



4 相関・回帰

4.5 回帰分析適用上の諸問題

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**



4.5 回帰分析適用上の諸問題

p.248

- (1) Anscombe (アンスコム) の例
- (2) 残差の検討
- (3) x に誤差を含む場合
- (4) 変数の選び方

補足 原点を通る回帰式

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

テキストの
該当ページ

★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



(1) Anscombe (アンスコム) の例

2 変数の関係を解析するとき,
グラフ化が重要であることを示す事例 (Anscombe, 1973)
多くの著書で引用されている

Anscombe の例

●事例 1

4組の2変数のデータ

(a) x1 と y1

(b) x1 と y2

(c) x1 と y3

(d) x4 と y4

それぞれの x、y の
平均, 標準偏差は同じ

本来、最初にグラフ化するが、
グラフ化しないで統計量を計算
4組の相関係数はほとんど同じ
強い相関がある

表示 4.5.1 Anscombe の例

	x1	y1	y2	y3	x4	y4
1	10	8.04	9.14	7.46	8	6.58
2	8	6.95	8.14	6.77	8	5.76
3	13	7.58	8.74	12.74	8	7.71
4	9	8.81	8.77	7.11	8	8.84
5	11	8.33	9.26	7.81	8	8.47
6	14	9.96	8.10	8.84	8	7.04
7	6	7.24	6.13	6.08	8	5.25
8	4	4.26	3.10	5.39	19	12.50
9	12	10.84	9.13	8.15	8	5.56
10	7	4.82	7.26	6.42	8	7.91
11	5	5.68	4.74	5.73	8	6.89
平均	9.000	7.501	7.501	7.500	9.000	7.501
標準偏差	3.317	2.032	2.032	2.030	3.317	2.031
相関係数		0.816 ^(a)	0.816 ^(b)	0.816 ^(c)		0.817 ^(d)

●単回帰分析

単回帰分析の結果

回帰係数、切片

その標準誤差

寄与率

回帰平方和

残差標準偏差

傾きと切片の t 値、 p 値

4 組の統計量はほぼ同じ

↓

いずれも同じ関係と

判断してよいか？

まずはグラフ化

(a) x_1 と y_1

	傾き	切片
回帰係数	0.500	3.000
標準誤差	0.118	1.125
寄与率	0.667	1.237
分散比	17.990	9
回帰平方和	27.510	13.763
t 値	4.241	2.667
p 値	0.002	0.026

(b) x_1 と y_2

	傾き	切片
回帰係数	0.500	3.001
標準誤差	0.118	1.125
残差標準偏差	0.666	1.237
残差の自由度	17.966	9
残差平方和	27.500	13.776
t 値	4.239	2.667
p 値	0.002	0.026

(c) x_1 と y_3

回帰係数	0.500	3.002
標準誤差	0.118	1.124
寄与率	0.666	1.236
分散比	17.972	9
回帰平方和	27.470	13.756
t 値	4.239	2.670
p 値	0.002	0.026

(d) x_4 と y_4

回帰係数	0.500	3.002
標準誤差	0.118	1.124
寄与率	0.667	1.236
分散比	18.003	9
回帰平方和	27.490	13.742
t 値	4.243	2.671
p 値	0.002	0.026

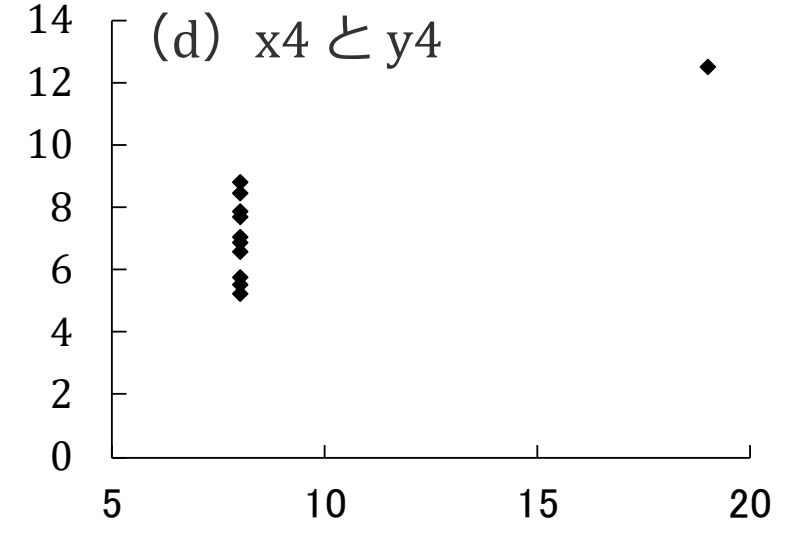
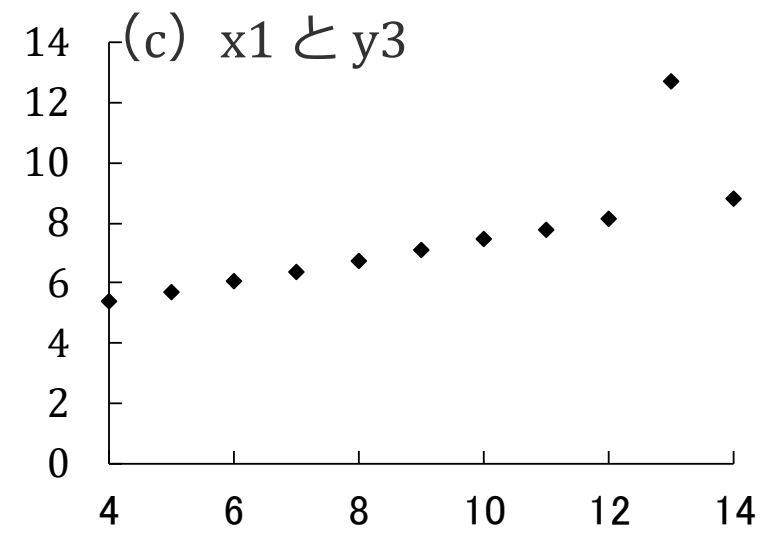
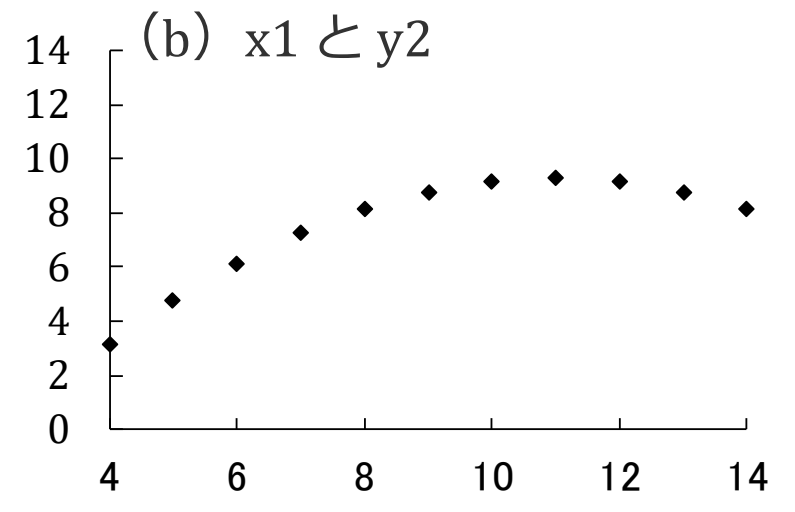
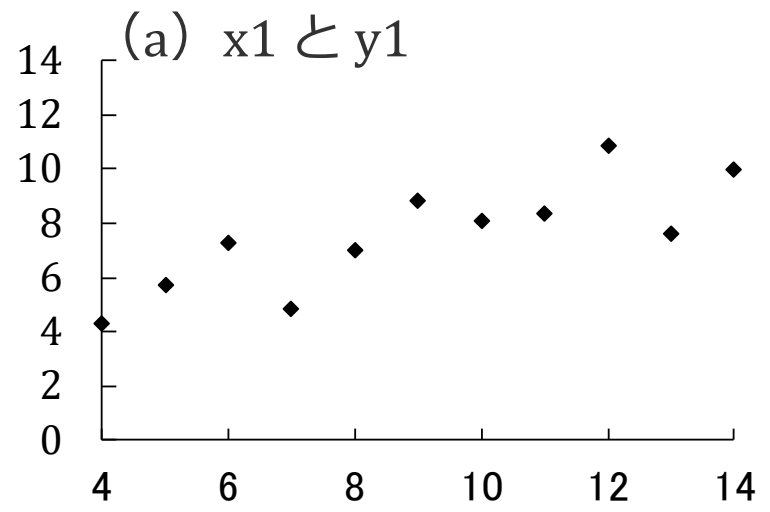


Anscombe の例

● 散布図

グラフから
2変数の関係が
同じではない
ことは明らか

表示 4.5.2 散布図と回帰直線 (Anscombe の例)



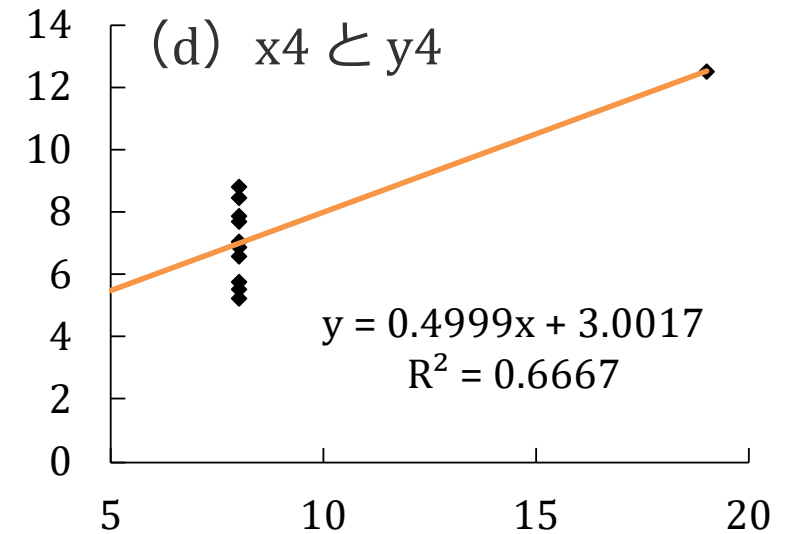
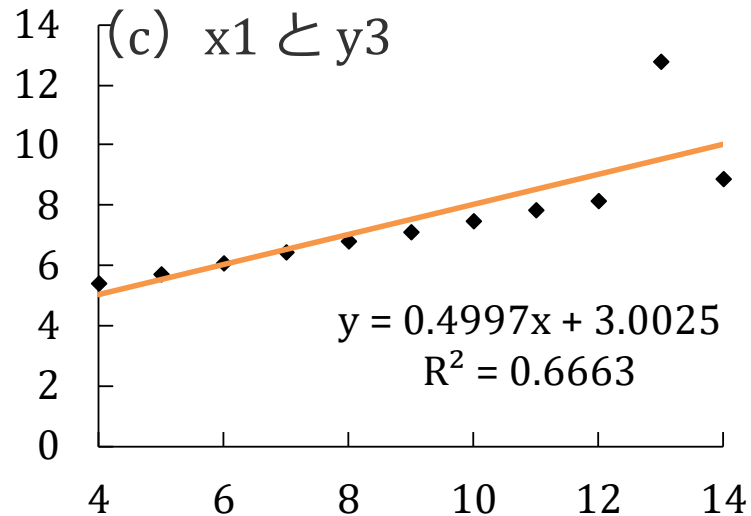
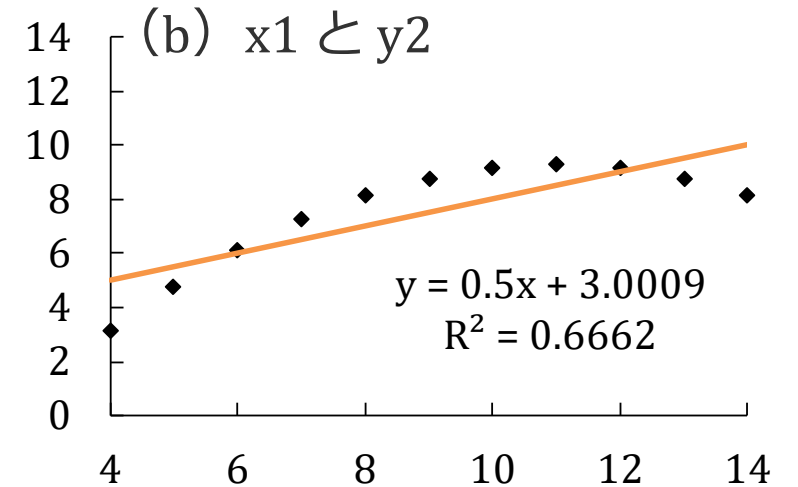
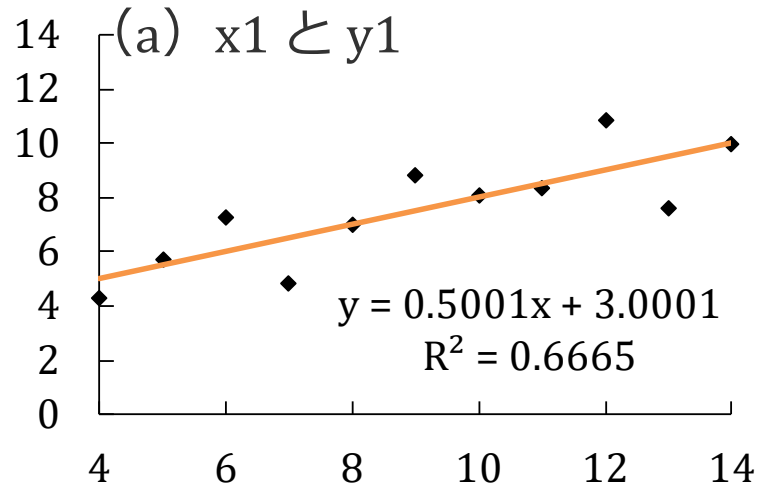
● 回帰直線

ほぼ同様の
回帰直線と
寄与率が
得られた



直線のあてはめ
は妥当ではない

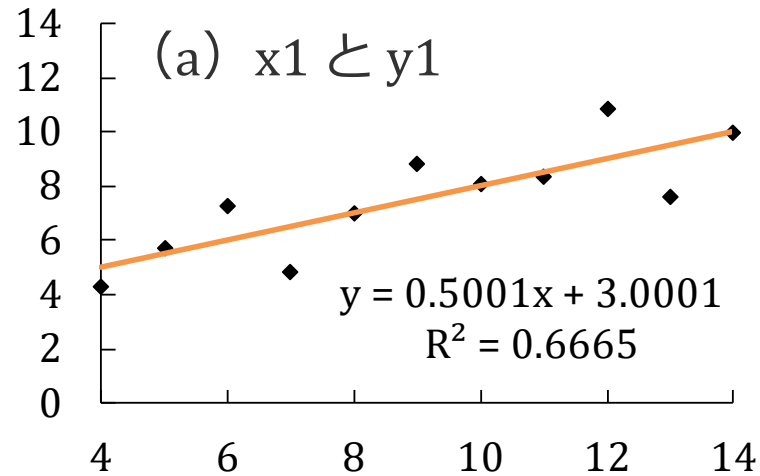
表示 4.5.2 散布図と回帰直線 (Anscombe の例)



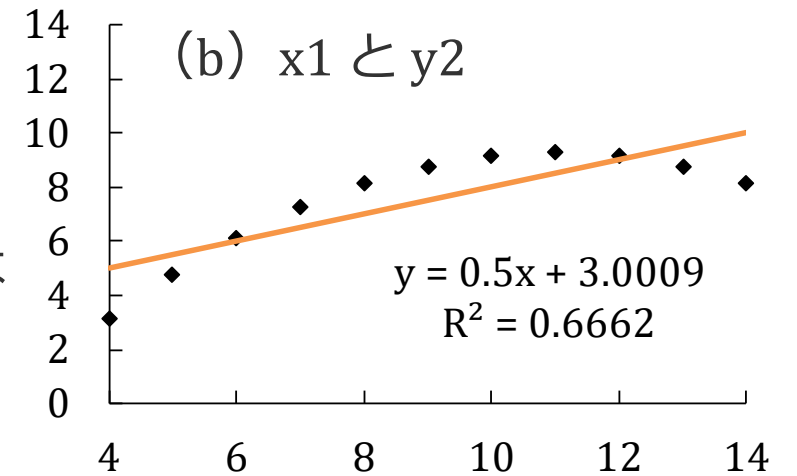
表示 4.5.2 散布図と回帰直線 (Anscombe の例)

●問題点

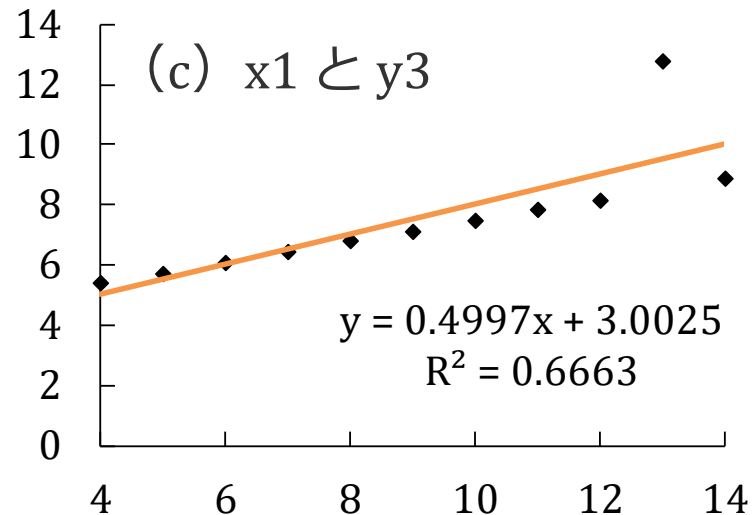
問題はない



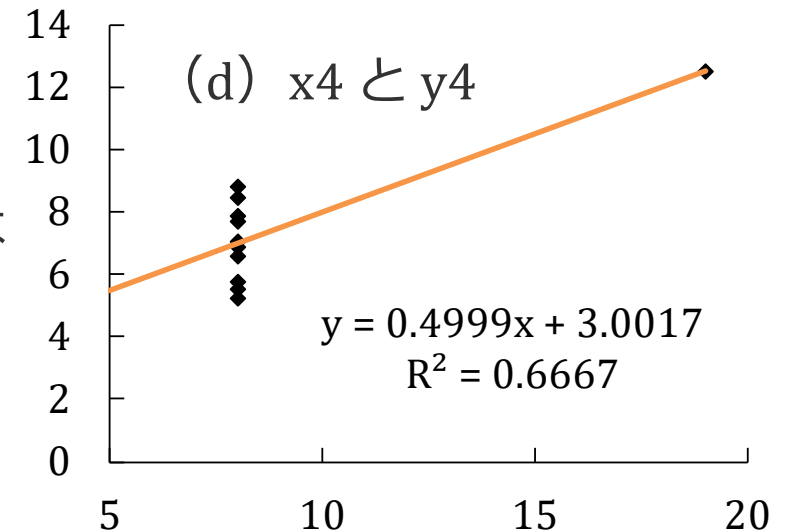
直線の
あてはめは
不適切
本来は
曲線関係



外れ値?
本来の直線と
異なる

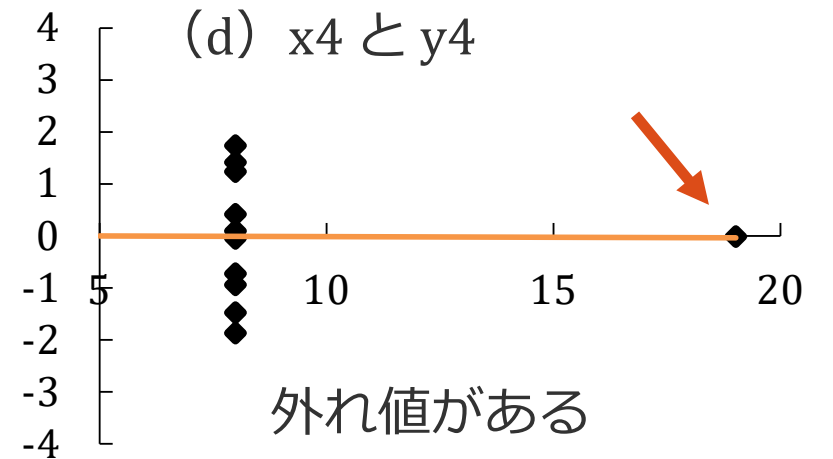
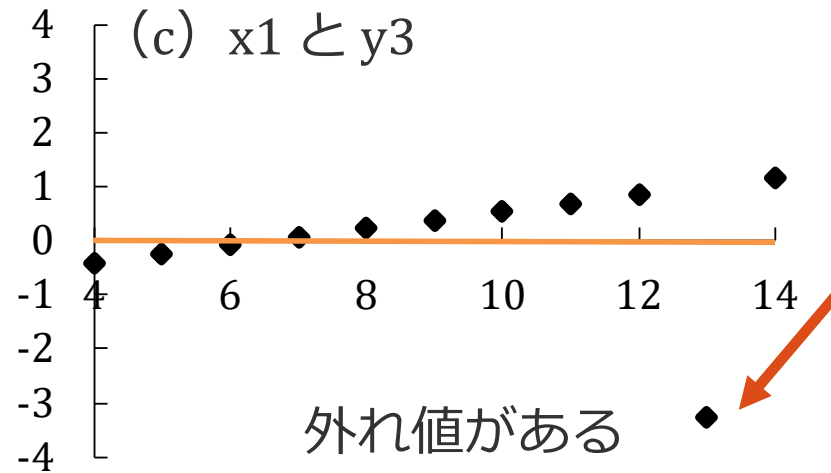
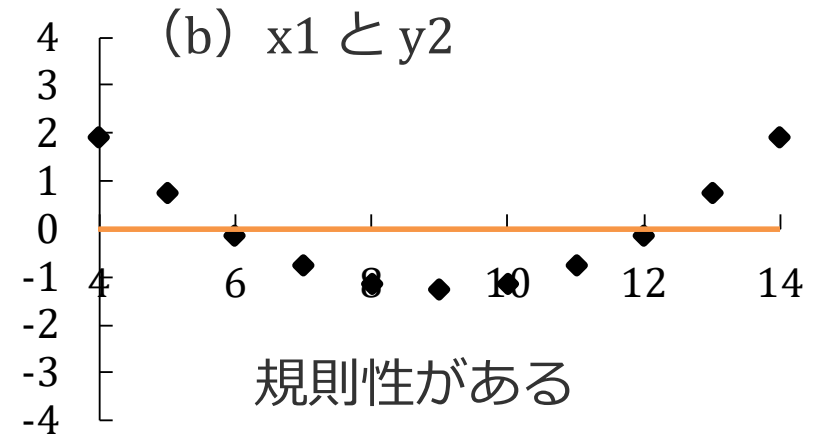
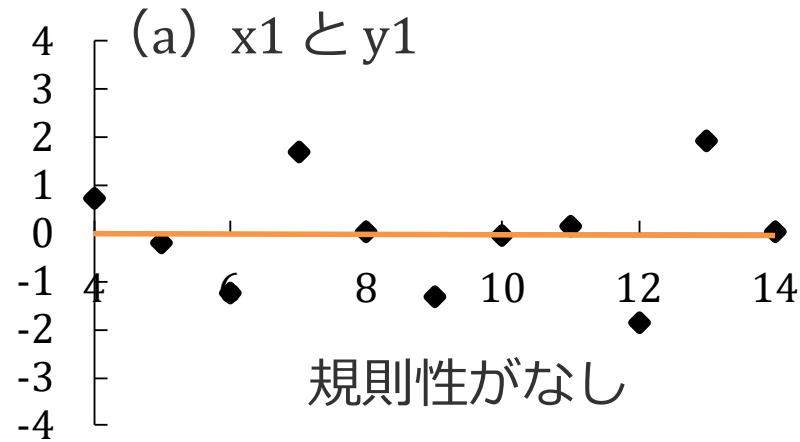


