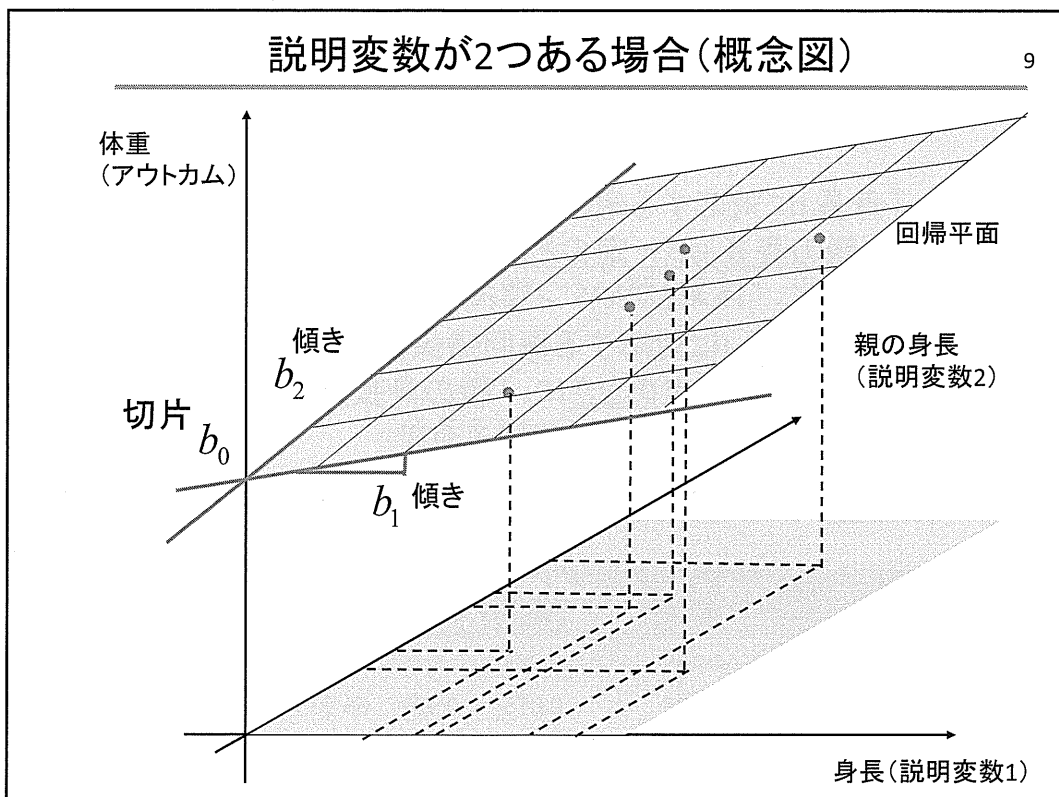


説明変数が2つある場合(概念図)

9



重回帰分析への拡張

10

「体重」を、「身長」と「親の身長」とで予測する場合

$$y_i = b_0 + b_1 x_{i1} + b_2 x_{i2} + e_i$$

x_{i1} iさんの1番目の説明変数=「身長」

x_{i2} iさんの2番目の説明変数=「親の身長」

y_i iさんのアウトカム=「体重」

b_0 切片(定数項)

b_1 「身長」に対応する(非標準化)偏回帰係数

b_2 「親の身長」に対応する(非標準化)偏回帰係数

} 推定すべきもの
(パラメタ)

重回帰分析の結果の解釈

11

b_1 「身長」に対応する偏回帰係数
=「親の身長」が一定のときに、「身長」が1だけ
増えると、「体重」がいくつ増えるか

b_2 「親の身長」に対応する偏回帰係数
=「身長」が一定のときに、「親の身長」が1だけ
増えると、「体重」がいくつ増えるか

b_1, b_2 は、交絡因子を統制したときの影響
力の大きさを表している

モデル	非標準化 係数	SE(非標準 化係数)	標準化 係数	t値	有意確率
b_0 切片	-0.469	0.083		-5.637	.000
b_1 要因A	0.997	0.015	0.992	66.935	.000
b_2 要因B	0.005	0.001	0.045	3.052	.003

