4 ロジスティック回帰分析

4.1 復習

テキスト

芳賀敏郎（2016）医薬品開発のための統計解析
第３部　非線形モデル 改訂版、サイエンティスト社、p.288

第３部　非線形モデル

1. 非線形最小２乗法（基礎）

    1.1 線形と非線形、1.2 非線形最小２乗法の基本的な考え方、1.3 指数曲線のあてはめ、

    1.4 Emaxモデルとロジスティック曲線

2. 非線形最小２乗法（応用）

    2.1 誤差を考慮した解析、2.2 効力比、2.3 併用効果（相乗・拮抗効果）、

    2.4 モデルの探索（複数の曲線の同時あてはめ）、2.5 薬物動態の解析

3. 計数値の解析

    3.1 ２項分布、3.2 割合の推定・検定と区間推定、3.3 割合の差の推定・検定と区間推定、

    3.4 多項分布（名義尺度）、3.5 多項分布（順序尺度）、3.6 要因が複数の場合

4. ロジスティック回帰分析

    4.1 復習、4.2 ロジスティック回帰分析（基本）、4.3 ロジスティック回帰分析（応用）

4.1 復習

p.235

テキストの
該当ページ

（1）JMP［二変量の関係］の出力
（2）ロジスティック曲線
（3）誤差の構造
（4）最も簡単なロジスティック回帰分析

使用するファイル

Excelファイル「改4Log回帰.xlsx」
JMPファイル「41-二変量.jmp」「33-割合の差.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

★PowerPoint の
スピーカーノートを、
PDF の注釈に変換してあります
ダウンロードしてご覧ください

はじめに

p.235

- ロジスティック曲線、ロジスティック回帰分析に関係する部分を復習
  - ロジスティック曲線のあてはめ
    - §1.4 Emax モデルとロジスティック曲線
    - §2.1 誤差を考慮した解析
    - §2.2 効力比
  - オッズ、オッズ比、対数オッズ（ロジット）、対数オッズ比
    - §3.3 割合の差の推定・検定と区間推定
    - §3.5 多項分布（順序尺度）
- ロジスティック回帰分析の入門
  - 誤差の構造
  - 最も簡単なロジスティック回帰分析

はじめに

p.235

- ロジスティック曲線、ロジスティック回帰分析に関係する部分を復習

  ロジスティック曲線のあてはめ

  「ロジスティック回帰分析」は、
  「ロジスティック曲線のあてはめ」の拡張ではない
  別の解析手法

  §1.4 Emax モデルとロジスティック曲線
  §2.1 誤差を考慮した解析
  §2.2 効力比

  オッズ、オッズ比、対数オッズ（ロジット）、対数オッズ比
  §3.3 割合の差の推定・検定と区間推定
  §3.5 多項分布（順序尺度）

- ロジスティック回帰分析の入門

  誤差の構造
  最も簡単なロジスティック回帰分析

p.235

（1） JMP［二変量の関係］の出力

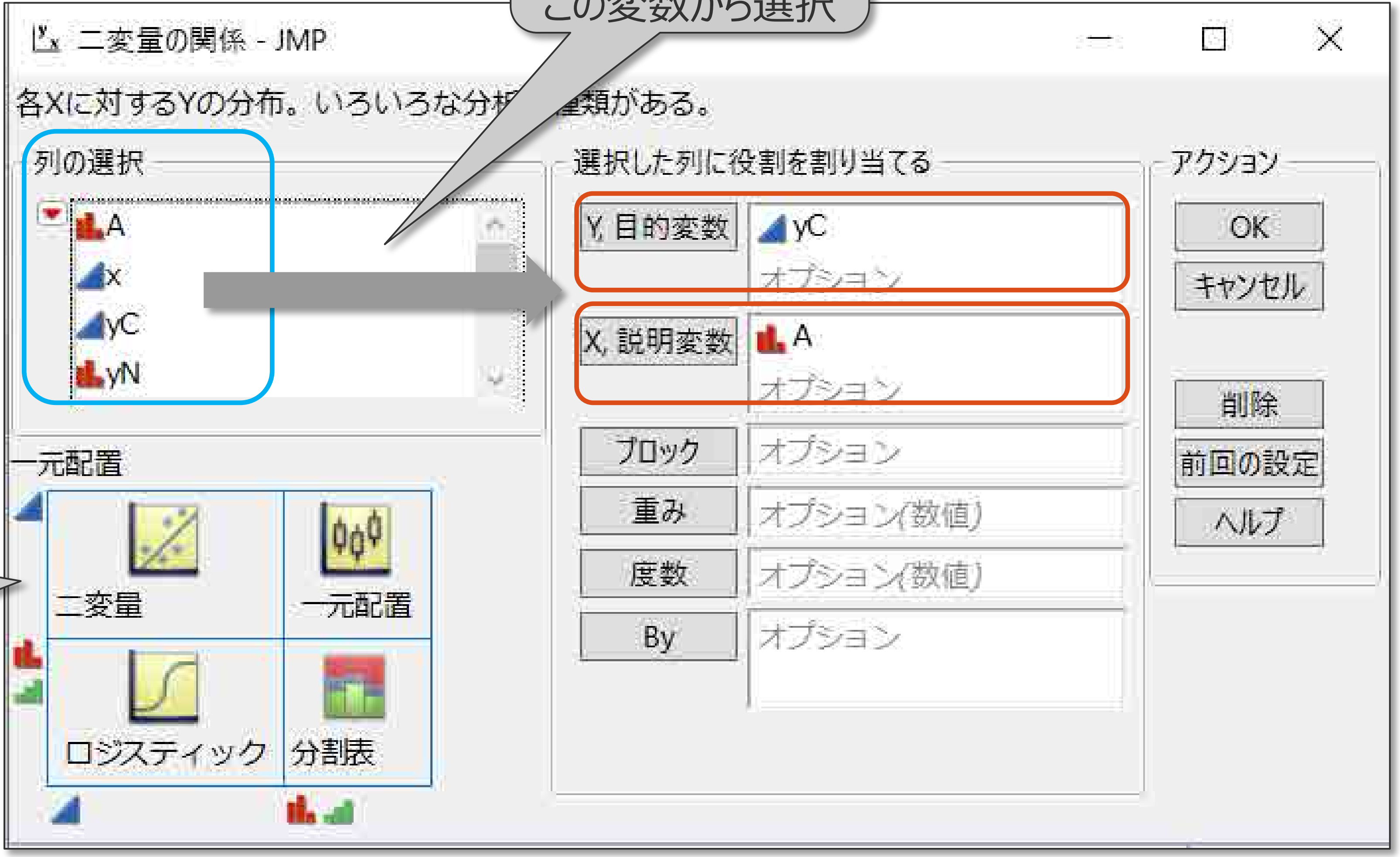
目的変数と説明変数の
４種類の組合せで異なる解析手法を選択

JMP［二変量の関係］の出力

- ［二変量の関係］のダイアログ

JMPファイル「41-二変量.jmp」
の読み込み
［二変量の関係］から
変数設定のダイアログを表示

Y, 目的変数
X, 説明変数
この変数から選択

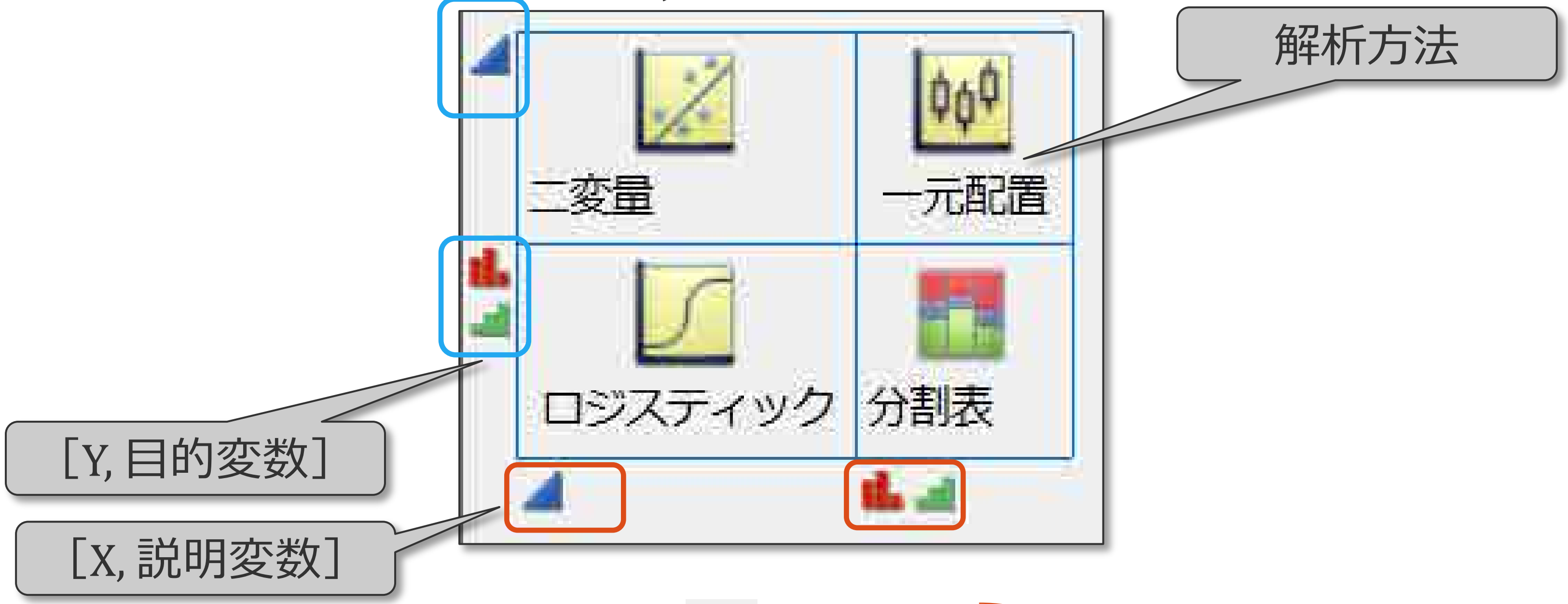


［Y, 目的変数］と
［X, 説明変数］に指定した
変数の尺度の組合せで、
解析方法が自動的に選択

JMP［二変量の関係］の出力

- ［二変量の関係］のダイアログ

表示 4.1.1 JMP の「二変量の関係」



名義尺度
順序尺度
連続尺度

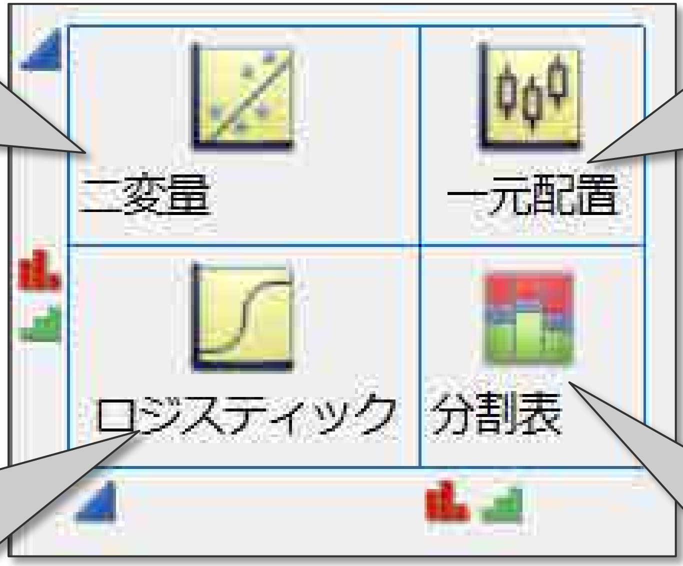
[二変量の関係]では
名義尺度と順序尺度との区別をしない
組合せ数は 2 × 2 = 4

JMP［二変量の関係］の出力

p.235

● ［二変量の関係］のダイアログ

表示 4.1.1 JMP の「二変量の関係」



［Y］：量的変数（連続尺度）
［X］：量的変数（連続尺度）

第１部　§4「回帰分析」
第２部　§2「１因子実験
（量的因子）」

［Y］：量的変数（連続尺度）
［X］：質的変数（名義尺度、
順序尺度）

第２部　§1
「１因子実験（質的因子）」

［Y］：質的変数（名義尺度、
順序尺度）
［X］：量的変数（連続尺度）

第３部　本章

［Y］：質的変数（名義尺度、
順序尺度）
［X］：質的変数（名義尺度、
順序尺度）

第３部　§3「分割表」

名義尺度
順序尺度
連続尺度

JMP［二変量の関係］の出力

p.236

●事例

目的変数２種類、説明変数２種類、計４種類の組合せ、繰り返し数10 の仮想データ

yC：Continuous
yN：Nominal

表示 4.1.2 ４つの組合せのデータ

A	x	1	2	3	4	5	6	7	8	9	10
A1	0	8	3	11	16	15	18	0	8	15	4
A2	1	23	11	29	15	16	14	22	22	[illegible]	22
A3	2	28	28	23	29	29	27	39	34	[illegible]	26
A4	3	48	32	42	35	30	39	37	43	38	43

yC

［目的変数］：yC, 量的変数
（連続尺度、 e.g. 体重、血糖値）

A	x	1	2	3	4	5	6	7	8	9	10
A1	0	0	0	0	0	0	1	0	0	0	0
A2	1	1	0	1	0	0	0	0	0	[illegible]	1
A3	2	1	1	0	1	1	0	1	1	1	0
A4	3	1	1	1	1	0	1	1	1	1	1

yN

［目的変数］：yN, 質的変数
（名義尺度、 e.g. 発疹の有無）

［説明変数］：x, 量的変数（連続尺度、e.g. 薬剤の投与量）

［説明変数］：A, 質的変数（名義尺度、 e.g. 薬剤の種類）

JMP［二変量の関係］の出力

p.236

●事例

［Y, 目的変数］

yC,量的変数

yN,質的変数

［X, 説明変数］

x, 量的変数： 二変量

A, 質的変数：一元配置

x, 量的変数：ロジスティック

A, 質的変数：分割表

表示 4.1.2 4 つの組合せのデータ

A	x	1	2	3	4	5	6	7	8	9	10
A1	0	8	3	11	16	15	18	0	8	15	4
A2	1	23	11	29	15	16	14	22	22	[illegible]	22
A3	2	28	28	23	29	29	27	39	34	[illegible]	26
A4	3	48	32	42	35	30	39	37	43	38	43

yC

［目的変数］：yC, 量的変数
（連続尺度、 e.g. 体重、 血糖値）

A	x	1	2	3	4	5	6	7	8	9	10
A1	0	0	0	0	0	0	1	0	0	0	0
A2	1	1	0	1	0	0	0	0	0	[illegible]	1
A3	2	1	1	0	1	1	0	1	1	1	0
A4	3	1	1	1	1	0	1	1	1	1	1

yN

［目的変数］：yN, 質的変数
（名義尺度、 e.g. 発疹の有無）

［説明変数］：x, 量的変数（連続尺度、 e.g. 薬剤の投与量）

［説明変数］：A, 質的変数（名義尺度、 e.g. 薬剤の種類）

JMP［二変量の関係］の出力

p.236

- ［二変量の関係］の実行

  JMPファイル「41-二変量.jmp」の読み込み

表示 4.1.2 4つの組合せのデータ

A	x	1	2	3	4	5	6	7	8	9	10
A1	0	8	3	11	16	15	18	0	8	15	4
A2	1	23	11	29	15	16	14	22	22	[illegible]	22
A3	2	28	28	23	29	29	27	39	34	[illegible]	26
A4	3	48	32	42	35	30	39	37	43	38	43

yC

A	x	1	2	3	4	5	6	7	8	9	10
A1	0	0	0	0	0	0	1	0	0	0	0
A2	1	1	0	1	0	0	0	0	0	[illegible]	1
A3	2	1	1	0	1	1	0	1	1	1	0
A4	3	1	1	1	1	0	1	1	1	1	1

yN

X: 説明変数

Y: 目的変数

	A	x	yC	yN	AA	xx	y	f
1	A1	0	8	0	A1	0	0	9
2	A1	0	3	0	A1	0	1	1
3	A1	0	11	0	A2	1	0	7
4	A1	0	16	0	A2	1	1	3
5	A1	0	15	0	A3	2	0	3
6	A1	0	18	1	A3	2	1	7
7	A1	0	0	0	A4	3	0	1
8	A1	0	8	0	A4	3	1	9
9	A1	0	15	0		•	•	•
10	A1	0	4	0		•	•	•
11	A2	1	23	1		•	•	•
12	A2	1	11	0		•	•	•
13	A2	1	29	1		•	•	•

JMP［二変量の関係］の出力

p.237

- ［二変量の関係］の実行

  Y, 目的変数：yC（量的変数、e.g. 体重）

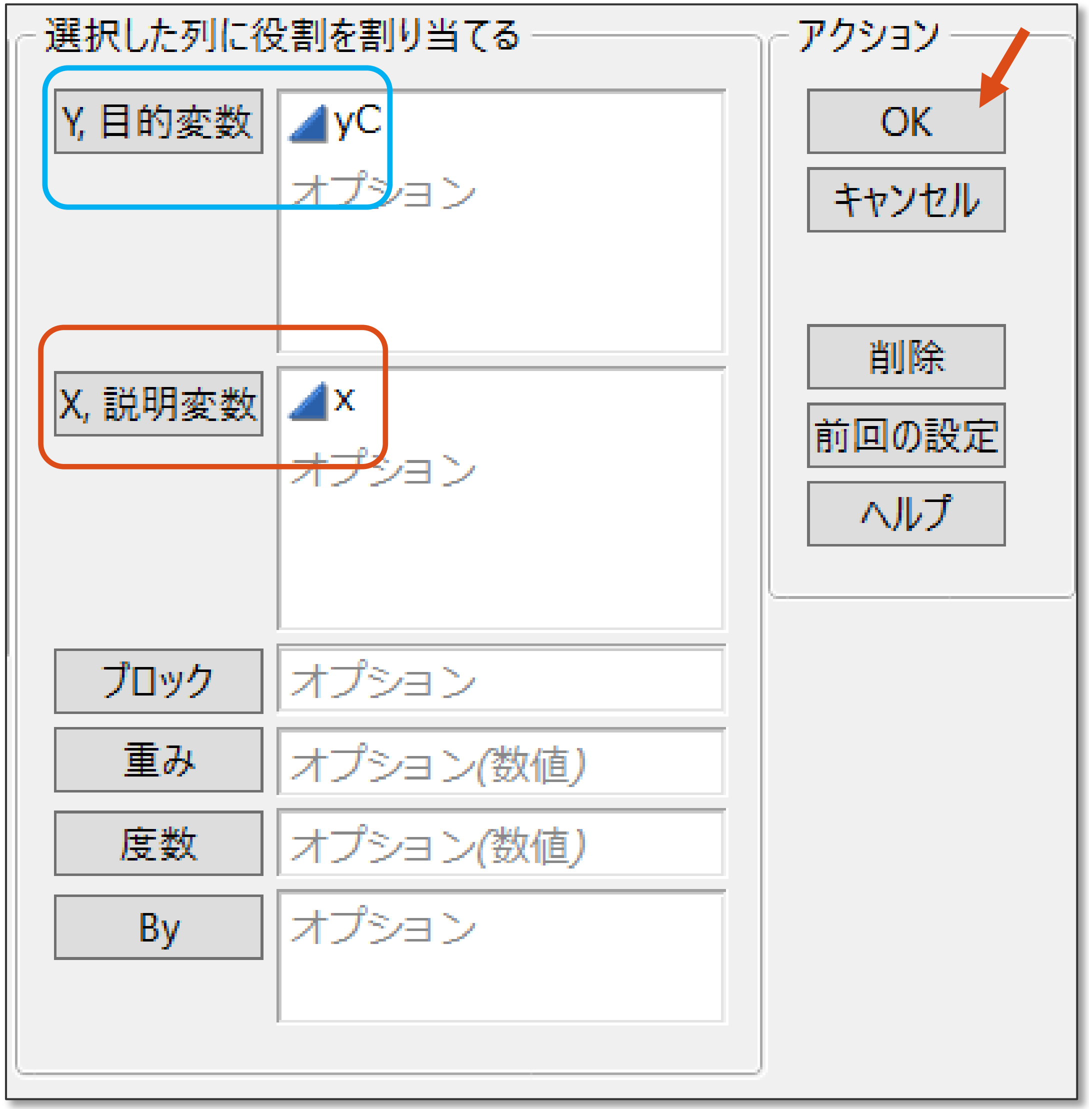
  X, 説明変数：x（量的変数、e.g. 薬剤の投与量）

表示 4.1.2 ４つの組合せデータ

A	x	1	2	3	4	5	6	7	8	9	10
A1	0	8	3	11	16	15	18	0	8	15	4
A2	1	23	11	29	15	16	14	22	22	[illegible]	2
A3	2	28	28	23	29	29	27	39	34	11	26
A4	3	48	32	42	35	30	39	37	43	38	43

yC

A	x	1	2	3	4	5	6	7	8	9	10
A1	0	0	0	0	0	0	1	0	0	0	0
A2	1	1	0	1	0	0	0	0	0	0	1
A3	2	1	1	0	1	1	0	1	1	1	0
A4	3	1	1	1	1	0	1	1	1	1	1