直線の
あてはめは
不適切
外れ値あり



●残差

x と残差 $(y_i - \hat{y}_i)$ の
プロット





●グラフ化（散布図）、残差の検討が重要

- (a) 特別の問題はない
- (b) 本来、曲線関係があるにもかかわらず、直線をあてはめている
残差に規則性があり、直線モデルをあてはめることは不適切
- (c) 1点の外れているため、本来あるべき直線とは異なった直線があてはめられている
外れた1点の原因と、残差に規則性が生じた原因を追究する必要がある
- (d) 1つの外れた値で回帰直線が決定されている
2変数は直線関係ではない

単回帰分析の適用が妥当な事例は (a) だけ

回帰式、寄与率などの数値（統計量）のみで判断してはならない

散布図を描き、個々の残差を検討することが重要

(c のケースは、第3部 §4.4 p.281で取り上げる)



(2) 残差の検討

単回帰モデルをあてはめた妥当性の検討
残差分散の推定
残差分析

●単回帰分析における残差の検討方法

残差からモデルの妥当性を判断

点予測

$$\text{残差 } (e_i) = \text{観測値 } (y_i) - \text{予測値 } (\hat{y}_i)$$

推定したモデルで説明できなかった部分

通常、残差は平均 (0) を中心に分布

$$\varepsilon_i \sim N(0, \sigma^2) \text{ を仮定したモデル}$$

基準化残差 (標準化残差)

残差を基準化 (標準偏差で割る) して比較

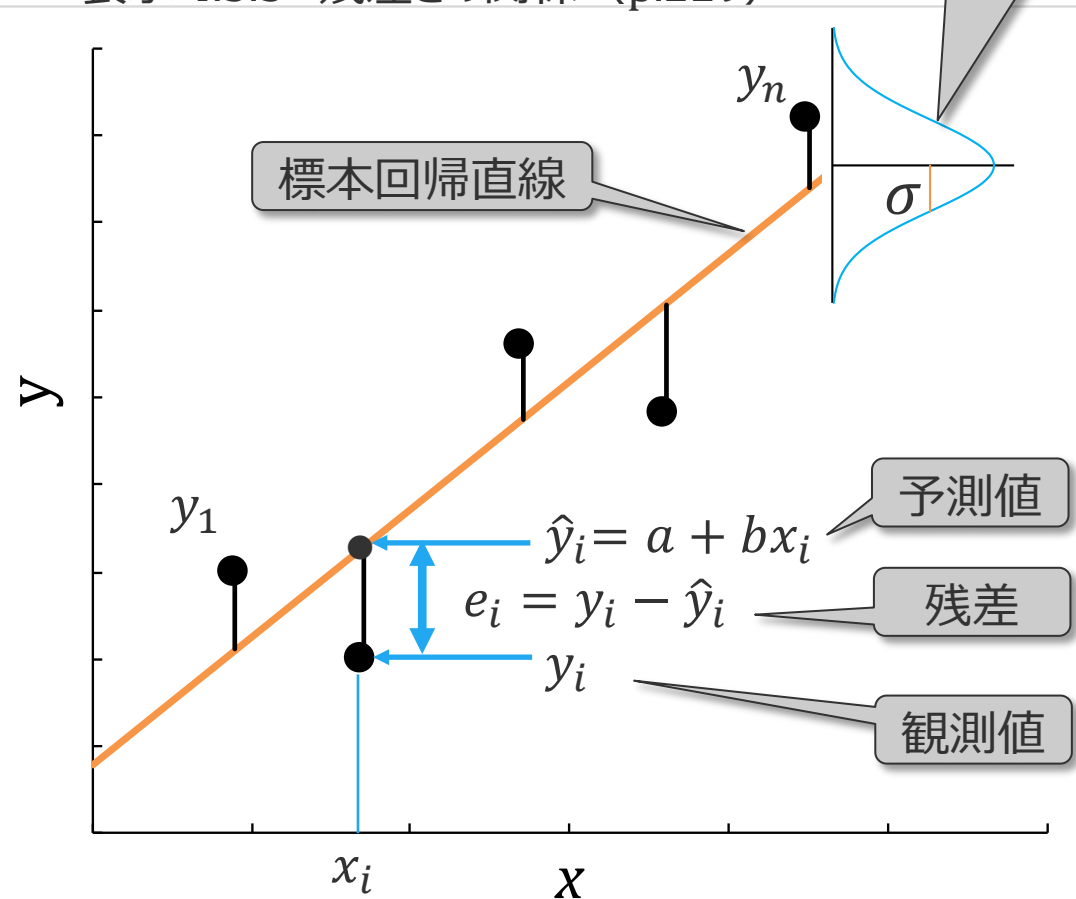
規則性や外れ値があれば要注意

グラフ化

xと残差の散布図、xと基準化残差の散布図

基準化残差の箱ひげ図など

表示 4.3.5 残差との関係 (p.229)



●残差の分散

単回帰モデル (母回帰式)

$$y_i = \alpha + \beta x_i + \varepsilon_i \quad \varepsilon_i \sim N(0, \sigma^2) \quad (4.3.3)$$

$$V[\varepsilon_i] = \sigma^2$$

標本回帰式 (母回帰式の推定式)

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

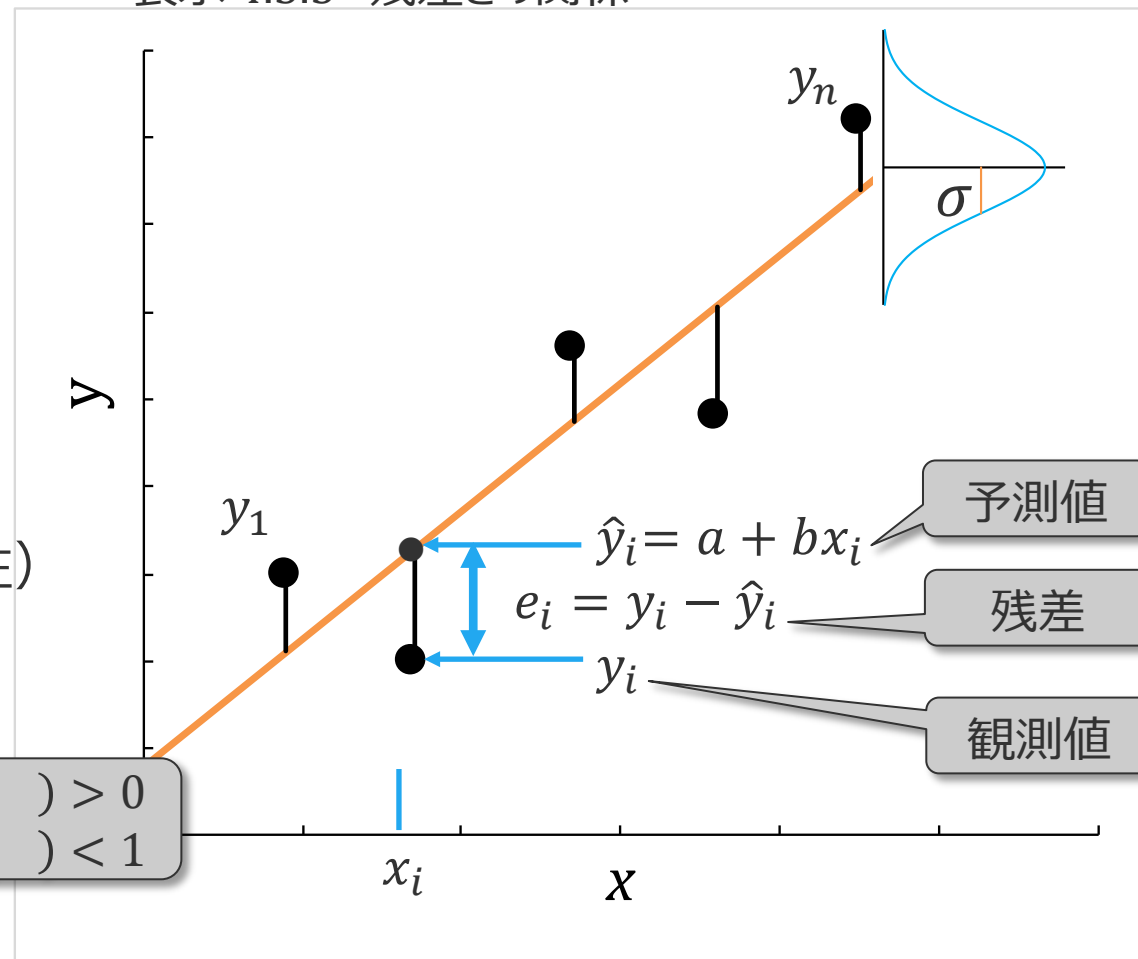
$$e_i = y_i - \hat{y}_i = y_i - \bar{y} - b(x - \bar{x}) \quad (\text{p.240})$$

$$V[e_i] = V[y_i - \bar{y} - b(x - \bar{x})] \quad (\text{p.249 脚注})$$

$$= \left(1 - \frac{1}{n} - \frac{(x_i - \bar{x})^2}{S_{xx}} \right) \sigma^2 \quad (4.5.1)$$

残差分散 $V[e_i]$ は誤差分散 σ^2 より小さい
 残差分散は x_i が \bar{x} から離れるほど小さい

表示 4.3.5 残差との関係



●残差の分散

単回帰モデル (母回帰式)

$$y_i = \alpha + \beta x_i + \varepsilon_i \quad \varepsilon_i \sim N(0, \sigma^2) \quad (4.3.3)$$

$$V[\varepsilon_i] = \sigma^2$$

誤差分散
誤差は独立
等分散

標本回帰式 (母回帰式の推定式)

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

$$e_i = y_i - \hat{y}_i = y_i - \bar{y} - b(x - \bar{x}) \quad (n.240)$$

$$V[e_i] = V[y_i - \bar{y} - b(x - \bar{x})]$$

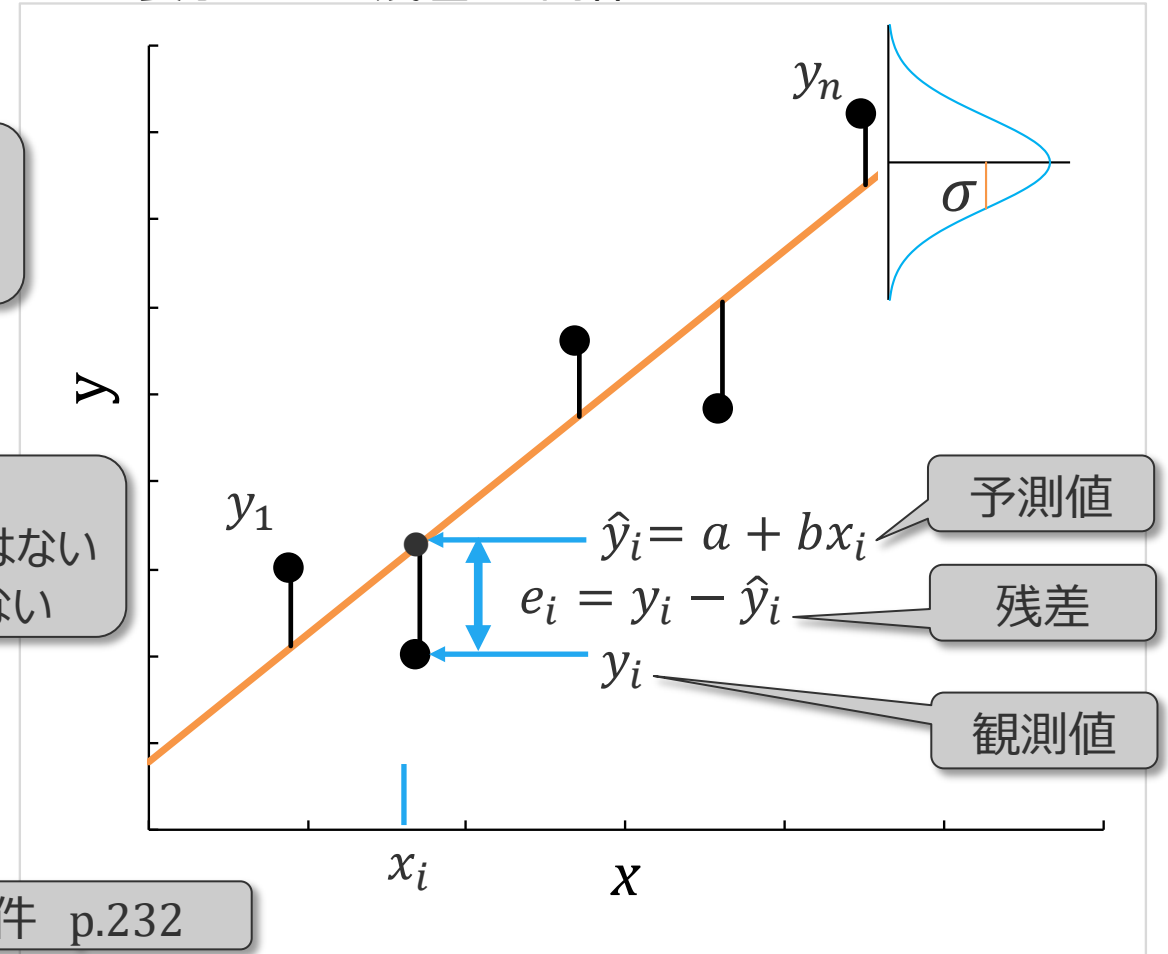
$$= \left(1 - \frac{1}{n} - \frac{(x_i - \bar{x})^2}{S_{xx}} \right) \sigma^2 \quad (4.5.1)$$

残差分散
残差は独立ではない
等分散ではない

$$\sum e_i = 0, \quad \sum e_i x_i = 0$$

残差の制約条件 p.232

表示 4.3.5 残差との関係



●残差の分散

残差 e_i の分散と観測値 y_i の分散の比較

$$V[\varepsilon_i] = \sigma^2 \quad (4.3.3)$$

$$V[e_i] = \left(1 - \frac{1}{n} - \frac{(x_i - \bar{x})^2}{S_{xx}}\right) \sigma^2 \quad (4.5.1)$$

$$V[\hat{y}_i] = \left(1 + \frac{1}{n} + \frac{(x_i - \bar{x})^2}{S_{xx}}\right) \sigma^2 \quad (4.4.5)$$

符号

符号

残差の分散は \bar{x} から離れるほど減少

観測値の分散は \bar{x} から離れるほど増加

残差分散は誤差分散 σ^2 より小さく、 \bar{x} からの距離によって変化するので等分散ではない

→ \bar{x} からの距離ごとに残差を比較するため、標準偏差で割って（基準化して）比較

平均値の残差（観測値と平均値の差）の分散（[§2.2](#) (4) p.78, [§2.7](#)(2) p.116)

平均値の残差の分散は
 σ^2 よりも小さくなる

$$V[(x_i - \mu)] = \sigma^2 \quad (2.2.1)$$

$$V[(x_i - \bar{x})] = V[e_i] = \frac{n-1}{n} \sigma^2 \quad (2.7.2)$$

●基準化残差（標準化残差）

基準化残差 e_i^* : 残差 e_i をその標準誤差 se_i で割って基準化（標準化）する

$$e_i = y_i - \hat{y}_i \quad (4.3.6)$$

$$e_i^* = \frac{e_i}{se_i} = \frac{e_i}{\sqrt{V[e_i]}}$$

$$V[e_i] = \left(1 - \frac{1}{n} - \frac{(x_i - \bar{x})^2}{S_{xx}} \right) \sigma^2 \quad (4.5.1)$$

利点 残差が平均 \hat{y}_i からどれだけ離れているかを、標準偏差の単位で公平に比較できる
基準化残差は標準正規分布に従うと期待されるので、

データポイント間の比較、外れ値の検出、モデルの妥当性検証などを行える
異なるスケールや単位を持つデータ間での比較が可能になる
残差プロットなどを作成して視覚的に残差の解析を行う

● 基準化残差（標準化残差）

$$V[e_i] = \left(1 - \frac{1}{n} - \frac{(x_i - \bar{x})^2}{S_{xx}}\right) \sigma^2 = \left(1 - \frac{1}{6} - \frac{(x_i - 5.0)^2}{50.0}\right) \times 0.834^2 \quad (4.5.1)$$

表示4.3.6 ソルバーの解（改変）

<i>i</i>	<i>x_i</i>	<i>y_i</i>	<i>y_i-hat</i>	<i>e_i</i>	<i>V[e_i]</i>	<i>se_i</i>	<i>e_i*</i>	<i> e_i*</i>
1	1	5	4.52	0.48	0.357	0.597	0.80	0.80
2	3	5	5.76	-0.76	0.524	0.724	-1.05	1.05
3	4	7	6.38	0.62	0.565	0.752	0.82	0.82
4	5	6	7.00	-1.00	0.579	0.761	-1.31	1.31
5	7	9	8.24	0.76	0.524	0.724	1.05	1.05
6	10	10	10.10	-0.10	0.232	0.481	-0.21	0.21
平均	5.0	7.0	7.00	0.00				
平方和	50.0	22.0						

事例 2

表示 4.3.8
LINEST関数の結果

	<i>x</i>	const	
<i>b</i>	0.620	3.900	<i>a</i>
<i>se</i>	0.118	0.681	<i>se</i>
<i>r²</i>	0.874	0.834	<i>sd</i>
<i>F</i>	27.655	4	<i>v</i>
<i>S_R</i>	19.220	2.780	<i>S_e</i>

残差の検討 (補足)

● 基準化残差 (標準化残差)

$$V[e_i] = \left(1 - \frac{1}{n} - \frac{(x_i - \bar{x})^2}{S_{xx}}\right) \sigma^2 = \left(1 - \frac{1}{6} - \frac{(x_i - 5.0)^2}{50.0}\right) \times 0.834^2 \quad (4.5.1)$$

$\hat{y}_i = 3.90 + 1.620x$

表示4.3.6 ソルバの解 (改変)

i	x_i	y_i	$y_i\text{-hat}$	e_i	$V[e_i]$	se_i	e_i^*	$ e_i^* $
1	1	5	4.52	0.48	0.357	0.597	0.80	0.80
2	3	5	5.76	-0.76	0.524	0.724	-1.05	1.05
3	4	7	6.38	0.62	0.565	0.752	0.82	0.82
4	5	6	7.00	-1.00	0.579	0.761	-1.31	1.31
5	7	9	8.24	0.76	0.524	0.724	1.05	1.05
6	10	10	10.10	-0.10	0.232	0.481	-0.21	0.21
	平均	5.0	7.0	7.00	0.00			
	平方和	50.0	22.0					

残差 $y_i - \hat{y}_i$

標準誤差 $\sqrt{V[e_i]}$

基準化残差 e_i/se_i

残差分散 $V[e_i]$

\bar{x}

S_{xx}

表示 4.3.8 LINEST関数の結果

標本回帰式 $\hat{y}_i = 3.90 + 0.620x_i$

$\sigma \sim$ 残差標準偏差

	x	const	
b	0.620	3.900	a
se	0.118	0.681	se
r^2	0.874	0.834	sd
F	27.655	4	v
S_R	19.220	2.780	S_e

残差の検討 (補足)

● 基準化残差 (標準化残差)

$$V[e_i] = \left(1 - \frac{1}{n} - \frac{(x_i - \bar{x})^2}{S_{xx}}\right) \sigma^2 = \left(1 - \frac{1}{6} - \frac{(x_i - 5.0)^2}{50.0}\right) \times 0.834^2 \quad (4.5.1)$$

残差分散 $V[e_i]$ 、標準誤差 se_i は \bar{x} から離れるほど小さくなる

標準誤差 < 残差標準偏差

表示4.3.6 ソルバーの解 (改変)

残差 $y_i - \hat{y}_i$ 標準誤差 $\sqrt{V[e_i]}$ 基準化残差 e_i/se_i

標本回帰式 $\hat{y}_i = 3.90 + 0.620x_i$ $\sigma \sim$ 残差標準偏差

i	x_i	y_i	$y_i - \hat{y}_i$	e_i	$V[e_i]$	se_i	e_i^*	$ e_i^* $
1	1	5	4.52	0.48	0.357	0.597	0.80	0.80
2	3	5	5.76	-0.76	0.524	0.724	-1.05	1.05
3	4	7	6.38	0.62	0.565	0.752	0.82	0.82
4	5	6	7.00	-1.00	0.579	0.761	-1.31	1.31
5	7	9	8.24	0.76	0.524	0.724	1.05	1.05
6	10	10	10.10	-0.10	0.232	0.481	0.21	0.21
平均	5.0	7.0	7.00	0.00				
平方和	50.0	22.0						

平均 \bar{x}

\bar{x}

S_{xx}

最大

残差標準誤差 0.834 より小
平均 \bar{x} から離れるほど減少

表示 4.3.8
LINEST関数の結果

	x	const	
b	0.620	3.900	a
se	0.118	0.681	se
r^2	0.874	0.834	sd
F	27.655	4	v
S_R	19.220	2.780	S_e