交絡因子への対処法

- 多変量解析
 - 重回帰分析
 - ロジスティック回帰分析 ←
 - Poisson 回帰分析
 - Cox回帰分析

アウトカムが二値変数の場合

13

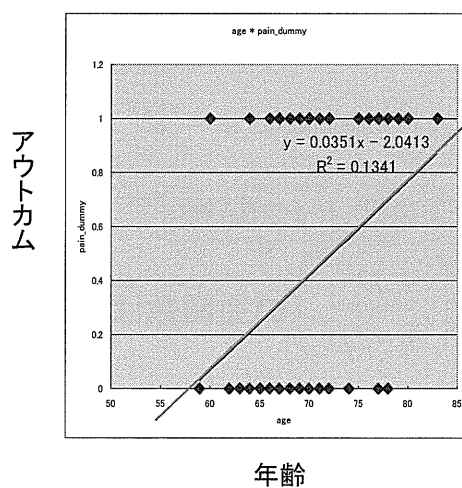
年齢	性別	喫煙	飲酒	投薬	症状
63	男性	1	4	1	1
48	男性	1	3	0	1
62	女性	0	2	0	1
45	女性	0	0	0	0
53	女性	0	0	0	1
58	男性	0	1	1	0
⋮	⋮	⋮	⋮	⋮	⋮



目的変数(アウトカム)が「あり/なし」
のような変数(二値変数)

回帰分析ではうまくいかない

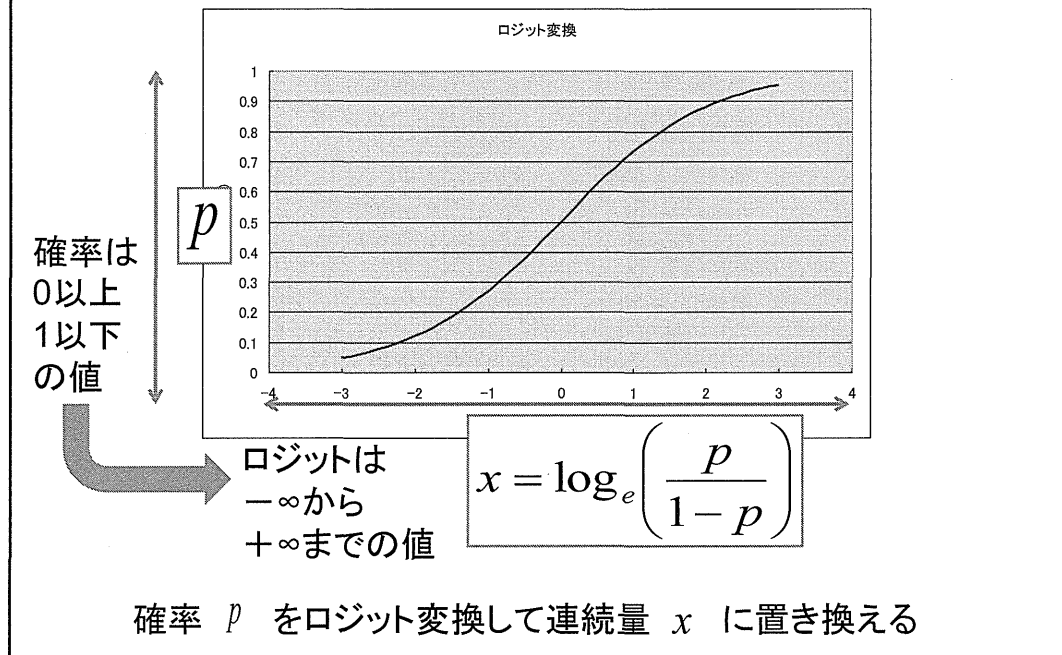
14



そもそも、このようなデータの誤差は「正規分布」しない！
→「生存」「死亡」のような2値データの場合は「二項分布」に従う

ロジット変換(ロジスティック変換)による モデル式の変更

15



ロジット関数の逆関数=ロジスティック関数

16

確率 p に関する予測のためには...

x がわかっている状況で, p の値を知りたい!

