移動表の分析：SAS、SPSS を用いたログリニア分析

平田　周一
日本労働研究機構

1. 対数線形モデル（loglinear model）とは何か。

a. 対数線形モデルの基本式

$\log F_{ij} = \mu + \lambda_i^R + \lambda_j^C + \lambda_{ij}^{RC}$ （1）2 変数（2 重クロス表）の場合

$\log F_{ijk} = \mu + \lambda_i^R + \lambda_j^C + \lambda_k^S + \lambda_{ij}^{RC} + \lambda_{ik}^{RL} + \lambda_{jk}^{CL} + \lambda_{ijk}^{RCL}$ （2）3 変数（3 重クロス表）の場合

対数線型モデルとは、多重クロス表を分析するための分析モデルで、上に示した様に、線形回帰式とよく似た形をとっているが、各セルの度数の対数をとったものを従属変数とする。(1) 式、(2) 式は、それぞれ、2 変数の場合、3 変数の場合の飽和モデル（saturated model）を示す。ここで、λ_i^R は行の変数の i 番目のカテゴリーの効果、λ_j^C は列の変数の j 番目のカテゴリーの効果を意味する。λ_{ij}^{RC} は i 行 j 列のセルに固有の効果で、交互作用項（interaction term）と呼ばれる。

あらためて定義すると、飽和モデルとは、クロス表における観察度数をカテゴリー変数、及びカテゴリー変数間のすべての組み合わせの関数として表現したもので、いずれの場合も、飽和モデルは観察度数を完全に再現する。

b. 対数線形モデルとカイ二乗検定

クロス表の分析には、カイ二乗値による検定が伝統的に用いられてきた。カイ二乗値は以下の式によって求めることができる。

$$\chi^2 = \sum_i \sum_j \frac{(f_{ij} - F_{ij})^2}{F_{ij}} \qquad (3)$$

ここで、 f_{ij} は i 行 j 列のセルにおける実際の観測値、 F_{ij} は期待値を表す。個々の期待値 F_{ij} は以下の式によって計算される。

$$F_{ij} = N * p_j * p_i \qquad (4)$$

ここで、 p_j は i 行の行比率、 p_i は j 列の列比率、N はクロス表全体の度数を表す。すなわち、期待度数とは、クロス表の周辺分布にしたがって全体の度数を分配したものであり、カイ二乗値とは、期待度数と実際の観測値の差を示す統計量である。したがって、カイ二乗値が小さければ、クロス表のセル内の度数は周辺度数によって決定されるが、カイ二乗値が十分に大きければ、周辺度数だけでは説明できない観測値が得られた、すなわち、行と列の変数の間に何らかの関連があるとされる。

対数線形分析と従来のカイ二乗検定の方法との間に根本的な違いはない。ここで、カイ二乗検定で期待度数を求める式（4）から対数線形モデルの基本式を導いてみよう。

実際のクロス表の観測度数は、周辺度数だけから求められる期待度数より大きかったり、少なかったりするので、i 行 j 列のセルに固有の期待度数からの偏差を e_{ij} とすると、観測度数 f_{ij} は、

$$\begin{aligned} f_{ij} &= F_{ij} + e_{ij} \\ &= N * p_i * p_j + e_{ij} \end{aligned}$$

という式で表される。ここで、$N * p_i * p_j + e_{ij} = N * p_i * p_j * \tau_{ij}$ となるように τ_{ij} をとり、右辺の

対数をとると、次のように 2 重クロス表の飽和モデルの式を得ることができる。

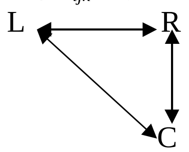
$$\begin{aligned}\log f_{ij} &= \log N + \log p_j + \log p_i + \log \tau_{ij} \\ &= \mu + \lambda_i^R + \lambda_j^C + \lambda_{ij}^{RC}\end{aligned} \qquad (6)$$

ただし、クロス表は周辺度数が固定されているので、パラメーターを識別するために制限を設けなければならない。この制限には、いくつかの方法がある。一つは、下に示すようにすべてのパラメーターを足すと 0 になるように標準化する方法で、ANOVA コーディングと呼ばれる。もう一つの方法は一つのパラメーターを 0 とし他のパラメーターはそこからの偏差とするもので、DUMMY コーディングと呼ばれる。いずれの場合も 1 つだけパラメーターが省略される。

