
変量効果を含む ロジスティック回帰

中外臨床研究センター
高橋 行雄

チュートリアルの目的

- ◆ 反応が2値の場合に対するSASの解析プロシジャの機能が拡張
- ◆ 多くの事例に応用が可能
- ◆ SASのマニュアル,あるいは,カテゴリカルデータの解析を扱っている成書では,不十分
- ◆ 世界的な規模で繰り返し実験が行われている臨床試験を事例
 - LOGISTICプロシジャのstrataステートメントによる条件付ロジスティック回帰
 - GENMODプロシジャを使ったGEE法
 - NLMIXEDプロシジャを使った変量効果を含むモデル

2値データに対する解析法の発展

最近の10年の変化

- ◆ PC版のSASバージョン6.03, 1993年10月に初版
 - FERQプロシジャおよびCATMODプロシジャ
 - 豊富な機能を有し, それ以前の単純な群間の有意差検定を主体にしてきた統計解析の枠組みを広げた.
- ◆ リリース6.07
 - 回帰分析の流れをくむLOGISTICプロシジャ
 - 多重分割表の解析の流れをくむCATMODプロシジャとを, どのように使い分けたら良いのか, それらの使い分けを模索した.
- ◆ リリース6.09
 - 一般化線形モデルのGENMODプロシジャ
 - 反応が計量値でも2値を含む計数値でも, どちらにも適用できる分散分析流のプロシジャ.

新たなチャレンジ: 変量効果

◆ リリース6.07

- 反応が計量値の場合MIXEDプロシジャ
 - 変量効果を常に含む分割実験の解析が手軽にできるようになった.

◆ バージョン8

- 反応が2値の場合でも, NLMIXEDプロシジャによって固定効果と変量効果含む場合でも解析が可能となった.
 - NLMIXEDプロシジャが, どの程度の問題まで適用できるかの見極めも今回の検討の課題である.

◆ バージョン9

- LOGISTICプロシジャは, でstrataステートメント
 - マッチングされた条件付ロジスティック回帰, PHREGプロシジャを使う必要がなくなった.
 - 臨床試験ごとに症例がマッチされたと考えることにより, 複数の臨床試験の併合の場合にも適用できる.

使いこなすために

- ◆ 典型的な事例に対する標準的なプログラムの例示
- ◆ 基本的な出力に即した解析結果の解釈
- ◆ 派生する応用事例を整備
- ◆ 回帰分析的なアプローチについては、日本語の成書も数多くある。
- ◆ 分散分析・共分散分析的アプローチも必須
 - 基本的な使い方を丁寧に示し、新しい機能の応用についても合わせて示す。

インフルエンザの発病予防

- ◆ アマンタジンのA型インフルエンザ発病予防.
- ◆ プラセボを対象とした臨床試験.
- ◆ Reumanら(1989)の論文
 - 異なる2つの試験結果が1論文として
 - 第1は, 18歳から55歳までの成人を対象にしたインフルエンザ発病の予防試験.
 - 第2の試験は, 18歳から40歳の志願者を対象とした**実験的**インフルエンザ発病の予防試験.

アマンタジンのインフルエンザ予防

Ex : 試験	用量 (mg)	発病あり	発病なし	<i>n</i>	発病率	相対リスク
Ex1 :	プラセボ	5	154	159	3.1%	100.0%
成人予防	100	2	157	159	1.3%	40.0%
	200	1	157	158	0.6%	20.1%
Ex2 :	プラセボ	11	8	19	57.9%	100.0%
実験的予防	50	4	16	20	20.0%	34.5%
	100	3	17	20	15.0%	25.9%
	200	2	17	19	10.5%	18.2%

実験の仮説

- ◆ 成人を対象にした予防試験
 - 予備試験の結果から発病率を30%と見積り
 - 適切な検出力が確保できる症例数と記載
 - 結果は、プラセボ群のインフルエンザ発病率が3.1%
- ◆ 実験的予防試験
 - プラセボのインフルエンザ発病率は90%以上
 - 50mg群では30%以下
 - 100と200mg群では20%以下
 - 検出力が99%と少数例でも効果の検証が可能と記載