一般に $y = \log_e(x)$ の逆関数は $x = e^y$

ロジスティック関数

$$x = \log_e \left(\frac{p}{1-p} \right) \text{ の逆関数は } p = \frac{1}{1+e^x}$$

$$p = \frac{1}{1+e^{b_0+b_1x_1}} \text{ を } p = \frac{1}{1+\exp(b_0+b_1x_1)}$$

と表記することも

ロジスティック回帰分析のモデル式

アウトカムが比率(確率)で表される場合に、
重回帰分析のモデル式

$$y_i = b_0 + b_1x_{i1} + b_2x_{i2} + b_3x_{i3} + \dots + b_jx_{ij} + e_i$$

の y_i の代わりに p_i (確率の値、=0以上1以下)を考え、

$$\log\left(\frac{p_i}{1-p_i}\right) = b_0 + b_1x_{i1} + b_2x_{i2} + b_3x_{i3} + \dots + b_jx_{ij} + e_i$$

をモデル式とする

→ 左辺は連続量(=データが「1」である確率に対応する値)

ロジスティック回帰分析のモデル式(続き)

回帰式

$$\log\left(\frac{\hat{p}_i}{1-\hat{p}_i}\right) = b_0 + b_1x_{i1} + b_2x_{i2} + b_3x_{i3} + \dots + b_jx_{ij}$$

を整理すると

$$\hat{p}_i = \frac{1}{1 + \exp\left(-\left(b_0 + b_1x_{i1} + b_2x_{i2} + b_3x_{i3} + \dots + b_jx_{ij}\right)\right)}$$

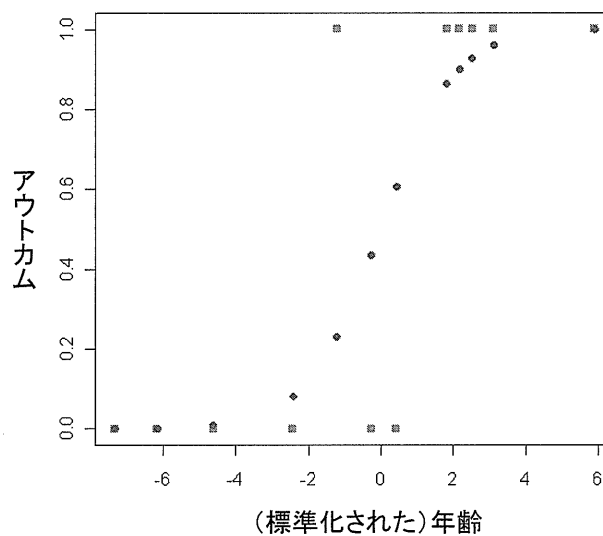
上式を回帰式とすれば、手持ちのデータ p_i x_{ij} より

b_j を推定し、説明変数ごとに影響力の大きさを推定できる

推定結果の図

赤点: データ

青点: 回帰によって
予測される
生存確率



モデルの適合度とステップワイズ法

Hosmer-Lemeshow 検定

$$\chi^2 = \sum_{k=1}^K \frac{(\text{群内実測値} - \text{群内での予測の平均})^2}{\text{群内での予測の分散}}$$