JMP［二変量の関係］の出力

p.237

- ［二変量の関係］の実行

  Y, 目的変数：yC（量的変数、e.g. 体重）

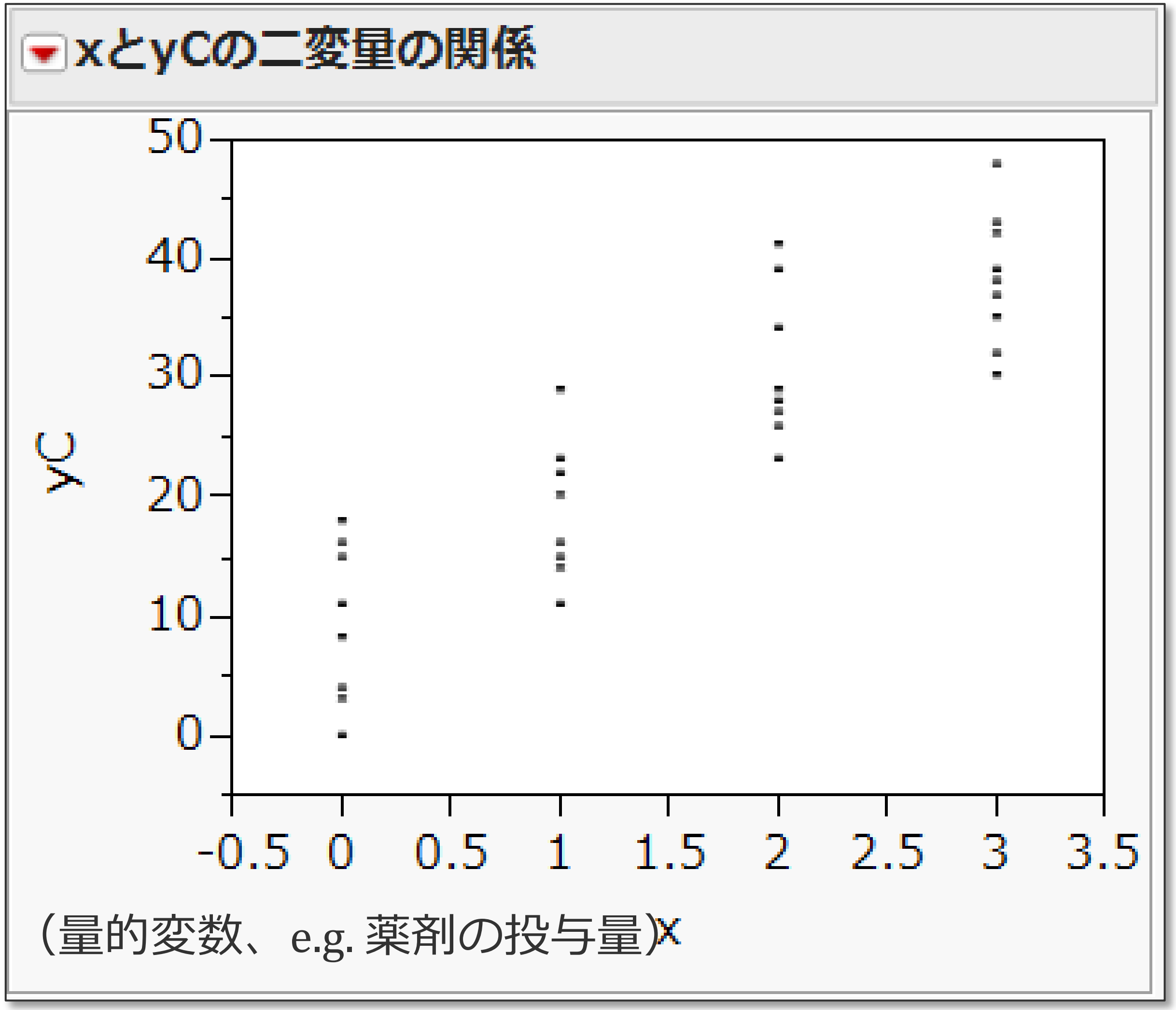
  X, 説明変数：x（量的変数、e.g. 薬剤の投与量）

表示 4.1.2 ４つの組合せデータ

A	x	1	2	3	4	5	6	7	8	9	10
A1	0	8	3	11	16	15	18	0	8	15	4
A2	1	23	11	29	15	16	14	22	22	[illegible]	[illegible]
A3	2	28	28	23	29	29	27	39	34	11	26
A4	3	48	32	42	35	30	39	37	43	38	43

yC

A	x	1	2	3	4	5	6	7	8	9	10
A1	0	0	0	0	0	0	1	0	0	0	0
A2	1	1	0	1	0	0	0	0	0	0	1
A3	2	1	1	0	1	1	0	1	1	1	0
A4	3	1	1	1	1	0	1	1	1	1	1



JMP［二変量の関係］の出力

- ［二変量の関係］の実行

  Y, 目的変数：yC（量的変数、e.g. 体重）

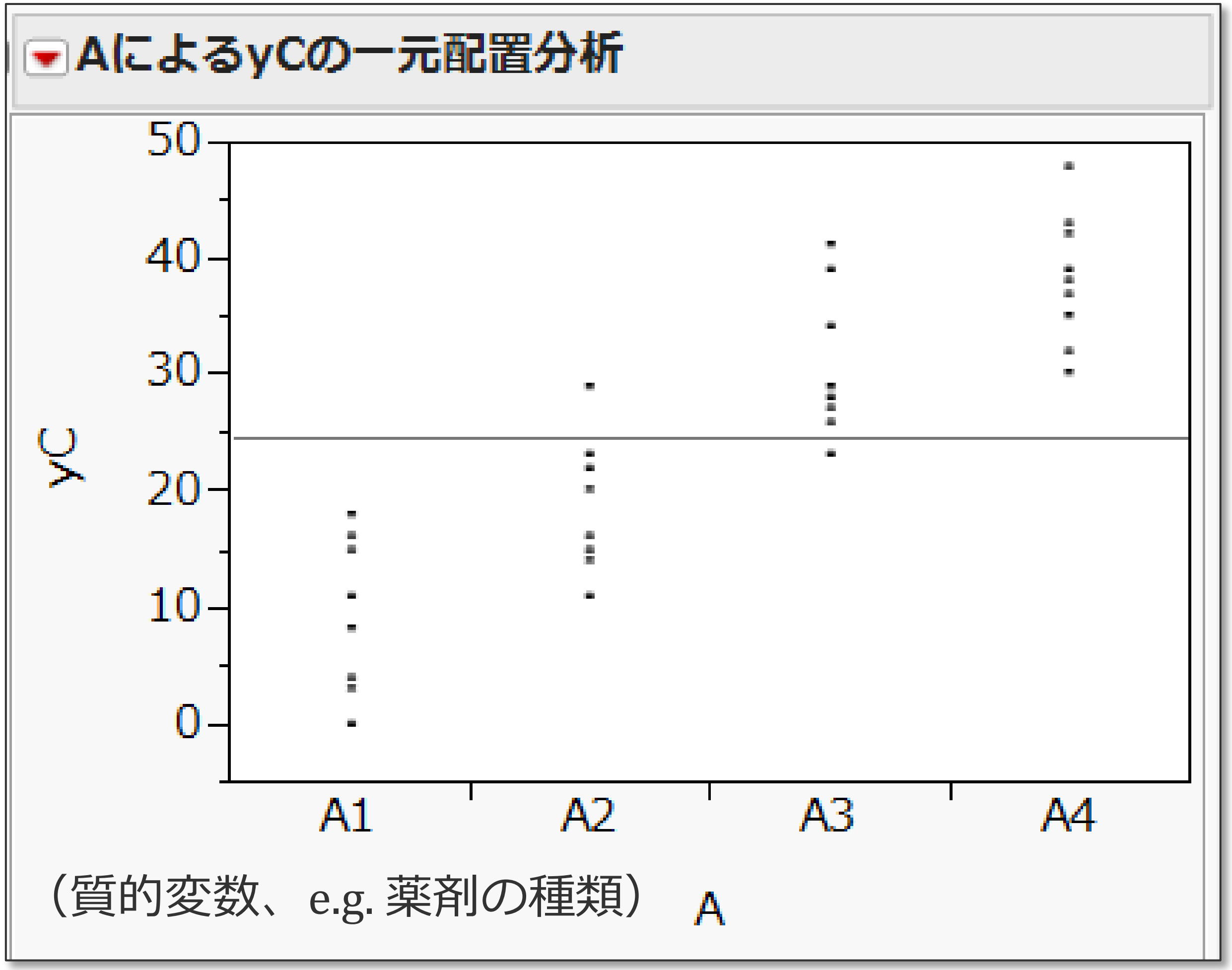
  X, 説明変数：A（質的変数、e.g. 薬剤の種類）

表示 4.1.2 ４つの組合せデータ

A	x	1	2	3	4	5	6	7	8	9	10
A1	0	8	3	11	16	15	18	0	8	15	4
A2	1	23	11	29	15	16	14	22	22	[illegible]	2
A3	2	28	28	23	29	29	27	39	34	11	26
A4	3	48	32	42	35	30	39	37	43	38	43

yC

A	x	1	2	3	4	5	6	7	8	9	10
A1	0	0	0	0	0	0	1	0	0	0	0
A2	1	1	0	1	0	0	0	0	0	0	1
A3	2	1	1	0	1	1	0	1	1	1	0
A4	3	1	1	1	1	0	1	1	1	1	1



JMP［二変量の関係］の出力

p.237

● ［二変量の関係］の実行

Y, 目的変数：yN（質的変数、e.g. 発疹の有無）

X, 説明変数：A（質的変数、e.g. 薬剤の種類）

注）「割合と「率」を区別しないで使用

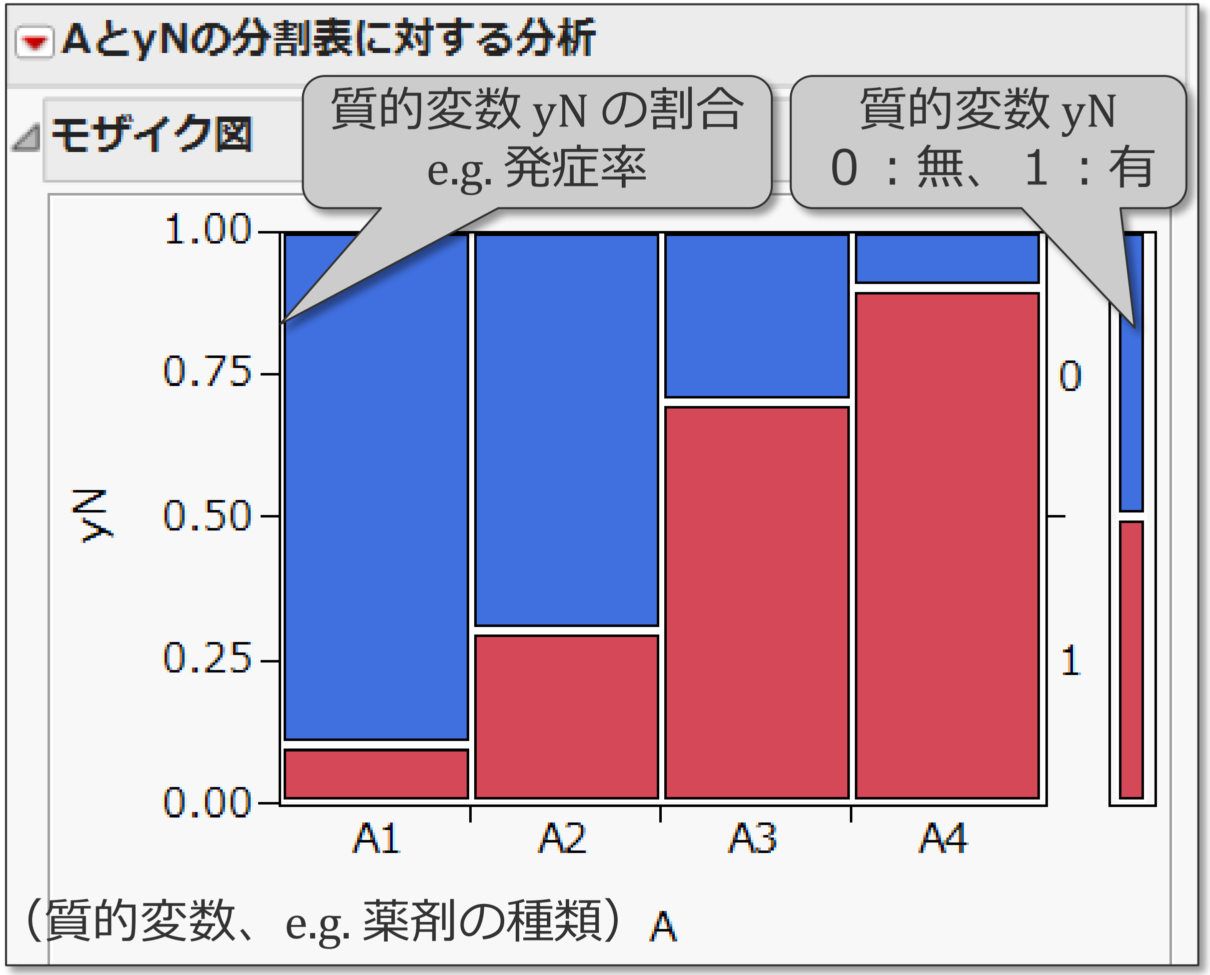
表示 4.1.2 ４つの組合せデータ

A	x	1	2	3	4	5	6	7	8	9	10
A1	0	8	3	11	16	15	18	0	8	15	4
A2	1	23	11	29	15	16	14	22	22	20	22
A3	2	28	28	23	29	29	27	39			
A4	3	48	32	42	35	30	39	37			

質的変数
名義尺度

A	x	1	2	3	4	5	6	7	8	9	10
A1	0	0	0	0	0	0	1	0	0	0	0
A2	1	1	0	1	0	0	0	0	0	0	1
A3	2	1	1	0	1	1	0	1	1		0
A4	3	1	1	1	1	0	1	1	1	1	1

yN



JMP［二変量の関係］の出力

- ［二変量の関係］の実行

  Y, 目的変数：yN（質的変数、e.g. 発疹の有無）

  X, 説明変数：x（量的変数、e.g. 薬剤の投与量）

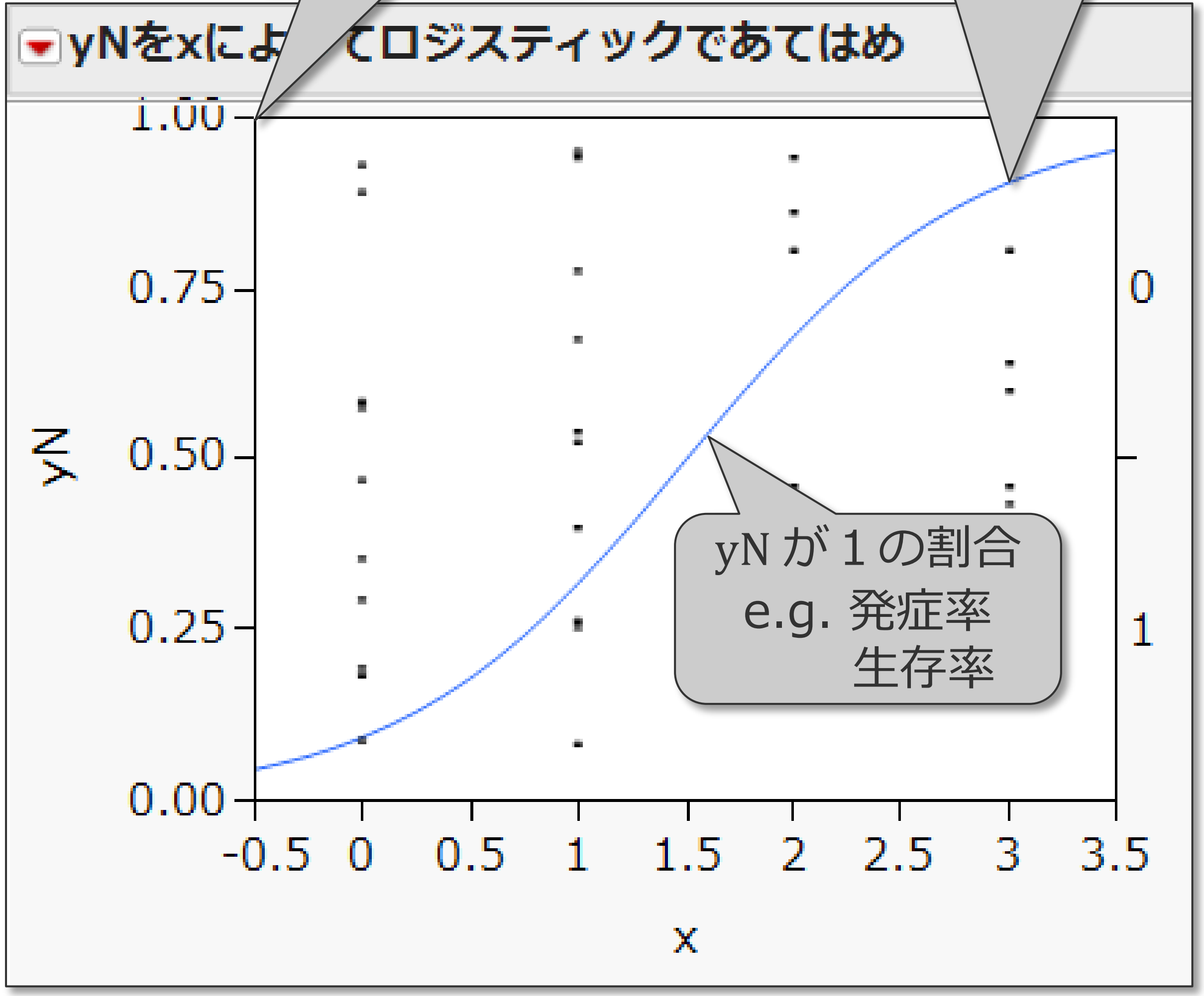
表示 4.1.2 4 つの組合せデータ

A	x	1	2	3	4	5	6	7	8	9	10
A1	0	8	3	11	16	15	18	0	8	15	4
A2	1	23	11	29	15	16	14	22	22	20	22
A3	2	28	28	23	29	29	27	39	[illegible]	[illegible]	[illegible]
A4	3	48	32	42	35	30	39	37	[illegible]	[illegible]	[illegible]

質的変数
名義尺度

A	x	1	2	3	4	5	6	7	8	9	10
A1	0	0	0	0	0	0	1	0	0	0	0
A2	1	1	0	1	0	0	0	0	0	0	1
A3	2	1	1	0	1	1	0	1	1	[illegible]	0
A4	3	1	1	1	1	0	1	1	1	1	1

yN



JMP［二変量の関係］の出力

p.237

- ［二変量の関係］の実行

  Y, 目的変数：yN（質的変数、e.g. 発疹の有無）

  X, 説明変数：x（量的変数、e.g. 薬剤の投与量）

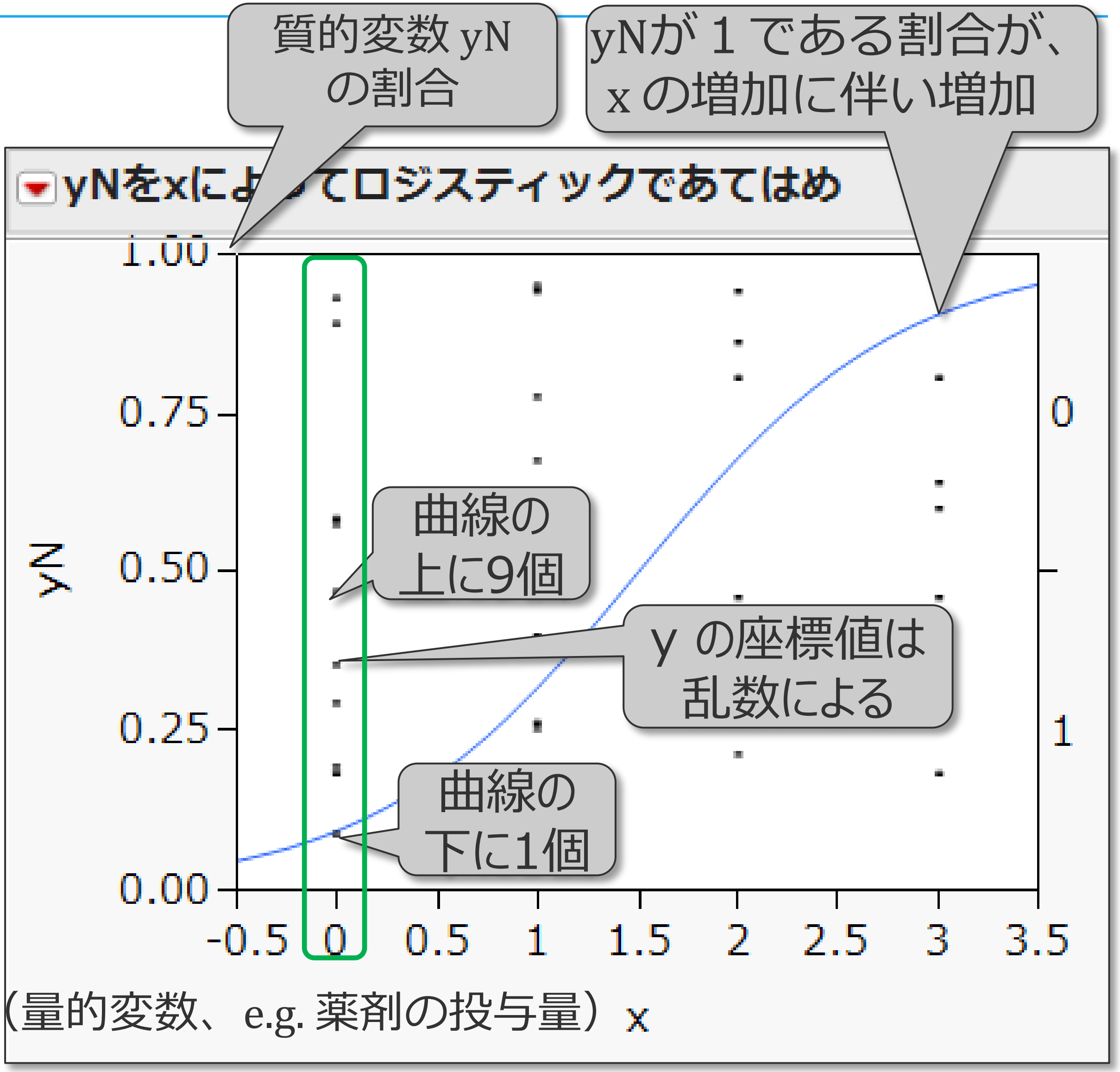
表示 4.1.2 ４つの組合せデータ

A	x	1	2	3	4	5	6	7	8	9	10
A1	0	8	3	11	16	15	18	0	8	15	4
A2	1	23	11	29	15	16	14	22	22	20	22
A3	2	28	28	23	29	29	27	39			
A4	3	48	32	42	35	30	39	37			

