4 相関・回帰

4.4 誤差を考慮した推定

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

4.4 誤差を考慮した推定

p.236

テキストの
該当ページ

（1）シミュレーション実験

（2）a、b の標準誤差

（3）β の仮説検定と区間推定

（4）予測値と y の区間推定

（5）逆推定

（6）分散分析

（7）JMPによる解析

使用するファイル

Excel ファイル「基本改4.xls」、JMP ファイル「4-相関3.jmp」
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

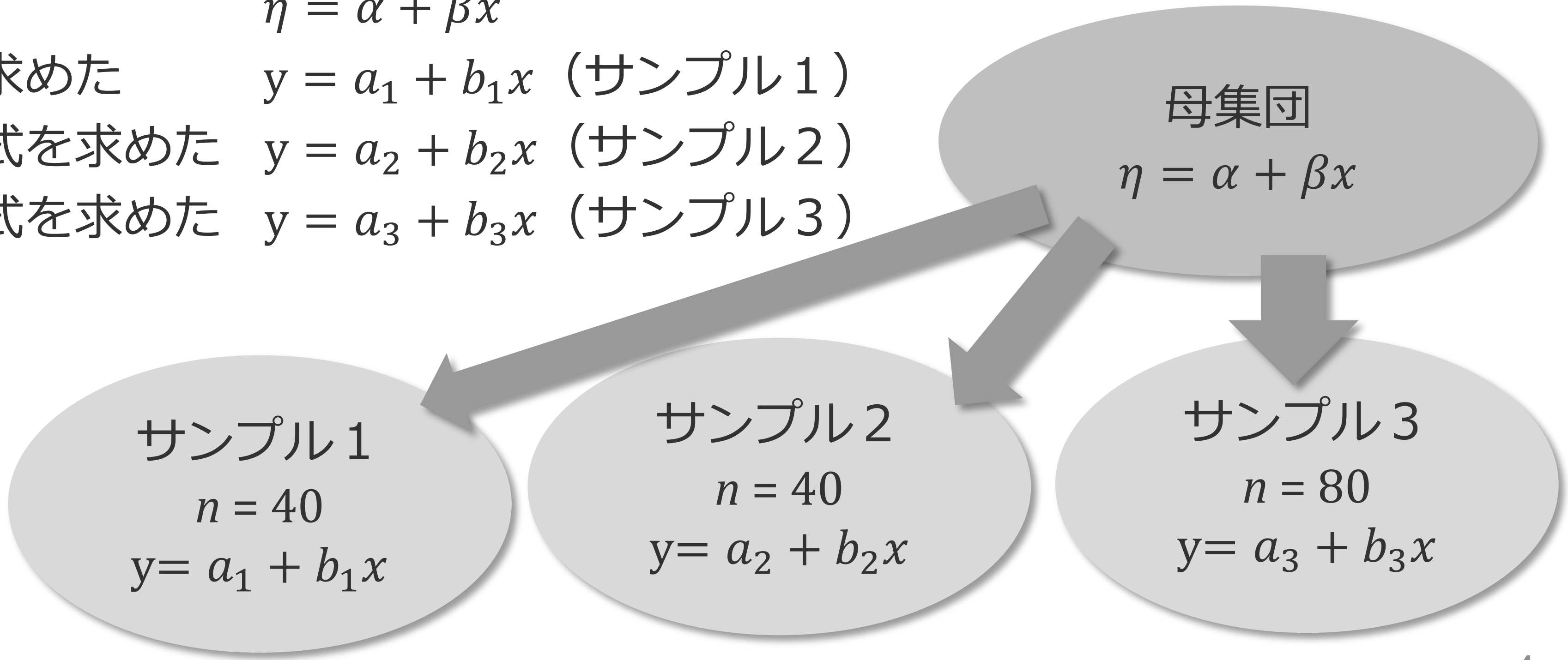
p.236

●単回帰モデル

$y = \eta + \varepsilon = \alpha + \beta x + \varepsilon \qquad \varepsilon \sim \mathrm{N}(0, \sigma^2) \qquad \longrightarrow$ 推定 $y = a + bx$ 　　前節 §4.3

●母集団の回帰式とサンプルの回帰式 （4.3.2）

ある母集団がある	$\eta = \alpha + \beta x$
40人を抽出して回帰式を求めた	$\mathrm{y} = a_1 + b_1 x$（サンプル１）
別の40人を抽出して回帰式を求めた	$\mathrm{y} = a_2 + b_2 x$（サンプル２）
別の80人を抽出して回帰式を求めた	$\mathrm{y} = a_3 + b_3 x$（サンプル３）



はじめに

p.236

●単回帰モデル

$$y = \eta + \varepsilon = \alpha + \beta x + \varepsilon \qquad \varepsilon \sim \mathrm{N}(0, \sigma^2) \qquad \longrightarrow \text{推定} \quad y = a + bx$$

目標　母数　α, β
実際　推定値 a, b
a, b の誤差？ 信頼性？

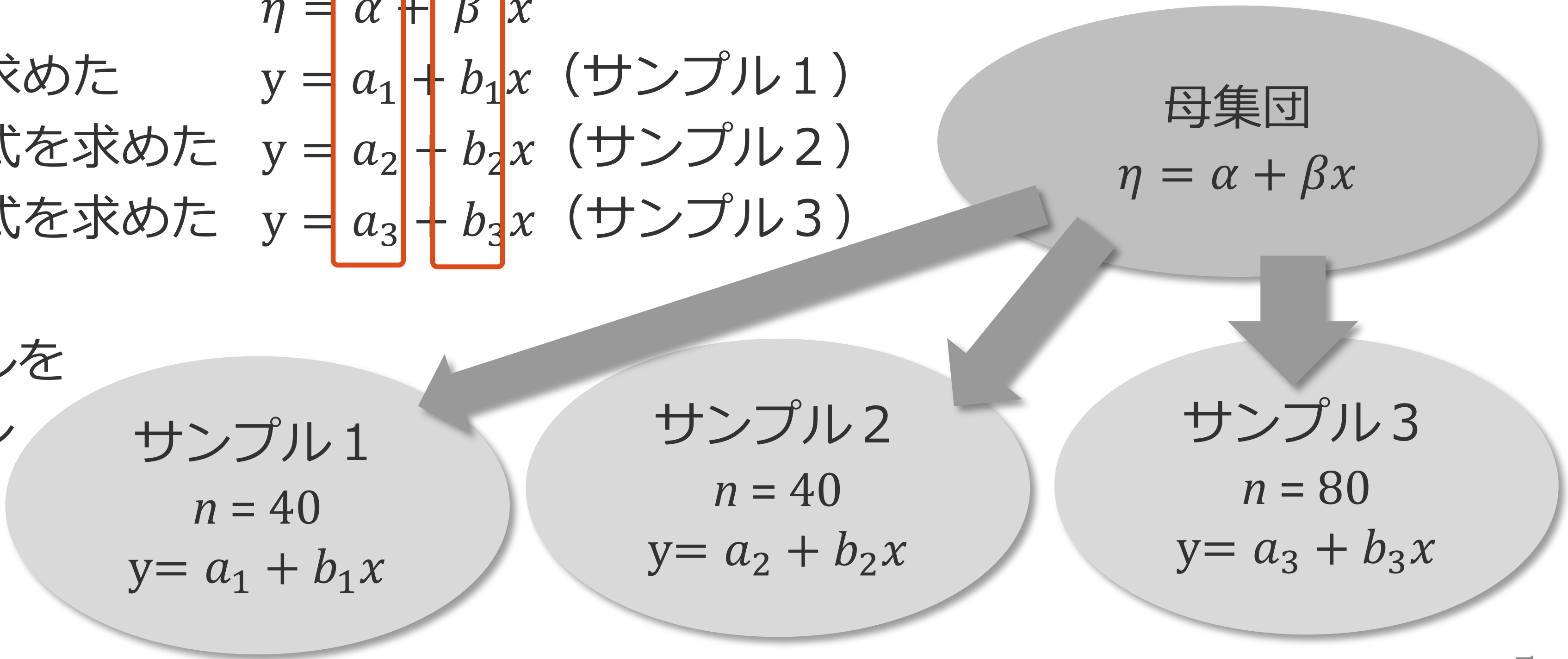
●母集団の回帰式とサンプルの回帰式

(4.2.2)

ある母集団がある	$\eta = \alpha + \beta x$
40人を抽出して回帰式を求めた	$y = a_1 + b_1 x$（サンプル１）
別の40人を抽出して回帰式を求めた	$y = a_2 + b_2 x$（サンプル２）
別の80人を抽出して回帰式を求めた	$y = a_3 + b_3 x$（サンプル３）

↓

既知の母集団からサンプルを
取り出すシミュレーション
　シミュレーション１
　シミュレーション２



p.236

（1）シミュレーション実験

シミュレーション１
シミュレーション２

シミュレーション１

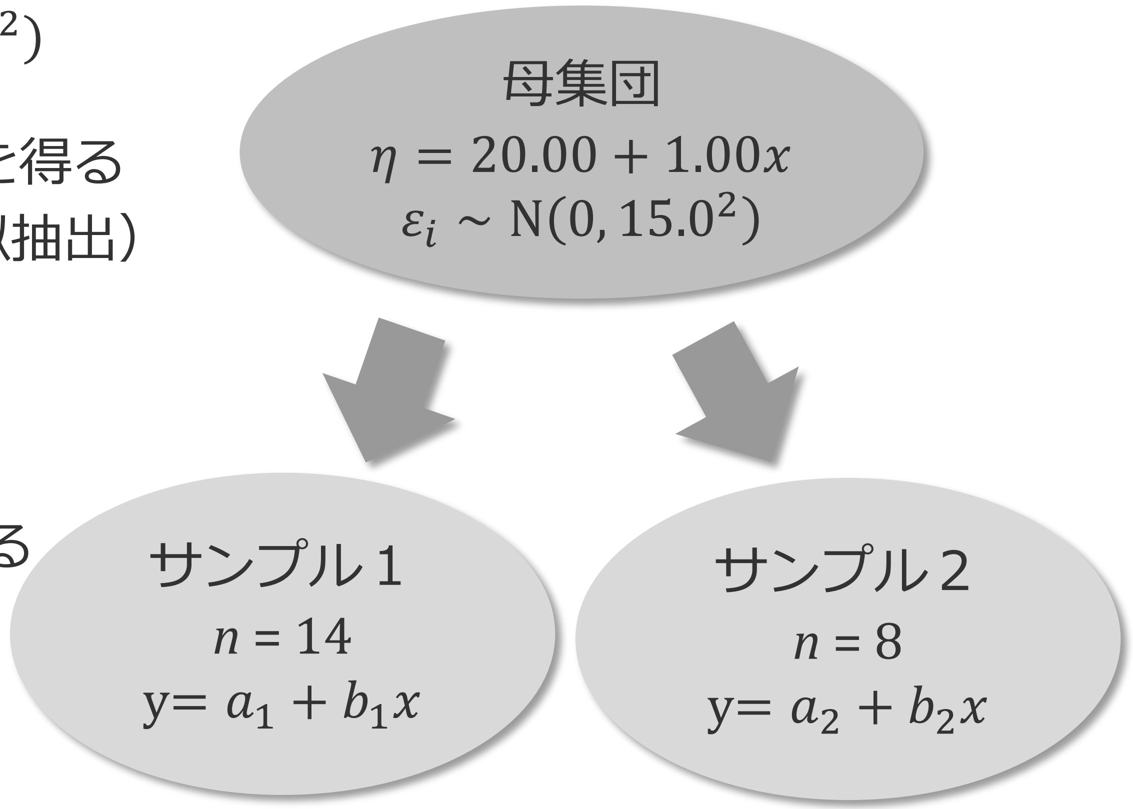
p.236

●シミュレーション１の内容と目的

母集団の回帰式がわかっていると仮定（通常は未知）

$$y_i = \alpha + \beta x_i + \varepsilon_i$$
$$= 20.00 + 1.00x_i + \varepsilon_i \qquad \varepsilon_i \sim \mathrm{N}(0, 15.0^2)$$

このモデルに従う乱数を発生させて、サンプルを得る
（母集団からサンプルを疑似抽出）



サンプルサイズ n は 14 または 8
a と b を計算する
a は20.00、b は 1.00 になることが期待される
実際に得られた a、b が
α、β とどのくらい異なるか確認

シミュレーション１

p.236

●Excelファイルの読み込み

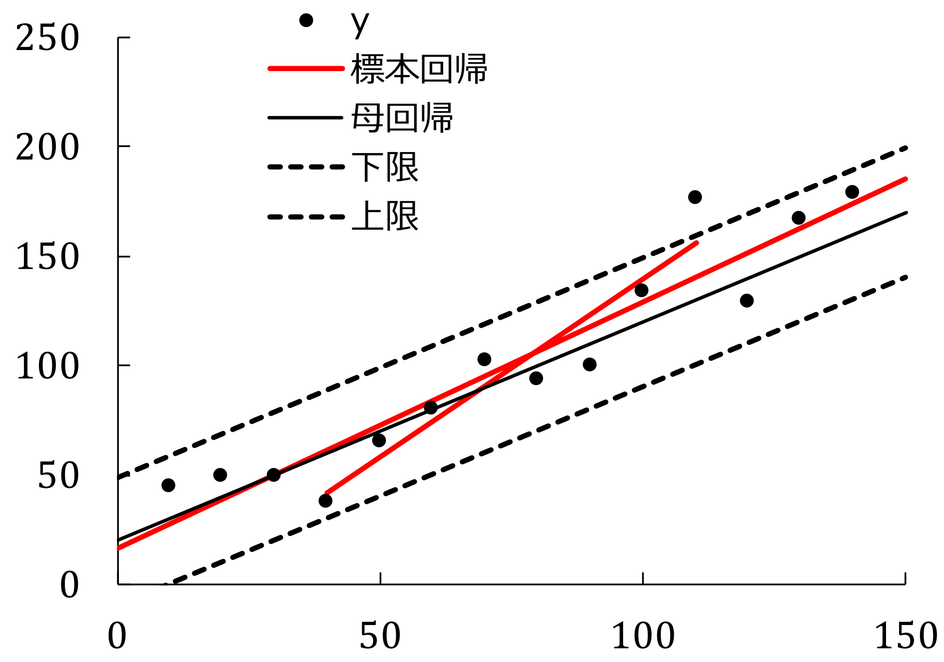
Excel ファイル「基礎改4.xls」、名前ボックスから「表示4.4.1」（Fig44_01）を選択

乱数発生には RAND 関数を利用
以下のタイミングで乱数が更新
・どこかのセルの内容が更新
・ファイルの保存、読み込み
・F9キーを押す

テキスト、Excel ファイル、この PowerPoint ファイルの数値はそれぞれ異なる

表示4.4.1 回帰モデルのシミュレーション

	A	B	C	D	E	F	G	H
3	x1	10			切片	傾き	σ	
4	dx	10		母回帰	20.00	1.00	15.0	
5				標本回帰	16.45	1.13	16.9	
6					-22.97	1.63	15.2	
7		x	y					
8	1	10.0	45.0					
9	2	20.0	49.8					
10	3	30.0	49.5					
11	4	40.0	37.9					
12	5	50.0	65.5					
13	6	60.0	80.4					
14	7	70.0	102.9					
15	8	80.0	94.0					
16	9	90.0	100.6					
17	10	100.0	134.3					
18	11	110.0	176.5					
19	12	120.0	129.3					
20	13	130.0	167.4					
21	14	140.0	179.0					



シミュレーション１

p.236

●図表の概要

表示4.4.1 回帰モデルのシミュレーション

	A	B	C	D	E	F	G	H
3	x1	10			切片	傾き	σ	
4	dx	10		母回帰	20.00	1.00	15.0	
5				標本回帰	16.45	1.13	16.9	
6					-22.97	1.63	15.2	
7		x	y					
8	1	10.0	45.0					
9	2	20.0	49.8					
10	3	30.0	49.5					
11	4	40.0	37.9					
12	5	50.0	65.5					
13	6	60.0	80.4					
14	7	70.0	102.9					
15	8	80.0	94.0					
16	9	90.0	100.6					
17	10	100.0	134.3					
18	11	110.0	176.5					
19	12	120.0	129.3					
20	13	130.0	167.4					
21	14	140.0	179.0					

α、β、σ の設定

a、b、s の計算値

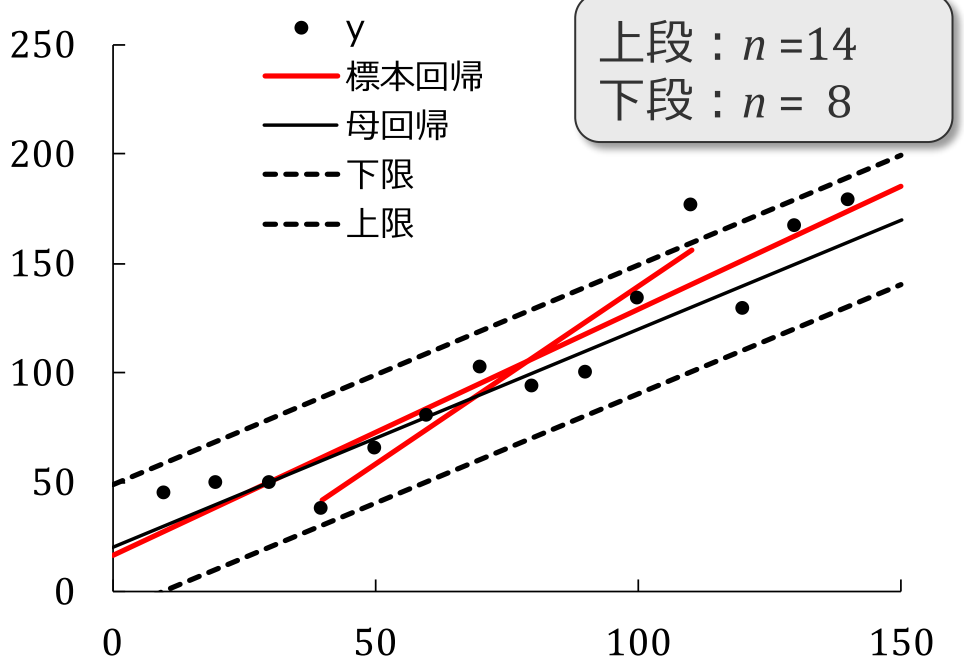
x の設定 10～140

y の乱数発生

14 組のデータ

８組のデータ

上段：n =14
下段：n = 8



シミュレーション１

p.236

● x の設定

セルB8:B21に x を設定

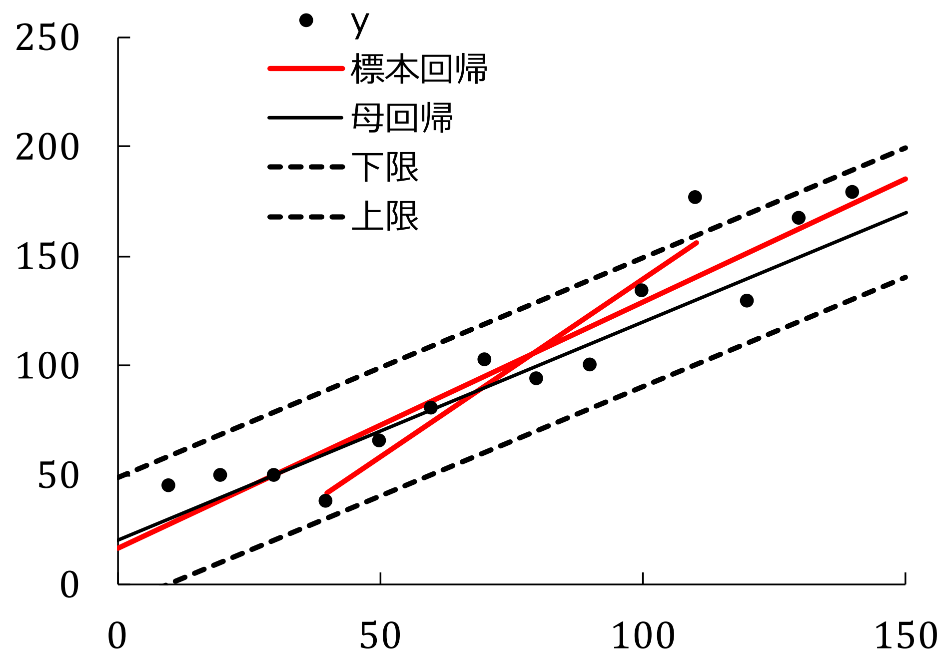
セルA3：１番目（B8）の値

セルA4：x の間隔

x の設定
10～140

表示4.4.1 回帰モデルのシミュレーション

	A	B	C	D	E	F	G	H
3	x1	10			切片	傾き	σ	
4	dx	10		母回帰	20.00	1.00	15.0	
5				標本回帰	16.45	1.13	16.9	
6					-22.97	1.63	15.2	
7		x	y					
8	1	10.0	45.0					
9	2	20.0	49.8					
10	3	30.0	49.5					
11	4	40.0	37.9					
12	5	50.0	65.5					
13	6	60.0	80.4					
14	7	70.0	102.9					
15	8	80.0	94.0					
16	9	90.0	100.6					
17	10	100.0	134.3					
18	11	110.0	176.5					
19	12	120.0	129.3					
20	13	130.0	167.4					
21	14	140.0	179.0					



シミュレーション１

p.236

● y の設定

セルC8:C21に y を発生

E4:G4 に指定したパラメータに従う正規乱数

= RAND()

0～１ の一様乱数を１つ発生

セルの変更、F9キー等で更新

= NORMSINV (RAND())

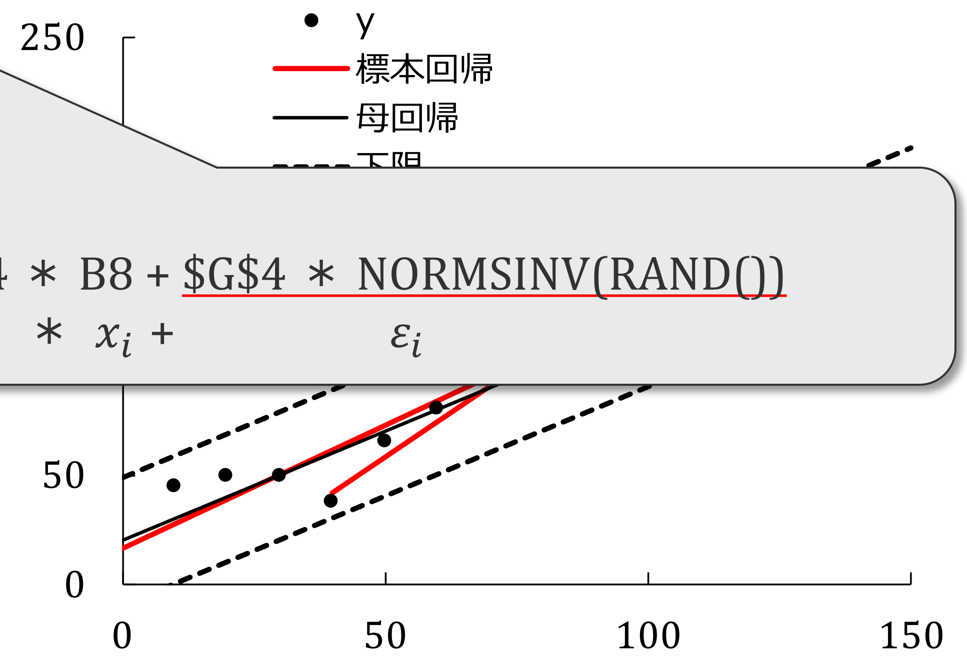
標準正規分布 $N(0, 1^2)$ に従う乱数に変換

σ（セルG4）を乗算

$N(0, 1^2) \rightarrow N(0, 15^2)$に変換

表示4.4.1 回帰モデルのシミュレーション

	A	B	C	D	E	F	G	H
3	x1	10			切片	傾き	σ	
4	dx	10		母回帰	20.00	1.00	15.0	
5				標本回帰	16.45	1.13	16.9	
6					-22.97	1.63	15.2	
7		x	y					
8	1	10.0	45.0					
9	2	20.0	49[illegible]					
10	3	30.0	49[illegible]					
11	4	40.0	37[illegible]					
12	5	[illegible]	[illegible]					
13	6	[illegible]	[illegible]					
14	7	[illegible]	[illegible]					
15	8	[illegible]	[illegible]					
16	9	90.0	100.6					
17	10	100.0	134.3					
18	11	110.0	176.5					
19	12	120.0	129.3					
20	13	130.0	167.4					
21	14	140.0	179.0					



セルC8

`= $E$4 + $F$4 ＊ B8 + $G$4 ＊ NORMSINV(RAND())`

$= \alpha + \beta * x_i + \varepsilon_i$

シミュレーション１

p.236

●標本回帰式の算出

セルE5:G6でパラメータを計算

（§4.3 参照）

=INTERCEPT (y の範囲, x の範囲)

切片 a（p.235）

=SLOPE (y の範囲, x の範囲)

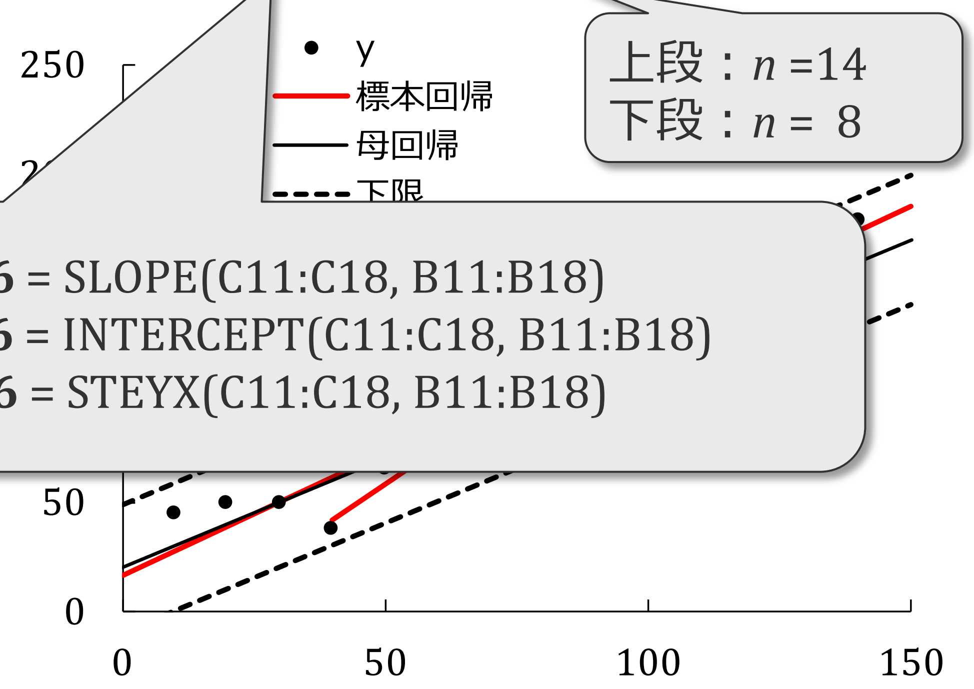
回帰係数 b（p.235）

=STEYX (y の範囲, x の範囲)