●基準化残差（標準化残差）

$$V[e_i] = \left(1 - \frac{1}{n} - \frac{(x_i - \bar{x})^2}{S_{xx}}\right) \sigma^2 = \left(1 - \frac{1}{6} - \frac{(x_i - 5.0)^2}{50.0}\right) \times 0.834^2 \quad (4.5.1)$$

x と残差 e_i の散布図、 x と基準化残差 e_i^* の散布図を描く

表示4.3.6 ソルバーの解（改変）

i	x_i	y_i	y_i -hat	e_i	$V[e_i]$	se_i	e_i^*	$ e_i^* $
1	1	5	4.52	0.48	0.357	0.597	0.80	0.80
2	3	5	5.76	-0.76	0.524	0.724	-1.05	1.05
3	4	7	6.38	0.62	0.565	0.752	0.82	0.82
4	5	6	7.00	-1.00	0.579	0.761	-1.31	1.31
5	7	9	8.24	0.76	0.524	0.724	1.05	1.05
6	10	10	10.10	-0.10	0.232	0.481	-0.21	0.21
平均	5.0	7.0	7.00	0.00				
平方和	50.0	22.0						

x と残差 e_i
の散布図

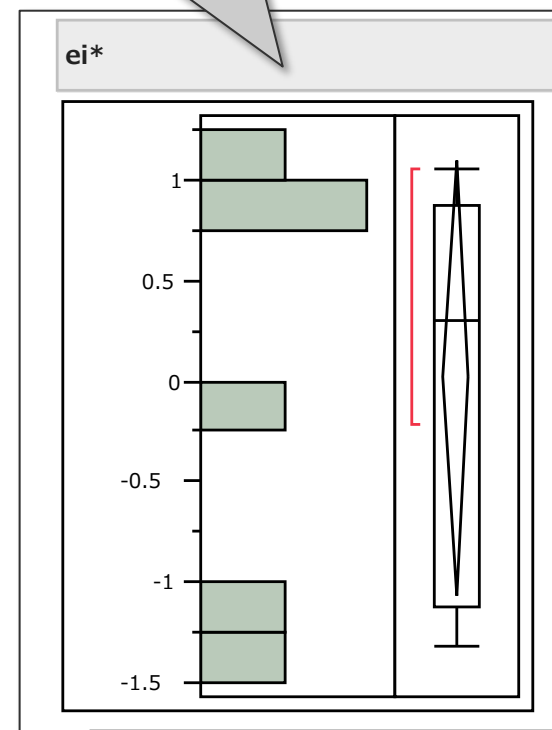
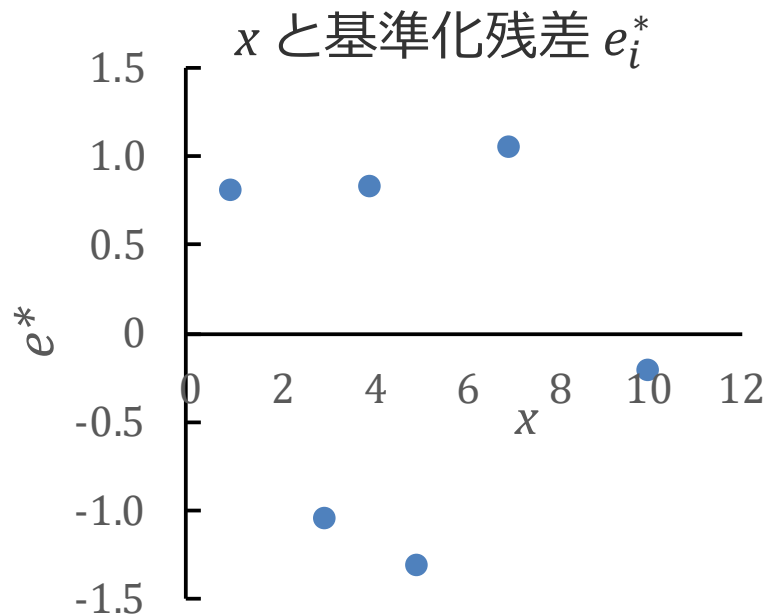
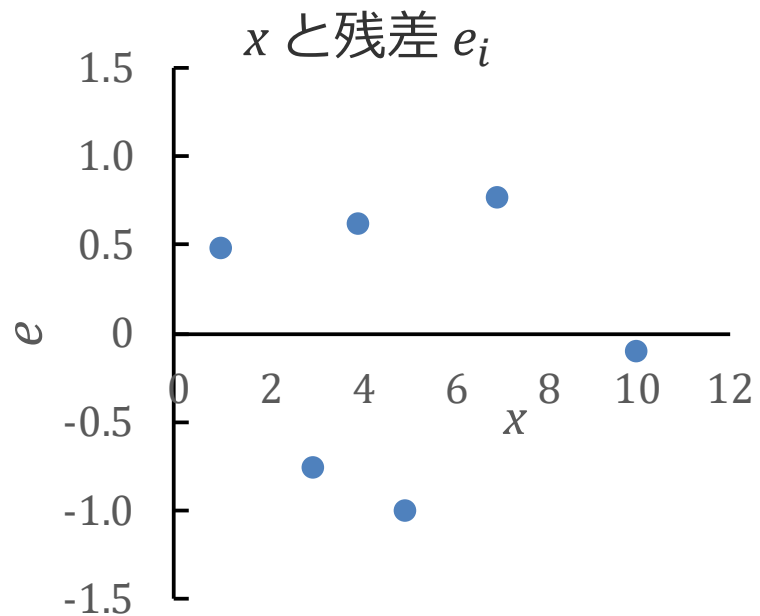
x と基準化残差 e_i^*
の散布図

● 基準化残差のグラフ化

x と残差 e_i の散布図、 x と基準化残差 e_i^* の散布図
周期性・規則性の有無、外れ値の有無を確認
回帰モデルのあてはめの妥当性を判断
(この事例はデータ不足)

基準化残差（標準化残差）
±2の範囲に基準化残差は含まれる
±3以上の基準化残差は外れ値として要確認

箱ひげ図で
外れ値などを確認



●基準化残差のグラフ化

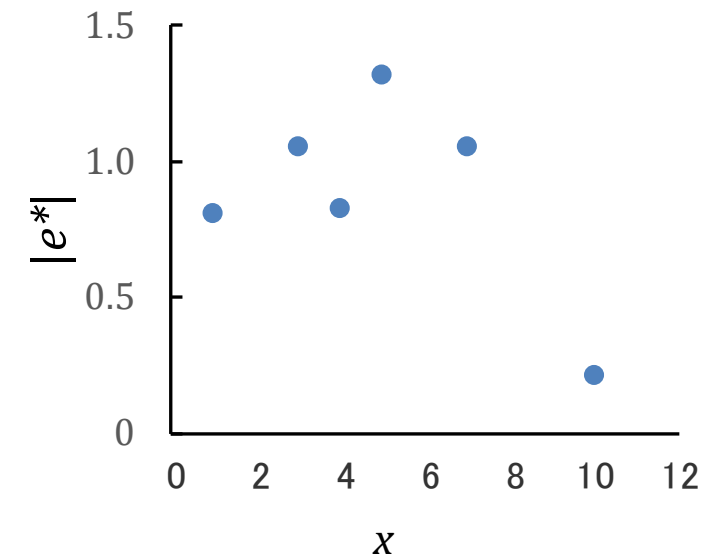
x と基準化残差の絶対値の散布図（Levene の検定の考え方の応用 [§3.3](#) p.143）

等分散性を確認（相関が有意であったり、規則性が見られると等分散が疑われる）

x の増加に伴いバラツキが大きくなるケースが少なくなく、この場合は正の相関関係になる

表示4.3.6 ソルバーの解（改変）

i	x	y	$y\text{-hat}$	e_i	$V[e_i]$	se	e_i^*	$ e_i^* $
1	1	5	4.52	0.48	0.357	0.597	0.80	0.80
2	3	5	5.76	-0.76	0.524	0.724	-1.05	1.05
3	4	7	6.38	0.62	0.565	0.752	0.82	0.82
4	5	6	7.00	-1.00	0.579	0.761	-1.31	1.31
5	7	9	8.24	0.76	0.524	0.724	1.05	1.05
6	10	10	10.10	-0.10	0.232	0.481	-0.21	0.21
平均	5.0	7.0	7.00	0.00				
平方和	50.0	22.0						



● $\hat{\eta}$ の標準誤差、 y の予測誤差

母平均の推定値の
分散と標準誤差

$$V[\hat{\eta}_i] = \left(\frac{1}{6} + \frac{(x_i - 5.0)^2}{50.0} \right) \times 0.834^2 \quad (4.4.3)$$

$$s.e. [\hat{\eta}_i] = \sqrt{V[\hat{\eta}_i]}$$

予測値の
分散と予測誤差

$$V[\hat{y}_i] = \left(1 + \frac{1}{6} + \frac{(x_i - 5.0)^2}{50.0} \right) \times 0.834^2 \quad (4.4.5)$$

$$s.e._{pred} [\hat{y}_i] = \sqrt{V[\hat{y}_i]}$$

表示4.3.6 ソルバーの解 (改変)

i	x	y	$y\text{-hat}$	e_i	$V[e_i]$	se	e_i^*	$ e_i^* $	$V[\eta_i\text{-hat}]$	se	$V[y_i]$	$se (pred)$
1	1	5	4.52	0.48	0.357	0.597	0.80	0.80	0.338	0.58	1.033	1.02
2	3	5	5.76	-0.76	0.524	0.724	-1.05	1.05	0.171	0.41	0.866	0.93
3	4	7	6.38	0.62	0.565	0.752	0.82	0.82	0.130	0.36	0.825	0.91
4	5	6	7.00	-1.00	0.579	0.761	-1.31	1.31	0.116	0.34	0.811	0.90
5	7	9	8.24	0.76	0.524	0.724	1.05	1.05	0.171	0.41	0.866	0.93
6	10	10	10.10	-0.10	0.232	0.481	-0.21	0.21	0.463	0.68	1.158	1.08
平均	5.0	7.0	7.00	0.00								
平方和	50.0	22.0										

$s.e. [\hat{\eta}_i] = \sqrt{V[\hat{\eta}_i]}$

$s.e._{pred} [\hat{y}_i] = \sqrt{V[\hat{y}_i]}$

●JMPによる基準化残差の計算

JMP ファイル「4-相関3.jmp」を読み込み

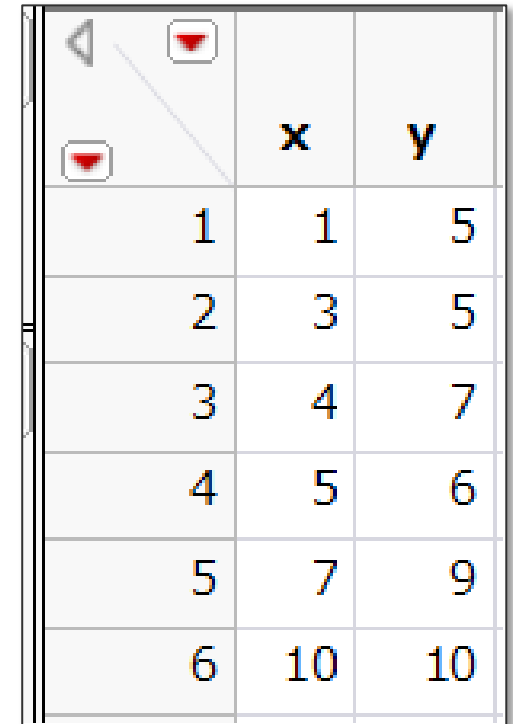
6組の x と y のデータ（表示4.3.6、表示4.4.3 のデータ）

JMP [二変量の関係]、[モデルのあてはめ]

表示4.3.6 ソルバーの解（改変）

i	x	y	$y\text{-hat}$	e_i	$V[e_i]$	se	e_i^*	$ e_i^* $
1	1	5	4.52	0.48	0.357	0.597	0.80	0.80
2	3	5	5.76	-0.76	0.524	0.724	-1.05	1.05
3	4	7						
4	5	6	7.00	-1.00	0.579	0.761	-1.31	1.31
5	7	9	8.24	0.76	0.524	0.724	1.05	1.05
6	10	10	10.10	-0.10	0.232	0.481	-0.21	0.21
平均	5.0	7.0	7.00	0.00				
平方和	50.0	22.0						

「4-相関3.jmp」のデータ



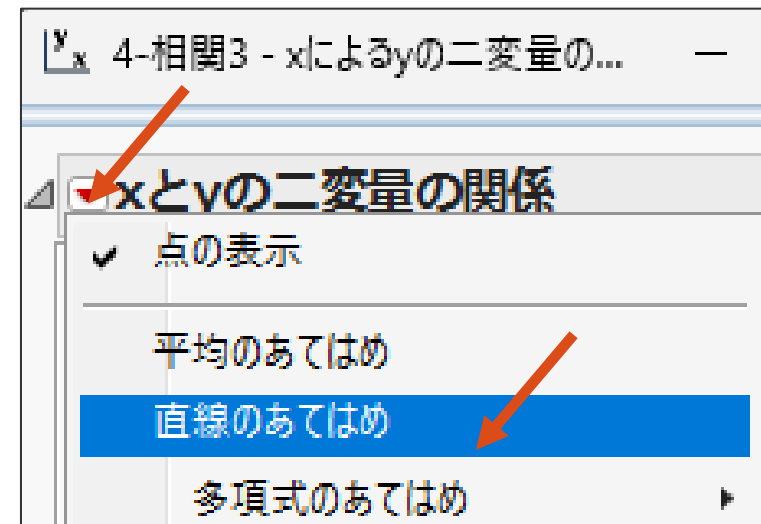
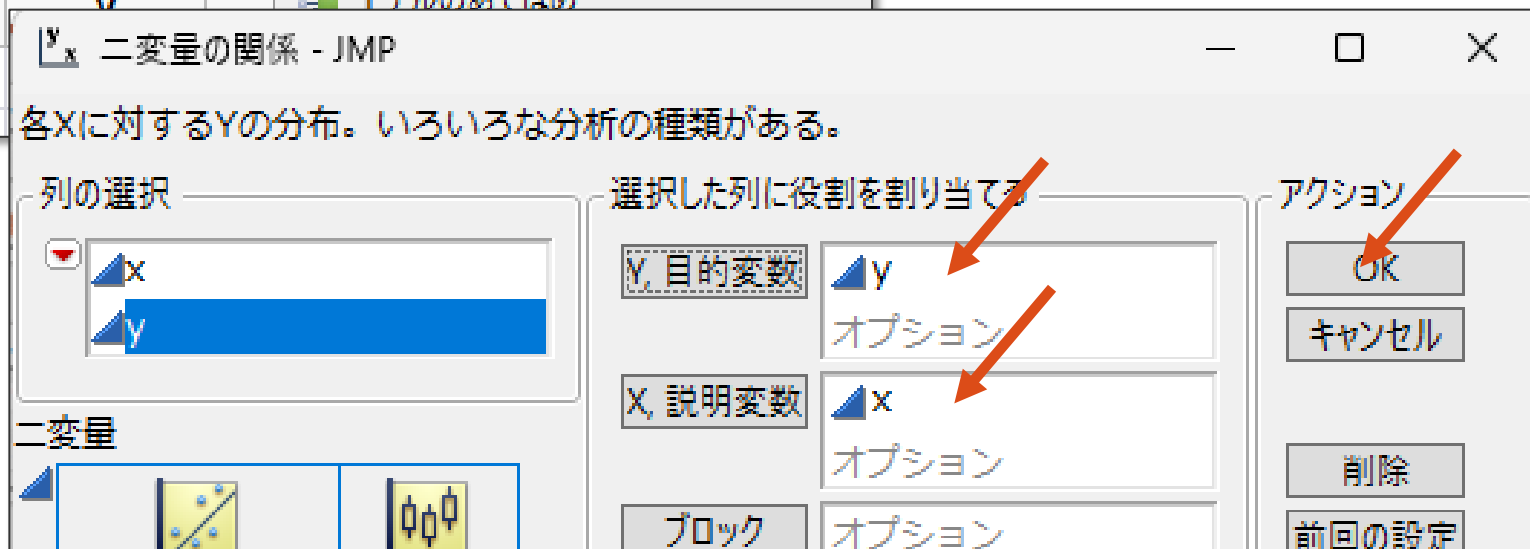
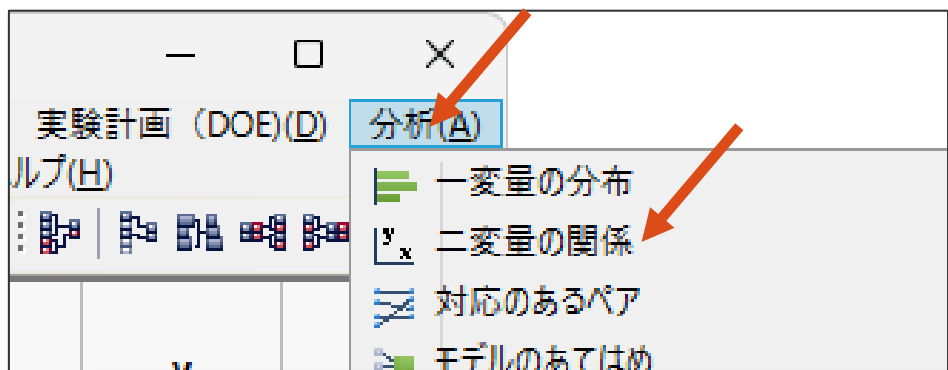
	x	y
1	1	5
2	3	5
3	4	7
4	5	6
5	7	9
6	10	10

●JMPによる基準化残差の計算：〔二変量の関係〕

〔分析〕 > 〔二変量の関係〕 > ダイアログボックス > ▼ 〔直線のあてはめ〕

〔Y、目的変数〕 : y

〔X、説明変数〕 : x



●JMP による基準化残差の計算： [二変量の関係]

JMP [二変量の関係]

散布図を描画

[直線のアてはめ]

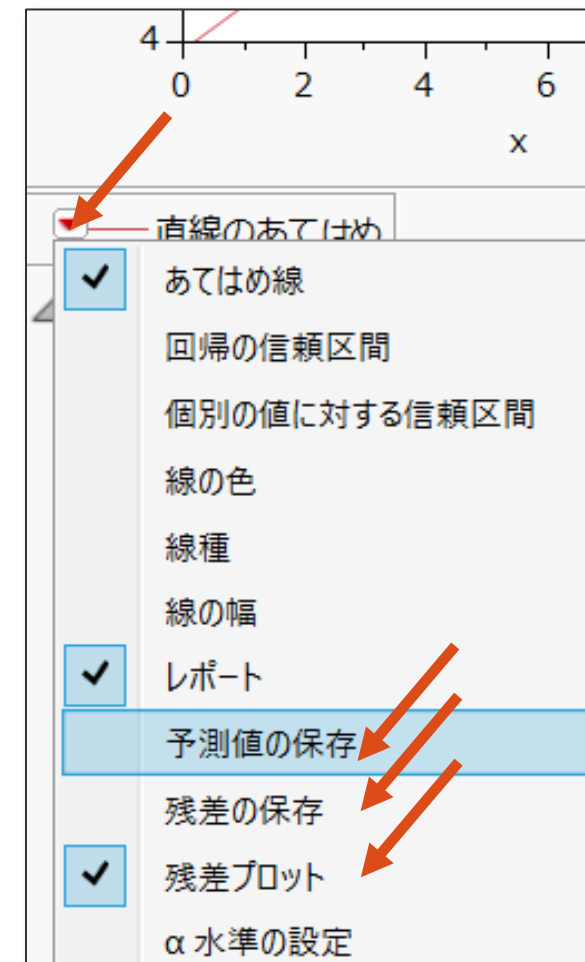
▼オプション> [予測値の保存]

▼オプション> [残差の保存]

▼オプション> [残差プロット]

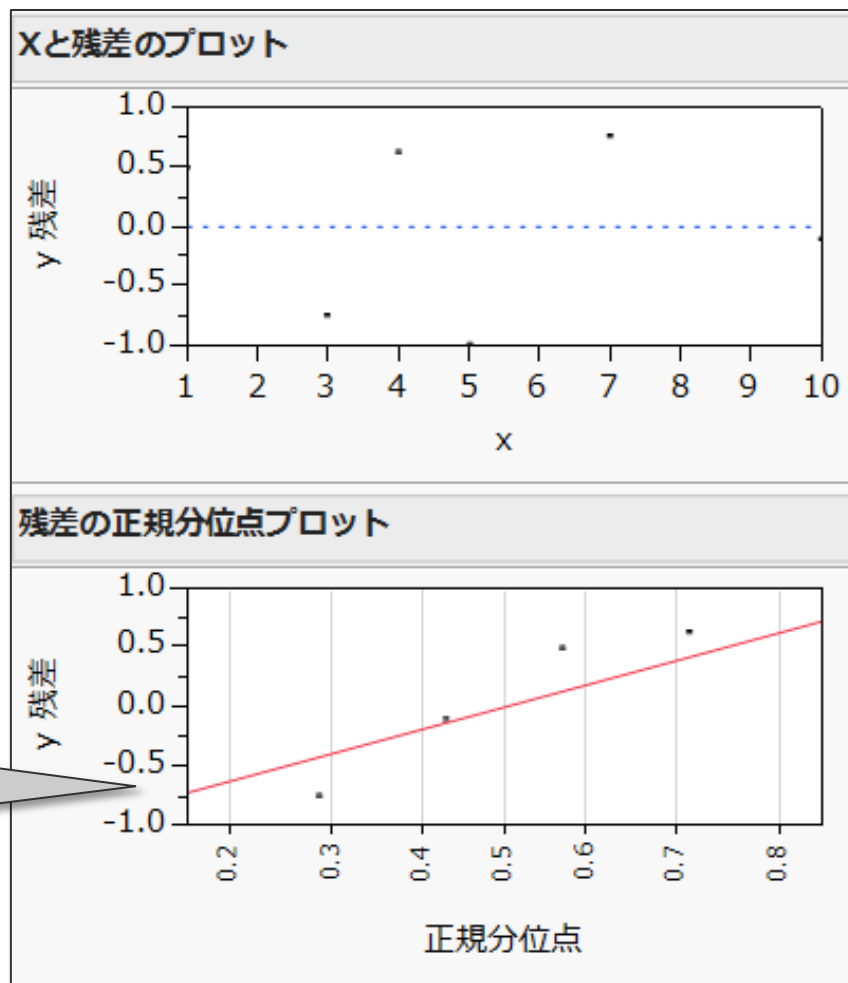
データテーブルに
新規の列が生成

	x	y	予測値 y	残差 y
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	-1
5	7	9	8.24	0.76
6	10	10	10.1	-0.1

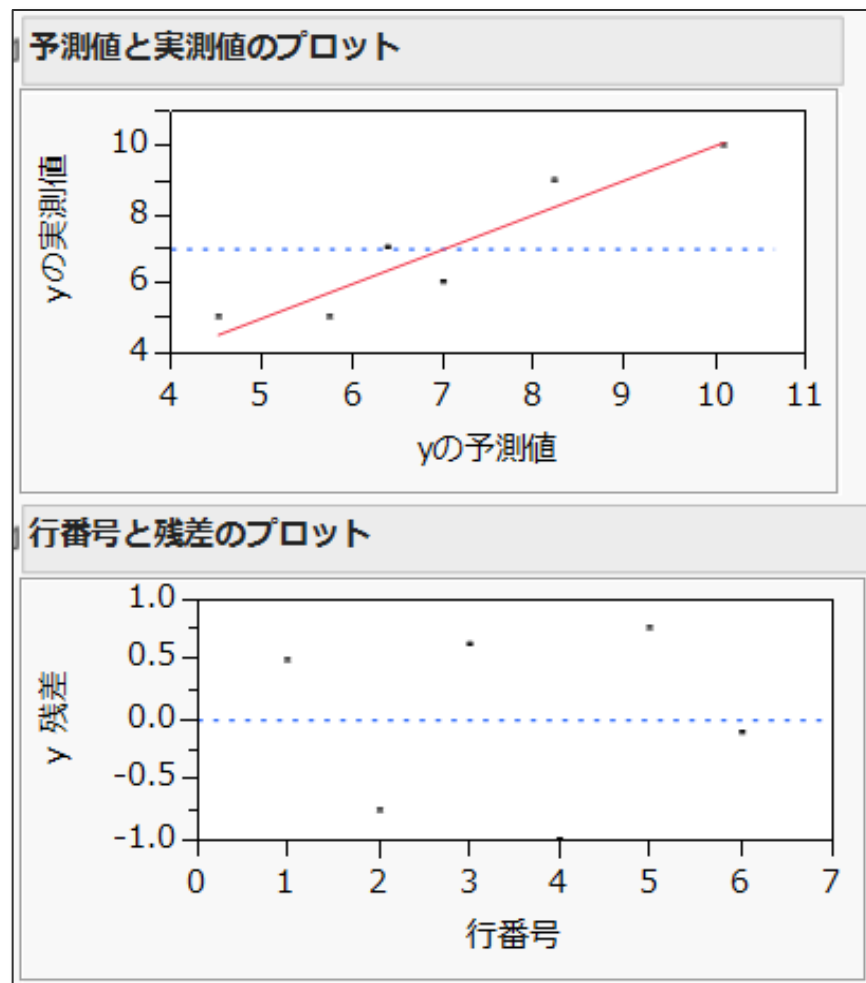


●JMPによる基準化残差の計算：[二変量の関係]