2値データの解析の基本

- ◆ SASの多くのプロシジャ
 - 1症例ごとに1行となるような形式
 - 同じ反応 y の症例数が n 例あるというような集計されたデータセットでも扱える。
- ◆ 2値反応データの場合,
 - 反応ありの症例数を r
 - 分母となる症例数を n
 - r/n を“反応” とすることも可能

SASのデータステップ

```
data amantadine ;  
    input Ex $ x r n ;  
datalines ;  
Ex1_成人    0    5    159  
Ex1_成人   100    2    159  
Ex1_成人   200    1    158  
Ex2_実験    0   11    19  
Ex2_実験   50    4    20  
Ex2_実験  100    3    20  
Ex2_実験  200    2    19  
;
```

シグモイド曲線のおてはめ

- ◆ シグモイド曲線,
 - 各種のプロシジャで自由に選択が可能
 - 正規分布曲線
 - ロジスティック分布曲線,
 - 極値分布曲線
- ◆ 反応が 2値の場合のシグモイド曲線
 - 用量 x が0mgに近づくに従って反応が0%
 - 用量 x が増加した場合には100%に漸近

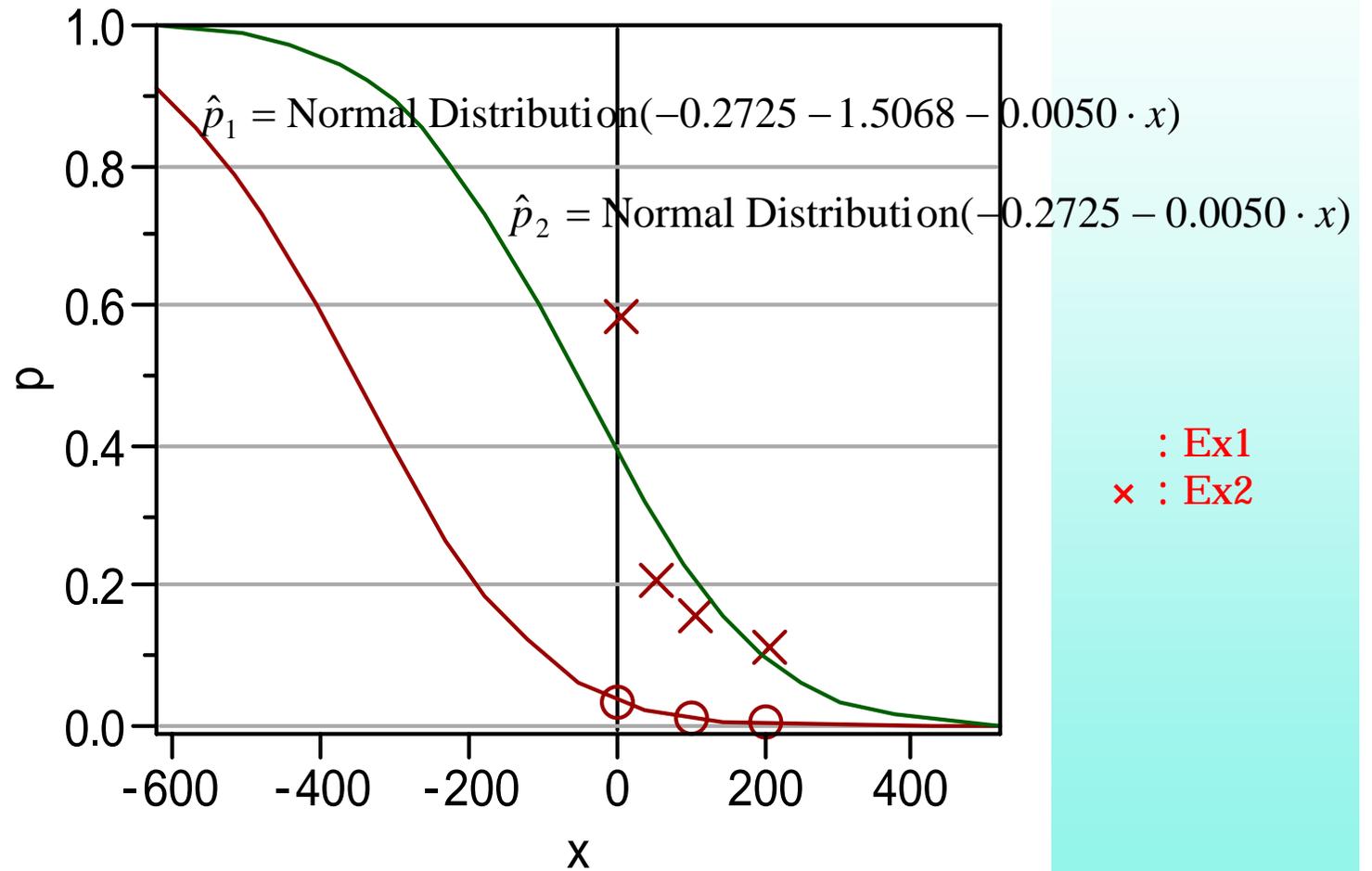
正規分布曲線のあてはめ

- ◆ Ex1_成人と Ex2_実験 に, 位置は異なるが同じ形状となる正規分布曲線のあてはめ

```
Proc genmod data=amantadine ;  
  class Ex ;  
  model r/n = Ex x / link = probit  
                                dist = binomial ;  
run ;
```

- ◆ class Ex ; 分類数に応じたインディケータ型ダミー変数が生成
- ◆ model r/n = Ex x ; 切片を含み4変数
- ◆ link = probit オプション: シグモイド曲線として正規分布曲線
- ◆ dist = binomial : 反応の誤差分布を, 2項分布

