

## Matched case-control 試験の解析におけるSASの利用

### 1. 序

反応が2値の場合に限定し、ケース・コントロール試験でのSASの利用法の実際を紹介したい。ここでは乳児の「突然死」のリスクについての解析事例を取りあげる。**Retro-spective case-control study**ではしばしば、リスク評価の精度を上げるために交絡因子でのマッチングが行われる。この研究では年齢を交絡因子と考え1例のケースと1例または2例のコントロールとマッチングが施された。この様な試験のデータのモデルに基づく解析には条件付きロジスティックモデルが利用されてきた。ここでは共変数を含めた種々の「一般化線形混合モデル」の利用方法を紹介する。

### 2. 用いたデータベース

1992 - 1995年にかけてScotlandで実施されたマッチド・ケースコントロール試験のデータをダウンロードして使用した(4)。

### 3. データセットの説明

**ID** : 患者がの **ID** でケースあるいはコントロールでユニークな番号  
**ID** がケースとコントロールで同じ症例でマッチング。

**GROUP** : **A=CASE**、 **B=CONTROL**

**GRP** : **1=CASE**、 **0=CONTROL**

**CASE** は **ID**が 1 から 300 まであり。

**CONTROL**は 1から5が欠損で 6 から 207 までである。

**TIME** : **1=** 同一病院でのケース発生前 **2=**同一病院でのケース発生後の可能性が考えられるが詳細は不明。また、通常条件付きロジスティック回帰では **PHREG** で一種のダミー変数 (**TIME** は **1** がケース、**2** はコントロール) を定義するが、このデータはそのような対応はない。

**ONE** : 全て1のダミー (これは**GENMOD** 用のもの)

**AGE** : 年齢 (週)

**SEAS** : **1=冬**、 **2=春・秋**、 **3=夏**

**DEPCAT** : **DEPRIVATION SCORE (1 - 7, 7は最も重篤)**

**SEX** : **1=女**、 **2=男**

**SLEEPN1** : **1=幼児用ベッドを使用**、 **0=幼児用ベッドの使用なし**

```

DATA    CH85;
INPUT  ID GROUP  $ GRP TIME ONE AGE SEAS DEPCAT SEX  SLEEPN1;
CARDS;
  1      A      1      1      1      9.1429      1      7      1      1
  2      A      1      1      1     39.2857      1      7      1      .
  3      A      1      1      1     31.7143      1      5      1      .
  4      A      1      2      1     10.2857      1      7      0      .
  5      A      1      2      1     12.0000      1      5      0      .
  6      A      1      1      1      9.1429      1      2      1      0
  7      A      1      1      1      5.0000      1      4      1      0
  8      A      1      2      1      1.1429      1      4      0      0
  9      A      1      1      1     17.5714      1      1      1      0
 76      A      1      2      1     13.8571      2      6      0      .
略-----
300      A      1      1      1     10.1429      2      4      1      .
-----
  6      B      0      1      1      9.1429      1      2      1      1
  6      B      0      2      1      9.1429      1      2      0      0
  7      B      0      1      1      5.0000      1      4      1      1
  7      B      0      1      1      5.0000      1      .      1      0
  8      B      0      2      1      1.1429      1      5      0      0
  8      B      0      2      1      1.1429      1      6      0      0
 11      B      0      1      1     10.7143      2      4      1      0
 11      B      0      2      1     10.7143      2      4      0      0
 12      B      0      1      1     16.0000      2      7      1      1
 12      B      0      1      1     16.0000      2      6      1      0
略-----
202      B      0      1      1      8.1429      1      1      1      1
203      B      0      1      1     12.0000      1      3      1      1
203      B      0      2      1     12.0000      1      3      0      1
207      B      0      2      1     40.4286      3      3      0      1
207      B      0      .      1     40.4286      3      3      .      1
;

```

#### 4. ケースのコントロールに対するリスクの推測プログラム

```

* ===== Binary response ===== ;
* Section 8.5, Binary analysis (sleeps in cot), Models 1-4;
OPTIONS LS=80 NODATE NONUMBER;
DATA a; SET ch85;
/* ----- 頻度集計 ----- */
  1          CASE のみ
  1 0          CASE=1, CONTROL= 1
  1 0 0        CASE=1, CONTROL= 2
  1 0 0 0      CASE=1, CONTROL= 3
-----*/
PROC SORT DATA= CH85; BY ID DESCENDING GRP;
DATA MATCHSET; SET CH85 ; BY ID DESCENDING GRP;