残差プロット



残差に正規性があれば、直線に並ぶ

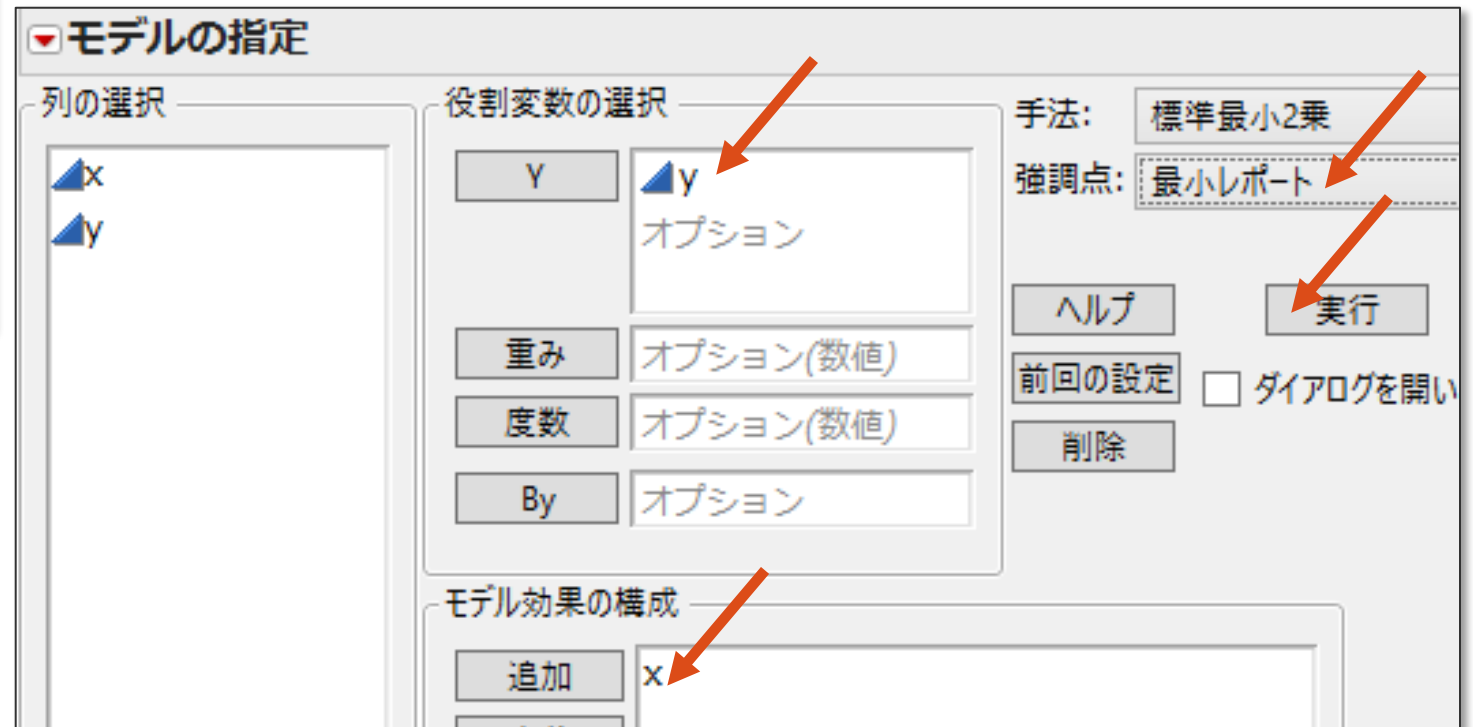
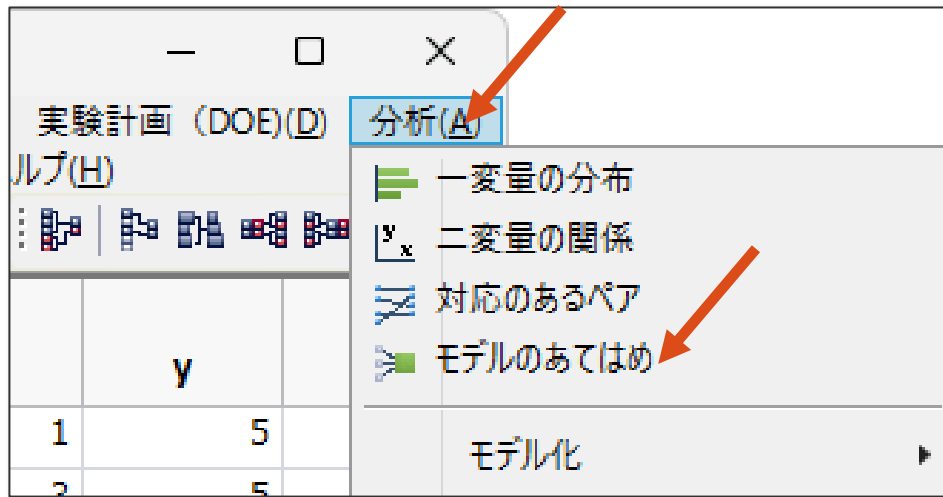


●JMPによる基準化残差の計算： [モデルのあてはめ]

[分析] > [モデルのあてはめ] > ダイアログボックス

[Y] : y

[モデルの効果の構成] : x



残差の検討（補足）

- JMP による基準化残差の計算： [モデルのあてはめ]

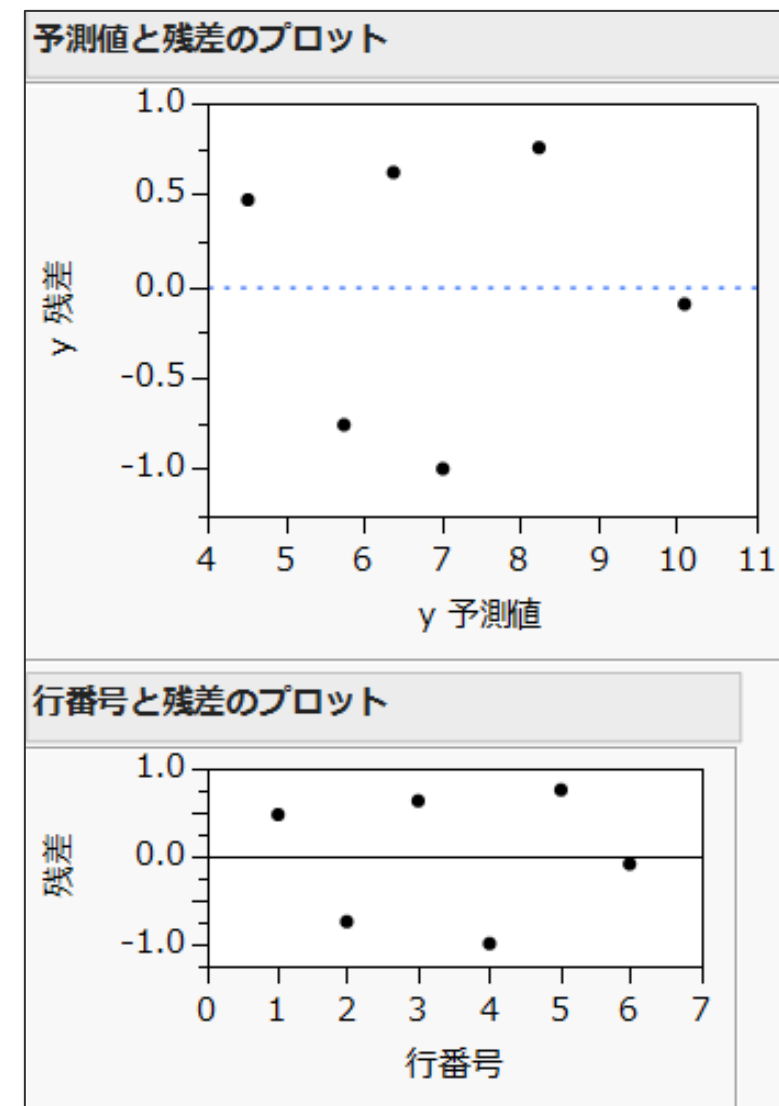
Model Diagnostics menu:

- 回帰レポート
- 推定値
- 要因のスクリーニング
- 因子プロファイル
- 行ごとの診断統計量
- 列の保存
- モデルダイアログ
- スクリプト

Plot options (checked):

- 回帰プロット
- 予測値と実測値のプロット
- 効果てこ比のプロット
- 予測値と残差のプロット
- 行番号と残差のプロット

項	推定値	標準誤差
切片	3.9	0.680686
x	0.62	0.117898



残差の検討（補足）

残差 y

- 回帰レポート
- 推定値
- 要因のスクリーニング
- 因子プロファイル
- 行ごとの診断統計量
- 列の保存
- モデルダイアログ
- スクリプト

分散分析

要因	自由度	
モデル	1	19.2
誤差	4	2.7
全体(修正済み)	5	22.0

予測式

- 予測値
- 残差
- 平均の信頼区間
- 個別の信頼区間
- スチューデント化残差
- ハット
- 予測値の標準誤差
- 残差の標準誤差
- 個々の標準誤差

	x	y	予測値 y	残差 y	スチューデント化残差	予測値の標準誤差 y	残差の標準誤差 y	個々の標準誤差 y
1	1	5	4.52	0.48	0.80361697	0.5815783	0.597299	1.0164808
2	3	5	5.76	-0.76	-1.0503344	0.4140450	0.723579	0.9308240
3	4	7	6.38	0.62	0.82464114	0.3601851	0.751842	0.9081482
4	5	6	7	-1	-1.3140086	0.3403429	0.761030	0.9004628
5	7	9	8.24	0.76	1.0503344	0.4140450	0.723579	0.9308240
6	10	10	10.1	-0.1	-0.2077630	0.6806859	0.481317	1.0762589

スチューデント化残差

基準化残差を改良して補正した残差

内部スチューデント化残差：すべてのデータを使用
JMP の出力

外部スチューデント化残差：外れ値と考えられる
データを除いて算出

残差の検討 (補足)

残差
 $y_i - \hat{y}_i$

標準誤差
 $\sqrt{V[e_i]}$

	x	y	予測値 y	残差 y	スチューデント化残差	予測値の標準誤差 y	残差の標準誤差 y	個々の標準誤差 y
1	1	5	4.52	0.48	0.8036169	0.5815783	0.597299	1.0164808
2	3	5	5.76	-0.76	-1.0503344	0.4140450	0.723579	0.9308240
3	4	7	6.38	0.62	0.8246411	0.3601851	0.751842	0.9081482
4	5	6	7	-1	-1.3140086	0.3403429	0.761030	0.9004628
5	7	9	8.24	0.76	1.0503344	0.4140450	0.723579	0.9308240
6	10	10	10.1	-0.1	-0.2077630	0.6806859	0.481317	1.0762589

表示4.3.6 ソルバーの解 (改変)

i	x	y	y-hat	e_i	$V[e_i]$	se	e_i^*	$ e_i^* $	$V[\eta_i\text{-hat}]$	se	$V[y_i\text{-hat}]$	se (pred)
1	1	5	4.52	0.48	0.357	0.597	0.80	0.80	0.338	0.58	1.033	1.02
2	3	5	5.76	-0.76	0.524	0.724	-1.05	1.05	0.171	0.41	0.866	0.93
3	4	7	6.38	0.62	0.565	0.752	0.82	0.82	0.130	0.36	0.825	0.91
4	5	6	7.00	-1.00	0.579	0.761	-1.31	1.31	0.116	0.34	0.811	0.90
5	7	9	8.24	0.76	0.524	0.724	1.05	1.05	0.171	0.41	0.866	0.93
6	10	10	10.10	-0.10	0.232	0.481	-0.21	0.21	0.463	0.68	1.158	1.08
平均	5.0	7.0	7.00	0.00								
平方和	50.0	22.0										



(3) x に誤差を含む場合

単回帰分析を適用する妥当性



x に誤差を含む回帰分析

●事例 3

血圧測定の効率化を図るために、医師による従来法の代わりに自動血圧計が使えるか確かめる
 50人の被験者の血圧を、医師による従来法 (x) と自動血圧計 (y) で同時に測定
 x 、 y の回帰式を得る ($n = 50$)

原点を通り、回帰係数 (傾き) が 1 であれば、
 自動血圧計を使うことに問題はないと考えられる

$$y = \alpha + \beta x = 0.0 + 1.0x \quad (4.5.2)$$

サンプル ($n = 50$) から推定された回帰式

$$y = a + bx = 25.198 + 0.7922x$$

回帰式は原点を通らず、回帰係数は 1 以下、
 自動血圧計が使えると判断できるか？
 単回帰分析で解析する

50 人の血圧測定値 (ランダムに測定)

被験者	従来法 (x)	自動血圧計 (y)
1	94.7	102.8
2	112.0	114.1
3	128.7	123.7
4	124.1	117.0
5	101.3	90.7
⋮	⋮	⋮
⋮	⋮	⋮
⋮	⋮	⋮
48	130.4	122.2
49	103.8	106.9
50	155.4	147.8

●事例 3 の単回帰分析

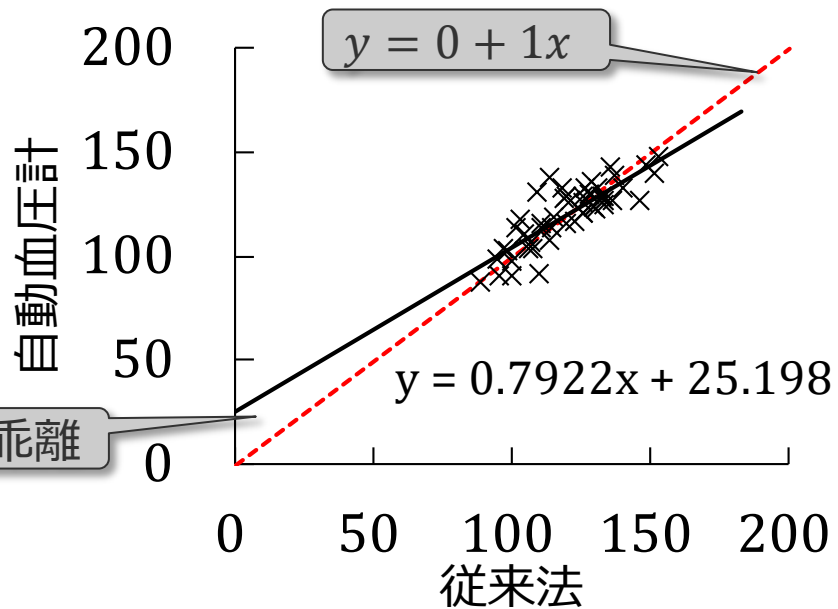
$\beta = 0$ ではないことに注意

母回帰係数と母切片の仮説検定 ($H_0 : \beta = 1$, $H_0 : \alpha = 0$) はいずれも有意 (有意水準 $\alpha = 0.05$)
 原点を通らず, 回帰係数は1より小さい

意味が異なる

→ この単回帰分析には、基本的な間違いがある

表示4.5.3 散布図と回帰分析 (1)



	x	const	
回帰係数	0.792	25.198	切片
その標準誤差	0.074	9.009	その標準誤差
寄与率	0.702	8.227	残差標準偏差
F比	113	48	残差自由度
回帰平方和	7653	3249	残差平方和
t 値	-2.789	2.797	
p 値	0.008	0.007	

$25.198 / 9.009 = 2.797$

$(0.792 - 1) / 0.074 = -2.789$

●単回帰分析における誤差の扱い

単回帰モデルでは y のみに誤差 ε を考慮

$$\eta = \alpha + \beta x \quad (4.3.2)$$

$$y = \eta + \varepsilon$$

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

$$\varepsilon \sim N(0, \sigma^2) \quad (\S 4.3)$$

y には
誤差 ε がある

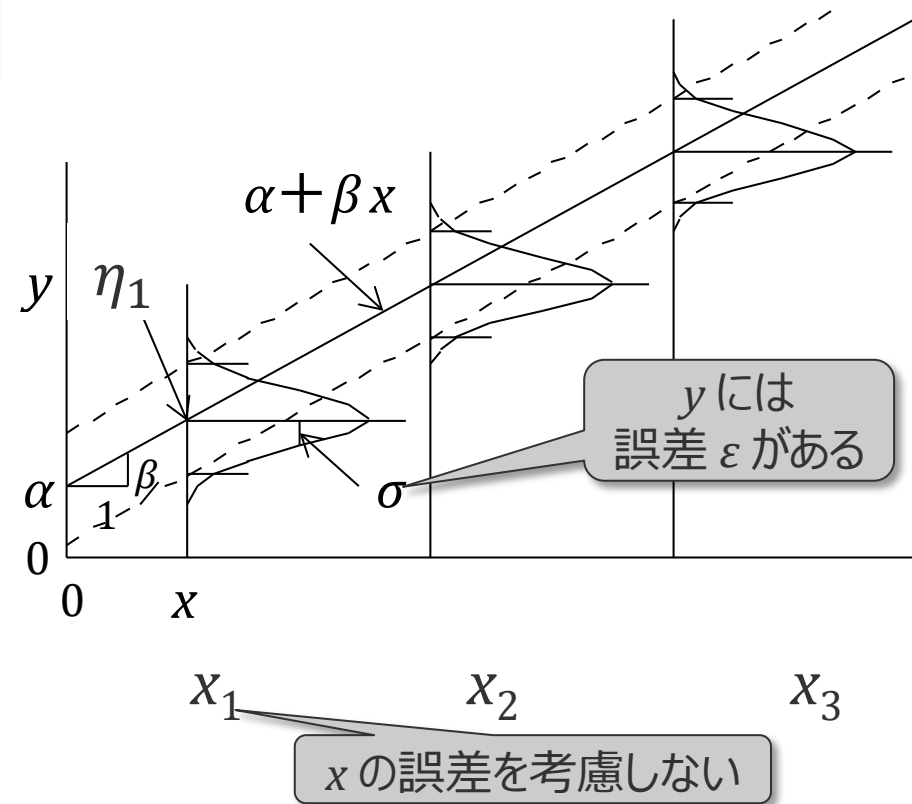
事例 3 には x, y の両方に同様の誤差がある

x : 医師の従来法による血圧の測定値

y : 自動血圧計による血圧の測定

y と同様の誤差が x にも含まれるので、
この事例に単回帰モデルをはてはめることは
妥当ではない

表示 4.3.2 回帰モデル (p.227)



x に誤差を含む回帰分析

● 回帰分析に及ぼす x と y の標準誤差の影響 (シミュレーション)

事例 4 (演習4.5.1)

Excel ファイル「基礎改4.xls」、名前ボックスで「表示4.5.3」を選択、セル「M1」に移動
50 人の血圧を、医師による従来法 (x) と自動血圧計 (y) で同時に測定するシミュレーション

	M	N	O	P	Q	R	S	T	U	V	W	X	Y	Z	AA	AB
1				μ	120	σ_x	σ_y									
2				σ	15	7.0	5.0									
3																
4																
5																
6																
7					x(真値)	x	y	平均	差	d						
8	-0.75	-1.56	-1.20	1	109	97.8	102.8	100.3	-4.9	-19.8						
9	0.14	-1.11	-1.59	2	122	114.2	114.1	114.2	0.2	-5.9						
10	0.68	-0.16	-1.29	3	130	129.0	123.7	126.4	5.3	6.3						
11	-0.25	0.88	0.16	4	116	122.3	117.0	119.7	5.3	-0.4						
12	-1.75	0.84	-0.60	5	94	99.6				4.9						
13	-0.92	-0.69	1.46	6	106	101.3				2.7						
14	-2.05	-0.18	-0.40	7	89	89.0				2.4						
15	-1.14	1.00	-2.34	8	103	109.9	91.2	100.5	18.7	-19.5						

	x	const	
回帰係数	0.792	25.198	切片
その標準誤差	0.074	9.009	その標準誤差
寄与率	0.702	8.227	残差標準偏差
F比	113	48	残差自由度
回帰平方和	7653	3249	残差平方和
t 値	-2.789	2.797	
p 値	0.008	0.007	

表示4.5.3 散布図と回帰分析(1)

x に誤差を含む回帰分析

●回帰分析に及ぼす x と y の標準誤差の影響 (シミュレーション)

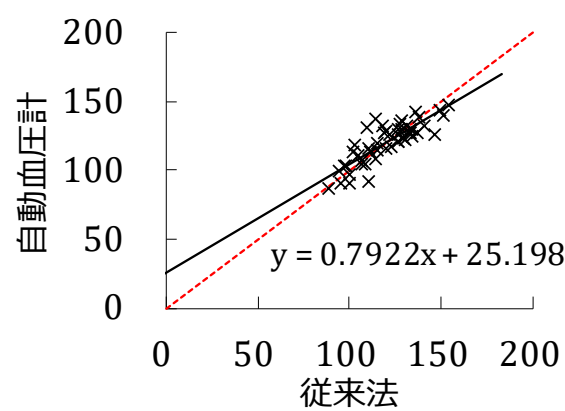
標準正規乱数 → x(真値) → σ_x の測定誤差を付加 → x (医師の従来法による測定値)
 → σ_y の測定誤差を付加 → y (自動血圧計による測定値)

回帰分析で $y = x$ が期待される

50 人の真の血圧 $N(120, 15^2)$ に従う乱数で設定

	M	N	O	P	Q	R	S	T	U	V	W	X	Y	Z	AA	AB
1					μ 120	σ_x	σ_y									
2					σ 15	7.0	5.0									
3	被験者番号															
4																
5	標準正規乱数															
6					x(真値)	x	平均	差	d							
7	-0.75	-1.56	-1.20	1	109	97.8	102.8	100.3	-4.9	-19.8						
8	0.14	-1.11	-1.59	2	122	114.2	114.1	114.2	0.2	-5.9						
9	0.68	-0.16	-1.29	3	130	129.0	123.7	126.4	5.3	6.3						
10	-0.25	0.88	0.16	4	116	122.3	117.0	119.7	5.3	-0.4						
11	-1.75	0.84	-0.60	5	94	99.6	90.7	95.1	8.9	-24.9						
12	-0.92	-0.69	1.46	6	106	101.3	113.5	107.4	-12.1	-12.7						
13	-2.05	-0.18	-0.40	7	89	88.0	87.3	87.6	0.7	-32.4						
14	-1.14	1.00	-2.34	8	103	109.9	91.2	100.5	18.7	-19.5						

$x(\text{真値}) = 120 + 15 \times \text{標準正規乱数}$
 (50 人の真の血圧とする)



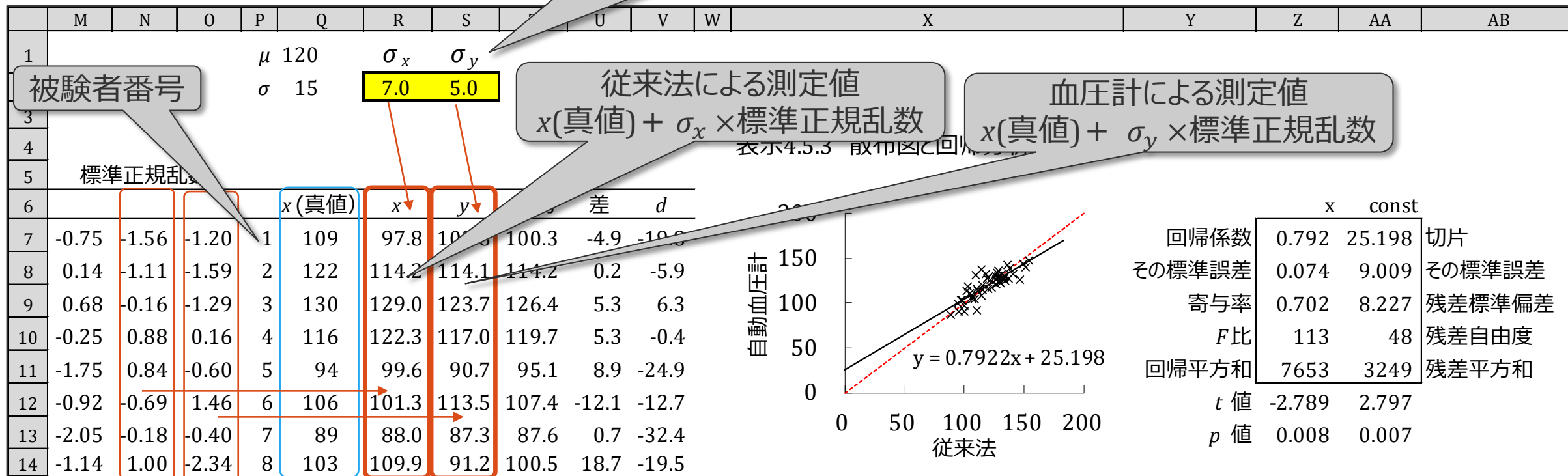
	x	const	
回帰係数	0.792	25.198	切片
その標準誤差	0.074	9.009	その標準誤差
寄与率	0.702	8.227	残差標準偏差
F比	113	48	残差自由度
回帰平方和	7653	3249	残差平方和
t 値	-2.789	2.797	
p 値	0.008	0.007	

x に誤差を含む回帰分析

●回帰分析に及ぼす x と y の標準誤差の影響 (シミュレーション)