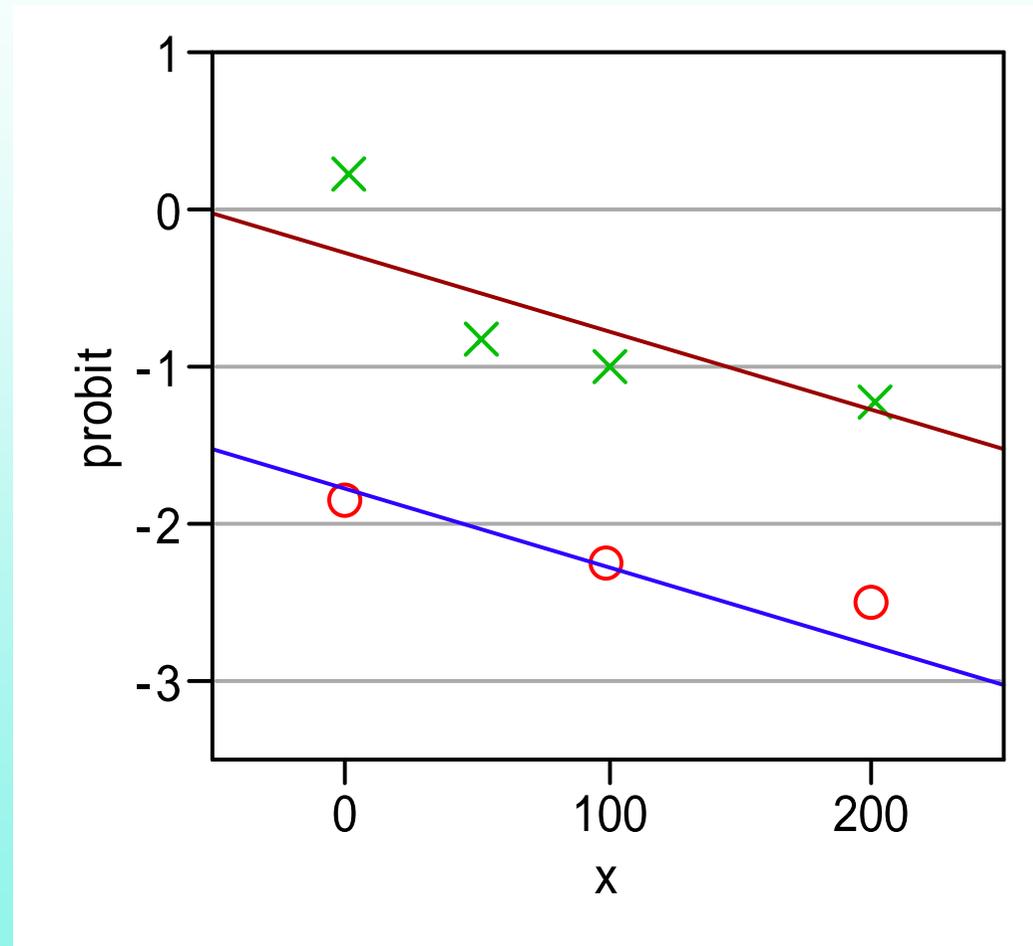
推定された正規分布曲線



あてはめの妥当性

- ◆ 用量 x としてはありえない負の領域を図示
- ◆ 正規分布曲線にかぎらず他のシグモイド曲線のあてはめでも共通
- ◆ 原因は対数用量としなかったため
 - 0mgの対数はマイナス無限大
 - 便宜的に x のままに
 - 回帰直線よりも曲線をあてはめたかった
- ◆ 解析モデルとして不適切

プロビットでの用量反応



: Ex1
× : Ex2

あてはめは妥当
なように見える

プロビット:

$$\eta_{ij} = \text{Normal Quatile}(p_{ij})$$

LOGISTICプロシジャの場合

```
proc logistic data=amantadine ;  
  class Ex / param=GLM ;  
  model r/n = Ex x / link=probit  
                                     technique=newton ;  
  
  run ;
```

- ◆ param=GLM : インディケータ型のダミー変数
- ◆ link=probit : シグモイド曲線に正規分布を
- ◆ technique=Newton : 解法をニュートン・ラフソン法に
 - LOGISTICプロシジャのデフォルトの解法がフィシャーのスコア法
 - GENMODプロシジャでは, 2階の偏微分を使うニュートン・ラフソン法
 - 2つの解法で, 回帰係数の推定値は一致, SEは完全には一致しない.

自然反応，免疫反応がある場合

- ◆ LOGISTIC , GENMOD プロシジャ
 - 解析できない
- ◆ 非線型回帰 NLMIXED プロシジャ
 - 解析できる
 - ダミー変数はデータステップで前もって生成する必要

NLMIXED プロシジャ

```