質的変数
名義尺度

A	x	1	2	3	4	5	6	7	8	9	10
A1	0	0	0	0	0	0	1	0	0	0	0
A2	1	1	0	1	0	0	0	0	0	0	1
A3	2	1	1	0	1	1	0	1	1		0
A4	3	1	1	1	1	0	1	1	1	1	1

yN



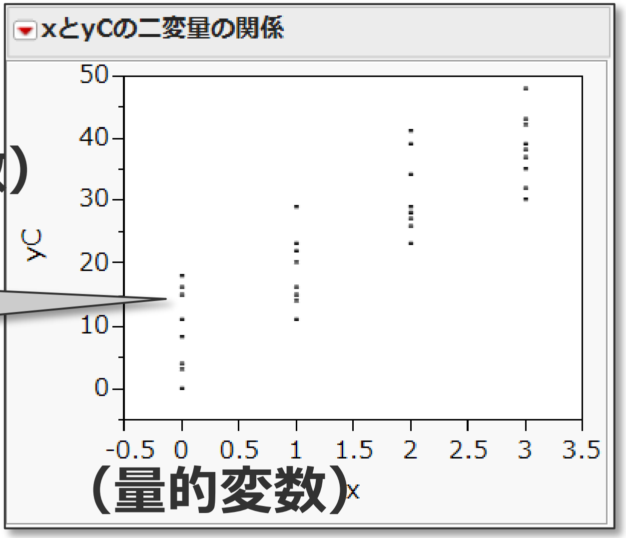
JMP［二変量の関係］の出力

p.237

表示 4.1.3 「二変量の関係」のグラフ出力

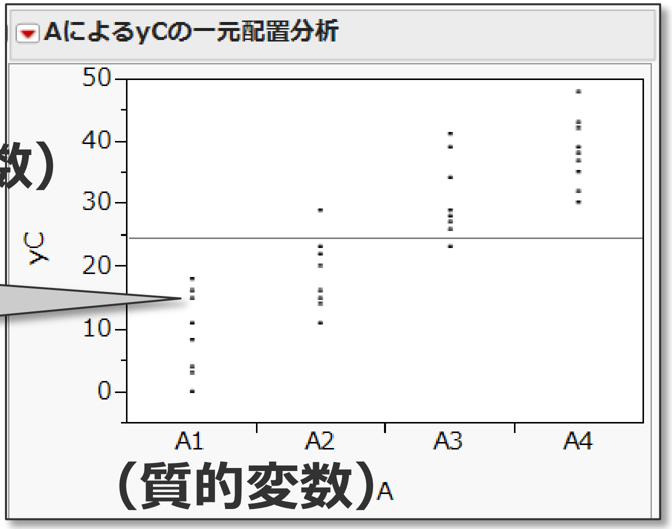
（量的変数）

点の位置は
y の値を示す



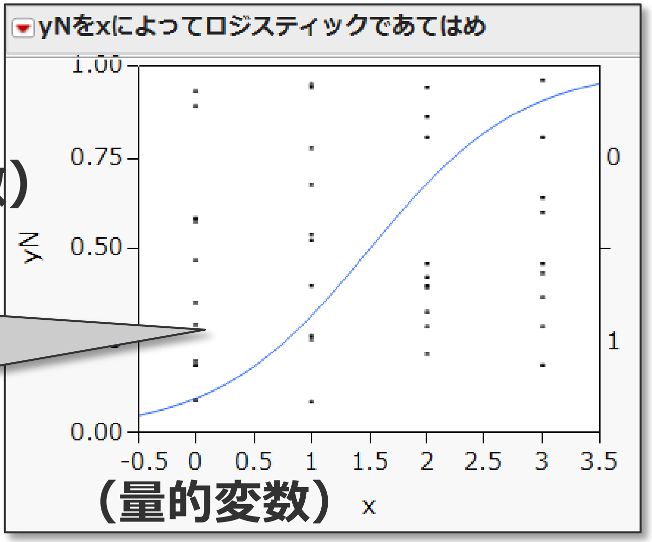
（量的変数）

点の位置は
y の値を示す



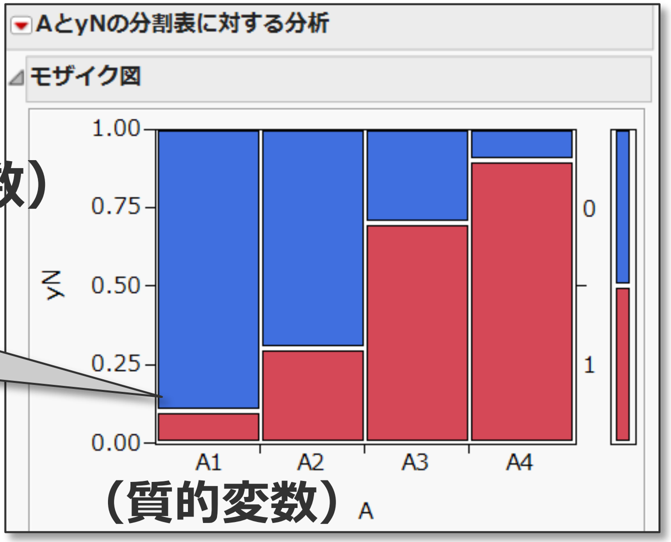
（質的変数）

yN の割合
$0 \leq yN \leq 1$
点の意味は
上図と異なる



（質的変数）

yN の割合
$0 \leq yN \leq 1$



p.237

（2）ロジスティック曲線

ロジット変換の意味
モデル式

ロジスティック曲線

p.237

●JMPによる解析

JMPファイル「41-二変量.jmp」の解析

名義尺度のカテゴリーの順序を指定（§3.3 ）

後に位置するカテゴリーが基準

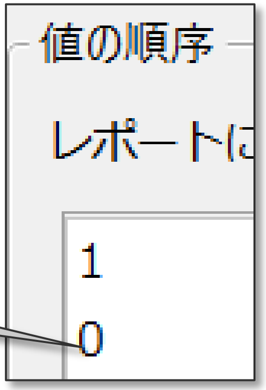
オッズでは分母になる

［分析］＞［二変量の関係］

yN：目的変数（質的変数）

x　：説明変数（量的変数）

［OK］をクリックして実行

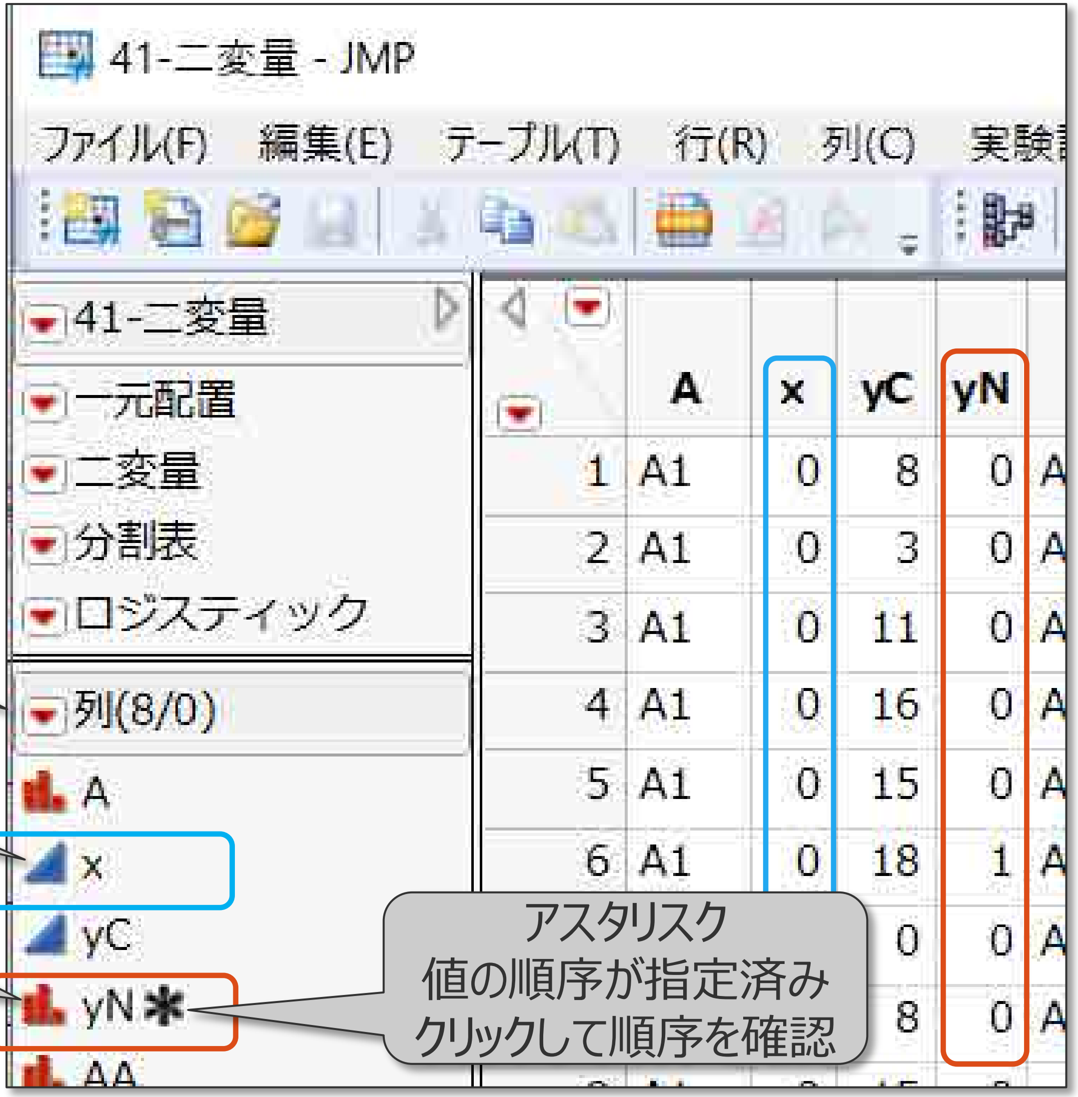


質的変数（名義尺度）
後（下）のカテゴリーが基準

列パネル

連測尺度

1, 0 であるが
名義尺度



アスタリスク
値の順序が指定済み
クリックして順序を確認

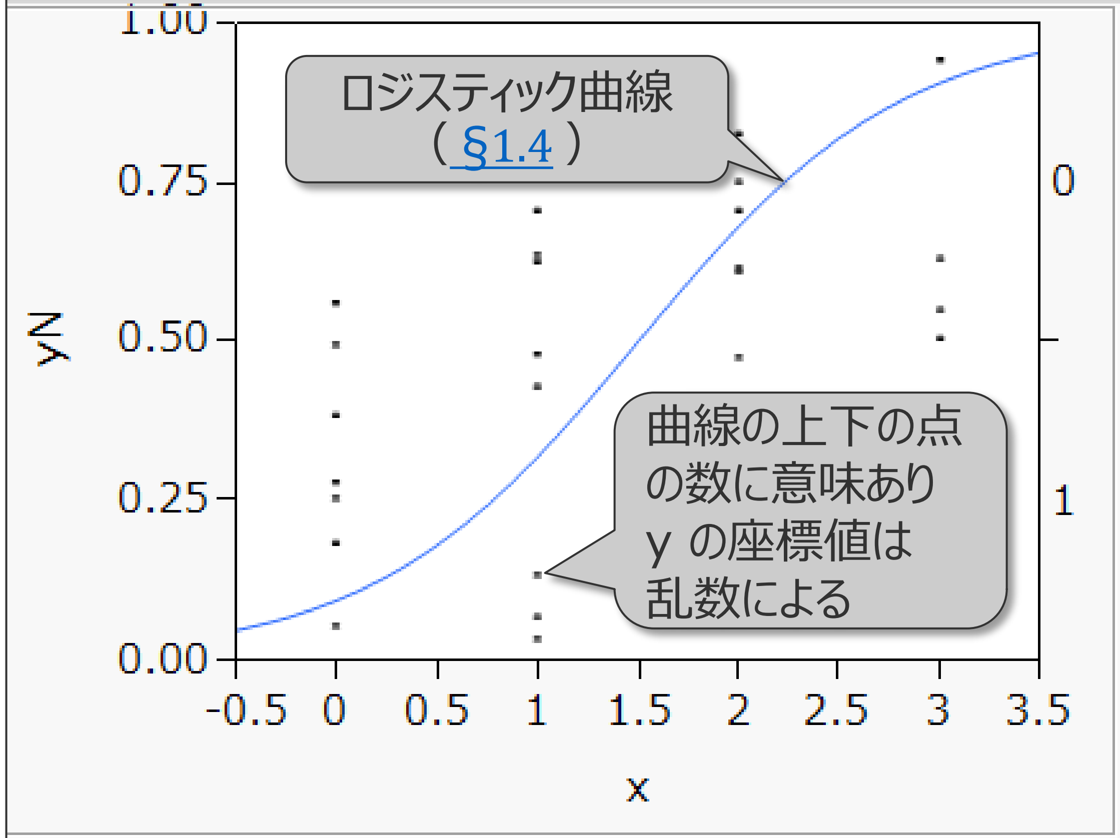
ロジスティック曲線

p.238

- ［二変量の関係］

表示 4.1.4 ロジスティック曲線と解析結果

yNをxによってロジスティックであてはめ



モデル全体の検定

モデル	(-1)*対数尤度	自由度	カイ2乗	p値(Prob>ChiSq)
差	8.984813	1	17.96963	<.0001*
完全	18.741074			
縮小	27.725887			

R2乗(U)	0.3241
AICc	41.8065
BIC	44.8599
オブザベーション(または重みの合計)	40

パラメータ推定値

項	推定値	標準誤差	カイ2乗	p値(Prob>ChiSq)
切片	-2.2708718	0.790164	8.26	0.0041*
x	1.51391451	0.4516065	11.24	0.0008*

推定値は次の対数オッズに対するものです： 1/0

ロジスティック曲線

●ロジスティック曲線とロジスティック回帰

量的変数 y が $0 \sim y_\infty$ の間で変化する S 字曲線

（§1.4）

$$y = \frac{y_\infty}{1 + \exp\left(-(a + bx)\right)} \qquad (4.1.1)$$

§1.4 Emax モデル
$X = \ln x$

割合 y が $0 \sim 1$ の間で変化、割合 p の期待値を π

$$\pi = \frac{1}{1 + \exp\left(-(\alpha + \beta x)\right)} \qquad (4.1.2)$$

パラメータの真値

π の推定値を p とする（推定値の式）

$$p = \frac{1}{1 + \exp\left(-(a + bx)\right)}$$

ロジスティック曲線

●ロジスティック曲線とロジスティック回帰

量的変数 y … 線

$$y = \frac{}{1 + \cdots}$$

$$p = \frac{1}{1 + \exp(-(a + bx))}$$

$$\frac{1}{p} = 1 + \exp(-(a + bx))$$

$$\exp(-(a + bx)) = e^{-(a+bx)}$$

$$-\ln e^{-(a+bx)} = a + bx$$

割合 y が 0 … と π

$$\pi = \frac{1}{1 + \exp(-(\alpha + \beta x))} \qquad (4.1.2)$$

割合 y が 0 ~ 1 の間で変化、 π の推定値を p

$$\boxed{p = \frac{1}{1 + \exp(-(a + bx))}}$$

pをロジット変換すると直線で表される

リスク $p = f/n$ とロジット（対数オッズ）z との関係は S 字曲線（§3.5）

$$z = \ln\left(\frac{p}{1-p}\right) \qquad (4.1.3)$$

$$= \ln\left(\frac{f/n}{(n-f)/n}\right) = \ln\left(\frac{f}{n-f}\right) \qquad (4.1.4)$$

$$z = \ln\left(\frac{p}{1-p}\right) = -\ln\left(\frac{1-p}{p}\right) = -\ln\left(\frac{1}{p} - 1\right)$$

$$= -\ln(1 + \exp(-(a + bx)) - 1)$$

$$= -\ln\left(\exp(-(a + bx))\right)$$

$$= a + bx \qquad (4.1.5)$$

ロジスティック曲線

●割合 p とロジット z の関係（スケールの変更）

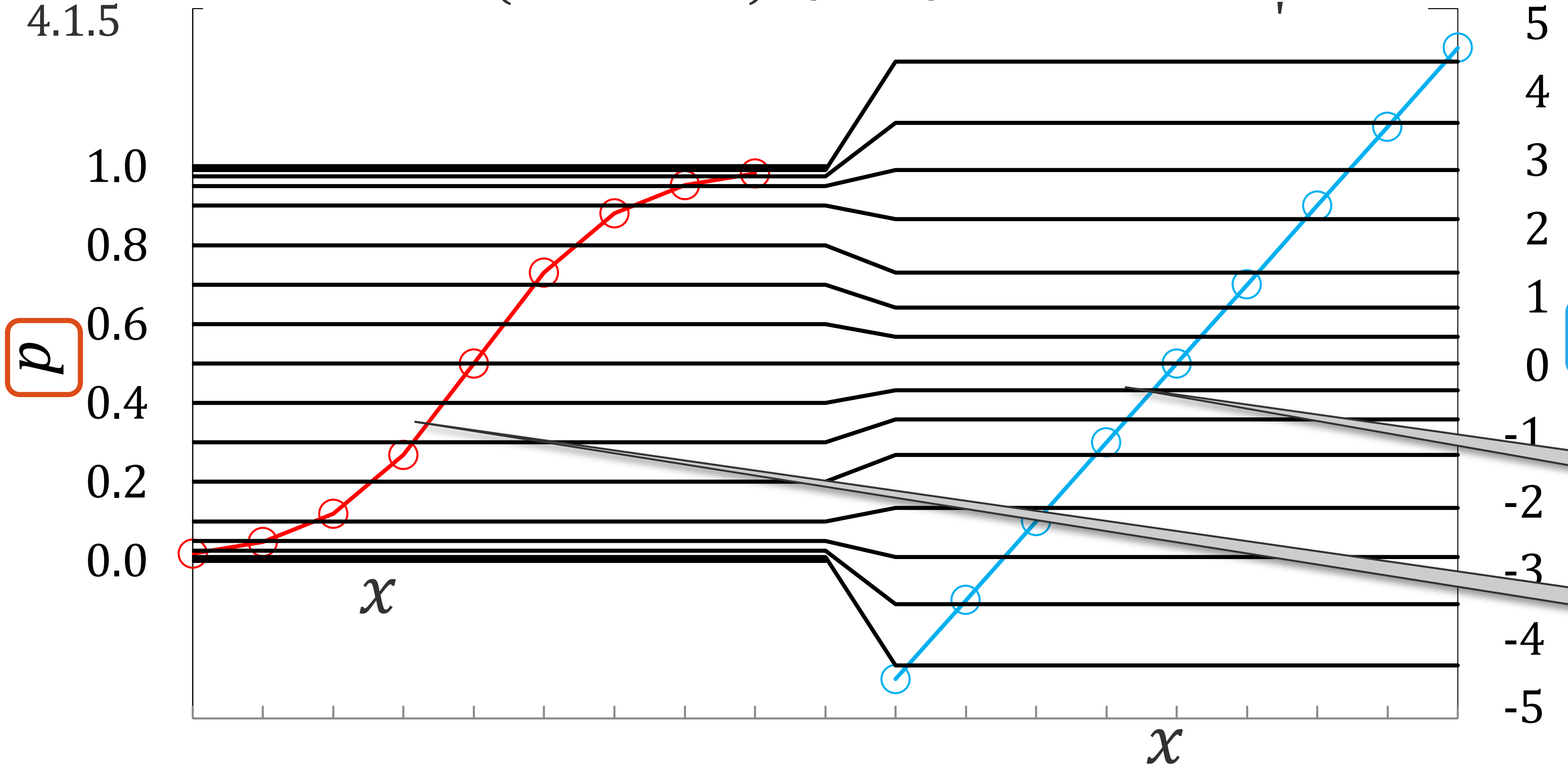
$$\boxed{\pi} = \frac{1}{1+\exp(-(\alpha+\beta x))} \quad (4.1.2)$$

