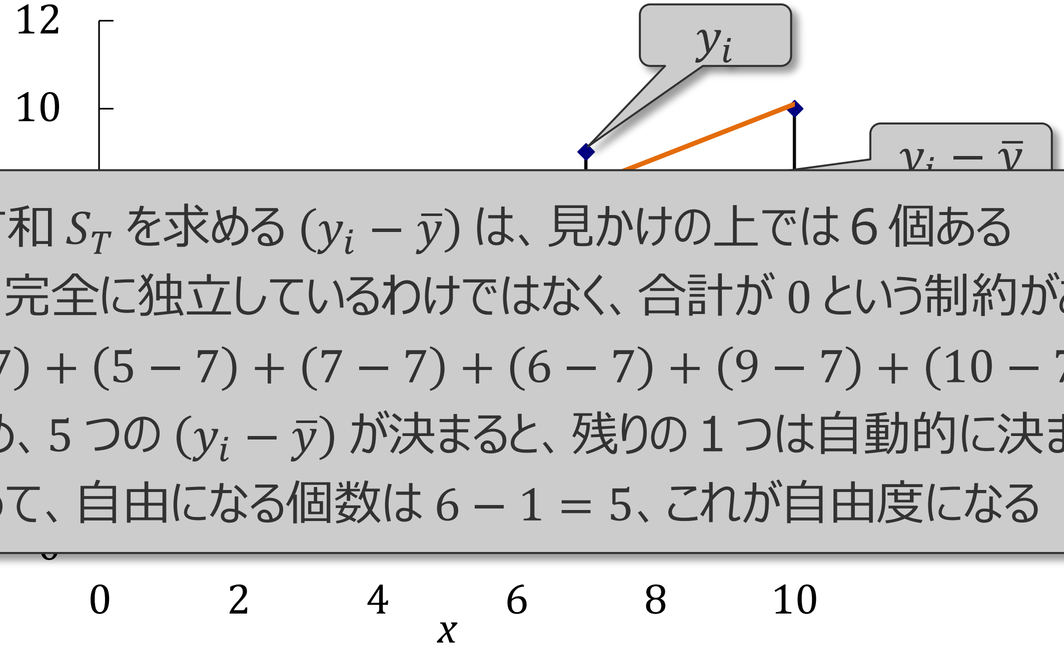
p.232

●総平方和とその自由度

総平方和 ＝ 回帰平方和 ＋ 残差平方和

$S_T = S_R + S_e$

$22.00 = 19.22 + 2.78$



総平方和 S_T を求める $(y_i - \bar{y})$ は、見かけの上では 6 個ある
しかし、完全に独立しているわけではなく、合計が 0 という制約がある

$(5-7)+(5-7)+(7-7)+(6-7)+(9-7)+(10-7)=0$

そのため、5 つの $(y_i - \bar{y})$ が決まると、残りの 1 つは自動的に決まる
したがって、自由になる個数は $6-1=5$、これが自由度になる

表示4.3.6　ソルバーの解（一部）

	L	M	N	O	P
3	i	x	y	y-hat	e
4	1	1	5	4.52	0.48
5	2	3	5	5.76	-0.76
6	3	4	7	6.38	0.62
7	4	5	6	7.00	-1.00
8	5	7	9	8.24	0.76
9	6	10	10	10.10	-0.10
10	平均	5.00	**7.00**	**7.00**	**0.00**
11	平方和	50.00	**22.00**	**19.22**	**2.78**