proc nlmixed data=amantadine2 df=9e+8 ;  
  parms beta0 0 beta1 0 beta3 0 ;  
  eta = beta0 + beta1*w1 + beta3*x ;  
  pi = probnorm(eta) ;  
  model r ~ binomial(n, pi) ;  
run ;
```

- ◆ $r \sim \text{binomial}(n, pi)$: r が, 観測数 n , 推定値 pi の二項分布
- ◆ $pi = \text{probnorm}(eta)$; pi は, 線形和 eta に対して正規分布曲線
- ◆ $df=9e+8$: t分布の自由度を大きくして正規分布とするため

NLMIXED プロシジャの出力

Parameter	Estimate	Standard Error	DF	t Value	Pr > t
beta0	-0.2725	0.1959	9E8	-1.39	0.1643
beta1	-1.5068	0.2199	9E8	-6.85	<.0001
beta3	-0.00497	0.001594	9E8	-3.12	0.0018

Output 1 GENMOD プロシジャによる回帰係数の推定値

Parameter		DF	Estimate	Standard Error
Intercept		1	-0.2725	0.1960
Ex	Ex1_成人	1	-1.5068	0.2199
Ex	Ex2_実験	0	0.0000	0.0000
x		1	-0.0050	0.0016
Scale		0	1.0000	0.0000

オッズ比による有害事象の比較

有害事象を軽減する用量

- ◆ 有害事象を減らす用量を探索することは、副次的な目的
- ◆ アマンタジン100mg群は、中枢神経系の有害事象を軽減しつつ予防効果が期待できる投与量かが副次的な目的
- ◆ 有害事象の投与量間のプロファイル比較をする場合の統計量としてオッズ比が適している。

