

# ロジスティック回帰における対応 のあるデータの分析

深澤 武志

# ロジスティック回帰モデル

## 二項モデル

- 応答変数 $Y_i$ は二項分布に従う
  - $n=1$  と  $P(Y_i=1)=\pi_i$
- $g(\pi_i)=x_i\beta$
- リンク関数としてロジット関数を用いる
  - その他にはプロビット、補対数対数
- 一般的には  $\text{logit}(\pi_i)=x_i\beta$  の形

# Matched Data ( 1 )

$$Y_{ij} \sim \text{Bin}(1, \pi_{ij})$$

i - 層 (strata)

j - 層内でのオブザベーション番号

例)

	有	無	
Trt	$y_{i0}$		1
Pla	$y_{i1}$		1

# Matched Data ( 2 )

各層における応答は次のうち、どれかになる。

2	1	1	0																
<table><tr><td>1</td><td>0</td></tr><tr><td>1</td><td>0</td></tr></table>	1	0	1	0	<table><tr><td>1</td><td>0</td></tr><tr><td>0</td><td>1</td></tr></table>	1	0	0	1	<table><tr><td>0</td><td>1</td></tr><tr><td>1</td><td>0</td></tr></table>	0	1	1	0	<table><tr><td>0</td><td>1</td></tr><tr><td>0</td><td>1</td></tr></table>	0	1	0	1
1	0																		
1	0																		
1	0																		
0	1																		
0	1																		
1	0																		
0	1																		
0	1																		

ロジスティック回帰モデル

# 理論的な説明

モデル:  $\text{logit}(\pi_{ij}) = \alpha_i + \beta x_{ij}$

各層における対数尤度(strata i)

$$(\sum y_{ij})\alpha_i + (\sum y_{ij}x_{ij})\beta + \text{定数}$$

$(\sum y_{ij})$ は $\alpha_i$ の十分統計量

$(\sum y_{ij})$ における条件付き分析を行う

の十分統計量  $\sum (\sum y_{ij}x_{ij})$

# 分析手法

- 条件付きロジスティック回帰分析
  - データ変換をおこなった後、LOGISTICプロシジャを実行する
  - LOGISTICプロシジャをSTRATAステートメントと共に実行する
- 正確条件付きロジスティック回帰分析
  - データ変換をおこなった後、LOGISTICプロシジャをEXACTステートメントと共に実行する
  - LOGISTICプロシジャをSTRATAとEXACTステートメントと共に実行する

```
DATA Data2;  
  SET Data1;  
  DROP id1 outcome1 trt1;  
  RETAIN id1 outcome1 trt1;  
  IF(id=id1) THEN DO;  
    trt=trt1-trt; outcome=outcome1-outcome;  
    IF outcome=0 THEN DELETE;  
    OUTPUT;  
  END;  
  ELSE DO;  
    id1=ID; trt1=trt; outcome1=outcome;  
  END;  
RUN;
```

- データ変換前

( Data1 )

OBS	ID	Outcome	trt
1	1	1	1
2	1	1	0
3	2	1	1
4	2	0	0
5	3	0	1
6	3	1	0
7	4	1	1
8	4	0	0
9	5	1	1
10	5	0	0
11	6	0	1
12	6	0	0

- データ変換後

( Data2 )

OBS	ID	Outcome	trt
1	1	0	1
2	2	1	1
3	3	-1	1
4	4	1	1
5	5	1	1
6	6	0	1



# 条件付きロジスティック回帰モデル データ変換後のLOGISTICプロシジャ

```
PROC LOGISTIC DATA=Data2  
                                DESCENDING;  
    MODEL outcome=trt / NOINT;  
RUN;
```

# 条件付きロジスティック回帰モデル LOGISTICプロシジャ+STRATAステートメント

```
PROC LOGISTIC DATA=Data1;  
  STRATA ID;  
  MODEL outcome(event='1')=trt;  
RUN;
```

# 理論的な説明

モデル:  $\text{logit}(\pi_{ij}) = \alpha_i + \beta x_{ij}$

における に対する十分統計量は

$$\sum (\sum y_{ij} x_{ij})$$

であるが、この例では  $x_{i0}$  は値0を取るので

$$\sum y_{i1}$$

となる。

正確条件付きロジスティック回帰  
LOGISTIC プロシジャ+EXACT ステートメント

```
PROC LOGISTIC DATA=Data2  
    DESCENDING EXACTONLY;  
    MODEL outcome=trt / NOINT;  
    EXACT trt / ESTIMATE=both;  
RUN;
```

# 正確条件付きロジスティック回帰 LOGISTICプロシジャ+STRATAとEXACTステートメント

```
PROC LOGISTIC DATA=Data1 EXACTONLY;  
  STRATA ID;  
  MODEL outcome(event='1')=trt;  
  EXACT trt / ESTIMATE=both OUTDIST=dist1;  
RUN;  
PROC PRINT data=dist1;  
RUN;
```

# 正確なp値の求め方(1)

Sufficient Statistics

Parameter	Value
trt	3

十分統計量は、このサンプルでは値3を取ることが出力される。

# 正確なp値の求め方(2)

OBS	trt	Count	Score	Prob
1	0	1	4	0.0625
2	1	4	1	0.2500
3	2	6	0	0.3750
4	3	4	1	0.2500
5	4	1	4	0.0625

Score統計量が1より大きい時のProbをたすことによって、  
正確なp値を得ることができる

# まとめ

条件付きロジスティック回帰モデルの分析方法

- 1 . データ変換後、LOGISTICプロシジャを実行
- 2 . データ変換後、LOGISTICプロシジャをEXACT  
ステートメントとともに実行
- 3 . LOGISTICプロシジャをSTRATAステートメントと  
ともに実行
- 4 . LOGISTICプロシジャをSTRATAとEXACTステー  
トメントとともに実行