２ １組のデータの解析

2.3 対数変換と対数正規分布

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

1.3 対数変換と対数正規分布

p.81

テキストの
該当ページ

（1）対数正規分布の例
（2）中心極限定理による対数正規分布の導出
（3）対数変換値の標準偏差
（4）その他の変換

使用するファイル
　Excel ファイル「基本改2.xls」
　JMP ファイル「2-演習.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 の出力を表示

★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

●対数変換と対数正規分布

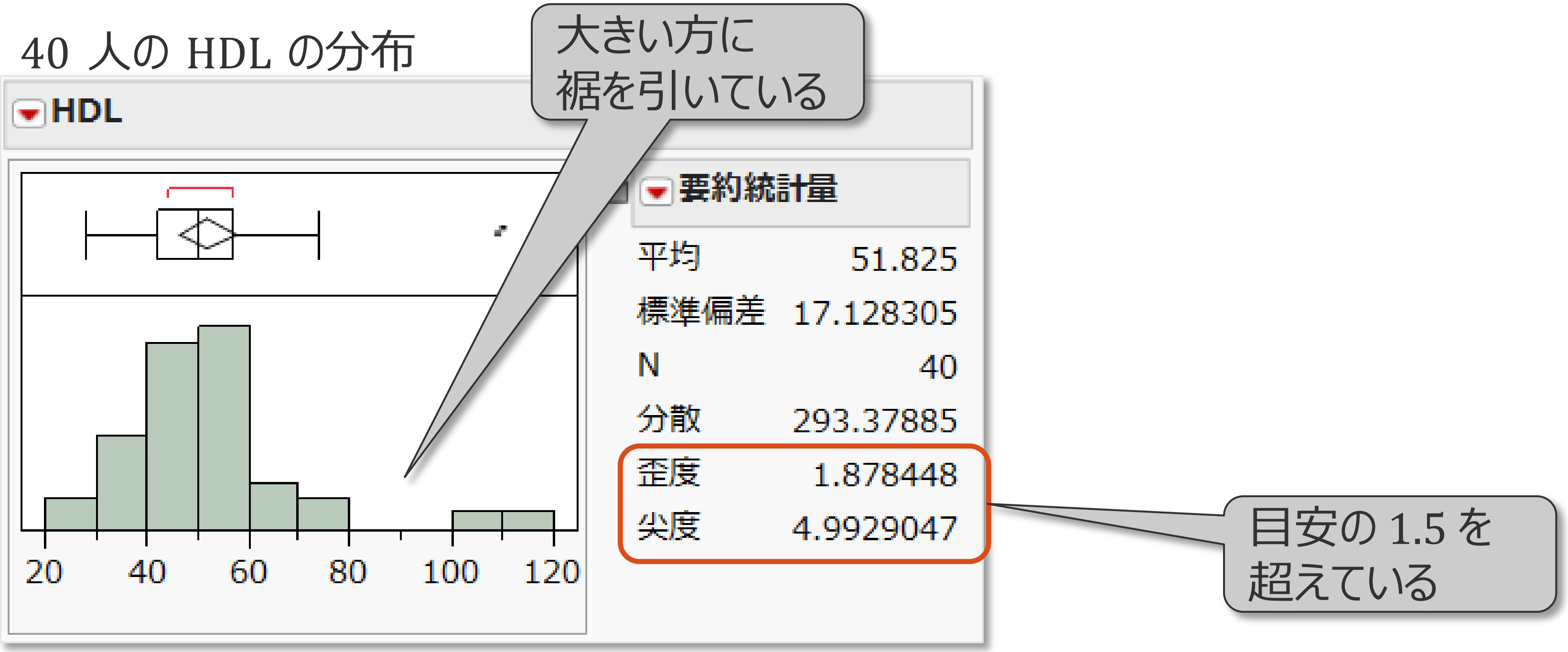
ひずみ（歪度）が正の値をとり，とがり（尖度）が大きく，大きい方に裾を引く分布

このような分布が少なくない

例えば、血液の生化学検査項目の HDL や酵素系成分（ALT, AST, ALP）の分布

（歪度と尖度が ±1.5 を超えると、正規分布からの外れ、小数個の外れ値の存在を検討、§2.1）

40 人の HDL の分布



はじめに

p.81

●対数変換と対数正規分布

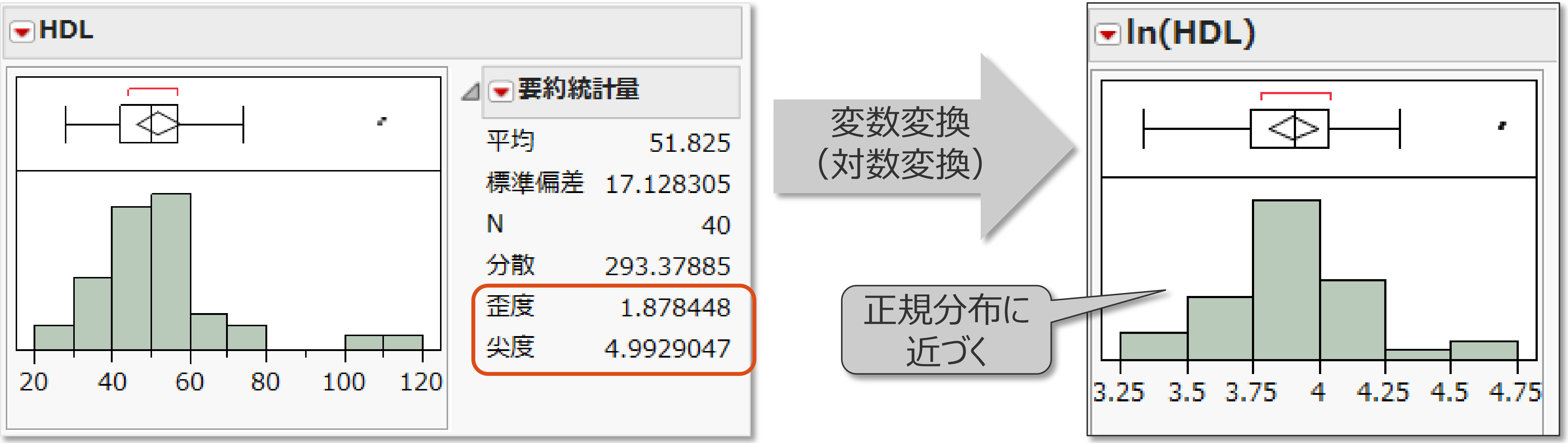
観測値を対数変換すると正規分布に近くなる場合、元の分布を「対数正規分布」という

統計解析の手法の多くは、データが正規分布に従い、等分散であることを条件としている

この条件が満たされないときに変数変換を検討する場面は少なくない

対数変換による等分散化（§3.4（2 組のデータの場合）、第 3 部 §2.1）

40 人の HDL の分布



はじめに

●対数

対数は、「ある数（a）を何回（X）掛け合わせると、別の数（x）になるか」を示す

$$a^X = x \quad \Longleftrightarrow \quad \log_a x = X \qquad (a > 0, a \neq 0)$$

a：底（てい）
X：a を底とする x の対数

例えば、$a = 2$ を $X = 3$ 回掛け合わせると $x = 8$ になる

$$2^3 = 8 \qquad \log_2 8 = 3$$

常用対数：$a = 10$ の対数

$$10^2 = 100 \qquad \log_{10} 100 = 2 \quad \rightarrow \quad \log 100 = 2$$

自然対数：$a = e = 2.71828\cdots$ の対数（e：ネイピア数）

$$e^2 = 7.389056\cdots \qquad \log_e 7.389056 \approx 2 \quad \rightarrow \quad \ln 7.389056 \approx 2$$

（数学の分野では自然対数を $\log 7.389056$ と表すこともある）

はじめに

p.120

●対数

2 を底とする x の対数

$\log_2 x$

2 を底とする 4 の対数は 2

$\log_2 4 = 2$、 $2^2 = 4$

$4^3 = 2^6 = 64$、 $\log_2 4 = 2$、 $\log_2 64 = 6$ なので

$\log_2 64 = \log_2 4^3 = 3 \times \log_2 4 = 3 \times 2 = 6$

乗算が加算に置き換えられる

$4 \times 8 = 32$

$2^2 \times 2^3 = 2^5$・・・指数の 2, 3, 5 は 4, 8, 32 の対数（底は 2）

$\log_2 4 + \log_2 8 = \log_2 32$ ・・・乗算が加算に置き換えられる

以上の性質は，底が 2 であることに限定されず，一般的に成立

2 を底とする x の対数

x	0.25	0.5	1	2	4	8
2 のべき乗	2^{-2}	2^{-1}	2^0	2^1	2^2	2^3
$\log_2(x)$	-2	-1	0	1	2	3

10を底とする x の対数、常用対数

x	0.01	0.1	1	10	100	1000
10 のべき乗	10^{-2}	10^{-1}	10^0	10^1	10^2	10^3
$\log_{10}(x)$	-2	-1	0	1	2	3

はじめに

p.120

●対数

常用対数：10を底とする x の対数　$\log_{10} x = \log x$

log 10 は 10^x の逆関数

$\sqrt{10} = 3.16\cdots$ の常用対数

$$\left(\sqrt{10}\right)^2 = 10^1 \quad \sqrt{10} = 10^{0.5} = 10^{\frac{1}{2}}$$

$$\log\sqrt{10} = \log 10^{0.5} = 0.5 \log_{10} 10 = 0.5$$

Excel 関数

=LOG(2)=0.30103・・・、=LOG(SQRT(10))=0.5、=LOG(10^2)=2

初項 0.1 、公比 $\sqrt{10}$ の等比数列

x	0.1	$0.1\times\sqrt{10}$	$0.1\times\sqrt{10}^2$	$1\times\sqrt{10}$	$\sqrt{10}\times\sqrt{10}$	$10\times\sqrt{10}$
x	0.1	$0.316\cdots$	1	$3.16\cdots$	10	$31.6\cdots$
$\log x$	-1	-0.5	0	0.5	1	1.5

2 を底とする x の対数

x	0.25	0.5	1	2	4	8
2 のべき乗	2^{-2}	2^{-1}	2^0	2^1	2^2	2^3
$\log_2(x)$	-2	-1	0	1	2	3

10を底とする x の対数、常用対数

x	0.01	0.1	1	10	100	1000
10 のべき乗	10^{-2}	10^{-1}	10^0	10^1	10^2	10^3
$\log_{10}(x)$	-2	-1	0	1	2	3

はじめに

p.120

●対数

自然対数：ネイピア数 e を底とする x の対数（e=2.71828･･･） $\log_e x = \ln x$

$\ln x$ は e^x の逆関数

$$e^X = x, \quad X = \ln x$$

指数を以下のように記述（指数 X の部分が分数や複雑な式になるとき、記述しやすい）

$$e^X = \exp(X)$$

Excel 関数

=LN(10)=2.302585･･･　　=LN(10.2)=2.322387･･･　　=LN(EXP(10))=10

	数式	Excel 関数	JMP 関数
a を底とする対数	$\log_{\mathrm{a}} x$	=LOG(x, a)	log a (x)
常用対数	$\log x$	=LOG(x)	log10(x)
自然対数	$\ln x$	=LN(x)	log(x)

はじめに

●対数

$a > 0, a \neq 1, m > 0, n > 0$　として

$\log_a 1 = 0,\qquad \log_a a = 1$

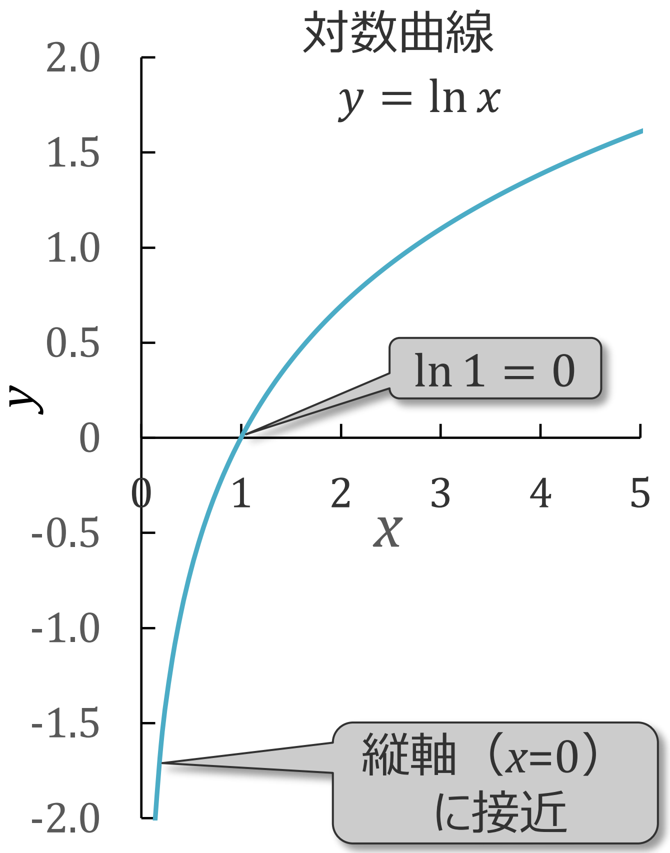
$\log_a mn = \log_a m + \log_a n$　（掛け算が足し算になる）

$\log_a \frac{m}{n} = \log_a m - \log_a n$　（割り算が引き算になる）

$\log_a m^r = r \log_a m$　（r 乗が r 倍になる）

$\log_a b = \log_c b / \log_c a$　（底の変換）

$\log_e x = \log_{10} x / \log_e e \sim 0.2303 \log_{10} x$
（常用対数と自然対数は定数倍の関係）



（1）対数正規分布の例

1 組のデータの解析

事例 1 ：20 個のデータ事例
事例 3 ：血液検査の HDL 値

対数正規分布の例（1）

p.81

●事例 1

Excelファイルの読み込みと表示

Excel ファイル「基礎改2.xls」

表示 2.3.1に移動

（名前ボックスで

Fig23_01 を選択）

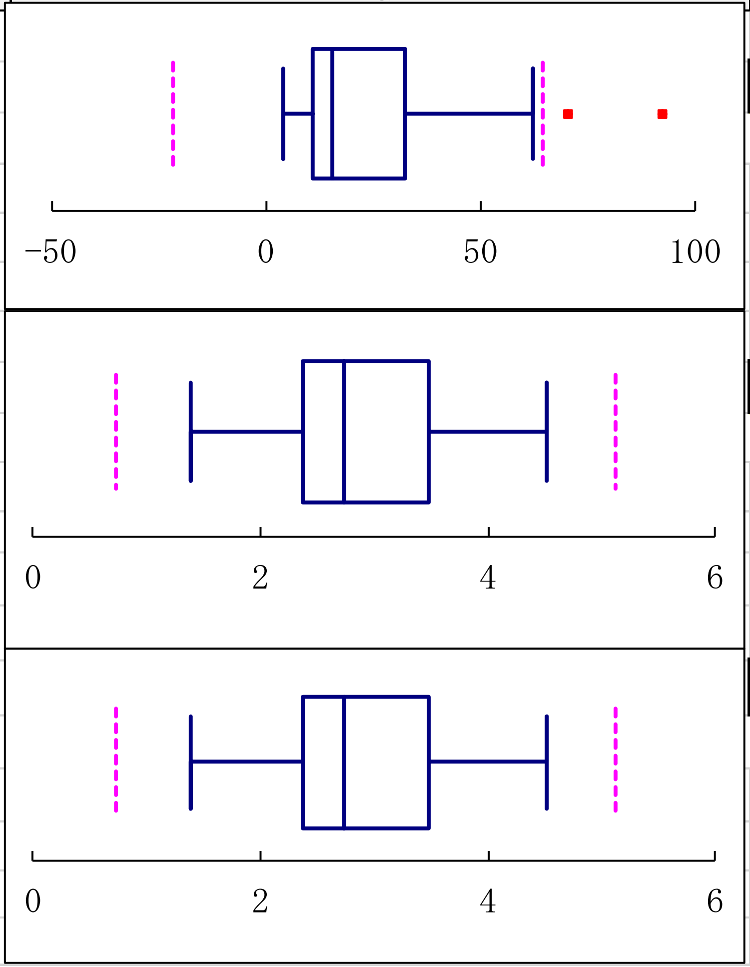
1 組のデータ

n=20

表示 2.3.1　x と ln(x) の分布と外れ値

	A	B	C	D	E	F
3		x		ln(x)		
4		14	9	2.64	2.20	
5		11	62	2.40	4.13	
6		4	35	1.39	3.56	
7		10	**92**	2.30	4.52	
8		10	7	2.30	1.95	
9		**70**	28	4.25	3.33	
10		13	12	2.56	2.48	
11		26	33	3.26	3.50	
12		17	32	2.83	3.47	
13		18	11	2.89	2.40	
14	平均		25.7		2.92	
15	標準偏差		23.5		0.82	
16	ひずみ		1.73		0.35	
17	変動係数		0.91			
18	棄却上限		85.8		5.0	
19						

G



H

K6
B4:C13

O6
D4:E13

S6
B4:C13　L

対数正規分布の例（1）

●事例１

１組のデータの事例

マクロによる箱ひげ図の作成

（§2.1）

対数変換前の値（上図）

20 個の観測値 x

ひずみが 1.73 ＞1.5（目安）

中央値が箱の左寄り

左のひげ ＜右のひげ

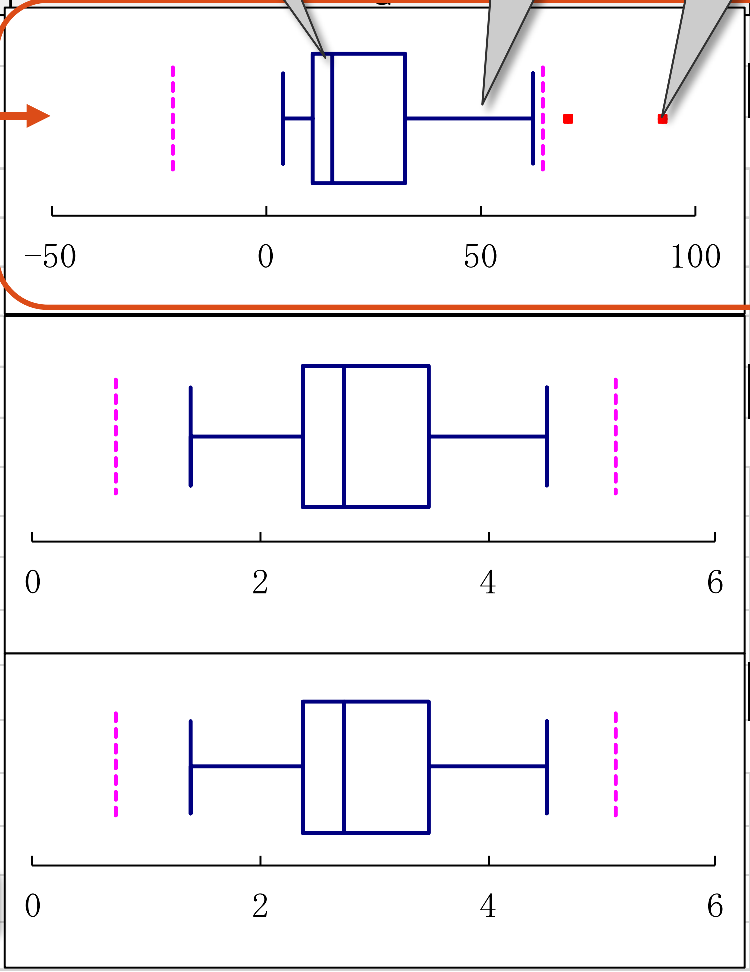
分布が右に裾を引いている

外れ値として２つ表示

表示 2.3.1　x と $\ln(x)$ の分布と外れ値

	A	B	C	D	E	F
3		x		ln(x)		
4		14	9	2.64	2.20	
5		11	62	2.40	4.13	
6		4	35	1.39	3.56	
7		10	**92**	2.30	4.52	
8		10	7	2.30	1.95	
9		**70**	28	4.25	3.33	
10		13	12	2.56	2.48	
11		26	33	3.26	3.50	
12		17	32	2.83	3.47	
13		18	11	2.89	2.40	
14	平均		25.7		2.92	
15	標準偏差		23.5		0.82	
16	ひずみ		1.73		0.35	
17	変動係数		0.91			
18	棄却上限		85.8		5.0	
19						