オッズ比によるプロファイルの比較

	Placebo	100mg	200mg
	発現率	オッズ比	オッズ比
全有害事象	30.8%	0.94	1.83
Central nervous	15.7%	0.91	2.27
Gastrointestinal	7.5%	1.00	1.48
Cardiovascular	0.6%	1.00	1.01
Whole body	15.1%	0.63	0.86

LOGISTICプロシジャによる解析

```
proc logistic data=amantadine_ad ;  
  class x (param=reference ref=first) ;  
  model r/n = x / link=logit expb ;  
  by ad ;
```

- ◆ param=reference ref=first) : 最初の水準を基準としたダミー変数
- ◆ link=logit : オッズ比はシグモイド曲線にロジスティック曲線
- ◆ expb : 対数オッズ推定値にたいし指数を取ってオッズの出力指示

GENMODプロシジャによる場合

```
proc genmod data=amantadine_ad ;  
  class x (param=reference ref=first) ;  
  model r/n = x / dist=binomial link=logit ;  
  estimate 'odds: 0 vs 100' x 1 0 / exp ;  
  estimate 'odds: 0 vs 200' x 0 1 / exp ;  
  by ad ;  
  run ;
```

- ◆ Estimate ステートメントで推定のための係数を指定し ,
exp オプションで指数をとった推定値も出力する設定をする

オッズ比の計算

Label	Estimate	Standard Error	Alpha	Confidence Limits
対数odds: 0 vs 100	-0.0982	0.3135	0.05	-0.7127 0.5163
Exp(対数odds: 0 vs 100)	0.9065	0.2842	0.05	0.4903 1.6758
対数odds: 0 vs 200	0.8196	0.2788	0.05	0.2731 1.3661
Exp(対数odds: 0 vs 200)	2.2695	0.6328	0.05	1.3140 3.9200