残差標準偏差 s（p.233）

表示4.4.1 回帰モデルのシミュレーション

	A	B	C	D	E	F	G	H
3	x1	10			切片	傾き	σ	
4	dx	10		母回帰	20.00	1.00	15.0	
5				標本回帰	16.45	1.13	16.9	
6					-22.97	1.63	15.2	
7		x	y					
8	1	10.0	45.0					
9	2	20.0	49.8					
10	3	30.0	49.5					
11	4	40.0	37.9					
12	5	50.0						
13	6	60.0						
14	7	70.0						
15	8	80.0						
16	9	90.0						
17	10	100.0	134.3					
18	11	110.0	176.5					
19	12	120.0	129.3					
20	13	130.0	167.4					
21	14	140.0	179.0					



上段：n =14
下段：n = 8

セルE6 = SLOPE(C11:C18, B11:B18)
セルF6 = INTERCEPT(C11:C18, B11:B18)
セルG6 = STEYX(C11:C18, B11:B18)

シミュレーション１

p.236

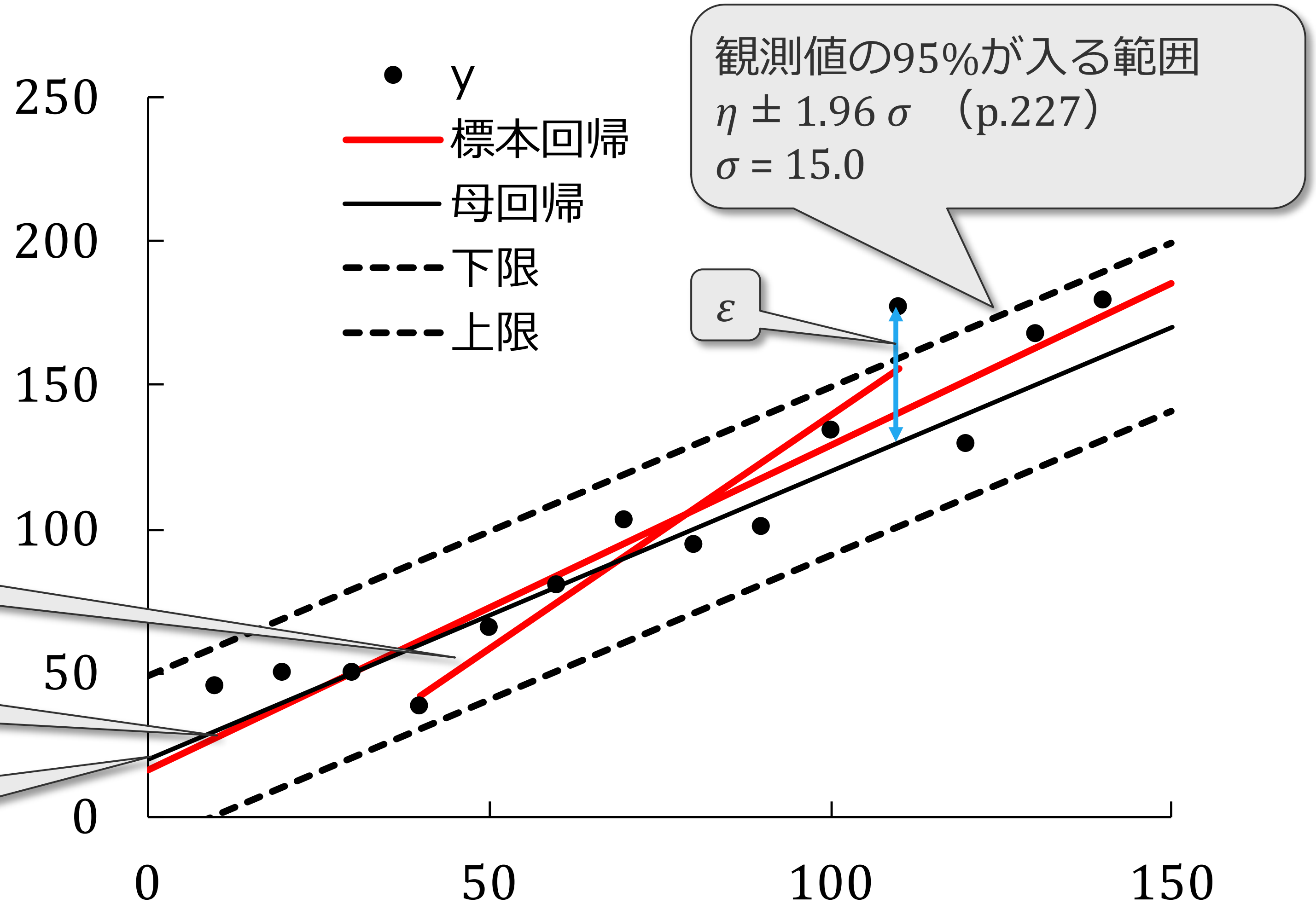
●回帰直線

回帰モデル

$$
\begin{aligned}
y_i &= \eta_i + \varepsilon_i \\
&= \alpha + \beta x_i + \varepsilon_i \\
&= 20.00 + 1.00 x_i + \varepsilon_i
\end{aligned}
$$

$$\varepsilon_i \sim \mathrm{N}(0, 15.0^2)$$

表示4.4.1 回帰モデルのシミュレーション



短い赤線：標本回帰直線（n= 8 ）

長い赤線：標本回帰直線（n=14 ）

黒線：母回帰直線 η = 20.00 + 1.00 x

シミュレーション１

p.236

●乱数の更新

F9キーを押して乱数を更新
回帰直線の変化を確認
n=14 と n=8 の二本の赤い直線の違いに注目

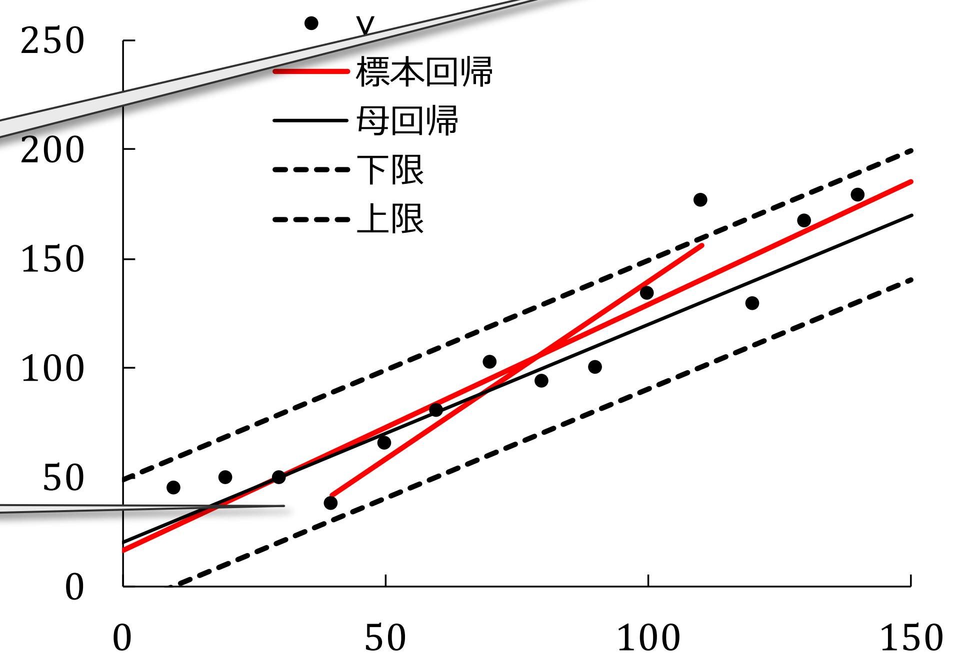
誤差の正規乱数が更新

パラメータが変化

２本の回帰直線が変化

表示4.4.1 回帰モデルのシミュレーション

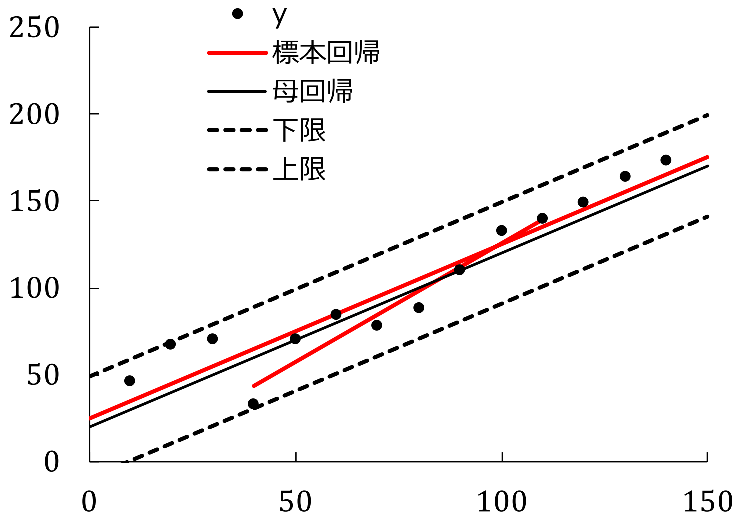
	A	B	C	D	E	F	G	H
3	x1	10			切片	傾き	σ	
4	dx	10		母回帰	20.00	1.00	15.0	
5				標本回帰	16.45	1.13	16.9	
6					-22.97	1.63	15.2	
7		x	y					
8	1	10.0	45.0					
9	2	20.0	49.8					
10	3	30.0	49.5					
11	4	40.0	37.9					
12		50.0	65.5					
13	6	60.0	80.4					
14	7	70.0	102.9					
15	8	80.0	94.0					
16	9	90.0	100.6					
17	10	100.0	134.3					
18	11	110.0	176.5					
19	12	120.0	129.3					
20	13	130.0	167.4					
21	14	140.0	179.0					

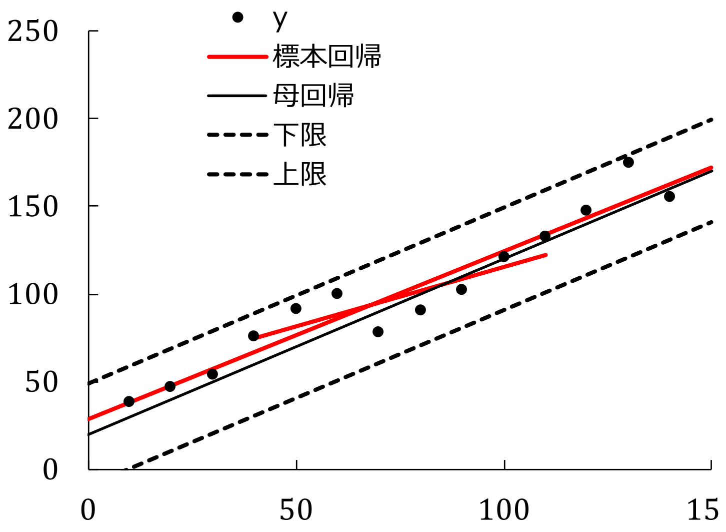


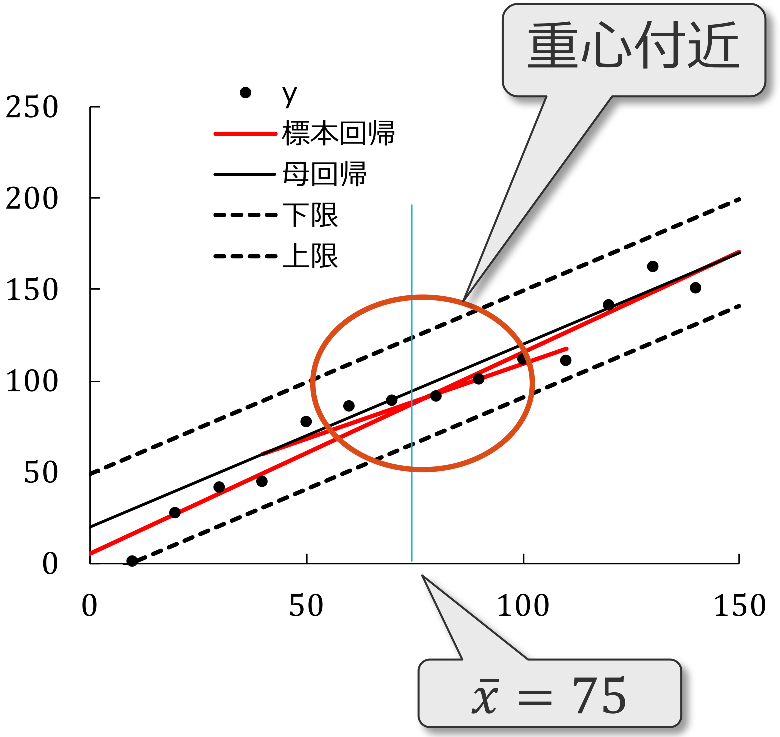
シミュレーション１

p.237

●標本回帰直線の動き







推定値 a と b はばらつく　（α =20.00, β =1.00）

n = 14 (長い線) と n = 8 (短い線) の回帰直線を比較すると、n = 8 の方が動きが大きい

（σ の推定値 : n = 8 がn =14より大）

重心（$\bar{x}, \bar{y}$）付近から離れたところで動きが大きい

p.237

（2） a、b の標準誤差

シミュレーション１
シミュレーション２

シミュレーション２

p.237

●シミュレーション２の内容と目的

母集団の回帰式がわかっていると仮定（通常は未知）

$$y_i = \alpha + \beta x_i + \varepsilon_i$$
$$= 20.00 + 1.00x_i + \varepsilon_i \qquad \varepsilon_i \sim \mathrm{N}(0, 15.0^2)$$

このモデルに従う乱数を発生させて、サンプルを得る

（母集団からサンプルを疑似抽出）

サンプルサイズ n は 14 または 8

<ここまでは、シミュレーション１と同じ>

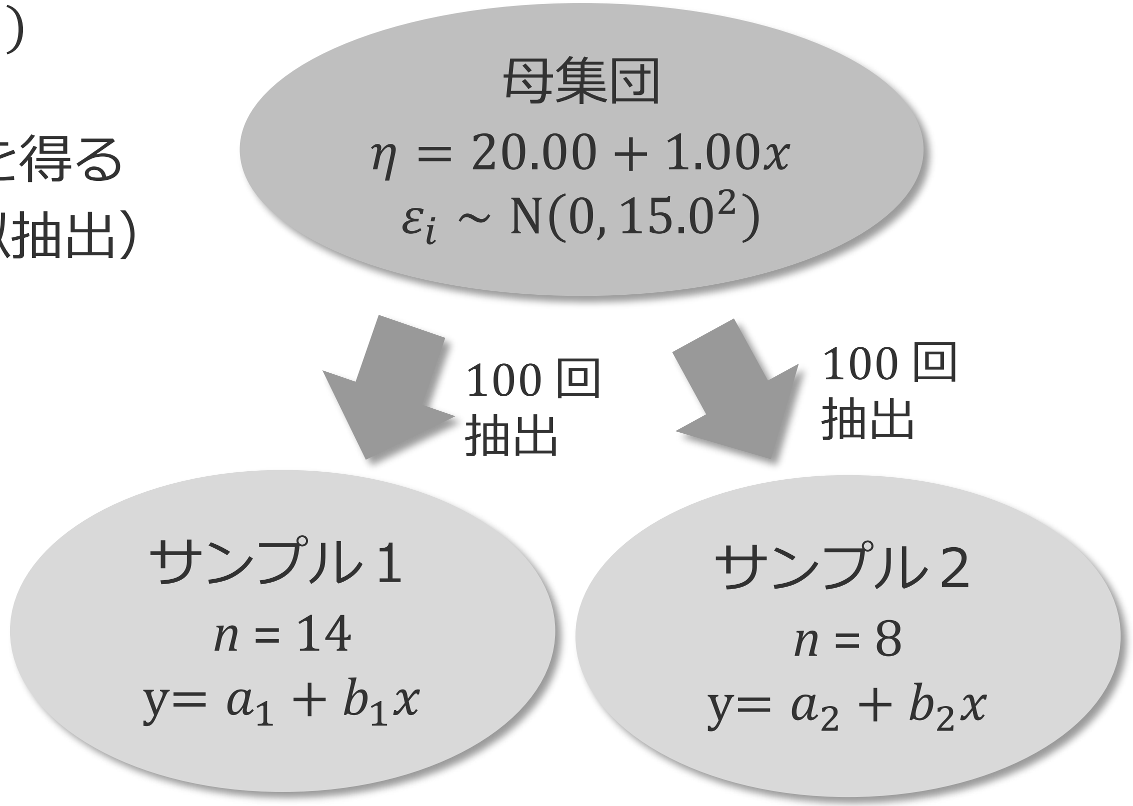
抽出を、それぞれ 100 回行う

a と b を計算する

a と b のバラツキ（標準偏差）を求める

理論値と比較

シミュレーション１で
n=14 と n=8 と比較
後者のばらつきが大きかった



シミュレーション２

● Excelファイルの読み込み

Excel ファイル「基礎改4.xls」、名前ボックスから「表示4.4.2」（Fig44_02）を選択

● シミュレーション１と２の内容と目的

シミュレーション１：表示4.4.1　$n = 8, n = 14$ の回帰式を数組得て、ばらつきを目視で確認

シミュレーション２：表示4.4.2　$n = 8, n = 14$ の回帰式を 100組得て、標準誤差を計算、理論値と比較

Excelのデータテーブル機能を利用（p96、p.263）

データテーブルの使い方は省略

表示4.4.2 100回のシミュレーションの過程と結果

	J	K	L	M	N	O	P	Q	R
3						n=14		n=8	
4						a	b	a	b
5	α	20.00	ｼﾐｭﾚｰｼｮﾝ		平均	19.717	1.002	18.353	1.026
6	β	1.00			標準偏差	8.177	0.099	21.854	0.303
7	σ	15.0	理論値		平均	20.000	1.000	20.000	1.000
8					標準偏差	8.468	0.099	18.151	0.231
10						n=14		n=8	
11		x	y			a	b	a	b
12	1	10	33.41		*	22.11	0.95	-1.58	1.26
13	2	20	47.21		1	12.32	1.13	-10.09	1.37

シミュレーション２

p.238

●シートの概要

表示4.4.2

	J	K	L	M	N	O	P	Q	R
3						n=14		n=8	
4						a	b	a	b
5	α	20.00	ｼﾐｭﾚｰｼｮﾝ 平均			19.717	1.002	18.353	1.026
6	β	1.00	標準偏差			8.177	0.099	21.854	0.303
7	σ	15.0	理論値 平均			20.000	1.000	20.000	1.000
8			標準偏差			8.468	0.099	18.151	0.231
10						n=14		n=8	
11		x	y			a	b	a	b
12	1	10	[illegible]		*	22.11	0.95	-1.58	1.26
13	[illegible]	20	47.21		1	12.32	1.13	-10.09	1.37
14	3	30	57.95		2	26.21	0.87	46.38	0.66
15	4	40	43.78		3	7.03	1.12	35.06	0.82
16	5	50	76.14		4	18.93	1.05	11.99	1.17
17	6	60	60.83		5	11.98	1.12	4.65	1.21
18	7	70	87.55		6	18.75	1.01	38.76	0.67
19	8	80	105.65		7	22.26	0.94	34.74	0.78
20	9	90	104.26		8	12.63	1.09	-4.41	1.35
21	10	100	127.60		9	23.48	0.89	22.47	1.07
22	11	110	138.19		10	32.89	0.85	16.64	1.02
23	12	120	128.34		11	30.25	0.87	31.00	0.82
24	13	130	157.26		12	15.39	1.05	1.08	1.34
25	14	140	143.61		13	15.24	1.10	-16.02	1.54
					⋮	⋮	⋮	⋮	⋮

母回帰式の α、β、σ

***a*, *b* の計算**
セルO12：n=14 の a
セルP12：n=8 の b
セルO12：n=14 の a
セルP12：n=8 の b

x の設定 10～140

y の乱数発生

シミュレーション２

p.238

●シートの概要

表示4.4.2

母回帰式の α、β、σ

	J	K	L	M	N	O	P	Q	R
3						n=14		n=8	
4						a	b	a	b
5	α	20.00	シミュレーション		平均	19.717	1.002	18.353	1.026
6	β	1.00			標準偏差	8.177	0.099	21.854	0.303
7	σ	15.0	理論値		平均	20.000	1.000	20.000	1.000
8					標準偏差	8.468	0.099	18.151	0.231

	J	K	L	M	N	O	P	Q	R
10						n=14		n=8	
11		x	y			a	b	a	b
[illegible]	[illegible]	10	23.41		*	22.1	0.95	-1.58	1.26
[illegible]	[illegible]	[illegible]	.21		1	12.3	1.13	-10.09	1.37
[illegible]	[illegible]	30	57.95		2	26.2	0.87	46.38	0.66
15	4	40	43.78		3	7.0	1.12	35.06	0.82
16	5	50	76.14		4	18.9	1.05	11.99	1.17
17	6	60	60.83		5	11.9	1.12	4.65	1.21
18	7	70	87.55		6	18.75	1.01	38.76	0.67
19	8	80	105.65		7	22.26	0.9	[illegible]	[illegible]
20	9	90	104.26		8	12.63	1.0	[illegible]	[illegible]
21	10	100	127.60		9	23.48	0.8	[illegible]	[illegible]
22	11	110	138.19		10	32.89	0.8	[illegible]	[illegible]
23	12	120	128.34		11	30.25	0.8	[illegible]	[illegible]
24	13	130	157.26		12	15.39	1.0	[illegible]	[illegible]
25	14	140	143.61		13	15.24	1.10	-16.02	1.54
					⋮	⋮	⋮	⋮	⋮

***a, b* の計算**

セルO12：n=14 の a

= INTERCEPT(L12:L25, K12:K25)

セルP12：n=8 の b

= SLOPE(L12:L25, K12:K25)

セルQ12：n=14 の a

= INTERCEPT(L15:L22, K15:K22)

セルR12 ：n=8 の b

= SLOPE(L15:L22, K15:K22)

乱数を 100 回更新して *a, b* を計算
計算結果の * の行を
その下の行から順に100回コピー
（データテーブルの機能）

y の乱数発生

シミュレーション２

p.238

●シートの概要

表示4.4.2

	J	K	L	M	N	O	P	Q	R
3						n=14		n=8	
4						a	b	a	b
5	α	20.00	シミュレーション		平均	19.717	1.002	18.353	1.026
6	β	1.00			標準偏差	8.177	0.099	21.854	0.303
7	σ	15.0	理論値		平均	20.000	1.000	20.000	1.000
8					標準偏差	8.468	0.099	18.151	0.231
9									
10						n=14		n=8	
11		x	y			a	b	a	b
12	1	10	33.41		*	22.1[illegible]	0.95	-1.58	1.26
13	2	20	47.21		1	12.3[illegible]	1.13	-10.09	1.37
14	3	30	57.95		2	26.2[illegible]	0.87	46.38	0.66
15	4	40	43.78		3	7.0[illegible]	1.12	35.06	0.82
16	5	50	76.14		4	18.9[illegible]	1.05	11.99	1.17
17	6	60	60.83		5	11.9[illegible]	1.12	4.65	1.21
18	7	70	87.55		6	18.75	1.01	38.76	0.67
19	8	80	105.65		7	22.26	0.94	34.74	0.7[illegible]
20	9	90	104.26		8	12.63	1.09	-4.41	1.35
21	10	100	127.60		9	23.48	0.89	22.47	1.07
22	11	110	138.19		10	32.89	0.85	16.64	1.02
23	12	120	128.34		11	30.25	0.87	31.00	0.82
24	13	130	157.26		12	15.39	1.05	1.08	1.34
25	14	140	143.61		13	15.24	1.10	-16.02	1.5[illegible]
					⋮	⋮	⋮	⋮	⋮