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

自由度：$\nu_T = n - 1 = 6 - 1 = 5$　（§2.1）

平方和の分解

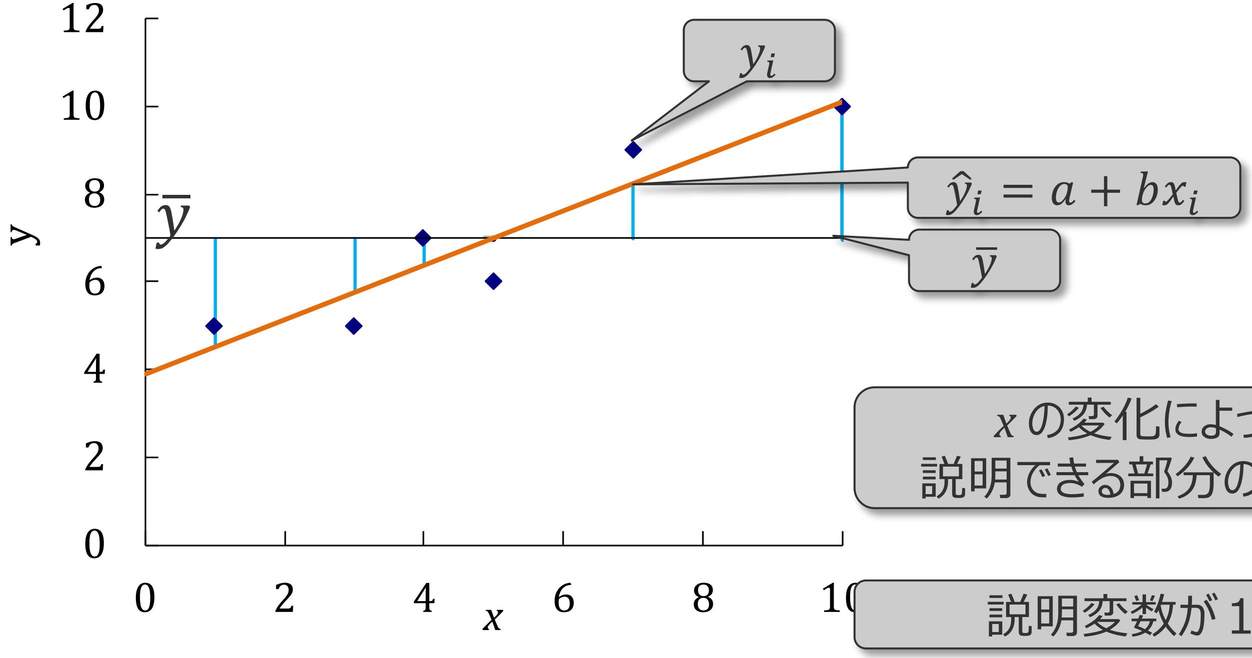
p.232

●回帰平方和とその自由度

総平方和	=	回帰平方和	+	残差平方和
S_t	=	S_R	+	S_e
22.00	=	19.22	+	2.78

表示4.3.6　ソルバーの解（一部）

	L	M	N	O	P
3	i	x	y	y-hat	e
4	1	1	5	4.52	0.48
5	2	3	5	5.76	-0.76
6	3	4	7	6.38	0.62
7	4	5	6	7.00	-1.00
8	5	7	9	8.24	0.76
9	6	10	10	10.10	-0.10
10	平均	5.00	**7.00**	**7.00**	**0.00**
11	平方和	50.00	**22.00**	**19.22**	**2.78**



x の変化によって説明できる部分の大きさ

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

説明変数が 1 個

自由度：$\nu_R = 1$

平方和の分解

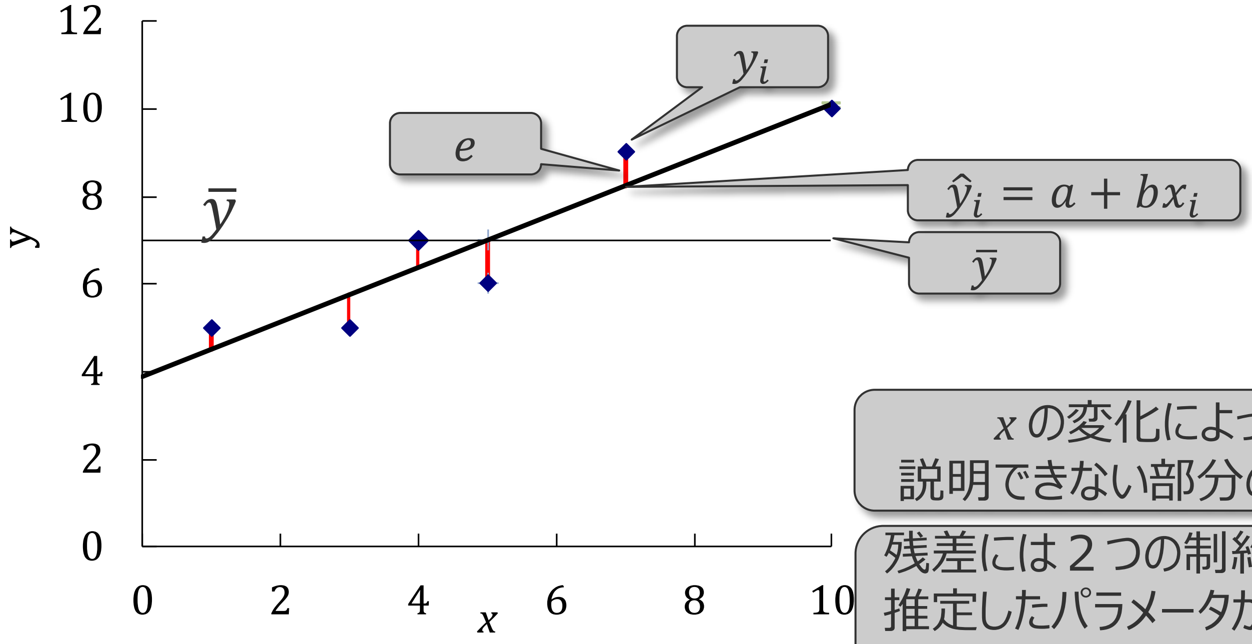
p.232

●残差平方和とその自由度

総平方和 = 回帰平方和 + 残差平方和

$S_t \quad = \quad S_R \quad + \quad S_e$

$22.00 \quad = \quad 19.22 \quad + \quad 2.78$



表示4.3.6　ソルバーの解（一部）

	L	M	N	O	P
3	i	x	y	y-hat	e
4	1	1	5	4.52	0.48
5	2	3	5	5.76	-0.76
6	3	4	7	6.38	0.62
7	4	5	6	7.00	-1.00
8	5	7	9	8.24	0.76
9	6	10	10	10.10	-0.10
10	平均	5.00	**7.00**	**7.00**	**0.00**
11	平方和	50.00	**22.00**	**19.22**	**2.78**