世界的規模で行われている 繰り返し実験

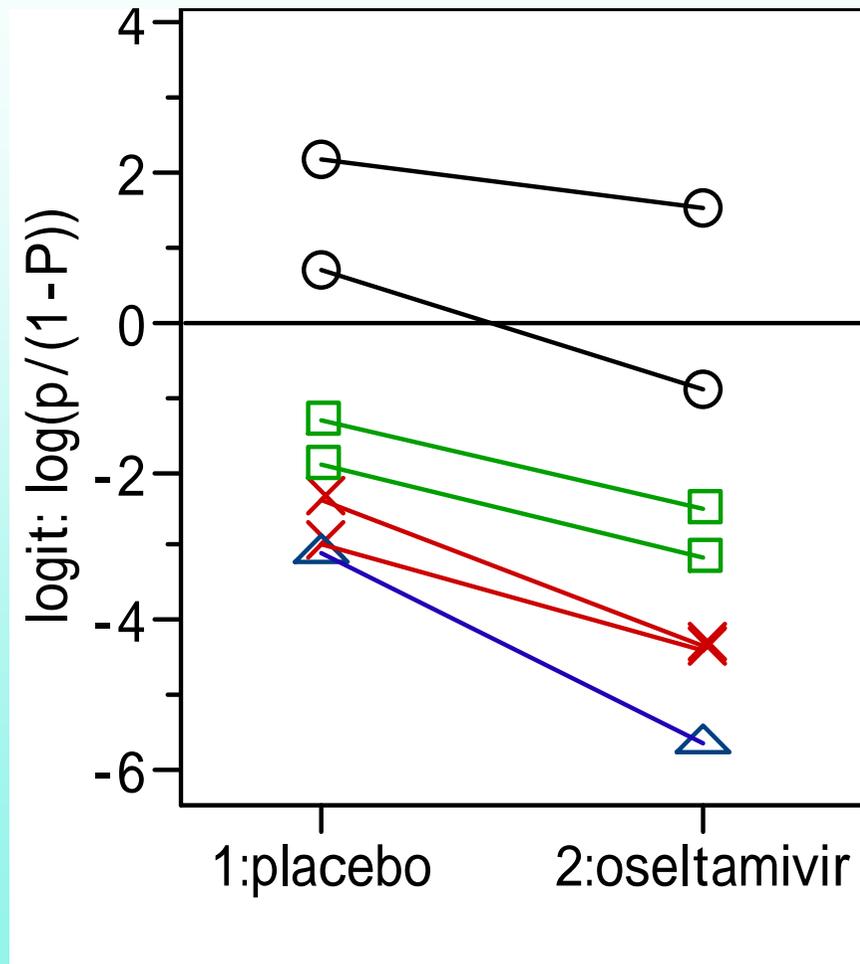
7つの臨床研究

- ◆ オセルタミビルインフルエンザの予防効果
- ◆ 7つの臨床研究は、独立に行われている
- ◆ 先行する研究デザインを踏襲
- ◆ 世界的規模で行われている繰り返し実験
- ◆ 異なる4つの対象集団
- ◆ 同様の薬効がオセルタミビルに期待できるのか
- ◆ 異なる対象集団は、実験計画法でいう標示因子
- ◆ それぞれの対象集団の中で、実験が2回反復

異なる 4 集団に対する試験

対象集団	No	B1: プラセボ群		B2: オセルタミビル群	
		<i>r</i>	<i>n</i>	<i>r</i>	<i>n</i>
A1:実験的研究	1	8	12	6	21
	2	17	19	32	39
A2:一般成人	3	25	519	13	1040
	4	13	155	2	153
A3 高齢者	5	12	272	1	276
A4:家族内	6	43	206	16	209
	7	33	258	10	244
全体		151	1441	80	1982

インフルエンザ発病率のロジット



- : A1 , 実験的研究
- × : A2 , 一般成人
- △ : A3 , 高齢者
- : A4 , 家族内

実験計画の観点

- ◆ 標示因子として因子Aを4水準
A1:実験的研究, A2:一般成人, A3:高齢者, A4:家族内
- ◆ 制御因子として因子Bでは2水準
B1:プラセボ群, B2:オセルタミビル群
- ◆ 因子Bは, 因子Aに対して枝分かれ, 2回反復
対象集団をランダムに B1:プラセボ とB2:オセルタミビル
インフルエンザの発病(有, 無)を観測
- ◆ それぞれの実験結果は1999年から2004年の5年
間で公表

GLM プロシジャ

```
proc glm data=ose4_01 ;  
  class A Ex B ;  
  model y = A Ex B A*B Ex*B / ss1 ;  
  test H=A E=Ex ;  
  freq n ;  
run ;
```

- ◆ 反応は, 0, 1 を計量値として便宜的に解析
- ◆ SS1 で逐次型平方和の分解を設定
- ◆ 1次単位の因子Aは, 1次誤差 Ex とするtestステートメントを使用

実験のランダム化の構造の把握

Output 6 GLM プロシジャの出力を整理した分散分析表

Source	DF	Type I SS	MS	F Value	Pr > F
A	3	41.18	13.725	8.96	0.0524
Ex	3	4.59	1.533	31.74	<.0001
B	1	3.31	3.311	68.66	<.0001
A*B	3	1.03	0.345	7.11	<.0001
Ex*B	3	0.64	0.214	4.44	0.0040
Error	3409	164.6	0.04		
Corrected Total	3422	215.4			

因子A の F 値は，誤差を Ex としている。
これは，test ステートメントで計算した結果である。

GENMODプロシジャによる解析

```
proc genmod data=oseltamivir4 ;  
  class A Ex B ;  
  model r/n = A Ex B A*B Ex*B  
          / link=logit dist=binomial type1 ;  
run ;
```

- ◆ modelステートメントオプションで type1 の指定
 - GLMプロシジャの ss1 と同じような逐次型の平方和の分解
- ◆ 対数尤度の 2倍で計算されるデビアンスの減少分
 - 因子の効果の大きさを表す統計量
- ◆ 枝分かれ型の分割実験
 - 対数尤度の 2倍を組替えことによる “分散分析” 表の作成