母回帰式の α、β、σ

100 個の a, b の平均と標準偏差

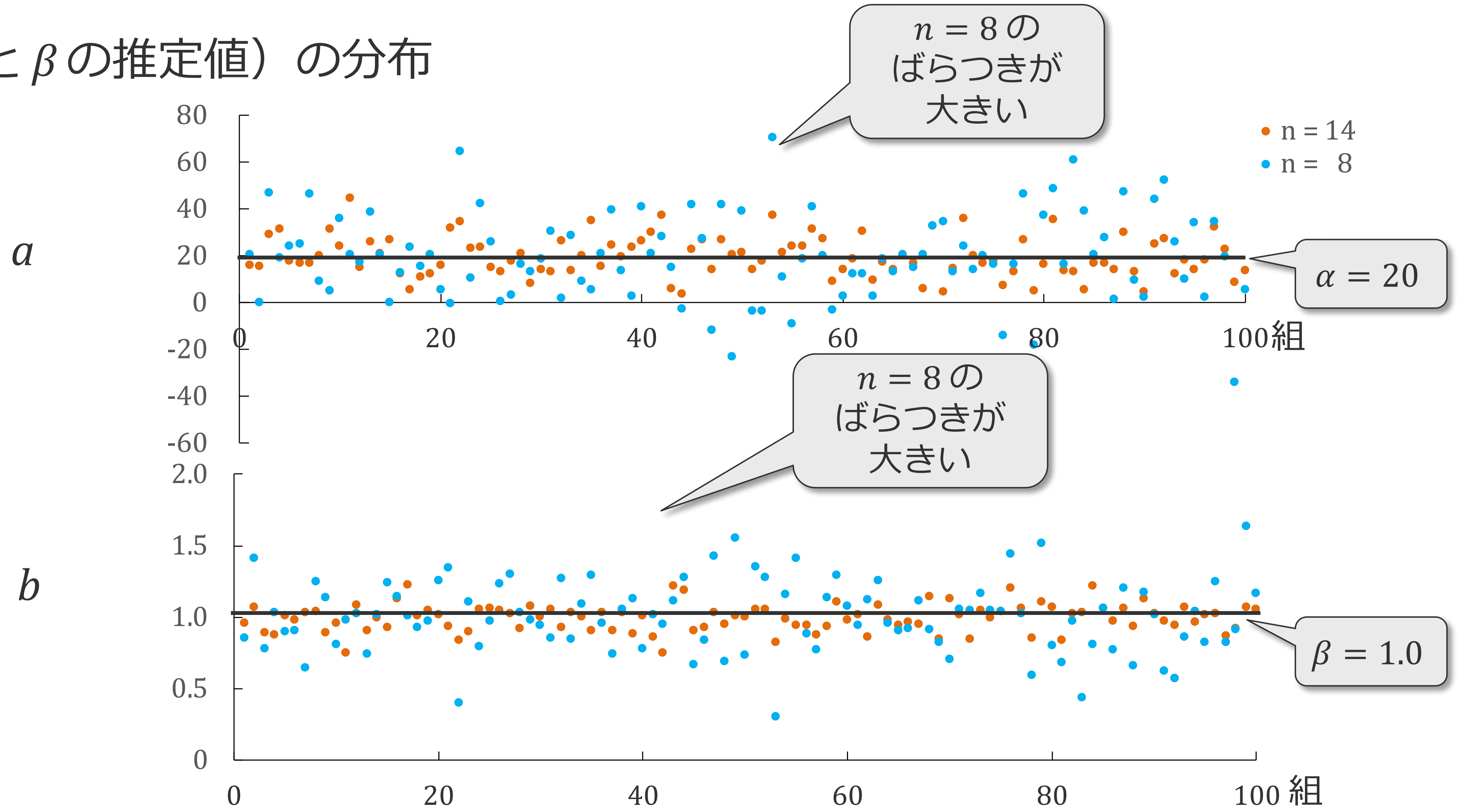
100 個の a, b

まずは、グラフ化

シミュレーション２

p.238

- a と b （α と β の推定値）の分布



シミュレーション 2

p.238

●シミュレーションの結果

n の大きい方が、a、b の推定値（シミュレーション）は理論値に近い

n の大きい方が、 a、b の標準偏差は小さい（シミュレーションと理論値ともに）

F9キーを押し、乱数を更新して確認

	J	K	L	M	N	O	P	Q	R
3						n=14		n=8	
4						a	b	a	b
5	α	20.00	ｼﾐｭﾚｰｼｮﾝ　平均			19.717	1.002	18.353	1.026
6	β	1.00	標準偏差			8.177	0.099	21.854	0.303
7	σ	15.0	理論値　平均			20.000	1.000	20.000	1.000
8			標準偏差			8.468	0.099	18.151	0.231
10						n=14		n=8	
11		x	y			a	b	a	b
12	1	10	33.41		*	22.11	0.95	-1.58	1.26
13	[illegible]	[illegible]	[illegible]		1	[illegible]	[illegible]	[illegible]	[illegible]
14	[illegible]	[illegible]	[illegible]		2	[illegible]	[illegible]	[illegible]	[illegible]
15	4	40	43.78		3	7.03	1.12	35.06	0.82

観測値（乱数）

100組の推定値

シミュレーション２

p.238

●シミュレーションの結果

n の大きい方が、a、b の推定値（シミュレーション）は理論値に近い
n の大きい方が、 a、b の標準偏差は小さい（シミュレーションと理論値ともに）
F9キーを押し、乱数を更新して確認

観測値から計算された
推定値の標準偏差
＝標準誤差（p.90）
↓
理論値

	J	K	L	M	N	O	P	Q	R
3						n=14		n=8	
4						a	b	a	b
5	α	20.00	ｼﾐｭﾚｰｼｮﾝ		平均	19.717	1.002	18.353	1.026
6	β	1.00			標準偏差	8.177	0.099	21.854	0.303
7	σ	15.0	理論値		平均	20.000	1.000	20.000	1.000
8					標準偏差	8.468	0.099	18.151	0.231
10						n=14		n=8	
11		x	y			a	b	a	b
12	1	10	33.41		*	22.11	0.95	-1.58	1.26
13	[illegible]	[illegible]	[illegible]		1	[illegible]	[illegible]	[illegible]	[illegible]
14	[illegible]	[illegible]	[illegible]		2	[illegible]	[illegible]	[illegible]	[illegible]
15	4	40	43.78		3	7.03	1.12	35.06	0.82

標準誤差

観測値（乱数）

100組の推定値

回帰係数（b）と切片（a）の標準誤差

p.239

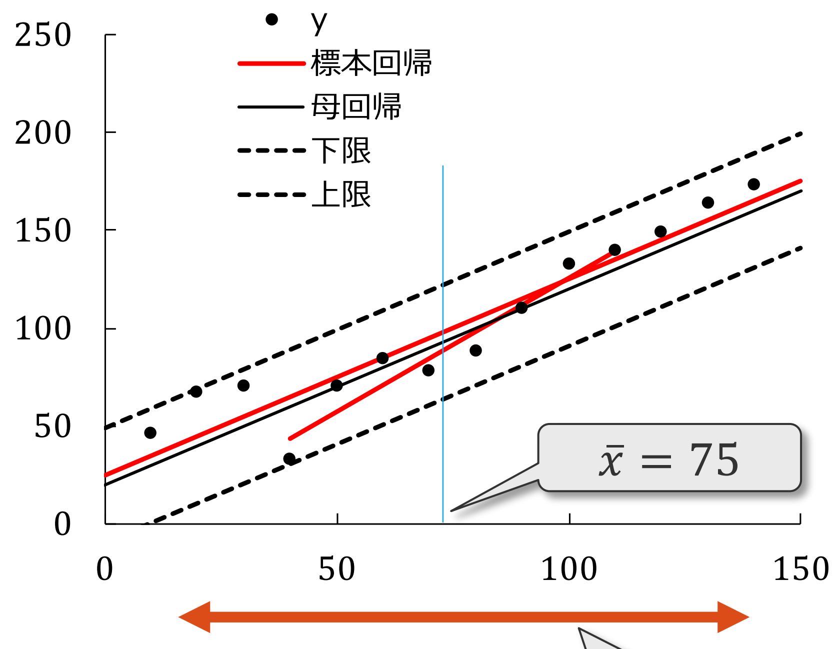
●理論値

$$s.e.[b] = \frac{\sigma}{\sqrt{S_{xx}}} \qquad (4.4.1)$$

$$S_{xx} = \sum_{i=1}^{n}(x_i - \bar{x})^2 \qquad ・・・平方和$$

$$s.e.[a] = \sqrt{\frac{1}{n} + \frac{\bar{x}^2}{S_{xx}}} \times \sigma \qquad (4.4.2)$$

表示4.4.1



x_iの変化の範囲が大きく、
nが大きいほど、S_{xx}は大きくなる
→ $s.e.[b]$、 $s.e.[b]$ は小さくなる（表示4.4.1 で確認済み）

理論値の求め方は省略（§4.6 (3) p.257 参照）

x_i を幅広く、かつ
たくさん設定することにより
標準誤差が小さくなる
（直線性がある範囲で）

回帰係数（b）と切片（a）の標準誤差

p.239

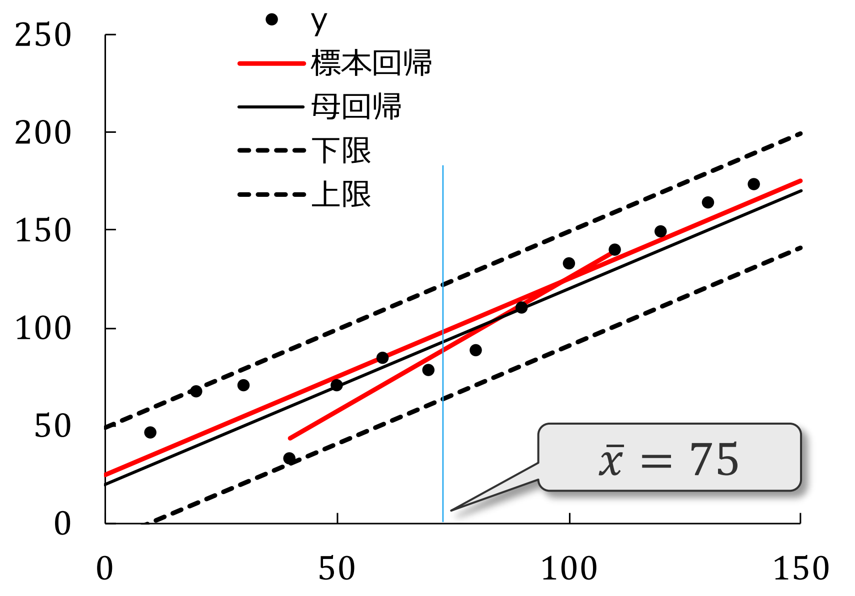
●理論値

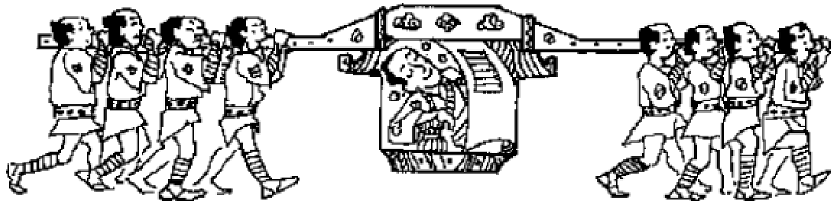
$$s.e.[b] = \frac{\sigma}{\sqrt{S_{xx}}} \qquad (4.4.1)$$

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

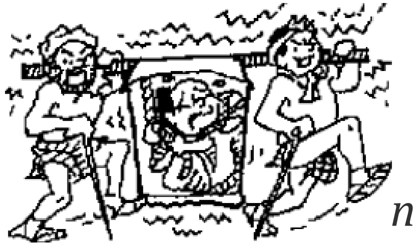
$$s.e.[a] = \sqrt{\frac{1}{n} + \frac{\bar{x}^2}{S_{xx}}} \times \sigma \qquad (4.4.2)$$

表示4.4.1





$n = 14$



$n = 8$

x_iの変化の範囲が大きく、
nが大きいほど、S_{xx}は大きくなる
→ $s.e.[b]$、$s.e.[b]$ は小さくなる（表示4.4.1 で確認済み）

理論値の求め方は省略（§4.6 (3) p.257 参照）

p.239 脚注

回帰係数（b）と切片（a）の標準誤差

p.239

● σ 既知の場合の計算事例（ 表示4.4.2 n=14の事例 ）

$$s.e.[b] = \frac{\sigma}{\sqrt{S_{xx}}} = \frac{15}{\sqrt{22750}} = 0.099 \qquad (4.4.1)$$

$$s.e.[a] = \sqrt{\frac{1}{n} + \frac{\bar{x}^2}{S_{xx}}} \times \sigma = \sqrt{\frac{1}{14} + \frac{75^2}{22750}} \times 15 = 8.468 \qquad (4.4.2)$$

表示4.4.2 100回のシミュレーションの過程と結果

	J	K	L	M	N	O	P	Q	R
3						n=14		n=8	
4						a	b	a	b
5	α	20.00	ｼﾐｭﾚｰｼｮﾝ		平均	19.717	1.002	18.353	1.026
6	β	1.00			標準偏差	8.177	0.099	21.854	0.303
7	σ	15.0	理論値		平均	20.000	1.000	20.000	1.000
8					標準偏差	8.468	0.099	18.151	0.231

表示4.4.2 の
x の平均値と残差平方和

	x_i	$(x_i-75)^2$
1	10	4,225
2	20	3,025
3	30	2,025
4	40	1,225
5	50	625
6	60	225
7	70	25
8	80	25
9	90	225
10	100	625
11	110	1,225
12	120	2,025
13	130	3,025
14	140	4,225
平均	75	22,750

回帰係数（b）と切片（a）の標準誤差

p.239

●σ 未知の場合の計算事例（表示4.3.6 の事例　p. 231）

$$s.e.[b] = \frac{\sigma}{\sqrt{S_{xx}}} = \frac{0.834}{\sqrt{50.0}} = 0.118 \quad (4.4.1)$$

母標準偏差 σ は分からない
推定値である残差標準偏差 s を用いる

$$s.e.[a] = \sqrt{\frac{1}{n} + \frac{\bar{x}^2}{S_{xx}}} \times \sigma = \sqrt{\frac{1}{6} + \frac{5.00^2}{50.00}} \times 0.834 = 0.681 \quad (4.4.2)$$

表示4.3.6 ソルバーの解

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

a	3.900
b	0.620
S	2.780

表示4.3.8 LINEST関数の結果（p.235）

	x	const	
回帰係数	0.620	3.900	切片
その標準誤差	0.118	0.681	その標準誤差
寄与率	0.874	0.834	残差標準偏差
F 比	27.655	4	残差自由度
回帰平方和	19.220	2.780	残差平方和

回帰係数（b）と切片（a）の標準誤差

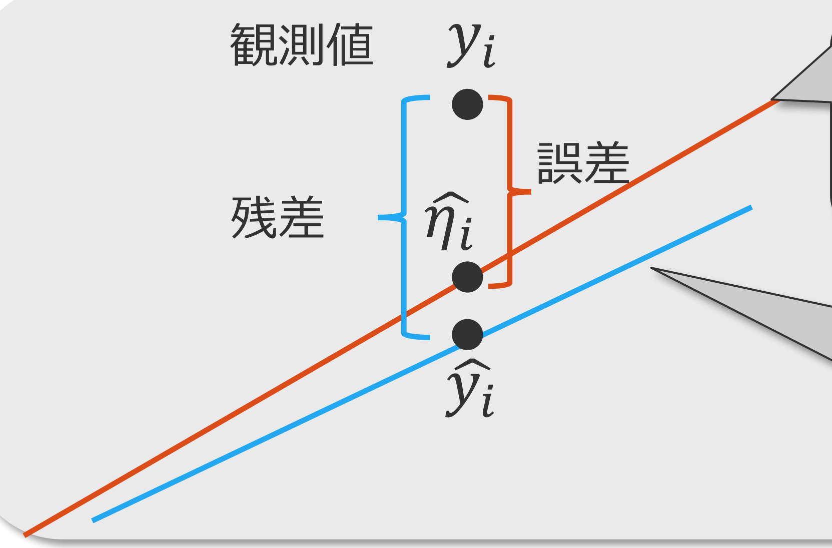
p.239

●残差と誤差

残差から誤差を推定

s は σ の推定値

（前節、p.233）



母回帰式（真の回帰式）

$y_i = \alpha + \beta x_i + \varepsilon_i \quad \varepsilon_i \sim \mathrm{N}(0, \sigma^2)$

$\hat{\eta}_i = \alpha + \beta x_i$

標本回帰式（推定した回帰式）

$y_i = a + b x_i + e_i$

$\hat{y}_i = a + b x_i$

表示4.3.6 ソルバーの解

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

a	3.900
b	0.620
S	2.780

表示4.3.8 LINEST関数の結果

	x	const	
回帰係数	0.620	3.900	切片
その標準誤差	0.118	0.681	その標準誤差
寄与率	0.874	0.834	残差標準偏差
F比	27.655	4	残差自由度
回帰平方和	19.220	2.780	残差平方和

p.240

（3）βの仮説検定と区間推定

推定したパラメータβの信頼性

β の仮説検定と区間推定

p.240

●帰無仮説　H_0：$\beta = 0$　の検定（H_1： $\beta \neq 0$）

x と y の間の回帰直線が水平であるかどうかを検定する

母相関係数が 0 であるかどうかの検定と同等

帰無仮説が棄却できなければ、x で y を説明できない

σ は未知なので、推定量である残差標準偏差 s を用いる

b が標準誤差 $s.e.[b]$ の何倍であるか、この値が t 分布（残差の自由度、$n-2$）に従う

$$t = \frac{b}{s.e.[b]} \qquad s.e.[b] = \frac{\sigma}{\sqrt{S_{xx}}} \sim \frac{s}{\sqrt{S_{xx}}} \qquad S_{xx} = \sum_{i=1}^{n} (x_i - \bar{x})^2$$

（4.4.1）

β の仮説検定

p.240

●帰無仮説　$H_0：\beta = 0$　の検定（$H_1：\beta \neq 0$）

$$t = \frac{b}{s.e.[b]} = \frac{0.620}{0.118} = 5.259$$

=TDIST(5.259, 4, 2) = 0.006

=T.DIST.2T(5.259, 4) = 0.006

有意水準0.05で β は有意
回帰直線は水平ではない
y は x によって有意に変化する

表示4.3.6 ソルバーの解

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

a	3.900
b	0.620
S	2.780

表示4.3.8 LINEST関数の結果

	x	const	
回帰係数	0.620	3.900	切片
その標準誤差	0.118	0.681	その標準誤差
寄与率	0.874	0.834	残差標準偏差
F比	27.655	4	残差自由度
回帰平方和	19.220	2.780	残差平方和

β の区間推定

p.240

●回帰係数 β の95%信頼区間（表示4.3.3のデータ）

$$-t(0.05,\nu_e) \le \frac{b-\beta}{s.e.[b]} \le t(0.05,\nu_e)$$

$$b - t(0.05,\nu_e)s.e.[b] \le \beta \le b + t(0.05,\nu_e)s.e.[b]$$

$$0.620 - 2.776 \times 0.118 \le \beta \le 0.620 + 2.776 \times 0.118$$

$$0.293 \le \beta \le 0.947$$

$t(0.05,\nu_e)$ は両側確率0.05の t 値

$= \mathrm{TINV}(0.05,4) = 2.776$　両側確率

$= \mathrm{T.INV}(0.975,4) = 2.776$　下側確率

$= \mathrm{T.INV.2T}(0.05,4) = 2.776$　両側確率

表示4.3.6 ソルバーの解

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

a	3.900
b	0.620
S	2.780

表示4.3.8 LINEST関数の結果

	x	const	
回帰係数	0.620	3.900	切片
その標準誤差	0.118	0.681	その標準誤差
寄与率	0.874	0.834	残差標準偏差
F比	27.655	4	残差自由度
回帰平方和	19.220	2.780	残差平方和

α の仮説検定

●切片 α の仮説検定（表示4.3.3のデータ）

$$t = \frac{a}{s.e.[a]} = \frac{3.900}{0.6807} = 5.729 \qquad (4.4.2)$$

=TDIST(5.729, 4, 2) = 0.005

=T.DIST.2T(5.729, 4) = 0.005

有意水準0.05で α は有意
切片はゼロと有意に異なる
原点を通らない

表示4.3.6 ソルバーの解

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

a	3.900
b	0.620
S	2.780

表示4.3.8 LINEST関数の結果

	x	const	
回帰係数	0.620	3.900	切片
その標準誤差	0.118	0.681	その標準誤差
寄与率	0.874	0.834	残差標準偏差
F比	27.655	4	残差自由度
回帰平方和	19.220	2.780	残差平方和

α の区間推定

●切片 α の95%区間推定（表示4.3.3のデータ）

$$-t(0.05,\nu_e) \le \frac{a-\alpha}{s.e.[a]} \le t(0.05,\nu_e)$$

$$a - t(0.05,\nu_e)s.e.[a] \le \alpha \le a + t(0.05,\nu_e)s.e.[a]$$

$$3.900 - 2.776 \times 0.681 \le \alpha \le 3.900 + 2.776 \times 0.681$$

$$2.010 \le \alpha \le 5.790$$

$t(0.05,\nu_e)$ は両側確率0.05の t 値

$= \mathrm{TINV}(0.05,4) = 2.776$ 両側確率

$= \mathrm{T.INV}(0.975,4) = 2.776$ 下側確率

$= \mathrm{T.INV.2T}(0.05,4) = 2.776$ 両側確率

表示4.3.6 ソルバーの解

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

a	3.900
b	0.620
S	2.780

表示4.3.8 LINEST関数の結果

	x	const	
回帰係数	0.620	3.900	切片
その標準誤差	0.118	0.681	その標準誤差
寄与率	0.874	0.834	残差標準偏差
F 比	27.655	4	残差自由度
回帰平方和	19.220	2.780	残差平方和

β、αの仮説検定と区間推定

●LINEST 関数の出力から算出

表示4.3.3のデータの解析結果

前節 §4.3 参照

$$t = \frac{b}{s.e.[b]} = \frac{0.620}{0.118} = 5.259$$

=TDIST(5.259, 4, 2) = 0.006

$$t = \frac{a}{s.e.[a]} = \frac{3.900}{0.681} = 5.727$$

=TDIST(5.727, 4, 2) = 0.005

$$0.293 \leq \beta \leq 0.947$$

$$2.010 \leq \alpha \leq 5.790$$

表示4.3.8（p.235）

他の場所にコピーして使える
（x とy の範囲を修正）

	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)

$ マークによる
絶対参照がない

p.240

（4）予測値と y の区間推定

回帰モデル

p.240

●回帰モデル

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

$$y = \eta + \varepsilon \qquad (\text{y は観測値})$$

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

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

● y の母平均 η の点推定と区間推定

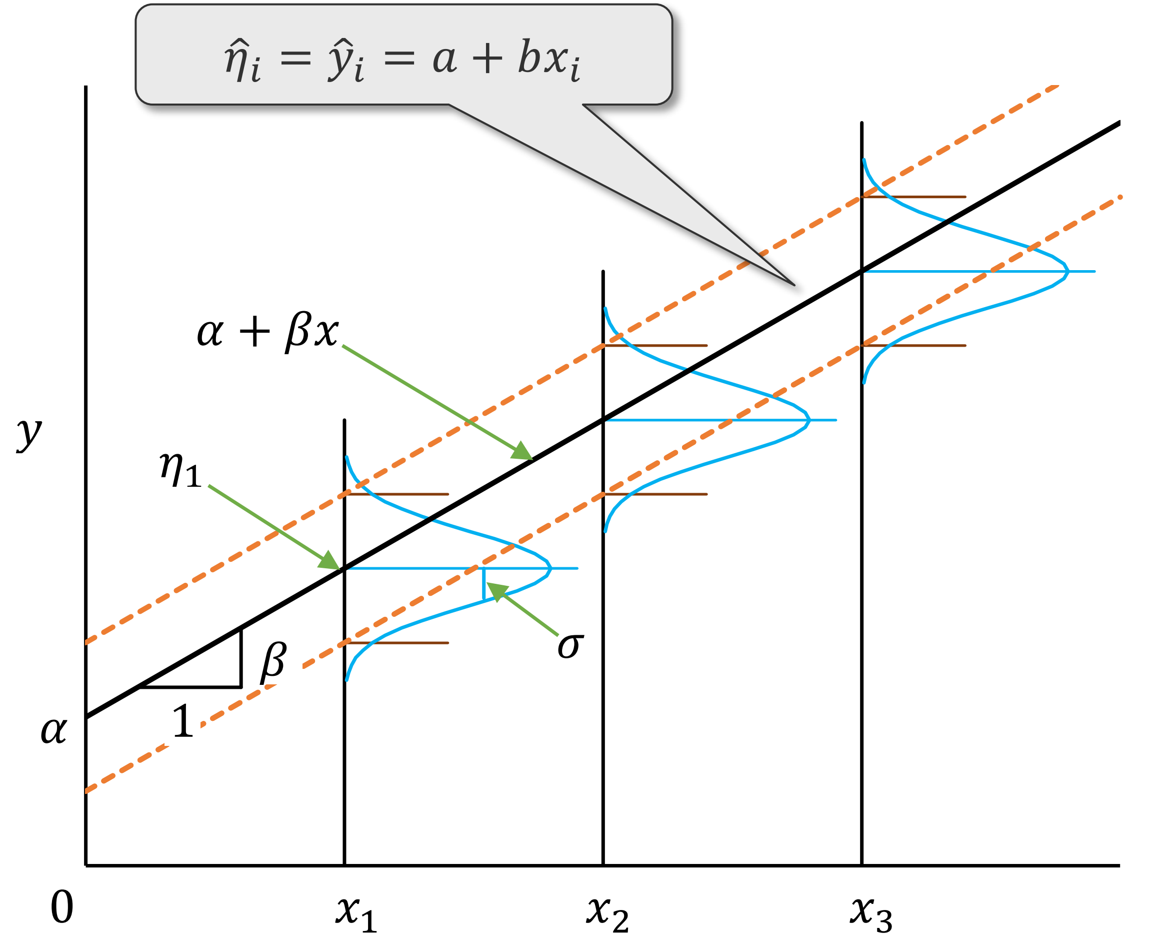
$$\hat{\eta}_i = \hat{y}_i = a + bx_i$$

a と b の標準誤差から

x がある特定の値 x_i をとったときの

母平均 $\hat{\eta}_i$ の区間推定を行う

→　母平均 $\hat{\eta}_i$ の分散が必要



表示 4.3.2　回帰モデル

予測値と y の区間推定

p.241

●母平均 $\hat{\eta}$ の分散

$$\hat{\eta} = \hat{y} = a + bx = \bar{y} + b(x - \bar{x})$$

$$\mathrm{V}[\hat{\eta}] = \mathrm{V}[\bar{y}] + \mathrm{V}[b] \times (x - \bar{x})^2$$

$$\mathrm{V}[\bar{y}] = \frac{1}{n}\sigma^2$$

$$\mathrm{V}[b] = \frac{1}{S_{xx}}\sigma^2$$

$$\mathrm{V}[\hat{\eta}] = \frac{1}{n}\sigma^2 + (x - \bar{x})^2 \frac{1}{S_{xx}}\sigma^2$$

$$= \left(\frac{1}{n} + \frac{(x - \bar{x})^2}{S_{xx}}\right)\sigma^2 \qquad (4.4.3)$$

a, b は独立ではないが、$\bar{y}$ と b は独立（無相関）
回帰分析の場合、x は誤差をもたない
x はある特定の値 → 「分散の加法性」を利用

$z = ax \quad \mathrm{V}[z] = a^2\mathrm{V}[x]$ （1.3.4）§1.3 p.23

$\sigma_{\bar{y}} = \frac{\sigma_y}{\sqrt{n}}$ （1.3.8） §1.3 p.24

$s.e.[b] = \frac{\sigma}{\sqrt{S_{xx}}}$ （4.4.1） p.238

予測値と y の区間推定

p.241

●母平均 $\hat{\eta}$ の95%信頼区間

Ve ：残差平均平方　s²　p.233 で推定
（誤差分散を残差分散で推定）

$$V[\hat{\eta}] = \left(\frac{1}{n} + \frac{(x-\bar{x})^2}{S_{xx}}\right)\sigma^2 \qquad (4.4.3)$$

$$\eta \sim y \pm t(0.05, \nu_e)\sqrt{\left(\frac{1}{n} + \frac{(x-\bar{x})^2}{S_{xx}}\right)V_e} = a + bx \pm t(0.05, \nu_e)\sqrt{\left(\frac{1}{n} + \frac{(x-\bar{x})^2}{S_{xx}}\right)V_e} \qquad (4.4.4)$$