標準正規乱数 → x(真値) → σ_x の測定誤差を付加 → x (医師の従来法による測定値)
 → σ_y の測定誤差を付加 → y (自動血圧計による測定値)

回帰分析で $y = x$ が期待される



x に誤差を含む回帰分析

- 回帰分析に及ぼす x と y の標準誤差の影響 (シミュレーション)
50 人の被験者について、それぞれの血圧の真値「x(真値)」を設定

	M	N	O	P	Q	R	S	T	U	V
1				μ 120		σ_x	σ_y			
2				σ 15		7.0	5.0			
3	標準正規乱数									
6					x(真値)	x	y	平均		
7	-0.75	-1.56	-1.20	1	109	97.8	102.8	100.3	-4.9	-19.8
8	0.14	-1.11	-1.59	2	122	114.2	114.1	114.2	0.2	-5.9
9	0.68	-0.16	-1.29	3	130	129.0	123.7	126.4	5.3	6.3
10	-0.25	0.88	0.16	4	116	122.3	117.0	119.7	5.3	-0.4
	0.75	0.84	-0.60	5	94	99.6	90.7	95.1	8.9	-24.9
	0.69	1.46		6	106	101.3	113.5	107.4	-12.1	-12.7
	0.18	-0.40		7	89	88.0	87.3	87.6	0.7	-32.4
	0.00	-2.34		8	103	109.9	91.2	100.5	18.7	-19.5
	0.62	2.03		9	122	118.0	132.5	125.2	-14.5	5.2

50 人の x(真値) 平均, 標準偏差を設定

標準正規乱数 (分析ツール)
M列 → 50人の x(真値)
N列 → x の測定誤差
O列 → y の測定誤差

$Q7 := \$Q\$1 + \$Q\$2 * M7$
x(真値) := 120 + 15 × 標準正規乱数

x に誤差を含む回帰分析

●回帰分析に及ぼす x と y の標準誤差の影響 (シミュレーション)

2つ測定法による測定値 (x,y) として、x(真値)にそれぞれ異なる測定誤差を付加、平均は同一

	M	N	O	P	Q	R	S	T	U	V
1				μ 120		σ_x	σ_y			
2				σ 15		7.0	5.0			
6	標準正規乱数				x(真値)	x	y	平均		
7	-0.75	-1.56	-1.20	1	109	97.8	102.8			
8	0.14	-1.11	-1.59	2	122	114.2	114.1	114.2		
	0.68	-0.16	-1.29	3	130	129.0	123.7	126.4		
	0.25	0.88	0.16	4	116	122.3	117.0	119.7		
	0.75	0.84	-0.60	5	94	99.6	90.7	95.1		
	0.92	-0.69	1.46	6	106	101.3	113.5	107.4		
	0.05	-0.18	-0.40	7	89	88.0	87.3	87.6		
	1.14	1.00	-2.34	8	103	109.9	91.2			
15	0.16	-0.62	2.03	9	122	118.0	132.5	125.2	-14.5	5.2

50人のx(真値)
平均,標準偏差を設定

標準正規乱数
M列→50人のx(真値)
N列→xの測定誤差
O列→yの測定誤差

x と y に異なる測定誤差を付加

Q7 := \$Q\$1 + \$Q\$2 * M7
x(真値) := 120 + 15 × 標準正規乱数
R7 := ROUND(\$Q7 + R\$2 * N7, 2)
x : 平均(真値) + σ_x × 標準正規乱数
S7 := ROUND(\$Q7 + S\$2 * O7, 2)
y : 平均(真値) + σ_y × 標準正規乱数

=ROUND(,2) : 小数点以下第3位を四捨五入

x に誤差を含む回帰分析

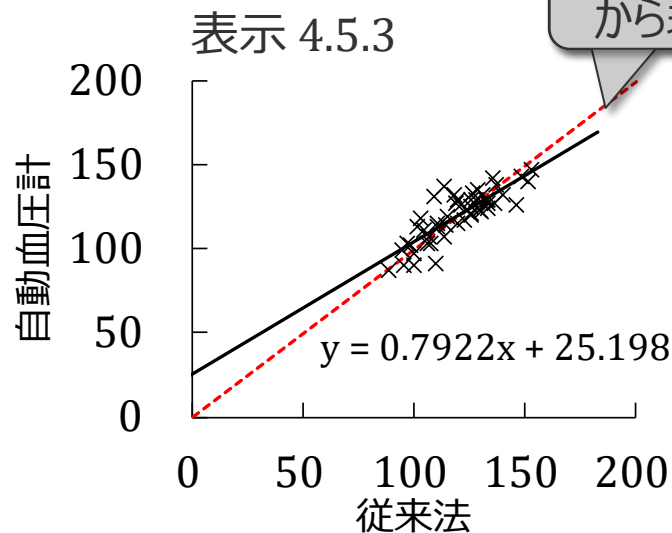
● x と y の標準偏差の影響

R2, S2

σ_x	σ_y
7.0	5.0

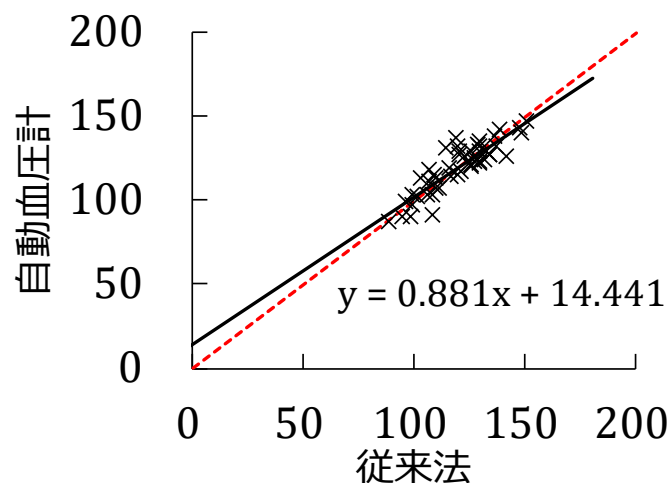
$H_0 : \beta=1$ 、 $H_0 : \alpha=0$ は棄却
当初の予想とは異なる結果

σ_x	σ_y
5.0	5.0



	x	const	
回帰係数	0.792	25.198	切片
その標準誤差	0.074	9.009	その標準誤差
寄与率	0.702	8.227	残差標準偏差
F比	113	48	残差自由度
回帰平方和	7653	3249	残差平方和
t 値	-2.789	2.797	
p 値	0.008	0.007	

$\alpha=0.05$
有意



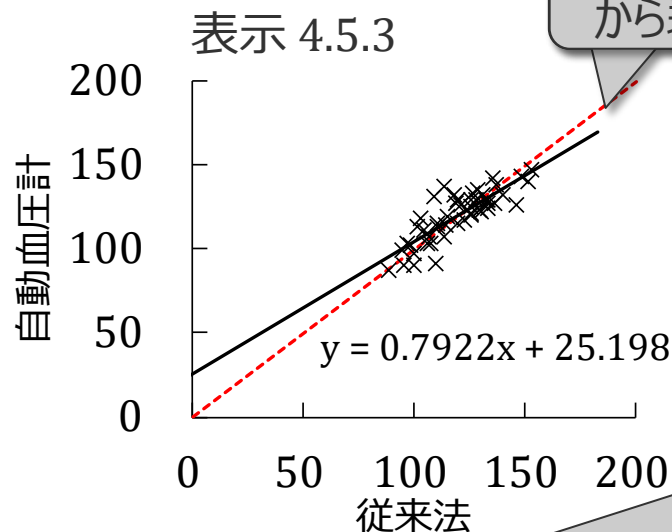
	x	const	
回帰係数	0.881	14.441	切片
その標準誤差	0.067	8.105	その標準誤差
寄与率	0.783	7.026	残差標準偏差
F比	173	48	残差自由度
回帰平方和	8533	2369	残差平方和
t 値	-1.776	1.782	
p 値	0.082	0.081	

x に誤差を含む回帰分析

● x と y の標準偏差の影響

R2, S2	σ_x	σ_y
	7.0	5.0

$H_0 : \beta=1$ 、 $H_0 : \alpha=0$ は棄却
当初の予想とは異なる結果



	x	const	
回帰係数	0.792	25.198	切片
その標準誤差	0.074	9.009	その標準誤差
寄与率	0.702	8.227	残差標準偏差
F比	113	48	残差自由度
回帰平方和	7653	3249	残差平方和
t 値	-2.789	2.797	
p 値	0.008	0.007	

$\alpha=0.05$
有意

$H_0 : \beta=0$ の検定 (§3.2 p.137)
(回帰係数 - 0) / 標準誤差 = $(0.792 - 0) / 0.074$

$H_0 : \beta=1$ の検定・・・今回の帰無仮説
(回帰係数 - 1) / 標準誤差 = $(0.792 - 1) / 0.074$

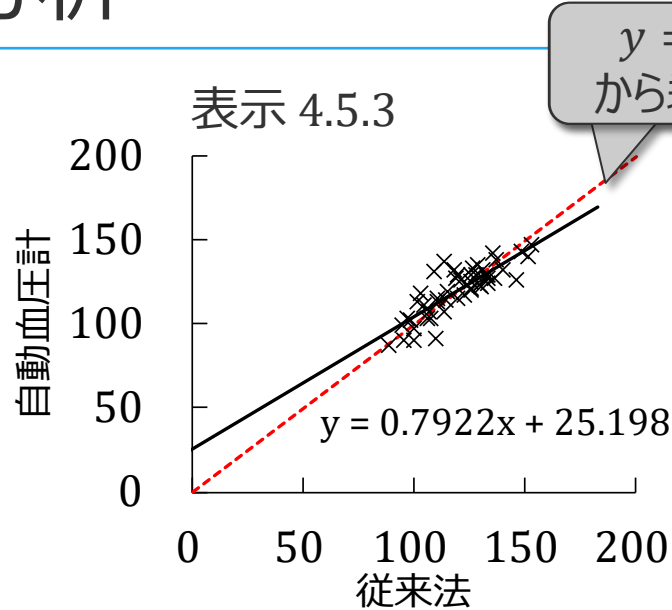
	x	const	
回帰係数	0.881	14.441	切片
その標準誤差	0.067	8.105	その標準誤差
寄与率	0.783	7.026	残差標準偏差
F比	173	48	残差自由度
平方和	8533	2369	残差平方和
t 値	-1.776	1.782	
p 値	0.082	0.081	

x に誤差を含む回帰分析

● x と y の標準偏差の影響

σ_x	σ_y
7.0	5.0

$H_0 : \beta=1$ 、 $H_0 : \alpha=0$ は棄却
当初の予想とは異なる結果



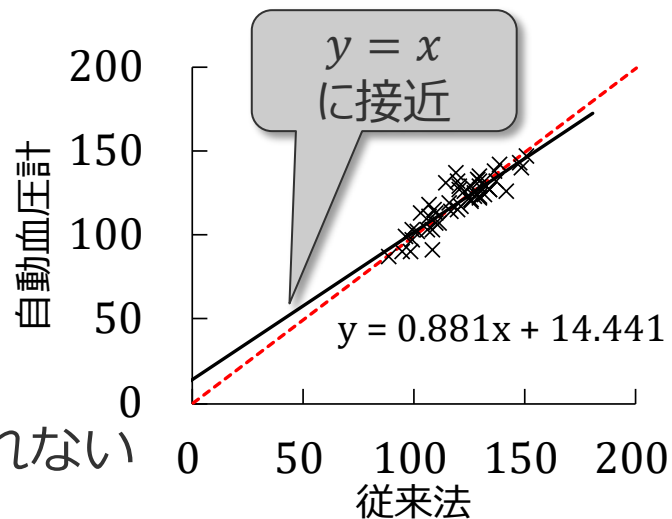
	x	const	
回帰係数	0.792	25.198	切片
その標準誤差	0.074	9.009	その標準誤差
寄与率	0.702	8.227	残差標準偏差
F比	113	48	残差自由度
回帰平方和	7653	3249	残差平方和
t 値	-2.789	2.797	
p 値	0.008	0.007	

$\alpha=0.05$
有意

R2

σ_x	σ_y
5.0	5.0

x の測定誤差を小さくすると
回帰式は $y=x$ に近づく
p 値が大きくなる



	x	const	
回帰係数	0.881	14.441	切片
その標準誤差	0.067	8.105	その標準誤差
寄与率	0.783	7.026	残差標準偏差
F比	173	48	残差自由度
回帰平方和	8533	2369	残差平方和
t 値	-1.776	1.782	
p 値	0.082	0.081	

有意
ではない

$H_0 : \beta=1$ 、 $H_0 : \alpha=0$ は棄却されない

x に誤差を含む回帰分析

● x と y の標準偏差の影響

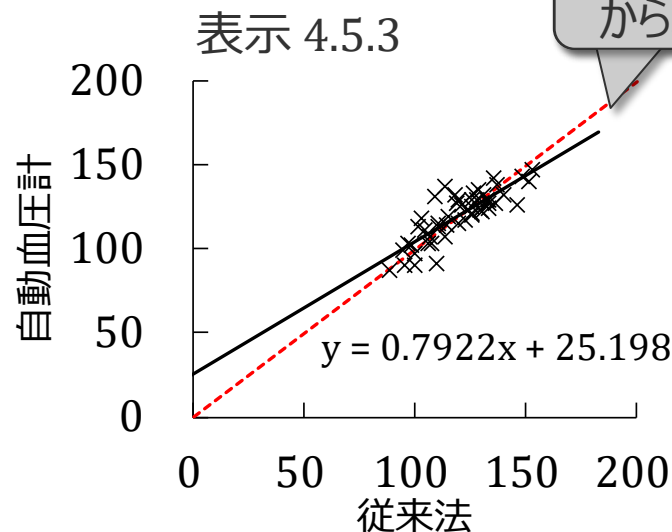
σ_x	σ_y
7.0	5.0

$H_0 : \beta=1$ 、 $H_0 : \alpha=0$ は棄却
当初の予想とは異なる結果

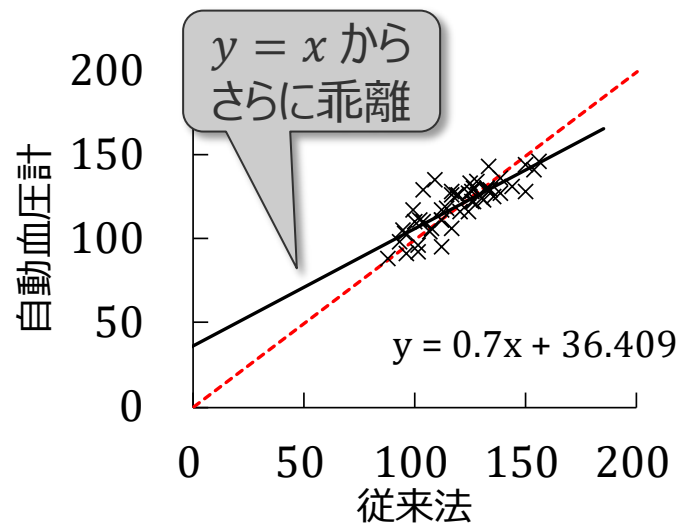
R2, S2

σ_x	σ_y
9.0	3.0

x の測定誤差を大きく、
y の測定誤差を小さくすると
回帰式は $y=x$ からさらに乖離
p 値が小さくなる

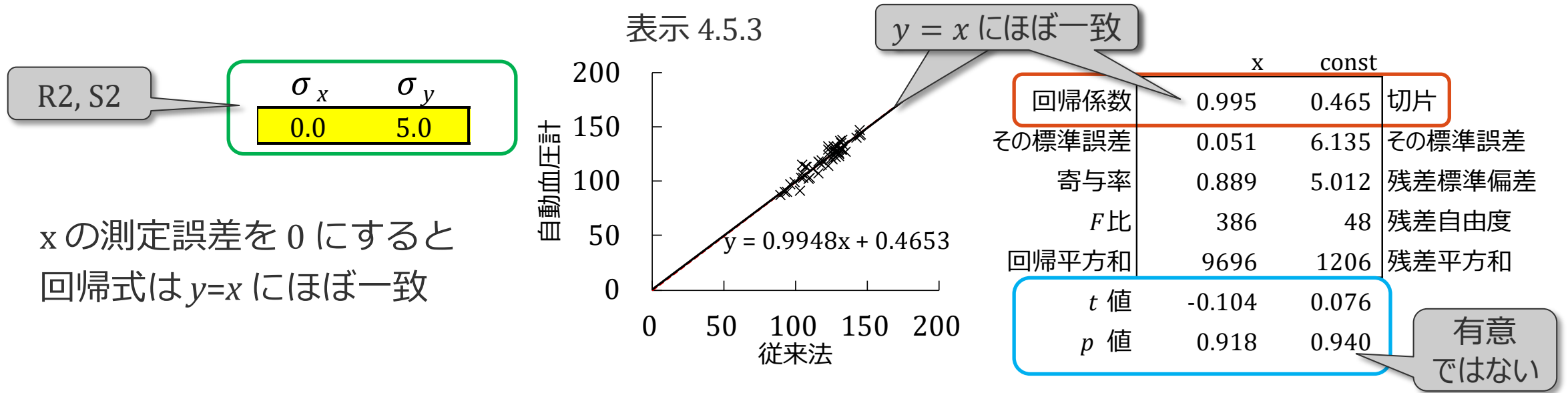


	x	const	
回帰係数	0.792	25.198	切片
その標準誤差	0.074	9.009	その標準誤差
寄与率	0.702	8.227	残差標準偏差
F比	113	48	残差自由度
回帰平方和	7653	3249	残差平方和
t 値	-2.789	2.797	
p 値	0.008	0.007	$\alpha=0.05$ 有意



	x	const	
回帰係数	0.700	36.409	切片
その標準誤差	0.072	8.666	その標準誤差
寄与率	0.665	8.421	残差標準偏差
F比	95	48	残差自由度
回帰平方和	6767	3404	残差平方和
t 値	-4.186	4.201	
p 値	0.000	0.000	高度に 有意

●回帰分析に及ぼす x と y の標準誤差の影響 (シミュレーション)



回帰式をあてはめるときは、

「x には誤差が含まれない」という前提が成立するかどうかを考える必要がある

横軸の値 x に誤差が含まれる場合、回帰係数（傾き）が小さくなる傾向がある → 別の解析方法



● 2つの分析法の一致性を評価する解析方法の1つ

Bland Altman 分析

横軸に平均値 $(x + y)/2$ 、縦軸に差 $(x - y)$ を取って散布図 (Bland Altman プロット) を描く

参照 [§3.5](#) 「対応のある場合の平均値の差の t 検定」 p.154

測定値の差の傾向やばらつきを確認し、2つの測定法がどの程度一致しているか解析

差が平均付近に集中しているか (ばらつきが一定か) ?

差が $\pm 1.96SD$ の範囲内に収まっている割合が高いか?

系統的バイアスがあるか?

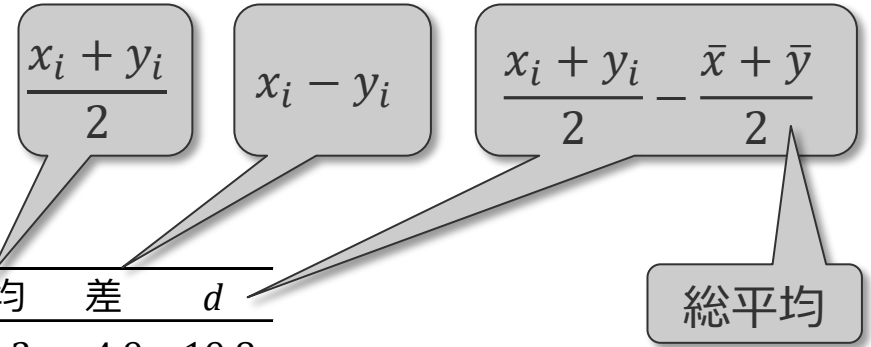
測定のスケールが同じであること、測定値の分布が正規分布に従うと見なせることが条件

●データ

横軸に平均値 $(x + y)/2$ 、縦軸に差 $(x - y)$ を取って散布図を描く

表示4.5.3 のデータ
(セル M1~)

	M	N	O	P	Q	R	S	T	U	V
1				μ	120	σ_x	σ_y			
2				σ	15	7.0	5.0			
3										
4										
5	標準正規乱数									
6					x (真値)	x	y	平均	差	d
7	-0.75	-1.56	-1.20	1	109	97.8	102.8	100.3	-4.9	-19.8
8	0.14	-1.11	-1.59	2	122	114.2	114.1	114.2	0.2	-5.9
9	0.68	-0.16	-1.29	3	130	129.0	123.7	126.4	5.3	6.3
10	-0.25	0.88	0.16	4	116	122.3	117.0	119.7	5.3	-0.4
11	-1.75	0.84	-0.60	5	94	99.6	90.7	95.1	8.9	-24.9
12	-0.92	-0.69	1.46	6	106	101.3	113.5	107.4	-12.1	-12.7
13	-2.05	-0.18	-0.40	7	89	88.0	87.3	87.6	0.7	-32.4
14	-1.14	1.00	-2.34	8	103	109.9	91.2	100.5	18.7	-19.5
15	0.16	-0.62	2.03	9	122	118.0	132.5	125.2	-14.5	5.2



●Bland Altman プロット

散布図

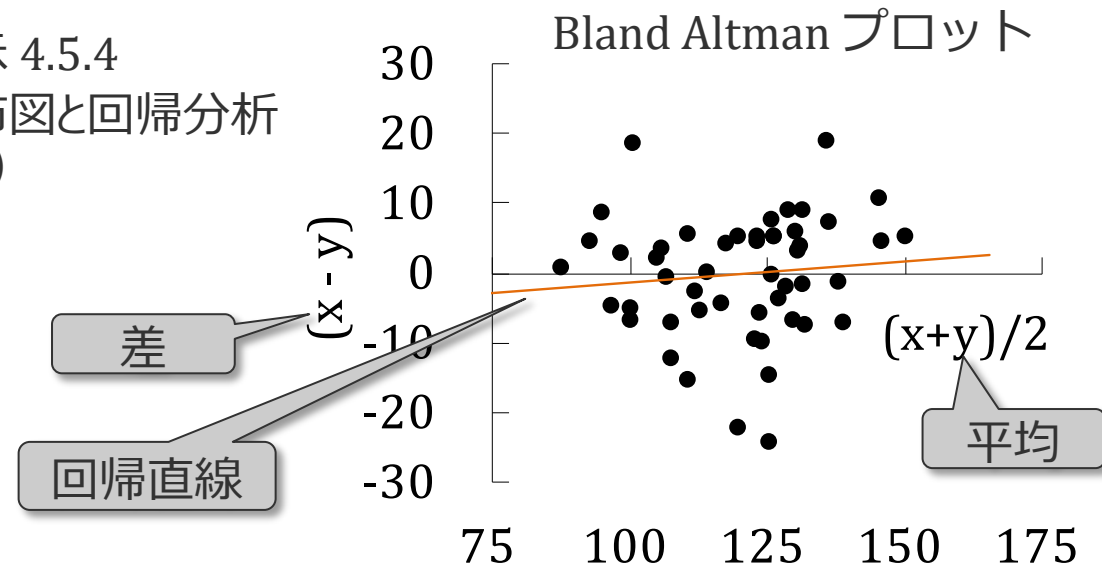
横軸：平均 $(x + y)/2$

縦軸：差 $(x - y)$ を取った散布図 ([§3.5](#) 参照)

回帰直線の描画

x	y	平均	差	d
97.8	102.8	100.3	-4.9	-19.8
114.2	114.1	114.2	0.2	-5.9
129.0	123.7	126.4	5.3	6.3
122.3	117.0	119.7	5.3	-0.4
...