x の変化によって説明できない部分の大きさ

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

残差には２つの制約
推定したパラメータが２つ（a, b）

自由度：$\nu_e = n - 2 = 6 - 2 = 4$

平方和の分解

p.232

●平和和の分解、寄与率（決定係数 R^2）

総平方和 ＝ 回帰平方和 ＋ 残差平方和

$$S_T = S_R + S_e$$

$$22.00 = 19.22 + 2.78$$

$S_R = 19.22$	$S_e = 2.78$

$S_T = 22.00$

S_R は、総平方和 S_T の内、x の変化で説明できる（回帰で説明できる）部分の大きさを表す

S_eは、x の変化で説明できない（回帰で説明できない）部分の大きさを表す

$$R^2 = \frac{S_R}{S_T} = \frac{19.22}{22.00} = 0.874 \qquad (4.3.14)$$

寄与率（決定係数）の定義

$$\sqrt{R^2} = \sqrt{0.874} = 0.935 = r \qquad (4.3.15)$$

相関係数

寄与率は回帰式がもつ説明力の指標：回帰平方和が大きいほど 1 に近づく（§4.4 参照）

寄与率（決定係数）の平方根は相関係数に一致（寄与率を R^2 と表記する理由）

平方和の分解

p.233

●平方和の自由度とその分解

平方和の自由度　ν

ギリシャ文字
ニュー

総平方和の自由度：n 個のデータ y_i の総平均との平方和なので、$n-1$ とする（§2.1）

$$\nu_T = n - 1 = 6 - 1 = 5$$

回帰平方和の自由度：用いた説明変数は１個（x）

$$\nu_R = 1$$

残差平方和の自由度：制約が２（e_i の平均は０、e_i と x_i とは無相関）

$$\nu_e = n - 2 = 6 - 2 = 4$$

この２は推定したパラメータの個数が２個（a, b）に対応

自由度 ν の分解

$$\nu_T = \nu_R + \nu_e \qquad (4.3.16)$$

平方和と自由度は分散分析で利用（次節、p.242）

平方和の分解

p.233

●平方和の自由度とその分解

平方和の自由度　ν

総平方和の自由度：n 個のデータ y_i の総平均との平方和なので、

$$\nu_T = n - 1 = 6 - 1 = 5$$

回帰平方和の自由度：用いた説明変数は１個 （x）

$$\nu_R = 1$$

残差平方和の自由度：制約が２（e_i の平均は０、 e_i と x_i とは無

$$\nu_e = n - 2 = 6 - 2 = 4$$

この２は推定したパラメータの個数が２個（a, b）に対応

自由度 ν の分解

$$\nu_T = \nu_R + \nu_e \qquad (4.3.16)$$

平方和と自由度は分散分析で利用（次節、p.242）

1 本の直線は 2 点で定まるため、
直線の自由度は本来 2 である
回帰直線は重心（$\bar{x}, \bar{y}$）を通るので
1 点は既に定まっている、残りは 1
永田（1996）

回帰式のパラメータは a, b の 2 個
回帰直線は重心（$\bar{x}, \bar{y}$）を通るので
回帰係数 b が決まると切片 a は
自動的に決まるので自由度は 1

残差平方和の自由度は、
総平方和の自由度から
回帰平方和の自由度を引く
$\nu_e = \nu_T - \nu_R$

誤差の標準偏差 σ の推定

p.233

●残差平均平方、残差標準偏差

$$\eta = \alpha + \beta x \qquad (\eta \text{ は母平均}) \qquad (4.3.2)$$

$$y = \eta + \varepsilon = \alpha + \beta x + \varepsilon$$

$$\varepsilon \sim \mathrm{N}(0, \sigma^2)$$

回帰モデルで、

誤差 ε のばらつきの大きさは標準偏差 σ

誤差 ε は x の変化で説明できない部分

残差平方和 S_e は x の変化で説明できない部分

→　標準偏差 σ を残差平方和 S_e から推定

表示4.3.6　ソルバーの解（一部）

	L	M	N	O	P
3	i	x	y	y-hat	e
4	1	1	5	4.52	0.48
5	2	3	5	5.76	-0.76
6	3	4	7	6.38	0.62
7	4	5	6	7.00	-1.00
8	5	7	9	8.24	0.76
9	6	10	10	10.10	-0.10
10	平均	5.00	**7.00**	**7.00**	**0.00**
11	平方和	50.00	**22.00**	**19.22**	**2.78**

総平方和 Total

回帰平方和 Regression

残差平方和 Error

$$V_e = S_e / \nu_e = 2.78/4 = 0.695$$

データ１個あたりの残差平方和 S_e （V_e ：残差平均平方、残差分散、誤差分散）

$\sigma \sim s = \sqrt{V_e} = \sqrt{0.695} = 0.834$ ・・・誤差の標準偏差（JMP では RMSE と表記）　（p.246）