G



H
K6
B4:C13
O6
D4:E13
S6
B4:C13 L

中央値

大きい方に裾を引く

外れ値として表示

外れ値として表示

正、＞1.5

対数正規分布の例（1）

p.81

●事例 1

対数変換値 ln(*x*) （中図）

ひずみが 0.35、0に近い

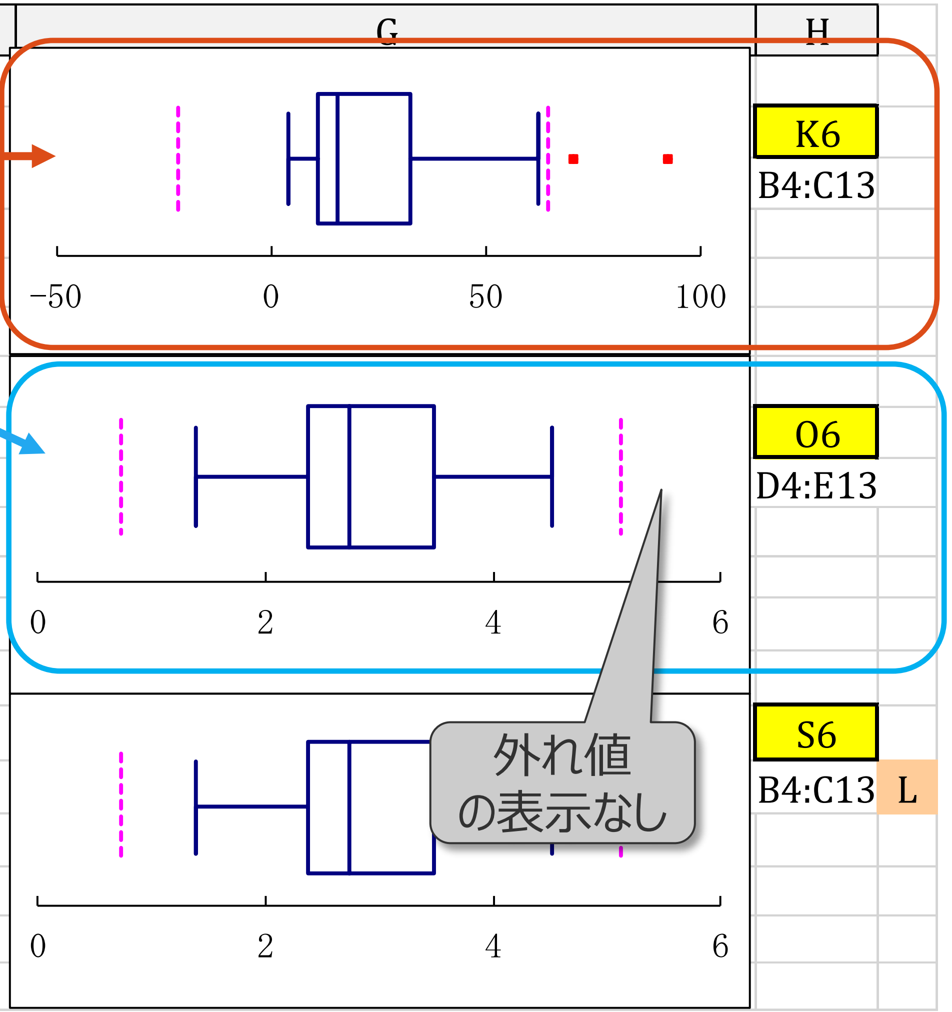
（正規分布に近い）

外れ値の表示がなくなった

表示 2.3.1　*x* と ln(*x*) の分布と外れ値

	A	B	C	D	E	F
3		x		ln(x)		
4		14	9	2.64	2.20	
5		11	62	2.40	4.13	
6		4	35	1.39	3.56	
7		10	**92**	2.30	4.52	
8		10	7	2.30	1.95	
9		**70**	28	4.25	3.33	
10		13	12	2.56	2.48	
11		26	33	3.26	3.50	
12		17	32	2.83	3.47	
13		18	11	2.89	2.40	
14	平均		25.7		2.92	
15	標準偏差		23.5		0.82	
16	ひずみ		1.73		0.35	
17	変動係数		0.91			
18	棄却上限		85.8		5.0	
19						

0 に近い



対数正規分布の例（1）

p.81

●事例１

対数変換値 ln(*x*)（中図）

ひずみが 0.35、0に近い

（正規分布に近い）

外れ値の表示がなくなった

対数変換値 ln(*x*)（下図）

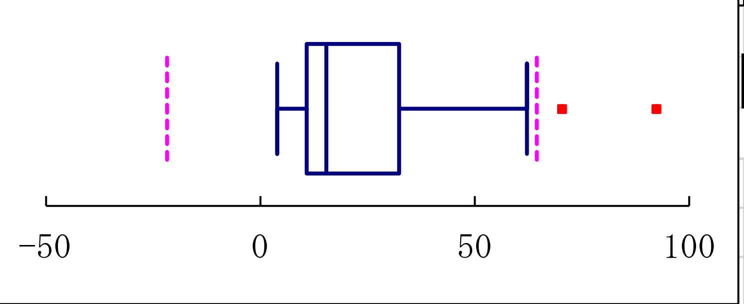
マクロのスイッチ「L」で

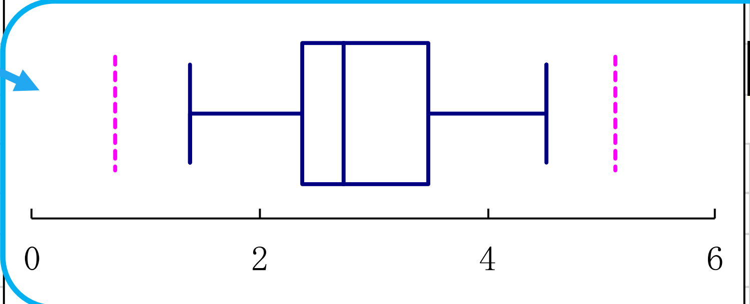
データを対数変換して

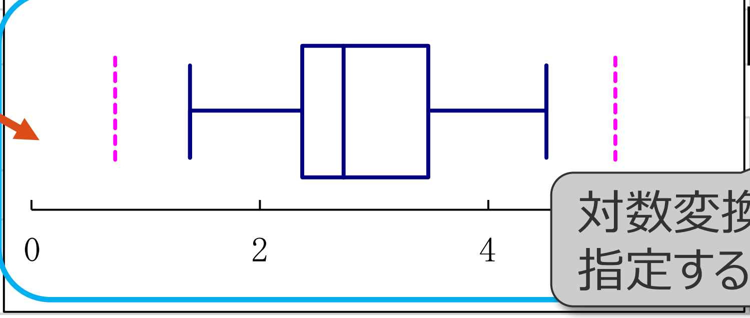
箱ひげ図を作成

表示 2.3.1　*x* と ln(*x*) の分布と外れ値

	A	B	C	D	E	F	G	H
3		x		ln(x)				
4		14	9	2.64	2.20			K6
5		11	62	2.40	4.13			B4:C13
6		4	35	1.39	3.56			
7		10	**92**	2.30	4.52			
8		10	7	2.30	1.95			
9		**70**	28	4.25	3.33			
10		13	12	2.56	2.48			O6
11		26	33	3.26	3.50			D4:E13
12		17	32	2.83	3.47			
13		18	11	2.89	2.40			
14	平均		25.7		2.92			
15	標準偏差		23.5		0.82			
16	ひずみ		1.73		0.35			S6
17	変動係数		0.91					B4:C13　L
18	棄却上限		85.8		5.0			
19								







対数変換を
指定する場所

対数正規分布の例（1）

●事例１

変動係数

x の変動係数：0.91

$$\mathrm{CV} = \frac{23.5}{25.7} = 0.91$$

$\ln(x)$ の変動係数

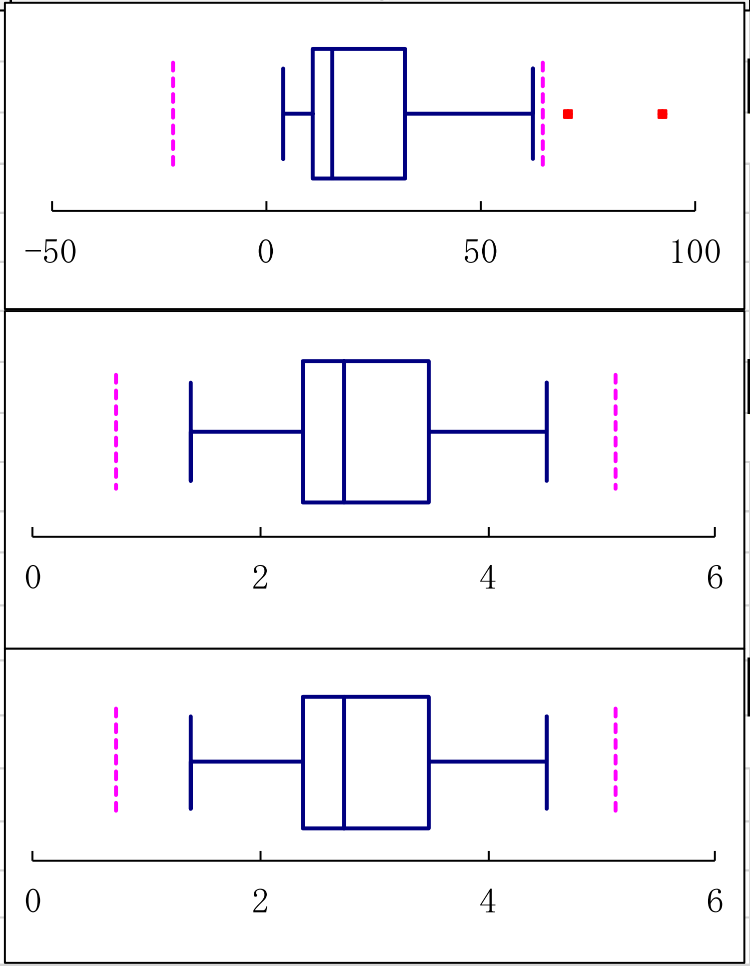
計算しない

対数変換後の変動係数は

意味を持たない

表示 2.3.1　x と $\ln(x)$ の分布と外れ値

	A	B	C	D	E	F	H	
3		x		ln(x)				
4		14	9	2.64	2.20		K6	
5		11	62	2.40	4.13		B4:C13	
6		4	35	1.39	3.56			
7		10	**92**	2.30	4.52			
8		10	7	2.30	1.95			
9		**70**	28	4.25	3.33			
10		13	12	2.56	2.48		O6	
11		26	33	3.26	3.50		D4:E13	
12		17	32	2.83	3.47			
13		18	11	2.89	2.40			
14	平均		25.7		2.92			
15	標準偏差		23.5		0.82			
16	ひずみ		1.73		0.35		S6	
17	変動係数		0.91				B4:C13	L
18	棄却上限		85.8		5.0			
19								



意味を持たないので空白

対数正規分布の例（2）

変動係数の利点

●対数変換値の変動係数は意味を持たない

事例２：５人の身長を cm と inchで表示 ：両者の平均と標準偏差は不一致、変動係数は一致

これを自然対数に変換して表示：両者の標準偏差は一致、変動係数は一致しない

対数変換値は間隔尺度（§2.1 (5)、p.60）になるので、

変動係数を計算してはならない、またさらに対数変換してはならない

対数変化前の変動係数と対数変換後の標準偏差がほぼ一致（後述）

x	ln(x)
0	
0.5	-0.69
5.8	1.76
10.6	2.36
30.8	3.43
60.8	4.11
93.7	4.54

単位の取り方が
異なっても CV は一致

	身長		ln(身長)	
	(cm)	(inch)	(cm)	(inch)
佐藤さん	165	65.0	5.11	4.17
鈴木さん	178	70.1	5.18	4.25
高橋さん	194	76.4	5.27	4.34
田中さん	173	68.1	5.15	4.22
中村さん	148	58.3	5.00	4.07
平均	171.6	67.6	5.14	4.21
標準偏差	16.9	6.662	0.100	0.100
変動係数	0.099	0.099	0.019	0.024

対数正規分布の例（2）

p.82

●対数変換値の変動係数は意味を持たない

変動係数の利点

事例２：５人の身長を cm と inchで表示 ：両者の平均と標準偏差は不一致、変動係数は一致

これを自然対数に変換して表示：両者の標準偏差は一致、変動係数は一致しない

対数変換値は間隔尺度（§2.1 (5)、p.60）になるので、

変動係数を計算してはならない、またさらに対数変換してはならない

対数変化前の変動係数と対数変換後の標準偏差がほぼ一致（後述）

x	ln(x)
0	
0.5	-0.69
5.8	1.76
10.6	2.36
30.8	3.43
60.8	4.11
93.7	4.54

絶対的な0が定義できない

負の値になることもある

単位の取り方が異なっても CV は一致

	身長		ln(身長)	
	(cm)	(inch)	(cm)	(inch)
佐藤さん	165	65.0	5.11	4.17
鈴木さん	178	70.1	5.18	4.25
高橋さん	194	76.4	5.27	4.34
田中さん	173	68.1	5.15	4.22
中村さん	148	58.3	5.00	4.07
平均	171.6	67.6	5.14	4.21
標準偏差	16.9	6.662	0.100	0.100
変動係数	0.099	0.099	× 0.019	0.024

単位によらず一致

誤り

対数正規分布の例（2）

p.82

●対数変換値の変動係数は意味を持たない

変動係数の利点

事例２：５人の身長を cm と inchで表示 ：両者の平均と標準偏差は不一致、変動係数は一致

これを自然対数に変換して表示：両者の標準偏差は一致、変動係数は一致しない

対数変換値は間隔尺度（§2.1 (5)、p.60）になるので、

変動係数を計算してはならない、またさらに対数変換してはならない

対数変化前の変動係数と対数変換後の標準偏差がほぼ一致（後述）

x	ln(x)
0	
0.5	-0.69
5.8	1.76
10.6	2.36
30.8	3.43
60.8	4.11
93.7	4.54

絶対的な0が定義できない

負の値になることもある

単位の取り方が異なっても CV は一致

	身長		ln(身長)	
	(cm)	(inch)	(cm)	(inch)
佐藤さん	165	65.0	5.11	4.17
鈴木さん	178	70.1	5.18	4.25
高橋さん	194	76.4	5.27	4.34
田中さん	173	68[illegible]	[illegible]15	4.22
中村さん	148	58.3	5.00	4.07
平均	171.6	67.6	5.14	4.21
標準偏差	16.9	6.662	0.100	0.100
変動係数	0.099	0.099	0.019	0.024

ほぼ一致

×

誤り

対数正規分布の例（3）

●事例３

JMP ファイル「2-演習.jmp」を読み込み

40人の健康診断のデータを利用

１組のデータの事例として HDL（HDL コレステロールの値）を取り上げる（課題 2.3）

１組のデータの事例

	No.	年齢	身長	体重	血圧(上)	血圧(下)	TC	HDL	中性脂肪	TN	血糖	BMI	血圧差
1	1	40	169	75	126	78	200	37	141	7.5	95	0.0026259	48
2	2	41	164	75	104	70	197	36	155	7.4	88	0.0027885	34
3	3	42	161	68	152	92	300	28	580	8.3	117	0.0026233	60
4	4	43	169	73	120	92	199	52	96	6.7	83	0.0025559	28
5	5	43	177	63	112	64	183	52	95	7.4	89	0.0020109	48
6	6	43	166	57	118	82	224	67	113	7.3	94	0.0020685	36
7	7	43	170	65	118	72	194	47	183	6.9	83	0.0022491	46

対数正規分布の例（3）

●事例 3

対数変換した列を作成

JMP に新しい列を作成、対数変換の式を入力（JMP で対数変換を行う）

対数変換の前と後のデータを比較

［一変量の分布］による解析

対数変換前

列を作成して
対数変換

	No.	年齢	身長	体重	血圧(上)	血圧(下)	TC	HDL	中性脂肪	TN	血糖	BMI	血圧差	
1	1	40	169	75	126	78	200	37	141	7.5	95	0.0026259	48	
2	2	41	164	75	104	70	197	36	155	7.4	88	0.0027885	34	
3	3	42	161	68	152	92	300	28	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	
4	4	43	169	73	120	92	199	52	96	6.7	83	0.0025559	28	
5	5	43	177	63	112	64	183	52	95	7.4	89	0.0020109	48	
6	6	43	166	57	118	82	224	67	113	7.3	94	0.0020685	36	
7	7	43	170	65	118	72	194	47	183	6.9	83	0.0022491	46	