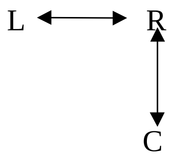
$$\sum \lambda_i^R = \sum \lambda_j^C = \sum \lambda_{ij}^{RC} = 0$$

このように、対数線型モデルとは、クロス表をセル内の度数を従属変数とする線型モデルに翻訳し、様々なモデルをたて、そのモデルが観察度数をよく説明するかを検討する方法である。(2) 式に示されている、3 重クロス表の飽和モデルから交互作用項を除くことによって、次の図のように様々なモデルをたてることができる。

$$\log F_{ijk} = \mu + \lambda_i^R + \lambda_j^C + \lambda_k^L + \lambda_{ij}^{RC} + \lambda_{ik}^{RL} + \lambda_{jk}^{CL}$$



$$\log F_{ijk} = \mu + \lambda_i^R + \lambda_j^C + \lambda_k^L + \lambda_{ij}^{RC} + \lambda_{ik}^{RL}$$



対数線形モデルを評価するには、モデルによって算出された期待値が観測値に近いほど観測値をよく説明するモデル、データに適合（fit）するモデルとする。データを評価するためにカイ二乗検定が行われるが、その際、従来のカイ二乗値ではなく尤度比（Likelihood Ratio）という統計量を用いる。カイ二乗値と尤度比は、サンプルの数が多ければ近似するが、尤度比は加法性をもっているという利点がある。

対数線型モデルの検定が従来のカイ二乗検定と異なっていることにとまどう者がいるが、従来のカイ二乗検定で有意だということは、交互作用項がない独立モデルはデータに適合しないことを意味している。

2. 移動表のログリニア分析

a. データ

移動表とは、おもに社会階層論の分野で世代間、世代内の階層移動、職業移動の研究のために用いられるもので、形式的には行変数と列変数のカテゴリーが同一である。以下の説明では、Hout（1983）で用いられているものと同じ、アメリカにおける世代間移動表をデータとして用いる。これは、もとは Featherman & Hauser（1978）に掲載されていたもので、実際の調査は 1973 年に労働力調査の補足調査として行われた。

表 1 に、用いられるデータの実数と完全移動モデルを適用した結果を示す。完全移動モデルとは、下に示す様に交互作用項を含まない、行変数と列変数の主効果のみを設定したモデルである。

$$\log F_{ij} = \mu + \lambda_i^R + \lambda_j^C \qquad (7)\ \text{完全移動モデル}$$

表 1 父親と息子の職業の関連（上:観測値　下:完全移動モデルでの期待値）

Father's Occupation	Son's Occupation: Upper Non-manual	Lower Non-manual	Upper Manual	Lower Manual	Farm	Total
Upper Nonmanual	1,414	521	302	643	40	2,920
Lower Nonmanual	724	524	254	703	48	2,253
Upper Manual	798	648	856	1,676	108	4,086
Lower Manual	756	914	771	3,325	237	6,003
Farm	409	357	441	1,611	1,832	4,650
Total	4,101	2,964	2,624	7,958	2,265	19,912

Father's Occupation	Son's Occupation: Upper Non-manual	Lower Non-manual	Upper Manual	Lower Manual	Farm	Total
Upper Nonmanual	601.39	434.66	384.80	1,167.00	332.15	2,920
Lower Nonmanual	464.02	335.37	296.90	900.43	256.28	2,253
Upper Manual	841.54	608.22	538.45	1,633.00	464.78	4,086
Lower Manual	1,236.36	893.58	791.07	2,399.15	682.84	6,003
Farm	957.70	692.18	612.78	1,858.41	528.94	4,650
Total	4,101	2,964	2,624	7,958	2,265	19,912

Chi-Square=7166.771　*df*=16　P=0.001
Likelihood Ratio Chi-Square=6170.130　*df*=16　P=0.001

```
DATA HOUT;
INPUT FATHER SON COUNT @@ ;
CARDS ;
    1  1  1414  1  2  521  1  3  302  1  4   643  1  5    40
    2  1   724  2  2  524  2  3  254  2  4   703  2  5    48
    3  1   798  3  2  648  3  3  856  3  4  1676  3  5   108
    4  1   756  4  2  914  4  3  771  4  4  3325  4  5   237
    5  1   409  5  2  357  5  3  441  5  4  1611  5  5  1832
; RUN  ;

DATA LIST FREE / FATHER SON COUNT.
WEIGHT BY COUNT.
BEGIN DATA
    1  1  1414  1  2  521  1  3  302  1  4   643  1  5    40
    2  1   724  2  2  524  2  3  254  2  4   703  2  5    48
    3  1   798  3  2  648  3  3  856  3  4  1676  3  5   108
    4  1   756  4  2  914  4  3  771  4  4  3325  4  5   237
    5  1   409  5  2  357  5  3  441  5  4  1611  5  5  1832
END DATA.
EXECUTE.
```