誤差の標準偏差 σ の推定

p.233

●残差平均平方、残差標準偏差

$$\eta = \alpha + \beta x \qquad (\eta \text{ は母平均}) \qquad (4.3.2)$$

$$y = \eta + \varepsilon = \alpha + \beta x + \varepsilon$$

$$\varepsilon \sim \mathrm{N}(0, \sigma^2)$$

回帰モデルで、

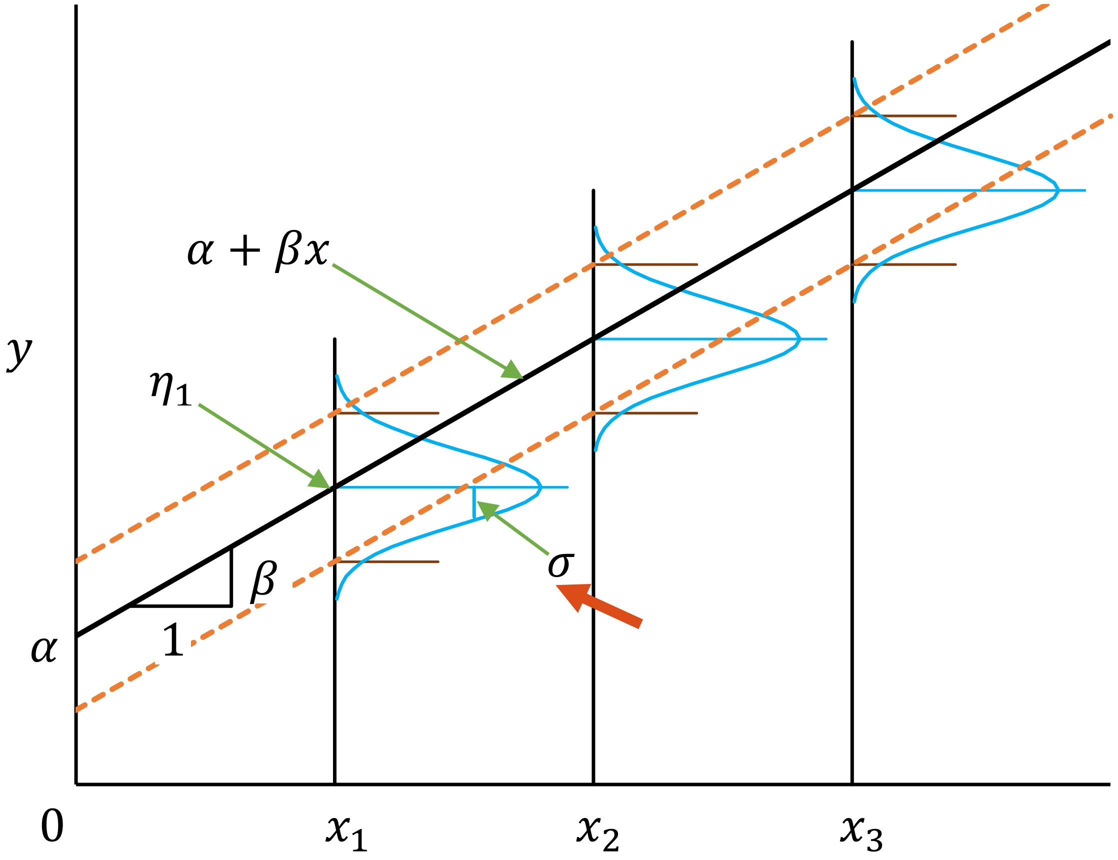
誤差 ε のばらつきの大きさは標準偏差 σ

誤差 ε は x の変化で説明できない部分

残差平方和 S_e は x の変化で説明できない部分

→　標準偏差 σ を残差平方和 S_e から推定

表示 4.3.2　回帰モデル



$$V_e = S_e/\nu_e = 2.78/4 = 0.695$$

データ１個あたりの残差平方和 S_e （V_e ：残差平均平方、残差分散、誤差分散）

$\sigma \sim s = \sqrt{V_e} = \sqrt{0.695} = 0.834$ ・・・誤差の標準偏差（JMP では RMSE と表記）（p.246）

p.233

（6） a, b, S_R などの計算式

最小２乗法による回帰式の推定

（3） 試行錯誤による回帰式の推定（最小２乗法）
（4）ソルバーによる回帰式の推定（最小２乗法）
（5）平方和の分解と σ^2 の推定（最小２乗法）
（6）数理統計学的な回帰式の推定（最小２乗法）
（7）LINEST 関数による回帰式の推定（最小２乗法）