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

n が大きくなると信頼区間は狭くなる

x　の範囲が広くなると信頼区間は狭くなる

平均値に近い x の場合ほど、信頼区間は狭くなる（x の平均値の所で最小）

予測値と y の区間推定

●母平均の95%信頼区間

$$\eta \sim a + bx \pm t\,(0.05, \nu)\sqrt{\left(\frac{1}{n} + \frac{(x-\bar{x})^2}{S_{xx}}\right)V_e} \qquad (4.4.4)$$

$$= 3.90 + 0.620x \pm t\,(0.05, 4)\sqrt{\left(\frac{1}{6} + \frac{(x-5.0)^2}{50.00}\right) \times 0.834^2}$$

表示4.3.6 ソルバーの解

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

a	3.900
b	0.620
S	2.780

表示4.3.8 LINEST関数の結果

	x	const	
回帰係数	0.620	3.900	
その標準誤差	0.118	0.681	
寄与率	0.874	0.834	残差標準偏差
*F*比	27.655	4	残差自由度
回帰平方和	19.220	2.780	残差平方和

予測値と y の区間推定

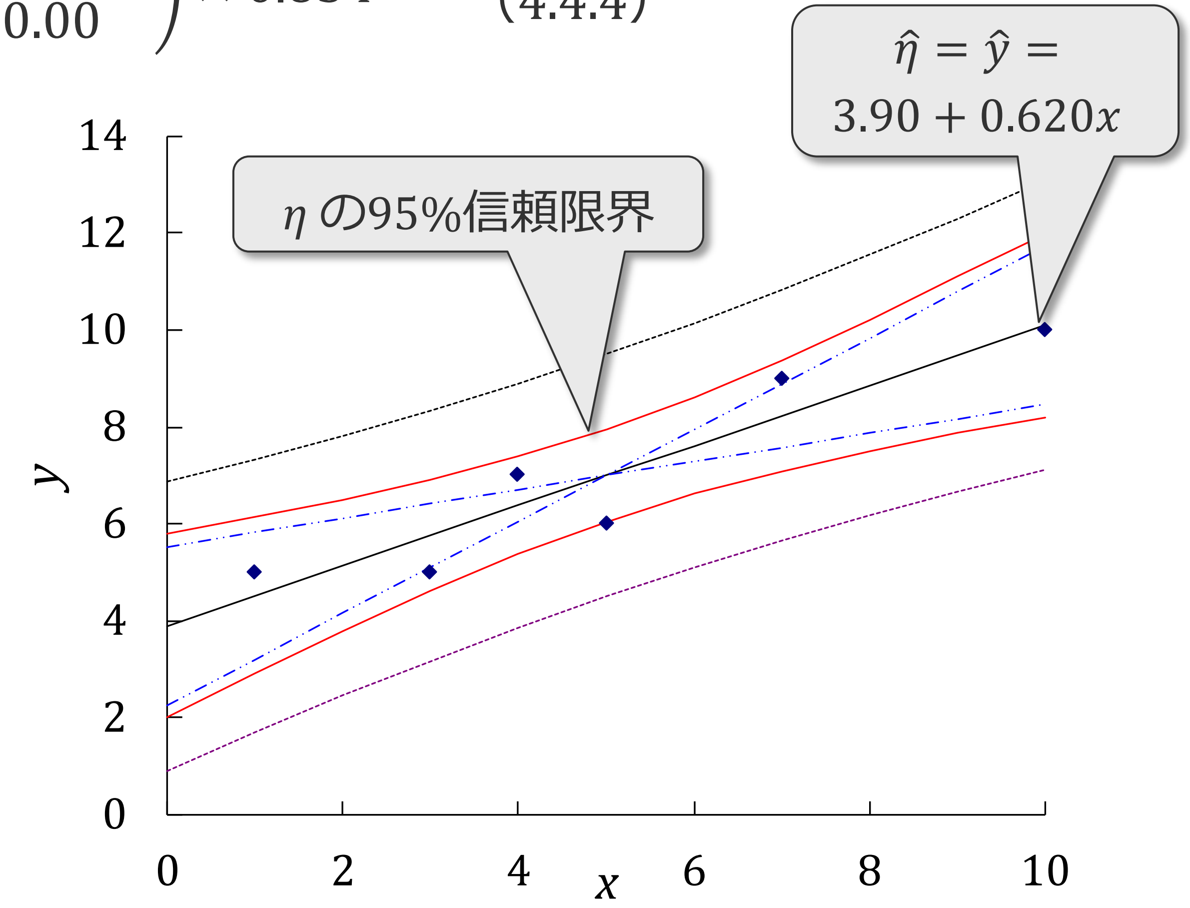
p.241

●母平均の95%信頼区間

$$\eta \sim 3.90 + 0.620x \pm t\,(0.05, 4)\sqrt{\left(\frac{1}{6} + \frac{(x-5.0)^2}{50.00}\right) \times 0.834^2} \qquad (4.4.4)$$

表示4.4.3 区間推定

x	y	y-hat	η の信頼限界 下限	η の信頼限界 上限	y の信頼限界 下限	y の信頼限界 上限
0		3.90	2.01	5.79	0.91	6.89
1	5	4.52	2.91	6.13	1.70	7.34
2		5.14	3.78	6.50	2.45	7.83
3	5	5.76	4.61	6.91	3.18	8.34
4	7	6.38	5.38	7.38	3.86	8.90
5	6	7.00	6.06	7.94	4.50	9.50
6		7.62	6.62	8.62	5.10	10.14
7	9	8.24	7.09	9.39	5.66	10.82
8		8.86	7.50	10.22	6.17	11.55
9		9.48	7.87	11.09	6.66	12.30
10	10	10.10	8.21	11.99	7.11	13.09



予測値と y の区間推定

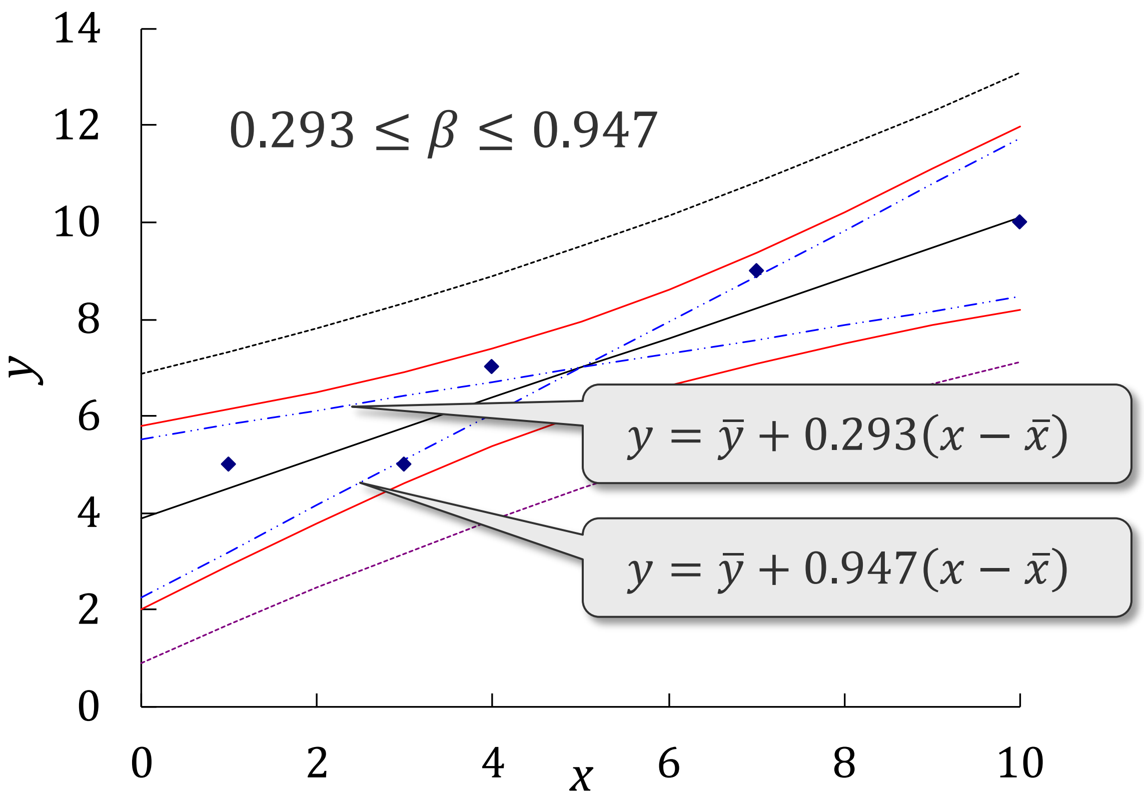
p.241

●母平均の95%信頼区間

$$\eta \sim 3.90 + 0.620x \pm t\,(0.05, 4)\sqrt{\left(\frac{1}{6} + \frac{(x - 5.0)^2}{50.00}\right) \times 0.834^2} \qquad (4.4.4)$$

表示4.4.3 区間推定

			η の信頼限界		y の信頼限界	
x	y	y -hat	下限	上限	下限	上限
0		3.90	2.01	5.79	0.91	6.89
1	5	4.52	2.91	6.13	1.70	7.34
2		5.14	3.78	6.50	2.45	7.83
3	5	5.76	4.61	6.91	3.18	8.34
4	7	6.38	5.38	7.38	3.86	8.90
5	6	7.00	6.06	7.94	4.50	9.50
6		7.62	6.62	8.62	5.10	10.14
7	9	8.24	7.09	9.39	5.66	10.82
8		8.86	7.50	10.22	6.17	11.55
9		9.48	7.87	11.09	6.66	12.30
10	10	10.10	8.21	11.99	7.11	13.09



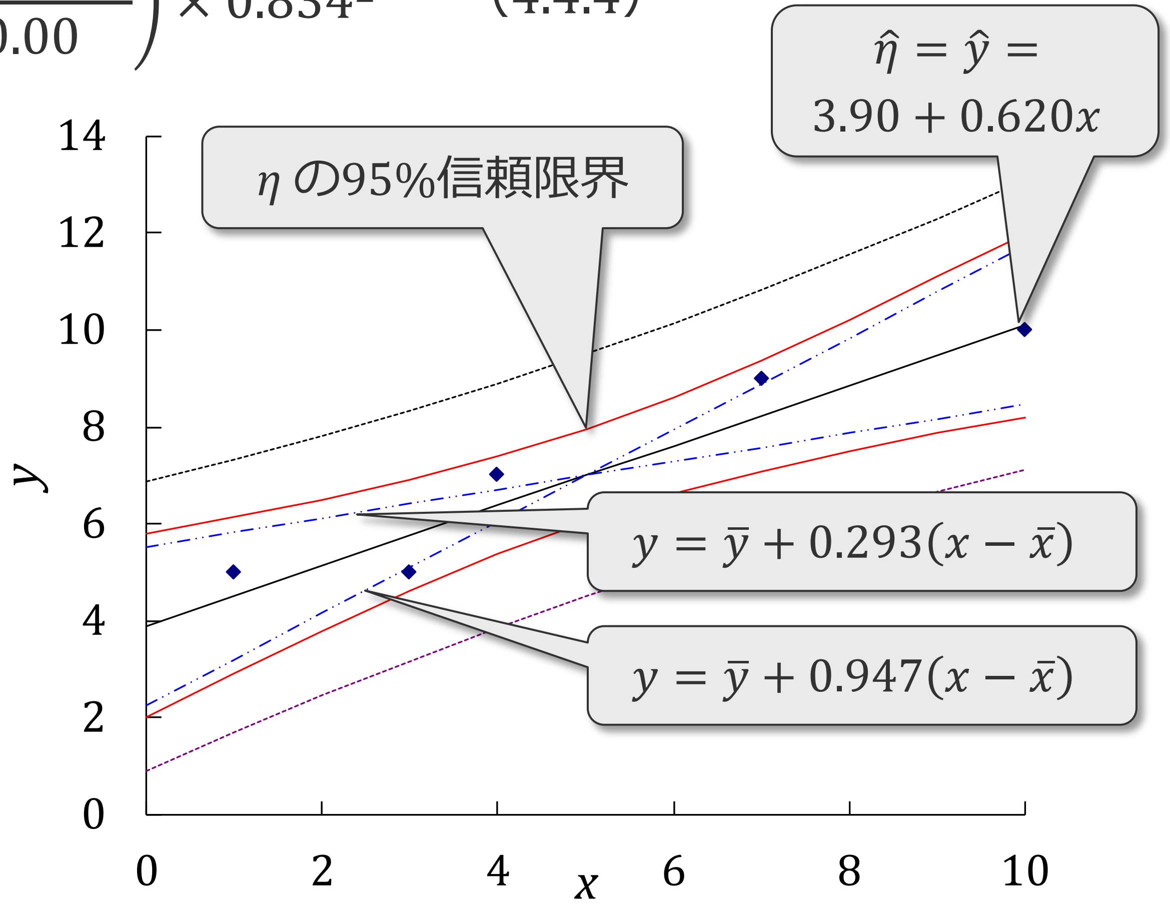
予測値と y の区間推定

p.241

●母平均の95%信頼区間

$$\eta \sim 3.90 + 0.620x \pm t\,(0.05, 4)\sqrt{\left(\frac{1}{6}+\frac{(x-5.0)^2}{50.00}\right)\times 0.834^2} \qquad (4.4.4)$$

x が $\bar{x}$ のところで区間推定の幅は最小
x が $\bar{x}$ から離れと、区間推定の幅は広くなる



母平均の95%区間推定ではなく、
個々の観測値の95%信頼区間は？

予測値と y の区間推定

p.241

●個々の観測値の95%信頼区間

$$y = \eta + \varepsilon \qquad \varepsilon \sim \mathrm{N}(0, \sigma^2) \qquad (4.3.3)$$

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

$$\mathrm{V}[\hat{\eta}] = \left(\frac{1}{n} + \frac{(x - \bar{x})^2}{S_{xx}} \right) \sigma^2 \qquad (4.4.3)$$

$$\mathrm{V}[y] = \mathrm{V}[\hat{\eta}] + V[\varepsilon] = \left(\frac{1}{n} + \frac{(x - \bar{x})^2}{S_{xx}} \right) \sigma^2 + \sigma^2 = \left(1 + \frac{1}{n} + \frac{(x - \bar{x})^2}{S_{xx}} \right) \sigma^2 \qquad (4.4.5)$$

$$y \sim (a + bx) \pm t(0.05, \nu_e) \sqrt{\left(1 + \frac{1}{n} + \frac{(x - \bar{x})^2}{S_{xx}} \right) V_e} \qquad (4.4.6)$$

ν：自由度
ニュー

ルートの中に１があるから、n が増えても変化は大きくない
n が大きくなると，ルートの中は Ve に近付き、t 分布は標準正規分布に近付く
→　双曲線は直線に近づく　$y \sim (a + bx) \pm 1.96s$　（s：残差標準偏差）

予測値と y の区間推定

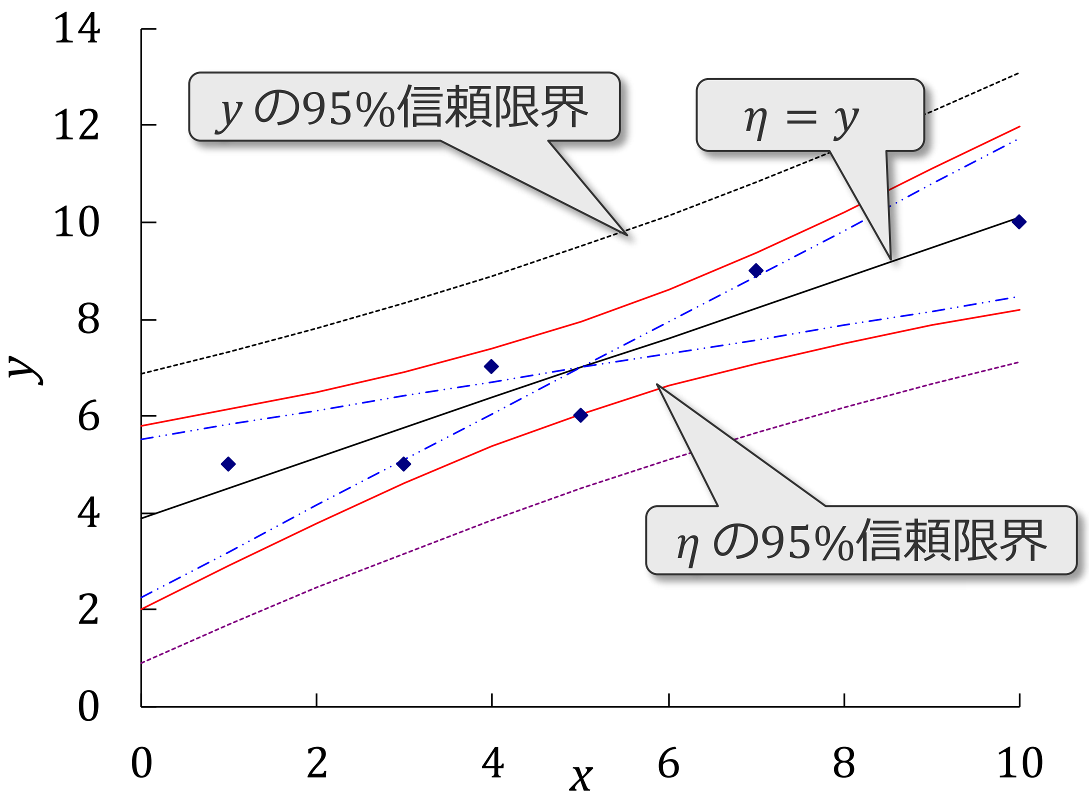
p.241

●個々の観測値の95%信頼区間

$$y \sim 3.90 + 0.620x \pm t\,(0.05, 4)\sqrt{\left(1 + \frac{1}{6} + \frac{(x - 5.0)^2}{50.00}\right) \times 0.834^2} \qquad (4.4.6)$$

表示4.4.3 区間推定

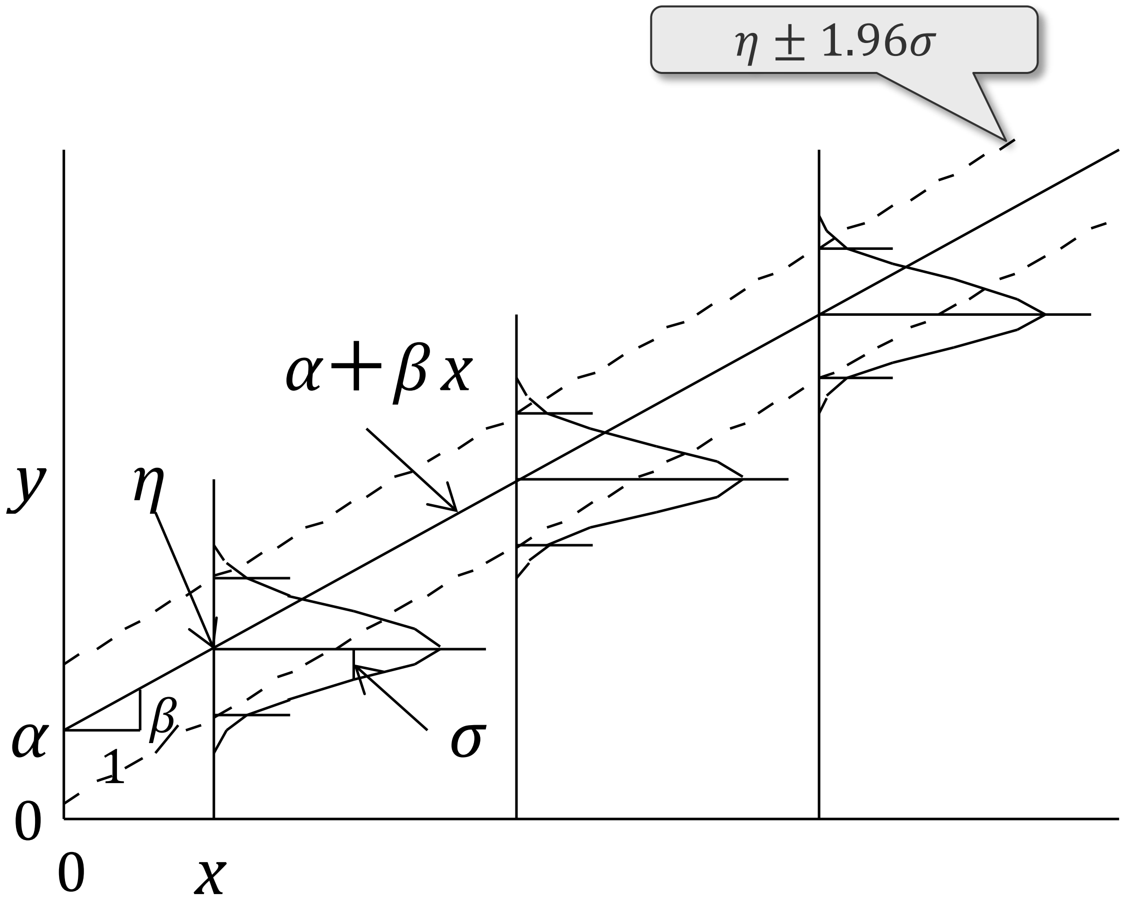
			η の信頼限界		y の信頼限界	
x	y	y-hat	下限	上限	下限	上限
0		3.90	2.01	5.79	0.91	6.89
1	5	4.52	2.91	6.13	1.70	7.34
2		5.14	3.78	6.50	2.45	7.83
3	5	5.76	4.61	6.91	3.18	8.34
4	7	6.38	5.38	7.38	3.86	8.90
5	6	7.00	6.06	7.94	4.50	9.50
6		7.62	6.62	8.62	5.10	10.14
7	9	8.24	7.09	9.39	5.66	10.82
8		8.86	7.50	10.22	6.17	11.55
9		9.48	7.87	11.09	6.66	12.30
10	10	10.10	8.21	11.99	7.11	13.09

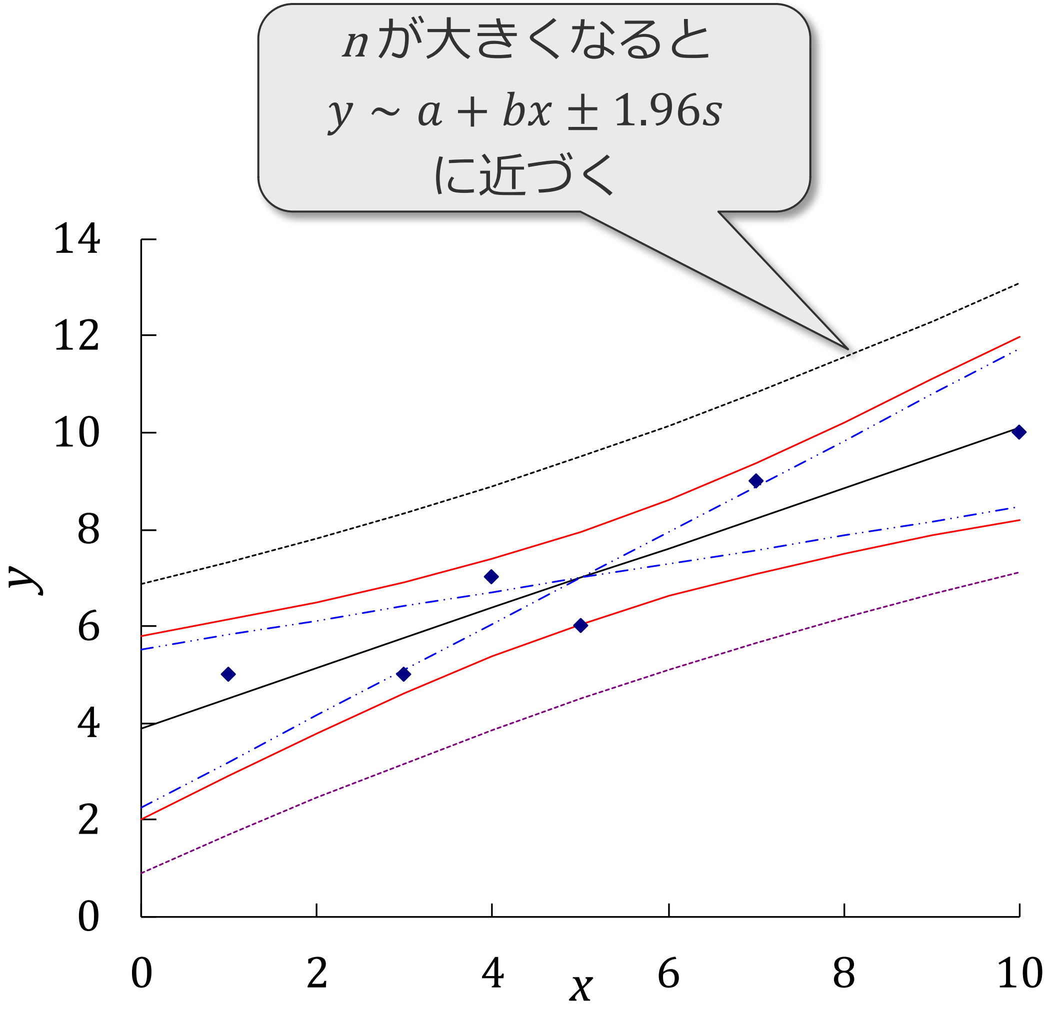


予測値と y の区間推定

p.241

●個々の観測値の95%信頼区間





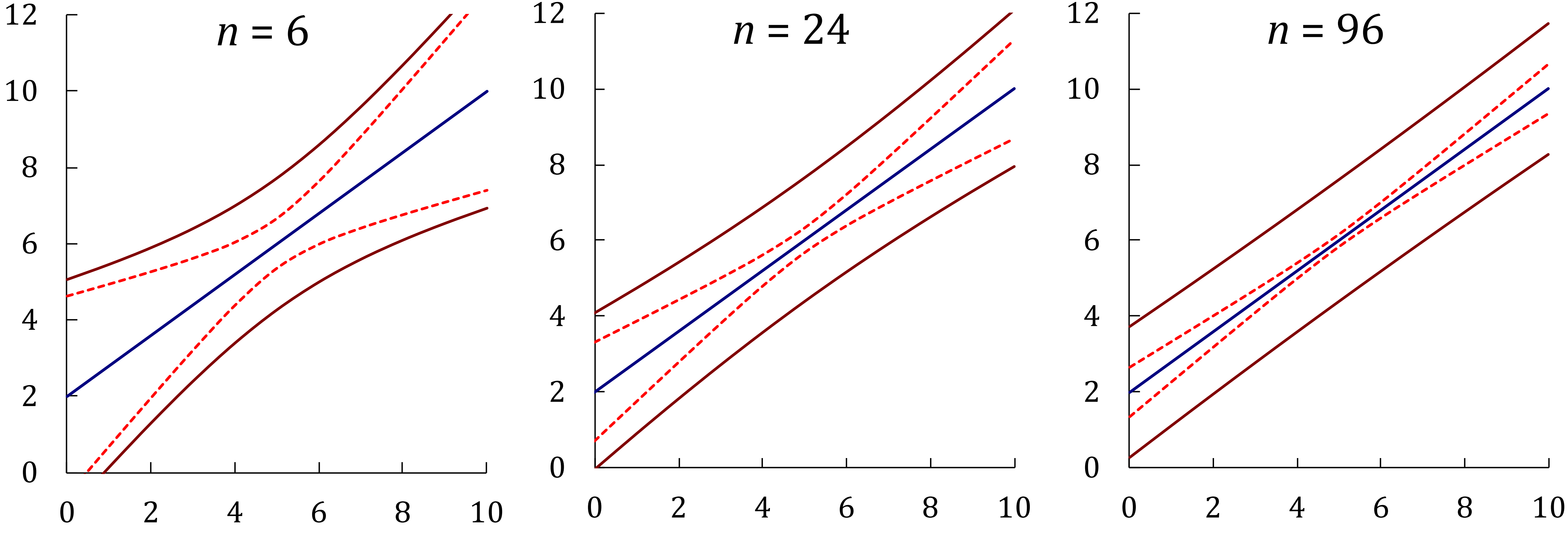
予測値と y の区間推定

p.242

●nと信頼区間の関係

n が４倍になるとη の95%信頼区間の幅はほぼ半分
外側の y の95%信頼区間の幅の変化は少ない

表示4.4.4 nによる信頼区間の変化



予測値と y の区間推定

p.242

- 推定と予測

推定：x がある特定の値のときの y の母平均 η の95%区間推定（母平均の信頼区間）

「回帰の信頼区間（JMP）」

$$\eta \sim a + bx \pm t\,(0.05, \nu)\sqrt{\left(\frac{1}{n} + \frac{(x-\bar{x})^2}{S_{xx}}\right)V_e} \qquad (4.4.4)$$