$$\boxed{\zeta} = \alpha + \beta x \quad (4.1.5 改変)$$

式(4.1.5) で z の代わりに ζ とした式

表示
4.1.5

p と z の関係を示している



ロジスティック曲線

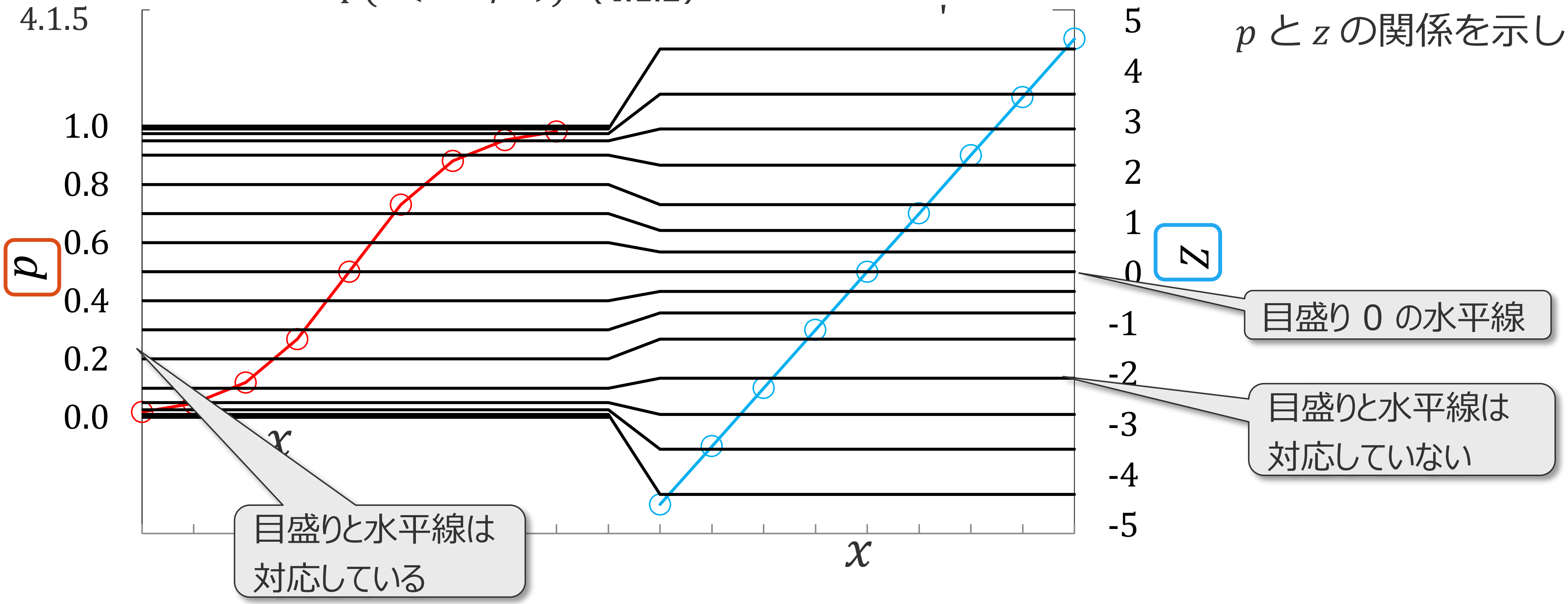
●割合 p とロジット z の関係（スケールの変更）

$$\boxed{\pi} = \frac{1}{1+\exp(-(\alpha+\beta x))} \quad (4.1.2)$$

$$\boxed{\zeta} = \alpha + \beta x \quad (4.1.5\text{ 改変})$$

式(4.1.5) で z の代わりに ζ とした式

表示
4.1.5



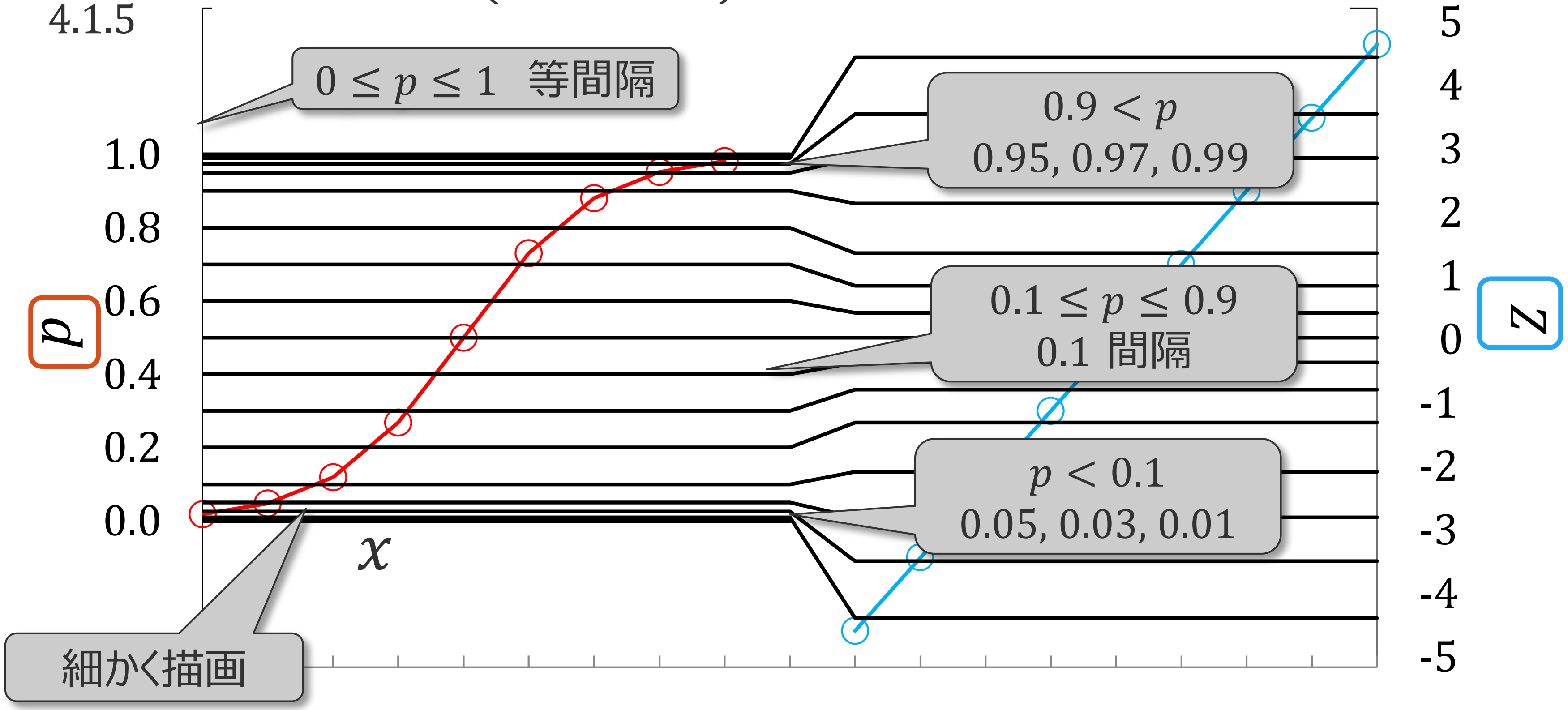
p と z の関係を示している

ロジスティック曲線

●割合 p とロジット z の関係（スケールの変更）

表示 4.1.5

$$\boxed{\pi} = \frac{1}{1+\exp(-(\alpha+\beta x))} \quad (4.1.2) \qquad \boxed{\zeta} = \alpha + \beta x \quad (4.1.5\ 改変)$$



p と z の関係を示している

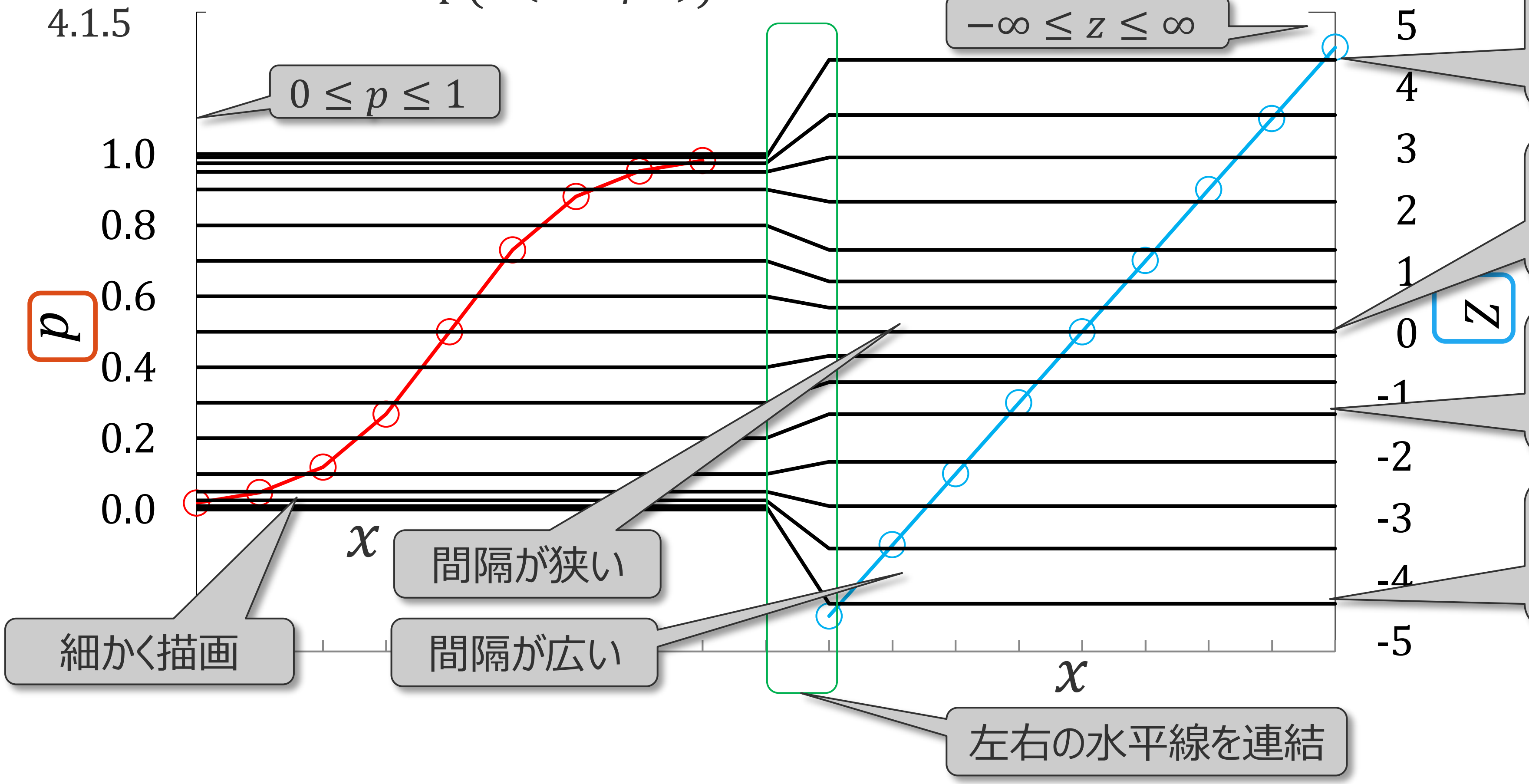
ロジスティック曲線

●割合 p とロジット z の関係（スケールの変更）

$$\boxed{\pi} = \frac{1}{1+\exp(-(\alpha+\beta x))} \quad (4.1.2)$$

$$\boxed{\zeta} = \alpha + \beta x \quad (4.1.5\ 改変)$$

$z = \ln\left(\frac{p}{1-p}\right)$ ロジット変換 p と z の対応



$z = \ln\left(\frac{0.99}{1-0.99}\right) = 4.6$

$z = \ln\left(\frac{0.5}{1-0.5}\right) = 0$

$z = \ln\left(\frac{0.2}{1-0.2}\right) = -1.4$

$z = \ln\left(\frac{0.01}{1-0.01}\right) = -4.6$

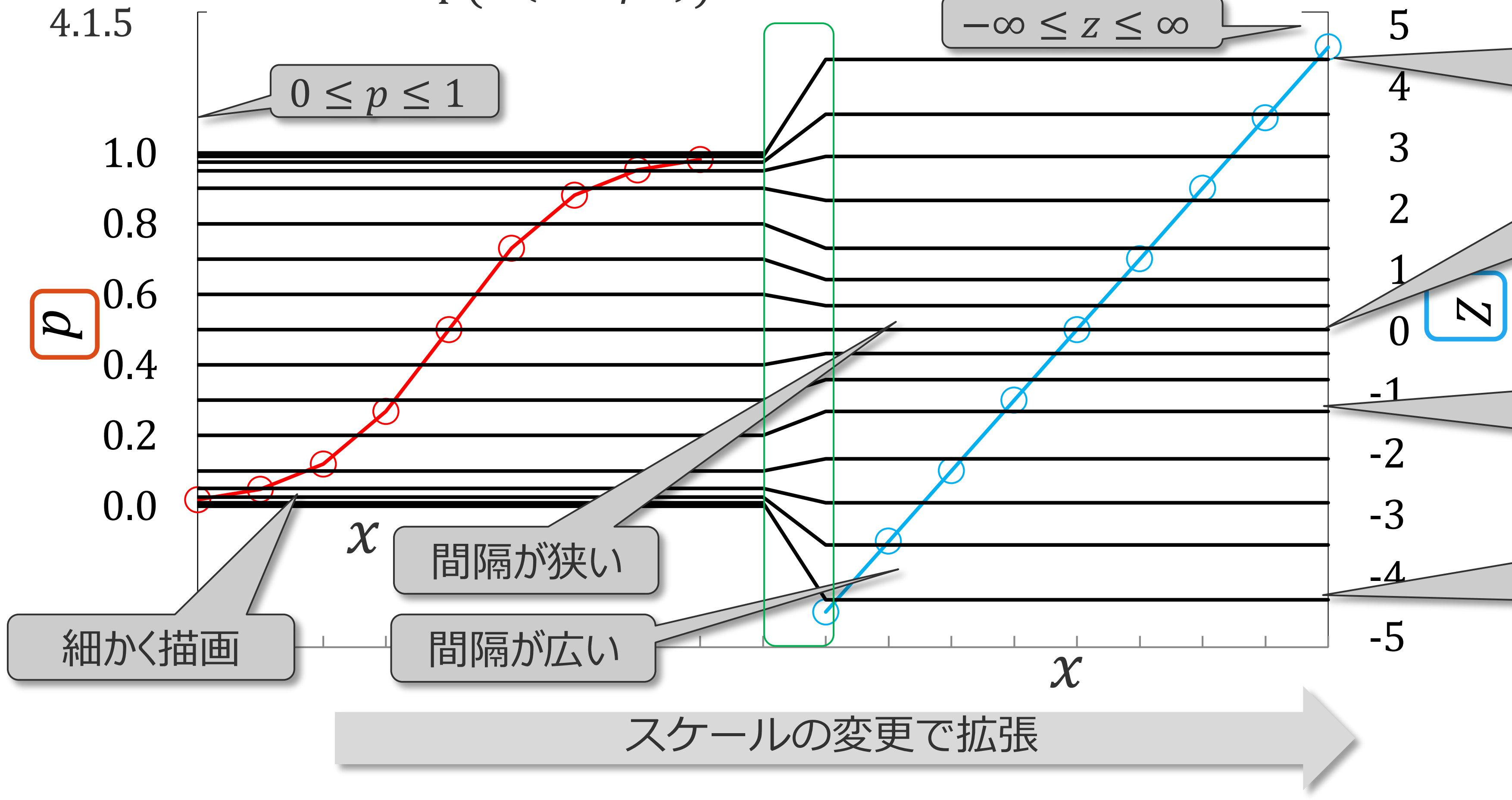
ロジスティック曲線

p.239

●割合 p とロジット z の関係（スケールの変更）

$$\boxed{\pi} = \frac{1}{1+\exp(-(\alpha+\beta x))} \quad (4.1.2) \qquad \boxed{\zeta} = \alpha + \beta x \quad (4.1.5\ 改変)$$

表示
4.1.5



$z = \ln\left(\frac{p}{1-p}\right)$ ロジット変換 p と z の対応

$z = \ln\left(\frac{0.99}{1-0.99}\right) = 4.6$

$z = \ln\left(\frac{0.5}{1-0.5}\right) = 0$

$z = \ln\left(\frac{0.2}{1-0.2}\right) = -1.4$

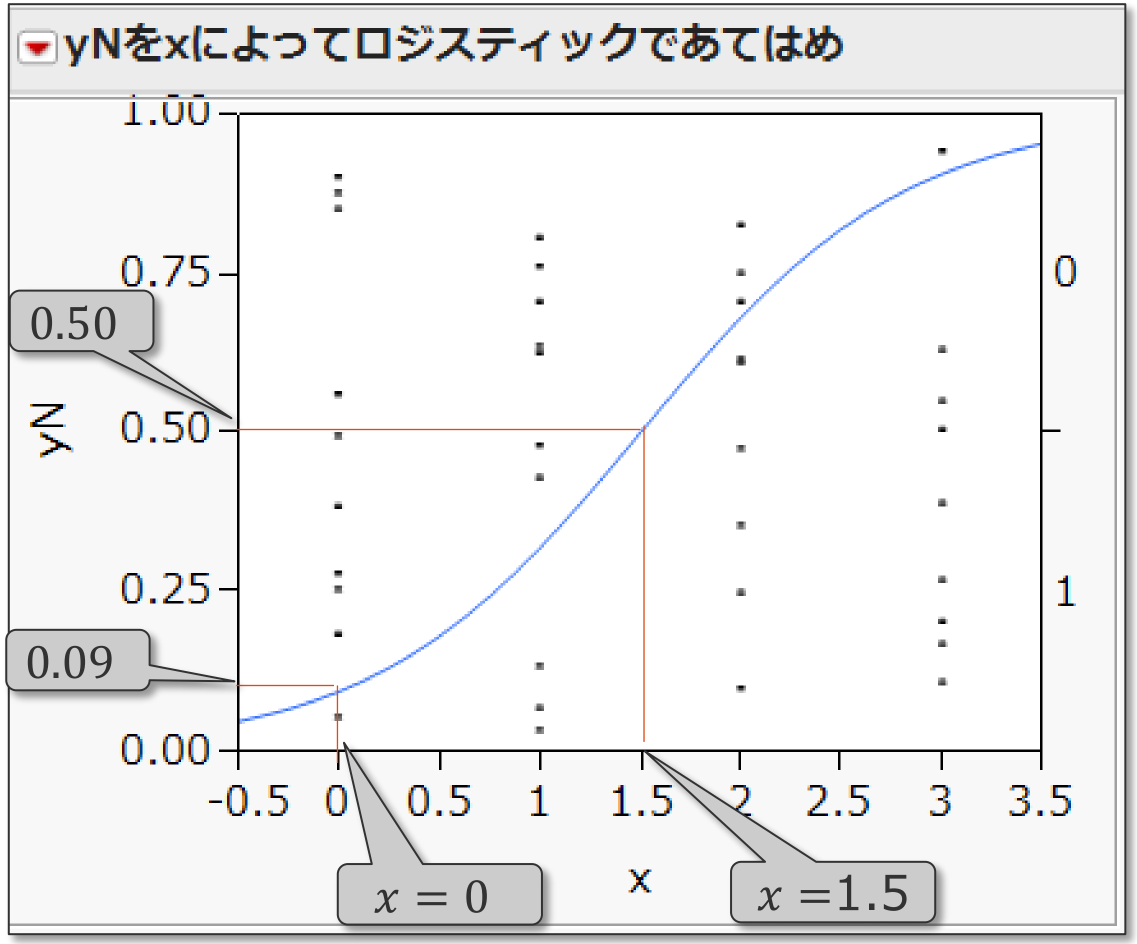
$z = \ln\left(\frac{0.01}{1-0.01}\right) = -4.6$

ロジスティック曲線

p.239

●パラメータ推定値

表示 4.1.4 ロジスティック曲線と解析結果



$$z = a + bx = -2.271 + 1.514x$$

$$p = \frac{1}{1 + \exp(-(-2.271 + 1.514 \times 0))}$$

$x = 0$

$$= \frac{1}{1 + \exp(-(-2.271))} = 0.09$$

$$p = \frac{1}{1 + \exp(-(-2.271 + 1.514 \times 1.5))}$$

$x = 1.5$

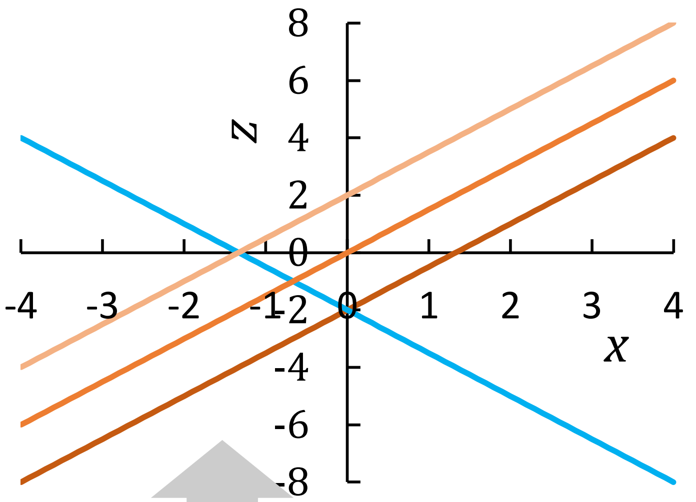
$$= \frac{1}{1 + \exp(0)} = \frac{1}{1 + 1} = 0.50$$

パラメータ推定値

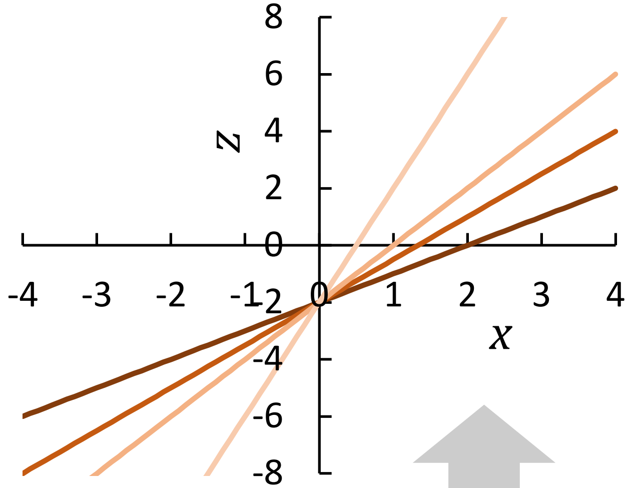
項	推定値	標準誤差	カイ2乗	p値(Prob>ChiSq)
切片	-2.2708718	0.790164	8.26	0.0041*
x	1.51391451	0.4516065	11.24	0.0008*

推定値は次の対数オッズに対するものです： 1/0

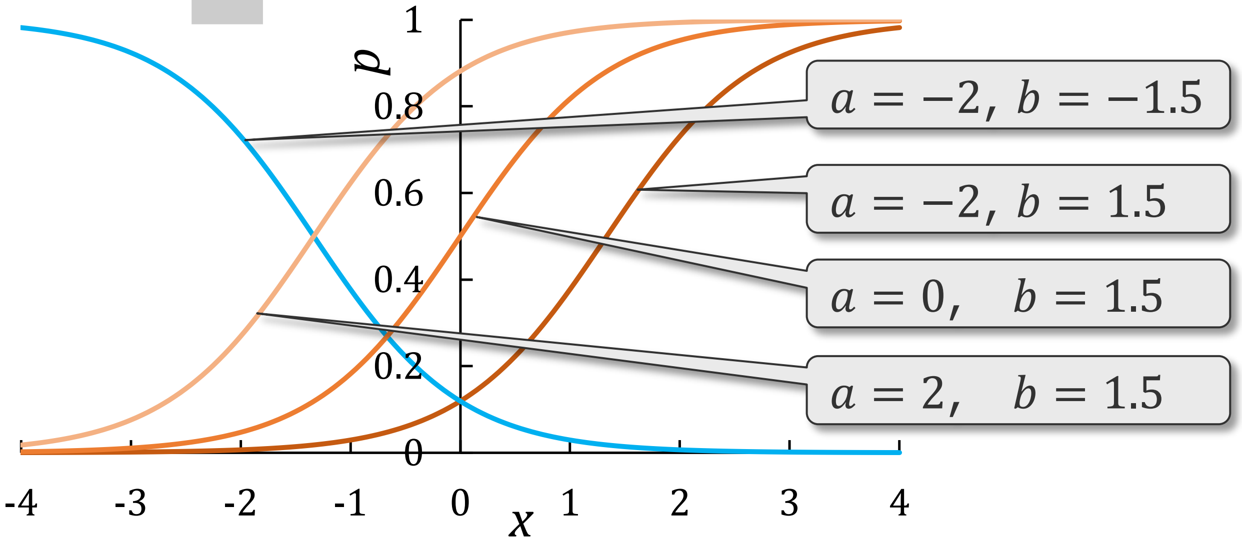
ロジスティック曲線

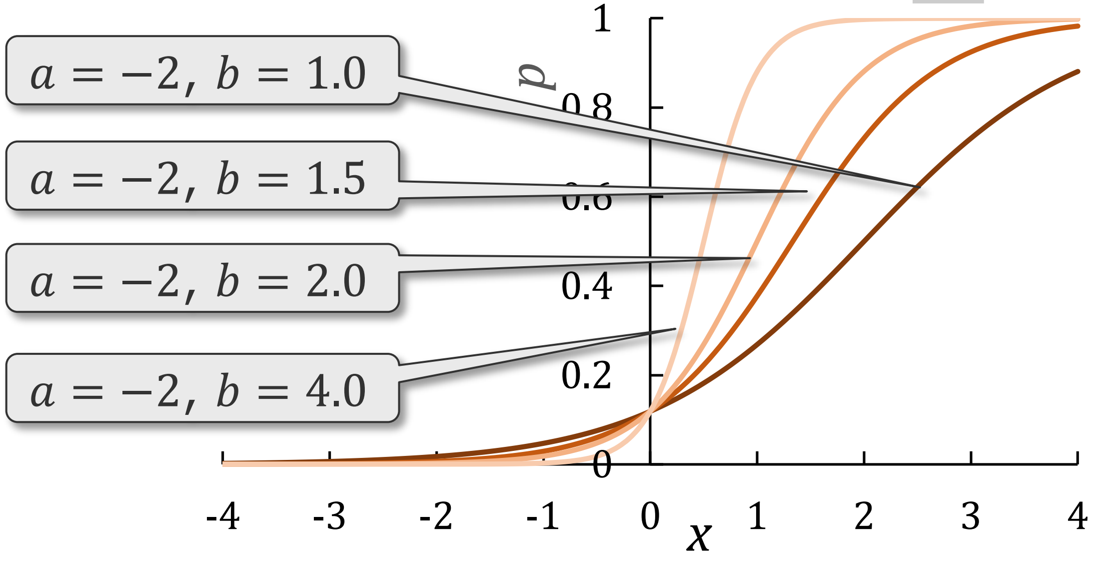


$$z = \ln\left(\frac{p}{1-p}\right) = a + bx$$



$$p = \frac{1}{1 + \exp\left(-(a + bx)\right)}$$





ロジスティック曲線

p.240

●ロジスティック回帰分析のモデル式と JMP での解析

パラメータに関して z が線形（§1.1）

p と z の関係

［二変量の関係］ $$p = \frac{1}{1+\exp\left(-(a+bx)\right)} \qquad z = a + bx \qquad (4.1.6)$$ 直線関係