テキストの項のタイトルとやや異なる

数理統計学的な回帰式の推定（最小２乗法）

p.233

●回帰係数と切片の推定式

最小２乗法に基づき、Excel のソルバーを利用して回帰式（切片 a、回帰係数 b）を推定

正規方程式による回帰式の解（§4.6 (1) p.255 参照）

$$b = \frac{S_{xy}}{S_{xx}} = \frac{31.0}{50.0} = 0.62 \qquad (4.3.17)$$

$$a = \bar{y} - b\bar{x} = 7.0 - 0.62 \times 5.0 = 3.90 \qquad (4.3.18)$$

平方和と積和から計算できる

S_{xy} ：x と y の積和

$$S_{xy} = \sum_{i=1}^{6} (x_i - \bar{x}.)(y_i - \bar{y}.)$$

S_{xx} ：x の平方和

$$S_{xx} = \sum_{i=1}^{6} (x_i - \bar{x}.)^2$$

この意味については次節（§4.4）で説明

表示4.3.3

i	x	y
1	1	5
2	3	5
3	4	7
4	5	6
5	7	9
6	10	10
平均	5.0	7.0
平方和	50.0	22.0
積和	31.0	

$S_{yy} = S_T$

S_{xx}

S_{xy}

数理統計学的な回帰式の推定（最小２乗法）

p.234

●別の回帰式の推定

$$y = a + bx \qquad (4.3.1)$$

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

回帰直線は重心$(\bar{x}, \bar{y})$を通る

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

$$y = a + bx = (\bar{y} - b\bar{x}) + bx = \bar{y} + b(x - \bar{x}) \qquad (4.3.19)$$

重心$(\bar{x}, \bar{y})$を通ることを示す回帰式

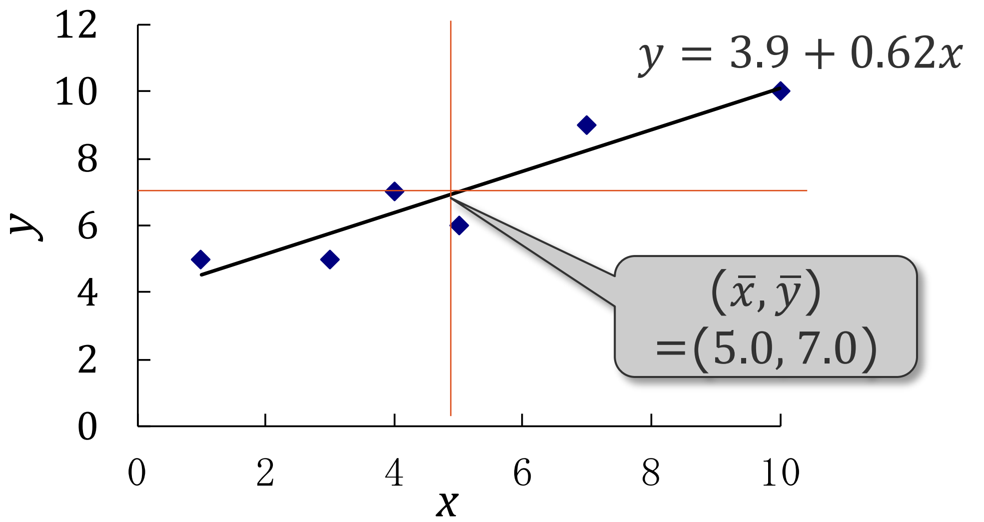
表示 4.3.6（一部）

	L	M	N	O	P	Q	R	S
3	i	x	y	y-hat	e			
4	1	1	5	4.52	0.48		a	3.900
5	2	3	5	5.76	-0.76		b	0.620
6	3	4	7	6.38	0.62		S	2.780
7	4	5	6	7.00	-1.00			
8	5	7	9	8.24	0.76			
9	6	10	10	10.10	-0.10			
10	平均	5.00	**7.00**	**7.00**	**0.00**			
11	平方和	50.00	**22.00**	**19.22**	**2.78**			

$y = 3.9 + 0.62x$　　式（4.3.1）の形式

↓↑

$y = 7.0 + 0.62 \times (x - 5.0)$　　式（4.3.19）の形式



重点を通ることを示す回帰式が適している場合もある
（切片に意味がないことが少なくない）

数理統計学的な回帰式の推定（最小２乗法）

p.234

●回帰平方和、残差平方和、寄与率の推定

$$r = \frac{S_{xy}}{\sqrt{S_{xx}S_{yy}}} \quad (4.2.5) \qquad R^2 = \frac{S_R}{S_T} \quad (4.3.14)$$

$$S_R = R^2 S_T = r^2 S_{yy}$$

$$= \left(\frac{S_{xy}}{\sqrt{S_{xx}S_{yy}}}\right)^2 \times S_{yy} = \frac{S_{xy}^2}{S_{xx}} \quad (4.3.20)$$

$$= \frac{31.0^2}{50} = 19.22$$

$$S_e = S_T - S_R = S_{yy} - S_R$$

式(4.3.13) から

$$= 22.00 - 19.22 = 2.78$$

	L	M	N	O	P	Q	R	S
3	i	x	y	y-hat	e			
4	1	1	5	4.52	0.48		a	3.900
5	2	3	5	5.76	-0.76		b	0.620
6	3	4	7	6.38	0.62		S	2.780
7	4	5	6	7.00	-1.00			
8	5	7	9	8.24	0.76			
9	6	10	10	10.10	-0.10			
10	平均	5.00	**7.00**	**7.00**	**0.00**			
11	平方和	50.00	**22.00**	**19.22**	**2.78**			