予測：x がある特定の値のときの個々の観測値 y の95%区間推定（観測値の予測区間）

「個々の値に対する信頼区間（JMP）」

$$y \sim a + bx \pm t(0.05, \nu_e)\sqrt{\left(1 + \frac{1}{n} + \frac{(x-\bar{x})^2}{S_{xx}}\right)V_e} \qquad (4.4.6)$$

ある x に対する y の母数の推定よりも、個々の対象の y の 値を予測する方が多く用いられる

p.242

（5）逆推定

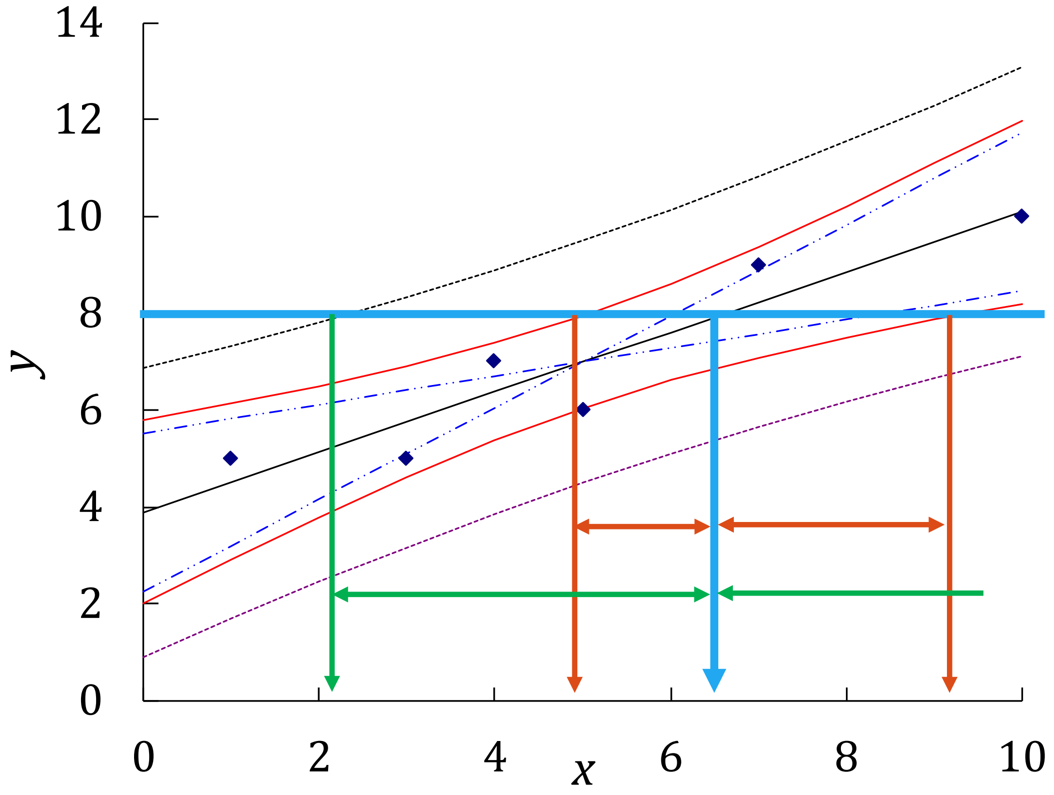
逆推定

p.242

●逆推定とは

回帰直線から、y の期待値η が特定の値になるx の点推定と信頼区間を求める

表示4.4.3 区間推定（改変）



y の期待値 η が 8 になるx の点推定と区間推定

y = 8 の水平線上の交点で、xを読み取る
信頼区間は等間隔ではない

説明変数xと目的変数yを交換して回帰させ、
x から y を求めると、上記とは異なった結果を
得るので、この方法は適切ではない

逆推定

p.242

- 計算による逆推定

x の点推定：回帰式を変形してx が求められる

$$y = a + bx = 3.9 + 0.62x$$

$$x = \frac{y - a}{b} = \frac{8 - 3.9}{0.62} = 6.61$$

x の信頼区間：計算は難しい

$$y = a + bx \pm t\ (0.05, \nu)\sqrt{\left(\frac{1}{n} + \frac{(x - \bar{x})^2}{S_{xx}}\right)V_e}$$

$$8 = 3.9 + 0.62x \pm 2.776 \times \sqrt{\left(\frac{1}{6} + \frac{(x - 5.0)^2}{50.0}\right) \times 0.834^2}$$

↑　ここら x を計算するのは難しい

逆推定

●Excel ゴールシークによる逆推定

ゴールシークについては、[§3.6](#) (4) p.163

前項(4) で区間推定を行った「表示4.4.3」を下方の別の場所にコピー（上から 7 行分）
これ以降、S29以降にコピーした状態で説明（上書き）

表示4.4.3　区間推定

	S	T	U	V	W	X	Y
15				η の信頼限界		y の信頼限界	
16	x	y	y-hat	下限	上限	下限	上限
17	0		3.90	2.01	5.79	0.91	6.89
18	1	5	4.52	2.91	6.13	1.70	7.34
19	2		5.14	3.78	6.50	2.45	7.83
20	3	5	5.76	4.61	6.91	3.18	8.34
21	4	7	6.38	[illegible]	7.38	3.86	8.90

表示4.4.5　逆推定

テキストではセル S29 以降に表示4.4.5 がある
ここに上書きして実習を行う

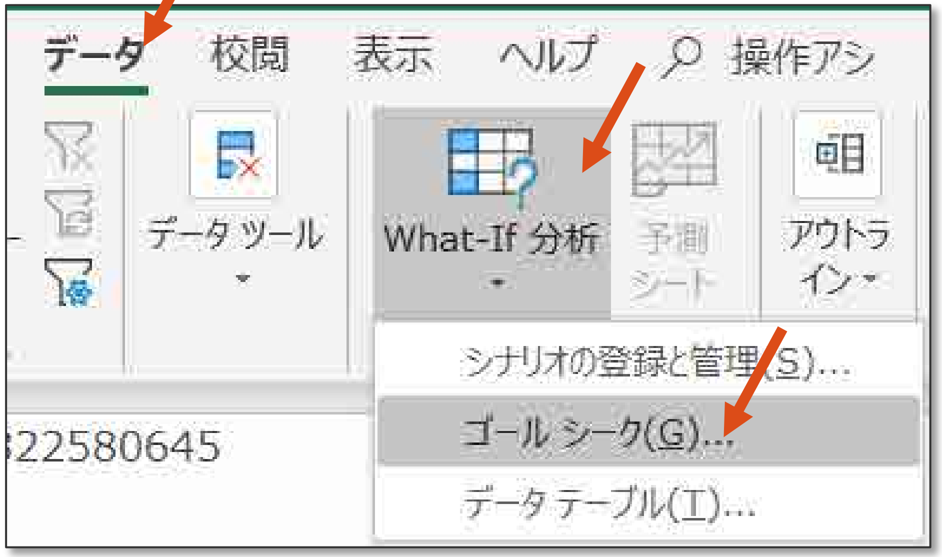
	S	T	U	[illegible]	W	X	Y
29				η の信頼限界		y の信頼限界	
30	x	y	y-hat	下限	上限	下限	上限
31	0		3.90	2.01	5.79	0.91	6.89
32	1	5	4.52	2.91	6.13	1.70	7.34
33	2		5.14	3.78	6.50	2.45	7.83
34	3	5	5.76	4.61	6.91	3.18	8.34
35	4	7	6.38	5.38	7.38	3.86	8.90

上から 7 行分を使用

逆推定

p.243

● Excel ゴールシークによる逆推定

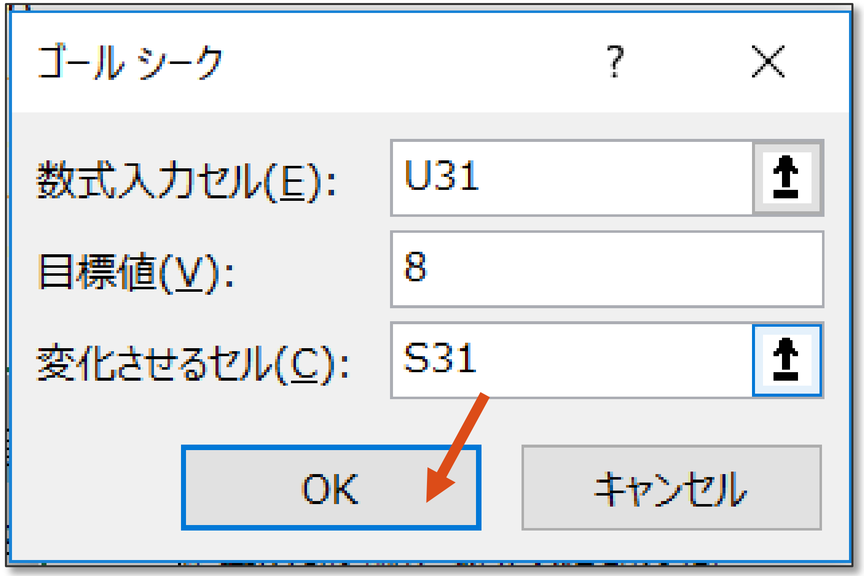


変化させるセル

数式入力セル

表示4.4.5
逆推定

	S	T	U	V	W	X	Y
29				η の信頼限界		y の信頼限界	
30	x	y	y-hat	下限	上限	下限	上限
31	6.61		**8.00**	6.92	9.08	5.44	10.56
32	9.38		9.72	**8.00**	11.44	6.84	12.60
33	5.09		7.06	6.11	**8.00**	4.55	9.56
34	12.13		11.42	8.90	13.94	**8.00**	14.84
35	2.34		5.35	4.07	6.64	2.71	**8.00**



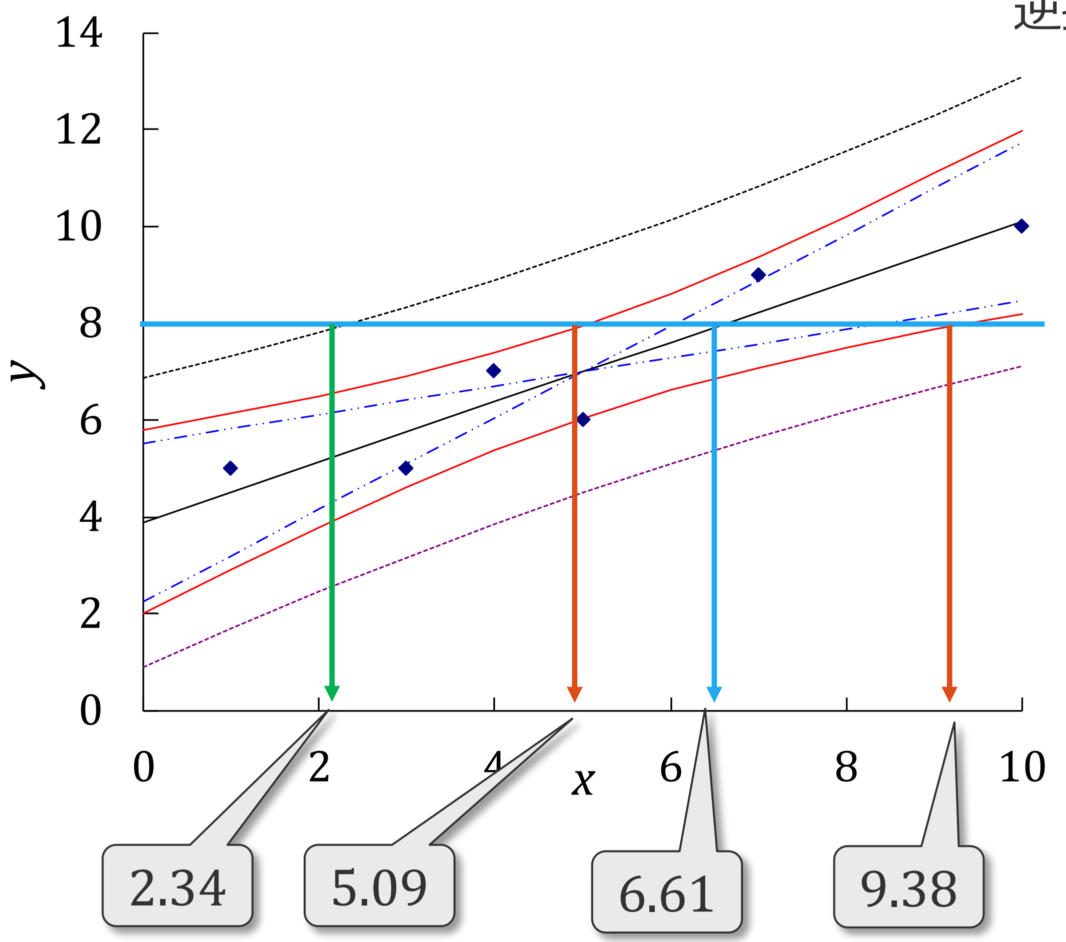
5つの組み合わせ

変化させるセル	数式入力セル
S31	U31
S32	V32
S33	W33
S34	X34
S35	Y35

逆推定

●Excel ゴールシークによる逆推定

表示4.4.5
逆推定



	S	T	U	V	W	X	Y
29				η の信頼限界		y の信頼限界	
30	x	y	y-hat	下限	上限	下限	上限
31	6.61		**8.00**	6.92	9.08	5.44	10.56
32	9.38		9.72	**8.00**	11.44	6.84	12.60
33	5.09		7.06	6.11	**8.00**	4.55	9.56
34	12.13		11.42	8.90	13.94	**8.00**	14.84
35	2.34		5.35	4.07	6.64	2.71	**8.00**

[母平均で考える]

η=8 になる x の点推定 x =6.61

η=8 になる x の95%信頼区間 x =[5.09, 9.38]

[観測値で考える]

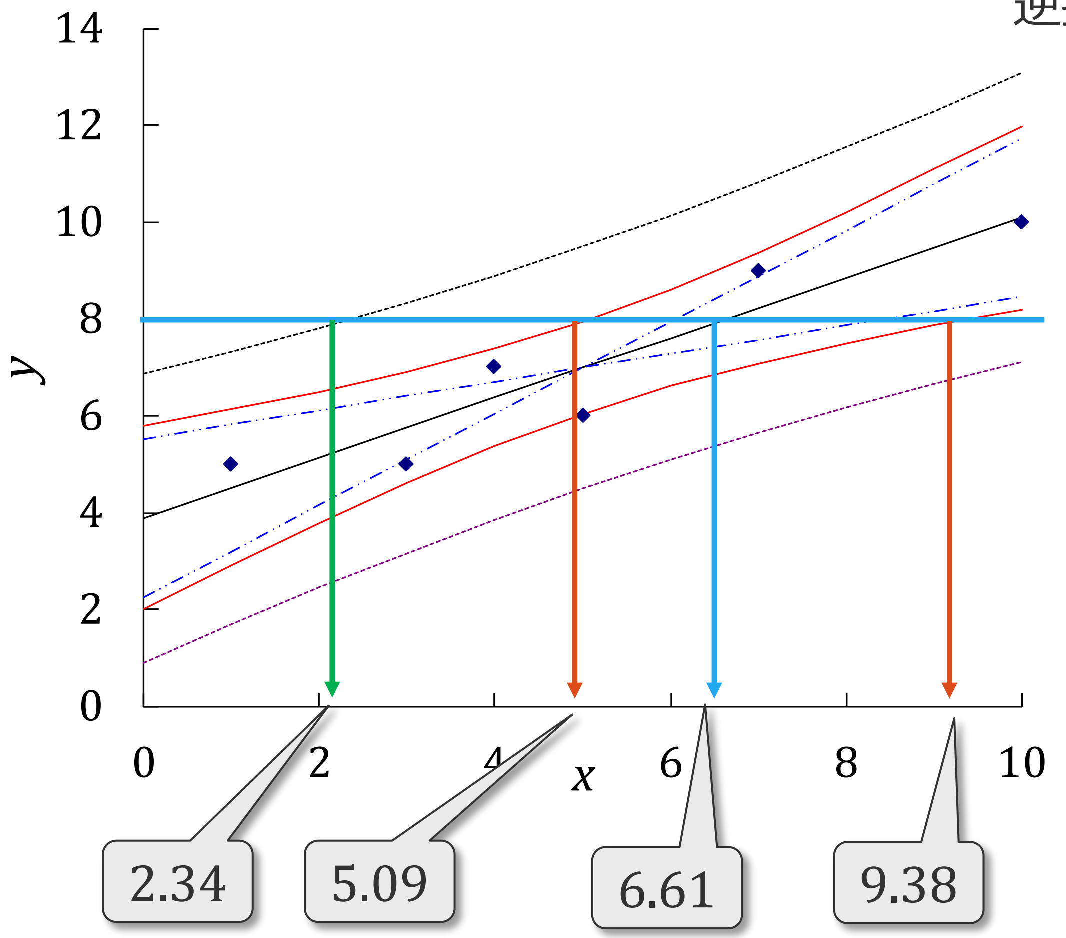
y=8 になる x の点推定 x =6.61

y=8 になる x の95%信頼区間 x =[2.34, 12.13]

逆推定

●Excel ゴールシークによる逆推定

表示4.4.5
逆推定



	S	T	U	V	W	X	Y
29				η の信頼限界		y の信頼限界	
30	x	y	y -hat	下限	上限	下限	上限
31	6.61		**8.00**	6.92	9.08	5.44	10.56
32	9.38		9.72	**8.00**	11.44	6.84	12.60
33	5.09		7.06	6.11	**8.00**	4.55	9.56
34	12.13		11.42	8.90	13.94	**8.00**	14.84
35	2.34		5.35	4.07	6.64	2.71	**8.00**

[化学分析の検量線]

濃度x が既知の検体について分析

分析計の表示値y を求めて回帰直線を得る

検体の表示値y から検体の濃度x を予測したい

→ 個々の検体の濃度を知りたい

→　外側 y の信頼区間を使う

p.244

（6）分散分析

分散分析

p.244

●x によって y が変化するかを検定

H_0：$\beta = 0$　の仮説検定　・・・・・・・・・ t 検定、§4.4(3)、p.240

H_0：回帰の平均平方=誤差の平均平方・・・分散分析による検定（F検定）

●平方和の分解（p.232）

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

$$22.00 = 19.22 + 2.78$$

●自由度の分解（p.233）

$$\nu_T \quad = \nu_R + \quad \nu_e$$

$$(6-1) = 1 \quad + (6-2) \qquad (4.3.16)$$

ν：自由度
ニュー

●分散分析表

表示4.4.6
分散分析表

要因	平方和（S）	自由度（ν）	平均平方（V）	F 比	p 値
回帰（R）	19.22	1	19.220	27.65	0.006
誤差（e）	2.78	4	0.695	1.00	
全体（T）	22.00	5			

分散分析

p.244

● F 検定

平均平方 = 平方和 ／ 自由度

y が x によって変化しない（H_0： β =0）とき、回帰の平均平方と誤差の平均平方が等しい

（第２部）

F 比＝ 回帰の平均平方 ／ 誤差の平均平方

帰無仮説が正しいという前提で、F比はF分布（§3.3.(1) p.141）に従う

=FDIST(27.65, 1, 4)=0.006 ＜0.05　帰無仮説は棄却される

誤差の行の「1.00」は、分母に用いたことを示す（本書独特）

$$F = \frac{19.220}{0.695} = 27.65$$

表示4.4.6
分散分析表

要因	平方和（S）	自由度（ν）	平均平方（V）	F 比	p 値
回帰（R）	19.22	1	19.220	27.65	0.006
誤差（e）	2.78	4	0.695	1.00	
全体（T）	22.00	5			

分母を明示
（本書独特）

分散分析

p.244

- 「分散分析の F 検定」と「回帰係数 b の t 検定」

  両者の p 値は等しい

  t=5.249　　=T.DIST.2T(5.259, 4)=0.006　（帰無仮説　H_0：$\beta=0$　の検定、§4.4(3) p.240）

  自由度 ν の t 分布の2乗は、自由度$(1,\nu)$の F 分布に一致（§3.8 補遺(2) p.187）

  $5.259^2=27.65$

  両者は同じ検定になる

  分散分析については、第２部で詳しく説明

表示4.4.6 分散分析表

要因	平方和（S）	自由度（ν）	平均平方（V）	F 比	p 値
回帰（R）	19.22	1	19.220	27.65	0.006
誤差（e）	2.78	4	0.695	1.00	
全体（T）	22.00	5			

分散分析

- 寄与率（決定係数、 R^2）、自由度調整寄与率

寄与率：平方和 S_T の中で S_Rで説明できる割合（S_eはx の変化で説明できない部分） §4.3

（回帰で説明できる部分の大きさ）

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

寄与率（決定係数）

$$R^2 = \frac{S_R}{S_T} = \frac{S_T - S_e}{S_T} = 1 - \frac{S_e}{S_T}$$

誤差率

平方和を
自由度で割って
平均平方に置き換え

$$R^{*2} = 1 - \frac{S_e/\nu_e}{S_T/\nu_T} = \frac{S_e/(n-1)}{S_T/(n-1)} = 1 - \frac{V_e}{V_T} = 1 - \frac{2.78/4}{22.00/5} = 0.842$$

・・・自由度調整寄与率
（p. 232、p.257）

表示4.4.6
分散分析表

要因	平方和（S）	自由度（ν）	平均平方（V）	F 比	p 値
回帰（R）	19.22	1	19.220	27.65	0.006
誤差（e）	2.78	4	0.695	1.00	
全体（T）	22.00	5			