```

```

KEEP ID MATCH; LENGTH MATCH_ $30;
RETAIN MATCH_ ;
  IF FIRST. ID THEN MATCH_=LEFT(GRP);
  ELSE MATCH_ = TRIM(LEFT(MATCH_)) || TRIM(LEFT(GRP));
  IF LAST. ID THEN DO;
    MATCH = INPUT(MATCH_ , 8. ) ; OUTPUT ; END;
PROC FREQ DATA=MATCHSET ;TABLES MATCH / ; RUN
* Model 1;
PROC GENMOD DATA=a;
CLASS id;
MODEL sleepn1/one = grp/ DIST=B;
REPEATED SUB=id/ TYPE=CS CORRW MODELSE;

```

<model 1 によるSAS のOUTPUTの抜粋>

GENMOD Procedure

```

DISTRIBUTION:BINOMIAL, LINK FUNCTION:LOGIT
DEPENDENT VARIABLE SLEEPN1
DEPENDENT VARIABLE ONE
observations used 421
Number of Event 175
Number of Trials 421
Missing values 56

```

```
LOG LIKELIHOOD -281.0617
```

Analysis of Parameter Estimates

Parameter	DF	Estimate	Std Err	ChiSquare	Pr>Chi
INTERCEPT	1	-0.1238	0.1208	1.0496	0.3056
GRP	1	-0.6527	0.2153	9.1942	0.0024
SCALE	0	1.0000	0.0000		

GEE MODEL INFORMATION

```

Correlation Structure Exchangeble
Subject Effects ID
Number of Clusters 201
Cluster with Missing Values 56

```

Working Correlation Matrix

	col1	col2	col3
row1	1	0.2939	0.2939
row2	0.2939	1	0.2939
row3	0.2939	0.2939	1

Analysis of GEE Parameter Estimates  
Empirical Standard Error Estimates

**Empirical 95% Confidence Limits**

Parameter	DF	Estimate	Std Err	Lower	Upper	Z	Pr> Z
INTERCEPT	1	-0.0880	0.1386	-0.3596	0.1837	-0.6345	0.5258
GRP	1	-0.6721	0.1904	-1.0453	-0.2988	-3.5290	0.0004
SCALE	0	1.0000	0.0000	.	.	.	.

**Analysis of GEE Parameter Estimates**  
**Model-based Standard Error Estimates**

**Empirical 95% Confidence Limits**

Parameter	DF	Estimate	Std Err	Lower	Upper	Z	Pr> Z
INTERCEPT	1	-0.0880	0.1386	-0.3596	0.1837	-0.6345	0.5258
GRP	1	-0.6721	0.1841	-1.0330	-0.3112	-3.6500	0.0003
SCALE	0	1.0008	0.0000	.	.	.	.

```
* Model 2;
PROC PHREG DATA=a NOSUMMARY;
MODEL time*sleepn1(0)=grp/
TIES=DISCRETE; STRATA id;
```

<model 2 によるSAS のOUTPUTの抜粋>

The PHREG Procedure						
Analysis of Maximum likelihood Estimates						
Variable	DF	parameter estimate	Standard Error	Wald Chi-Square	Pr> Chi-Square	Risk Ratio
sleepn1	1	-0.332214	0.717	0.24538		1.83298
0.1758						

```
* Model 2';
* 条件付ロジスティック回帰分析(付録) ;
data a ; set a;
status= 2- grp;
/*注意：status は CASE=1, CONTROL=2 でケースよりコントロールが大
きくなければならない。
GRPでは case は1, control は0 なので,status は caseで1, controlでは2となる*/
```

```
PROC PHREG DATA=a NOSUMMARY;
MODEL status*grp(0)=sleepn1/TIES=DISCRETE;
STRATA id;
```

<model 2' による SASのoutput の抜粋>

```
Dependent variables      Status
Dependent variables      GRP
Censoring Handlings      Discrete
```

Score 14.695 with 1 df (p=0.0002)

Variable	DF	Parameter Estimate	Standard Error	Wald Chi-Square	Pr>Chi-Square	Risk Ratio
Sleepn1	1	-1.043126	0.28746	13.16826	0.0003	0.352

```
* Model 3;
PROC GENMOD DATA=a;
MODEL sleepn1/one = grp seas age/ DIST=B;
```

<model 3 による SAS の output の抜粋>

```
DISTRIBUTION:BINORMAL, LINK FUNCTION:LOGIT
DEPENDENT VARIABLE SLEEPN1
DEPENDENT VARIABLE ONE
observations used 421
Number of Event 175
Number of Trials 421
Missing values 56

LOG LIKELIHOOD -209.5423
```