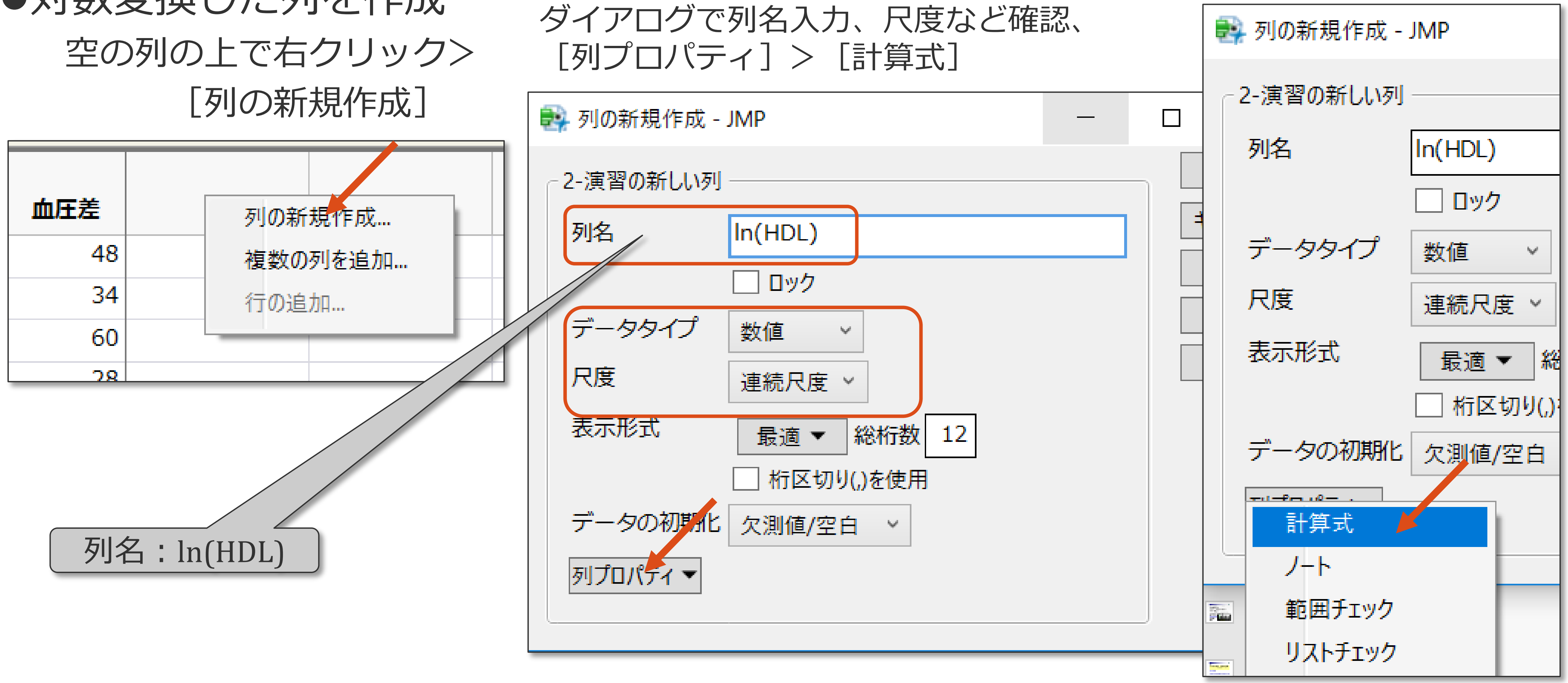
対数正規分布の例（3）

p.82

●対数変換した列を作成

空の列の上で右クリック>［列の新規作成］

ダイアログで列名入力、尺度など確認、［列プロパティ］＞［計算式］



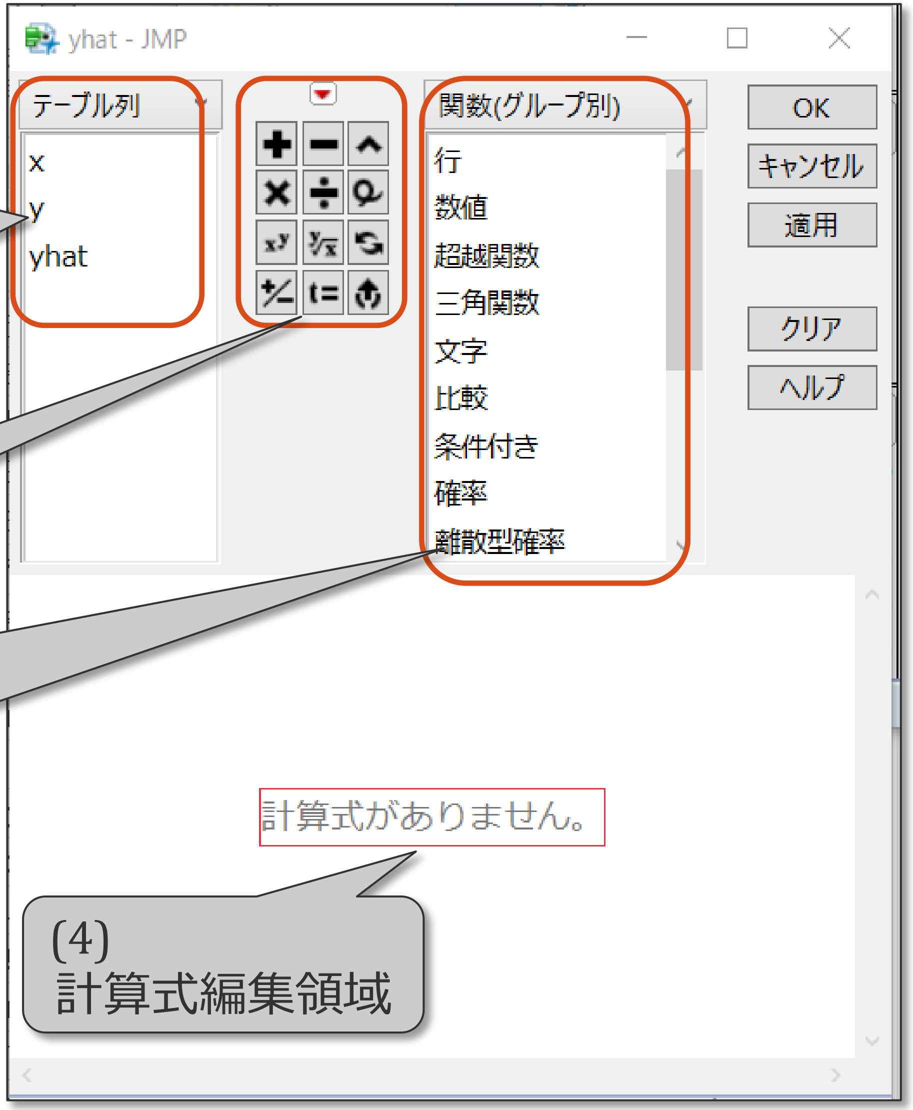
対数正規分布の例（3）

p.82

● 対数変換した列を作成

計算式エディタによる計算式の入力

計算式エディタ



(1)
定数
テーブル変数
テーブル列

(2)
キーパッドの演算子

(3)
超越関数
三角関数
条件付き
比較　など

(4)
計算式編集領域

「関数」、超越関数、Log、「テーブル列」、HDL

対数正規分布の例（3）

p.82

●対数変換した列を作成

計算式エディタによる計算式の入力

関数（グループ別）＞超越関数＞ Log

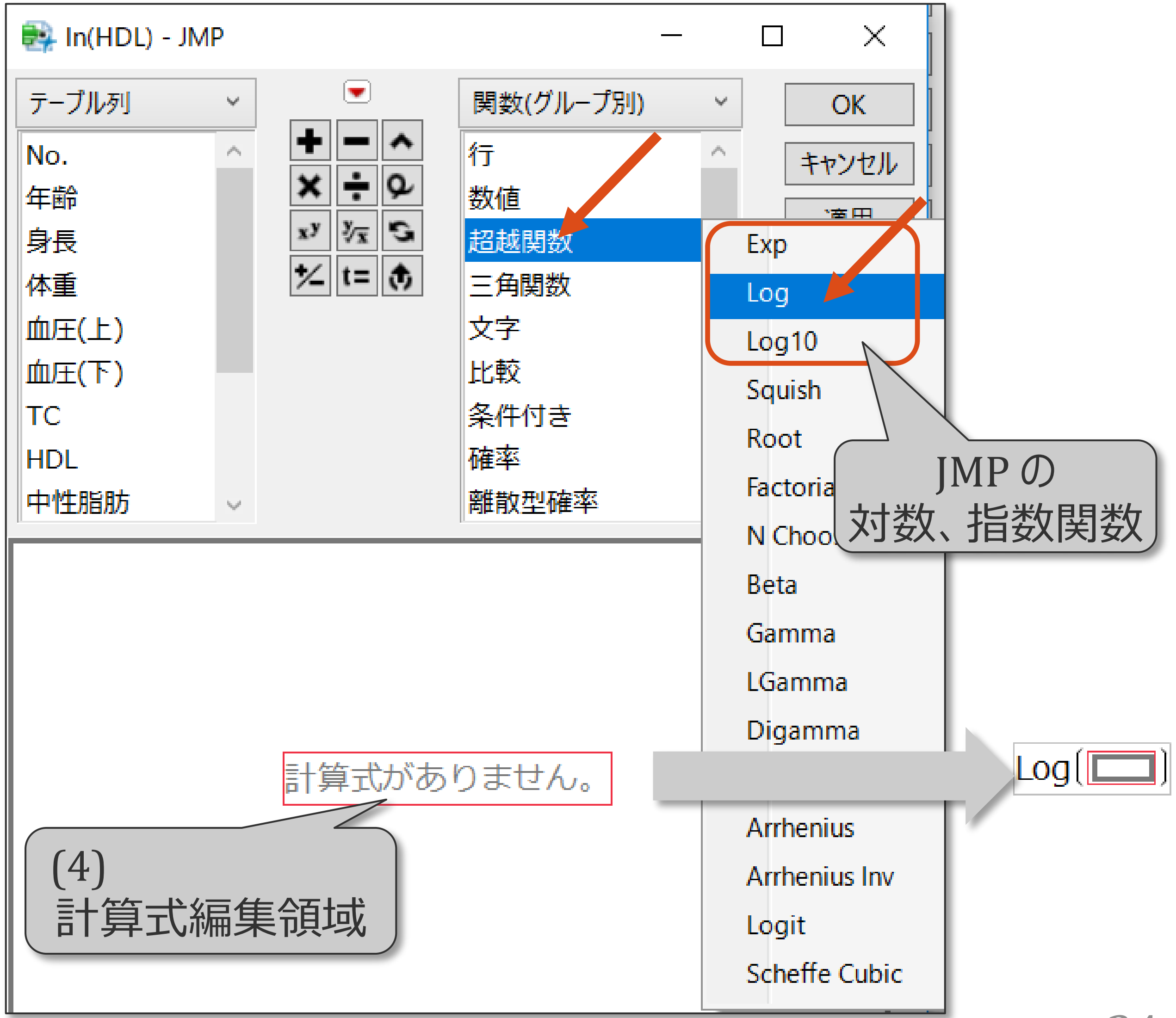
JMPの対数関数、指数関数

自然対数：Log → 選択

常用対数：Log10

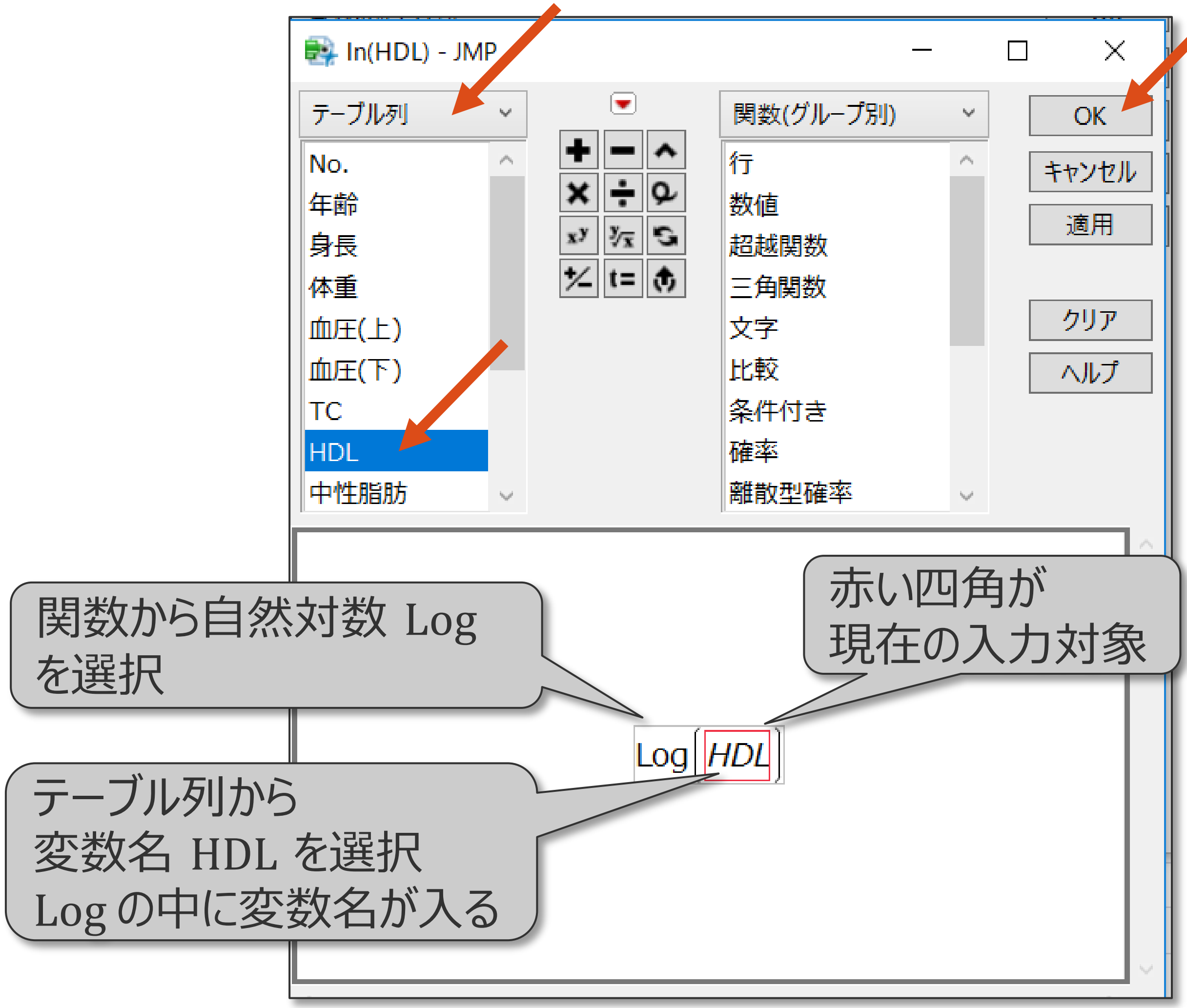
指数関数：Exp

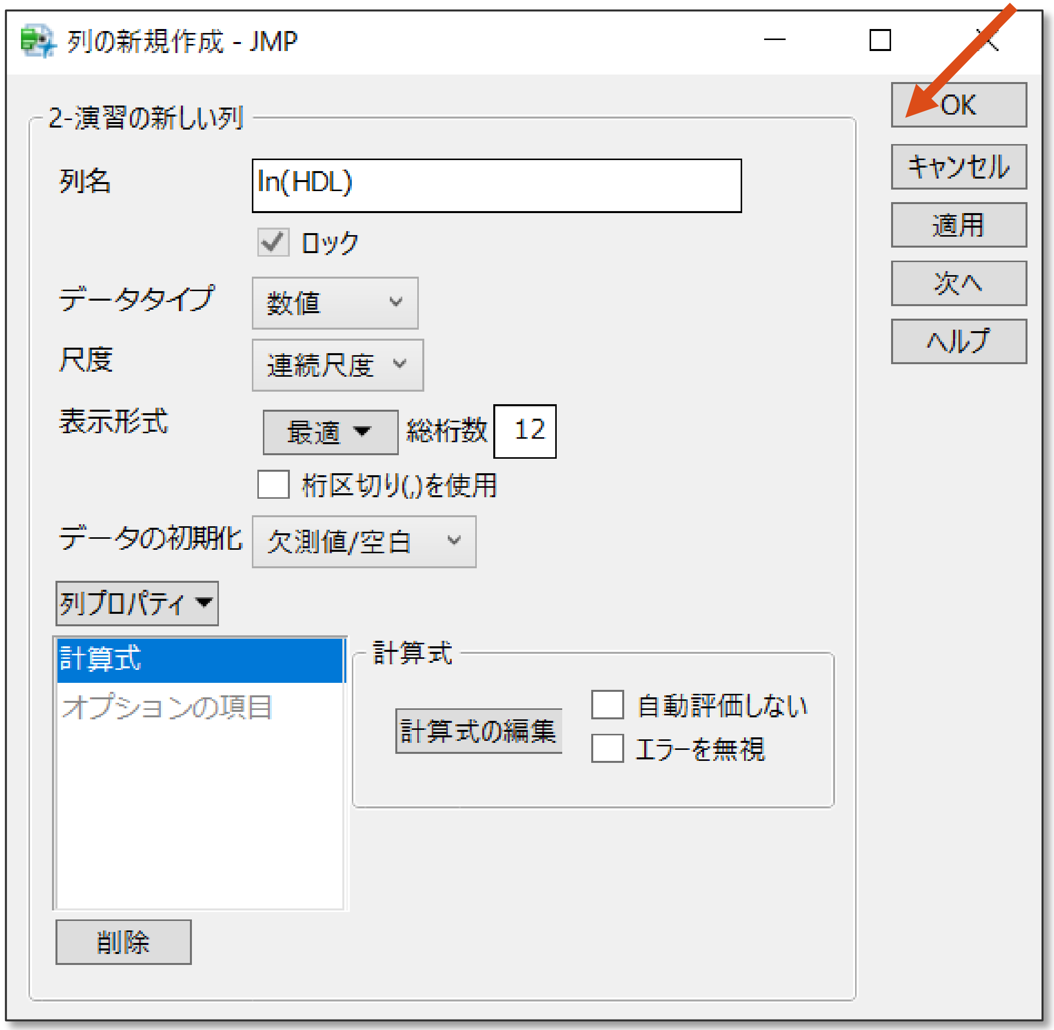
「関数」、超越関数、Log、「テーブル列」、HDL



対数正規分布の例（3）

●対数変換した列を作成





対数正規分布の例（3）

p.82

●対数変換した列を作成

JMP 以外で対数変換をして、
その変換値を JMP に張り付けた場合、
桁落ちによる計算誤差が問題になる場合あり
対数変換などの変数変換は
解析する統計ソフト（JMP）で行う

列の新規作成
対数変換値

TC	HDL	中性脂肪	TN	血糖	BMI	血圧差	ln(HDL)
200	37	141	7.5	95	0.002625	48	3.6109179126
197	36	155	7.4	88	0.002788	34	3.5835189385
300	28	580	8.3	117	0.002623	60	3.3322045102
199	52	96	6.7	83	0.002555	28	3.9512437186
183	52	95	7.4	89	0.002010	48	3.9512437186
224	67	113	7.3	94	0.002068	36	4.2046926194
194	47	183	6.9	83	0.002249	46	3.8501476017
171	59	109	7.4	80	0.002258	42	4.0775374439
154	38	161	7.5	93	0.002515	38	3.6375861597
213	63	169	7.3	103	0.002431	52	4.1431347264
185	49	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	3.8918202981
179	45	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	3.8066624898
196	74	84	7.3	90	0.002497	32	4.3040650932
209	49	140	7.2	82	0.002469	44	3.8918202981
157	110	74	6.9	90	0.002249	64	4.7004803658
196	32	337	7.3	94	0.002584	44	3.4657359028
198	48	109	7.9	80	0.002551	48	3.8712010109
167	29	136	7.2	82	0.002314	46	3.36729583
223	52	79	7.2	84	0.002002	36	3.9512437186
173	109	93	7.2	85	0.002019	52	4.6913478822
153	70	50	7.2	72	0.001806	30	4.248495242