（7） JMPによる解析

JMP ［二変量の関係］

p.245

●JMPファイルの読み込みと表示

JMP ファイル「4-相関3.jmp」を読み込み

	x	y
1	1	5
2	3	5
3	4	7
4	5	6
5	7	9
6	10	10

●データと目的

表示4.3.3、表示4.3.4、表示4.3.6、表示4.4.3のデータ

６組の x と y のデータ

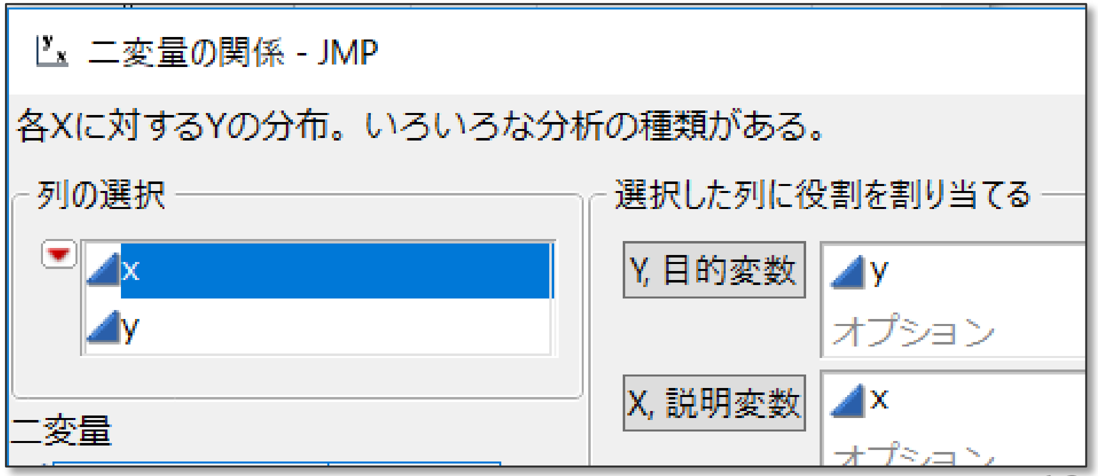
単回帰分析を行う

●二変量の関係

［分析］＞［二変量の関係］

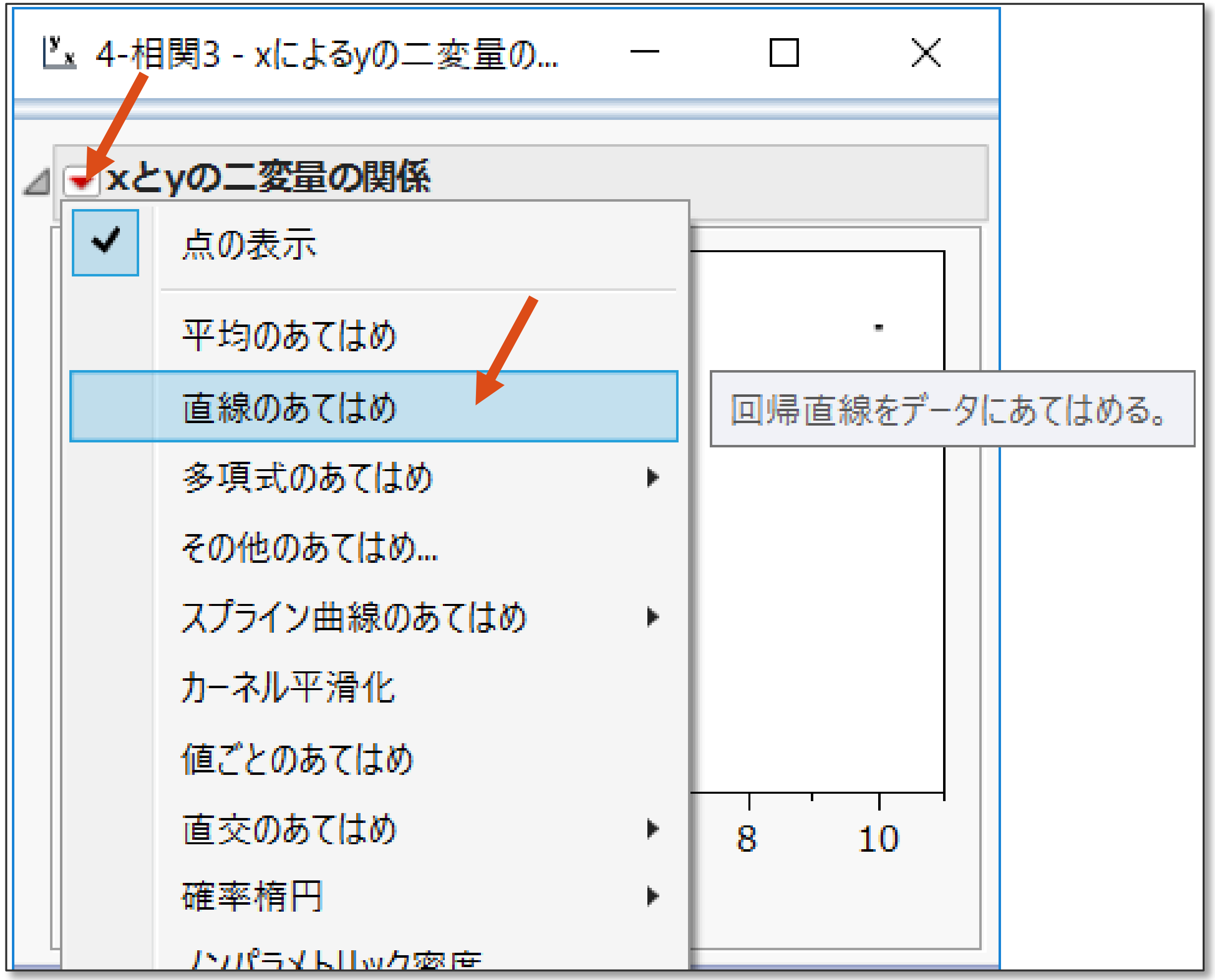
［Y、目的変数］：y

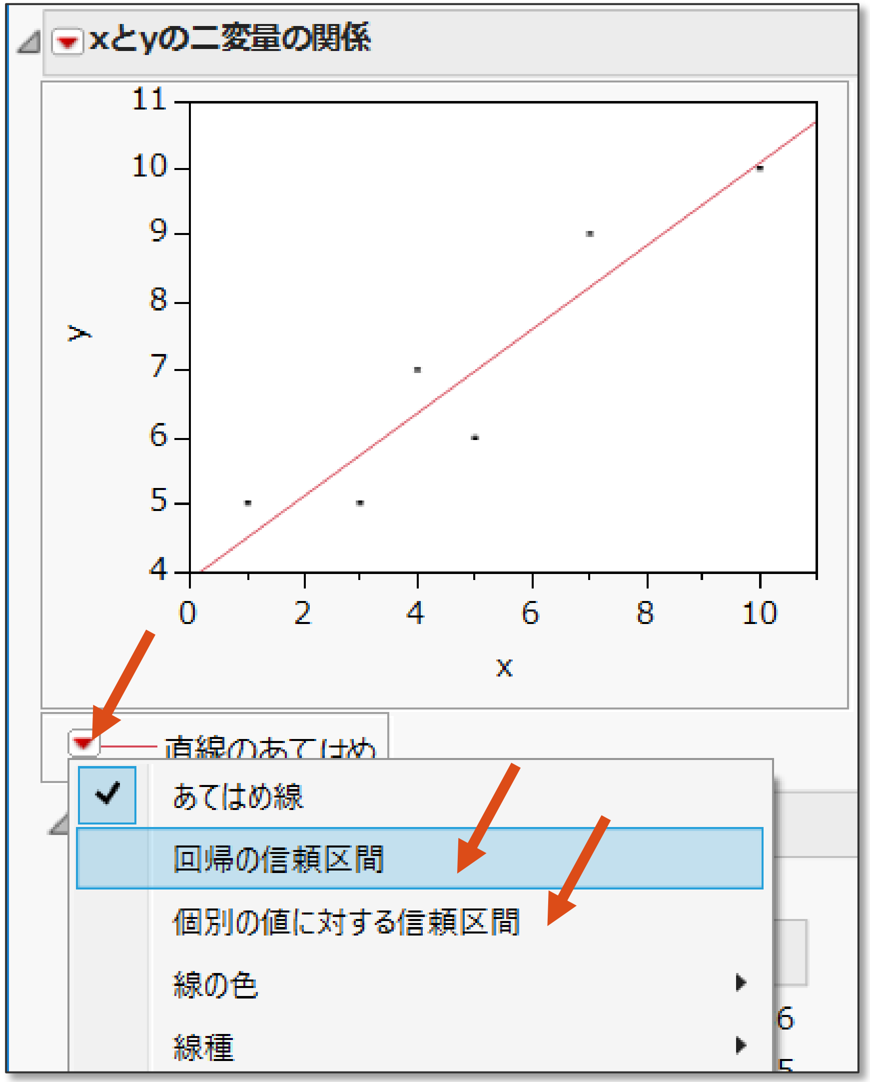
［X、説明変数］：x



JMP ［二変量の関係］

●単回帰分析（直線のあてはめ）

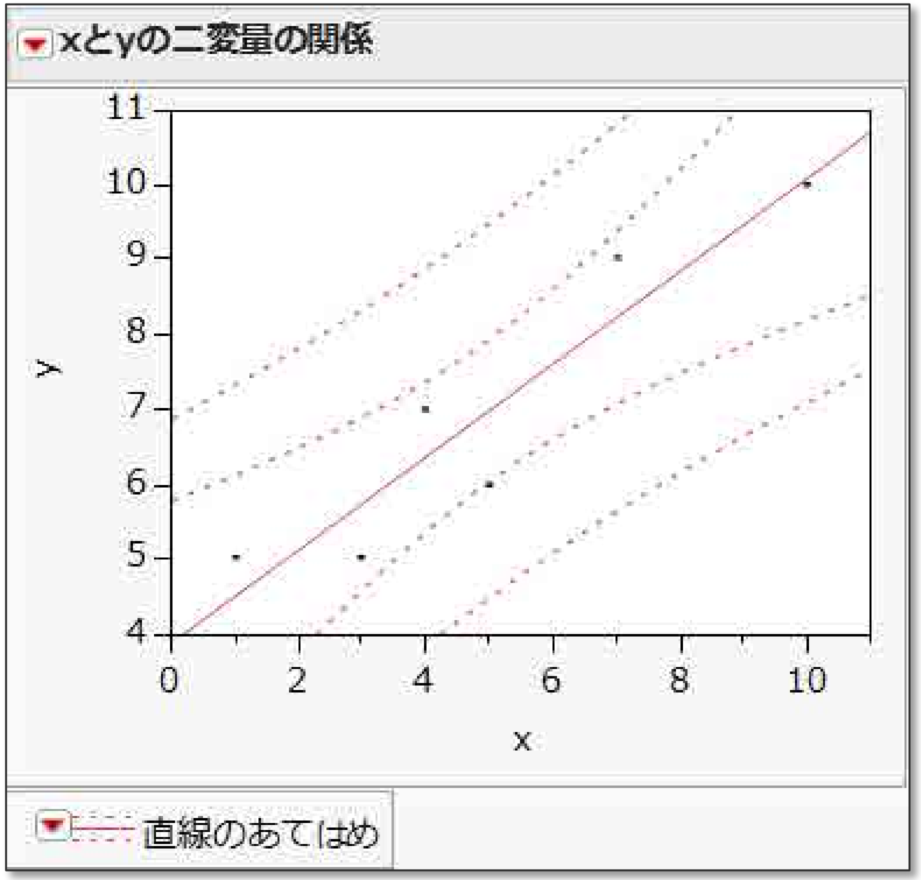




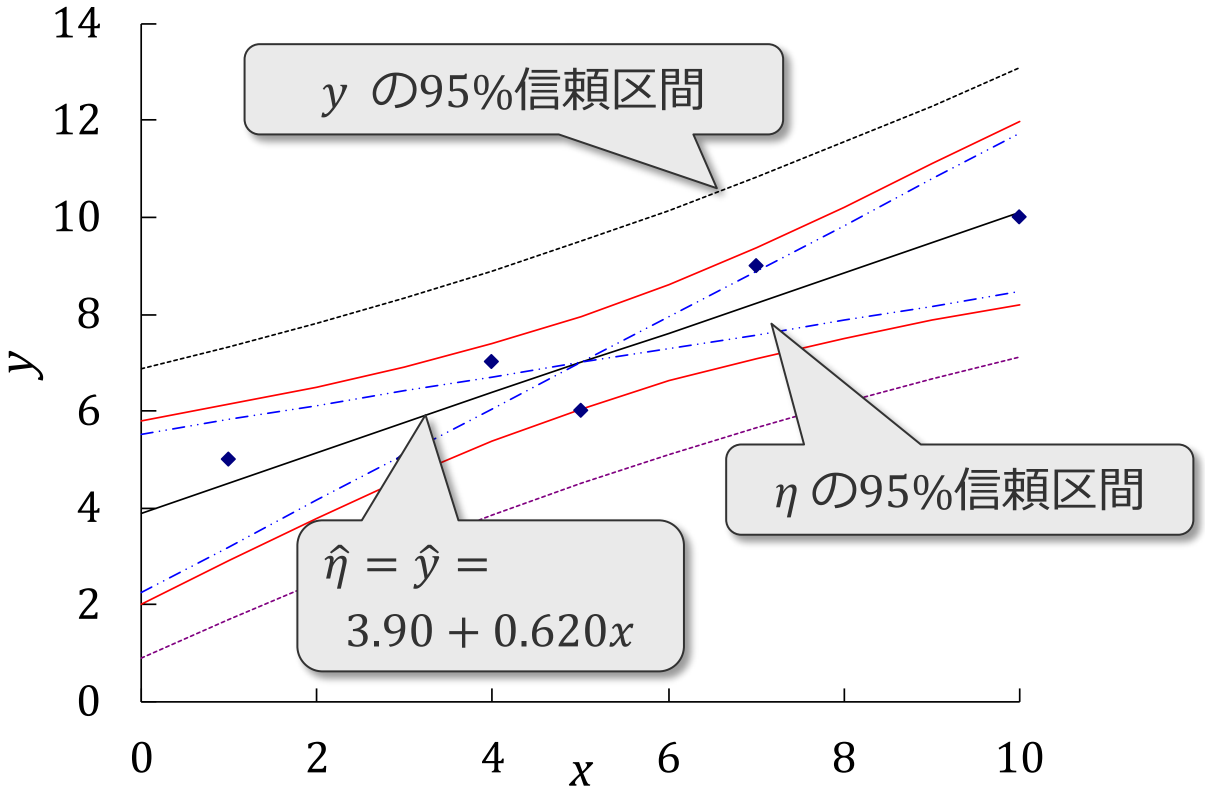
JMP ［二変量の関係］

●信頼区間

表示4.4.7 JMP「二変量の関係」による回帰式



表示4.4.3 区間推定



JMP ［二変量の関係］

p.246

●解析結果

表示4.4.7
JMP「二変量の関係」
による回帰式

直線のあてはめ

y = 3.9 + 0.62*x

あてはめの要約

R2乗	0.873636
自由度調整R2乗	0.842045
誤差の標準偏差(RMSE)	0.833667
Yの平均	7
オブザベーション(または重みの合計)	6

分散分析

要因	自由度	平方和	平均平方	F値
モデル	1	19.220000	19.2200	27.6547
誤差	4	2.780000	0.6950	p値(Prob>F)
全体(修正済み)	5	22.000000		0.0063*

パラメータ推定値

項	推定値	標準誤差	t値	p値(Prob>|t|)
切片	3.9	0.680686	5.73	0.0046*
x	0.62	0.117898	5.26	0.0063*

回帰式
（p.228）

寄与率（決定係数）
自由度調整寄与率

残差標準偏差
（p.233）
RMSE ：
Root Mean Square Error

分散分析表
（p.244）

仮説検定（p.240）
H_0 ：β=0
H_0 ：α=0

回帰モデル

p.227

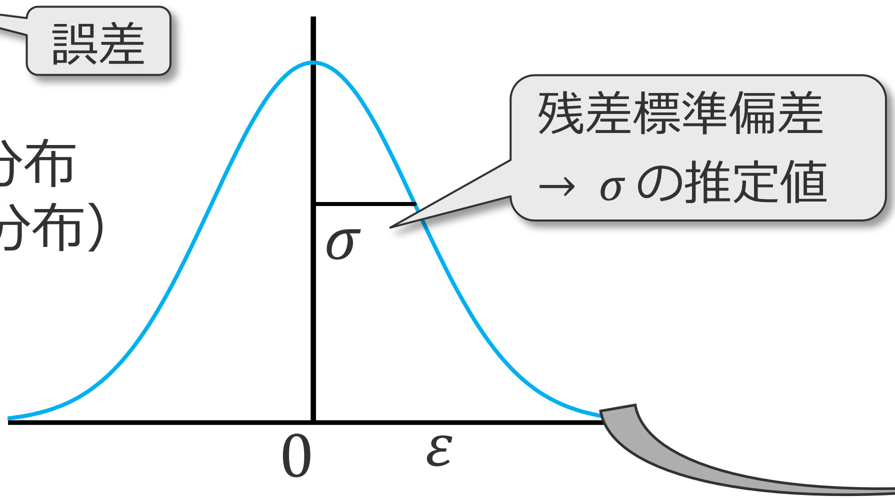
●単回帰モデルの視覚的説明

$\eta = \alpha + \beta x$　（η ：母平均）　§4.3

$y = \eta + \varepsilon$　（y ：観測値）

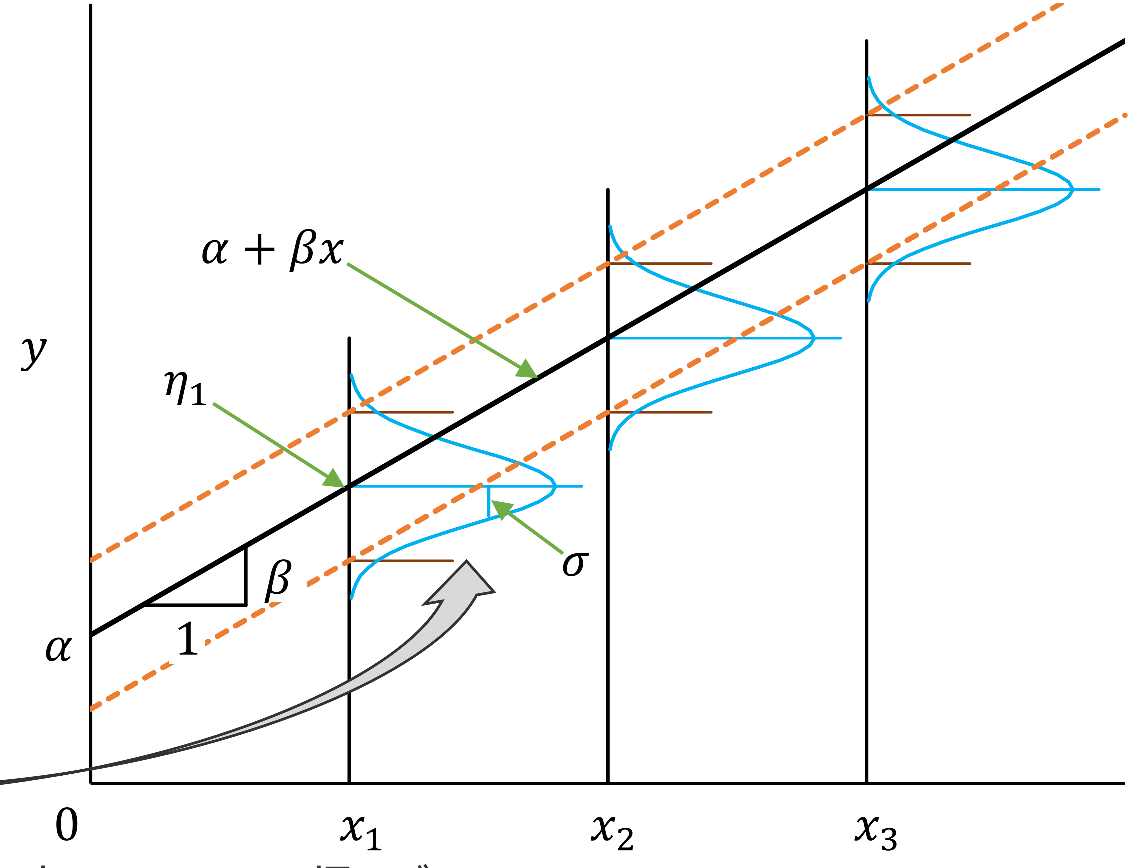
$y = \eta + \varepsilon = \alpha + \beta x + \varepsilon$

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



BMI が 特定の値 x_1 の人の血糖 y_1 を測定

$\varepsilon = y_1 - \eta_1$



表示 4.3.2　回帰モデル

JMP ［二変量の関係］

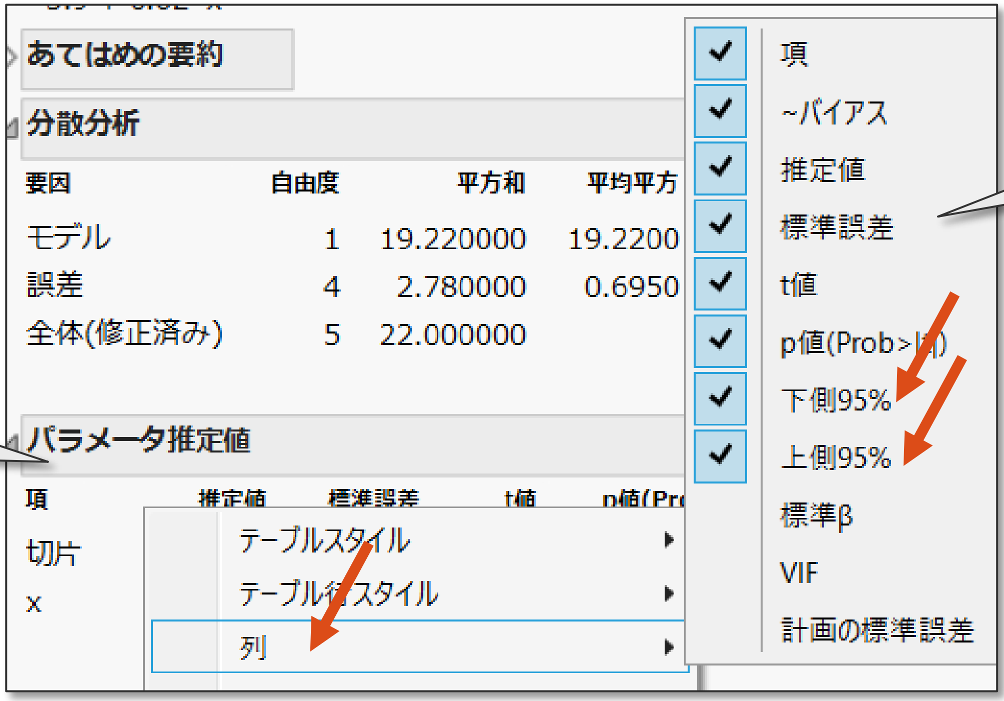
p.246

●信頼区間

母回帰係数の95%信頼区間

母切片の95%信頼区間

「パラメータ推定値」の表の上で右クリック



オプションメニュー

パラメータ推定値

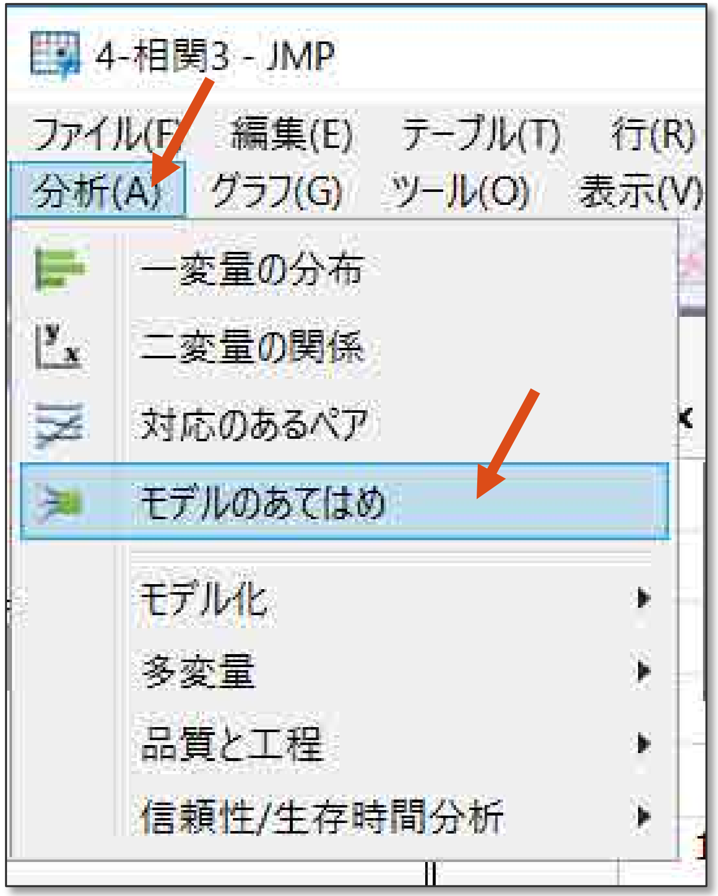
項	推定値	標準誤差	t値	p値(Prob>|t|)	下側95%	上側95%
切片	3.9	0.680686	5.73	0.0046*	2.0101129	5.7898871
x	0.62	0.117898	5.26	0.0063*	0.2926619	0.9473381

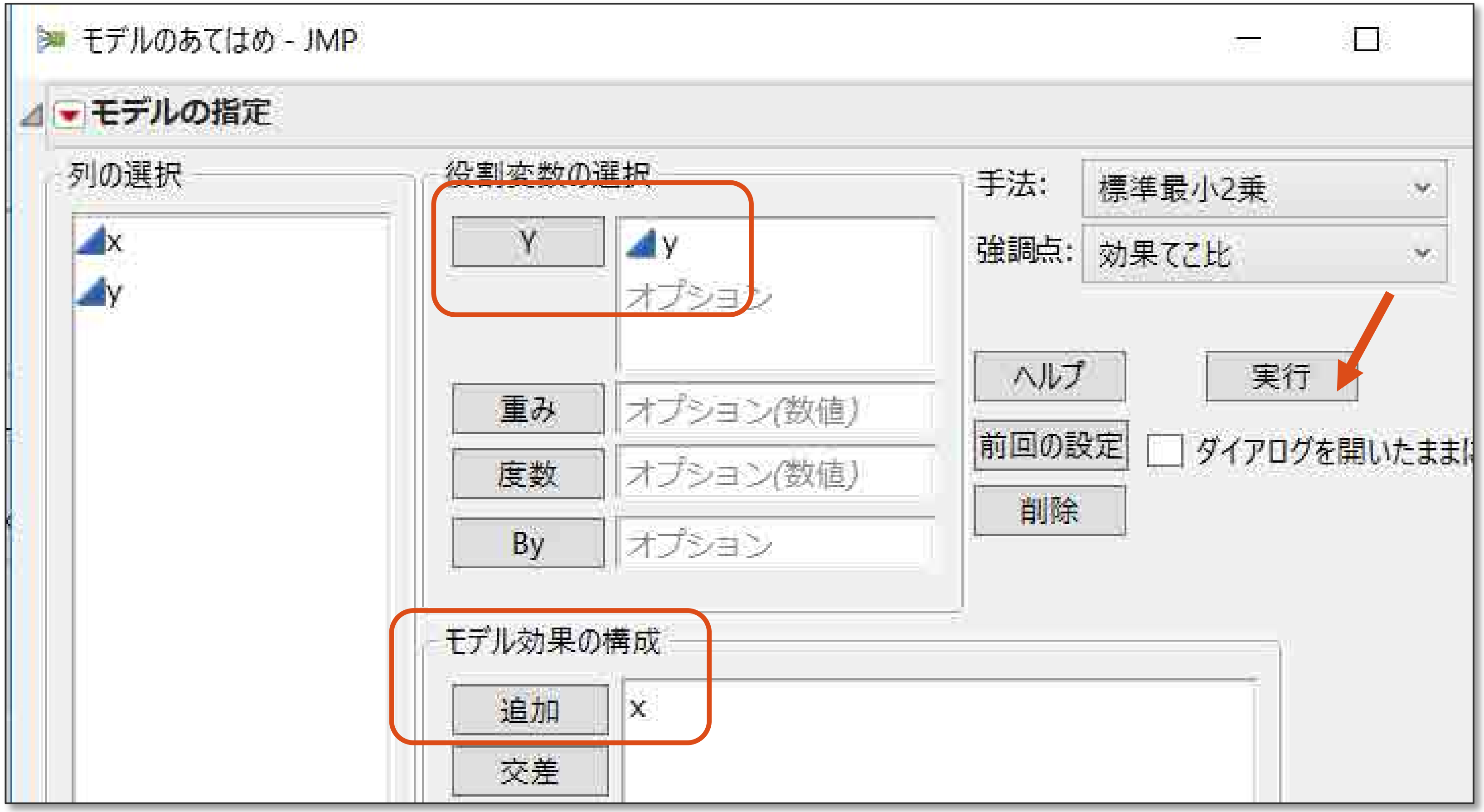
表示4.4.7
JMP「二変量の関係」
による回帰式

JMP［モデルのあてはめ］

●単回帰分析

［分析］＞［モデルのあてはめ］＞ダイアログ





JMP ［モデルのあてはめ］

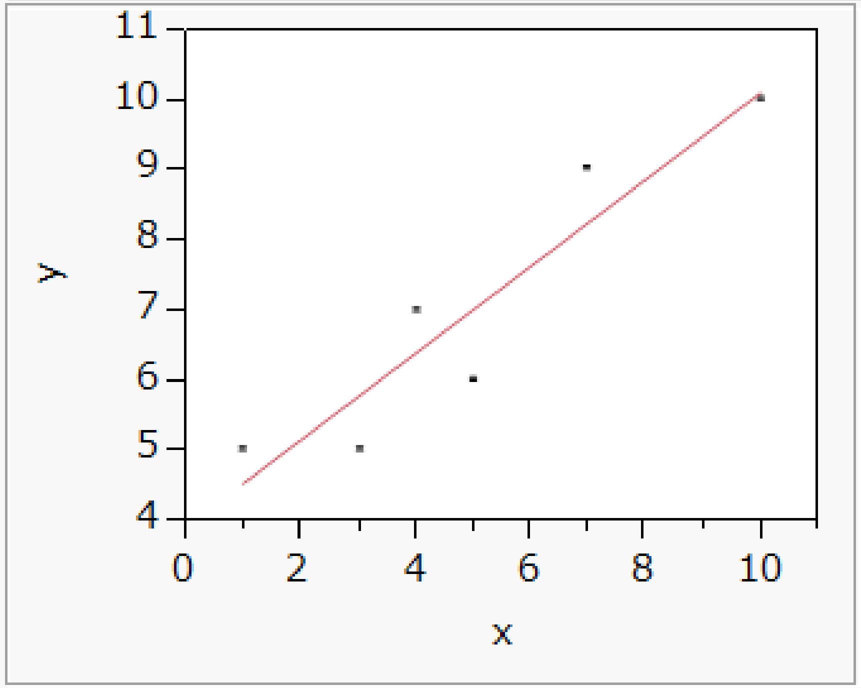
p.246

●単回帰分析

応答 y

モデル全体

回帰プロット



あてはめの要約

R2乗	0.873636
自由度調整R2乗	0.842045
誤差の標準偏差(RMSE)	0.833667
Yの平均	7
オブザベーション(または重みの合計)	6