表示 4.5.4
散布図と回帰分析
(2)



	x	const	
回帰係数	0.061	-7.600	切片
その標準誤差	0.086	10.360	その標準誤差
寄与率	0.010	8.822	残差標準偏差
F 比	0.507	48	残差自由度
回帰平方和	39.428	3736	残差平方和
t 値	0.712	-0.734	
p 値	0.480	0.467	

●Bland Altman プロットと母回帰係数の仮説検定

回帰直線の傾き：回帰直線が水平：2つの測定値の間の差は一定

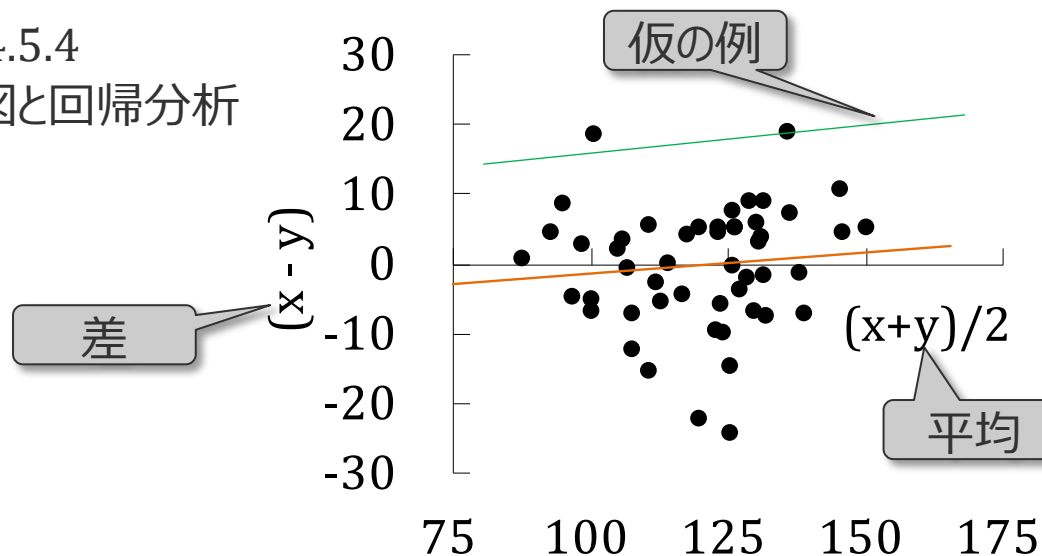
回帰直線が斜め：2つの測定値の間の差が一定ではない、

回帰係数の検定（2変数の差が一定か否かの検定）

帰無仮説 $H_0: \beta=0$ の p 値は0.480, 水平ではないとはいえない（水平であると推測）

2つの測定法が一致しているか？ → x と y が等しいか？ = x と y の差の平均が0か？

表示 4.5.4
散布図と回帰分析
(2)



	x	const	
回帰係数	0.061	-7.600	切片
その標準誤差	0.086	10.360	その標準誤差
寄与率	0.010	8.822	残差標準偏差
F比	0.507	48	残差自由度
回帰平方和	39.428	3736	残差平方和
t 値	0.712	-0.734	
p 値	0.480	0.467	

$H_0: \beta=0$

●切片からの解析

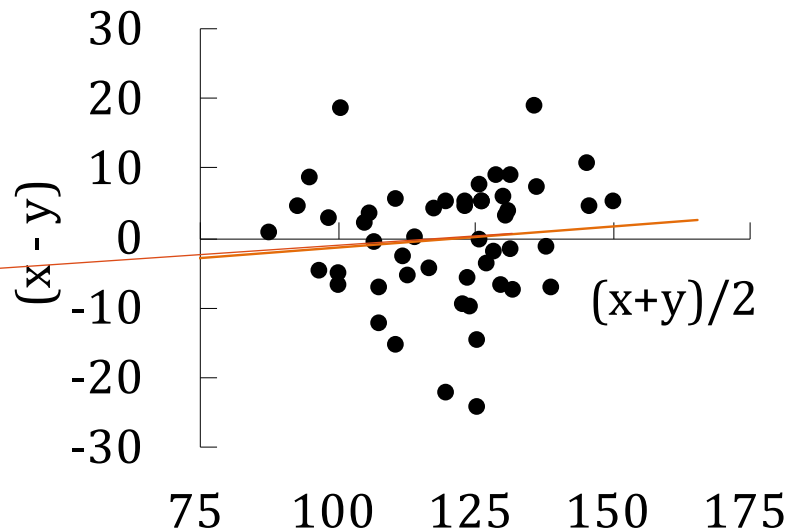
回帰直線の切片：切片が 0 であれば、2つの測定法の差が 0 であるといえるか
 → x と y の差の平均が 0 か否かの検定にはならない

母切片の仮説検定

帰無仮説 $H_0: \alpha=0$ は、 $x=0$ のときの差 $(x-y)$ が 0 という検定

p 値 0.467 帰無仮説は棄却されない：

表示 4.5.4
 散布図と回帰分析
 (2)



横軸が 0 のとき
 切片が 0 か否か

	x	const	
回帰係数	0.061	-7.600	切片
その標準誤差	0.086	10.360	その標準誤差
寄与率	0.010	8.822	残差標準偏差
F比	0.507	48	残差自由度
回帰平方和	39.428	3736	残差平方和
t 値	0.712	-0.734	
p 値	0.480	0.467	

平均の差が 0 の
 検定ではない

Bland Altman 分析

● 差の平均値の仮説検定

2 変数の差の平均の検定

横軸：平均から総平均 $(\bar{x} + \bar{y})/2$ を引いた値 (d)

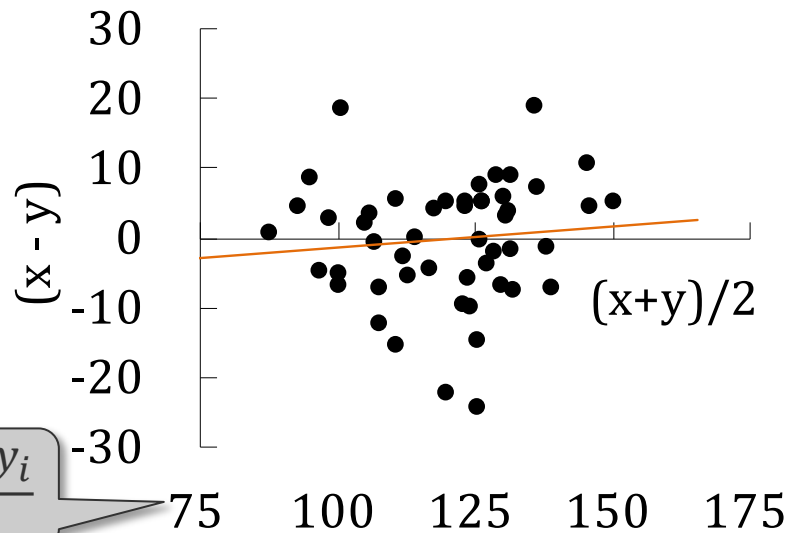
縦軸：差 $(x - y)$

この散布図を画き、回帰分析切片が差の平均値に一致、
切片の検定 (H_0 : 切片=0) を行う

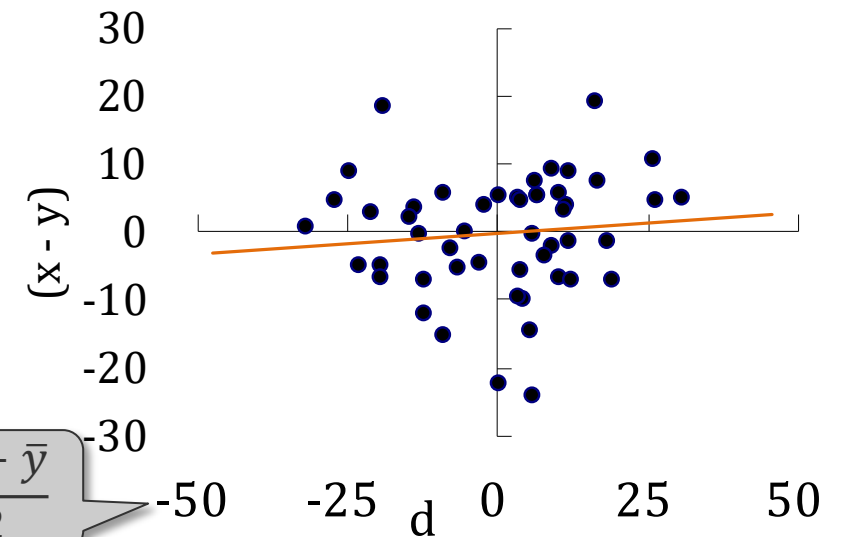
$$d = \frac{x_i + y_i}{2} - \frac{\bar{x} + \bar{y}}{2}$$

x	y	平均	差	d
97.8	102.8	100.3	-4.9	-19.8
114.2	114.1	114.2	0.2	-5.9
129.0	123.7	126.4	5.3	6.3
122.3	117.0	119.7	5.3	-0.4
...

表示 4.5.4
散布図と回帰分析
(2)



$$d = \frac{x_i + y_i}{2} - \frac{\bar{x} + \bar{y}}{2}$$



Bland Altman 分析

● 差の平均値の仮説検定

帰無仮説 $H_0 : (x - y) = 0$

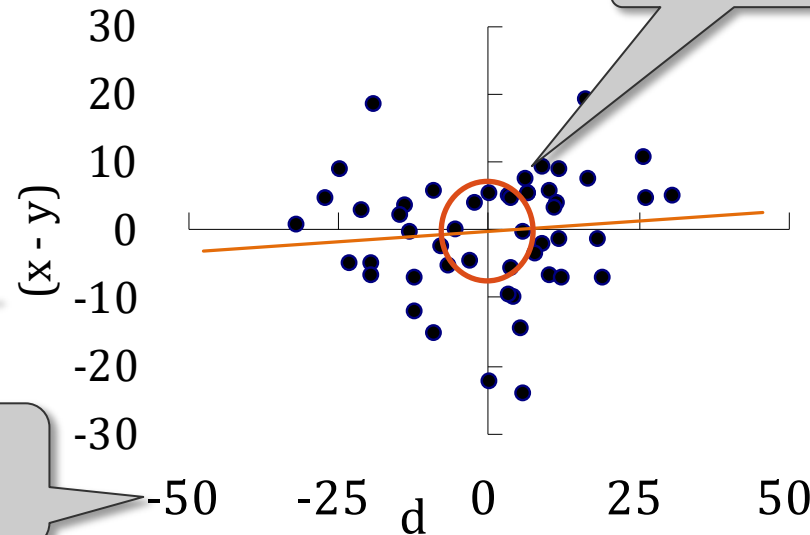
対立仮説 $H_1 : (x - y) \neq 0$

p 値 0.824 帰無仮説は棄却されない
2つの測定値の平均値は等しいと判断される

$$d = \frac{x_i + y_i}{2} - \frac{\bar{x} + \bar{y}}{2}$$

x	y	平均	差	d
97.8	102.8	100.3	-4.9	-19.8
114.2	114.1	114.2	0.2	-5.9
129.0	123.7	126.4	5.3	6.3
122.3	117.0	119.7	5.3	-0.4
...

Excel ファイル
「表示4.5.4」の下



差 $x - y$

$$d = \frac{x_i + y_i}{2} - \frac{\bar{x} + \bar{y}}{2}$$

	x	const	
回帰係数	0.061	-0.279	切片
その標準誤差	0.086	1.248	その標準誤差
寄与率	0.010	8.822	残差標準偏差
F 比	0.507	48	残差自由度
回帰平方和	39.428	3736	残差平方和
t 値	0.712	-0.224	
p 値	0.480	0.824	

切片の検定

Bland Altman 分析

●JMP [対応のあるペア]

Excel の x と y (R6:S56) を JMP にコピー/ペースト (列名含む)

JMP [編集] > [列名とともにコピー]

[分析] > [対応のあるペア] (§3.5 参照)

	P	Q	R	S	T
1	μ	120	σ_x	σ_y	
2	σ	15	7.0	5.0	
3					
4					
5					
6	x (真値)		x	y	平均
7	1	109	97.8	102.8	100.3
8	2	122	114.2	114.1	114.2
9	3	130	129.0	123.7	126.4
10	4	116	122.3	117.0	119.7
11	5	94	99.6	90.7	95.1
12	6	106	101.3	113.5	107.4
13	7	89	88.0	87.3	87.6

データ
セル R6:S56

	x	y
1	97.8	102.8
2	114.2	114.1
3	129	123.7
4	122.3	117
5	99.6	90.7
6	101.3	113.5
7	88	87.3

無題 - JMP

ファイル(F) 編集(E) テーブル(T) 行(R)

分析(A) グラフ(G) ツール(O) 表示(V)

- 一変量の分布
- y_x 二変量の関係
- 対応のあるペア**
- モデルのあてはめ

モデル化 ▶

多変量 ▶

品質と工程 ▶

信頼性/生存時間分析 ▶

y 6

7

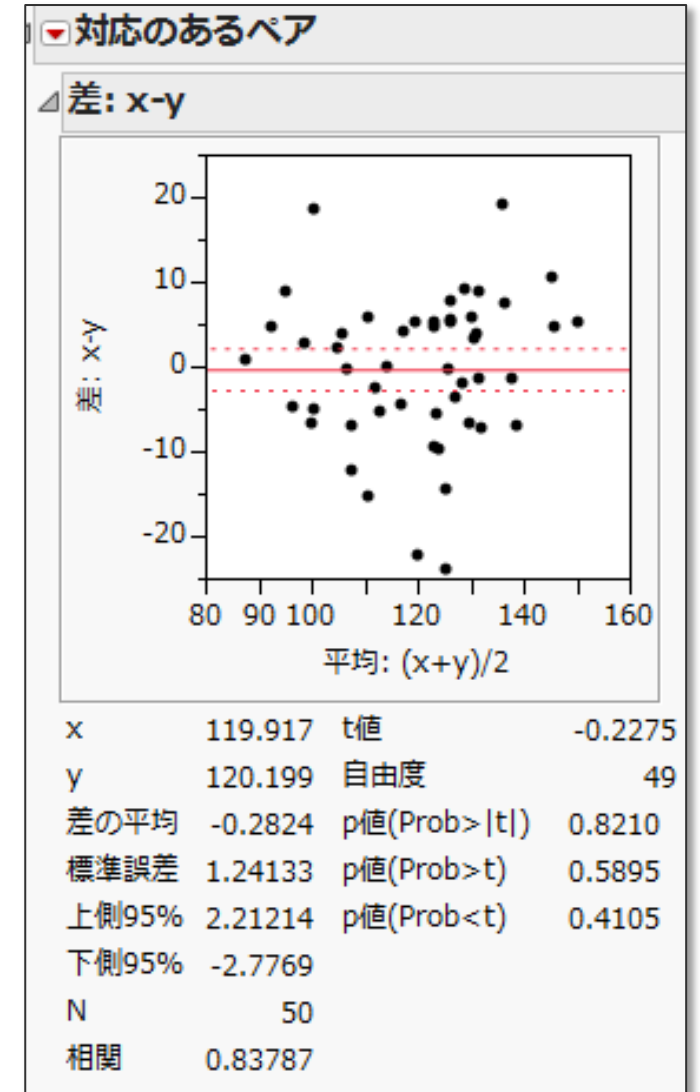
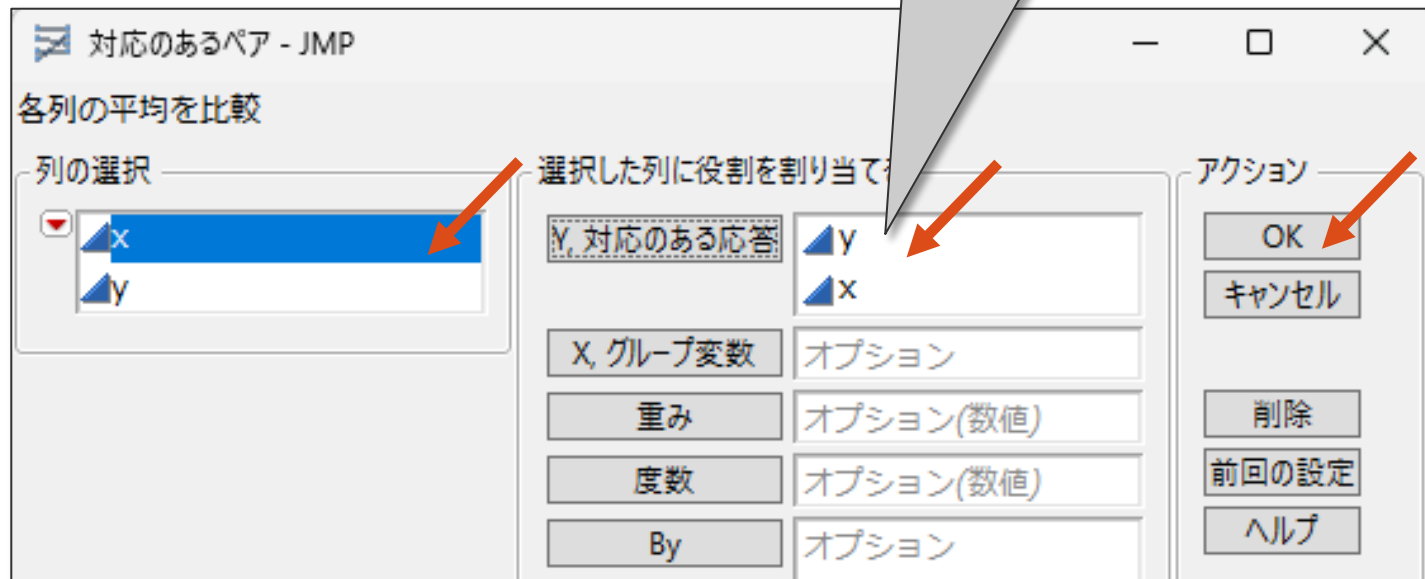
Bland Altman 分析

●JMP [対応のあるペア]

JMP [分析] > [対応のあるペア] (§3.5 参照)

ダイアログボックスで、2列を設定

(差を $x - y$ にするために、 y を上、 x を下に設定)

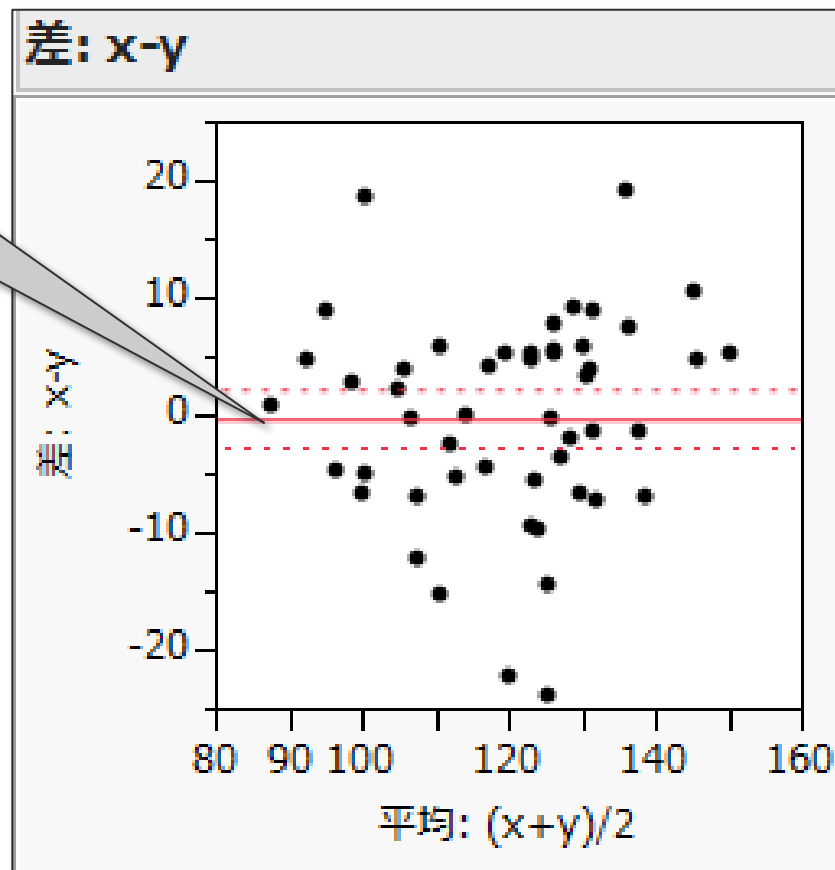
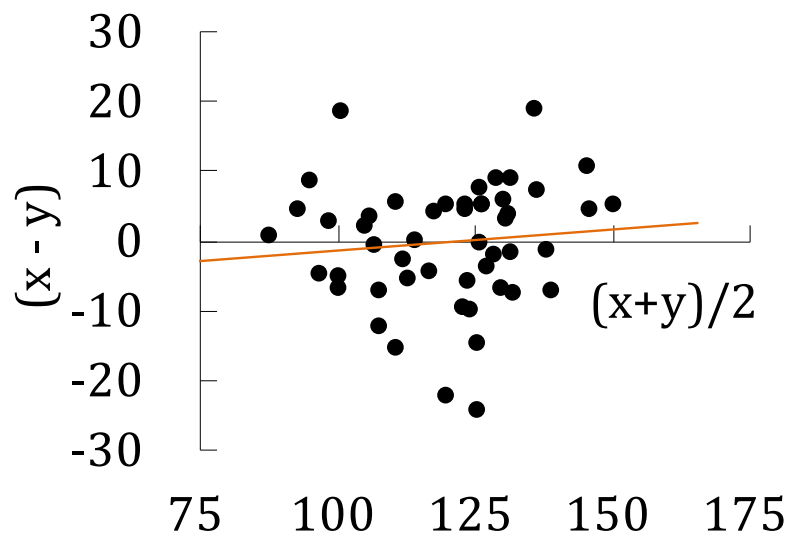


● Bland Altman プロット

新しい JMP には、Bland Altman 分析 に関わる様々なメニューがある
2つの測定法を比較する分析手法として適している

$x - y$ の差の平均と
95% 信頼区間

表示 4.5.4 散布図回帰分析 (2)





(4) 変数の選び方

回帰分析に用いる説明変数の選び方

●事例 5

40名の健康診断の検査について、検査項目の間の相関係数行列を得た (p.218) 血糖値が高くならないように、血圧のコントロールの重要性が認識されてきた 血糖と血圧(上)には弱い相関があり、血糖と血圧(下)はほぼ無相関、血圧(上)のみが血糖のコントロールに重要か？

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

血圧 (上) : 収縮期の血圧
血圧 (下) : 拡張期の血圧

●事例 5

- (1) 血糖の上のみ $\text{血糖} = 59.59 + 0.24 \times \text{血糖(上)}$ $s.d. = 9.00$ $t = 2.385$ $p = 0.02$
- (2) 血糖の上下の差 $\text{血糖} = 66.47 + 0.50 \times \text{血糖(差)}$ $s.d. = 8.40$ $t = 3.489$ $p = 0.001$
- (3) 血糖の上下、両方 $\text{血糖} = 63.16 + 0.50 \times \text{血糖(上)} - 0.45 \times \text{血糖(下)}$ $s.d. = 8.50$
 (重回帰式) $t = -2.373$ $p = 0.023$ 、 $t = 3.442$ $p = 0.001$