対数変換

対数正規分布の例（3）

●対数変換の前と後のデータを比較

対象：「HDL」「ln(HDL)」

［一変量の分布］による解析（§2.1）

［分析］＞［一変量の分布］

＞［Y,列］に「HDL」「ln(HDL)」を選択

ヒストグラム、要約統計量を表示

（分散、歪度、尖度、変動係数を追加表示）

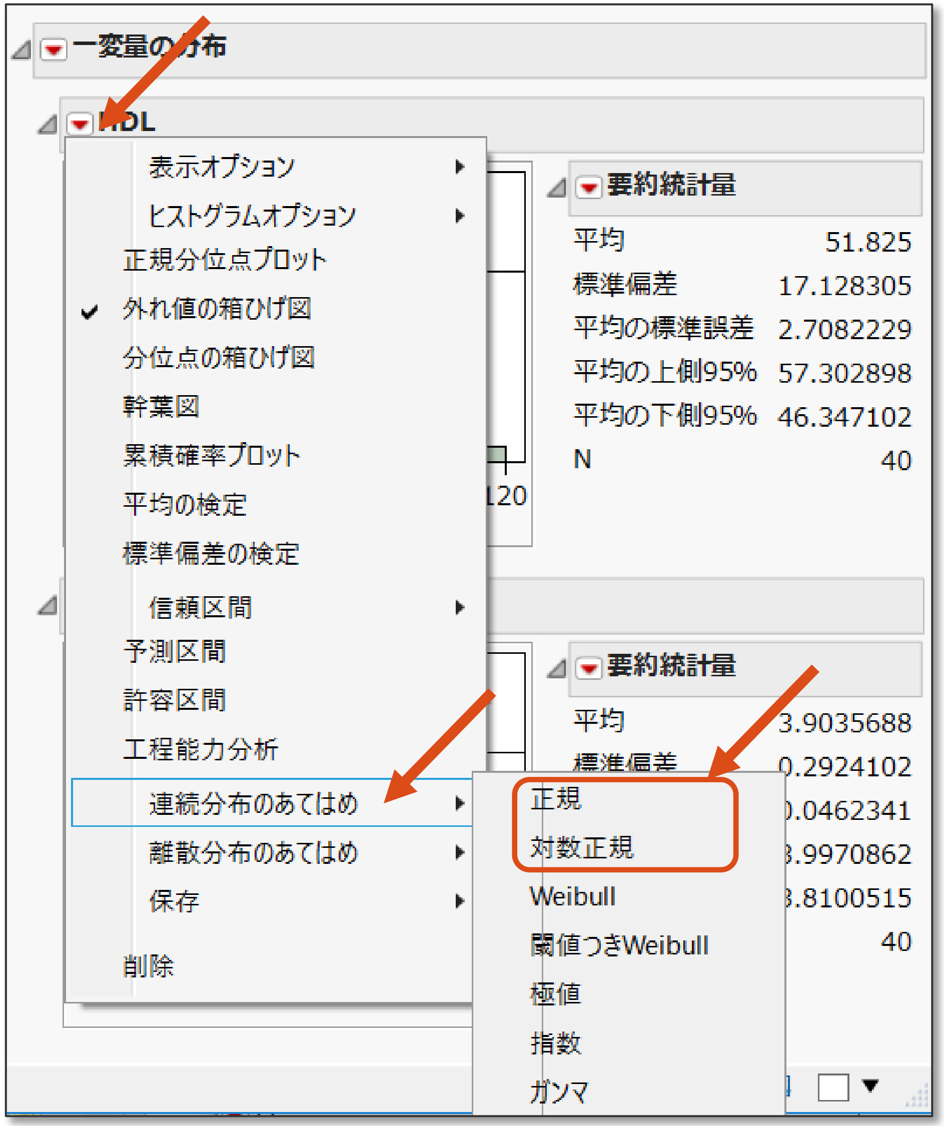
▼＞［積み重ねて表示］

クリックしてチェックを入れる

▼＞［工程能力分析］［連続分布のあてはめ］

＞［正規］、［対数正規］

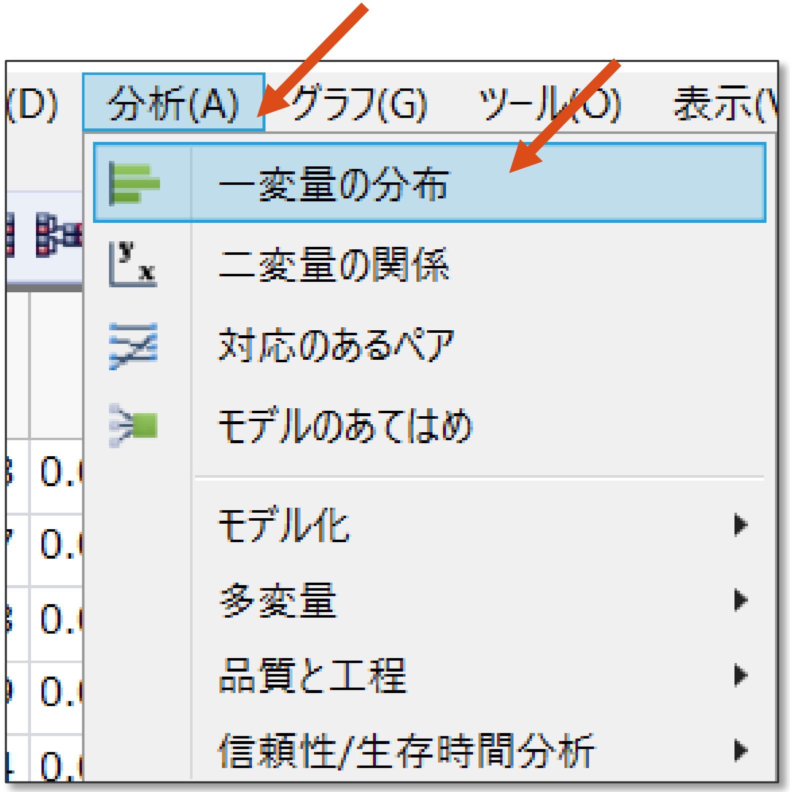
ただし、「ln(HDL)」は［正規］のみ

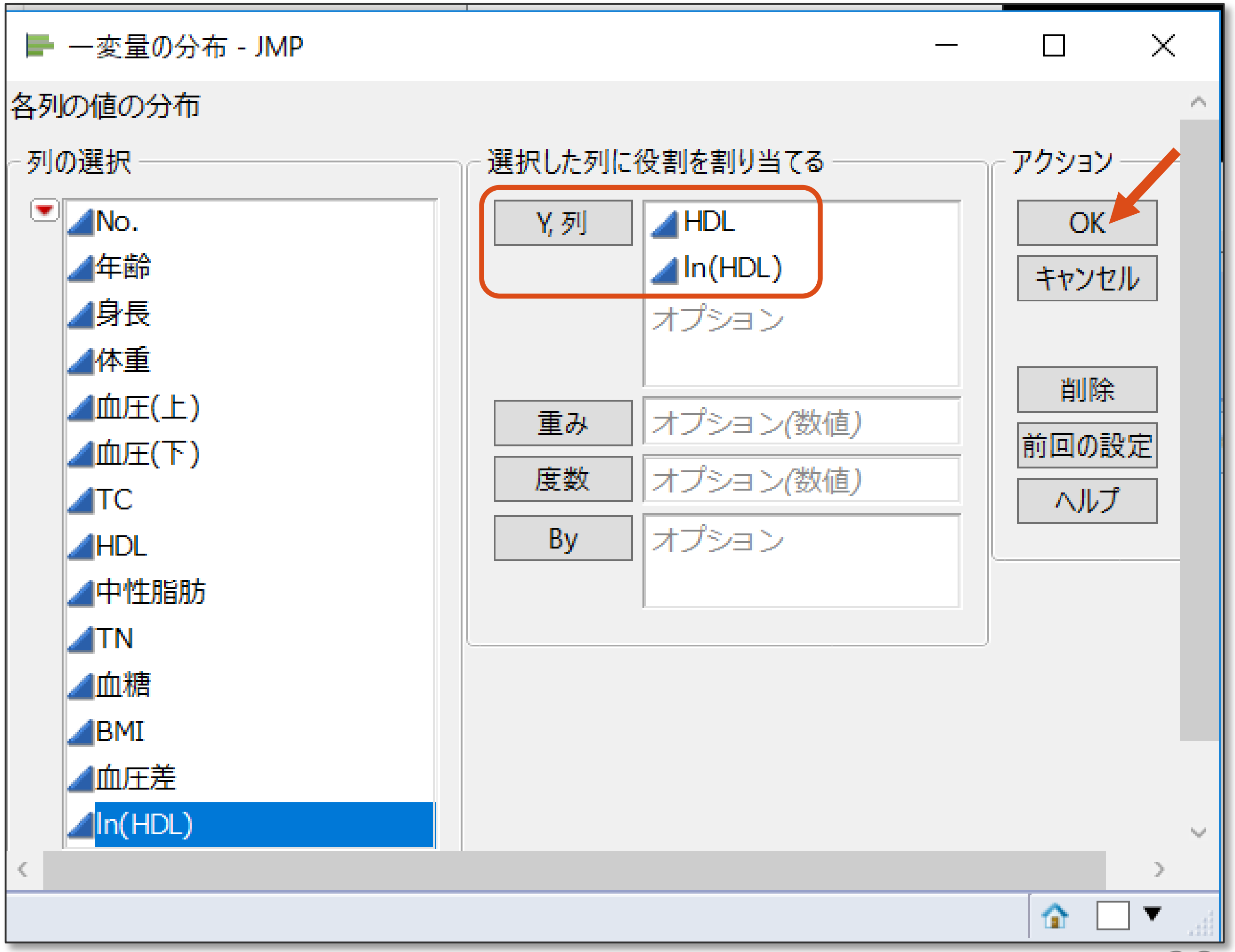


対数正規分布の例（3）

●対数変換の前と後のデータを比較

［分析］＞［一変量の分布］





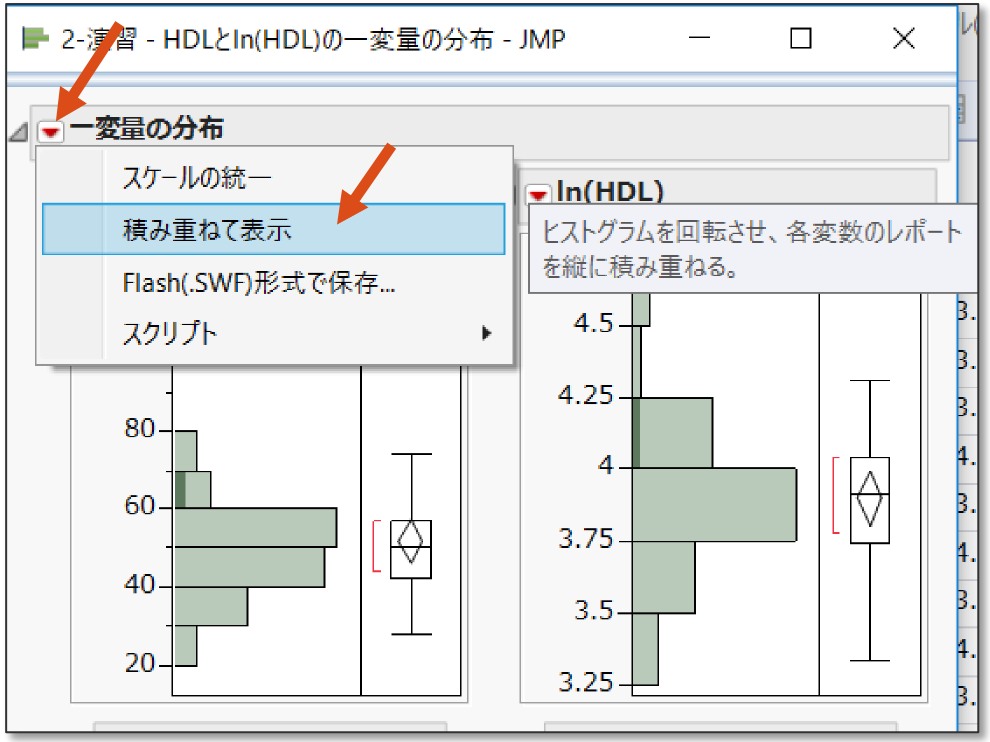
対数正規分布の例（3）

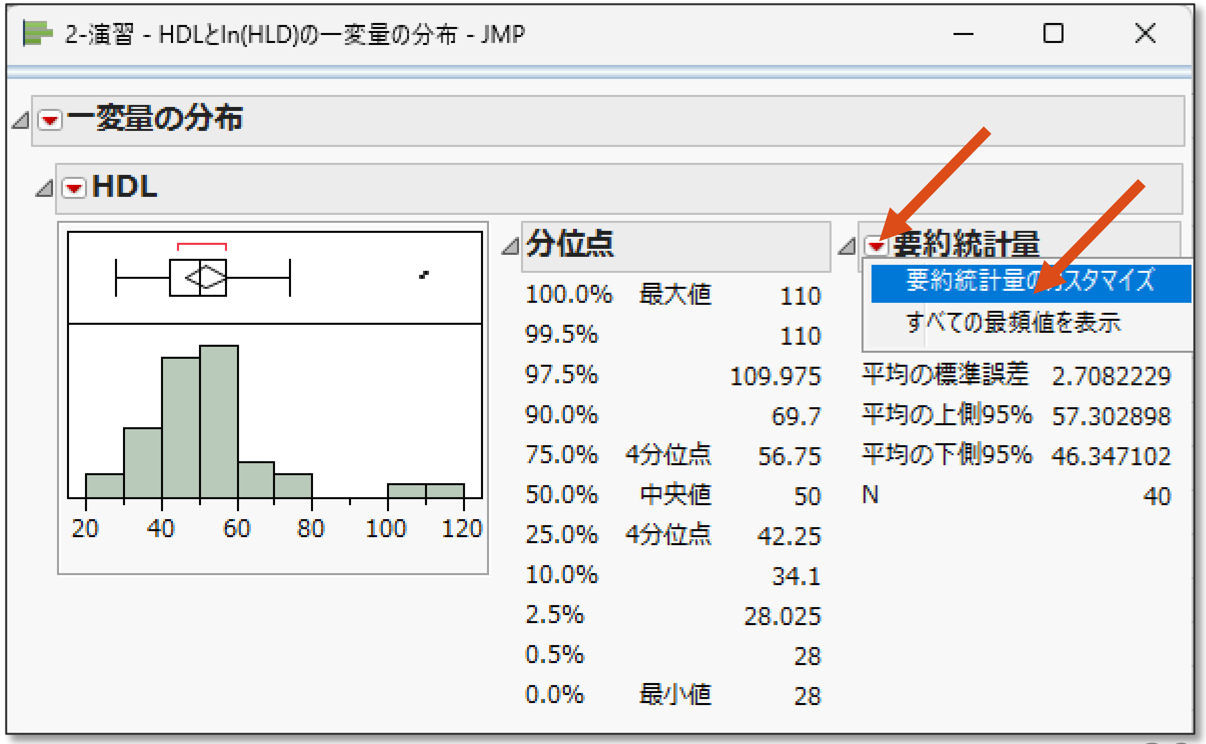
p.83

●対数変換の前と後のデータを比較

［一変量の分布］の▼>［積み重ねて表示］

［要約統計量］の▼>［要約統計量のカスタマイズ］
分散、歪度、尖度、変動係数にチェックマーク





対数正規分布の例（3）

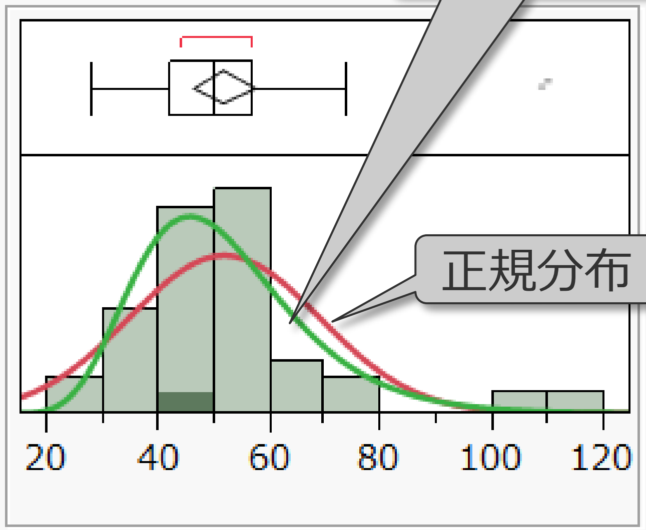
p.83

●対数変換の前と後のデータを比較

- 左右非対称、大きい方に裾を引く分布
- 尖度、歪度はともに 1.5 よも大きい
- 中央値と平均値がずれている

- 左右対称、正規分布の形を表している
- 尖度、歪度ともに 1.5 より 0 に近い
- 尖度は1.5に近いが、更なる対数変換は不可

HDL



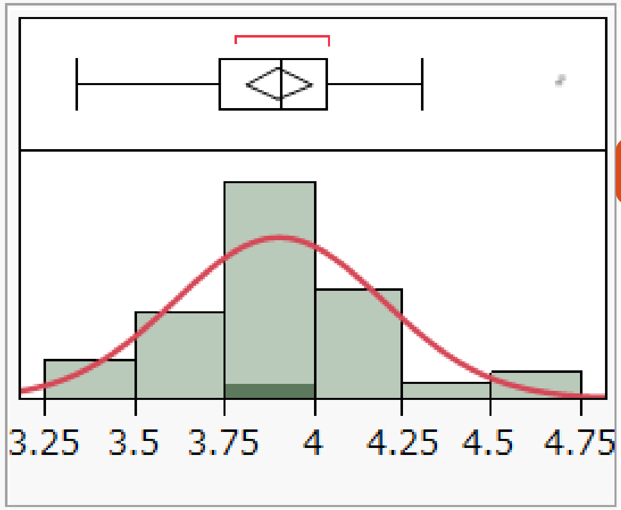
正規(51.825,17.1283)
対数正規(3.90357,0.28873)

要約統計量

平均	51.825
標準偏差	17.128305
平均の標準誤差	2.7082229
平均の上側95%	57.302898
平均の下側95%	46.347102
N	40
分散	293.37885
歪度	1.878448
尖度	4.9929047
変動係数	33.050276

＞1.5

ln(HDL)



正規(3.90357,0.29241)

要約統計量

平均	3.9035688
標準偏差	0.2924102
平均の標準誤差	0.0462341
平均の上側95%	3.9970862
平均の下側95%	3.8100515
N	40
分散	0.0855037
歪度	0.5982563
尖度	1.3926864

＜1.5

変動係数を非表示

対数正規分布の例

p.84

- 研究開発の対象となる評価項目の分布

取り上げる評価項目について、多くの観測値を集めて分布を特定しておくことが望ましい

事例３は $n = 40$、この 程度の観測値では分布を間違いなく特定することは難しい

医薬品開発の場合、健常者のデータ（たとえば、企業の定期検診のデータ）で分布を特定

過去の実験データ（n が数十程度）が複数個あるときは，それらを併合して解析

平均，標準偏差が実験によって異なる場合は，基準化してから併合

関心事項はひずみ、とがり）

基準化（標準化）：観測値から平均値を引いて標準偏差で除す（p.28）

研究開発の評価項目の分布に対応して変数変換を行うなど、適切な統計解析を実施

（2）中心極限定理による対数正規分布の導出

賭けは法律的に問題あり
単なる事例、推奨している意図はない

中心極限定理による対数正規分布の導出

p.84

●変数 x の変化

変数 x の変化は、たくさんの変動原因に起因する誤差（ e_i ）が積み重なった結果

$$x = \mu + e_1 + e_2 + \cdots + e_i + \cdots + e_n \qquad (2.3.1)$$

個々の変動要因に起因する誤差（e_i）の分布がどのようなものであっても、
それらを集めた変数 x の分布は n が大きくなると正規分布に近づく

（中心極限定理、§1.3 ）

複数の変動要因の積み重なりが、和ではなく、積である場合を考える → 事例４

$$x = \mu \times c_1 \times c_2 \times \cdots \times c_i \times \cdots \times c_n \qquad (2.3.2)$$

中心極限定理による対数正規分布の導出

p.84

●事例４

資金（μ ）を元手に賭けを行う

毎回、その時点の所持金の1/10を賭ける

負ければ賭け金を失う（×0.9）

勝てば賭け金の倍が戻る（×1.1）、勝率 1/2

20回賭けた後の残金（x）の分布を考える

$$x = \mu \times c_1 \times c_2 \times \cdots \times c_i \times \cdots \times c_{20}$$

$$c_i = 0.9 \text{ or } 1.1 \qquad (2.3.2)$$

（例）

$$x = \mu \times 1.1 \times 0.9 \times 1.1 \times \cdots \times 1.1$$

$$= \mu \times 1.350$$

全敗は $0.9^{20} = 0.12$、全勝は $1.1^{20} = 6.72$

中心極限定理による対数正規分布の導出

p.84

●事例４

資金（μ）を元手に賭けを行う

毎回、その時点の所持金の1/10を賭ける

負ければ賭け金を失う（×0.9）

勝てば賭け金の倍が戻る（×1.1）、勝率 1/2

20回賭けた後の残金（x）の分布を考える

$$x = \mu \times c_1 \times c_2 \times \cdots \times c_i \times \cdots \times c_{20}$$
$$c_i = 0.9 \text{ or } 1.1 \qquad (2.3.2)$$

（例）

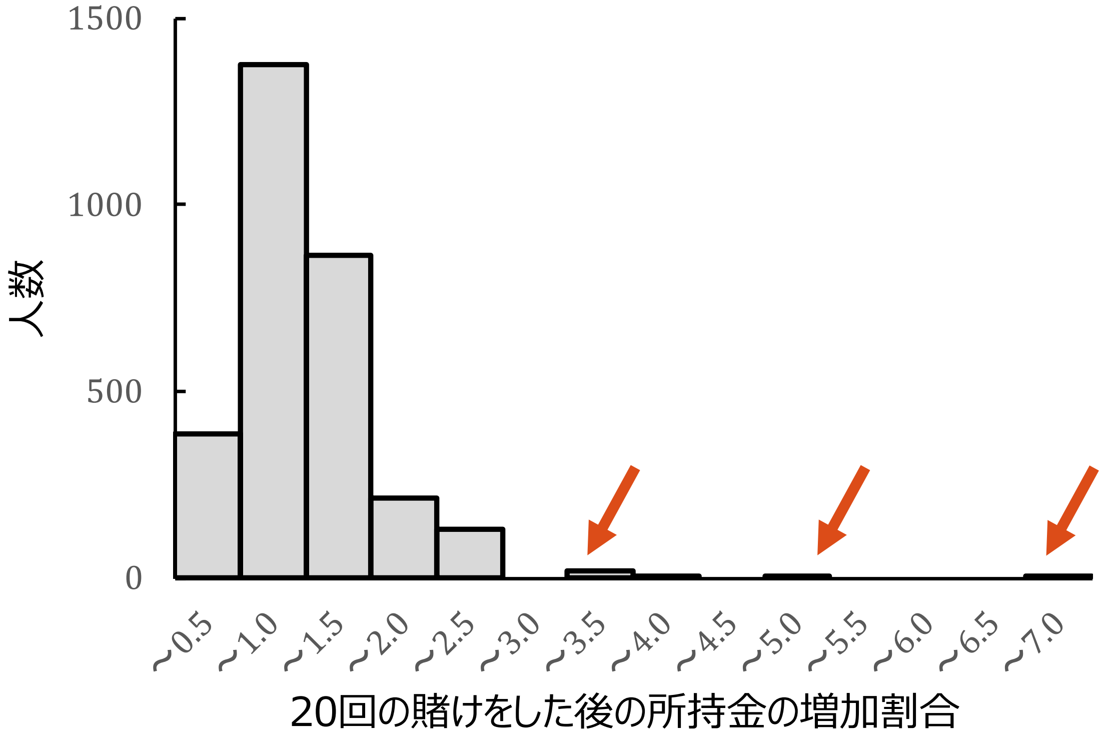
$$x = \mu \times 1.1 \times 0.9 \times 1.1 \times \cdots \times 1.1$$
$$= \mu \times 1.350$$