分散分析

要因	自由度	平方和	平均平方	F値
モデル	1	19.220000	19.2200	27.6547
誤差	4	2.780000	0.6950	p値(Prob>F)
全体(修正済み)	5	22.000000		0.0063*

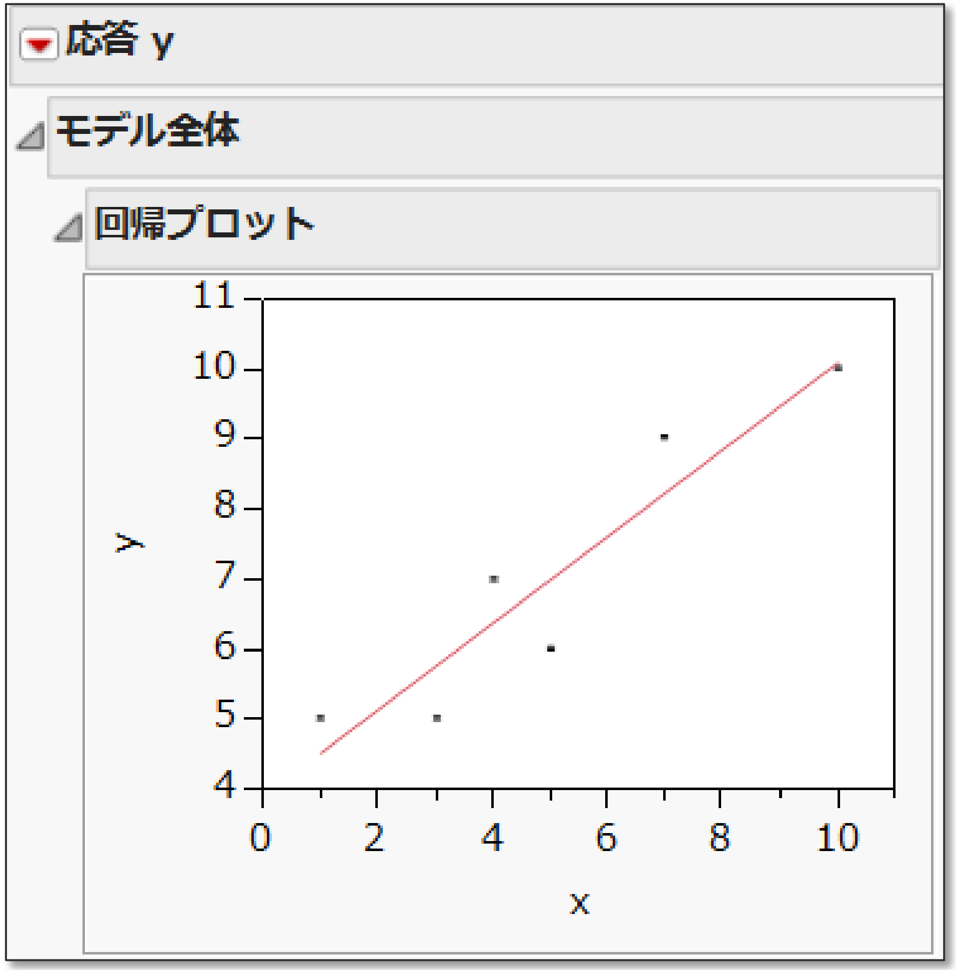
パラメータ推定値

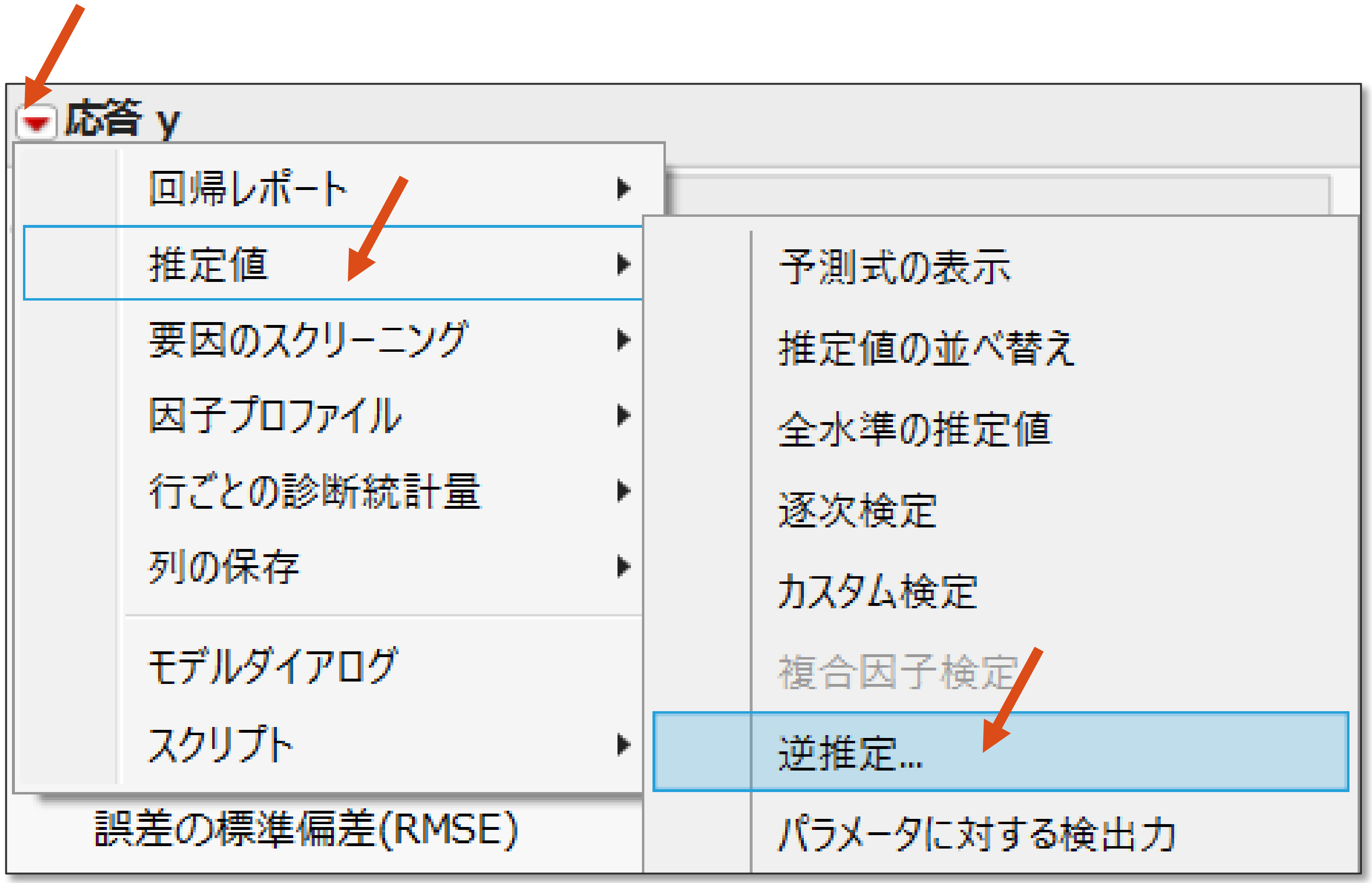
項	推定値	標準誤差	t値	p値(Prob>|t|)
切片	3.9	0.680686	5.73	0.0046*
x	0.62	0.117898	5.26	0.0063*

JMP ［モデルのあてはめ］

p.246

●逆推定



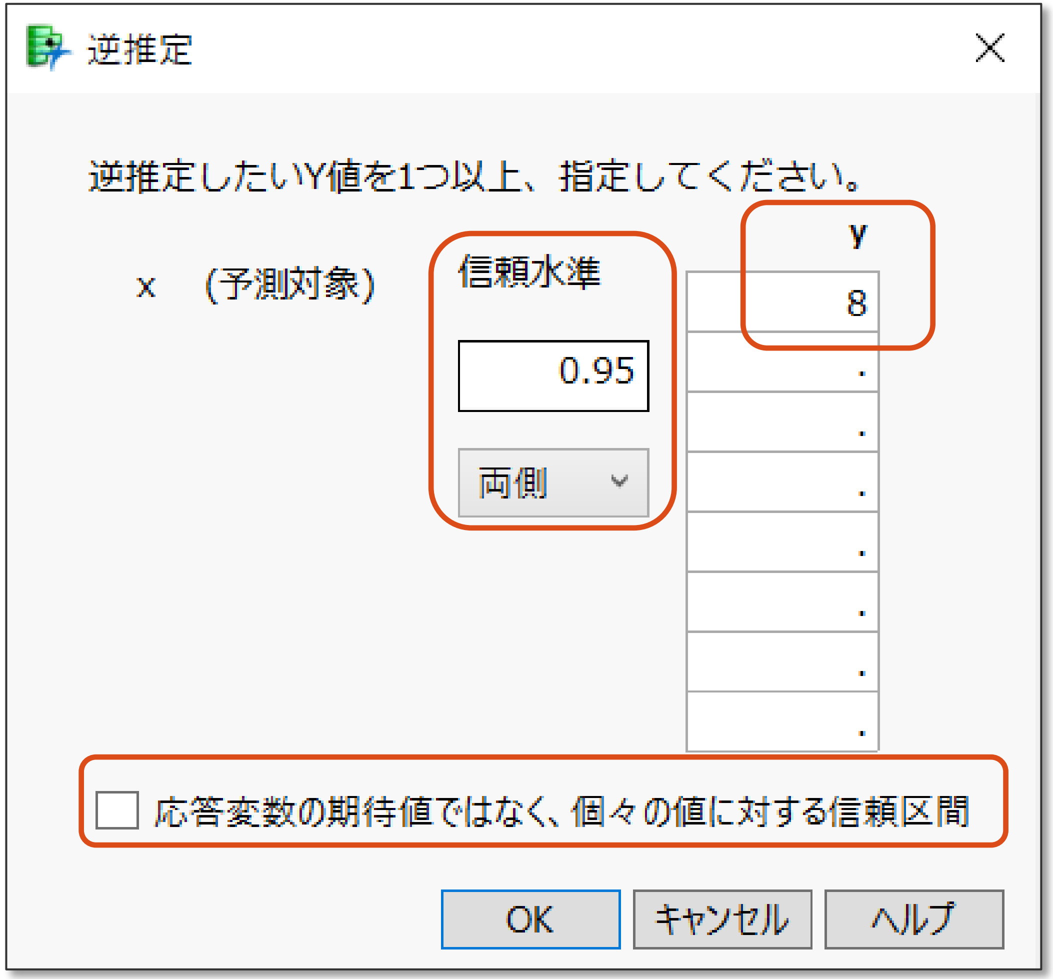


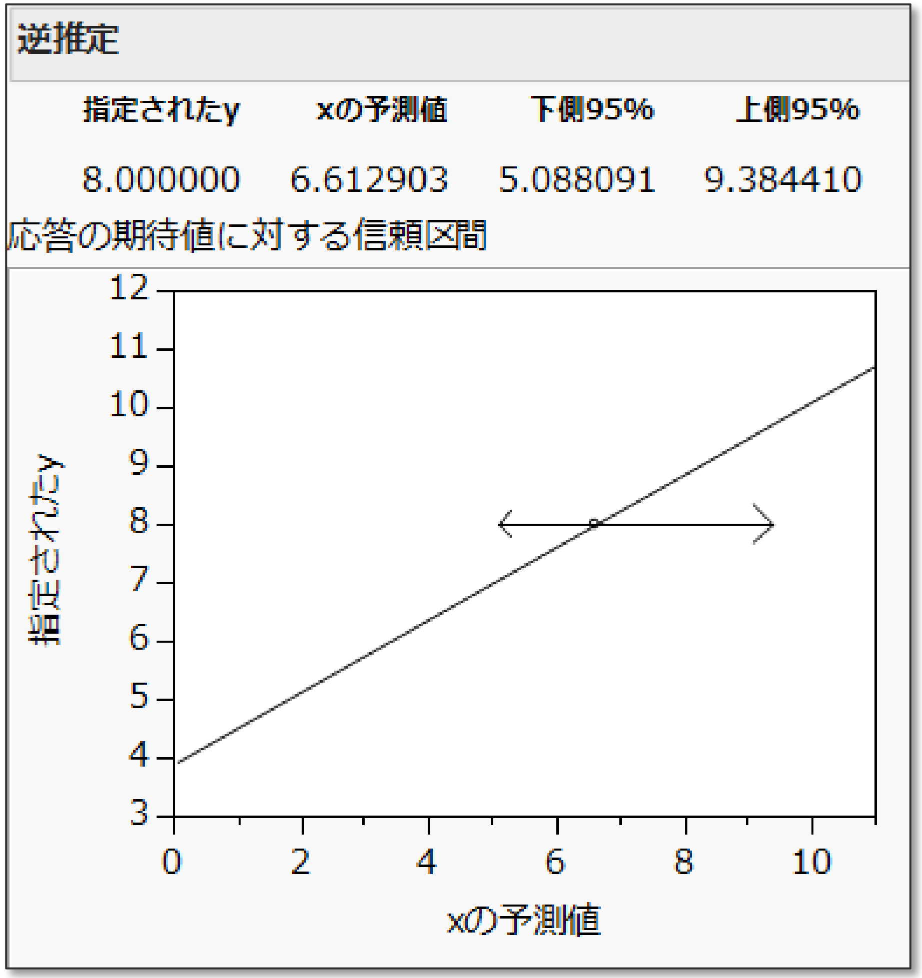
JMP ［モデルのあてはめ］

p.246

●逆推定

表示4.4.8 JMP「モデルまあてはめ」による逆推定（1）





JMP［モデルのあてはめ］

p.247

●逆推定

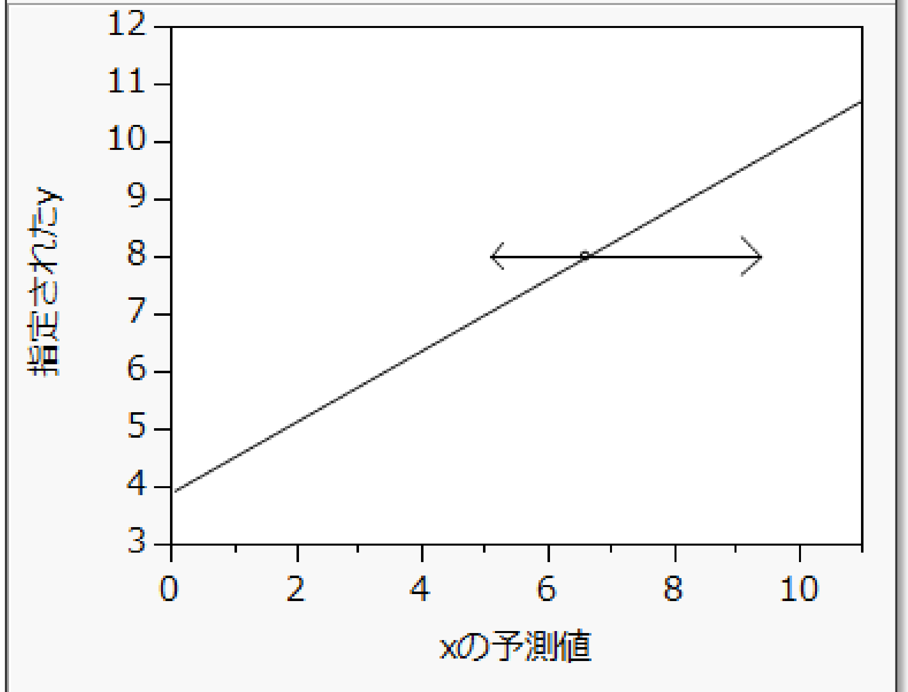
応答の期待値に対する信頼区間
（母平均 η の信頼限界）

表示4.4.8

逆推定

指定されたy	xの予測値	下側95%	上側95%
8.000000	6.612903	5.088091	9.384410

応答の期待値に対する信頼区間



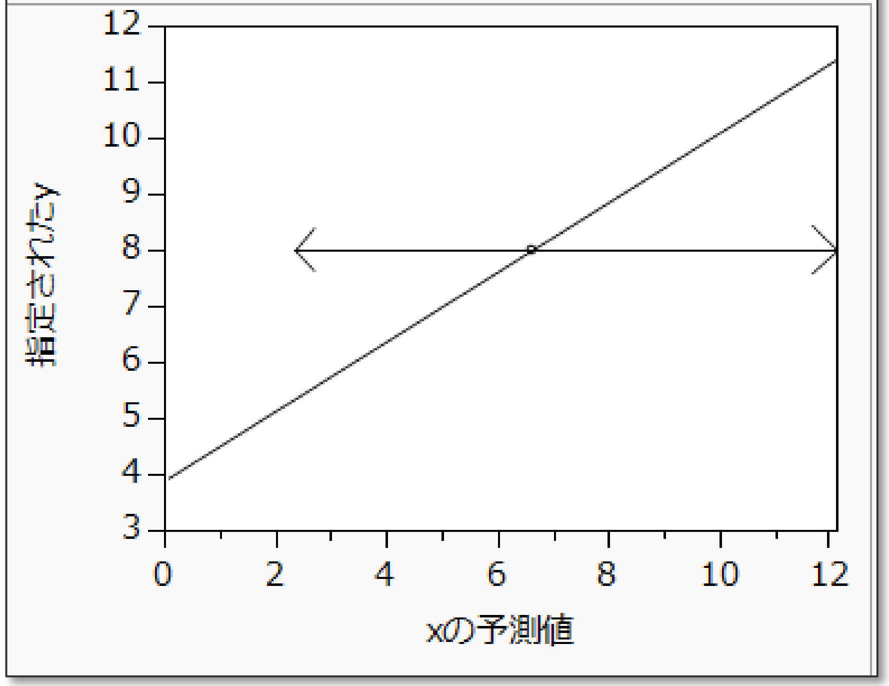
個別の応答に対する信頼区間
（観測値 y の信頼限界）

表示4.4.9

逆推定

指定されたy	xの予測値	下側95%	上側95%
8.000000	6.612903	2.343567	12.12893

個別の応答に対する信頼区間



JMP ［モデルのあてはめ］

●逆推定

応答の期待値に対する信頼区間
（η の信頼限界）

表示4.4.8

逆推定

指定されたy	xの予測値	下側95%	上側95%
8.000000	6.612903	5.088091	9.384410

応答の期待値に対する信頼区間

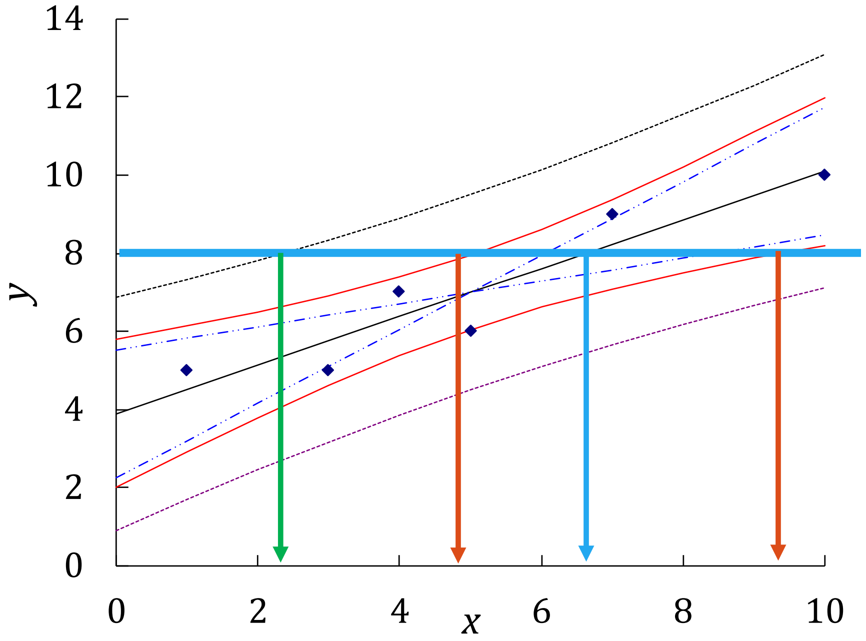
個別の応答に対する信頼区間
（y の信頼限界）

表示4.4.9

逆推定

指定されたy	xの予測値	下側95%	上側95%
8.000000	6.612903	2.343567	12.12893

個別の応答に対する信頼区間



表示4.4.5　逆推定

	S	T	U	V	W	X	Y
29				η の信頼限界		y の信頼限界	
30	x	y	y-hat	下限	上限	下限	上限
31	6.61		**8.00**	6.92	9.08	5.44	10.56
32	9.38		9.72	**8.00**	11.44	6.84	12.60
33	5.09		7.06	6.11	**8.00**	4.55	9.56
34	12.13		11.42	8.90	13.94	**8.00**	14.84
35	2.34		5.35	4.07	6.64	2.71	**8.00**