例 1　SAS（上）と SPSS（下）によるデータの読み込み手続き

b. 準完全移動モデル（Quasi-Perfect Mobility Model）

移動表を分析する時、完全独立モデルが適合する事は非常に稀にしか起きない。その第一の理由は、移動表では、移動元と移動先で同じカテゴリーに属するものが多いことにある。移動表に限らず、行変数と列変数が同じカテゴリーを持つクロス表では、行と列のカテゴリーが一致するセル、すなわち、対角線上のセルにおける観測度数が高くなる。このようなクロス表を分析するモデルとして Goodman（1979）によって提唱されたのが準完全移動モデル（Quasi-Perfect Mobility Model）である。簡単にいえば、準完全移動モデルは表の対角線上に位置するセルを除いた部分では独立モデルが適用されると仮定する。すなわち、対角線以外のセルでは交互作用項のない独立モデルが適用

され、対角線上のセルでは、以下のように交互作用項のあるモデルが適用される。

$$\log F_{ij} = \mu + \lambda_i^R + \lambda_j^C$$

$$\log F_{ij} = \mu + \lambda_i^R + \lambda_j^C + \lambda_{ij}^{RC} \qquad （但し、i = j）$$

準完全移動モデルを実際に行うために、対角線上のセルをブロックするという方法がよく用いられる。下図は、SPSS を用いた例では、/CWEIGHT=というサブコマンド以下に示されている対角線上のセルには 0、対角線以外のセルには 1 が配置されているマトリックスによって、対角線をブロックする。SPSS のマニュアルによれば、“ /CWEIGHT= ”はセルのウェイトを指定するサブコマンドで、図のようなマトリックスを指定すると対角線上の観測度数、期待度数の双方ともに 0 が割り当てられる。

表 2 に準完全移動モデルによる期待値と尤度比による検定結果が示されている。有意水準は 1 パーセント未満なので、準完全移動モデルは適合しないことがわかるが、完全移動モデルの尤度比の値と比べると、5486.79（6170.13－683.34）ポイント小さくなっている。これは自由度 5 で有意、すなわち、準完全移動モデルによって統計的に有意に適合度が上昇したことがわかる。

```
LOGLINEAR
   FATHER(1, 5) SON(1, 5)
   /CWEIGHT= (0 1 1 1 1
                          1 0 1 1 1
                          1 1 0 1 1
                          1 1 1 0 1
                          1 1 1 1 0)
   /PRINT=FREQ RESID ESTIM
   /DESIGN FATHER, SON.
```

例 2　SPSS の LOGLINEAR を用いた準完全移動モデル

表 2 準完全移動モデルによる期待値

Father's Occupation	Son's Occupation: Upper Nonmanual	Lower Nonmanual	Upper Manual	Lower Manual	Farm	Total
Upper Nonmanual	1,414	344.00	285.54	811.49	64.96	2,920
Lower Nonmanual	419.43	524	321.81	914.55	73.21	2,253
Upper Manual	754.79	697.68	856	1,645.78	131.74	4,086
Lower Manual	934.37	863.67	716.88	3,325	163.09	6,003
Farm	578.41	534.64	443.77	1,261.18	1,832	4,650
Total	4,101	2,964	2,624	7,958	2,265	19,912