全敗は $0.9^{20} = 0.12$、全勝は $1.1^{20} = 6.72$

賭けを 3,000 人が行ったシミュレーション

多くは 1.0 付近に集中

右に裾が伸びる分布、3倍、5倍、6倍の人も出現



中心極限定理による対数正規分布の導出

p.84

●変数 x の変化

変数 x の変化：原因 e が和として影響する場合

$$x = \mu + e_1 + e_2 + \cdots + e_n \qquad (2.3.1)$$

中心極限定理により、x の分布は正規分布に近づく

変数 x の変化：原因 c が積として影響する場合

$$x = \mu \times c_1 \times c_2 \times \cdots \times c_n \qquad \text{（賭けの例）} \qquad (2.3.2)$$

$$\ln(x) = \ln(\mu) + \ln(c_1) + \ln(c_2) + \cdots + \ln(c_n) \qquad (2.3.3)$$

$$= \ln(\mu) + \ln(1 + e_1) + \cdots + \ln(1 + e_n) \qquad (2.3.4)$$

$$c_i = 1 + e_i \qquad e_i = c_i - 1$$

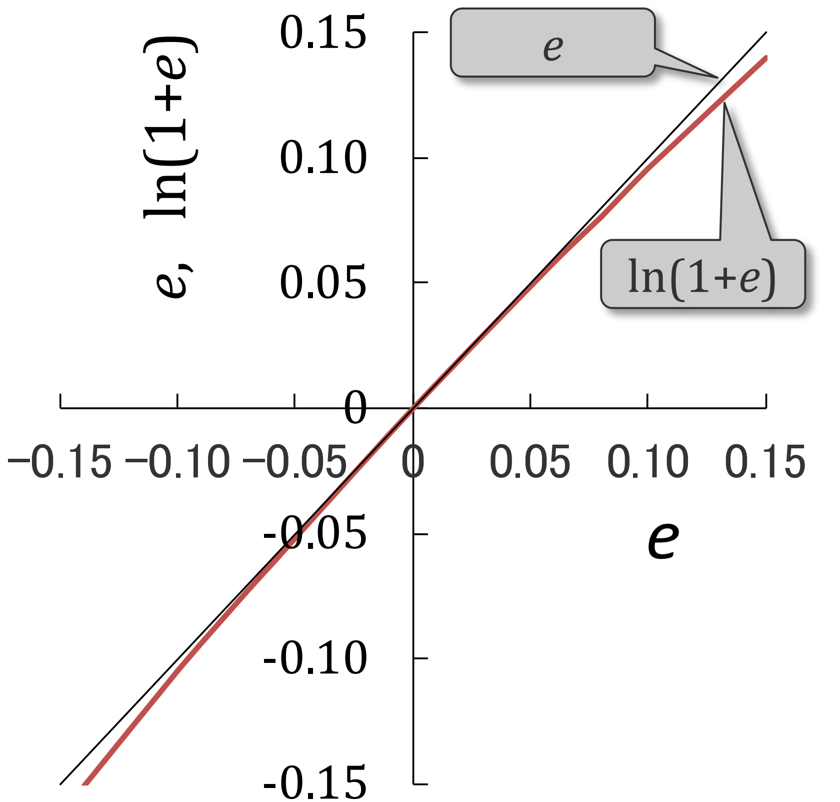
賭けの例では $c_i = 0.9$ or 1.1, $e_i = -0.1$ or 0.1

e_i が 0 に近いとき、$\ln(1 + e_i) \approx e_i$ (2.3.5, 右図)

$$\ln(x) = \ln(\mu) + e_1 + e_2 + \cdots + e_n \qquad (2.3.6)$$

中心極限定理により、ln(x) の分布は正規分布に近づく

e	-0.150	0.090	0.100	0.150
ln(1+e)	-0.163	0.086	0.095	0.140



中心極限定理による対数正規分布の導出

p.84

●変数 x の変化

e	-0.100	-0.050	0.000	0.050	0.100
ln(1+e)	-0.105	-0.051	0.000	0.049	0.095

変数 x の変化：原因 e が和として影響する場合

$$x = \mu + e_1 + e_2 + \cdots + e_n \qquad (2.3.1)$$

中心極限定理により、x の分布は正規分布に近づく

変数 x の変化：原因 c が積として影響する場合

$$x = \mu \times c_1 \times c_2 \times \cdots \times c_n \qquad \text{（賭けの例）} \qquad (2.3.2)$$

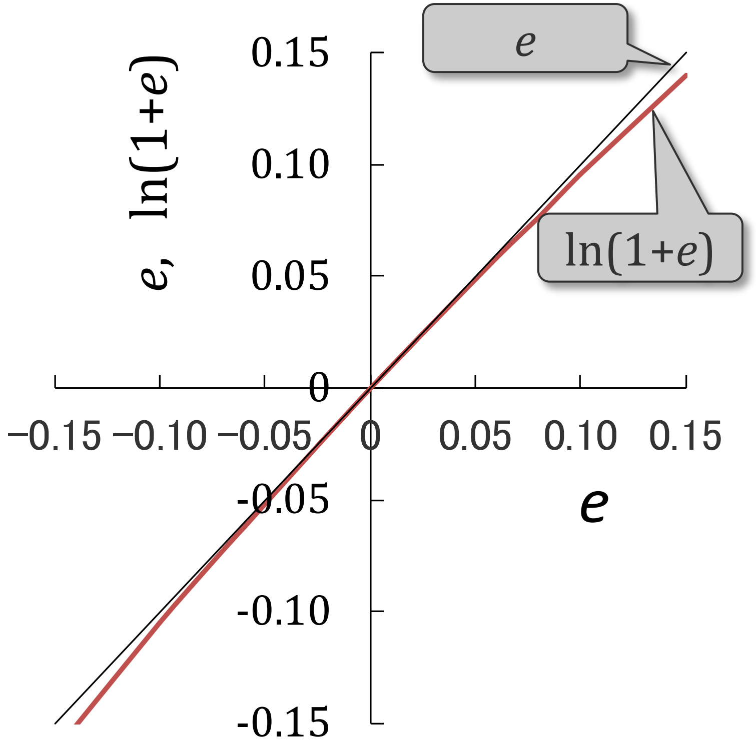
$$\ln(x) = \ln(\mu) + \ln(c_1) + \ln(c_2) + \cdots + \ln(c_n) \qquad (2.3.3)$$

$$= \ln(\mu) + \ln(1+e_1) + \cdots + \ln(1+e_n) \qquad (2.3.4)$$

$$c_i = 1 + e_i \qquad e_i = c_i - 1$$

賭けの例では $c_i = 0.9 \text{ or } 1.1, \quad e_i = \pm 0.1$

e_i が 0 に近いとき、$\ln(1+e_i) \approx e_i$ (2.3.5, 右図)



$$\ln(x) = \ln(\mu) + e_1 + e_2 + \cdots + e_n \qquad (2.3.6)$$

式(2.3.1) と同じ

中心極限定理により、ln(x) の分布は正規分布に近づく（事例 4 ）

中心極限定理による対数正規分布の導出

p.85

●事例５

飲酒の影響が、ALT が 20 の人（普通）と 80 の人（肝機能障害あり）に及ぼす影響

（ALTは肝臓機能の指標）

左図：２人に同程度の影響がある

20 の人は+2 の 22 に上昇

80 の人は+2 の 82 に上昇

中図：２人に同程度の影響がある

20 の人は 1.1 倍の 22 に上昇

(22/20=1.1)