JMP ［二変量の関係］

p.271

●回帰直線と確率楕円（演習4..4.2）　表示4.7.7

JMP ファイル「4-相関3.jmp」

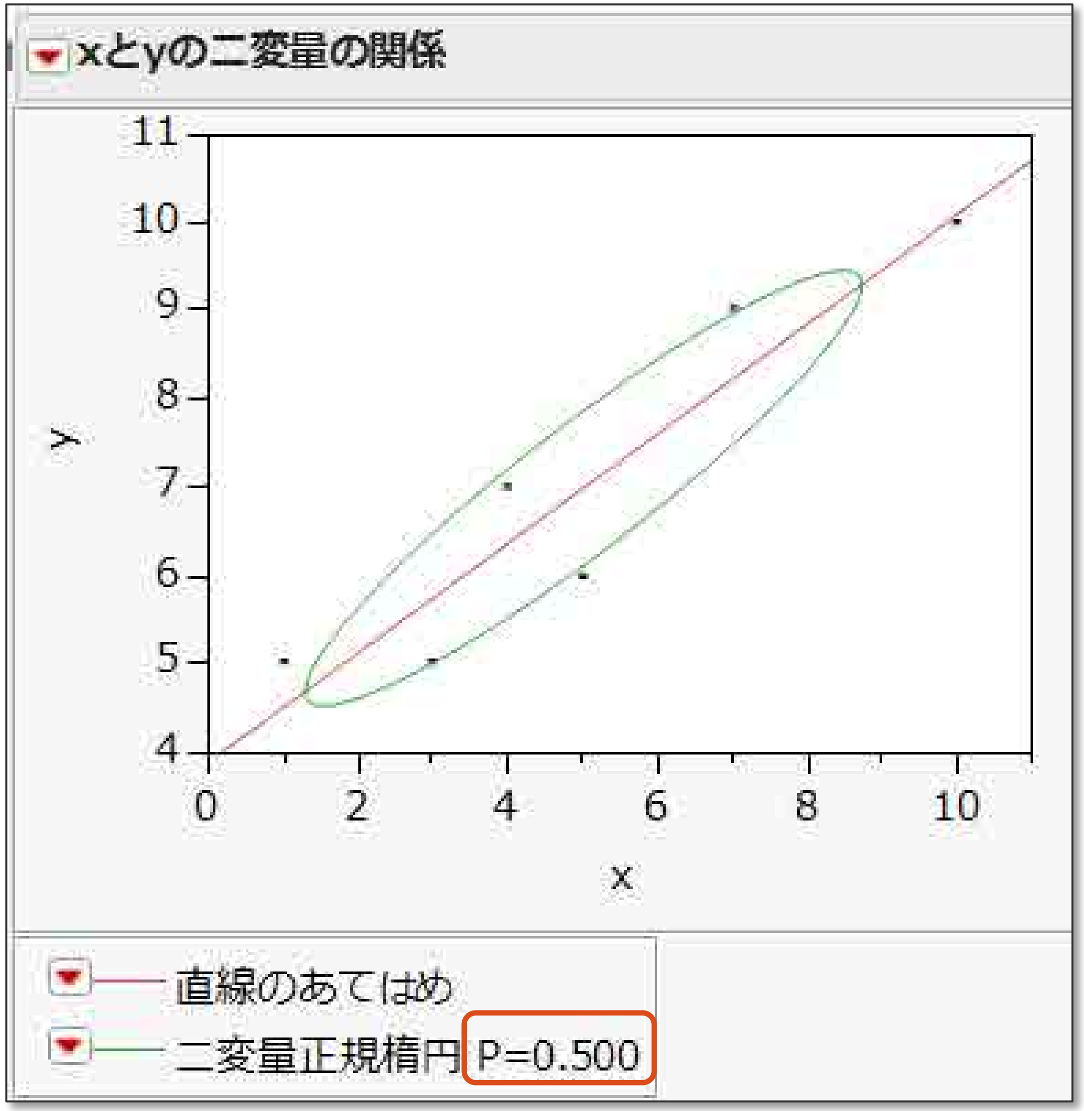
［分析］＞［二変量の関係］＞y→Y, x→X

▼オプション>［直線のあてはめ］

▼オプション>［確率楕円］＞［0.50］

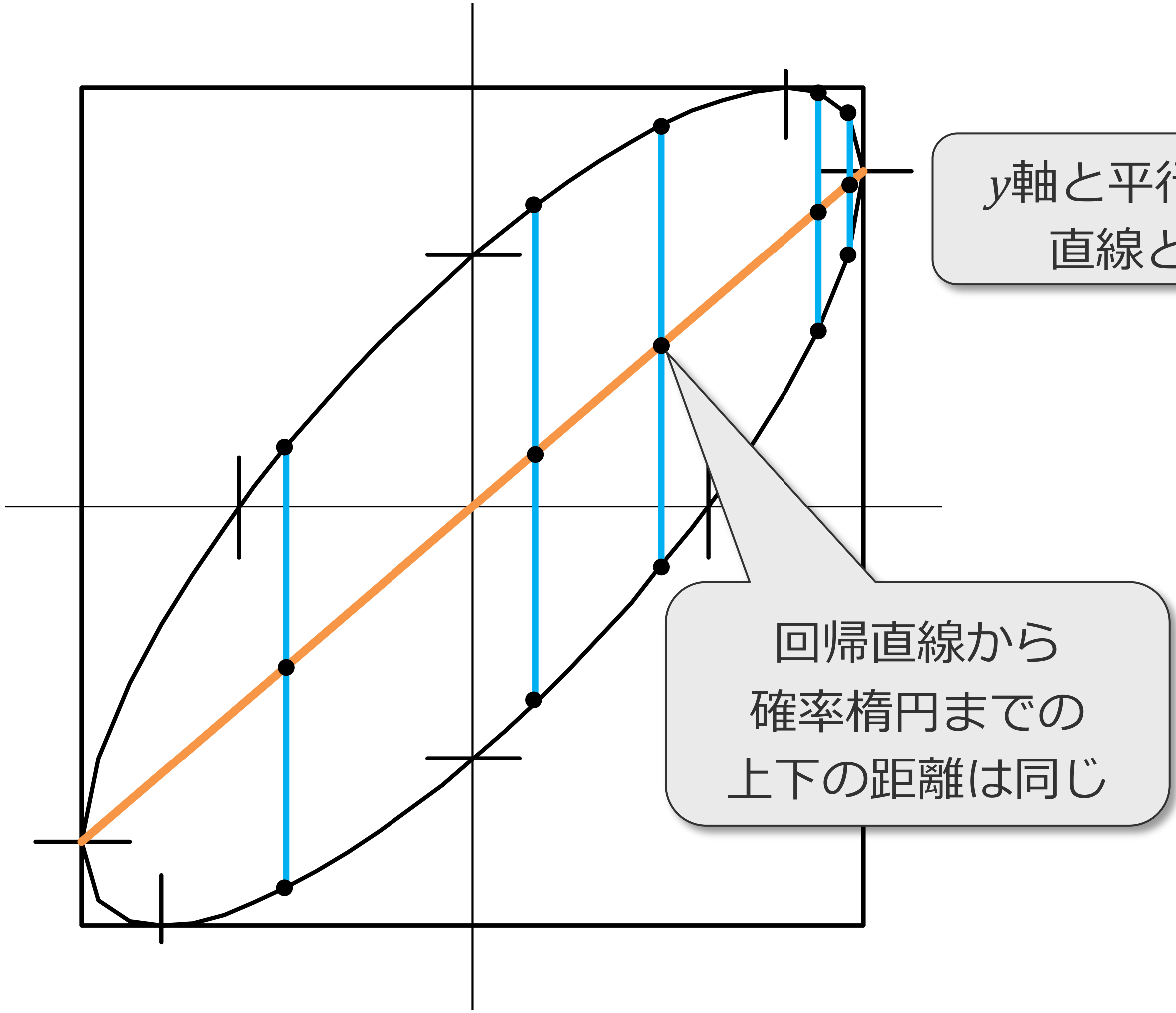
回帰直線と確率楕円の関係を確認する

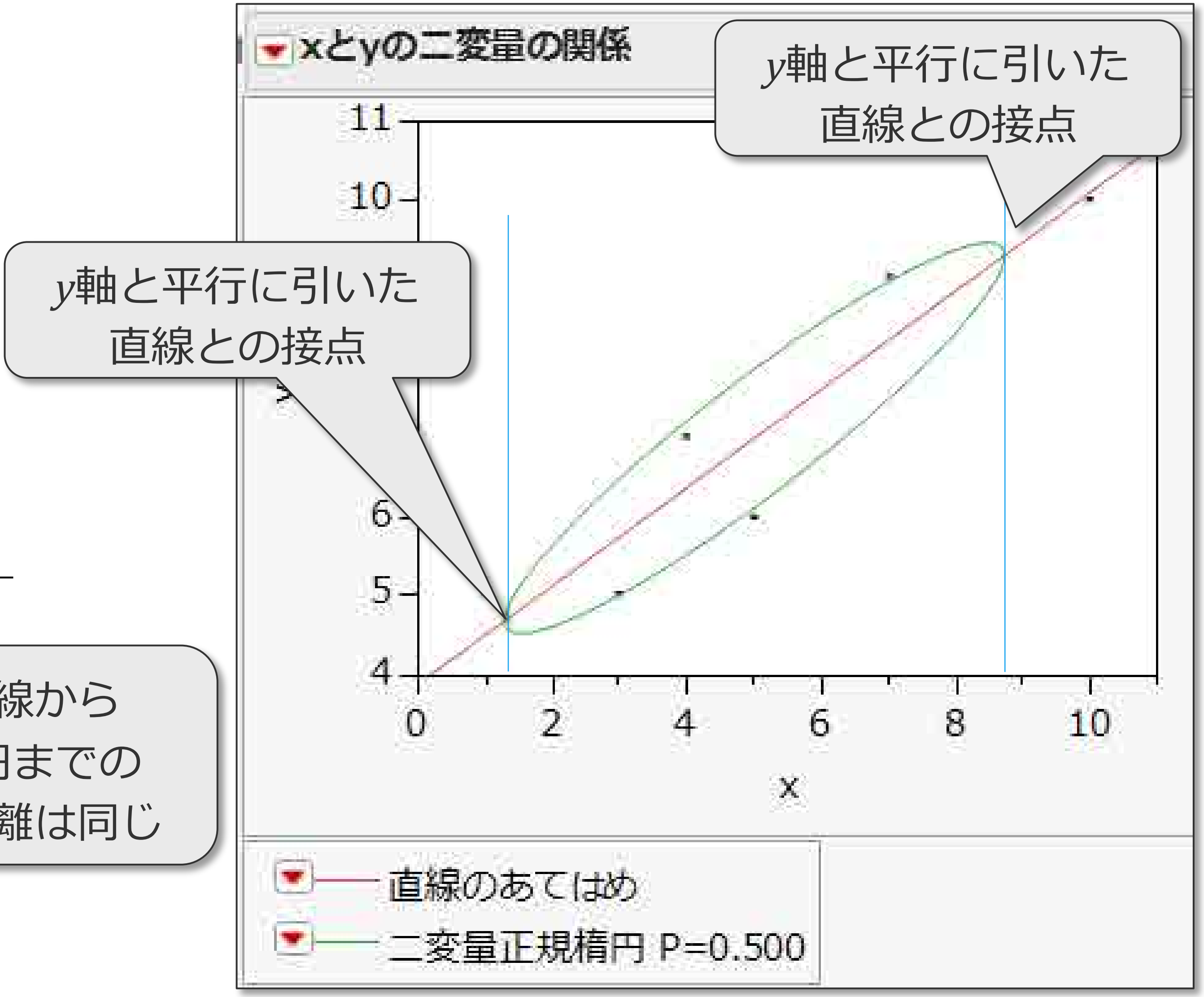
	x	y
1	1	5
2	3	5
3	4	7
4	5	6
5	7	9
6	10	10



JMP ［二変量の関係］

●回帰直線と確率楕円





JMP ［二変量の関係］

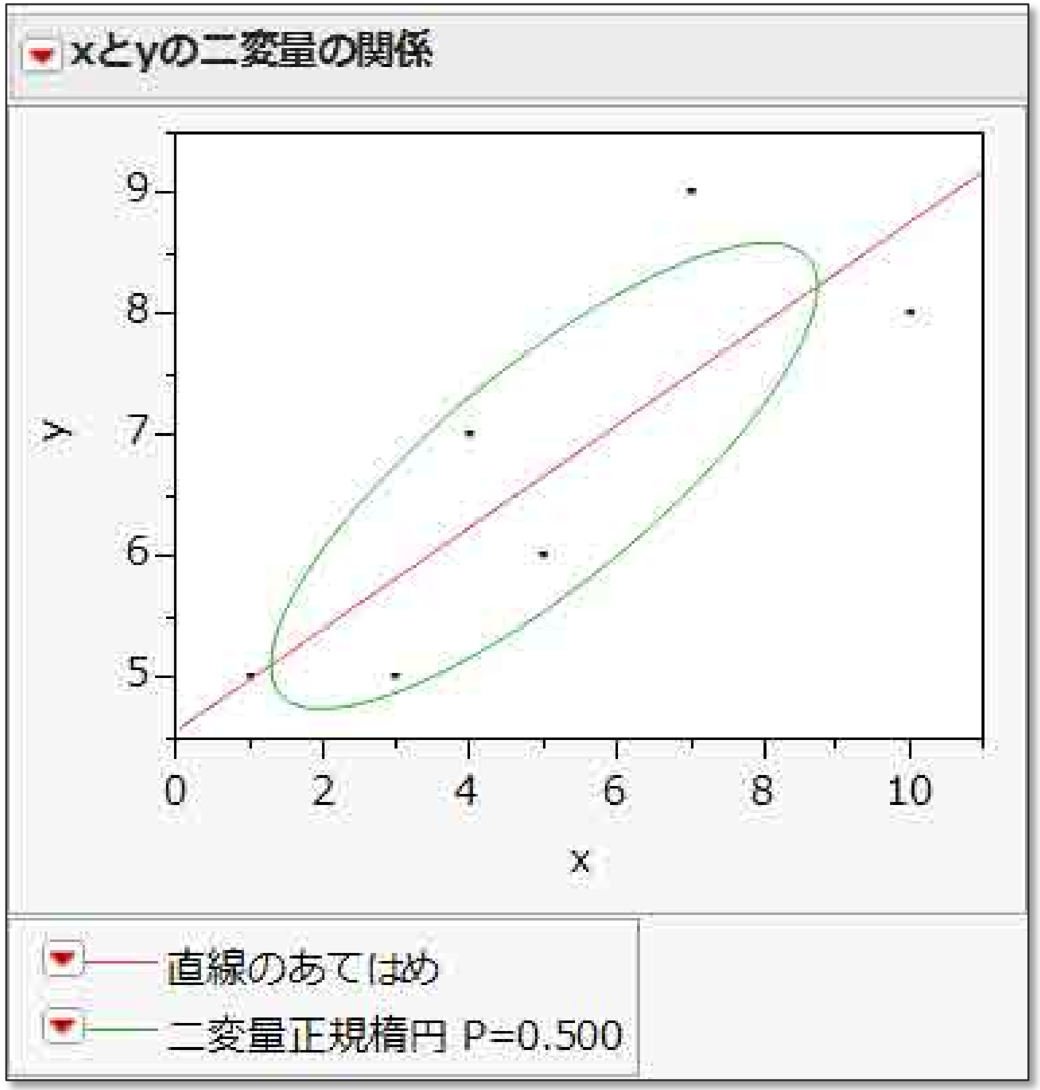
p.271

●回帰直線と確率楕円

	x	y
1	1	5
2	3	5
3	4	7
4	5	6
5	7	9
6	10	8

10 → 8
変更

表示4.7.7

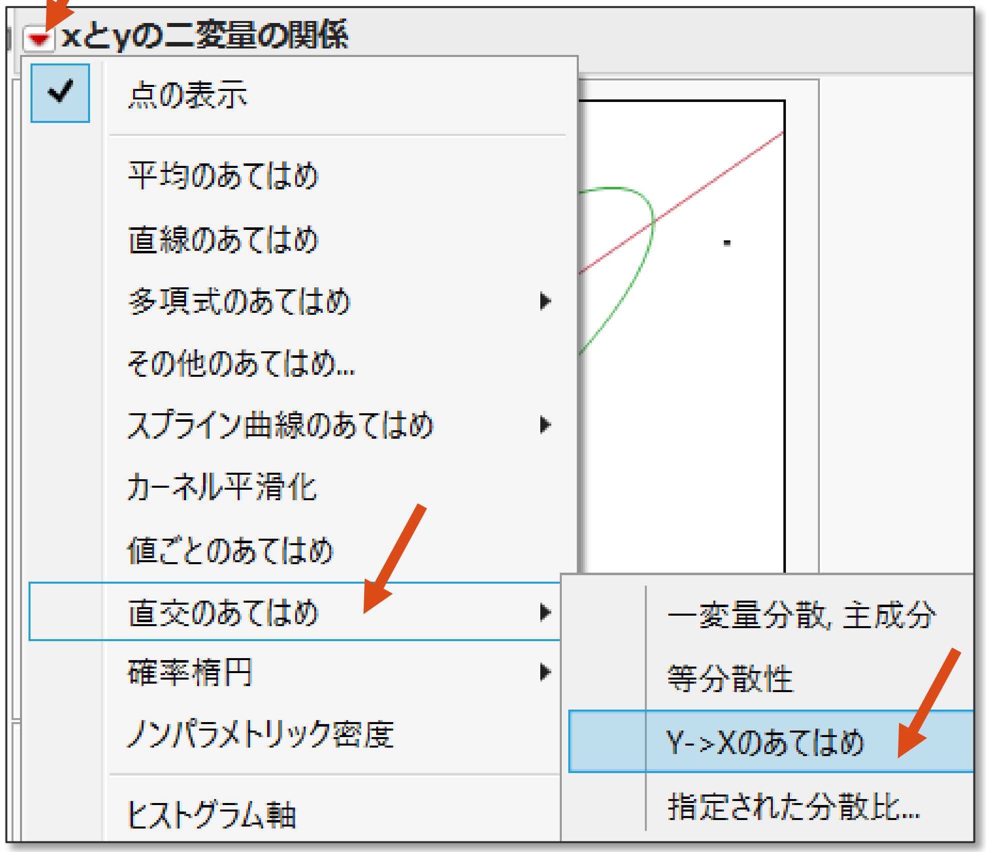


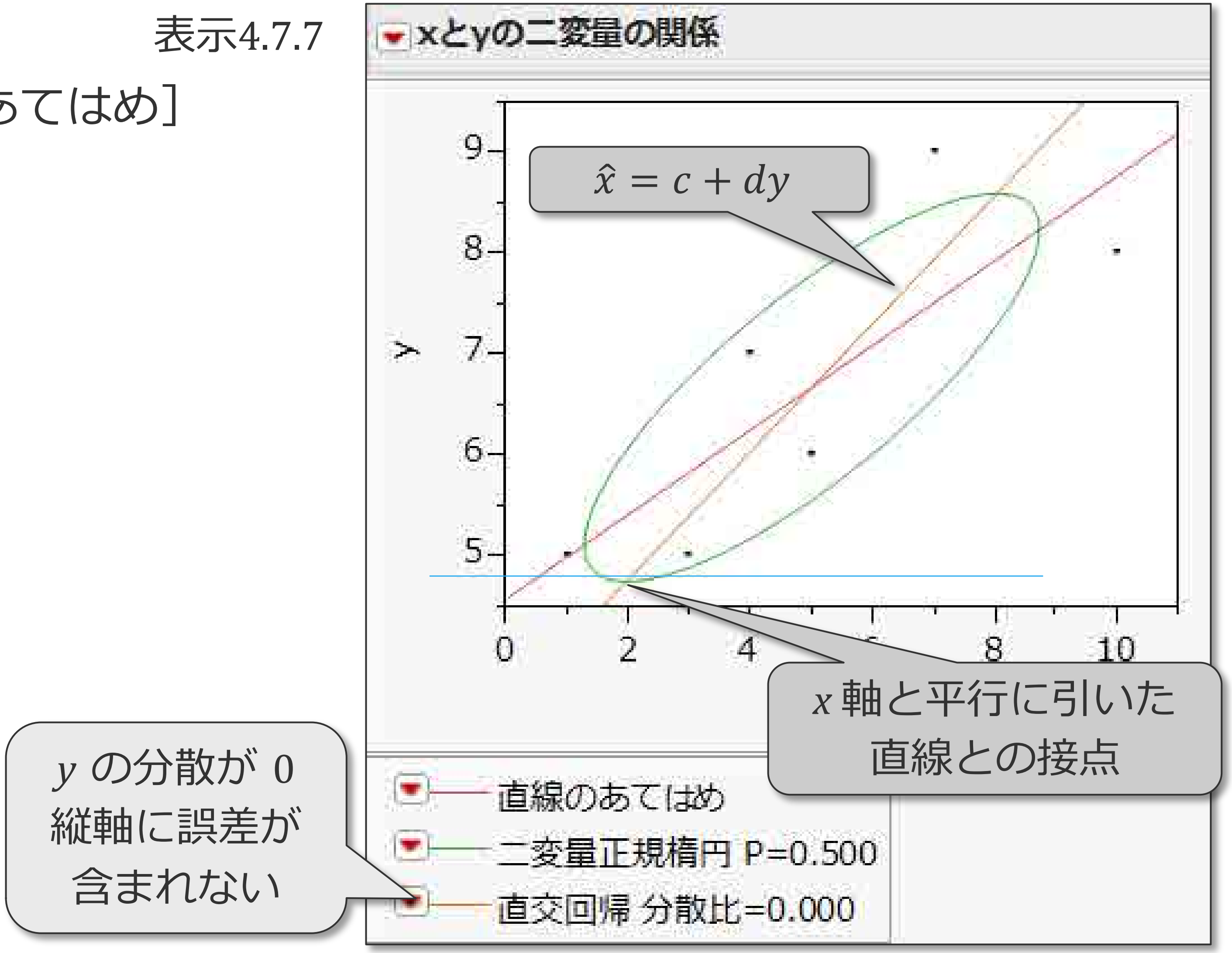
JMP ［二変量の関係］

●回帰直線と確率楕円

表示4.7.7

▼＞［直交のあてはめ］＞［Y->Xのあてはめ］

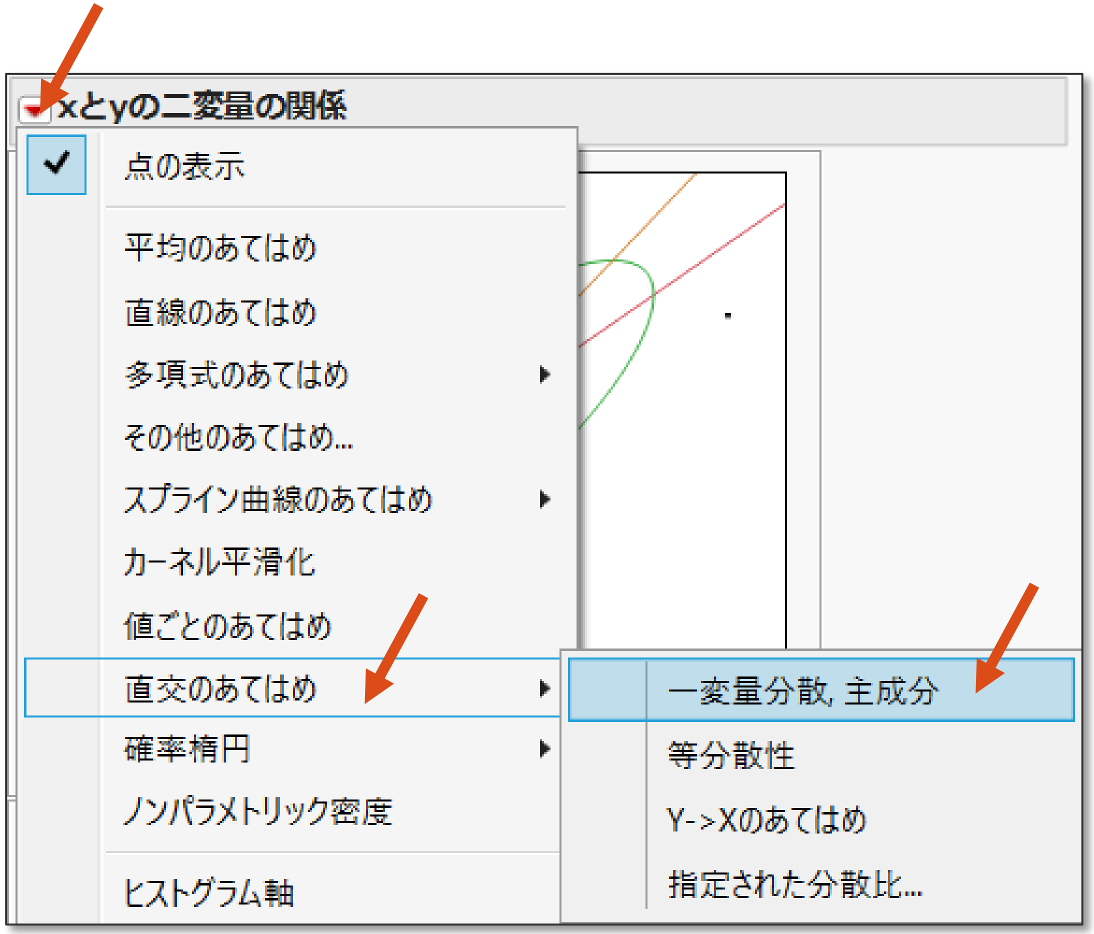




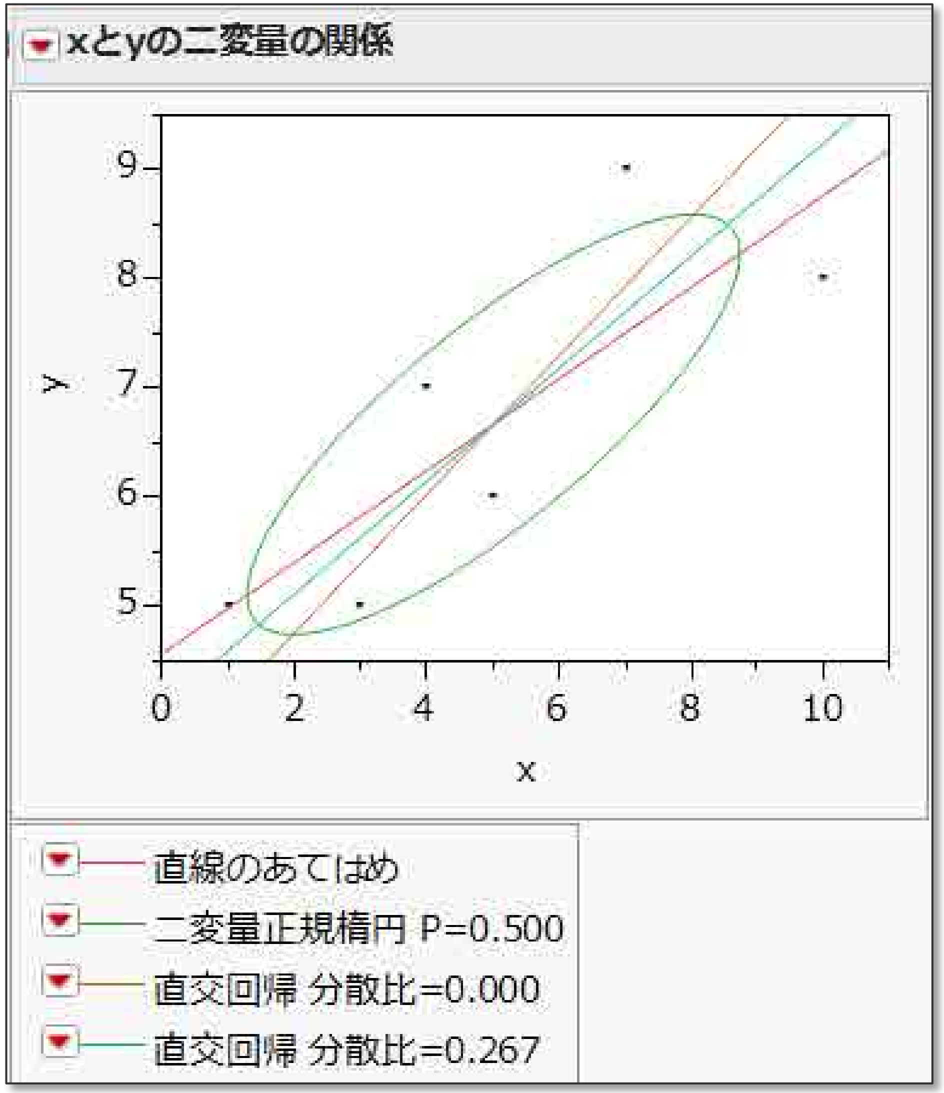
JMP ［二変量の関係］

●回帰直線と確率楕円

▼＞［直交のあてはめ］＞［一変量分散、主成分］



表示4.7.7



まとめ

p.247

- 回帰式 $y=a+bx$ の推定

  データには誤差が含まれるため，
  得られた推定値 a，b の値にも誤差が含まれる
  これをシミュレーションで認識

- 推定値 a，b の標準誤差

- yの予測値の標準誤差

- 逆推定（y から x を推定）

  区間推定は左右非対称

- 作成　片瀬雅彦
- 監修　松本一彦、長谷文雄
- 作成時期　2019年3月4日
- 改訂　2019年7月1日、2022年5月31日