［モデルのあてはめ］ $$p = \frac{1}{1+\exp\left(-(a+bx+cx^2)\right)} \qquad z = a + bx + cx^2 \qquad (4.1.7)$$ ２次式の関係

$$p = \frac{1}{1+\exp\left(-(a+bx_1+cx_2)\right)} \qquad z = a + bx_1 + cx_2 \qquad (4.1.8)$$ 多項式の関係

$$p = \frac{1}{1+\exp\left(-(a+b\sqrt{x})\right)} \qquad z = a + b\sqrt{x} \qquad (4.1.9)$$ z の平方根と直線関係

パラメータに関して z が非線形（§1.1）

［非線形回帰］ $$p = \frac{1}{1+\exp\left(-b(x-x_{50})\right)} \qquad z = b(x - x_{50})$$

ロジスティック曲線

p.240

●ロジスティック回帰分析の手法の選択

モデル式の線形、非線形と解析方法

z がパラメータに関して線形な場合　：JMP［二変量の関係］、JMP［モデルのあてはめ］

z がパラメータに関して非線形な場合：JMP［非線形回帰］・・・ §4.2 (6) 以降

JMP でロジスティック回帰分析を行うには、モデル式により３つの中から適切な手法を選択

［二変量の関係］＜［モデルのあてはめ］＜［非線形回帰］

適用範囲は、右の手法ほど広い

式の入力や設定などは、右の手法ほど複雑になる（作業の手間が必要になる）

次節以降では、この３つの方法の解析が説明されるので、このことに留意して理解する

p.240

（3）誤差の構造

最小２乗法
非線形最小２乗法
最尤法

誤差の構造

●単回帰モデル（第１部）

期待値

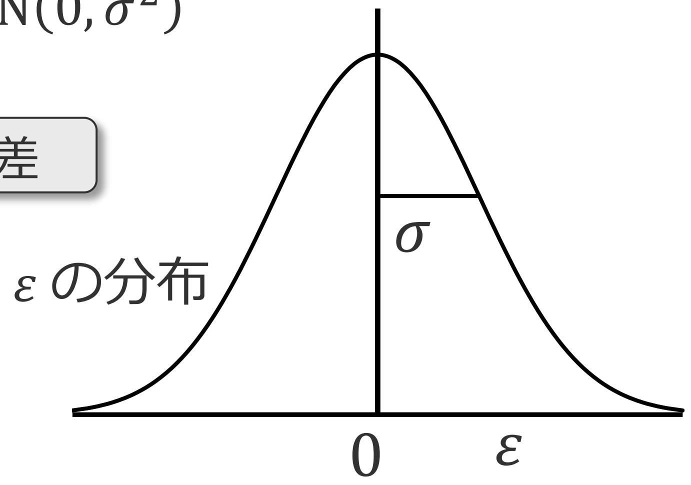
$\eta = \alpha + \beta x$　　（η ：母平均、α：母切片、β：母回帰係数）（4.3.2　第１部 §4.3 ）

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

$y = \eta + \varepsilon = \alpha + \beta x + \varepsilon$　　（4.3.3　第１部）・・・単回帰モデル

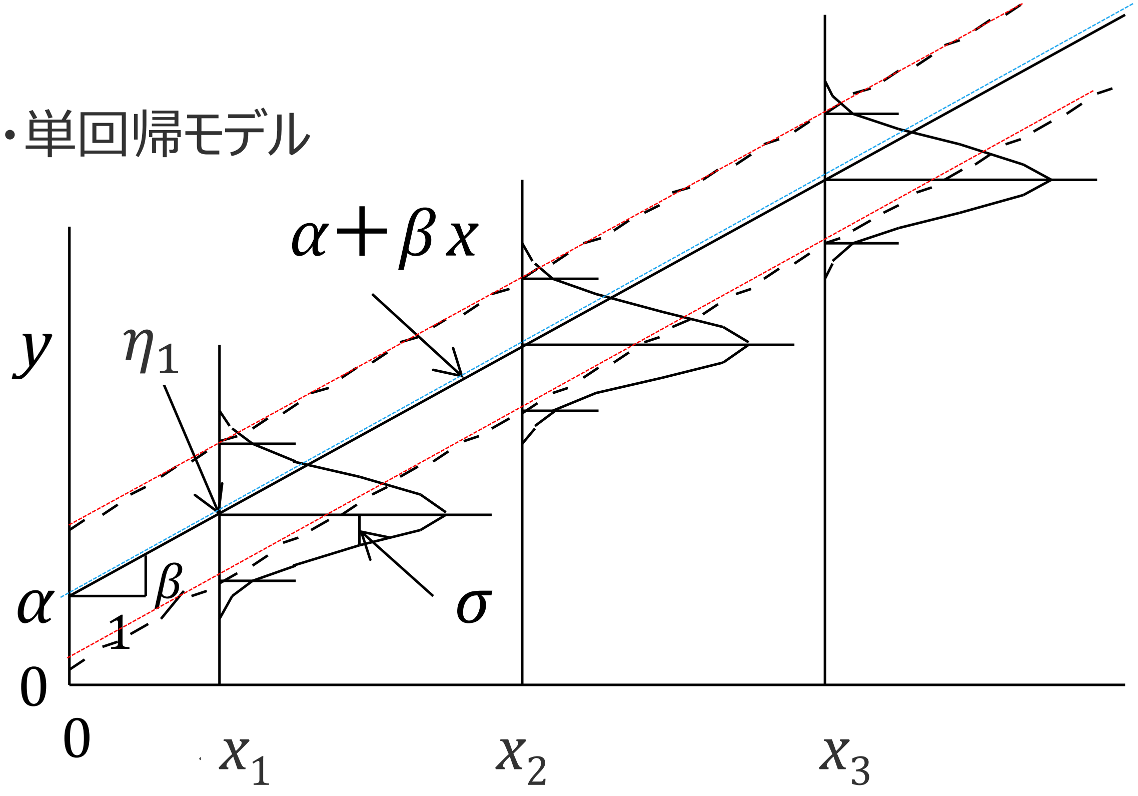
$\varepsilon \sim \mathrm{N}(0, \sigma^2)$

誤差



特定の値 x_1 を多数集めてきたときの y_1 の分布

$\varepsilon = y_1 - \eta_1$



表示 4.3.2　回帰モデル（第１部）

誤差の構造

●単回帰モデル（第１部§4.3）

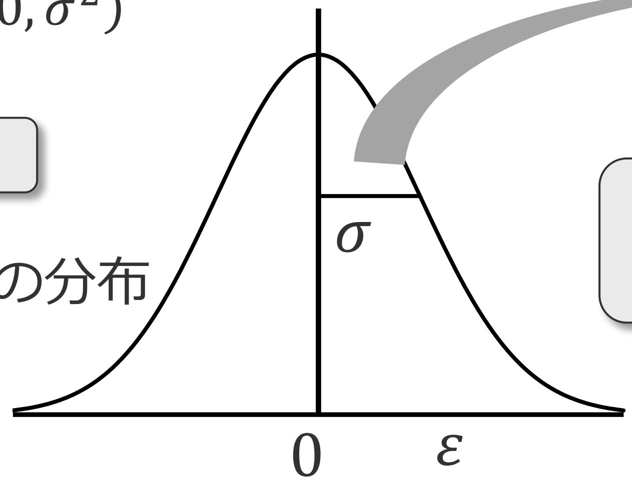
$\eta = \alpha + \beta x$　（η ：母平均、α：母切片、β：母回帰係数）（4.3.2　第１部 §4.3 ）

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

$y = \eta + \varepsilon = \alpha + \beta x + \varepsilon$　（4.3.3　第１部）・・・単回帰モデル

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

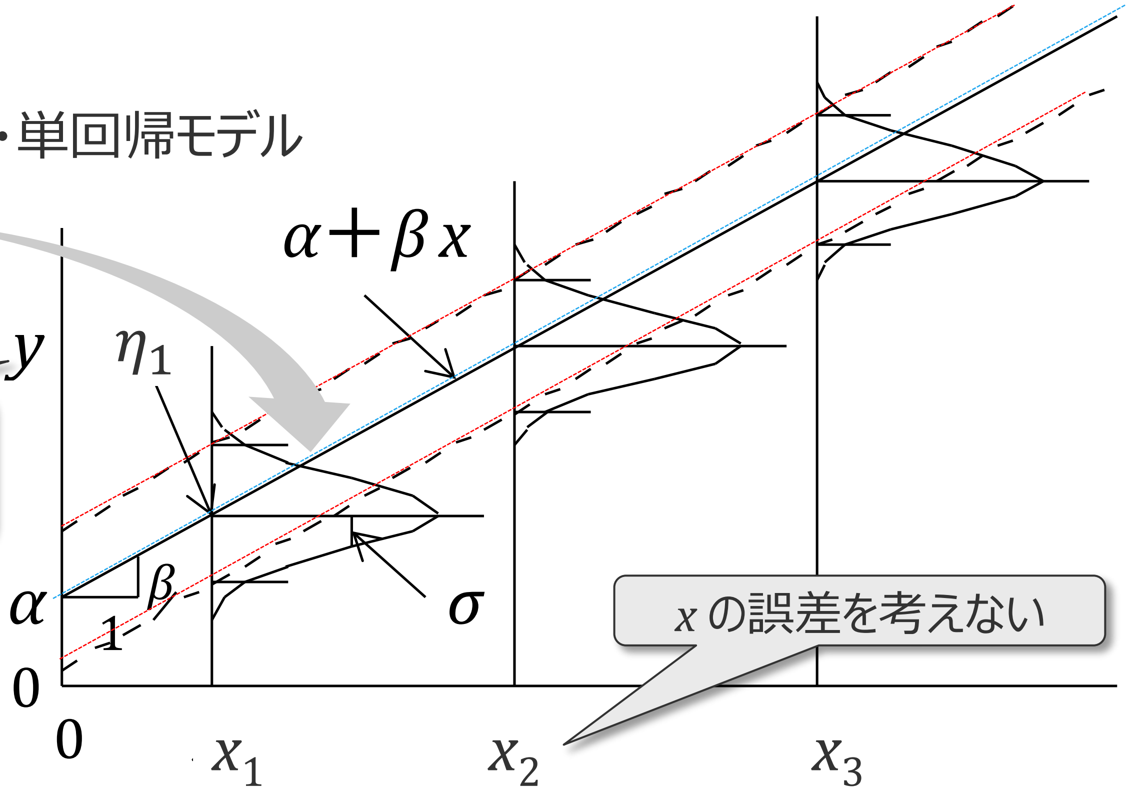
誤差



特定の値 x_1 を多数集めてきたときの y_1 の分布

$\varepsilon = y_1 - \eta_1$

y には
誤差 ε がある



x の誤差を考えない

表示 4.3.2　回帰モデル（第１部）

誤差の構造

p.241

●ロジスティック曲線の誤差

（1）量的変数の観測値 y は式(4.1.1)に誤差が加わったもの

誤差は x によらず一定（等分散）で、正規分布に従う → 最小２乗法を利用（§1、§2）

ロジスティック曲線のあてはめ

$$y = \frac{y_\infty}{1+\exp\left(-(a+bx)\right)} \qquad (4.1.1)$$

（2）割合の観測値 p は式(4.1.6) に誤差が加わったもの

誤差は２項分布（離散分布）、非対称、分散は一定ではない → 最小２乗法は使えない

ロジスティック回帰分析

$$p = \frac{1}{1+\exp\left(-(a+bx)\right)} \qquad (4.1.6)$$

（2）は（1）の拡張ではない

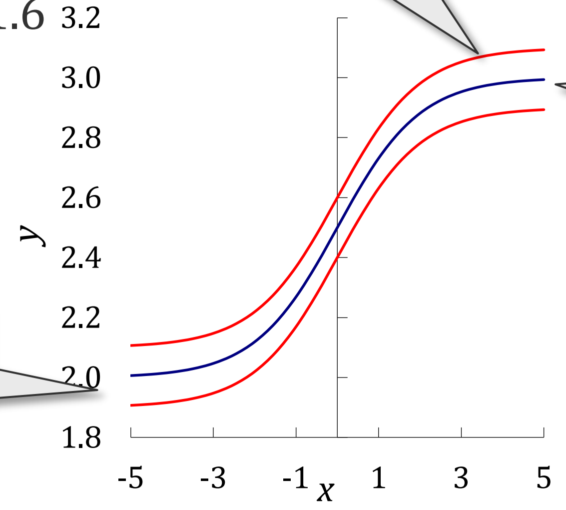
誤差の構造

p.241

S字曲線

表示4.1.6

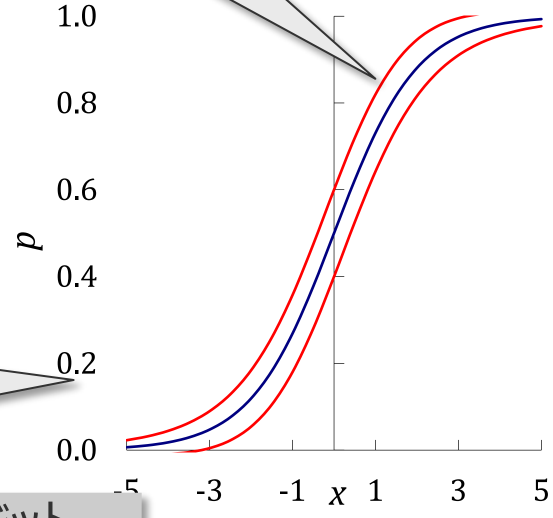
標準偏差の幅



期待値（母平均）
の変化

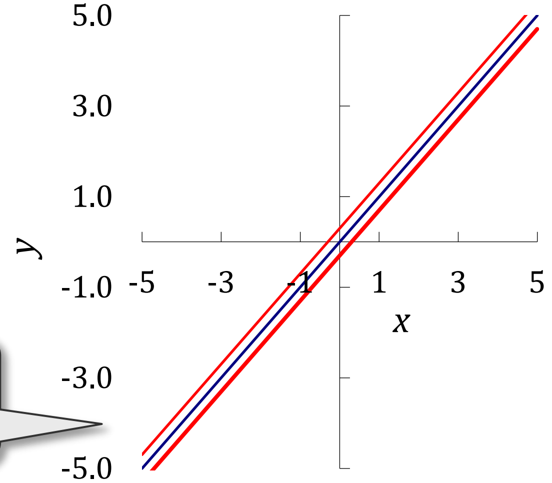
縦軸は量的変数
（誤差は等分散）

標準偏差の幅



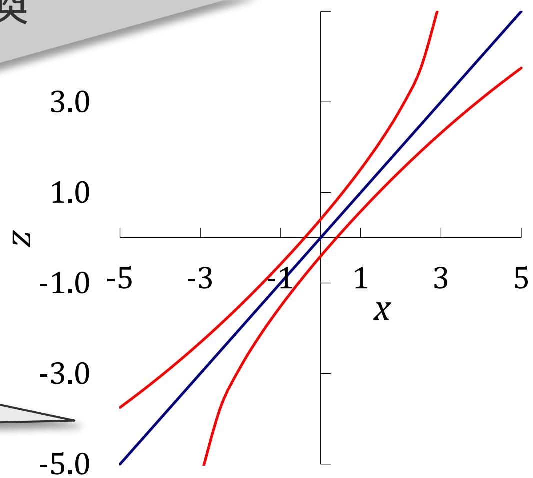
縦軸は割合
（誤差は不等分散）

直線



縦軸は量的変数
（誤差は等分散）

ロジット
変換



縦軸はロジット
（誤差は不等分散）

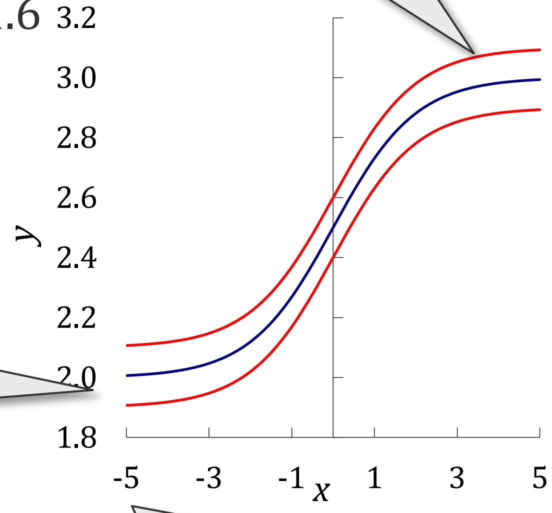
誤差の構造

p.241

S 字曲線

表示4.1.6

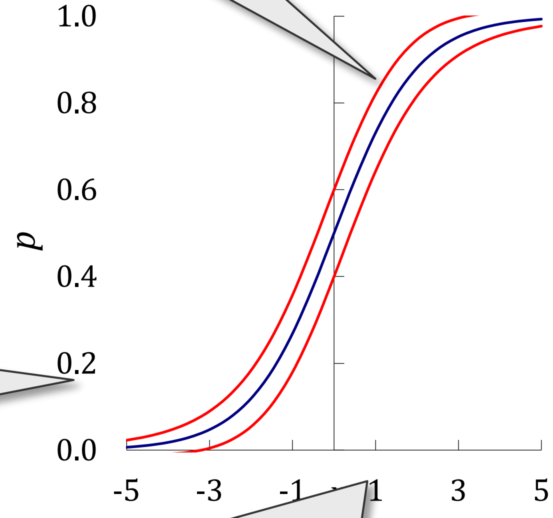
標準偏差の幅



縦軸は量的変数
（誤差は等分散）

誤差構造が違う

標準偏差の幅



縦軸は割合
（誤差は不等分散）

$y_{-\infty} = 2, y_{\infty} = 3$ のロジスティック曲線
上下の曲線は、$\pm\sigma = \pm 0.1$ の幅の曲線
（幅はどこでも同じ）
３本の曲線は上下に移動すると重なる
（等分散性が成立）
誤差の分布は正規分布（対称）

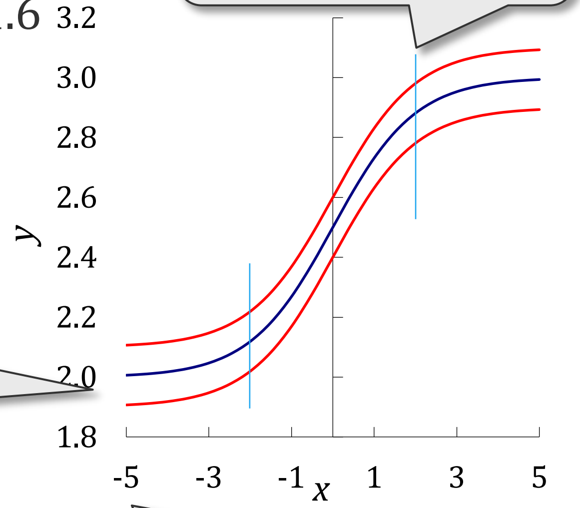
$0 \leq p \leq 1$のロジスティック曲線
上下の曲線は、２項分布の標準偏差の幅
標準偏差は $p = 0.5$ のとき最大、
0 と 1 に近づくと狭くなる（等分散性が不成立）
誤差の分布は非対称の山型（図にはない）、
内側($p = 0.5$ の方向) の方に裾を引く

誤差の構造

p.241

S 字曲線

表示4.1.6



どこでも同じ
正規分布

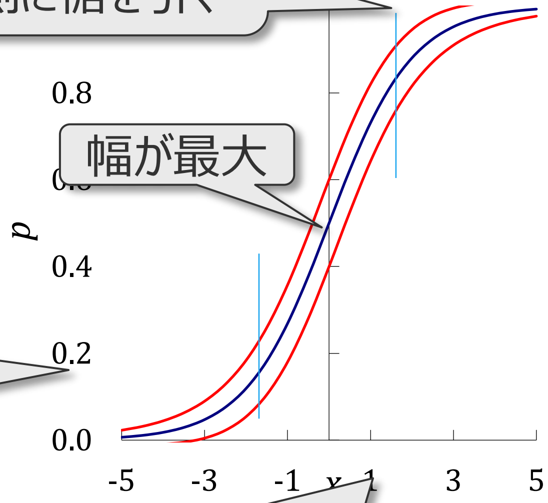
縦軸は量的変数
（誤差は等分散）

誤差構造が違う

分布は非対称
内側に裾を引く

幅が最大

縦軸は割合
（誤差は不等分散）



$y_{-\infty} = 2, y_{\infty} = 3$ のロジスティック曲線
上下の曲線は、$\pm\sigma = \pm 0.1$ の幅の曲線
（幅はどこでも同じ）
３本の曲線は上下に移動すると重なる
（等分散性が成立）
誤差の分布は正規分布（対称）

$0 \leq p \leq 1$のロジスティック曲線
上下の曲線は、２項分布の標準偏差の幅
標準偏差は $p = 0.5$ のとき最大、
0 と 1 に近づくと狭くなる（等分散性が不成立）
誤差の分布は非対称の山型（図にはない）
内側($p = 0.5$ の方向) の方に裾を引く

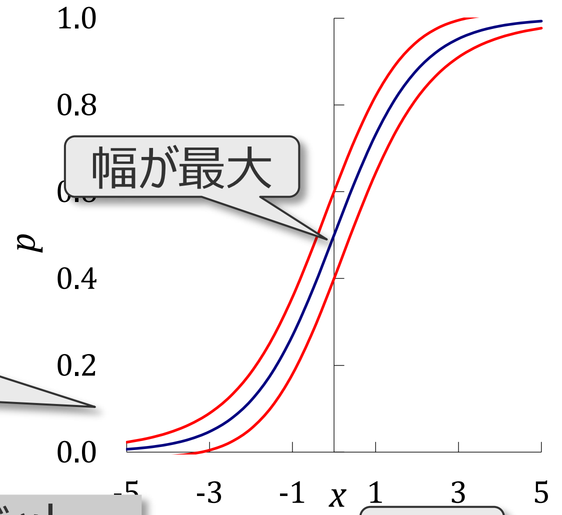
誤差の構造

p.241

S 字曲線

$0 \le p \le 1$のロジスティック曲線
上下の曲線は、2 項分布の標準偏差の幅
標準偏差は $p = 0.5$ のとき最大、0 と 1 付近で狭い
（等分散性が不成立）
誤差の分布は非対称の山型（図にはない）、
内側($p = 0.5$の方向) に裾を引く