L^2=683.34; df=11; p< .01.

```
Observed, Expected Frequencies and Residuals.
      Factor           Code           OBS count  EXP count  Residual  Std Resid
  FATHER                 1
   SON                    1                  .0         .0       .00        .00
   SON                    2               521.0      344.0    176.98       9.54
   SON                    3               302.0      285.6     16.45        .97
   SON                    4               643.0      811.5   -168.49      -5.91
   SON                    5                40.0       65.0    -24.96      -3.10
  FATHER                 2
   SON                    1               724.0      419.4    304.55      14.87
   SON                    2                  .0         .0       .00        .00
   SON                    3               254.0      321.8    -67.82      -3.78
   SON                    4               703.0      914.6   -211.55      -7.00
   SON                    5                48.0       73.2    -25.21      -2.95
  FATHER                 3
   SON                    1               798.0      754.8     43.18       1.57
   SON                    2               648.0      697.7    -49.70      -1.88
   SON                    3                  .0         .0       .00        .00
   SON                    4              1676.0     1645.8     30.21        .74
   SON                    5               108.0      131.7    -23.75      -2.07
  FATHER                 4
   SON                    1               756.0      934.3   -178.32      -5.83
   SON                    2               914.0      863.6     50.38       1.71
   SON                    3               771.0      716.8     54.15       2.02
   SON                    4                  .0         .0       .00        .00
   SON                    5               237.0      163.1     73.92       5.79
  FATHER                 5
   SON                    1               409.0      578.4   -169.42      -7.04
   SON                    2               357.0      534.7   -177.65      -7.68
   SON                    3               441.0      443.8     -2.78       -.13
   SON                    4              1611.0     1261.2    349.83       9.85
   SON                    5                  .0         .0       .00        .00
- - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - -
Goodness-of-fit test statistics
    Likelihood ratio chi square =   683.34316  DF (UNADJUSTED) = 16  P =  .000
                                               DF (ADJUSTED)   = 11  P =  .000
             Pearson chi square =   720.73549  DF (UNADJUSTED) = 16  P =  .000
                                               DF (ADJUSTED)   = 11  P =  .000
```

例 3　SPSS の Loglinear コマンドを用いた準完全移動モデルのアウトプット（部分）

対角線上のセルにだけ割り当てられた交互作用項を独立変数に加えたモデルを適用することによっても、準完全移動モデルを実行できる。

```
1.  COMPUTE I=1.
2.  DO REPEAT VAR=DI1 TO DI5.
3.  COMPUTE VAR=(FATHER=I AND SON=I).
4.  COMPUTE I=I+1.
5.  END REPEAT.
6.  GENLOG FATHER SON WITH DI1 DI2 DI3 DI4 DI5
        /PRINT FREQ ESTIM
        /PLOT NONE
        /DESIGN FATHER SON DI1 DI2 DI3 DI4 DI5.
```

例 4　SPSS の GENLOG コマンドを用いた準完全移動モデル

表 3 準完全移動モデルにおける対角線上のセルの交互作用パラメーター

Father's Occupation	Son's Occupation Upper Nonmanual	Lower Nonmanual	Upper Manual	Lower Manual	Farm
Upper Nonmanual	1.335				
Lower Nonmanual		0.301			
Upper Manual			0.391		
Lower Manual				0.490	
Farm					2.899

```
DATA ;
SET HOUT ;
GOODMAN=1 ;
IF FATHER=SON THEN GOODMAN=0 ;
PROC GENMOD ;
CLASS FATHER SON;
WEIGHT GOODMAN ;
MODEL COUNT=FATHER SON /DIST=POI LINK=LOG PREDICTED ;
TITLE 'QUASI-PERFECT MOBILITY' ;
RUN ;
```

例 5　SAS の GENMOD プロシジャーを用いた準完全移動モデル

c.トポロジカルモデルまたはレヴェルモデル

この方法は、Featherman と Hauser が提唱したものだが、アイデアは、交互作用項をいくつかのグループにまとめる点にある。下に示したマトリックスにあるように、Featherman と Hauser は、交互作用項、すなわち、父親の職業と息子の職業の凝集性を 5 つのグループに分けた。ただし、このグルーピング変数は順序変数ではなく、1 から 5 までの値の大きさは意味を持たない。

表 4 Featherman & Hauser によるレヴェルマトリックス

	Upper Nonmanual	Lower Nonmanual	Upper Manual	Lower Manual	Farm
Upper Nonmanual	2	4	5	5	5
Lower Nonmanual	3	4	5	5	5
Upper Manual	5	5	5	5	5
Lower Manual	5	5	5	4	4
Farm	5	5	5	4	1