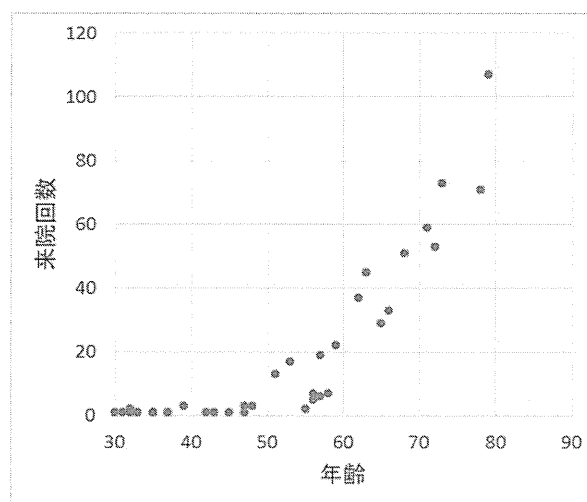
ステップワイズ法

説明変数を出し入れし、モデル適合度指標(デビアンس)の変化から「もっとも説明力の高いモデル」を探索的に見出す分析手法

交絡因子への対処法

- 多変量解析
 - 重回帰分析
 - ロジスティック回帰分析
 - Poisson 回帰分析 ◀
 - Cox回帰分析

カウントデータのための分析: 来院回数データ 22



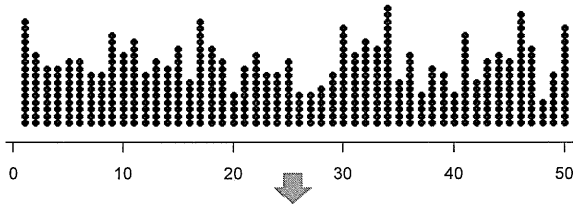
- ある期間内に, ある病棟に来院した患者の年齢と, 来院回数をプロット
 - 一定期間内に何回イベント(来院)が発生したかを数え上げたデータを「カウントデータ」と呼ぶ

カウントデータの特徴

- 「1回」など, 少数の人が多い
- 「1回」「2回」というように, 離散的な値をとる

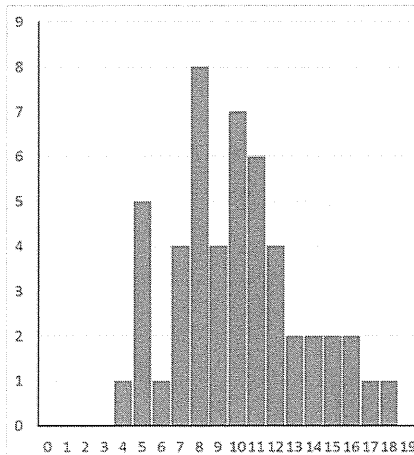
カウントデータのための分析: Poisson回帰 23

1日に平均10回の「来院」
 が起こる病棟において、
 50日の来院発生回数を
 プロット



来院頻度は、正規分布し
 そうにない・・・

発生回数	頻度
0	
1	
2	
3	
4	1
5	5
6	1
7	4
8	8
9	4
10	7
11	6
12	4
13	2
14	2
15	2
16	2
17	1
18	1
19	

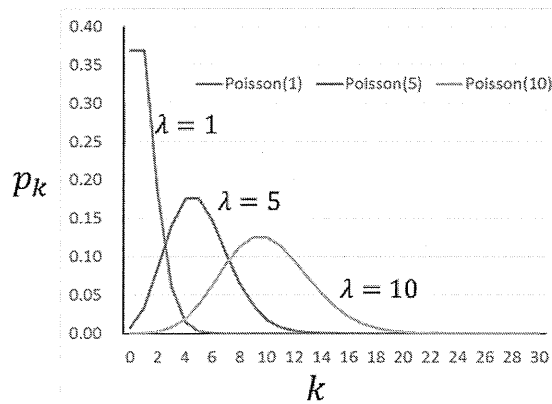


Poisson分布 24

1単位時間(1時間, 1日, 1月...)に平均 λ 回の事象がランダムに生起するとき、
 この事象が1単位時間に k 回発生する確率は、

$$p_k = \frac{\lambda^k - e^{-\lambda}}{k!} \quad \text{となる.}$$

この分布を「Poisson (ポアソン) 分布」と呼ぶ。



1単位時間に発生する
 回数の期待値は、 λ .
 $E[k] = \lambda$

1単位時間に発生する
 回数の分散は、 λ .
 $V[k] = \lambda$

平均が大きければ、
 分散も大きい。

「1単位時間の事象発生数」をアウトカムとした回帰分析

回帰式

$$\log(\lambda_i) = b_0 + b_1x_{i1} + b_2x_{i2} + b_3x_{i3} + \dots + b_jx_{ij}$$