S_e

表示4.3.3

i	x	y
1	1	5
2	3	5
3	4	7
4	5	6
5	7	9
6	10	10
平均	5.0	7.0
平方和	50.0	22.0
積和	31.0	

S_{xx}　$S_{yy} = S_T$　S_{xy}

p.234

（7） LINEST 関数による解法

最小２乗法による回帰式の推定

テキストの項のタイトルとやや異なる

Excel の LINEST 関数による回帰式の推定（最小２乗法）

p.234

●データ

Excel ファイル「基礎改4.xls」、名前ボックスから「表示4.3.8」（Fig43_08）を選択

表示4.3.8　LINEST 関数の結果

	L	M	N	O	
21		x	const		
22	回帰係数	0.620	3.900	切片	
23	その標準誤差	0.118	0.681	その標準誤差	
24	寄与率	0.874	0.834	残差標準偏差	
25	F比	27.655	4	残差自由度	
26	回帰平方和	19.220	2.780	残差平方和	
27	t 値	5.259	5.730	= M22 / M23	
28	p 値	0.006	0.005	= TDIST(ABS(M27), N25, 2)	
29	下限	0.293	2.010	= M22 －M31 * M23	
30	上限	0.947	5.790	= M22 ＋M31 * M23	
31	$t(\alpha)$, α	2.776	0.05	= TINV(N31, N25)	

Excel の LINEST 関数による回帰式の推定（最小２乗法）

p.234

●Excelの LINEST 関数

戻り値を配列で返すことが前提になっている関数の一つ

この他に、FREQUENCY（§2.2）など

線形回帰モデルの解など、

最小２乗法が使われる広範囲の解析に利用できる

（Linear Estimation：線形推定）

実際の出力表示は四角で囲った部分

（M22:N26）

周囲の表示は実行後に追加

第２部、第３部でも使う重要ツール

単回帰分析を例に使い方を習得

枠内が出力表示

周囲は追加

表示4.3.8　LINEST 関数の結果

	L	M	N	O
21		x	const	
22	回帰係数	0.620	3.900	切片
23	その標準誤差	0.118	0.681	その標準誤差
24	寄与率	0.874	0.834	残差標準偏差
25	F比	27.655	4	残差自由度
26	回帰平方和	19.220	2.780	残差平方和
27	t 値	5.259	5.730	= M22 / M23
28	p 値	0.006	0.005	= TDIST(ABS(M27), N25, 2)
29	下限	0.293	2.010	= M22 －M31 * M23
30	上限	0.947	5.790	= M22 ＋M31 * M23
31	$t(\alpha)$, α	2.776	0.05	= TINV(N31, N25)

Excel の LINEST 関数による回帰式の推定（最小 2 乗法）

p.234

●LINEST 関数の使い方（補足）

(A) Excel 2019 以前：静的配列を利用した配列数式（CSE 数式）

古い入力方法（テキストで説明あり）

関数の入力時、Ctrl+Shift+Enter の

3つのキーを同時に押して確定する

(B) Microsoft 365 の Excel、Excel 2021

：動的配列を利用した配列数式

配列の処理を自動的に行う「スピル」

という機能を利用（ブログ 参照）

新しい入力方法（テキストで説明なし）

スピル機能が利用できる Excel ならば、

(B) の方法を使う

(A) (B) ともに
同じ結果を得る
計算式の扱いは異なる

表示4.3.8　LINEST 関数の結果

	L	M	N	O	
21		x	const		
22	回帰係数	0.620	3.900	切片	
23	その標準誤差	0.118	0.681	その標準誤差	
24	寄与率	0.874	0.834	残差標準偏差	
25	F 比	27.655	4	残差自由度	
26	回帰平方和	19.220	2.780	残差平方和	
27	t 値	5.259	5.730	= M22 / M23	
28	p 値	0.006	0.005	= TDIST(ABS(M27), N25, 2)	
29	下限	0.293	2.010	= M22 －M31 * M23	
30	上限	0.947	5.790	= M22 ＋M31 * M23	
31	$t(\alpha)$, α	2.776	0.05	= TINV(N31, N25)	