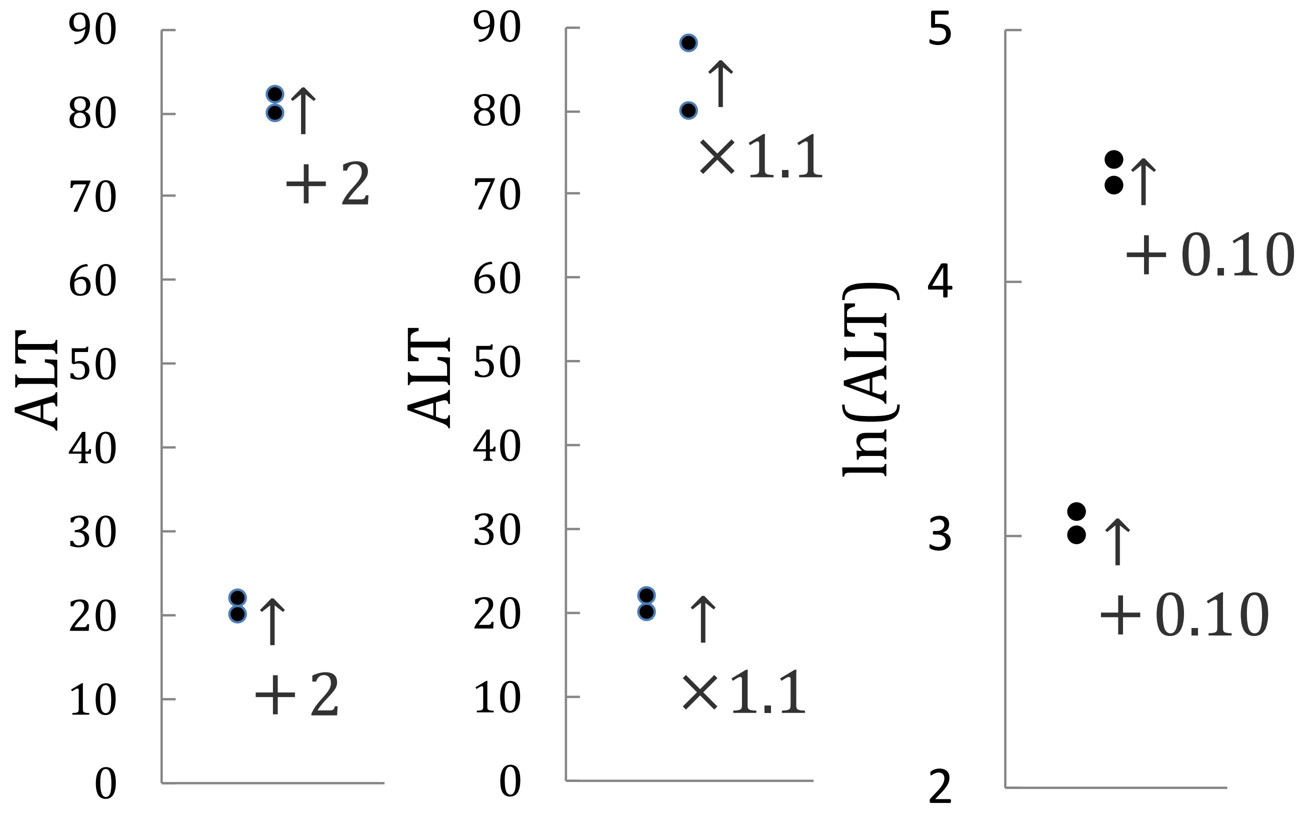
80 の人は 1.1 倍の 88 に上昇

右図（中図の縦軸を対数変換）

ln(20) の人は ln(22) に+0.10 の上昇

ln(80) の人は ln(88) に+0.10 の上昇

影響を対数変換値の和として考える



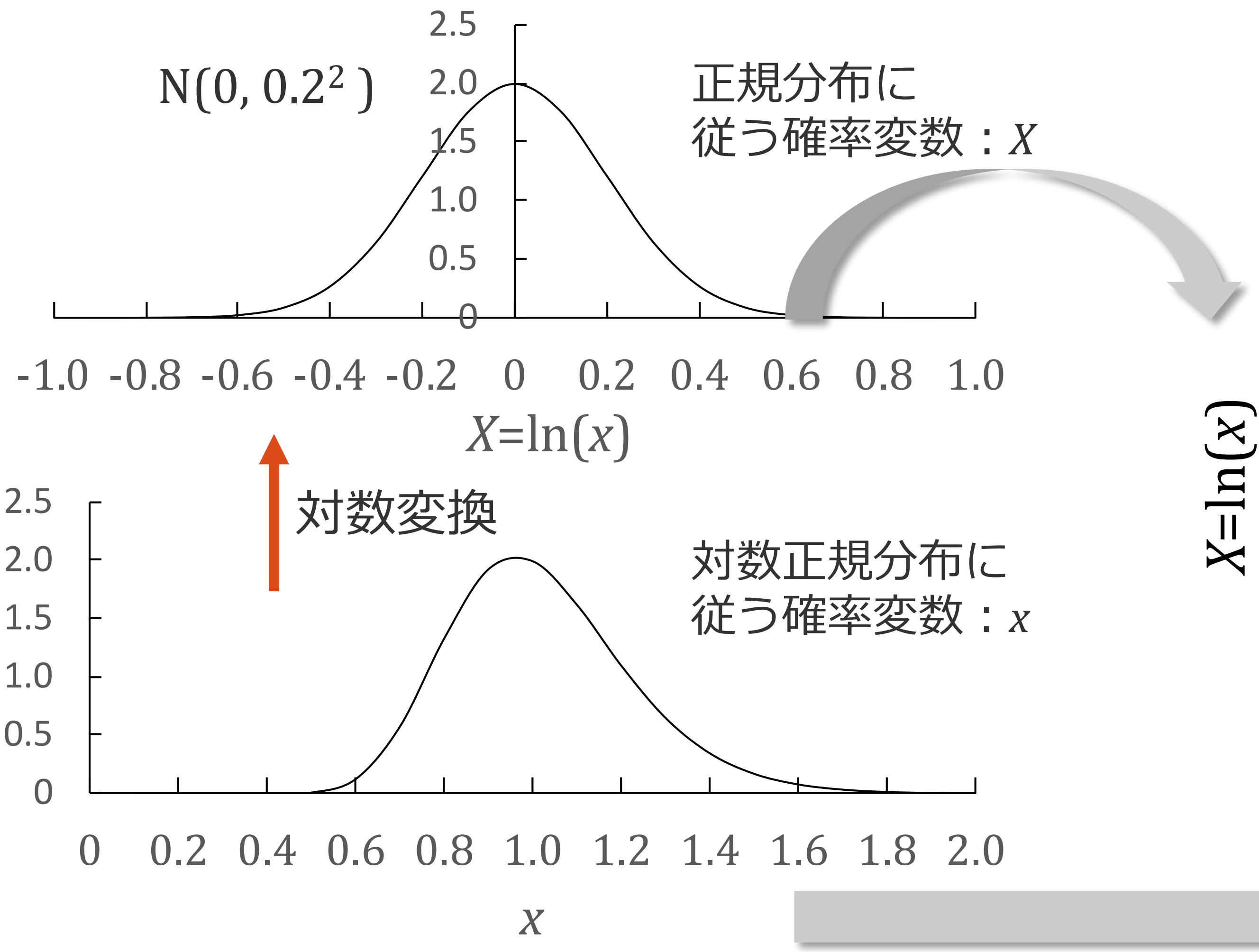
p.85

（3）対数変換値の標準偏差

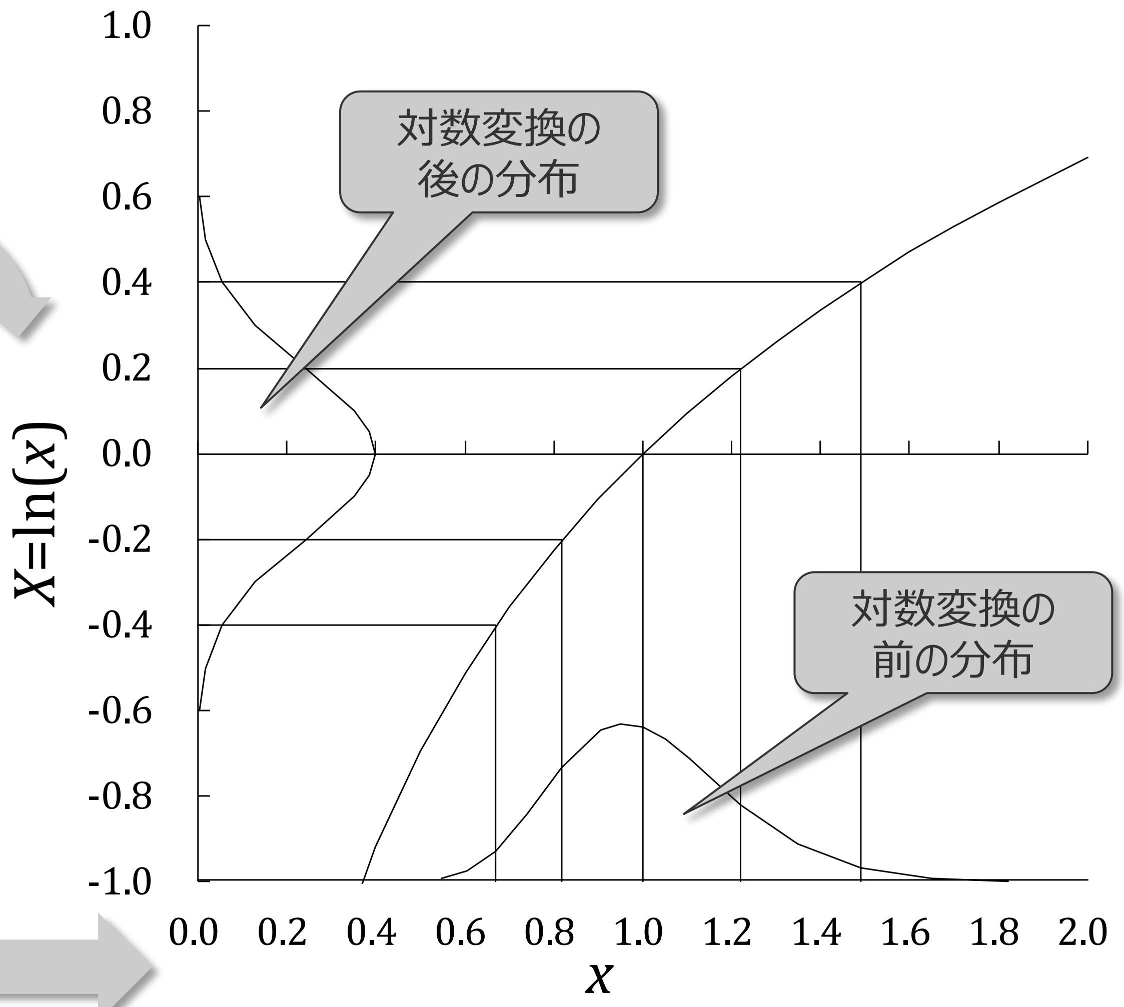
対数変換後の標準偏差と
変換前の変動係数の関係

対数変換値の標準偏差

●対数正規分布と対数変換後の正規分布



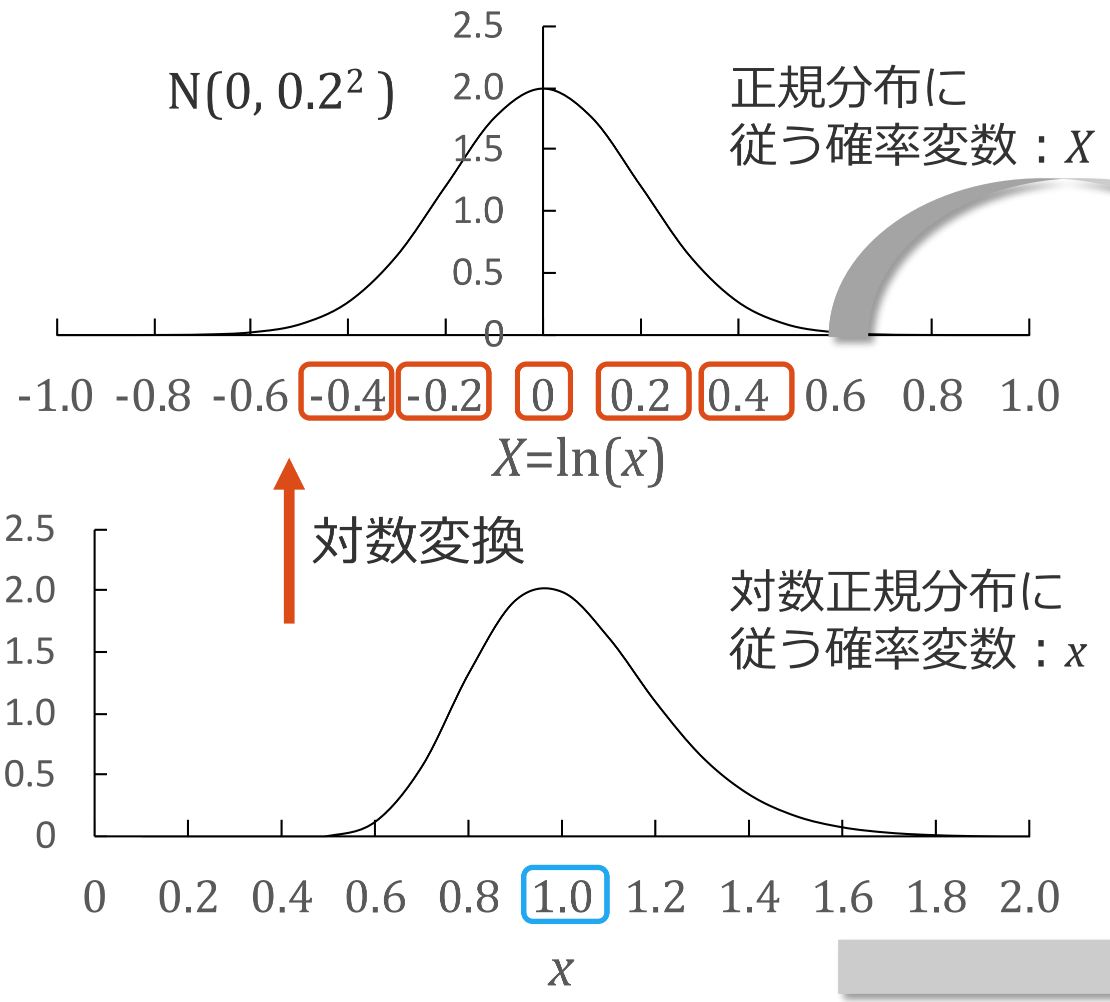
表示 2.3.4 x, $X = \ln(x)$ の関係と x, X の分布



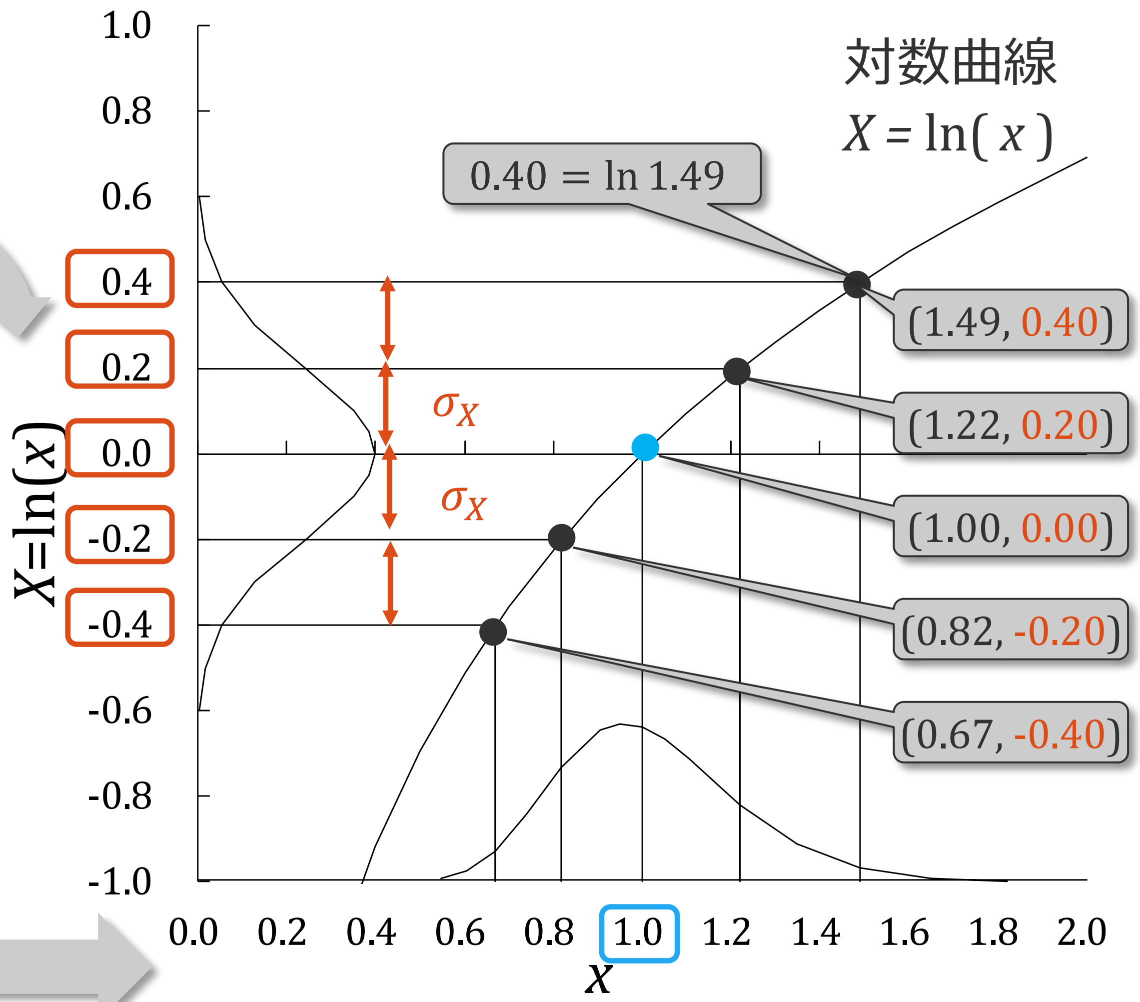
対数変換値の標準偏差

p.85

●対数正規分布と対数変換後の正規分布



表示 2.3.4 x, $X = \ln(x)$ の関係と x, X の分布



対数変換値の標準偏差

p.85

●対数正規分布と対数変換後の正規分布

両者の確率分布

X = ln(x) の分布：正規分布　N(0, 0.2^2)

=NORM.DIST(－0.4, 0, 0.2)　=　0.023

=NORM.DIST(－0.2, 0, 0.2)

－0.023 = 0.136

=NORM.DIST(0, 0, 0.2)

－0.023－0.136 = 0.341

***x* の分布**：対数正規分布 自然対数変換値

=LOGNORM.DIST(0.670, 0, 0.2) = 0.023

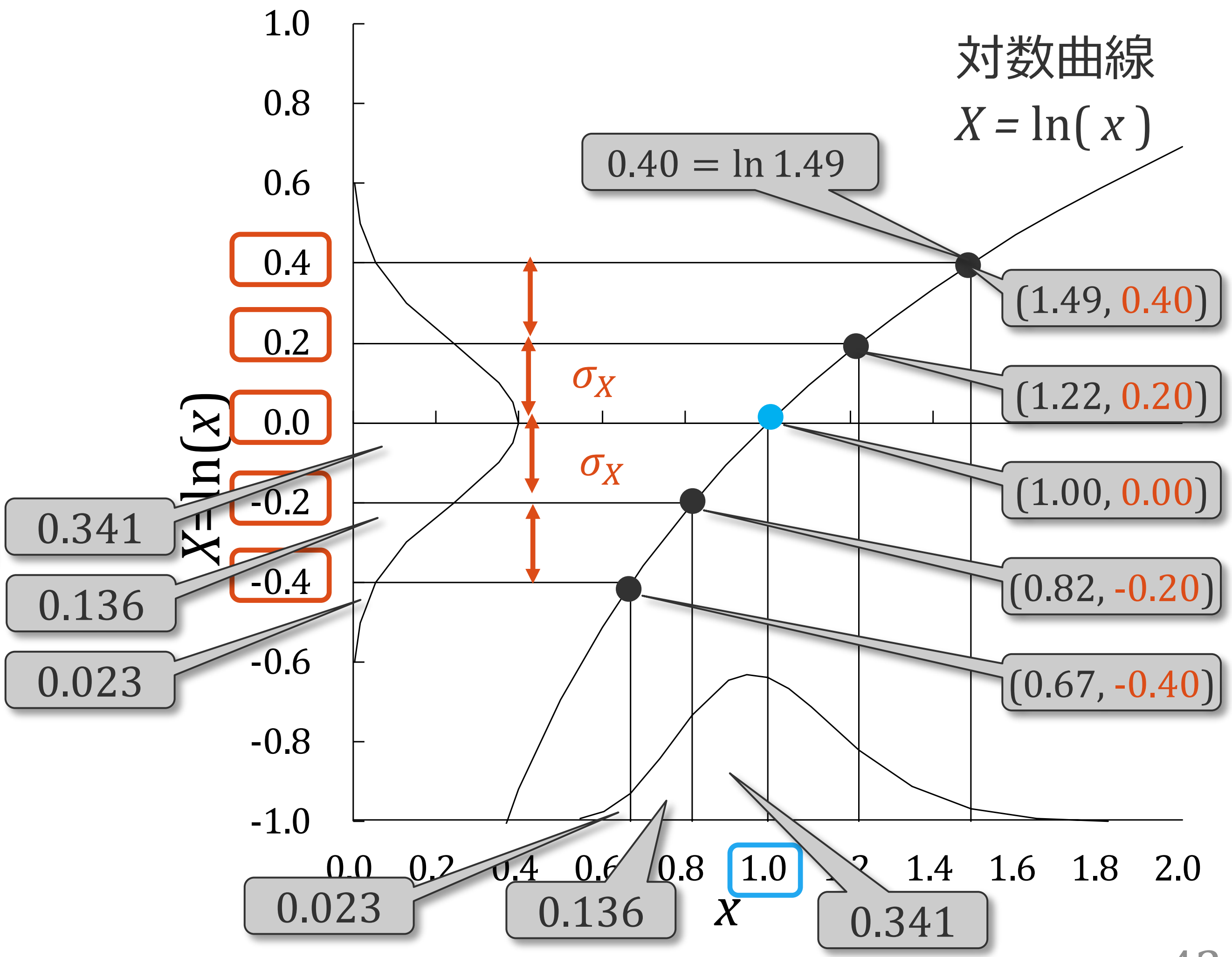
=LOGNORM.DIST(0.819, 0, 0.2)

－ 0.023 = 0.136

=NORM.DIST(1.000, 0, 0.2)

－0.023－0.136 = 0.341

表示 2.3.4 *x*, *X* = ln(*x*) の関係と *x*, *X* の分布



対数変換値の標準偏差

p.85

●対数正規分布と対数変換後の正規分布

両者の対応する部分の確率は同じ（同じ色）

X = ln(x) の分布：正規分布　N(0, 0.2^2)

左右対称

５本の線は等間隔（間隔は σ_X）

***x* の分布**：対数正規分布

５本の線の間隔は右ほど広い

左側の高さは高く、右側の高さは低い

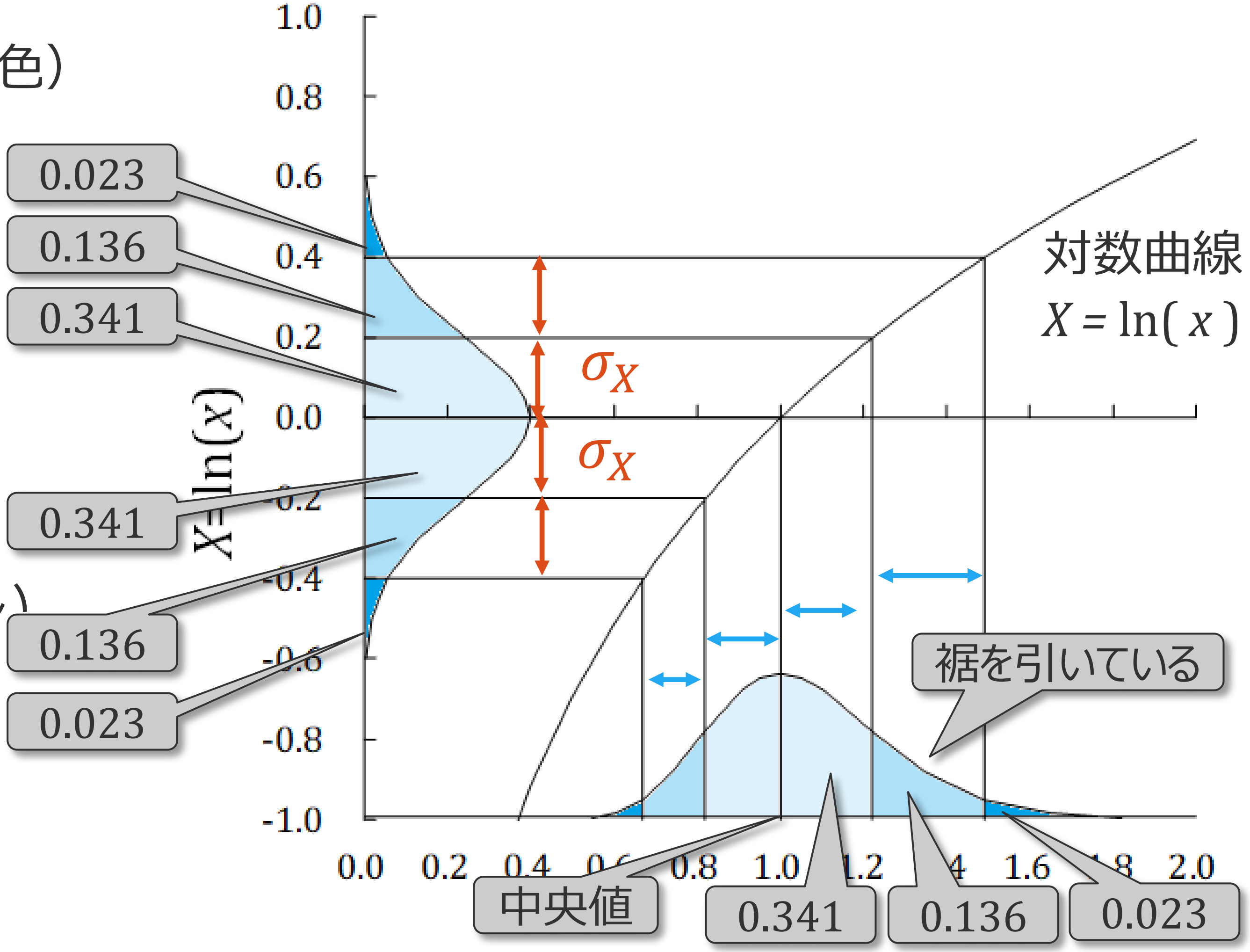
右方向に裾を引いている

左右は非対称であるが同じ面積

中央の $x = 1$ は平均ではなく中央値

（平均値は 1.02）

表示 2.3.4 x, $X = \ln(x)$ の関係と x, X の分布



対数変換値の標準偏差

p.86

●対数正規分布と対数変換後の正規分布

対数曲線に $x=1, X=0$ で接する直線を引く

接線の傾きは $X=\ln(x)$ を x で微分して求める

$$\frac{d}{dx}\ln(x)=\frac{1}{x}$$

$x=1.0, X=0.0$での接線の傾きは 1.0

この接線の傾きは X の σ_X と x の $\sigma_x{}^*$ の比

（$\sigma_x{}^*$ ：対数正規分布の近似標準誤差）

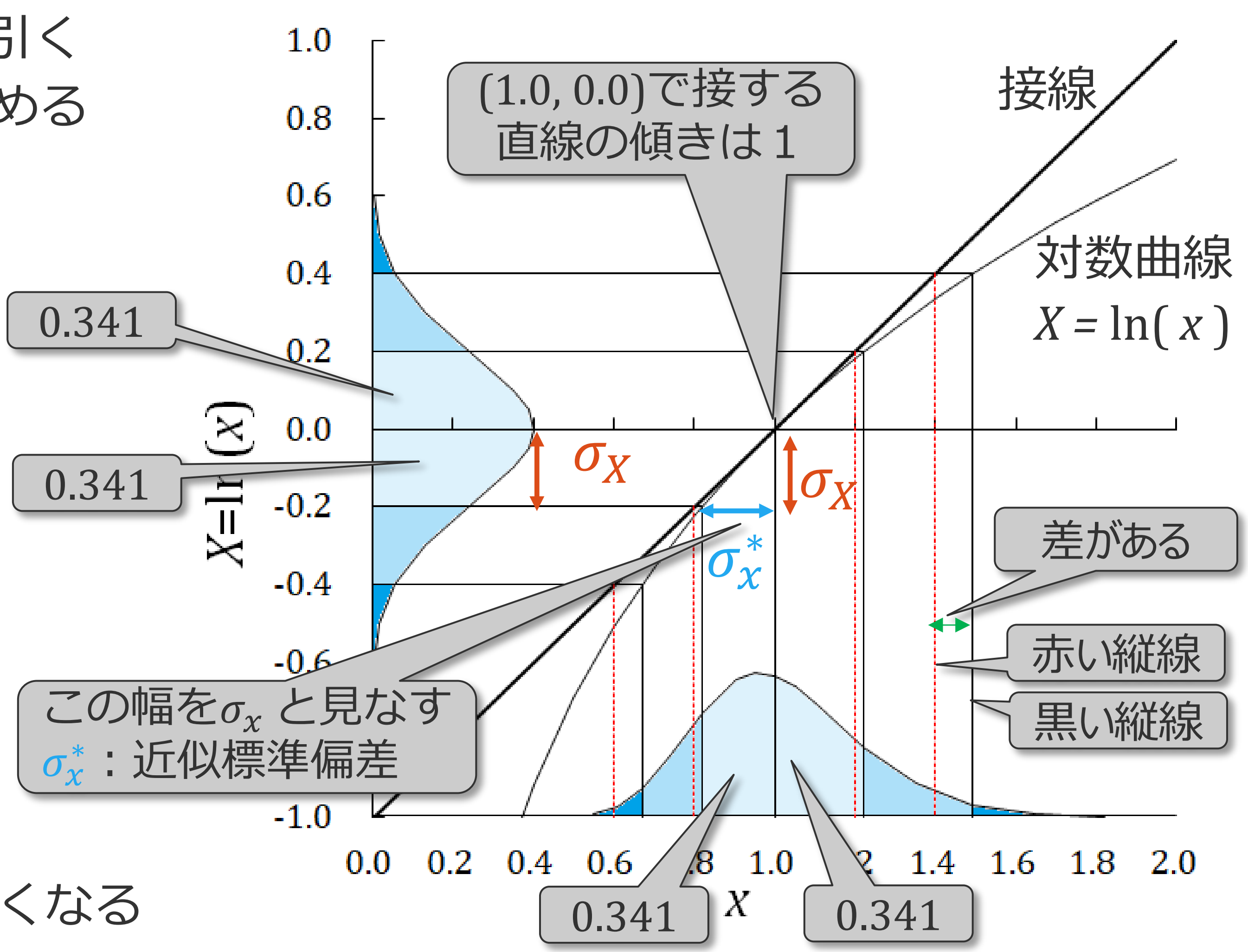
$$\frac{\sigma_X}{\sigma_x^*}=\frac{1}{x}$$

対数曲線からの「黒い縦線」と

接線からの「赤い縦線」には差がある

中心から離れるほど差が広がり近似が悪くなる

表示 2.3.4 x, $X=\ln(x)$ の関係と x, X の分布



対数変換値の標準偏差

p.86

●対数変換後の標準偏差 σ_X と変換前の変動係数 CV

対数変換後の標準偏差を σ_X 、対数変換前の近似標準偏差を σ_x^* とすると

$$\frac{\sigma_X}{\sigma_x^*} = \frac{1}{x}$$

対数変換前の期待値を μ_x、標準偏差を σ_xとすると

$$\frac{\sigma_X}{\sigma_x} \approx \frac{\sigma_X}{\sigma_x^*} = \frac{1}{x} = \frac{1}{\mu_x}$$

$$\sigma_X \approx \frac{\sigma_x}{\mu_x} = \mathrm{CV}[x] \qquad (2.3.7)$$

対数変換した X の標準偏差は、
対数変換前の x の変動係数にほぼ一致

応用事例（第３部 p.67）

事例２：５人の身長を cm と inchで表示

	身長		ln(身長)	
	(cm)	(inch)	(cm)	(inch)
佐藤さん	165	65.0	5.11	4.17
鈴木さん	178	70.1	5.18	4.25
高橋さん	194	76.4	5.27	4.34
田中さん	173	68	5	4.22
中村さん	148	58.3	5.00	4.07
平均	171.6	67.6	5.14	4.21
標準偏差	16.9	6.662	$\boxed{0.100}$	$\boxed{0.100}$
変動係数	$\boxed{0.099}$	$\boxed{0.099}$		