#### Analysis of Parameter Estimates

Parameter	DF	Estimate	Std Err	ChiSquare	Pr>Chi
INTERCEPT	1	-2.5500	0.4393	33.6968	0.0001
GRP	1	-1.0051	0.2693	13.9298	0.0002
SEAS	1	0.1366	0.0175	0.5778	0.4472
AGE	1	0.1587	0.0175	82.4572	0.0001
SCALE	0	1.0000	0.0000	.	.

```
* Model 4;
PROC GENMOD DATA=a;
MODEL sleepn1/one = grp/ DIST=B;
RUN;
```

<model 4 による SAS の OUTPUT の抜粋> <-----これはモデル1の最初部分と同じ

```
The GENMOD Procedure
DISTRIBUTION:BINORMAL, LINK FUNCTION:LOGIT
DEPENDENT VARIABLE SLEEPN1
DEPENDENT VARIABLE ONE
observations used 421
Number of Event 175
Number of Trials 421
Missing values 56

LOG LIKELIHOOD -281.0617
```

## Analysis of Parameter Estimates

Parameter	DF	Estimate	Std Err	ChiSquare	Pr>Chi
INTERCEPT	1	-0.1238	0.1208	1.0496	0.305
GRP	1	-0.6527	0.2153	9.1942	0.0024
SCALE	0	1.0000	0.0000	.	.

## 5. 成績の解釈

データセット中のペアーの分布  
各ケース症例を年齢でマッチングした場合のペアー

1	45	例 (ケースのみでマッチング相手なし)
10	36	例 (ケース1: コントロール 1) -----> 1:1 MATCHING DATA
100	120	例 (ケース1: コントロール 2) -----> 1:2 MATCHING DATA

LOGIT SCALE におけるグループ効果 (ケース/コントロール: 変数はGRP) の検討

	LOGIT SCALE			
	ESTIMATE	SE	ODDS RATIO	95 C.I.
MODEL 1	-0.67	0.19	0.51	0.35-0.74
MODEL 2	-0.33	0.25	0.72	0.44-1.16
MODEL 3	-1.01	0.27	0.37	0.22-0.62
MODEL 4	-0.65	0.22	0.52	0.34-0.80

年齢によるマッチングを考慮した解析結果

	LOGIT SCALE		
	ESTIMATE	SE	ODDS RATIO
MODEL 2'	-1.043	0.29	0.352

その結果、モデル2以外はいずれも幼児用ベッド使用下での睡眠は幼児用ベッド以外 (乳母車、バスケット、籠等) での睡眠と比べリスクが減少することが示唆された。年齢でマッチングしたことを考慮し、1:1, 1:2 を含めたマッチング (NMマッチング) ではより顕著に幼児用ベッド使用によるリスクの減少が現れている。しかし、本に示されたモデル2ではリスクの有意な軽減は認められなかった。TIME 変数の作成方法が正確に分らないので解釈は困難である。

マッチド・ケースコントロール試験では conditional logistic regression modelより random effects model による解析の方が好んで用いられる。それは conditional logistic regression ではケースもコントロールも「共に反応あり」は情報ロスとなり、ケースあるいはコントロールのみで反応をみた症例数で分析されるという特徴も一つの理由である。混合モデルの共分散構造には compound symmetry covariance が利用される。モデル1,2,3,4 はそれぞれ異なるグループ効果を示す。

これは睡眠の場所が年齢と関係していることによるためかもしれない。モデル1とモデル2は正確に年齢でマッチングされている。しかし、結果が異なる理由はモデル2は睡眠の場所は統計モデル上では無視されているためである。

モデル3は年齢の効果が有意である。ただし、年齢がロジスチックスケール上での線形であるとの仮定がある。モデル4は年齢の影響が除かれていないため適切とは言えない。モデル4の標準誤差はモデル3より小さいが、これはケースとコントロールで脱落の発生がアンバランスであり適切とは言えない。よって、結果モデル3では標準誤差が大きくなったと考えられる。これらの解析結果は混合効果モデルを利用する価値を表している。

我々が期待したように、固定効果モデルに比べ混合効果モデルは群効果の標準誤差を減少させた。

## 6. 終わりに

リスクの評価に混合モデルが利用された。ここでは全てのデータが適応された。

マッチングとは**CASE**と交絡因子の分布が同様になるように**CONTROL**を選ぶことである。よって一般には分析のパワーは高いと言われている。しかし、交絡因子はまず「暴露変数と関係すること」、更に「反応変数とも関係するもの」でなければならない。マッチングする場合は当然ながらマッチングさせた因子のリスクは検討出来ない。複数の因子でマッチングすることも可能であるが、マッチングされない症例がその分増えるのでマッチング因子の数には限界がある。