Excel の LINEST 関数による回帰式の推定（最小２乗法）

p.234

●LINEST 関数の使い方（A）

シート「§4.2モデルの推定(2)」で操作

（1） ５行２列の出力領域を範囲指定して反転させる

（2）反転させた状態で関数を入力　= LINEST(y 範囲, x 範囲, ，TRUE)、入力の確定は (3) で実行

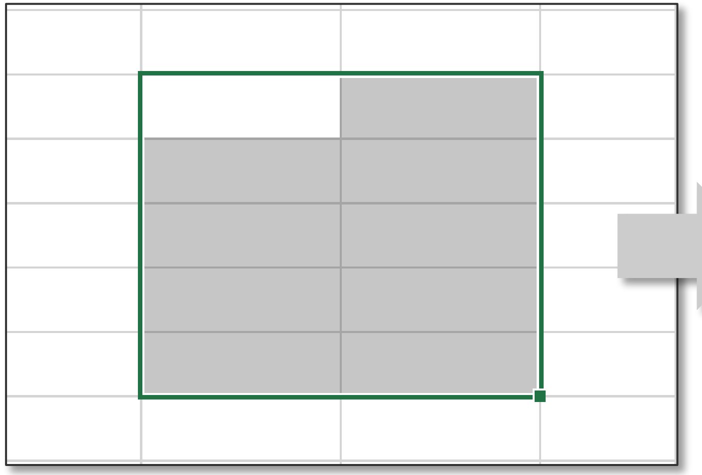
（3）Ctrl キー・Shift キーを同時に押しながら Enter キーを押す