ほぼ一致

対数変換値の標準偏差

●対数変換後の標準偏差 σ_X と変換前の変動係数 CV

Excel ファイル「基礎改2.xls」表示 2.3.4を表示

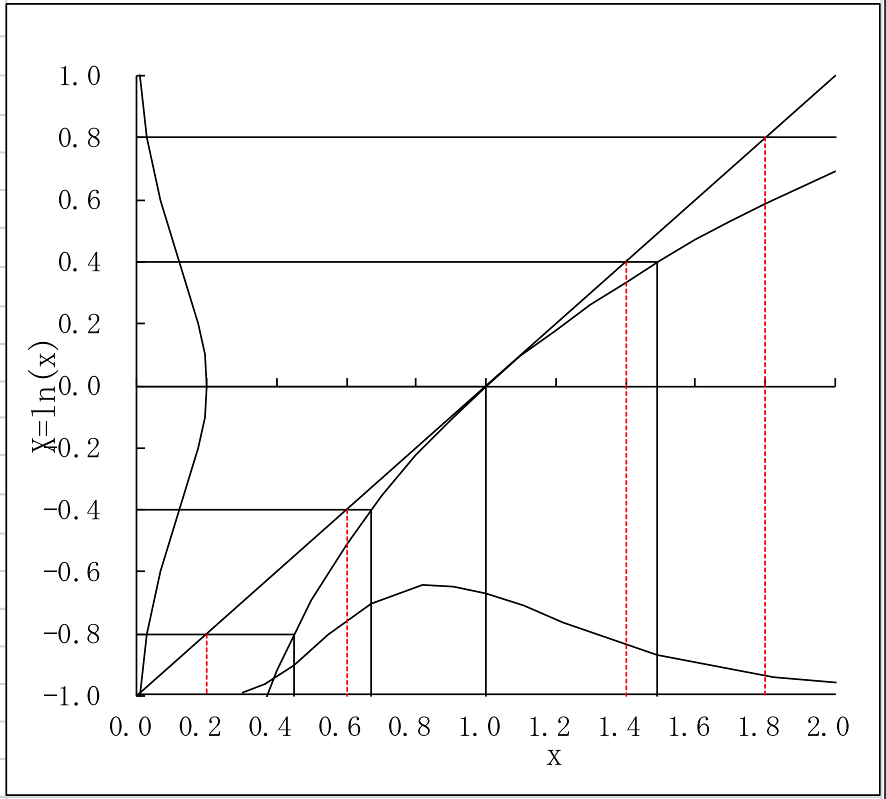
X の標準偏差 σ_X を変化させ（0.1～0.4）、近似の程度を確認する

$$\sigma_X \approx \frac{\sigma_x}{\mu_x} = \mathrm{CV}[x] \qquad (2.3.7)$$

μ x	1		
σ X	0.4		
	−1		
		x	ln(x)
		0.1	−2.30
		0.2	−1.61
		0.3	−1.20
		0.4	−0.92
		0.5	−0.69
		0.6	−0.51
		0.7	−0.36
		0.8	−0.22
		0.9	−0.11
		1.0	0.00
		1.1	0.10
		1.2	0.18
		1.3	0.26
		1.4	0.34
		1.5	0.41
		1.6	0.47
		1.7	0.53
		1.8	0.59
		2.0	0.69

σ_X

表示2.3.4 x, X=ln(x)の関係とx, Xの分布

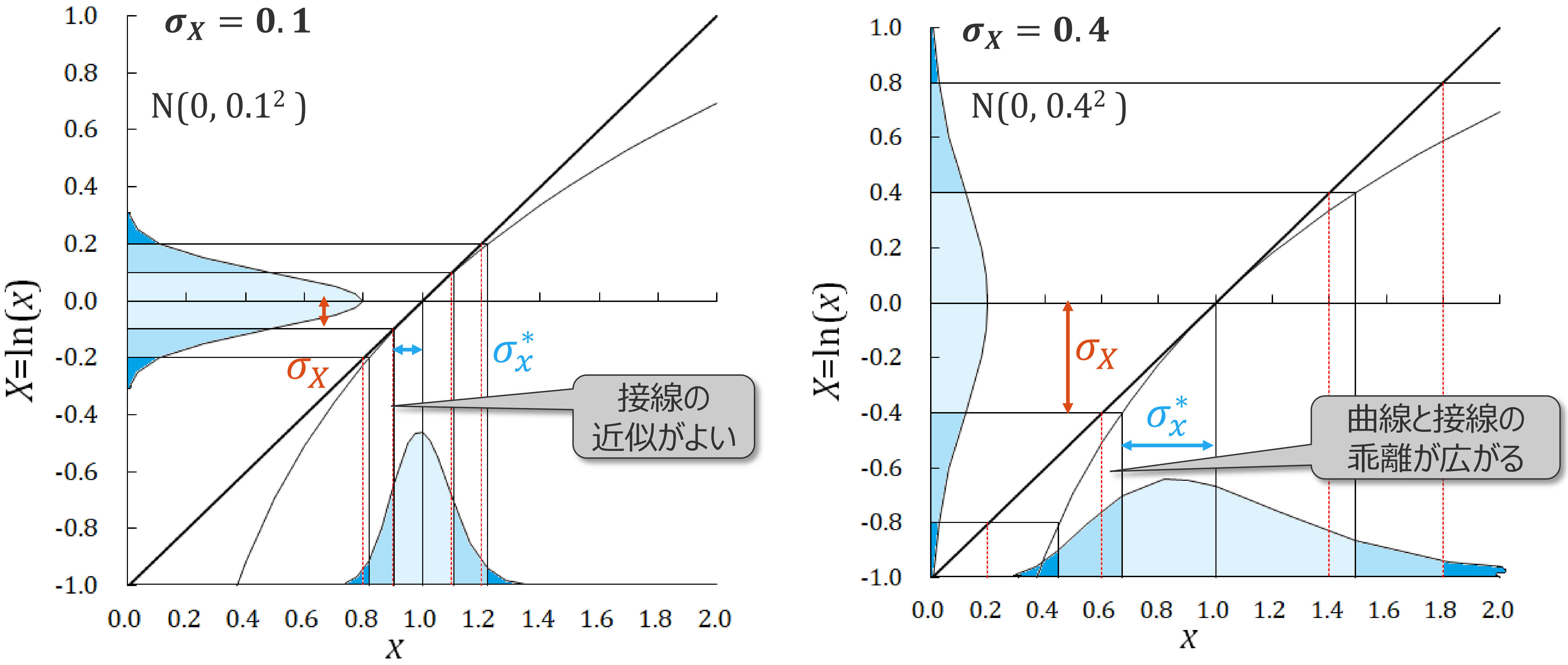


対数変換値の標準偏差

p.86

●対数変換後の標準偏差 σ_X と変換前の変動係数 CV

$$\sigma_X \approx \frac{\sigma_x}{\mu_x} = \mathrm{CV}[x] \quad (2.3.7)$$



対数変換値の標準偏差

p.86

●対数変換後の標準偏差 σ_X と変換前の変動係数 CV

表示 2.3.1

事例 1

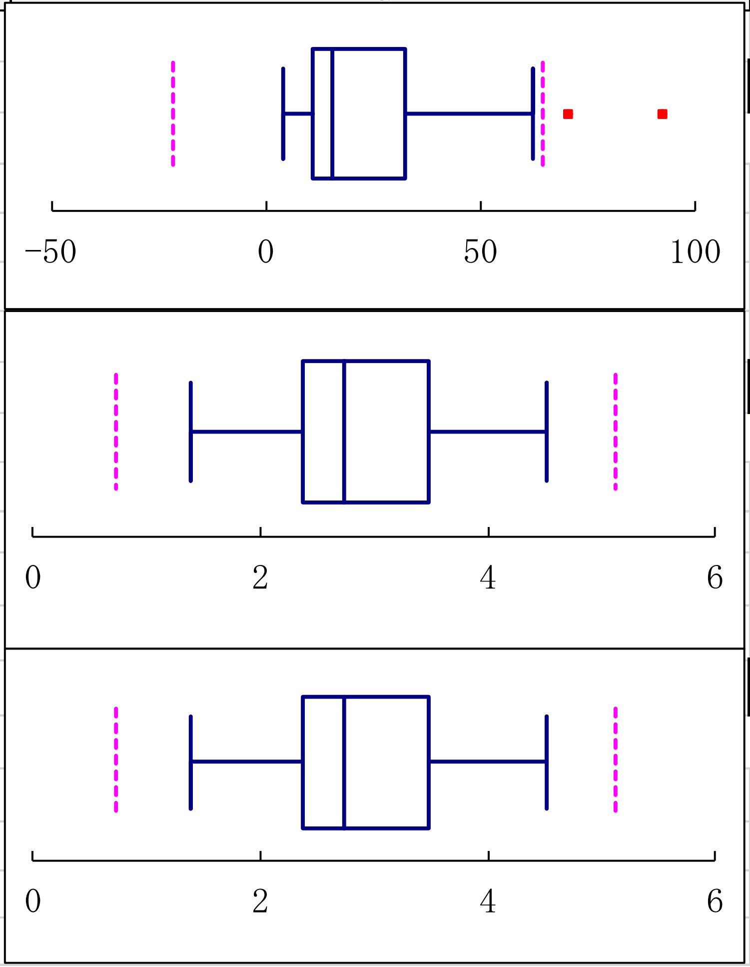
対数変換前の CV が 0.91 と大きいため、対数変換後の標準偏差 0.82 とかなり異なる

$$\sigma_X \approx \frac{\sigma_x}{\mu_x} = \mathrm{CV}[x] \qquad (2.3.7)$$

$$\sigma_X \approx 0.82$$

$$\mathrm{CV}[x] = \frac{\sigma_x}{\mu_x} \approx \frac{23.5}{25.7} = 0.91$$

	A	B	C	D	E	F	G	H	
3		x		ln(x)					
4		14	9	2.64	2.20			K6	
5		11	62	2.40	4.13			B4:C13	
6		4	35	1.39	3.56				
7		10	**92**	2.30	4.52				
8		10	7	2.30	1.95				
9		**70**	28	4.25	3.33				
10		13	12	2.56	2.48			O6	
11		26	33	3.26	3.50			D4:E13	
12		17	32	2.83	3.47				
13		18	11	2.89	2.40				
14	平均		25.7		2.92				
15	標準偏差		23.5		0.82				
16	ひずみ		1.73		0.35			S6	
17	変動係数		0.91					B4:C13	L
18	棄却上限		85.8		5.0				
19									



p.86

（4）その他の変換

対数変換
平方根変換
逆数変換
角変換
Box-Cox 変換

変数変換

p.86

●変数変換の種類

得られたデータのタイプに応じて、データの分布を正規分布に近づける様々な変数変換がある

対数変換　$X = \log(x)$　$X = \ln(x)$

生物的な変量にはこのタイプが多い

データに 0 が含まれる場合、0 を除く最小値の半分の値に置き換える

あるいは、全ての値に有効桁数の半分の値を加算する（山村, 2002）

整数の場合： $\ln(x+0.5)$
小数点第 1 位まである場合： $\ln(x+0.05)$

平方根変換　$X = \sqrt{x}$

生物的な変量にはこのタイプもある

逆数変換　$X = 1/x$

x：抵抗（X：コンダクタンス）

角変換（逆正弦変換）　$X = \sin^{-1}\sqrt{x}$

x：比率データ

変数変換

●変数変換の種類

得られたデータのタイプに応じて、データの分布を正規分布に近づける様々な変数変換がある

対数変換　$X = \log(x)$　$X = \ln(x)$

生物的な変量にはこのタイプが多い

データに 0 が含まれる場合、0 を除く最小値の半分の値に置き換える

あるいは、全ての値に有効桁数の半分の値を加算する（山村, 2002）

平方根変換　$X = \sqrt{x}$

生物的な変量にはこのタイプもある

データのひずみが 1.5 以上の場合（大きい方に裾を引く）、
対数変換を適用
この対数変換値のひずみが負になる場合（小さい方に裾を引く），
対数変換は行き過ぎと考えられるので、対数変換ではなく、
平方根変換を適用すると左右対称の分布に近づく場合がある

逆数変換　$X = 1/x$

x：抵抗（X：コンダクタンス）

角変換（逆正弦変換）　$X = \sin^{-1}\sqrt{x}$

x：比率データ

Box-Cox 変換

p.123

● Box-Cox 変換

べき乗変換（x ：変換前の値、 y ：変換後の値、 p ：変換パラメータ）

$y = x^p$

$p = 2$	$y = x^2$	2 乗変換
$p = 0.5$	$y = x^{0.5} = \sqrt{x}$	平方根変換
$p = -1$	$y = x^{-1} = 1/x$	逆数変換

Box-Cox 変換

べき乗変換を修正して、より一般的なべき乗変換に拡張した方法（Box and Cox, 1964）

$$y = (x^p - 1)/p \quad (p \neq 0) \qquad (2.7.4)$$

$$y = \ln x \quad (p = 0) \qquad (2.7.5)$$

（p：変換パラメータ、λで表すことが多い）

Box-Cox 変換

● Box-Cox 変換

べき乗変換（x ：変換前の値、 y ：変換後の値、 p ：変換パラメータ）

累乗　：指数が自然数
べき乗：指数が実数

$y = x^p$

$p = 2$	$y = x^2$	2 乗変換
$p = 0.5$	$y = x^{0.5} = \sqrt{x}$	平方根変換
$p = -1$	$y = x^{-1} = 1/x$	逆数変換

Box-Cox 変換

べき乗変換を修正して、より一般的なべき乗変換に拡張した方法（Box and Cox, 1964）

$$y = (x^p - 1)/p \quad (p \neq 0) \qquad (2.7.4)$$

$$y = \ln x \quad (p = 0) \qquad (2.7.5)$$

（p：変換パラメータ、λで表すことが多い）

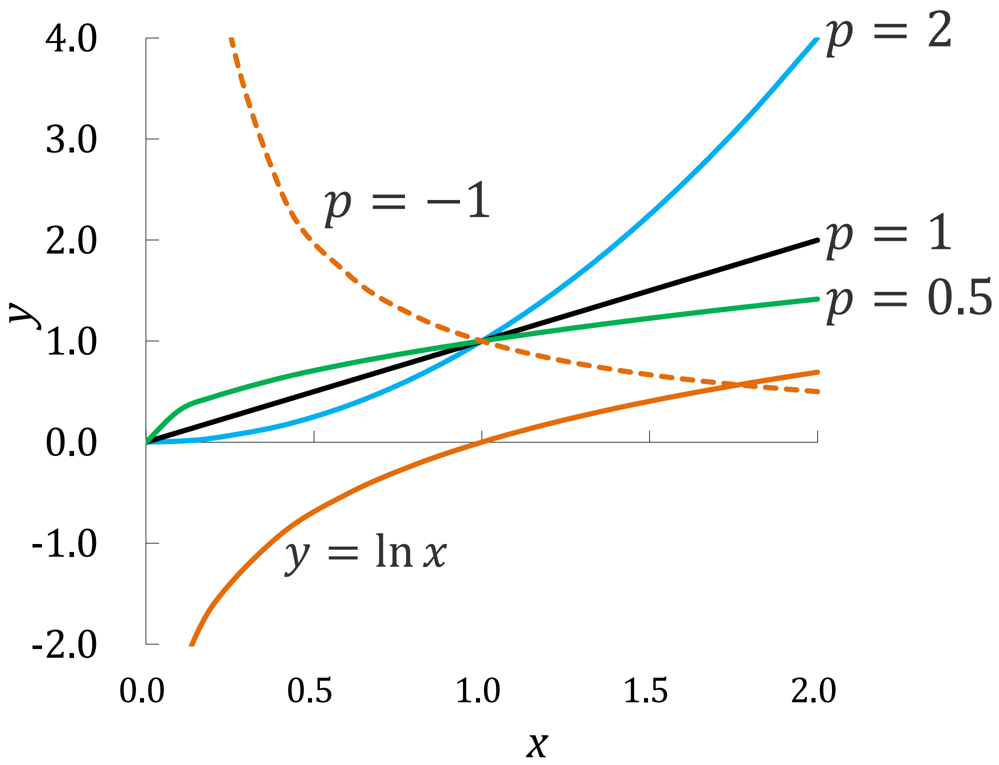
Box-Cox 変換

p.123

●べき乗変換と対数変換

べき乗変換		$y = x^p$
２乗変換	$(p = 2)$	$y = x^2$
無変換	$(p = 1)$	$y = x^1 = x$
平方根変換	$(p = 0.5)$	$y = x^{0.5} = \sqrt{x}$
逆数変換	$(p = -1)$	$y = x^{-1} = 1/x$
対数変換		$y = \ln x$

表示 2.7.5 Box-Cox 変換（左）



$x = 0\sim2.0$ で変化させて変換値を計算（逆変換と対数変換：$x = 0$ のとき無限大になるので空白とする）

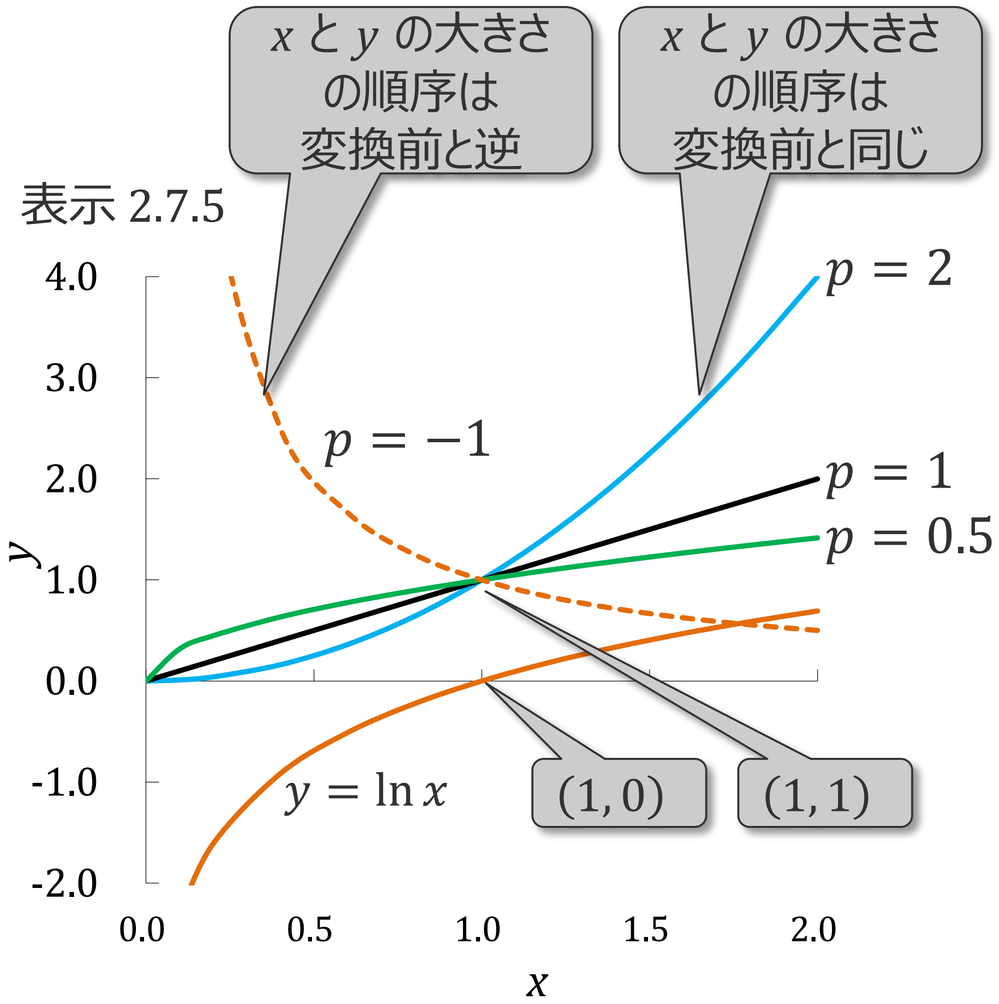
Box-Cox 変換

p.123

●べき乗変換と対数変換

べき乗変換		$y = x^p$
２乗変換	（$p = 2$）	$y = x^2$
無変換	（$p = 1$）	$y = x^1 = x$
平方根変換	（$p = 0.5$）	$y = x^{0.5} = \sqrt{x}$
逆数変換	（$p = -1$）	$y = x^{-1} = 1/x$
対数変換		$y = \ln x$