縦軸は割合
（誤差は不等分散）



ロジット
変換

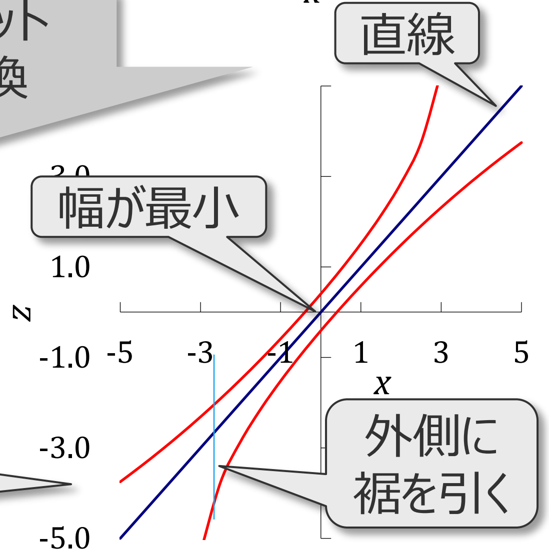
直線

右上の縦軸をロジット変換、中心の曲線は直線に変化
ロジットの分散は $p = 0.5$ で最小、0 と 1 付近で広い

$$V[logit] = \frac{1}{n\pi(1-\pi)}$$

（等分散性が不成立）
（3.3.9 §3.3 ）
誤差の分布は非対称の山型（図にはない）、
外側($p = 0.5$ と反対側) に裾を引く

縦軸はロジット
（誤差は不等分散）



誤差の構造

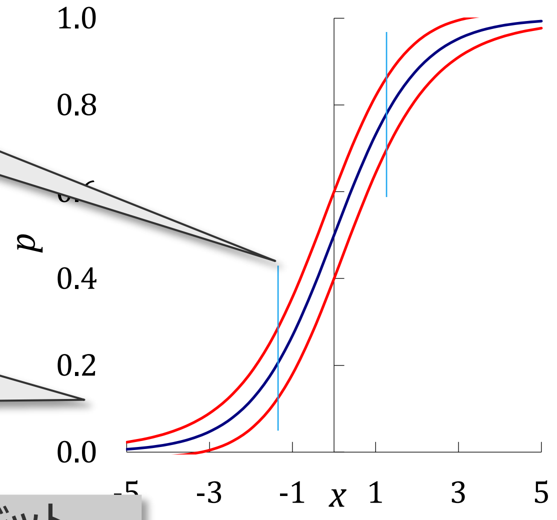
p.241

S 字曲線

$0 \leq p \leq 1$のロジスティック曲線
上下の曲線は、2 項分布の標準偏差の幅
標準偏差は $p = 0.5$ のとき最大、0 と 1 付近で狭い
（等分散性が不成立）
誤差の分布は非対称の山型（図にはない）、
内側($p = 0.5$の方向) に裾を引く

分布は非対称
内側に裾を引く

縦軸は割合
（誤差は不等分散）



ロジット
変換

直線

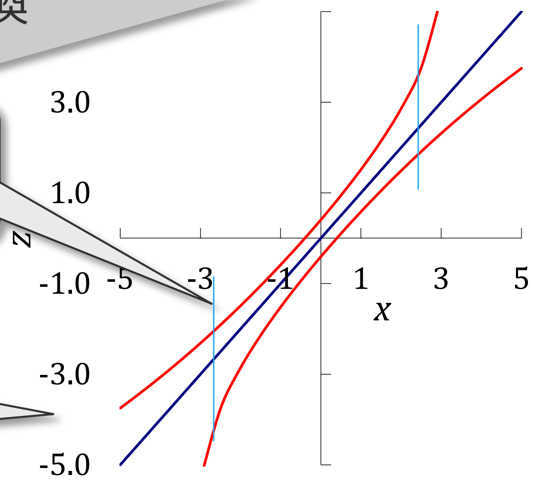
右上の縦軸をロジット変換、中心の曲線は直線に変化
ロジットの分散は $p = 0.5$ で最小、0 と 1 付近で広い

$$\mathrm{V}[logit] = \frac{1}{n\pi(1-\pi)}$$

（等分散性が不成立）
(3.3.9 §3.3)
誤差の分布は非対称の山型（図にはない）、
外側($p = 0.5$ と反対側) に裾を引く

分布は非対称
外側に裾を引く

縦軸はロジット
（誤差は不等分散）



誤差の構造

直線関係
いずれの x においても、誤差の幅は同じ
（等分散性が成立）
（第 1 部 §4.3 ）

右上の縦軸をロジット変換、中心の曲線は直線に変化
ロジットの分散は $p = 0.5$ で最小、0 と 1 付近で広い

$$V[logit] = \frac{1}{n\pi(1-\pi)}$$

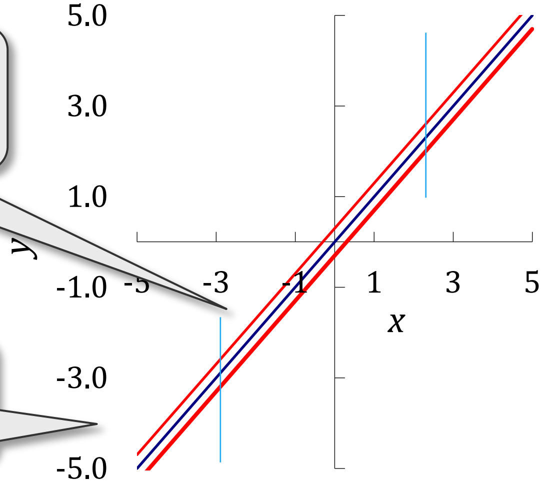
（等分散性が不成立）
（3.3.9 §3.3 ）

誤差の分布は非対称の山型（図にはない）、
外側($p = 0.5$ と反対側) に裾を引く

直線

どこでも同じ
正規分布

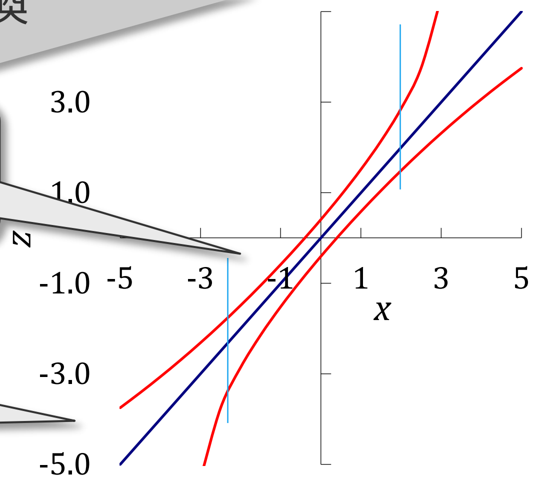
縦軸は量的変数
（誤差は等分散）



ロジット
変換

分布は非対称
外側に裾を引く

縦軸はロジット
（誤差は不等分散）



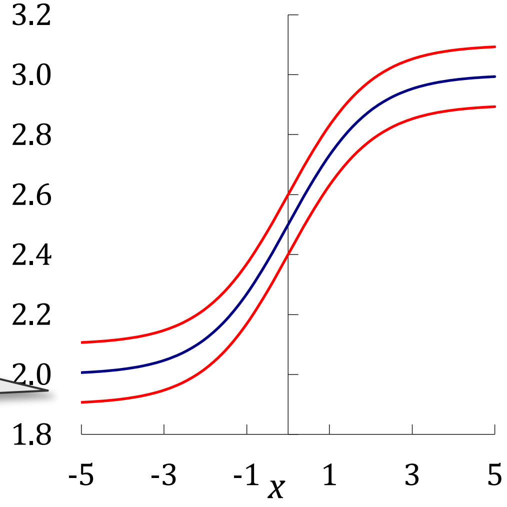
誤差の構造

p.241

誤差構造が違う

S字曲線

表示4.1.6

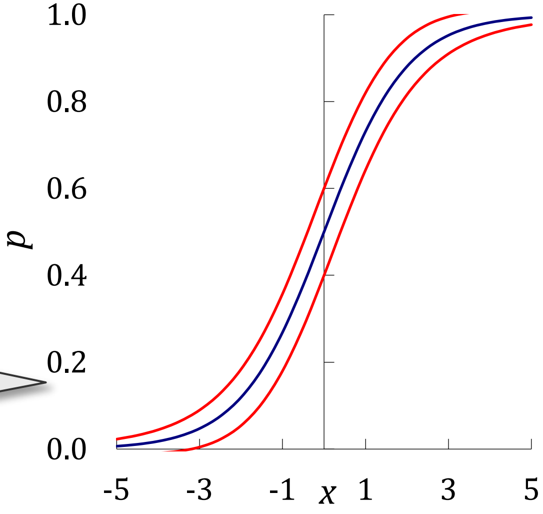


非線形最小２乗法を利用

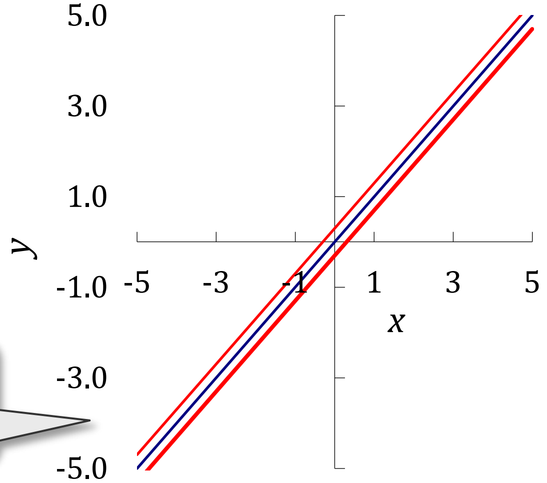
縦軸は量的変数
（誤差は等分散）

非線形最小２乗法ではなく
最尤法を利用

縦軸は割合
（誤差は不等分散）



直線

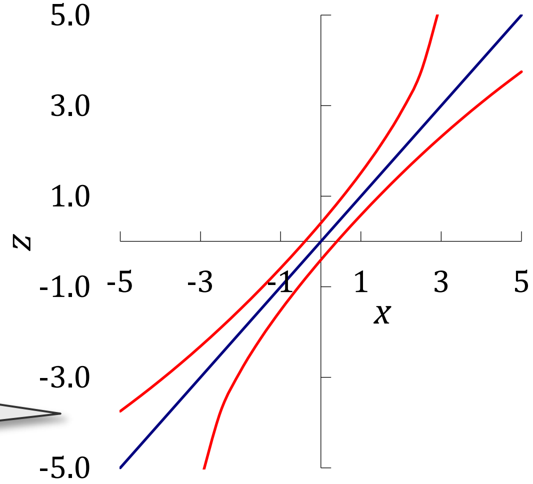


直線回帰分析を利用
（最小２乗法）

縦軸は量的変数
（誤差は等分散）

直線回帰分析ではなく
最尤法を利用

縦軸はロジット
（誤差は不等分散）



誤差の構造

- 最小２乗法と最尤法

最小２乗法

定量データで、等分散性が仮定できる場合

残差平方和　$S = \sum(y - \hat{y})^2$ を最小にするようにパラメータを決定

最尤法

割合データなど、等分散性が仮定できない場合

－対数尤度 の和を最小にするようにパラメータを決定

p.242

（4）最も簡単なロジスティック回帰分析

2 点のロジスティック回帰分析

最も簡単なロジスティック回帰分析

p.242

●事例

新薬開発の過程で得られたデータ（仮想データ、§3.3 表示3.3.1）

旧処方では 15 例中 4 例が有効、新処方では 10 例中 7 例が有効であった

2 × 2 分割表の解析（［二変量の関係］［モデルのあてはめ］）⇔ ロジスティック回帰分析

表示 4.1.7 割合の差の比較用データ

処方	有効	無効	合計	有効率	ロジット
旧処方	4	11	15	0.27	-1.0116
新処方	7	3	10	0.70	0.8473
合計	11	14	25	0.44	

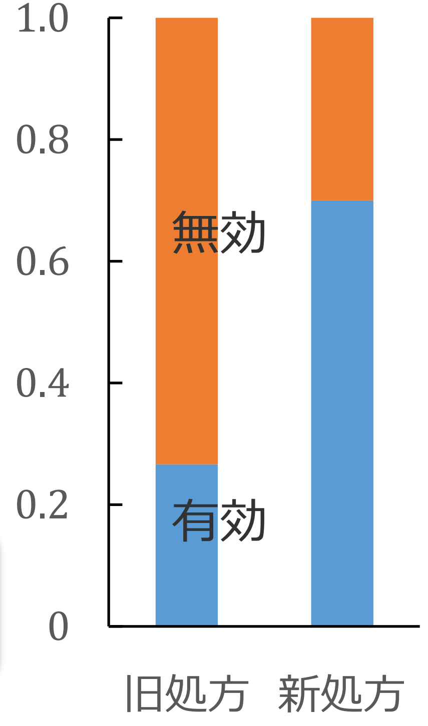
$$\ln\left(\frac{0.27}{1-0.27}\right) = \ln\left(\frac{4}{11}\right) = -1.0116$$

無効が基準

$$\ln\left(\frac{0.70}{1-0.70}\right) = \ln\left(\frac{7}{3}\right) = 0.8473$$

処方	x	効果	度数
旧処方	-1	有効	4
旧処方	-1	無効	11
新処方	1	有効	7
新処方	1	無効	3

質的変数をダミー変数（連続尺度）で表す



最も簡単なロジスティック回帰分析

- JMPファイルの読み込みと表示

  JMP ファイル「33-割合の差.jmp」を読み込み

  「効果」「x」「度数」を使う

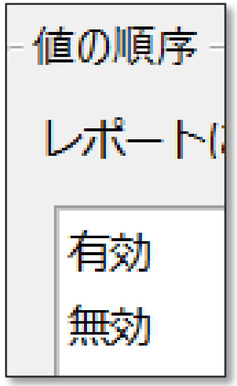
  「効果」の値の順序を設定（有効、無効）

- ［二変量の関係］の手順　　§3.3 参照

  ［Y, 目的変数］：「効果」

  ［X, 説明変数］：「x」（ダミー変数）

  ［度数］：「度数」　　↑連続尺度

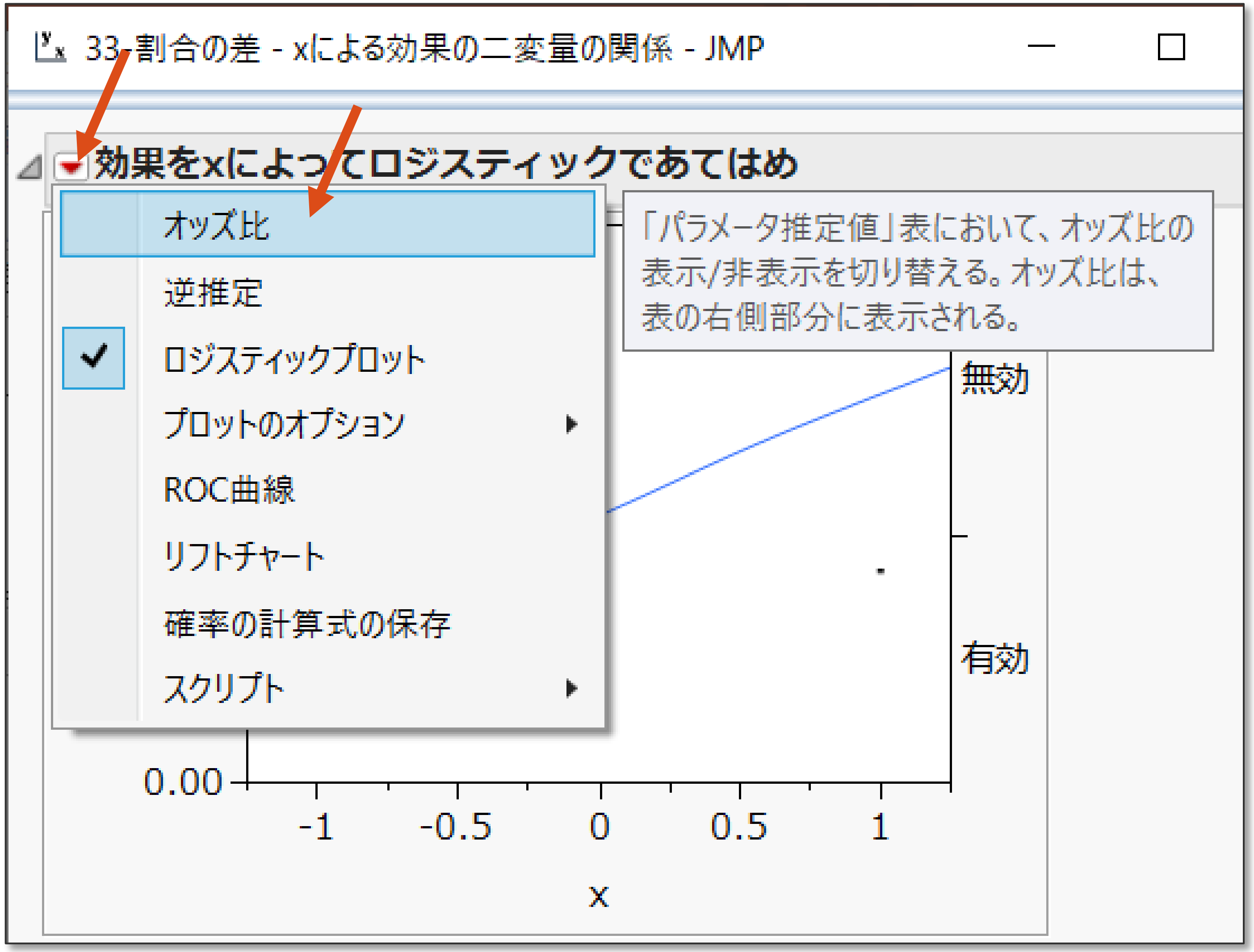


	処方o	x	効果o	度数	処方	効果
1	旧処方	-1	有効	4	旧処方	有効
2	旧処方	-1	無効	11	旧処方	無効
3	新処方	1	有効	7	新処方	有効
4	新処方	1	無効	3	新処方	無効