表示 4.5.5
血糖と血糖の関係

(1) 血糖(上)

0.24	59.59
0.10	12.02
0.13	9.00
5.69	38
461	3080

t 2.385
p 0.022

(2) 血糖(差)

0.50	66.47
0.14	6.33
0.24	8.40
12.18	38
859	2682

t 3.489
p 0.001

(3) 血糖(下) 血糖(上)

-0.45	0.50	63.16
0.19	0.14	11.45
0.25	8.50	#N/A
6.01	37	#N/A
868	2673	#N/A

t -2.373 3.442
p 0.023 0.001

自由度
40 - 3 = 37

●事例 6

40名の健康診断の検査について、検査項目の間の相関係数行列を得た (p.218)
 血糖 (y) を、身長と体重から計算される BMI (x) から推定した (§4.3(1) p.225)
 血糖 (y) を、身長 (x_1) と体重 (x_2) から直接的に推定する (重回帰式を求める)

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	血压(上)	血压(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							
血压(上)	0.103	0.118	0.279	1						
血压(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

BMI (Body Mass Index)
 体重(kg) ÷ 身長(m)²
 標準 : 22、肥満 : 25以上

●事例 6

重回帰分析

$$Y = 3.801 - 1.270X_1 + 0.532X_2$$

$$\log_{10} y = 3.801 - 1.270 \log_{10} x_1 + 0.532 \log_{10} x_2$$

$$y = 10^{3.801} x_1^{-1.270} x_2^{0.532}$$

$$X_1 = \log_{10} x_1$$

$$X_2 = \log_{10} x_2$$

$$Y = \log_{10} y$$

表示4.5.6
血糖と
体重・身長
の関係
(n=40)

	x1	x2	y	X1	X2	Y
No.	身長	体重	BMI	身長	体重	血糖
1	169	75	26.3	2.228	1.875	1.978
2	164	75	27.9	2.215	1.875	1.944
3	161	68	26.2	2.207	1.833	2.068
4	169	73	25.6	2.228	1.863	1.919
5	177	63	20.1	2.248	1.799	1.949
6	166	57	20.7	2.220	1.756	1.973
7	170	65	22.5	2.230	1.813	1.919
8	163	60	22.6	2.212	1.778	1.903
9	157	62	25.2	2.196	1.792	1.968
10	166	67	24.3	2.220	1.826	2.013

	log(体重)	log(身長)	const
係数	0.532	-1.270	3.801
se(係数)	0.143	0.517	1.061
r^2	0.285	0.039	#N/A
F	7.385	37	#N/A
SR	0.022	0.055	#N/A

重回帰分析

	BMI	const
係数	2.161	38.002
se(係数)	0.564	13.143
r^2	0.278	8.200
F	14.658	38
SR	985.616	2555.159

単回帰分析

下部は省略

●事例 6

重回帰分析：血糖 (y) を身長 (x_1) と体重 (x_2) に回帰

$$Y = 3.801 - 1.270X_1 + 0.532X_2$$

$$\log_{10} y = 3.801 - 1.270 \log_{10} x_1 + 0.532 \log_{10} x_2$$

$$y = 10^{3.801} x_1^{-1.270} x_2^{0.532}$$

係数の比

$$b_1 : b_2 = -1.270 : 0.532 \approx -2 : 1$$

この比は、肥満度の指標として $\text{BMI} = x_2/x_1^2 = x_1^{-2}x_2^1$ が用いられている妥当性を示唆

単回帰式：血糖 (y) を BMI ($x = x_2/x_1^2$) に回帰

$$y = 38.0 + 2.16x = 38.0 + 2.16 x_2/x_1^2 = 38.0 + 2.16 x_1^{-2} x_2^1 \quad (\text{表示4.3.1})$$

$$\begin{aligned} X_1 &= \log_{10} x_1 \\ X_2 &= \log_{10} x_2 \\ Y &= \log_{10} y \end{aligned}$$

$$\begin{aligned} \text{BMI} \\ &= \text{体重}(x_2) / \text{身長}(x_1)^2 \end{aligned}$$



●変数の取り方

(1) 1 変量で説明する $y = a + bx$

(2) 2 変量からなる指数を用いる . . . $y = a + b(x_1/x_2^2)$
(血糖をBMIで回帰)

(3) 2 変量の差 (比) を用いる . . . $y = a + b(x_2 - x_1)$ $y = a + b(x_2/x_1)$
(血糖を血圧の上下の差で回帰)

(4) 2 変量を同時に用いる $y = b_0 + b_1x_1 + b_2x_2$
(血糖を身長と体重で回帰、血糖を血圧の上と下で回帰、重回帰式)

(5) 2 次式を用いる $y = b_0 + b_1x + b_2x^2$
(第 2 部 [§2.2](#) で取り上げる)

(6) 1 変量で説明する $y = bx$
(原点を通る回帰直線) . . . この後で説明 (§4.6(4))

どのような説明変数を使うかは、
解析者の固有技術で判断



補足：原点を通る回帰式

4.6 補遺 (4)



原点を通る回帰式

●平方和、積和の用語

偏差平方和（平方和）

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

偏差積和

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

2乗和

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

積和

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

●事例 7

原点を通る単回帰分析

薬剤が無投与のときの薬効が 0 である場合、回帰直線が原点を通るモデルのあてはめを考える
($x = 0, y = 0$)

回帰モデル (母回帰式)

$$y_i = \beta x_i + \varepsilon_i \quad (4.6.17)$$

推定式 (標本回帰式)

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

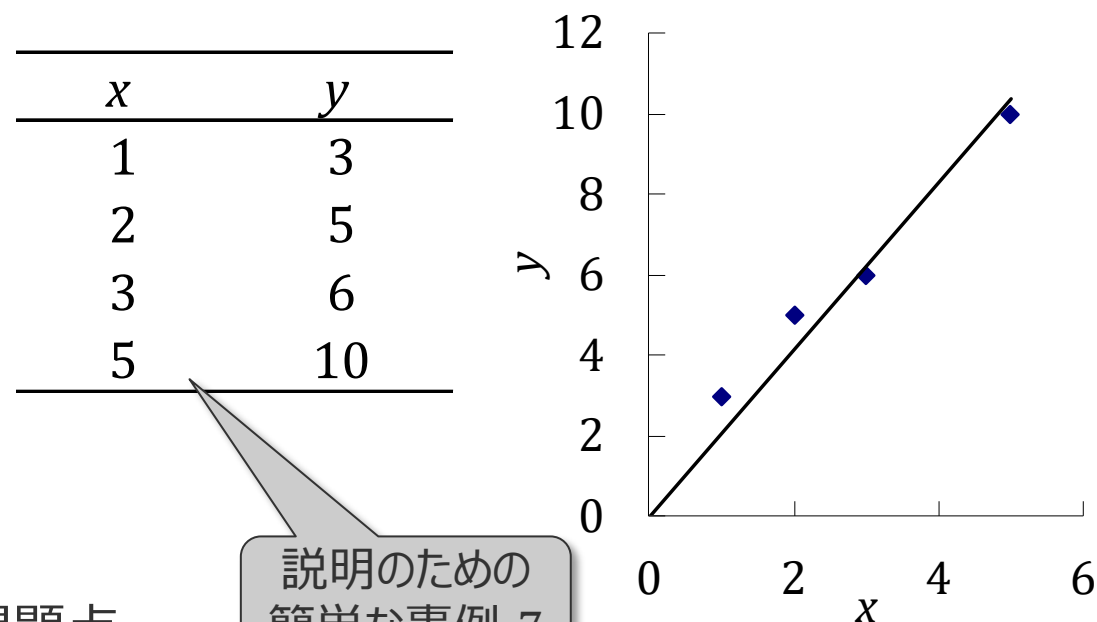
単回帰分析

Excelによる回帰分析 (散布図、LINEST関数)

JMP [二変量の関係] による回帰分析

回帰係数、寄与率 (決定係数) の計算方法と問題点

表示 4.6.1 原点を通る回帰式 (一部、改変)



説明のための
簡単な事例 7

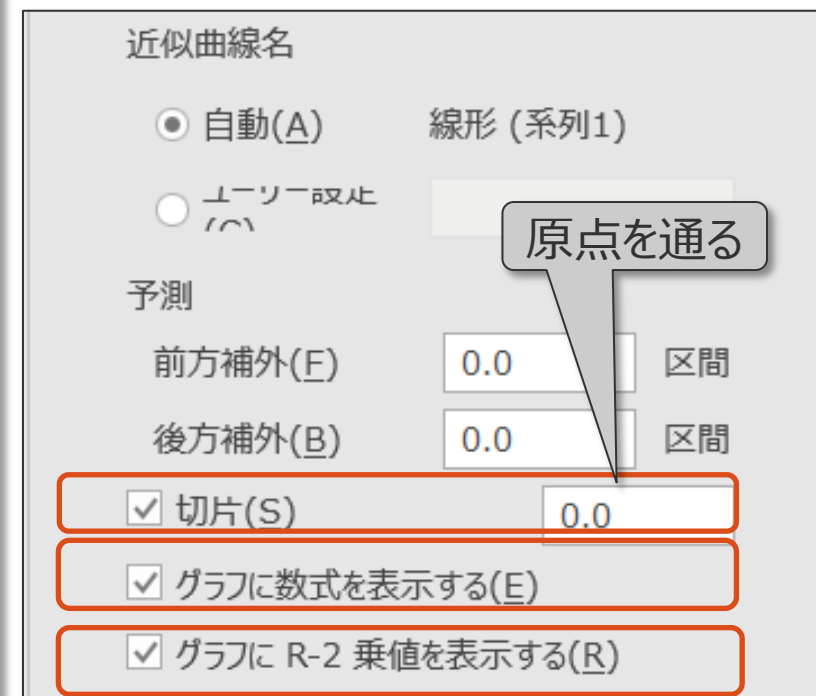
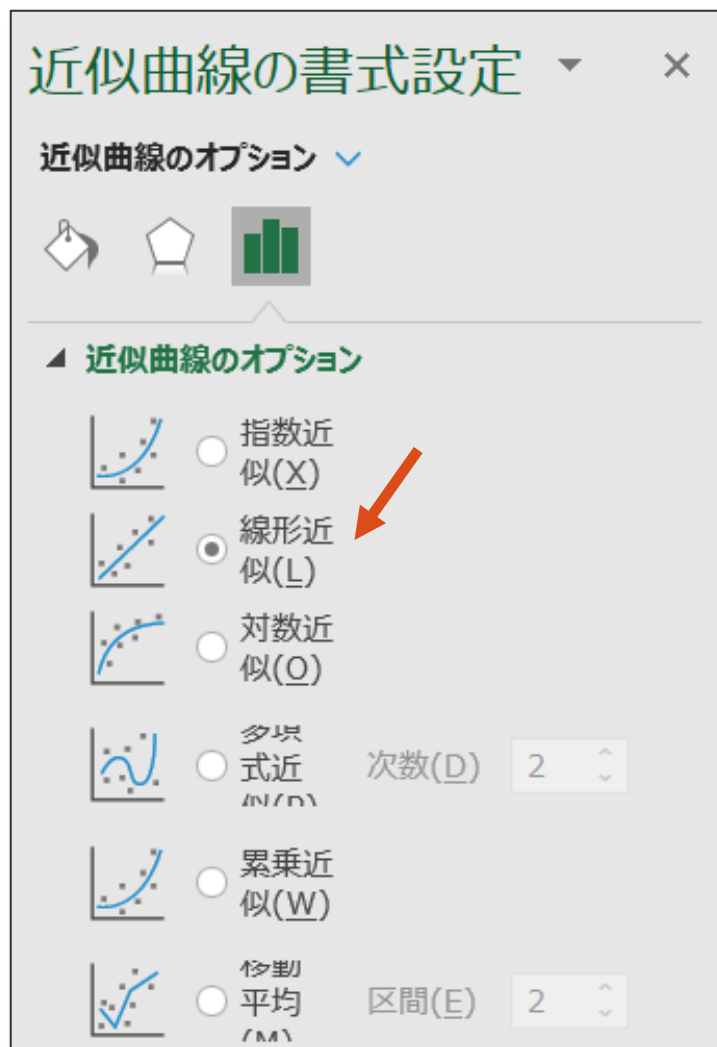
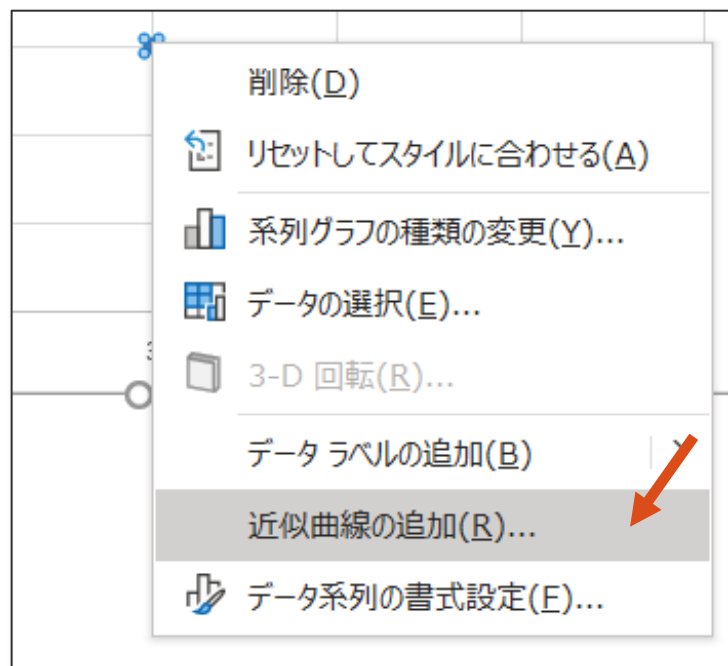
●Excel の散布図

散布図の作成 ([§4.1](#))

図中のマークを右クリック

[近似曲線の追加]

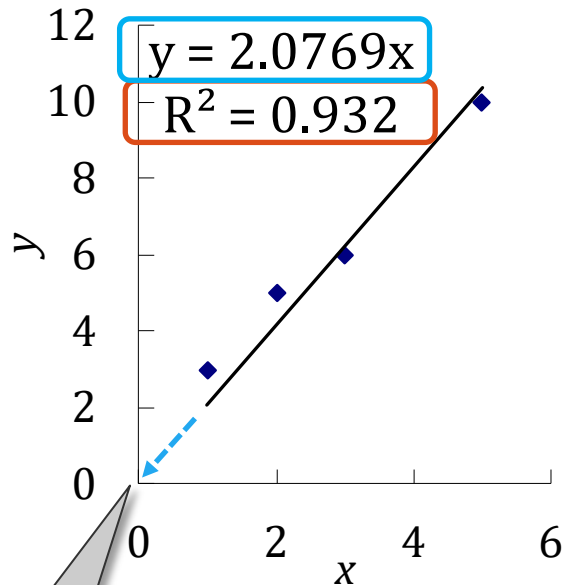
> [近似曲線の書式設定]



●Excel の散布図と LINEST 関数の結果

表示 4.6.1

原点を通る回帰式 (一部、改変)



原点を通る

スピルによる LINEST 関数の実行 (§4.3, [ブログ](#) 参照)

x	y
1	3
2	5
3	6
5	10

=LINEST(K7:K10,J7:J10,FALSE,TRUE)

y 列
x 列
切片を 0 に指定

結果を出力する
スペースを確保

$\hat{y} = 2.0769x$

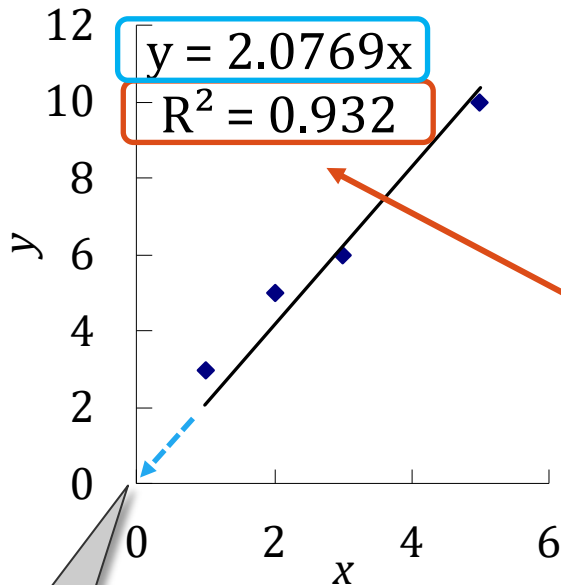
	x	const	
回帰係数	2.07692	0	切片
その標準誤差	0.12297	#N/A	
寄与率	0.98959	0.76795	標準偏差
F 比	285.261	3	残差自由度
回帰平方和	168.231	1.76923	残差平方和

●Excel の散布図と LINEST 関数の結果

スピルによる LINEST 関数の実行 (§4.3, [ブログ](#) 参照)

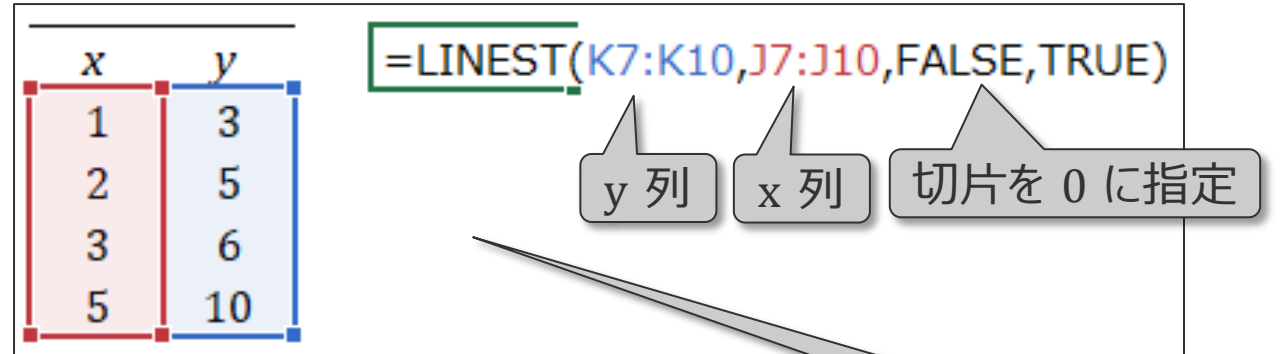
表示 4.6.1

原点を通る回帰式 (一部、改変)



原点を通る

散布図の寄与率と
LINEST 関数の寄与率
が異なる
(定義が異なるため)



結果を出力する
スペースを確保

$$\hat{y} = 2.0769x$$

	x	const	
回帰係数	2.07692	0	切片
その標準誤差	0.12297	#N/A	
寄与率	0.98959	0.76795	標準偏差
F 比	285.261	3	残差自由度
回帰平方和	168.231	1.76923	残差平方和

原点を通る回帰式

(a) 原点を通るという制約がない

No.	x	y	yhat	e
1	1	3	3.00	0.00
2	2	5	4.71	0.29
4	3	6	6.43	-0.43
5	5	10	9.86	0.14
平均	2.75	6.00		
2乗和積和				
平方和積和	8.75	26.00	25.71	0.29
		15.00		
a		1.29		
b		1.71		
R^2		0.989		

$$\hat{y} = a + bx$$

(b) 原点を通る (散布図のR^2)

No.	x	y	yhat	e
1	1	3	2.08	0.92
2	2	5	4.15	0.85
3	3	6	6.23	-0.23
4	5	10	10.38	-0.38
平均		6.00		
2乗和積和	39.00			
		81.00		
平方和積和		26.00	37.74	1.77
b		2.08		
R^2		0.932		

$$\hat{y} = bx$$

(c) 原点を通る (LINEST関数のR^2)

No.	x	y	yhat	e
1	1	3	2.08	0.92
2	2	5	4.15	0.85
3	3	6	6.23	-0.23
4	5	10	10.38	-0.38
平均				
2乗和積和	39.00	170.00	168.23	1.77
		81.00		
平方和積和				
b		2.08		
R^2		0.990		

$$\hat{y} = bx$$

$$b = \frac{\sum (x_i - \bar{x})(y_i - \bar{y})}{\sum (x_i - \bar{x})^2} = \frac{15.0}{8.75} = 1.71$$

$$R^2 = 1 - S_e/S_T \quad (\S 4.3)$$

$$= 1 - 0.29/26 = 0.989$$

$$b = \frac{\sum x_i y_i}{\sum x_i^2} = \frac{81.0}{39.0} = 2.08$$

$$R^2 = 1 - S_e/S_T$$

$$= 1 - 1.77/26 = 0.932$$

$$b = \frac{\sum x_i y_i}{\sum x_i^2} = \frac{81.0}{39.0} = 2.08$$

$$R^2 = 1 - S_e/S_T$$

$$= 1 - 1.77/170 = 0.990$$

原点を通る回帰式

(a) 原点を通るという制約がない

No.	x	y	yhat	e
1	1	3	3.00	0.00
2	2	5	4.71	0.29
4	3	6	6.43	-0.43
5	5	10	9.86	0.14
平均	2.75	6.00		
2乗和積和				
平方和積和	8.75	26.00	25.71	0.29
a		1.29		
b		1.71		
R^2		0.989		

(b) 原点を通る (散布図のR^2)

No.	x	y	yhat	e
1	1	3	2.08	0.92
2	2	5	4.15	0.85
3	3	6	6.23	-0.23
4	5	10	10.38	-0.38
平均				
2乗和積和	39.00	170.00	168.23	1.77
平方和積和	81.00			
a				
b		2.08		
R^2		0.932		

(c) 原点を通る (LINEST関数のR^2)

No.	x	y	yhat	e
1	1	3	2.08	0.92
2	2	5	4.15	0.85
3	3	6	6.23	-0.23
4	5	10	10.38	-0.38
平均				
2乗和積和	39.00	170.00	168.23	1.77
平方和積和	81.00			
a				
b		2.08		
R^2		0.990		

$S_T = S_R + S_e$
 $26.00 = 25.71 + 0.29$
 R^2 は S_T に占める S_R の割合 .77

平均を引く
 R^2 0.932

$$b = \frac{\sum (x_i - \bar{x})(y_i - \bar{y})}{\sum (x_i - \bar{x})^2} = \frac{15.0}{8.75} = 1.71$$

$$R^2 = 1 - S_e/S_T \quad (\S 4.3)$$

$$= 1 - 0.29/26 = 0.989$$

$$b = \frac{\sum x_i y_i}{\sum x_i^2} = \frac{81.0}{39.0} = 2.08$$

$$R^2 = 1 - S_e/S_T$$

$$= 1 - 1.77/26 = 0.932$$

$$b = \frac{\sum x_i y_i}{\sum x_i^2} = \frac{81.0}{39.0} = 2.08$$

$$R^2 = 1 - S_e/S_T$$

$$= 1 - 1.77/170 = 0.990$$

原点を通る回帰式

(a) 原点を通るという制約がない

平和和の分解				
$S_T \neq S_R + S_e$				
$26.00 \neq 37.74 + 1.77$				
R^2 は S_T に占める S_R の割合ではない				
平均	2.75	6.00		
2乗和積和				
平方和積和	8.75	26.00	25.71	0.29
		15.00		
a			1	
b			1	
R^2			0.989	

(b) 原点を通る (散布図の R^2)

No.	x	y	yhat	e
1	1	3	2.08	0.92
2	2	5	4.15	0.85
3	3	6	6.23	-0.23
4	5	10	10.38	-0.38
平均		6.00		
2乗和積和	39.00			
		81.00		
平方和積和			26.00	37.74
				1.77
			S_T	
b		2.08		
R^2			0.932	

(c) 原点を通る (LINEST関数の R^2)