表示 2.7.5



べき乗変換の曲線を 1 だけ下に移動すると，対数変換の曲線と同様に $x = 1, y = 0$ を通る

Box-Cox 変換

p.123

●Box-Cox 変換の定義

			Box-Cox 変換
べき乗変換		$y = x^p$	$y = (x^p - 1)/p \quad (p \neq 0) \qquad (2.7.4)$ $y = \ln x \quad (p = 0) \qquad (2.7.5)$
2乗変換	$(p = 2)$	$y = x^2$	$y = (x^2 - 1)/2$
無変換	$(p = 1)$	$y = x^1 = x$	$y = (x^1 - 1)/1 = x - 1$
平方根変換	$(p = 0.5)$	$y = x^{0.5} = \sqrt{x}$	$y = (x^{0.5} - 1)/0.5 = 2\sqrt{x} - 2$
対数変換	$\boxed{(p = 0)}$	$y = \ln x$	$y = \ln x$
逆数変換	$(p = -1)$	$y = x^{-1} = 1/x$	$y = (x^{-1} - 1)/(-1) = 1 - 1/x$

修正

対数変換を
$p = 0$ に
位置付け

対数変換を
$p = 0$ に
位置付け

Box-Cox 変換

●Box-Cox 変換の定義

1 を引くことにより
(1, 0) を通る

			Box-Cox 変換		
べき乗変換		$y = x^p$	$y = (x^p - 1)/p$	$(p \neq 0)$	(2.7.4)
		修正	$y = \ln x$	$(p = 0)$	(2.7.5)
２乗変換	$(p = 2)$	$y = x^2$	$y = (x^2 - 1)/2$		
無変換	$(p = 1)$	$y = x^1 = x$	$y = (x^1 - 1)/1 = x - 1$		
平方根変換	$(p = 0.5)$	$y = x^{0.5} = \sqrt{x}$	$y = (x^{0.5} - 1)/0.5 = 2\sqrt{x} - 2$		
対数変換	$(p = 0)$	$y = \ln x$	$y = \ln x$		
逆数変換	$(p = -1)$	$y = x^{-1} = 1/x$	$y = (x^{-1} - 1)/(-1) = 1 - 1/x$		

対数変換を
$p = 0$ に
位置付け

対数変換を
$p = 0$ に
位置付け

p で割ると符合が逆になり
x と y の大小の順序は元のまま

Box-Cox 変換

●べき乗変換とBox-Cox 変換

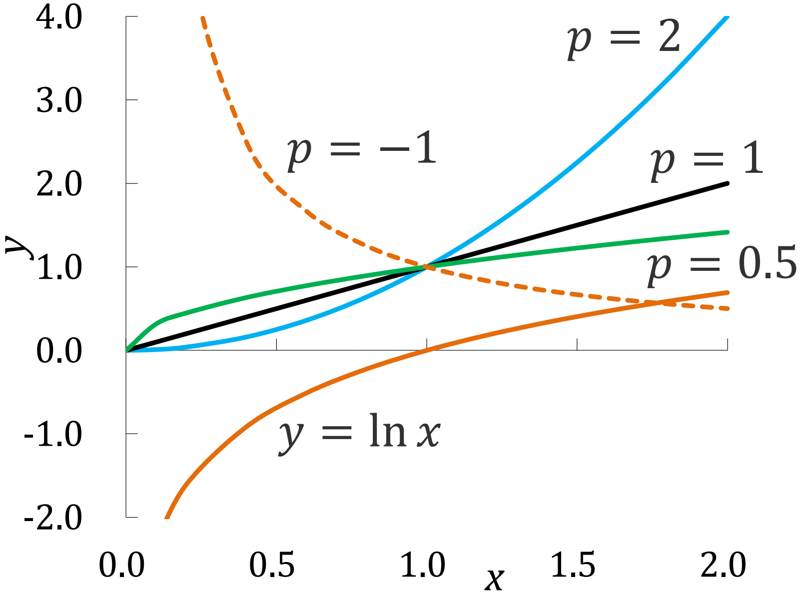
Box-Cox 変換の特徴

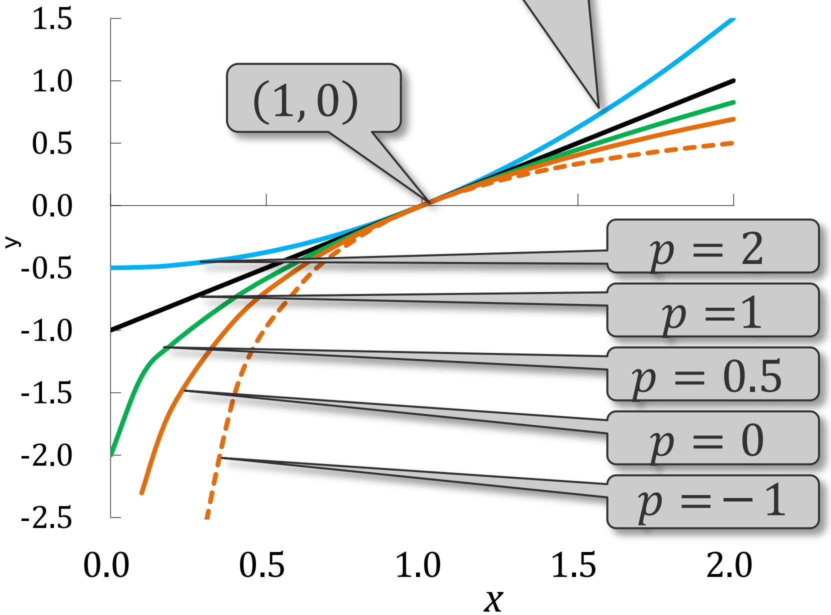
$x = 1, y = 0$ を通る

x と y の大小の順序は
変換前と同じ

傾きはいずれも正

表示 2.7.5 Box-Cox 変換





２乗変換	($p = 2$)	$y = x^2$	$y = (x^2 - 1)/2$
無変換	($p = 1$)	$y = x^1 = x$	$y = (x^1 - 1)/1 = x - 1$
平方根変換	($p = 0.5$)	$y = x^{0.5} = \sqrt{x}$	$y = (x^{0.5} - 1)/0.5 = 2\sqrt{x} - 2$
対数変換	($p = 0$)	$y = \ln x$	$y = \ln x$
逆数変換	($p = -1$)	$y = x^{-1} = 1/x$	$y = (x^{-1} - 1)/-1 = 1 - 1/x$

対数変換を
$p = 0$ に
位置付け

Box-Cox 変換

p.124

●対数変換と $p = 0$ の対応

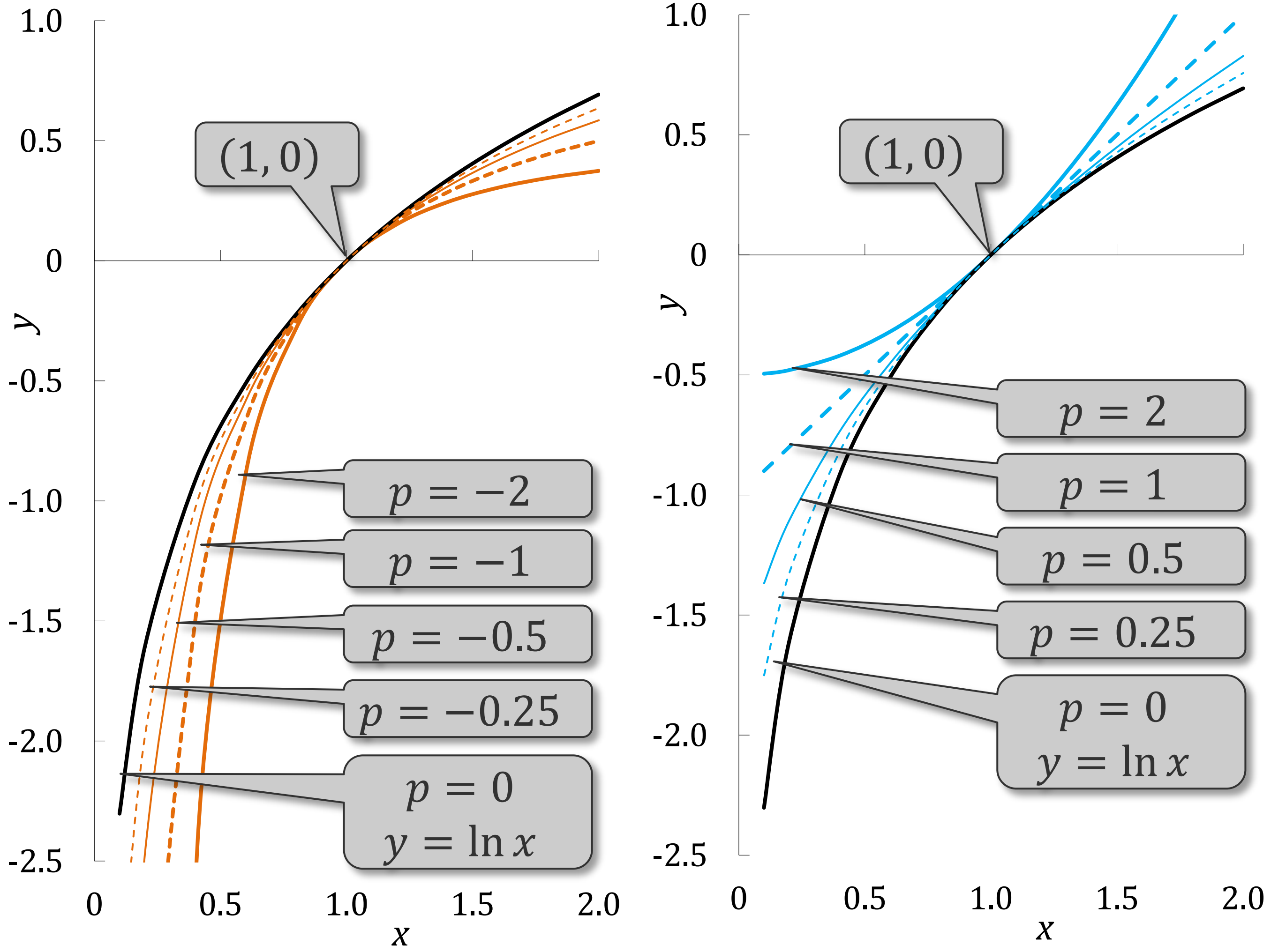
$$y = (x^p - 1)/p \quad (p \neq 0) \qquad (2.7.4)$$
$$y = \ln x \quad (p = 0) \qquad (2.7.5)$$

式(2.7.4) で、
$p = -2,\ -1,\ -0.5, -0.25$ にすると
曲線は $y = \ln x$ の曲線に接近

$p = 2,\ 1,\ 0.5, 0.25$ にすると
曲線は $y = \ln x$ の曲線に接近

対数変換は平方根変換（$p = 0.5$）と逆数変換（$p = -1$）の間に位置し、$p = 0$ に対応する

表示 2.7.5 Box-Cox 変換（右、改変）



Box-Cox 変換

●対数変換と $p = 0$ の対応

x^p をテーラー展開する

$$x^p = e^{p \ln x}$$

$$= 1 + p \ln x + \frac{1}{2}(p \ln x)^2 + \cdots$$

p が 0 に近いとき

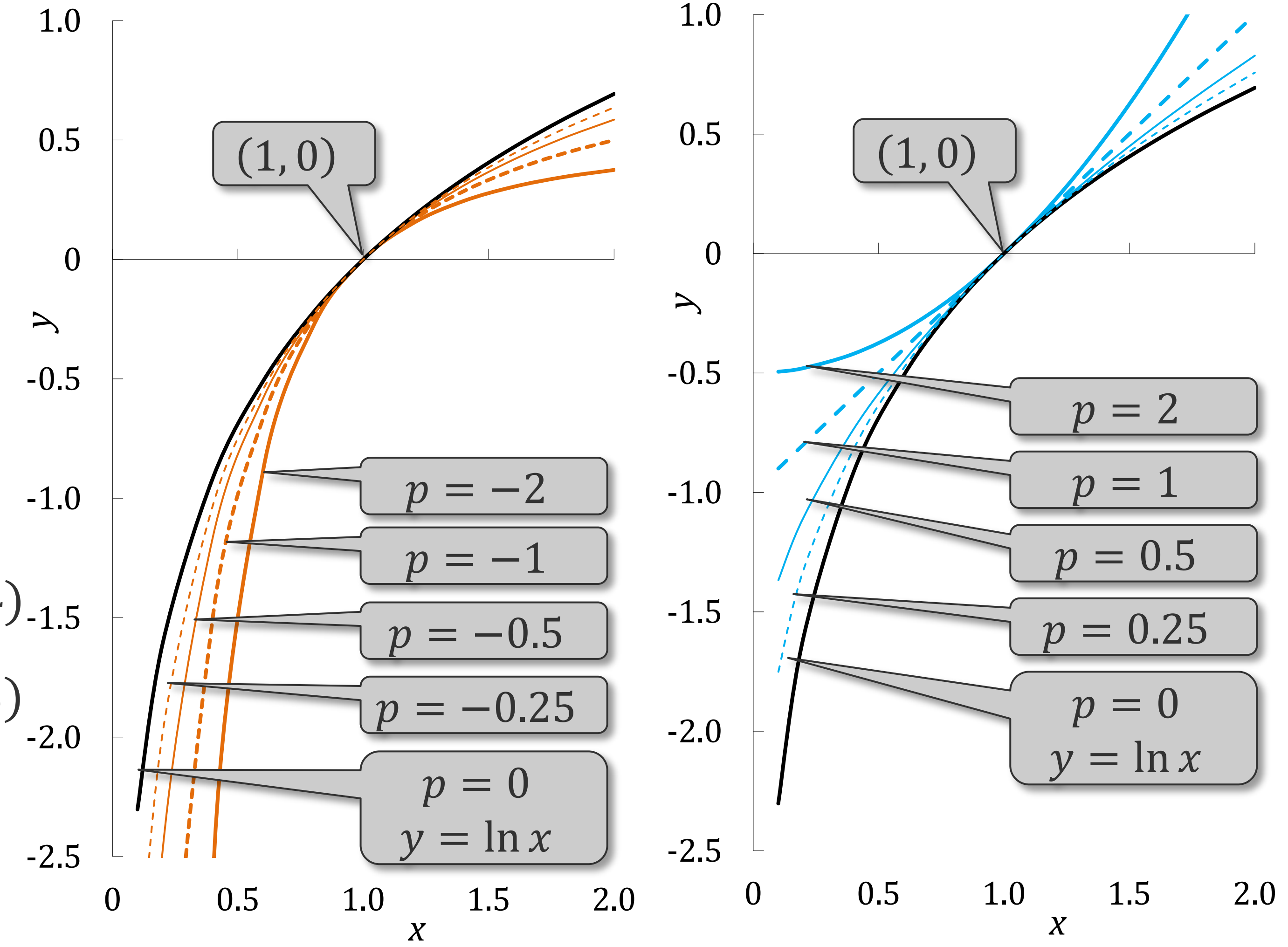
$$x^p \approx 1 + p \ln x$$

$$y = (x^p - 1)/p \qquad (2.7.4)$$

$$\approx (1 + p \ln x - 1)/p = \ln x \qquad (2.7.5)$$

対数変換は平方根変換（$p = 0.5$）と逆数変換（$p = -1$）の間に位置し、$p = 0$ に対応する

表示 2.7.5 Box-Cox 変換（右、改変）



Box-Cox 変換

●変換パラメータ p の探索

$$y = (x^p - 1)/p \quad (p \neq 0) \qquad (2.7.4)$$
$$y = \ln x \quad (p = 0) \qquad (2.7.5)$$

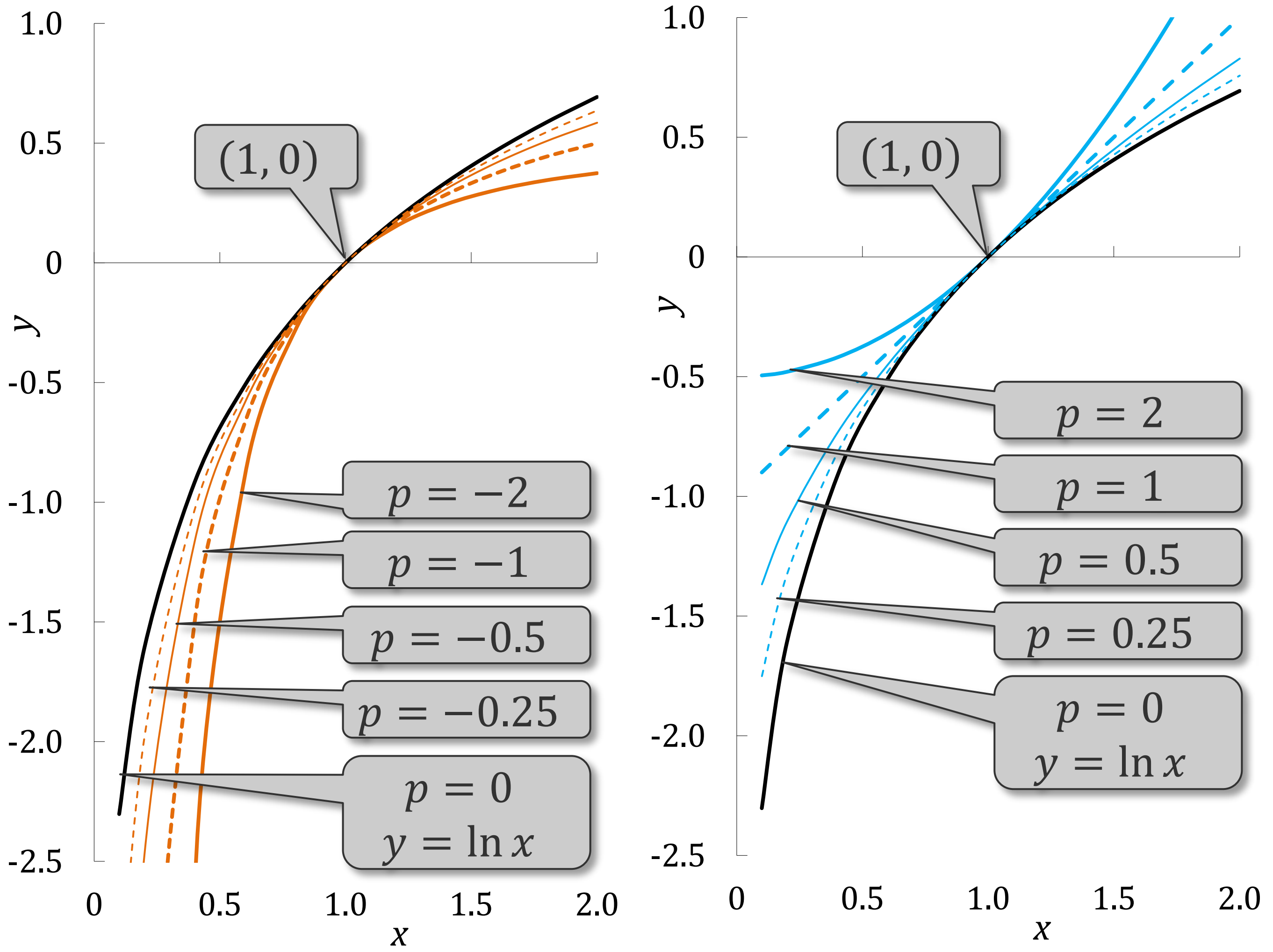
変換値を正規分布に近づけるように
最適な p を最尤推定法で求める
-2~2 の間で 0.2 間隔程度で探索
統計ソフト（JMP、R など）が必須

変換値が常に正規分布に従うわけでは
ないので、変換結果を慎重に評価する

たとえば $p = 0.43$ が得られた場合
その固有技術的な意味づけは難しい

機械的に適用することは好ましくない

表示 2.7.5 Box-Cox 変換（右、改変）



まとめ

p.87

●データの正規性

統計解析の手法の多くは，データが正規分布をするということが条件
しかし，現実のデータは正規分布をするとは限らない
このようなとき，変数変換を検討する必要がある.

●変数変換

対数変換が代表的な変数変換
種々の変数変換の方法があるが，先ず，対数変換をマスター

- 引用

  山村光司（2004）正しい分散分析結果を導くための変数変換法、植物防疫、56：436-441.

- 作成　片瀬雅彦
- 監修　松本一彦、長谷文雄
- 作成時期　2018年10月20日
- 改訂　2019年4月8日、2024年9月19日