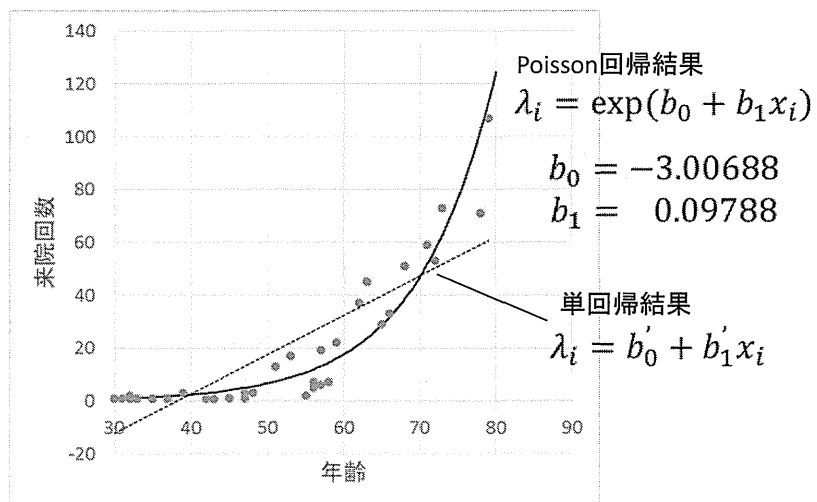
予測結果が0を下回ることを防ぐためにlogをとる

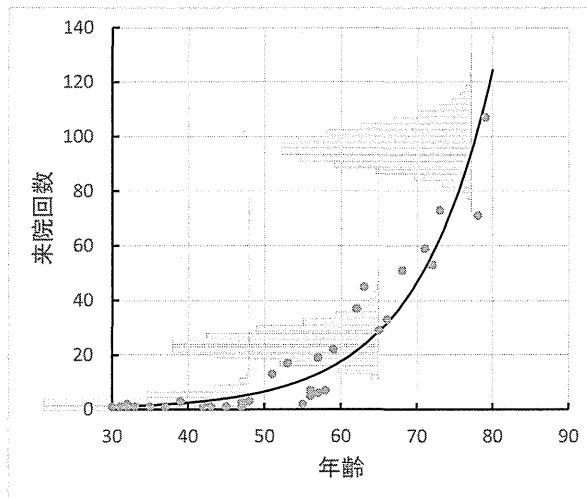
i 番目の個体の誤差が「平均 λ_i のPoisson分布に従う」という仮定をおく

$$\lambda_i = \exp(b_0 + b_1x_{i1} + b_2x_{i2} + b_3x_{i3} + \dots + b_jx_{ij})$$