No.	x	y	yhat	e
1	1	3	2.08	0.92
2	2	5	4.15	0.85
3	3	6	6.23	-0.23
4	5	10	10.38	-0.38
平均				
2乗和積和	39.00	170.00	168.23	1.77
		81.00		
平方和積和				
b		2.08		
R^2			0.990	

平均を引かない

$$b = \frac{\sum (x_i - \bar{x})(y_i - \bar{y})}{\sum (x_i - \bar{x})^2} = \frac{15.0}{8.75} = 1.71$$

$$R^2 = 1 - S_e/S_T \quad (\S 4.3)$$

$$= 1 - 0.29/26 = 0.989$$

$$b = \frac{\sum x_i y_i}{\sum x_i^2} = \frac{81.0}{39.0} = 2.08$$

$$R^2 = 1 - S_e/S_T$$

$$= 1 - 1.77/26 = 0.932$$

$$b = \frac{\sum x_i y_i}{\sum x_i^2} = \frac{81.0}{39.0} = 2.08$$

$$R^2 = 1 - S_e/S_T$$

$$= 1 - 1.77/170 = 0.990$$

原点を通る回帰式

(a) 原点を通るという制約がない

平和和の分解				
$S_T \neq S_R + S_e$				
$26.00 \neq 37.74 + 1.77$				
R^2 は S_T に占める S_R の割合ではない				
平均	2.75	6.00		
2乗和積和				
平方和積和	8.75	26.00	25.71	0.29
		15.00		
a			1	
b			1	
R^2			0.989	

(b) 原点を通る (散布図の R^2)

2乗和の分解				
$S_T = S_R + S_e$				
$170.00 = 168.23 + 1.77$				
R^2 は S_T に占める S_R の割合				
No.				
1				
2				
3				
4				
平均		6.00		
2乗和積和	39.00			
	81.00			
平方和積和		26.00	37.74	1.77
				S_T
b			2.08	
R^2			0.932	

(c) 原点を通る (LINEST関数の R^2)

No.	x	y	yhat	e
1	1	3	2.08	0.92
2	2	5	4.15	0.85
3	3	6	6.23	-0.23
4	5	10	10.38	-0.38
平均				
2乗和積和	39.00	170.00	168.23	1.77
	81.00			S_T
平方和積和				
b			2.08	
R^2			0.990	

平均を引かない

$$b = \frac{\sum (x_i - \bar{x})(y_i - \bar{y})}{\sum (x_i - \bar{x})^2} = \frac{15.0}{8.75} = 1.71$$

$$R^2 = 1 - S_e/S_T \quad (\S 4.3)$$

$$= 1 - 0.29/26 = 0.989$$

$$b = \frac{\sum x_i y_i}{\sum x_i^2} = \frac{81.0}{39.0} = 2.08$$

$$R^2 = 1 - S_e/S_T$$

$$= 1 - 1.77/26 = 0.932$$

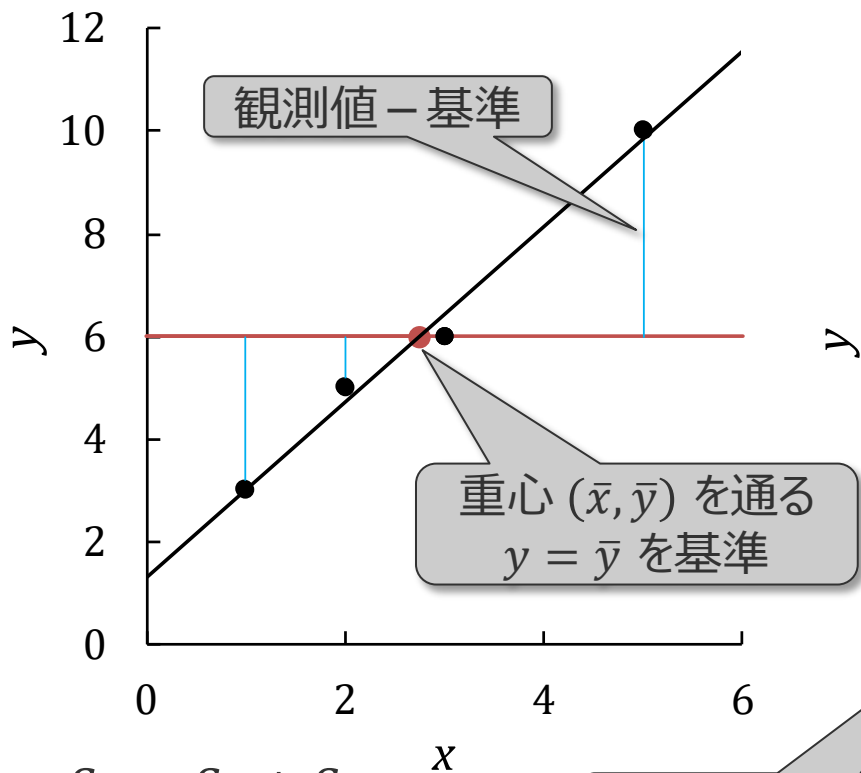
$$b = \frac{\sum x_i y_i}{\sum x_i^2} = \frac{81.0}{39.0} = 2.08$$

$$R^2 = 1 - S_e/S_T$$

$$= 1 - 1.77/170 = 0.990$$

原点を通る回帰式

(a) 原点を通るという制約がない

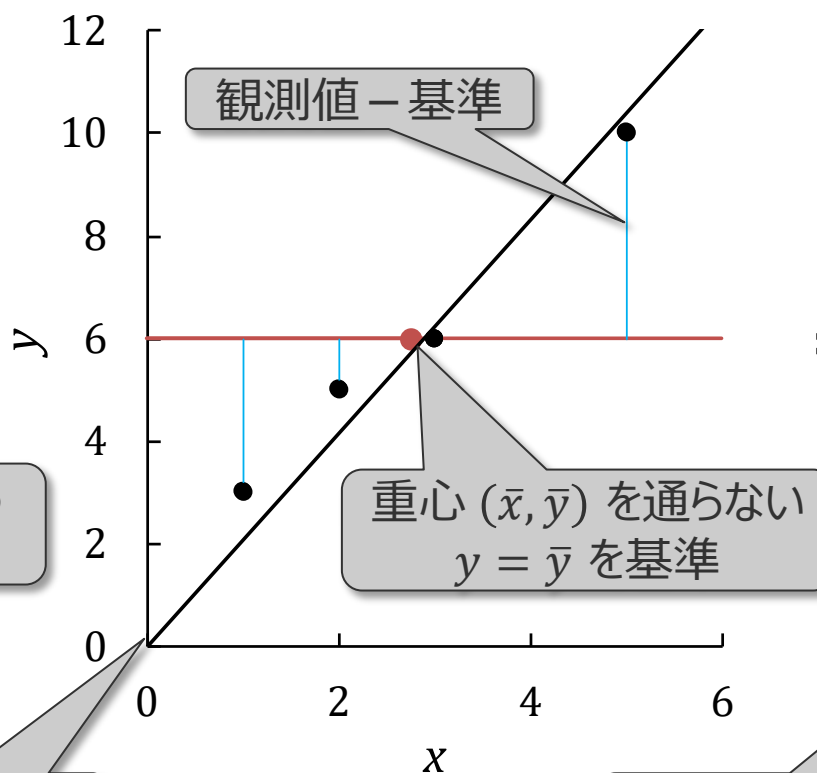


$$S_T = S_R + S_e$$

$$R^2 = 1 - S_e/S_T$$

$$= 0.989 \dots \text{正しい}$$

(b) 原点を通る (散布図のR^2)

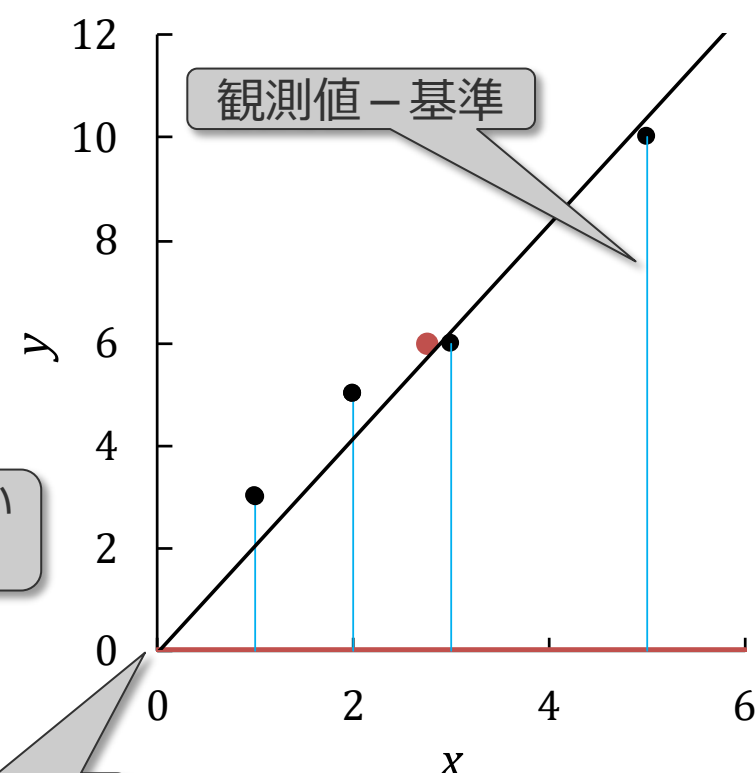


$$S_T \neq S_R + S_e$$

$$R^2 = 1 - S_e/S_T$$

$$= 0.932 \dots \text{曖昧}$$

(c) 原点を通る (LINEST関数のR^2)



$$S_T = S_R + S_e$$

$$R^2 = 1 - S_e/S_T$$

$$= 0.990 \dots \text{正しい}$$

原点を通る回帰式

(a) 原点を通るという制約がない

No.	x	y	yhat	e
1	1	6	4.80	1.20
2	3	5	5.91	-0.91
4	3	6	7.03	-1.03
5	5	10	9.26	0.74
平均	2.75	6.75		
2乗和積和				
平方和積和	8.75	14.75	10.86	3.89
a		3.69		
b		1.11		
R^2		0.737		

(b) 原点を通る (散布図のR^2)

No.	x	y	yhat	e
1	1	6	2.15	3.85
2	2	5	4.31	0.69
3	3	6	6.46	-0.46
4	5	10	10.77	-0.77
平均		6.75		
2乗和積和	39.00			
平方和積和		14.75	40.59	16.08
b		2.15		
R^2		-0.090		

(c) 原点を通る (LINEST関数のR^2)

No.	x	y	yhat	e
1	1	6	2.15	3.85
2	2	5	4.31	0.69
3	3	6	6.46	-0.46
4	5	10	10.77	-0.77
平均				
2乗和積和	39.00	197.00	180.92	16.08
平方和積和		84.00		
b		2.15		
R^2		0.918		

$$b = \frac{\sum (x_i - \bar{x})(y_i - \bar{y})}{\sum (x_i - \bar{x})^2} = \frac{9.75}{8.75} = 1.11$$

$$R^2 = 1 - S_e/S_T = 1 - 3.89/14.75 = 0.737$$

$$b = \frac{\sum x_i y_i}{\sum x_i^2} = \frac{84.0}{39.0} = 2.15$$

$$R^2 = 1 - S_e/S_T = 1 - 16.08/14.75 = -0.090$$

$$b = \frac{\sum x_i y_i}{\sum x_i^2} = \frac{84.0}{39.0} = 2.15$$

$$R^2 = 1 - S_e/S_T = 1 - 16.08/197 = 0.918$$

原点を通る回帰式

●Excel の寄与率の表示

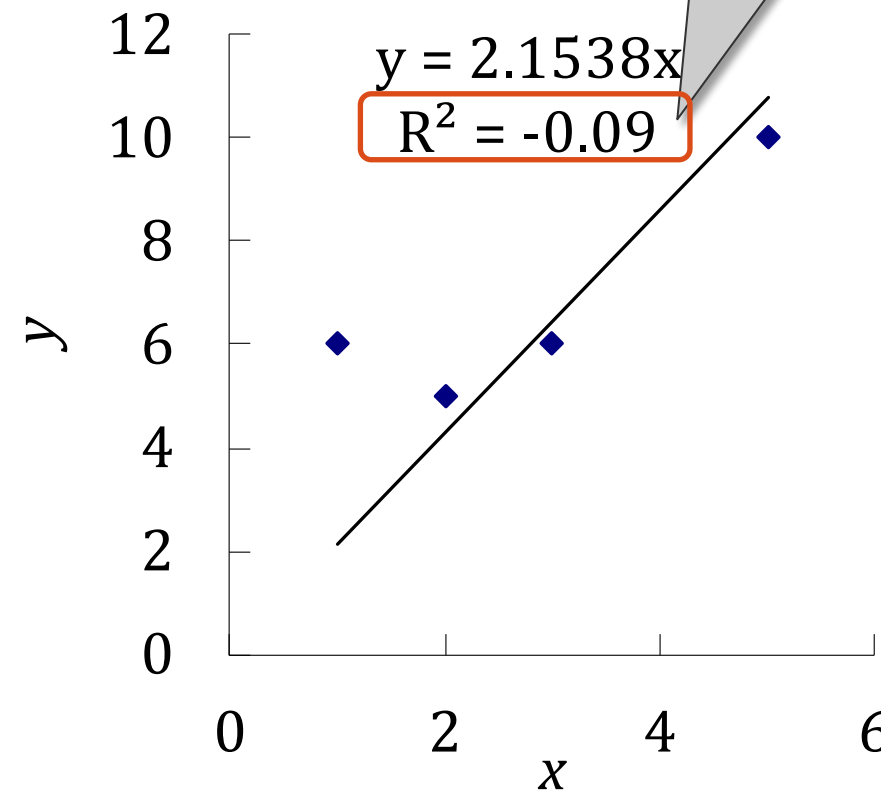
LINEST 関数の結果

x	y
1	6
2	5
3	6
5	10

3 → 6

	x	const	
回帰係数	2.1538	0	切片
その標準誤差	0.3707	#N/A	その標準誤差
寄与率	0.9184	2.3149	残差標準偏差
F比	33.761	3	残差自由度
回帰平方和	180.92	16.077	残差平方和

散布図の表示

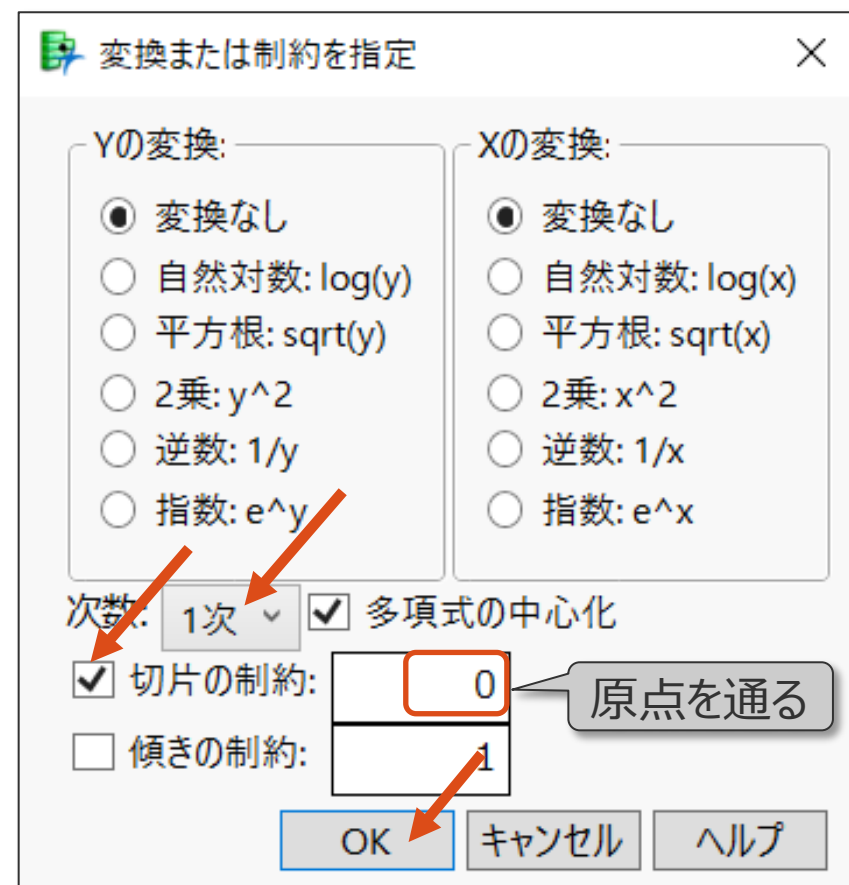
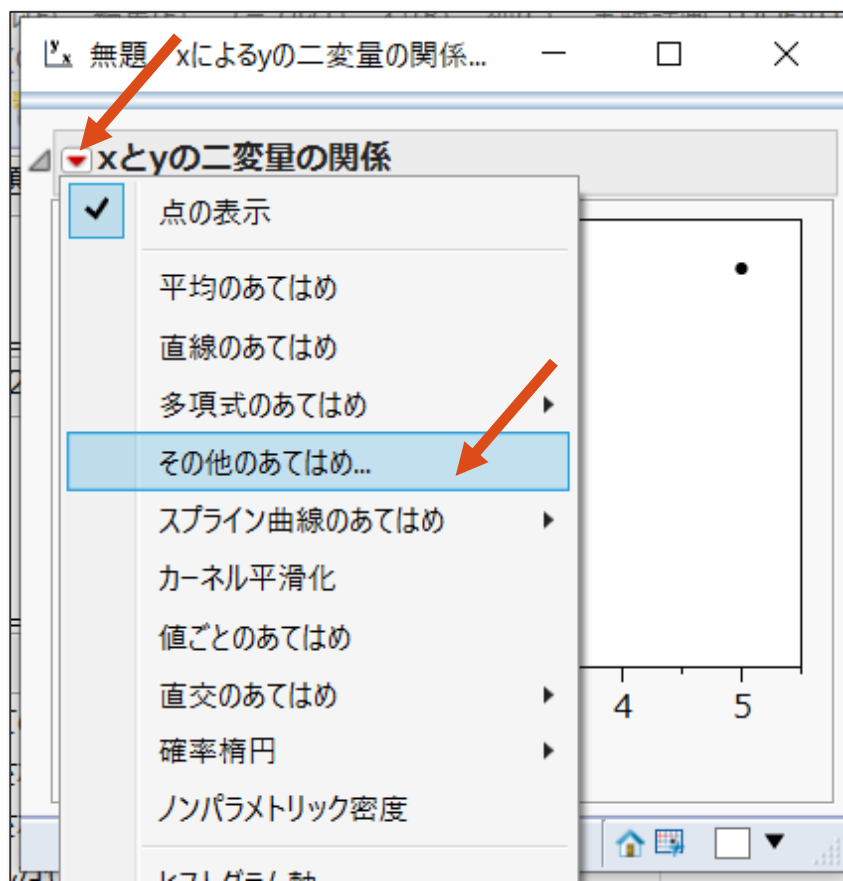


原点を通る回帰式

●JMP [二変量の関係]

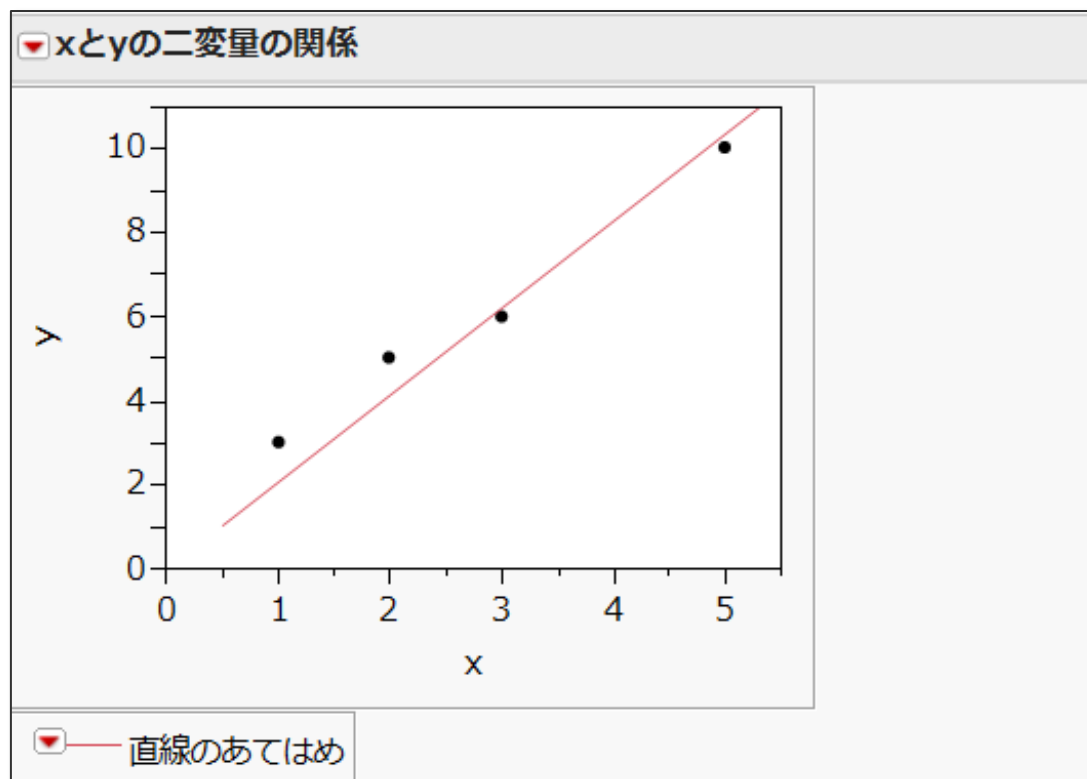
トップメニュー [分析] > [二変量の関係]
▼ [その他のあてはめ] > [切片の制約]

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



●JMP [二変量の関係]

[あてはめの要約] で [R2乗] と [自由度調整R2乗] は空白
原点を通る回帰分析では、寄与率で判断しないようにしている



原点を通る回帰式

空白

直線のあてはめ	
$y = 0 + 2.0769231 * x$	
あてはめの要約	
R2乗	.
自由度調整R2乗	.
誤差の標準偏差(RMSE)	0.767948
Yの平均	6
オブザベーション(または重みの合計)	4

原点を通る回帰式

●JMP [二変量の関係]

平方和に $y=0$ を基準とした 2 乗和を使用
 ($y = \bar{y}$ を基準とした偏差平方和ではない)
 $S_T = S_R + S_e$ が成立
 $177.00 = 168.23 + 1.77$

分散分析				
要因	自由度	平方和	平均平方	F値
モデル	1	168.23077	168.231	285.2609
誤差	3	1.76923	0.590	p値(Prob>F)
全体(修正済み)	4	170.00000		0.0005*

平方和や検定は、 $Y=0$ に固定した時のものです。

パラメータ推定値				
項	推定値	標準誤差	t値	p値(Prob> t)
切片 制約	0	0	.	.
x	2.0769231	0.12297	16.89	0.0005*

切片は 0

(c) 原点を通る (LINEST関数のR^2)

No.	x	y	yhat	e
1	1	3	2.08	0.92
2	2	5	4.15	0.85
3	3	6	6.23	-0.23
4	5	10	10.38	-0.38
平均				
2 乗和	39.00	170.00	168.23	1.77
積和	81.00			
平方和				
積和				
b				
		2.08		
R^2				
		0.990		

●回帰分析適用上の諸問題

散布図の重要性

Anscombe の例はそれを具体的に示す

回帰式の基本 (§4.3(2)) を忘れずに

単回帰モデルの前提

- (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)

重回帰式

複数の説明変数を使って回帰式を求めた
重回帰分析, 多変量解析に発展する手法



- 作成 片瀬雅彦
- 監修 松本一彦、長谷文雄
- 作成時期 2019年3月14日
- 改訂 2019年7月1日、2022年5月31日
2025年5月9日