逐次型デビアンズ

LR Statistics For Type 1 Analysis

Source	Deviance	DF	Chi-Square	Pr > ChiSq
Intercept	<u>434.3730</u>			
A	<u>100.3931</u>	3	<u>333.98</u>	<.0001
Ex	71.0333	3	29.36	<.0001
B	3.7123	1	67.32	<.0001
A*B	1.1845	3	2.53	0.4703
Ex*B	0.0000	3	1.18	0.7567

- ◆ 因子Aの統計量は2つのデビアンズの差から

$$\chi_A^2 = D_T - D_A = 434.3730 - 100.3931 = 333.98$$

ロジスティック回帰の分散分析表

表5 ロジスティック回帰の分散分析表スタイル

要因	自由度	χ^2 値	F 値	p 値
A	3	333.98	11.37	0.0212
Ex	3	29.36		
B	1	67.32		<.0001
A*B	3	2.53		0.4703
Ex*B	3	1.18		0.7567
2 次誤差				
全体	13	434.37		

$$F_A = \chi_A^2 / \chi_{Ex}^2 = 333.98 / 29.36 = 11.37, p = 0.0212$$

オッズ比の推定

試験 Ex ごとのオッズ比

```
proc logistic data=oseltamivir4 ;  
    model r/n = B / link=logit technique=newton expb ;  
    by Ex ;  
run ;
```

共通オッズ比

```
proc logistic data=oseltamivir4 ;  
    class Ex (para=ref ref=first);  
    model r/n = Ex B / link=logit technique=newton expb ;  
run ;
```

固定効果の共通 オッズ比

対象集団	No. 著者 年	信 頼 区 間		
		オッズ比	L95	U95
A1:実験的研究	1:Hayden 1999 A 型	0.200	0.043	0.922
	2:Hayden 2000 B 型	0.538	0.100	2.880
A2:一般成人	3:Hayden 1999	0.250	0.127	0.493
	4:柏木 2000	0.145	0.032	0.652
A3 高齢者	5:WV15825	0.079	0.010	0.610
A4:家族内	6:Welliver 2001	0.314	0.171	0.579
	7:Hayden 2004	0.291	0.140	0.605
共通	最尤法：固定効果	0.258	0.182	0.365

条件付ロジスティック回帰

- ◆ バージョン9から
- ◆ LOGISTICプロシジャのstrataステートメント
- ◆ 条件付ロジスティック回帰による共通オッズ比

```
proc logistic data=oseltamivir4 ;    /* 条件付き */  
  class Ex ;  
  model r/n = B / link=logit technique=newton expb ;  
  strata Ex ;  
  run ;
```

結果の比較

- ◆ 最尤法による共通オッズ比 (95%信頼区間)
0.258倍 (0.182, 0.365)
- ◆ 試験Exをstrataステートメントで層
0.259倍 (0.184, 0.367)

- ◆ 結果は同様

4つの対象集団を 変量効果としたモデル

試験を変量効果とする解析

◆ GENMOD プロシジャ

- repeated ステートメントを使う GEE 法

◆ NLMIXED プロシジャ

- 対数オッズを変量効果として, 正規分布に従うとみなした解析

GENMODプロシジャ

```
proc genmod data=oseltamivir4 ;    /* 変量 GEE */  
  class Ex (param=ref ref=first) ;  
  model r/n = B / dist=binomial link=logit ;  
  repeated subject=Ex / type=ind ;  
  estimate '0 vs 1' B 1 / exp ;  
run ;
```

GEE

- ◆ modelステートメントでは, $r/n = x$
- ◆ repeatedステートメントでsubject=Ex
- ◆ 試験Ex間の相関構造は, 各試験は独立
 - 相関は0と仮定, type=ind
- ◆ 共通オッズ比は, estimateステートメントのexpオプションで推定値の指数を計算
- ◆ 薬剤Bは(0, 1)の計量値
 - 薬剤Bの回帰係数の推定値
 - プラセボに対するオセルタミビル対数の差
estimate '0 vs 1' B 1 / exp ;