## 文献

1. Maura E. Stokes et al.(1995): **Categorical data Analysis using the SAS SYSTEM, p275.**  
Chapter 10, Conditional Logistic regression .
2. SAS Institute(version 6): **Logistic Regression Examples using the SAS System. p119.**  
Example 14. analysis of N:M matched data.
3. 浜田知久馬: **SUGI-J (1994) : 527-540**
4. Helen Brown and Robin Prescott (1999) : **Applied Analysis in Medicine, John Wiley & sons ltd.** データをダウンロードする Web site: [www.med.ed.ac.uk/phs/mixed/](http://www.med.ed.ac.uk/phs/mixed/)

## 付録

条件付きロジスチック回帰分析は **N:M** マッチングのデータに適応される。浜田は頻度データの形式で**PHREG**を使用する方法を紹介している。浜田の例(**1:1 MATCED DATA**)で確かめてみたが本質的に結果に差異はなかった。その結果、回帰係数 = **-1.0986, SE=0.57735, p=0.0571**であった。一方、ここで紹介された混合効果モデル、たとえばモデル1を適応する場合、この例ではケース、コントロールともに反応なし(00)が**48**例、また共に反応あり(11)が**16**例あるため、これらのタイの数が反映し、**GRP**の回帰係数は**-0.4796**とリスク比の点推定値自体は小さくなるが、その標準誤差(**SE=0.3487**)は小さくなり、その結果 **P=0.042**と有意差が付くことは興味深い。

```
DATA A;
  INPUT CASE CONTROL W @@ ;
  DO I=1 TO W; OUTPUT; END;
  CARDS;
  0 0 48
  1 1 16
  1 0 12
  0 1 4
```

- \* 0 0 48 とは CASEも CONTROLも暴露無し
- \* 1 1 16 とは CASEも CONTROLも暴露あり
- \* 1 0 12 とは CASEは暴露ありで CONTROLは暴露なし
- \* 0 1 4 とは CASEはなしで CONTROLは暴露あり

```
DATA B ; SET A ;
ID=_N_ ;
T=1 ;
GRP=0 ; SLEEPN1= CASE ; OUTPUT ;
GRP=1 ; SLEEPN1=CONTROL ; OUTPUT ;
```

\* ————ここで示した条件付きロジスティック回帰での指定方法 (A) ————

```
DATA C; SET B;
STATUS=2-GRP;
PROC SORT DATA=C; BY ID;
PROC PRINT DATA=C; VAR GRP SLEEPN1 STATUS T;
PROC PHREG DATA=C;
STRATA ID;
MODEL STATUS*GRP(0)=SLEEPN1 ; RUN;
```

\* ————変数 T の利用 (B) ————

```
PROC PHREG DATA=C;
STRATA ID8;
MODEL T*GRP(1) = SLEEPN1 / TIES =DISCRETE ; RUN;
```

- \* T は CASE と CONTROL で「共通の値をとる」時間変数
- \* GRPは CASE=0, CONTROL=1

\* ————モデル1を適応 (C) ————

```
PROC GENMOD DATA=C;
CLASS ID ;
MODEL SLEEPN1 / ONE = GRP / DIST = B ;
REPEATED SUB=ID / TYPE =CS CORRW MODELSE ; RUN;
```

\* ————その他の方法 (LOGISTIC PROCEDURE) ———— ;

1:1 マッチングでは McNemar 検定が利用される。SAS ではCASE とCONTROLの差をとり、これを新たに独立変数とする。従属変数は全てのペアで同じ値をとる変数 (たとえば1) を定義し、「切片なしの LOGISTIC 分析」にて実施できる。

## アウトプット

(A)

## The PHREG Procedure

Variable	DF	parameter estimate	Standard Error	Wald Chi-Square	Pr> Chi-Square	Risk Ratio
sleepn1	1	-1.098612	0.333	0.57735		3.62085
	0.0571					

(B)

Variable	DF	parameter estimate	Standard Error	Wald Chi-Square	Pr> Chi-Square	Risk Ratio
sleepn1	1	1.098612	3.000	0.57735		3.62085
	0.0571					

(C)

Subject Effect	ID (80 levels)
Number of Cluster	80
	col1 col2
row1	1 0.5517
row2	0.5517 1

## Analysis of GEE Parameter Estimate

Empirical Standard Error Estimate	95% Confidence Limits
	Lower Upper Z Pr> Z
INTERCEPT -0.6190 0.2344	-1.0785 -0.1596 -2.641 0.0083
GRP -0.4796 0.2360	-0.9420 -0.0171 -2.032 0.0421
Scale 1.0063	

## Analysis of GEE Parameter Estimate

Model Based Standard Error Estimate	95% Confidence Limits
	Lower Upper Z Pr> Z
INTERCEPT -0.6190 0.2359	-1.0814 -0.1567 -2.624 0.0087
GRP -0.4796 0.2356	-0.9414 -0.0177 -2.035 0.0418
Scale 1.0063	