上式を回帰式とすれば、手持ちのデータ λ_i x_{ij} より

b_j を推定し、説明変数ごとに影響力の大きさを推定できる





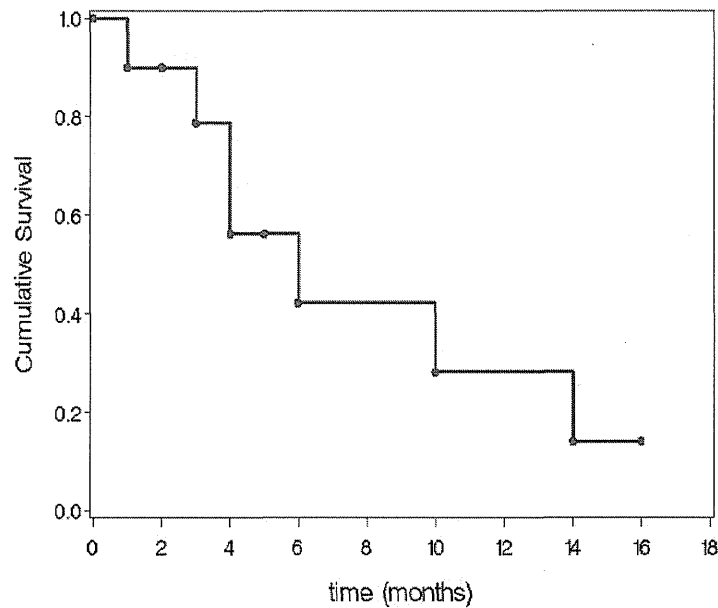
誤差分散としてPoisson分布を仮定したほうがよい

交絡因子への対処法

- 多変量解析
 - 重回帰分析
 - ロジスティック回帰分析
 - Poisson 回帰分析
 - Cox回帰分析 ◀

生存曲線

29



COX比例ハザードモデルの意義

30

「生存時間」という概念を加味した形での回帰分析

- 生存データの分析で、複数の「要因」のデータが各患者からとれている時、アウトカムと関連性が高い要因がどれなのかを推定したい

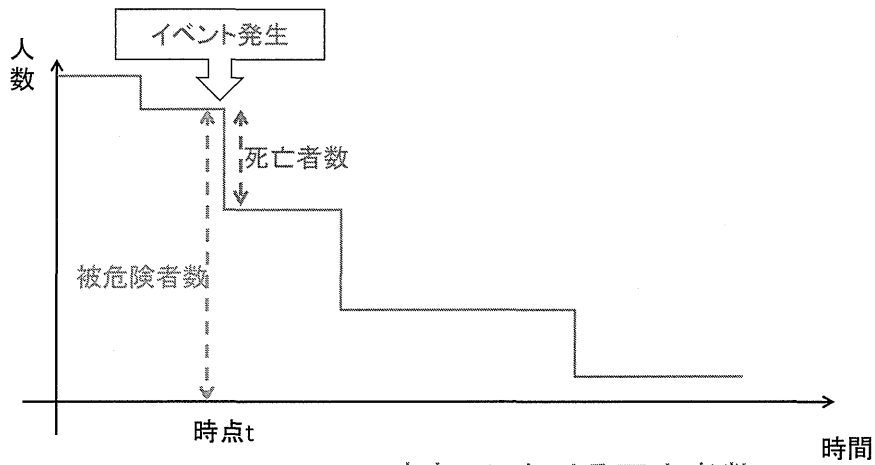
→この場合の「アウトカム」とは、何か？

- 累積生存曲線の傾きを大きくしている要因はどれかを知りたい(なぜなら、傾きが大きい＝そこで生存率が大きく変化しているのだ)

つまり・・・「アウトカム」＝「累積生存曲線の傾きの大きさ」ということを、モデルで表現したい

「生存率の変化」を定義する

31



$$\text{時点 } t \text{ におけるハザード} = \frac{\text{時点 } t \text{ における死亡者数}}{\text{時点 } t \text{ における被危険者数}}$$

t から次のイベントが起きた時点 (t+1) の間における生存者数の変化の大きさを示す

比例ハザードモデル

32

モデルの式: 各個体について以下の式を仮定

$$\lambda(t | x_1 \cdots x_p) = \lambda_0(t) \cdot \exp(\beta_1 x_1 + \beta_2 x_2 + \cdots + \beta_p x_p)$$

ハザード関数
(アウトカム)

ロジスティック回帰分析でおなじみの
「共変量 x と係数 β 」の項

基準ハザード関数
(baseline hazard function)

アウトカムがハザード関数になるような回帰分析
= 比例ハザードモデル

比例ハザードモデル

33

モデルの式: 各個体について以下の式を仮定

$$\lambda(t | x_1 \cdots x_p) = \lambda_0(t) \cdot \exp(\beta_1 x_1 + \beta_2 x_2 + \cdots + \beta_p x_p)$$

ハザード関数
(アウトカム)

ロジスティック回帰分析でおなじみの
「共変量 x と係数 β 」の項

基準ハザード関数
(baseline hazard function)

アウトカムがハザード関数になるような回帰分析
= 比例ハザードモデル

比例ハザードモデル

34

モデルの式:

$$\lambda(t | x_1 \cdots x_p) = \lambda_0(t) \cdot \exp(\beta_1 x_1 + \beta_2 x_2 + \cdots + \beta_p x_p)$$

ハザード関数
(アウトカム)

ロジスティック回帰分析でおなじみの
「共変量 x と係数 β 」の項

基準ハザード関数
(baseline hazard function)