GEEの結果

- ◆ 試験を固定効果にした最尤法
0.258倍 (0.182 , 0.365)
- ◆ GENMODプロシジャによるオッズ比と95%
信頼区間は,
0.3593倍 (0.1856 , 0.6958)
 - 信頼区間は拡大
 - 高い点推定値
 - その理由については今後の課題

NLMIXED プロシジャ

```
proc nlmixed data=oseltamivir4 ; /* 変量 NLMIXED */
  parms beta0 -2 beta1 -1 s1=1 ;
  eta = beta0 + beta1*B + u ;
  pi   = exp(eta) / (1+exp(eta)) ;
  model r ~ binomial(n,pi);
  random u ~ normal(0, s1*s1) subject=Ex ;
  predict eta out=eta;
  predict pi out=p;
run;
```

何を変量効果とするのか

- ◆ $\eta = \beta_0 + \beta_1 * B + u$;
- ◆ u が , 試験 Ex 間での変量効果
 - 平均が0 , 分散が $s_1 * s_1$ の正規分布と仮定
- ◆ randomステートメントで指定
 - `random u ~ normal(0, s1*s1) subject=Ex ;`
 - 現在は , 変量効果として正規分布のみ

NLMIXEDプロシジャによる推定結果

- ◆ 対数オッズ比 -1.336
- ◆ 指数を取りオッズ比に戻す
 - 0.282倍 (0.171, 0.404)
 - 妥当な推定値
 - 試験を固定効果にした最尤法(再掲)
0.258倍 (0.182, 0.365)
- ◆ Predict ステートメントにより, 固定効果と変量効果を含む推定値の出力
 - 変量効果が適切であったか残差を再計算
 - モデルの妥当性の検証

NLMIXEDによる推定値との残差

A	Ex	B	r	n	p	Predicted Value	差 ($\hat{p} - p$)
A1	1:Hayden_1999	0	8	12	0.67	0.6140	-0.0526
A1	1:Hayden_1999	1	6	21	0.29	0.2948	0.0091
A1	2:Hayden_2000	0	17	19	0.89	0.9303	0.0355
A1	2:Hayden_2000	1	32	39	0.82	0.7780	-0.0425
A2	3:Hayden_1999	0	25	519	0.05	0.0479	-0.0003
A2	3:Hayden_1999	1	13	1040	0.01	0.0130	0.0005
A2	4:柏木_2000	0	13	155	0.08	0.0775	-0.0064
A2	4:柏木_2000	1	2	153	0.01	0.0216	0.0085
A3	5:WV15825	0	12	272	0.04	0.0389	-0.0052
A3	5:WV15825	1	1	276	0.00	0.0105	0.0069
A4	6:Welliver_2001	0	43	206	0.21	0.2174	0.0087
A4	6:Welliver_2001	1	16	209	0.08	0.0680	-0.0085
A4	7:Hayden_2004	0	33	258	0.13	0.1312	0.0033
A4	7:Hayden_2004	1	10	244	0.04	0.0382	-0.0028

今後の課題

割愛した課題

- ◆ 反応に過大発散 (Over-dispersion) がある場合の問題
- ◆ LOGISTIC プロシジャ
 - scale オプション: Deviance 法, Pearson 法, Williams 法のそれぞれの適用の問題の明確化と選択基準の目安
- ◆ GENMOD プロシジャ
 - scale オプション: Deviance 法, Pearson 法
- ◆ 2値の経時データの解析への応用
 - GENMOD プロシジャでの GEE の適用も魅力的である。

自然反応や免疫反応がある場合

- ◆ NLMIXEDプロシジャ
 - 自然反応がある場合
 - 自然死などで反応が0%とならない場合
 - 自然反応
 - 用量を増やしても反応が100%に達しない場合
 - 自然反応と自然反応が両方ある場合
- ◆ シグモイド曲線として左に裾を引く極値分布曲線が反応曲線として妥当である場合
 - LOGISTICプロシジャ, GENMODプロシジャ, NLMIXEDプロシジャでの例示
- ◆ 機会があるごとに, これらの課題に対して紹介する

ご清聴ありがとうございました