最も簡単なロジスティック回帰分析

p.242

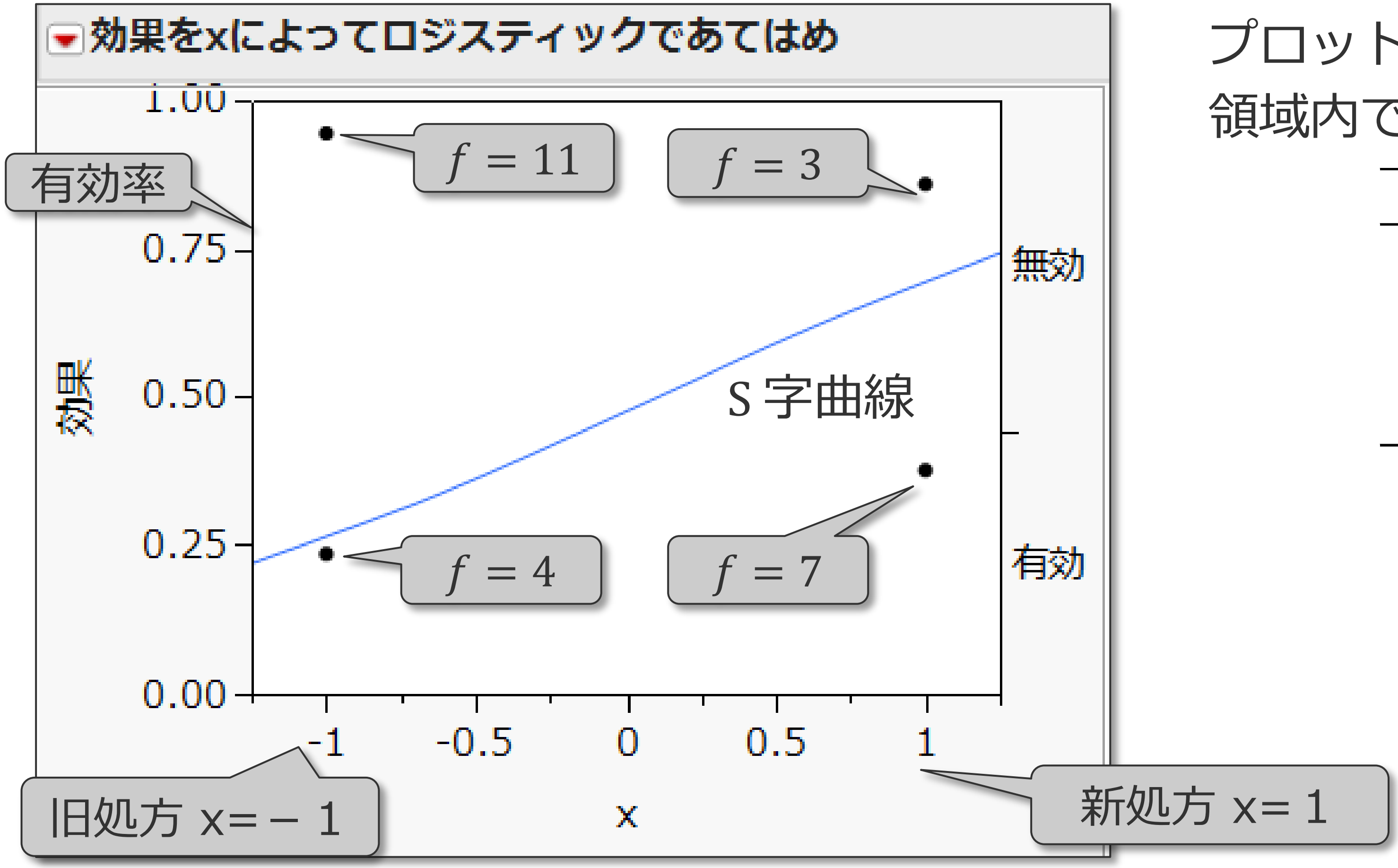
- ［二変量の関係］の手順
  - ▼オプション>［オッズ比］



最も簡単なロジスティック回帰分析

● ［二変量の関係］の結果

表示 4.1.8 JMP［二変量の関係］による解析



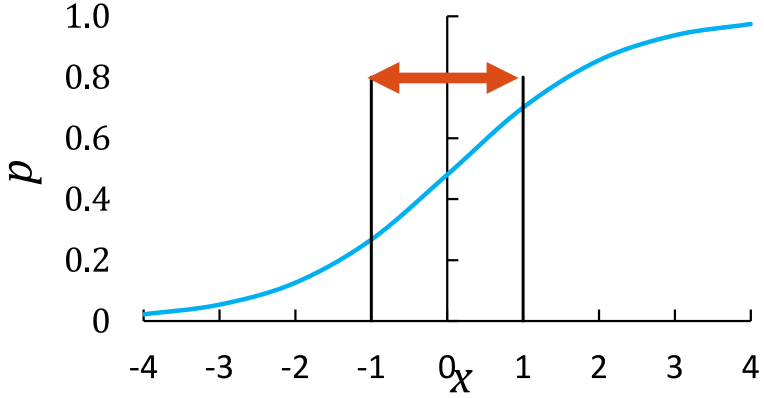
S 字曲線（x の範囲が狭いので直線に近い）

プロットは度数（f）の点数が重なっている

プロットは有効の領域と無効の領域に位置する

領域内での y の座標値は乱数で決まる

処方	x	効果	度数
旧処方	-1	有効	4
旧処方	-1	無効	11
新処方	1	有効	7
新処方	1	無効	3

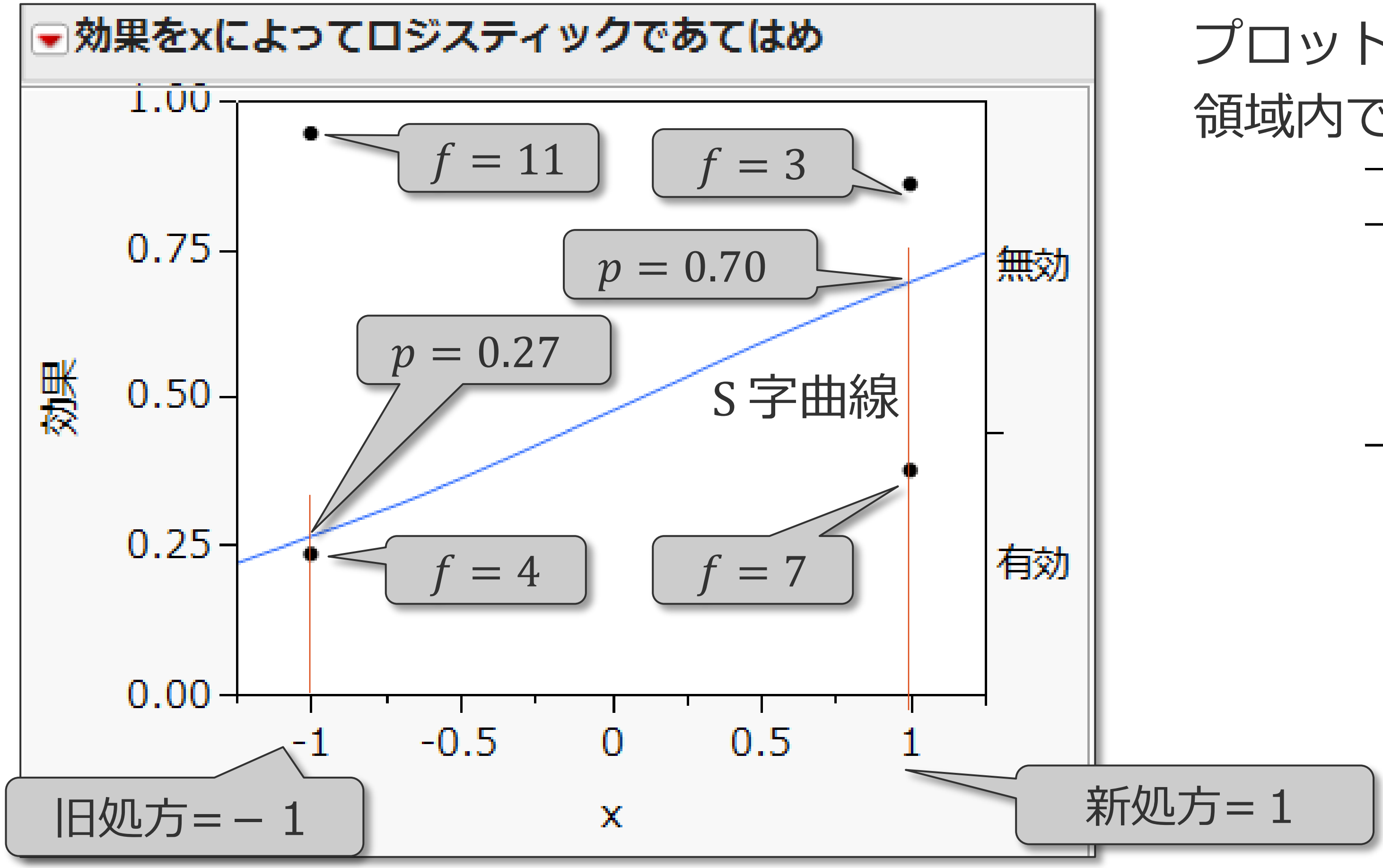


最も簡単なロジスティック回帰分析

p.243

● ［二変量の関係］の結果

表示 4.1.8 JMP［二変量の関係］による解析

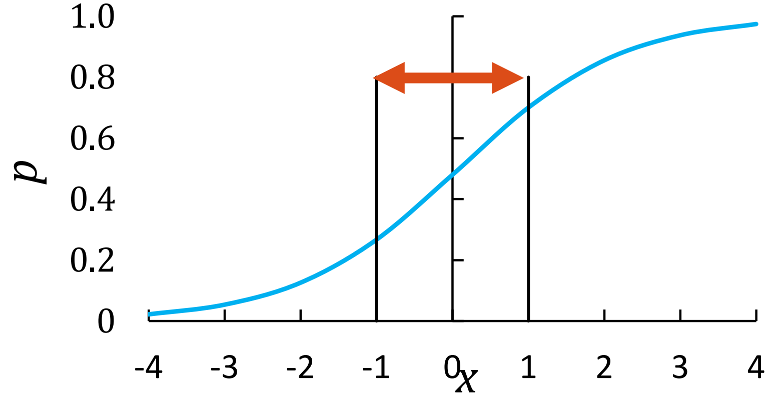


S 字曲線（x の範囲が狭いので直線に近い）
プロットは度数（f）の点数が重なっている
プロットは有効の領域と無効の領域に位置する
領域内での y の座標値は乱数で決まる

処方	x	効果	度数
旧処方	-1	有効	4
旧処方	-1	無効	11
新処方	1	有効	7
新処方	1	無効	3

4/15 = 0.27

7/10 = 0.70

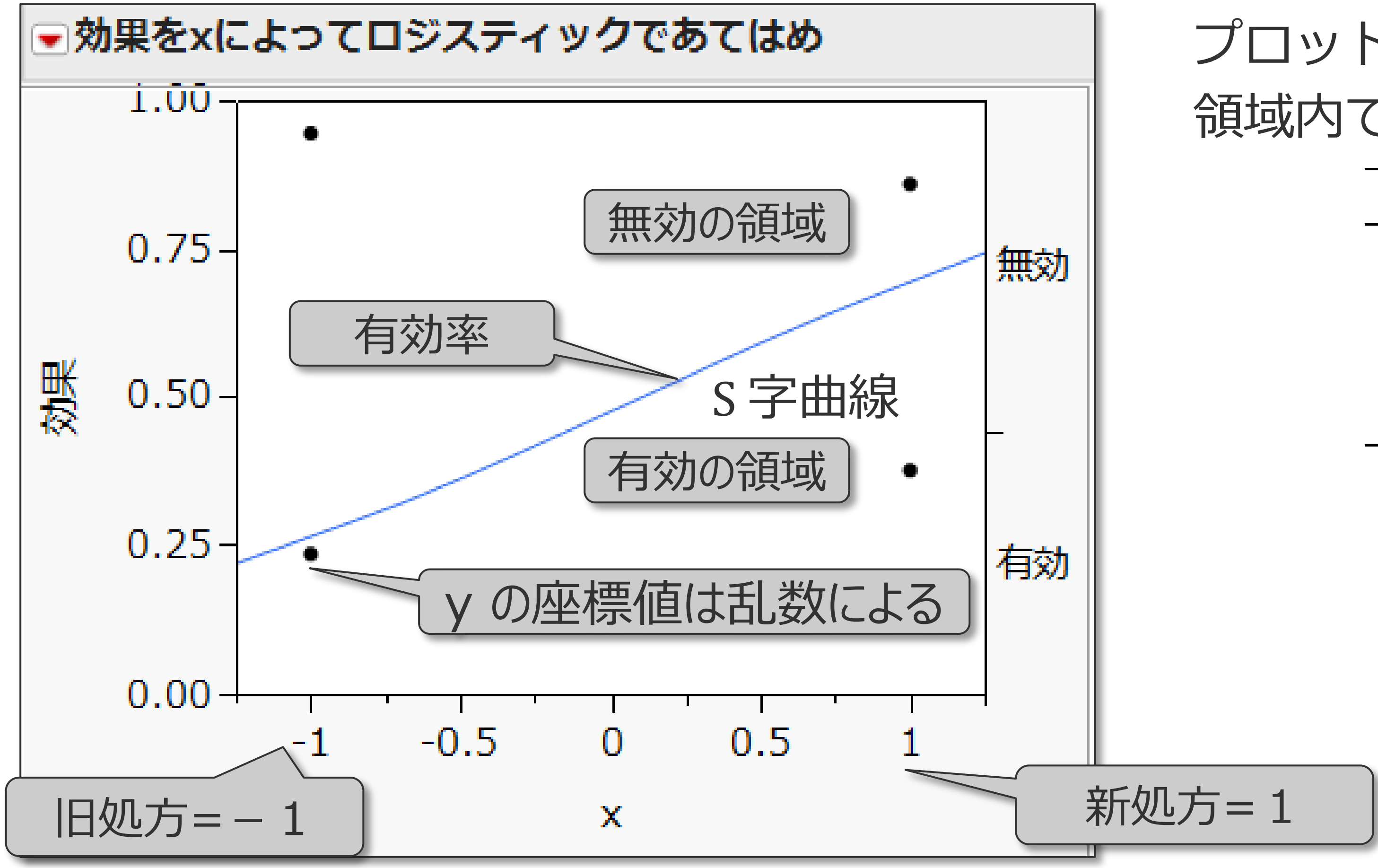


最も簡単なロジスティック回帰分析

p.243

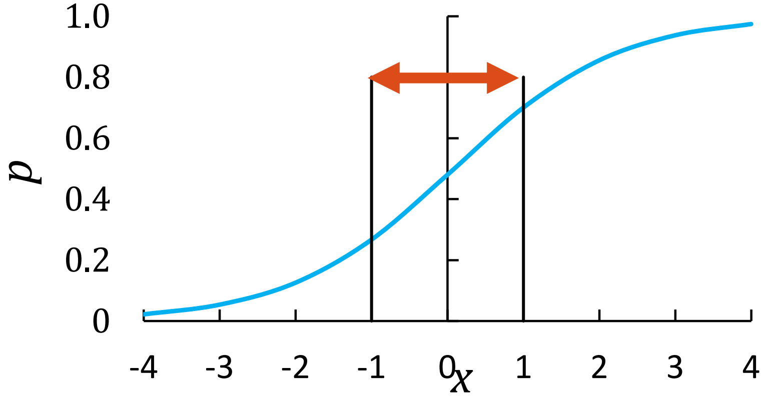
- ［二変量の関係］の結果

表示 4.1.8 JMP［二変量の関係］による解析



S 字曲線（x の範囲が狭いので直線に近い）
プロットは度数（f）の点数が重なっている
プロットは有効の領域と無効の領域に位置する
領域内での y の座標値は乱数で決まる

処方	x	効果	度数
旧処方	-1	有効	4
旧処方	-1	無効	11
新処方	1	有効	7
新処方	1	無効	3

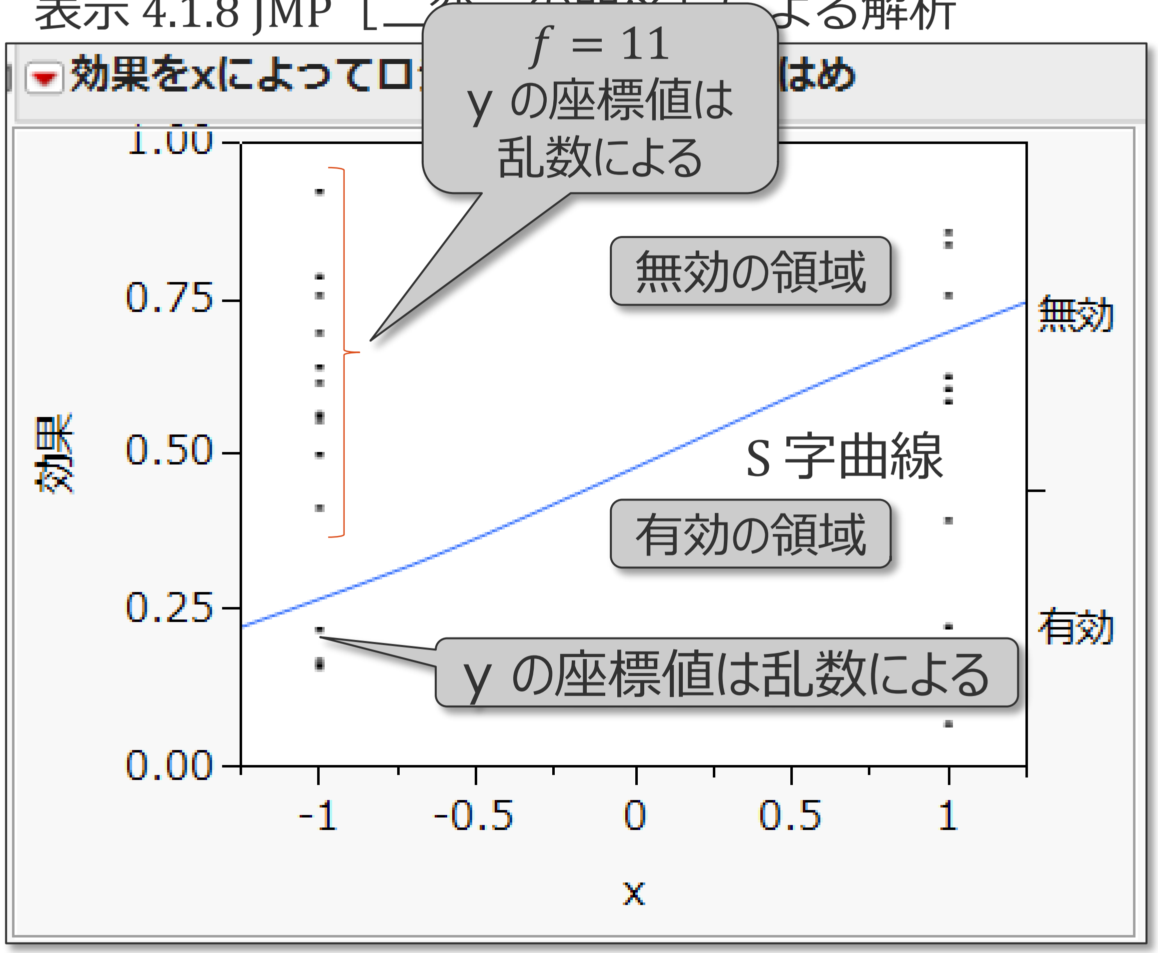


最も簡単なロジスティック回帰分析

p.243

- ［二変量の関係］の結果

表示 4.1.8 JMP［二変量の関係］による解析



処方	x	効果	度数
旧処方	-1	有効	4
旧処方	-1	無効	11
新処方	1	有効	7
新処方	1	無効	3

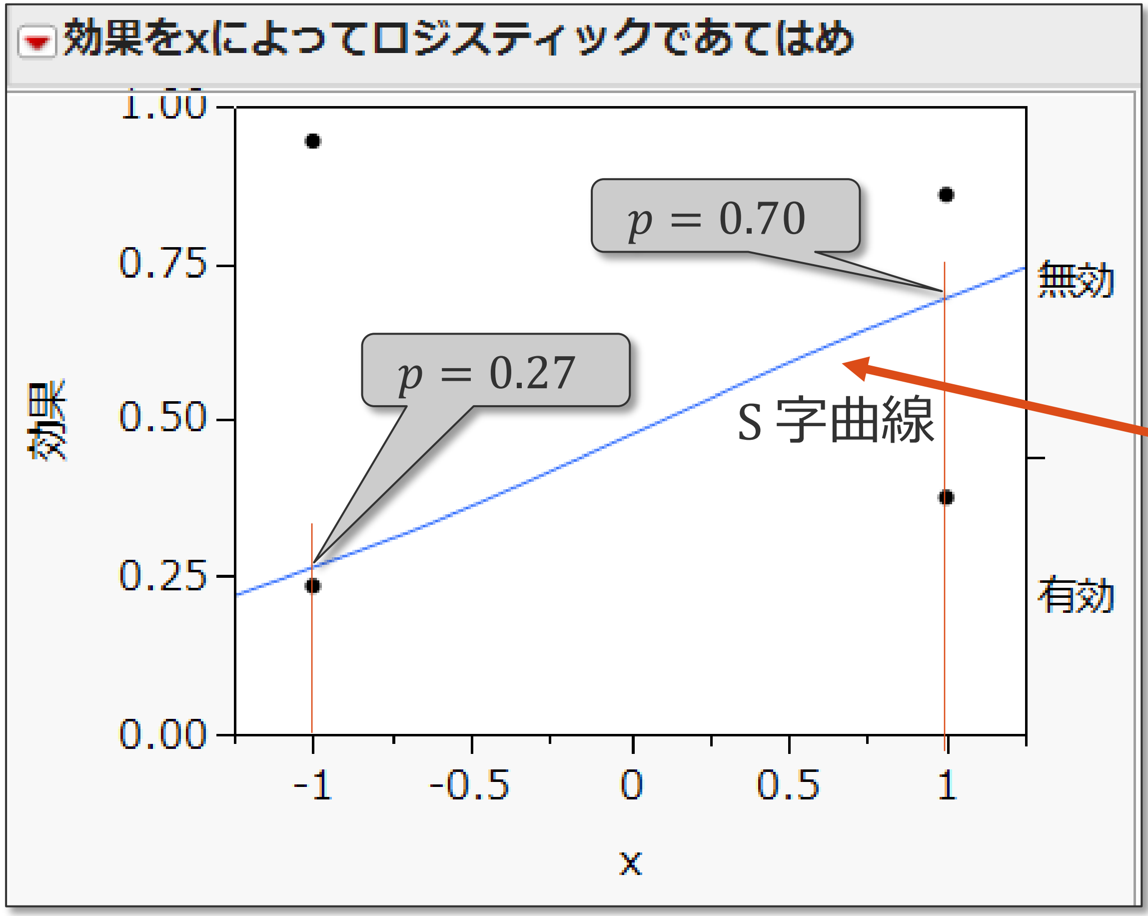
	x	効果
1	-1	有効
2	-1	有効
3	-1	有効
4	-1	有効
5	-1	無効
6	-1	無効
7	-1	無効
8	-1	無効
9	-1	無効
10	-1	無効
11	-1	無効