ハザード関数のうち、
時間のみに依存し、
共変量とは関係ない部分

ハザード関数のうち、
共変量のみ依存し、
時間とは関係ない部分

具体例：肺癌による死亡と抗がん剤との関係

35

アウトカム：肺癌による死亡

共変量 x : 抗がん剤の投与あり=1 なし(プラセボ)=0

$$\text{モデルの式: } \lambda(t|x) = \lambda_0(t) \cdot \exp(\beta_1 x)$$

抗がん剤投与の場合のハザード関数

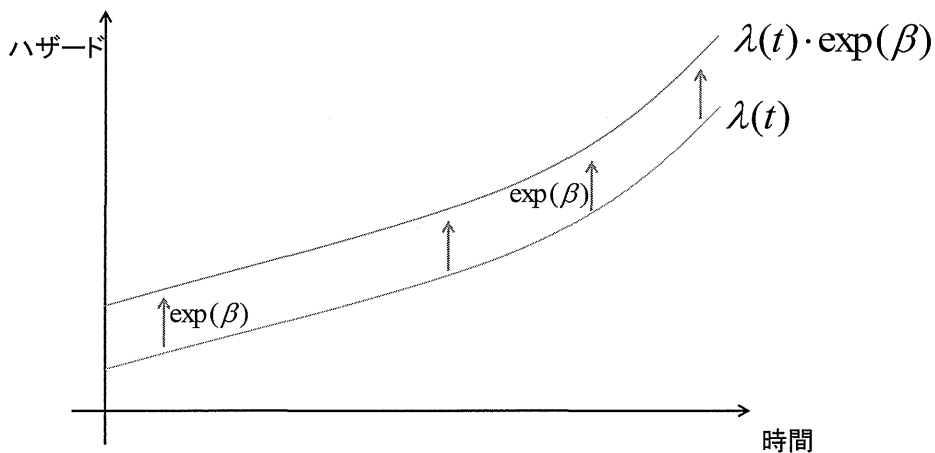
$$\begin{aligned} \lambda(t) &= \lambda_0(t) \cdot \exp(\beta \cdot 1) \\ &= \lambda_0(t) \cdot \exp(\beta) \end{aligned}$$

抗がん剤投与なしの場合のハザード関数(基準ハザード関数)

$$\begin{aligned} \lambda(t) &= \lambda_0(t) \cdot \exp(\beta \cdot 0) \\ &= \lambda_0(t) \end{aligned}$$

比例ハザードモデル：モデルの仮定

36



どの時点においても、抗がん剤を投与した方が、
基準ハザード関数よりも $\exp(\beta)$ だけ生存率大きい、
ということ表現しているモデル

比例ハザードモデル:モデルの仮定

37

実際のデータから,このような平行移動関係がみられるかを,
分析前に調べておく必要がある
(比例ハザード性の検証)

Kaplan-Meier 曲線を抗がん剤あり・なしの2群について描く
曲線が交差している場合,
「時間の経過とともに抗がん剤が効く」というような関係がみられる
(時間に依存して抗がん剤の効果が見られる)
→ 比例ハザードモデルの仮定を満たしていない

仮定が満たされない場合...

- Kaplan-Meier 曲線が交差していない部分についてのみ分析
- 時間の要素を一部入れたモデルによる分析
(ただし,「ハザード比 $=\exp(\beta)$ = 時間によらない相対リスク」
ではなくなる)

比例ハザードモデル:パラメタの解釈

38

説明変数が1単位上昇する → リスクが $\exp(\beta)$ となる

回帰係数 β の \exp がハザード比になる
(説明変数が名義尺度の場合)

ハザード比...説明変数が「1」の個体におけるハザードと,
説明変数が「0」の個体におけるハザードの比

[説明変数が間隔尺度・比率尺度の場合は,
「個体Aと,説明変数が1単位だけ大きな個体Bとのハザードの比」]

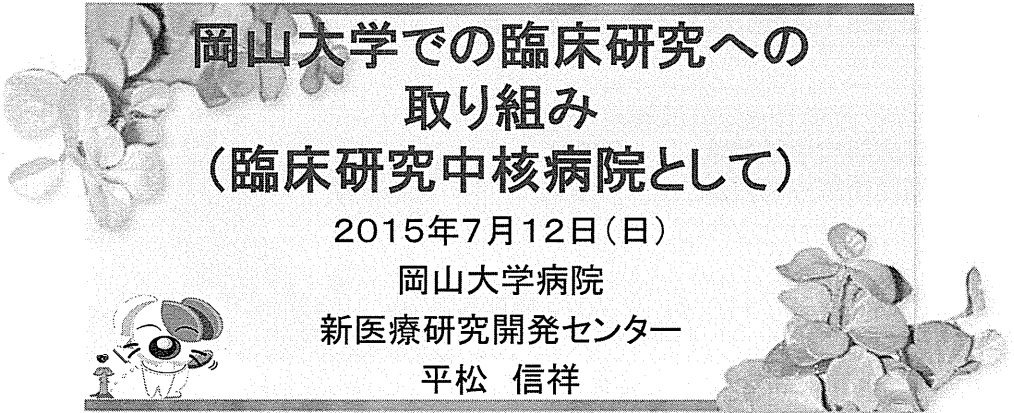
(ロジスティック回帰分析の「オッズ比」と同じ解釈で良い)