（4）指定した範囲に戻り値が表示される（数式には中括弧 { } で囲まれる）

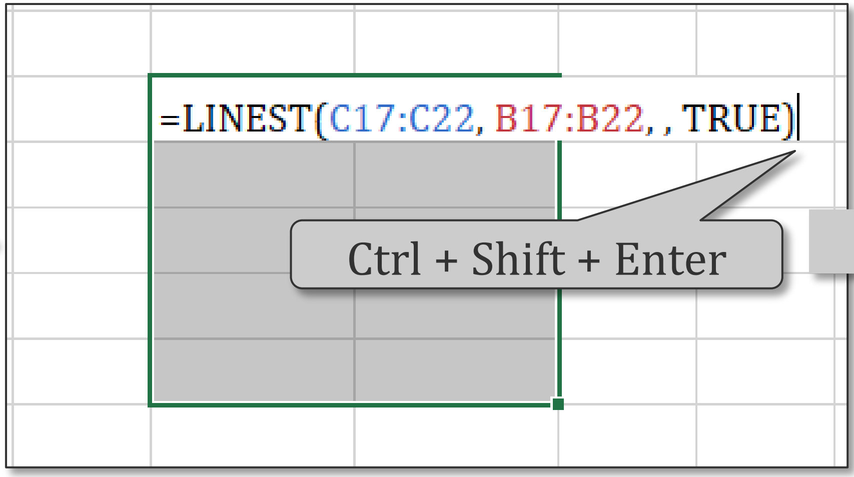
y 範囲 ：C17:C22
x 範囲 ：B17:B22

カンマが２つ
引数 1 個省略

（1） ５行２列を範囲指定



（2）= LINEST(y 範囲, x 範囲, ，TRUE)
（3）Ctrl キー + Shift キー + Enter キー



（4）出力結果

計算式は
自動的に
{} で囲まれる

0.62	3.9
0.1179	0.68069
0.87364	0.83367
27.6547	4
19.22	2.78

Excel の LINEST 関数による回帰式の推定（最小２乗法）

p.234

●LINEST 関数の使い方（B）

シート「§4.2モデルの推定(2)」で操作

（1） ５行２列の出力領域で、左端のトップの位置をクリックします。

（2）関数を入力、= LINEST(y 範囲, x 範囲, , TRUE)、ENTER キーを押して確定

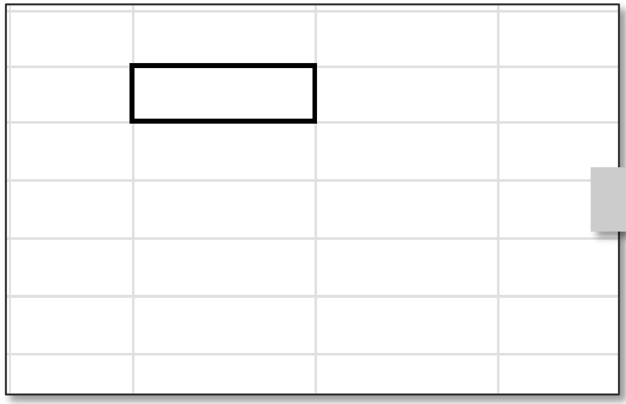
（3）戻り値が表示される

y 範囲：C17:C22
x 範囲：B17:B22

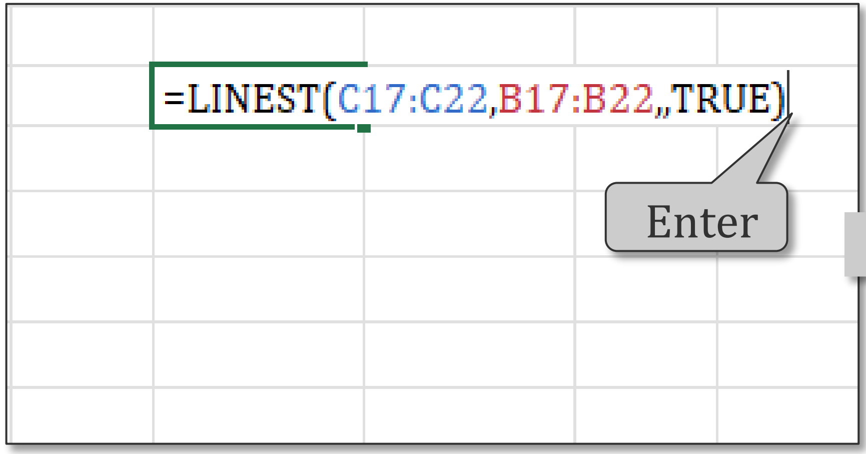
カンマが２つ
引数 １ 個省略

計算式は
自動的に
{}で囲まれない

（1） ５行２列で、左端の
トップのセルをクリック

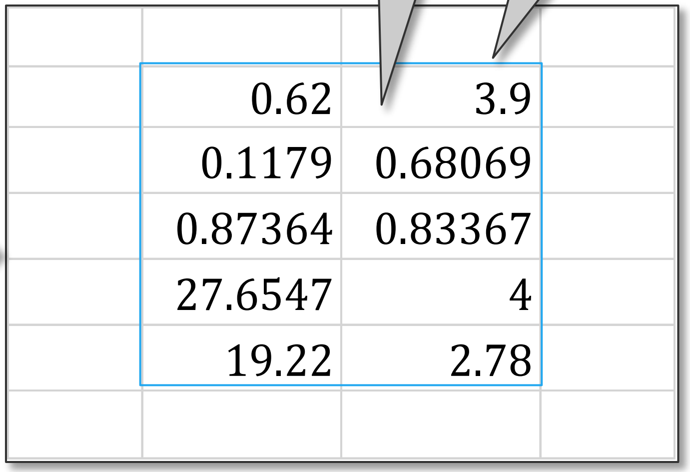


（2）= LINEST(y 範囲, x 範囲, , TRUE)
Enter キーで確定



（3）出力結果

青い線



0.62	3.9
0.1179	0.68069
0.87364	0.83367
27.6547	4
19.22	2.78

Excel の LINEST 関数による回帰式の推定（最小２乗法）

p.235

●LINEST 関数による単回帰分析

LINEST 関数の出力

周囲に説明を追加

表示4.3.8　LINEST 関数の結果

	L	M	N	O	
21		x	const		
22	回帰係数	0.620	3.900	切片	
23	その標準誤差	0.118	0.681	その標準誤差	
24	寄与率	0.874	0.834	残差標準偏差	
25	F比	27.655	4	残差自由度	
26	回帰平方和	19.220	2.780	残差平方和	
27	t 値	5.259	5.730	= M22 / M23	
28	p 値	0.006	0.005	= TDIST(ABS(M27), N25, 2)	
29	下限	0.293	2.010	= M22 －M31 * M23	
30	上限	0.947	5.790	= M22 ＋M31 * M23	
31	$t(\alpha)$, α	2.776	0.05	= TINV(N31, N25)	

$b = S_{xy}/S_{xx}$

$R^2 = S_R/S_T$

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

次節で説明

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

$s = \sqrt{V_e} = \sqrt{\dfrac{S_e}{\nu_e}}$

$\nu_e = n - 2$

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

Excel の LINEST 関数による回帰式の推定（最小２乗法）

p.235

●回帰分析に関する Excel 関数

単回帰直線に関する推定値

切片	=INTERCEPT (y の範囲, x の範囲)
回帰係数（傾き）	=SLOPE (y の範囲, x の範囲)
残差標準偏差	=STEYX (y の範囲, x の範囲)
寄与率（決定係数）	=RSQ (y の範囲, x の範囲)

単回帰分析のみに利用可
重回帰分析には適用できない

	A	B	C	D	E	F	G	H
1	i	x	y			x	const	
2	1	1	5		回帰係数	0.620	3.900	切片
3	2	3	5		その標準誤差	0.118	0.681	その標準誤差
4	3	4	7		寄与率	0.874	0.834	残差標準偏差
5	4	5	6		*F* 比	27.655	4.000	残差自由度
6	5	7	9		回帰平方和	19.220	2.780	残差平方和
7	6	10	10					
8				切片	3.900	=INTERCEPT(C2:C7, B2:B7)		
9				回帰係数	0.620	=SLOPE(C2:C7, B2:B7)		
10				残差標準偏差	0.834	=STEYX(C2:C7, B2:B7)		
11				寄与率	0.874	=RSQ(C2:C7, B2:B7)		

まとめ

p.235

- 単回帰モデル

単回帰モデルの理解と利用

回帰直線の意味

回帰直線の求め方

寄与率 R^2 の意味

→ 各自が演習4.3.1を何度も試み、演習4.3.2を考えて自分のものにする

LINEST関数の利用

統計解析の基礎となる考え方

最小２乗法、平方和の分解、ソルバー、LINEST関数など

第２部、第３部でも取り上げられる

- 引用

  永田　靖（1996）統計的方法のしくみ、日科技連

- 作成　片瀬雅彦
- 監修　松本一彦、長谷文雄
- 作成時期　2019年2月28日
- 改訂　2019年6月8日、2022年5月31日
  2025年3月25日