最も簡単なロジスティック回帰分析

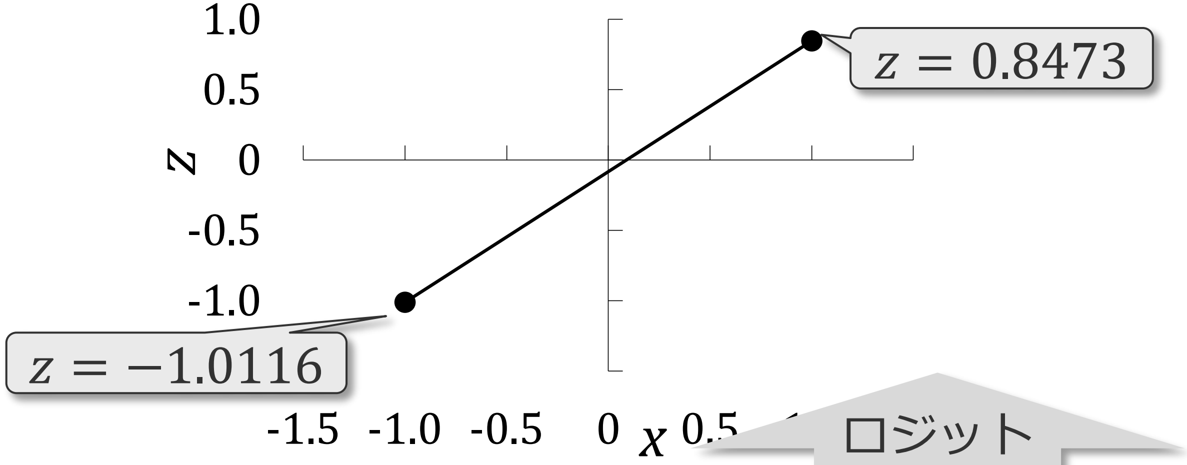
p.243

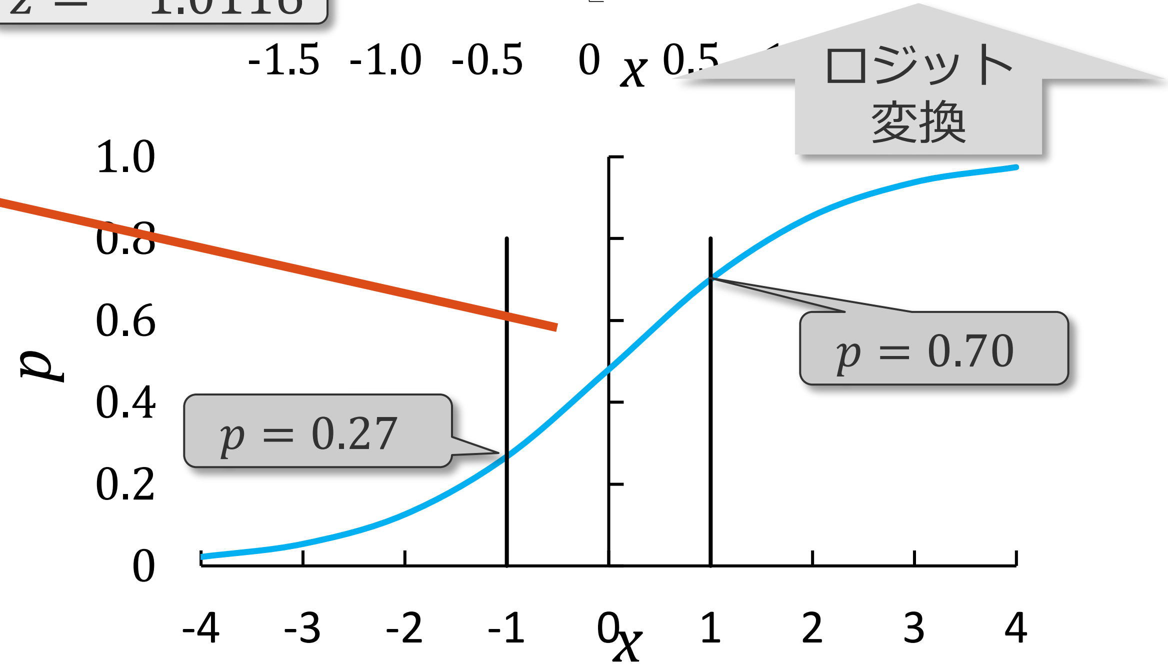
● ［二変量の関係］の結果

表示 4.1.8 JMP［二変量の関係］による解析



表示 4.1.9 縦軸にロジットをとったグラフ



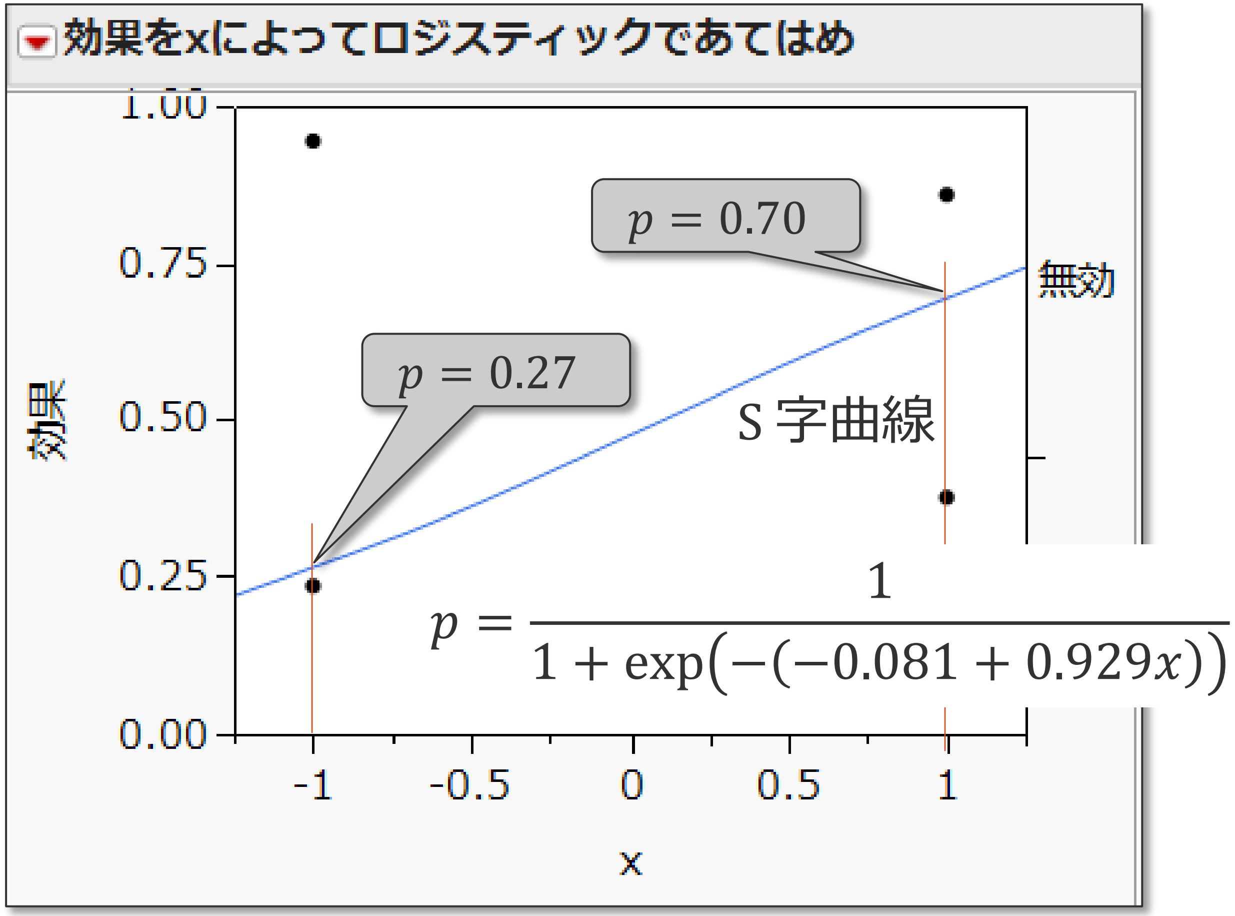


最も簡単なロジスティック回帰分析

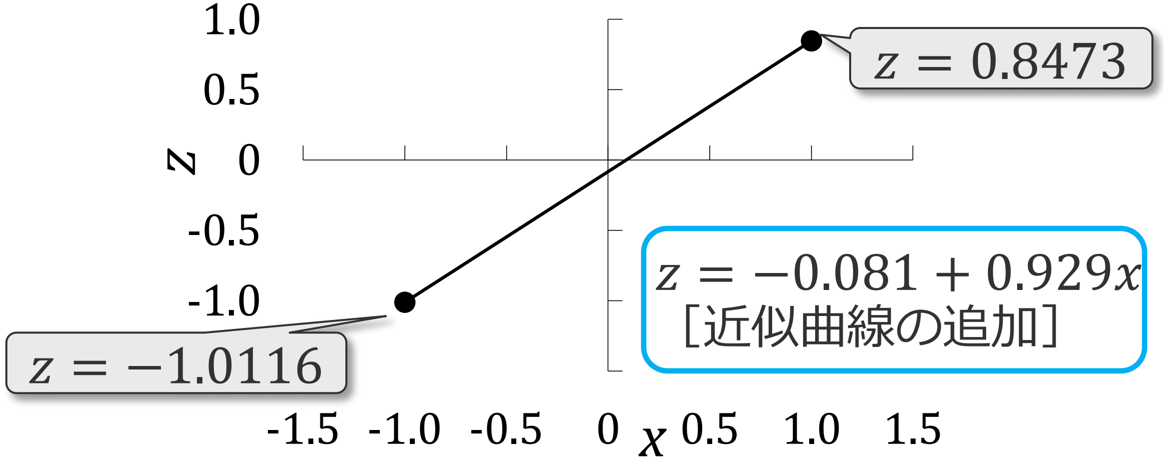
p.244

● ［二変量の関係］の結果

表示 4.1.8 JMP［二変量の関係］による解析



表示 4.1.9 縦軸にロジットをとったグラフ



表示 4.1.7

処方	有効	無効	合計	有効率	ロジット
旧処方	4	11	15	0.27	-1.0116
新処方	7	3	10	0.70	0.8473
合計	11	14	25	0.44	

右上のロジスティック回帰分析の結果が
左の散布図の近似直線と一致するのは
２×２分割表の場合だけ（一般には成立しない）

最も簡単なロジスティック回帰分析

p.244

- ［二変量の関係］の結果

$$p = \frac{1}{1 + \exp(-(-0.081 + 0.929x))}$$

$$z = -0.081 + 0.929x$$

Excel［近似曲線の追加］

表示3.3.12（§3.3 2 × 2 分割表を [モデルのあてはめ] で解析）

パラメータ推定値

項	推定値	標準誤差	カイ2乗	p値(Prob>ChiSq)
切片	-0.0821515	0.4519678	0.03	0.8558
処方[新処方]	0.92944939	0.4519678	4.23	0.0397*

推定値は次の対数オッズに対するものです： 有効/無効

表示 4.1.8
[二変量の関係]による解析
（ロジスティック回帰分析）

パラメータ推定値

項	推定値	標準誤差	カイ2乗	p値(Prob>ChiSq)	単位オッズ比	オッズ比
切片	-0.0821515	0.4519678	0.03	0.8558	.	.
x	0.92944939	0.4519678	4.23	0.0397*	2.53311403	6.41666667

推定値は次の対数オッズに対するものです： 有効/無効

最も簡単なロジスティック回帰分析

p.244

● ［二変量の関係］の結果

尤度比検定のカイ２乗値.

表示 4.1.8（ロジスティック回帰分析）

モデル全体の検定

モデル	(-1)*対数尤度	自由度	カイ2乗	p値(Prob>ChiSq)
差	2.340874	1	4.681749	0.0305*
完全	14.807371			
縮小	17.148245			

R2乗(U)	0.1365
AICc	34.1602
BIC	36.0525
オブザベーション(または重みの合計)	25

表示 3.3.5 （§3.3、２×２分割表［二変量の関係］）

検定

N	自由度	(-1)*対数尤度	R2乗(U)
25	1	2.3408744	0.1365

検定	カイ2乗	p値(Prob>ChiSq)
尤度比	4.682	0.0305*
Pearson	4.573	0.0325*

表示 3.3.12
（§3.3、２×２分割表［モデルのあてはめ］）

モデル全体の検定

モデル	(-1)*対数尤度	自由度	カイ2乗	p値(Prob>ChiSq)
差	2.340874	1	4.681749	0.0305*
完全	14.807371			
縮小	17.148245			

効果の尤度比検定

要因	パラメータ数	自由度	尤度比カイ2乗	p値(Prob>ChiSq)
処方	1	1	4.68174883	0.0305*

最も簡単なロジスティック回帰分析

p.244

- ［二変量の関係］：オッズ比

2 × 2 分割表のオッズ比（新処方／旧処方）

$$OR = (7 \times 11)/(4 \times 3) = 6.417$$

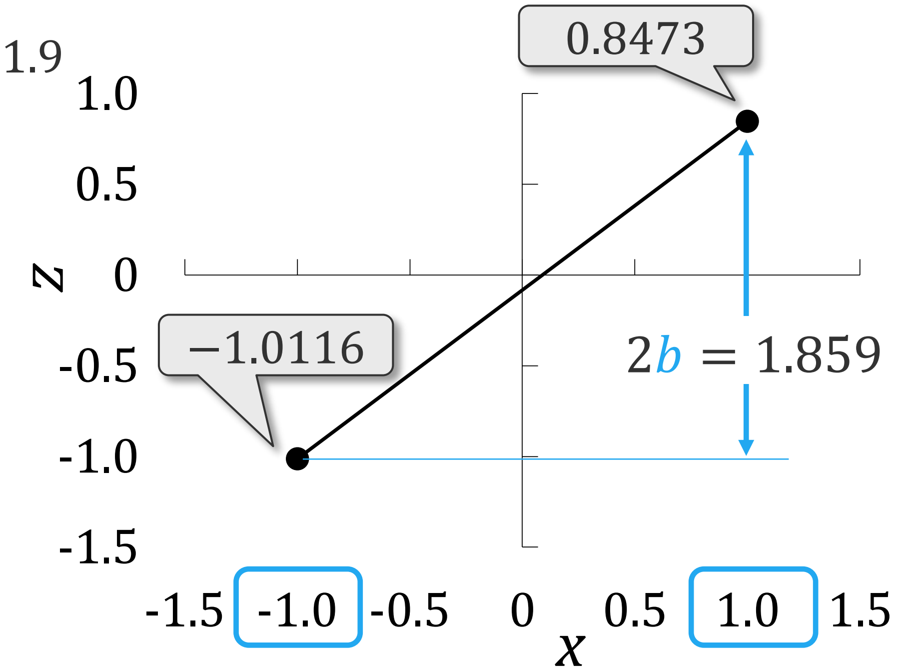
新処方と旧処方のロジットの差（対数オッズ比）は $2b$

$$\ln OR = 2b \rightarrow OR = \exp(2b) = \exp(1.8589) = 6.417$$

表示 4.1.7

処方	有効	無効	合計	有効率	ロジット
旧処方	4	11	15	0.27	-1.0116
新処方	7	3	10	0.70	0.8473
合計	11	14	25	0.44	

表示 4.1.9



表示 4.1.8

パラメータ推定値

項	推定値	標準誤差	カイ2乗	p値(Prob>ChiSq)	単位オッズ比	オッズ比
切片	-0.0821515	0.4519678	0.03	0.8558	.	.
x	0.92944939	0.4519678	4.23	0.0397*	2.53311403	6.41666667

推定値は次の対数オッズに対するものです： 有効/無効

最も簡単なロジスティック回帰分析

p.244

- ［二変量の関係］：単位オッズ比

  2 × 2 分割表のオッズ比（新処方／旧処方）

  $OR = (7 \times 11)/(4 \times 3) = 6.417$

  新処方と旧処方のロジットの差（対数オッズ比）は $2b$

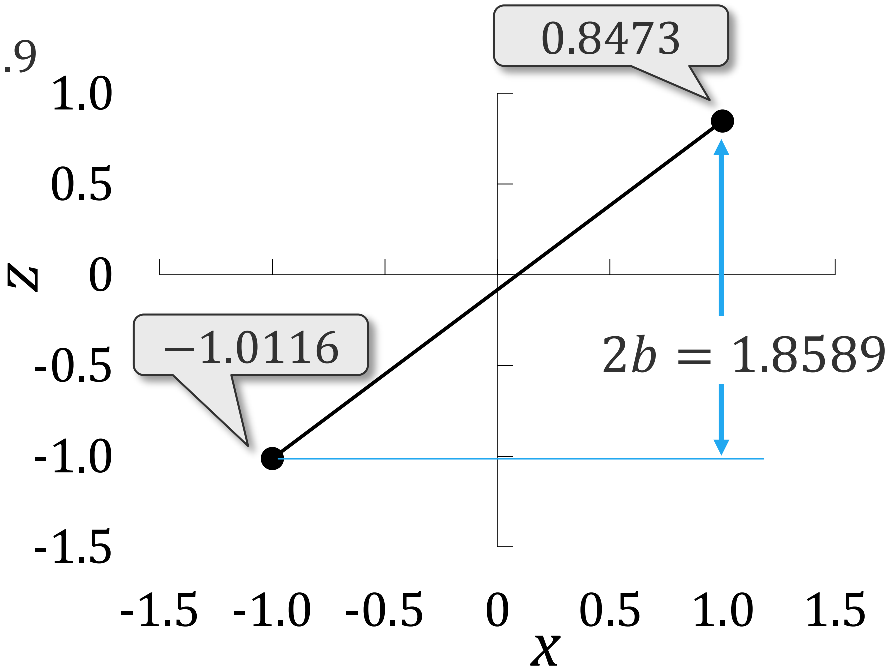
  $\ln OR = 2b \;\rightarrow\; OR = \exp(2b) = \exp(1.8589) = 6.417$

  単位オッズ比は x が 1 変化したときのオッズ比

  $\ln OR = b \;\rightarrow\; OR = \exp(b) = \exp(0.9294) = 2.533$

  ダミー変数の取り方で、値が変わる

表示 4.1.9



表示 4.1.8

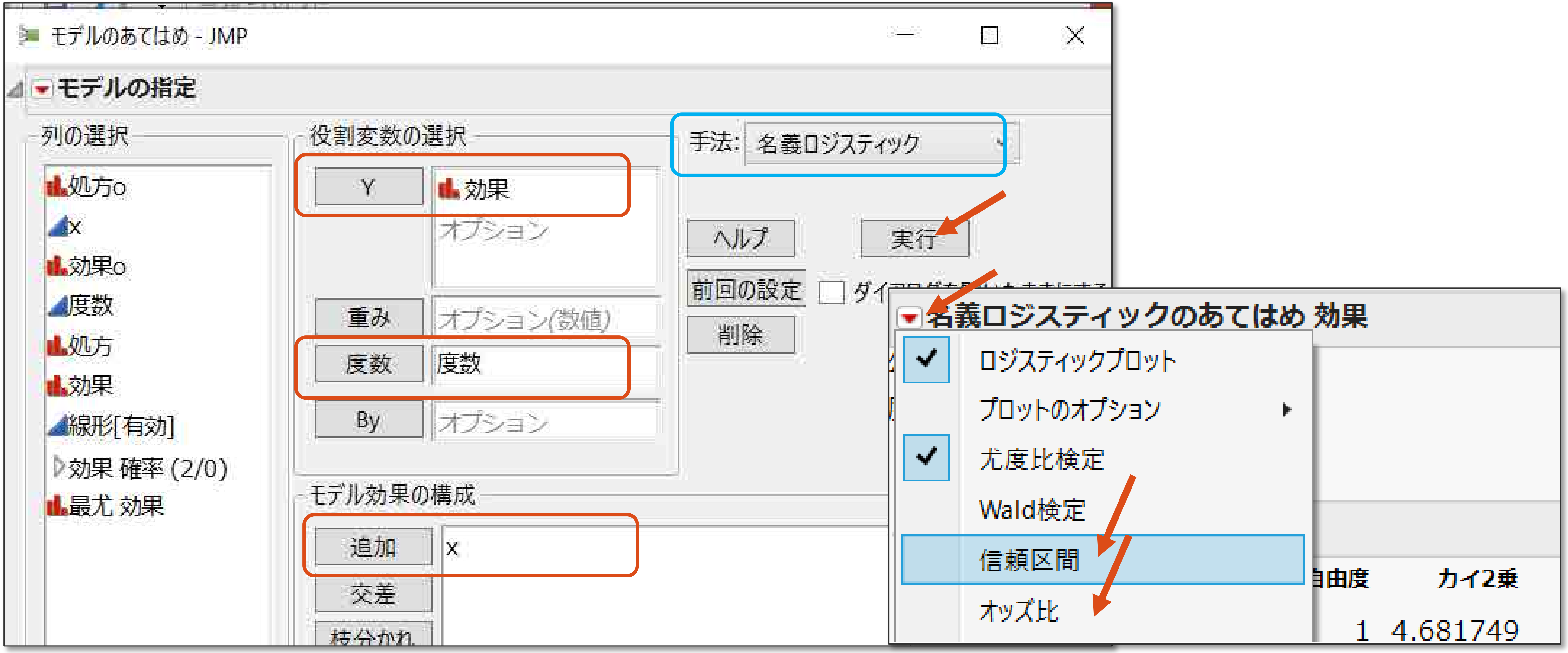
パラメータ推定値

項	推定値	標準誤差	カイ2乗	p値(Prob>ChiSq)	単位オッズ比	オッズ比
切片	-0.0821515	0.4519678	0.03	0.8558	.	.
x	0.92944939	0.4519678	4.23	0.0397*	2.53311403	6.41666667

推定値は次の対数オッズに対するものです： 有効/無効

最も簡単なロジスティック回帰分析

●JMP［モデルのあてはめ］の手順



最も簡単なロジスティック回帰分析

p.245

- ［モデルのあてはめ］の結果

表示 4.1.10 （改変）

名義ロジスティックのあてはめ 効果

勾配で収束しました, 4回の反復

度数: 度数

反復回数

モデル全体の検定

モデル	(-1)*対数尤度	自由度	カイ2乗	p値(Prob>ChiSq)
差	2.340874	1	4.681749	0.0305*
完全	14.807371			
縮小	17.148245			

R2乗(U)	0.1365
AICc	34.1602
BIC	36.0525
オブザベーション(または重みの合計)	25

［二変量の関係］と同じ結果

最も簡単なロジスティック回帰分析

p.245

- ［モデルのあてはめ］：パラメータ推定値

［二変量の関係］と同じ結果

表示 4.1.10

パラメータ推定値

項	推定値	標準誤差	カイ2乗	p値(Prob>ChiSq)	下側95%	上側95%
切片	-0.0821515	0.4519678	0.03	0.8558	-0.9758179	0.8445593
x	0.92944939	0.4519678	4.23	0.0397*	0.08489983	1.89037498

推定値は次の対数オッズに対するものです： 有効/無効

推定値の共分散

効果の尤度比検定

要因	パラメータ数	自由度	尤度比カイ2乗	p値(Prob>ChiSq)
x	1	1	4.68174883	0.0305*

最も簡単なロジスティック回帰分析

p.245

- ［モデルのあてはめ］：オッズ比、単位オッズ比

表示 4.1.10

オッズの計算

有効と無効のどちらを分母にするかは任意

値の順序　　有効、無効

オッズ　　　有効対無効（有効／無効）

オッズ比の計算

新処方と旧処方の

どちらを分母にするかは任意

旧処方：－１、新処方：１

オッズ比　　新処方／旧処方

$$OR = (7 \times 11)/(4 \times 3) = 6.417$$

オッズ比

効果: 有効対無効のオッズ比に対して

オッズ比の検定と信頼区間は、尤度比に基づいて計算されています。

分母と分子を
逆にしたオッズ比

単位オッズ比

連続変数が1単位だけ変化した場合

項	オッズ比	下側95%	上側95%	逆数
X	2.533114	1.088608	6.621851	0.394771

範囲オッズ比

連続変数が範囲全体で変化した場合

項	オッズ比	下側95%	上側95%	逆数
X	6.416667	1.185067	43.84891	0.1558442

最も簡単なロジスティック回帰分析

p.245

- ［モデルのあてはめ］：オッズ比、単位オッズ比

表示 4.1.10

単位オッズ比
　x が 1 単位だけ変化した場合のオッズ比
　このオッズ比が 1 から離れるほど
　　x の目的変数への影響度は 大きい

範囲オッズ比
　x の範囲全体（最小値～最大値）で
　　変化した場合のオッズ比
　実験での x の取り方で値は変る

オッズ比

効果: 有効対無効のオッズ比に対して

オッズ比の検定と信頼区間は、尤度比に基づいて計算されています。

分母と分子を
逆にしたオッズ比

単位オッズ比

連続変数が1単位だけ変化した場合

項	オッズ比	下側95%	上側95%	逆数
x	2.533114	1.088608	6.621851	0.394771

範囲オッズ比

連続変数が範囲全体で変化した場合

項	オッズ比	下側95%	上側95%	逆数
x	6.416667	1.185067	43.84891	0.1558442

まとめ

p.245

- ロジスティック回帰分析

  カテゴリカルな応答に対して使われるモデル

  （§1.4 ロジスティック曲線のあてはめの拡張ではない、誤差の構造が異なる）

  目的変数を割合 p とし、範囲を 0～1 に制限

  x と p の間にロジスティック曲線を仮定

  応答が 2 水準で、応答の分布が 2 項分布に従うと仮定

  （正規分布ではない、等分散性が成立しない）

  推定方法には「最尤法」が用いられる

- JMP によるロジスティック回帰分析

  [二変量の関係]

  [モデルのあてはめ]

  [非線形回帰]

- 作成　片瀬雅彦
- 監修　松本一彦、長谷文雄
- 作成時期　2020年9月16日
- 改訂　2021年3月21日、2022年11月20日