- 多変量解析を行うことで、交絡因子を調整した影響力指標を得ることができる
- データ(の背後にある誤差)の分布に何を仮定するかによって、適用する多変量解析手法は変わる
- パラメタ推定は統計ソフト(SPSS, JMP, STATA, R等)にお任せ

- 複雑なモデルよりも、単純なモデルを目指す(説明変数がたくさんあるモデルは解釈が難しい)
- モデルのあてはまりの指標にも注意

- 先行研究がよいお手本になる
- 研究デザインの検討の際、知っている統計手法の種類が多いと有利

ご清聴に感謝申し上げます



岡山大学での臨床研究への
取り組み
(臨床研究中核病院として)

2015年7月12日(日)

岡山大学病院

新医療研究開発センター

平松 信祥

本日のお話

1. 臨床研究中核病院について
2. 橋渡し事業拠点としての取り組み
3. AROとしてのサポート



国際水準の臨床研究メガホスピタル機能の強化

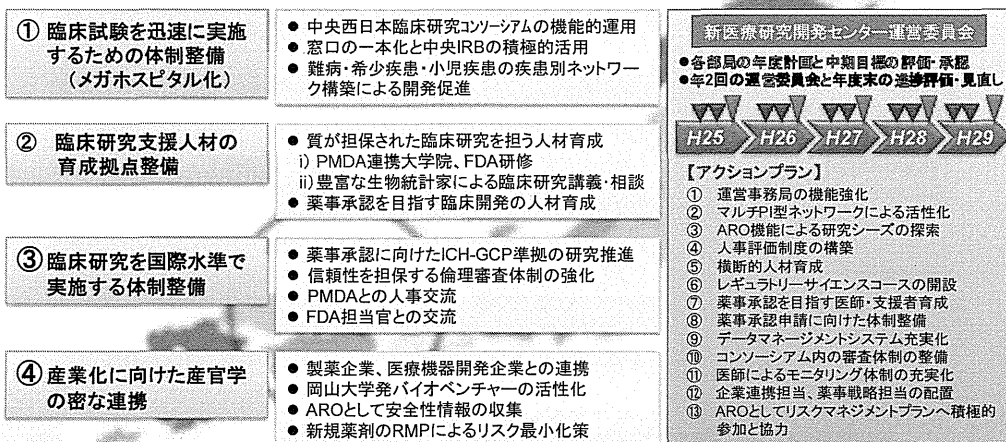
事業概要

- 規制当局との密接な連携を持ち、AROとして日本発の医薬品・医療機器を開発。
- ICH-GCPIに基づく医師主導治験・臨床試験を実施(難病・希少疾患・小児疾患を重点課題)。
- 市販後医薬品の適応拡大や多剤併用のエビデンスを創出。
- これらの成果を産業化につなげ医療産業クラスター形成を目指す。

強化課題

目標

進捗管理



橋渡し事業拠点

- 拠点外シーズを含むシーズ探索、選考
- シーズの実用化に向けた開発戦略の策定支援
- シーズの進捗管理、評価
- シーズの国際展開を目指した、国際的に通用する基準での支援、海外の共同研究機関対応、海外規制当局対応支援
- 企業とのマッチング活動支援
- 戦略的な知的財産の確保・活用の支援
- 迅速な治験完遂を可能とする等を目的としたネットワーク構築

シーズ育成機能の強化

シーズを開発段階に応じて以下の3つに分類し外部資金獲得までの育成費用を拠点に支援



※支援期間は原則1年間(推進委員会による審査に基づき継続可)とし、他の競争的資金等の外部資金獲得に努めることを条件とする。

シーズ提案から研究費支援決定までの流れ