SASのGENMODプロシジャーを用いてFHJモデルを実行するには、次の例にあるように、このマトリックスに相応する変数を独立変数に加えればよい。SPSSを用いるときは、準完全移動モデルと同様に、グルーピングレヴェルに対応するダミー変数を作成し、これを独立変数に加える。ただし、5つのレヴェルのうち、一つは基準変数となるので、作成するダミー変数の数は4である。

```
DATA ;
SET HOUT ;
IF FATHER=1 AND SON=1 THEN FH=2 ;
IF FATHER=1 AND SON=2 THEN FH=4 ;
IF FATHER=1 AND SON>2 THEN FH=5 ;
IF FATHER=2 AND SON=1 THEN FH=3 ;
IF FATHER=2 AND SON=2 THEN FH=4 ;
IF FATHER=2 AND SON>2 THEN FH=5 ;
IF FATHER=3                THEN FH=5 ;
IF FATHER=4 AND SON<4 THEN FH=5 ;
IF FATHER=4 AND SON>3 THEN FH=4 ;
IF FATHER=5 AND SON<4 THEN FH=5 ;
IF FATHER=5 AND SON=4 THEN FH=4 ;
IF FATHER=5 AND SON=5 THEN FH=1 ;
PROC GENMOD ;
   CLASS FATHER SON FH ;
   MODEL COUNT=FATHER SON FH /DIST=POI LINK=LOG PREDICTED ;
   TITLE "FEATHERMAN-HAUSER MODEL" ;
RUN;
```

```
DO REPEAT VAR1=FH1 TO FH4.
COMPUTE VAR1=0.
END REPEAT.
IF (FATHER=5 & SON=5) FH1=1.
IF (FATHER=1 & SON=1) FH2=1.
IF (FATHER=2 & SON=1) FH3=1.
IF (FATHER=1 & SON=2) FH4=1.
IF (FATHER=2 & SON=2) FH4=1.
IF (FATHER=4 & SON=4) FH4=1.
IF (FATHER=4 & SON=5) FH4=1.
IF (FATHER=5 & SON=4) FH4=1.
GENLOG FATHER SON WITH FH1 FH2 FH3 FH4
/PRINT FREQ ESTIM
/PLOT NONE
/DESIGN FATHER SON FH1 FH2 FH3 FH4.
```

例 6　SAS（上）、SPSS（下）を用いたトポロジカルモデル

表 5 FH モデルによる期待値

Father’s Occupation	Son’s Occupation: Upper Nonmanual	Lower Nonmanual	Upper Manual	Lower Manual	Farm
Upper Nonmanual	1414.00	515.08	297.43	649.06	44.41
Lower Nonmanual	724.00	522.95	301.97	658.97	45.09
Upper Manual	754.23	740.04	777.93	1697.62	116.16
Lower Manual	812.27	796.98	837.79	3328.23	227.73
Farm	396.35	388.89	408.80	1624.01	1832.00

L^2=66.57; df=12; p< .01.

表 6 FH モデルの交互作用パラメーター

Father’s Occupation	Son’s Occupation Upper Nonmanual	Lower Nonmanual	Upper Manual	Lower Manual	Farm
Upper Nonmanual	1.590	0.599			
Lower Nonmanual	0.905	0.599			
Upper Manual					
Lower Manual				0.599	0.599
Farm				0.599	3.402

c. レヴェルモデルを用いた様々な分析モデルの表現

SPSSを用いてトポロジカルモデルを実行するには、それぞれのレヴェルに対応した交互作用項を設定し、それらを独立変数に加えた。この方法は、準完全移動モデルにも当てはめることができた。実は、移動表を分析するための多くのモデルはレヴェルモデルを用いて表現できる。表8には、Hout（1983）で社会的距離モデル（Social Distance Models）と呼ばれている、準完全移動モデルを拡張した幾つかのモデルについて、指定すべきパラメーターが示されている。

表7 社会的距離モデルのパラメーター（Hout (1983) p.30）

1				
	1			
		1		
			1	
				1

QPM-C

1	2			
2	1	2		
	2	1	2	
		2	1	2
			2	1

SD-C

1	2			
3	1	2		
	3	1	2	
		3	1	2
			3	1

D-C

1				
	2			
		3		
			4	
				5

QPM

1	6			
6	2	6		
	6	3	6	
		6	4	6
			6	5

QSD-C

1	6			
7	2	6		
	7	3	6	
		7	4	6
			7	5

QD-C

1	6			
6	2	7		
	7	3	8	
		8	4	9
			9	5

QSD

1	6			
7	2	8		
	9	3	10	
		11	4	12
			13	5

QD

SASのGENMODプロシジャーを用いて準完全移動モデルを実行するには、前に示したセルにウェイトをかける方法とは別に、トポロジカルモデルと同様のマトリックスを独立変数に加えるという方法がある。次の例に示されている様に、QIというマトリックスを行変数、列変数とは別のカテゴリー変数として独立変数に加えることによって準完全移動モデルを適用できる。このQIというマトリックスは、上の表でQPMモデルで示されているマトリックスと同等[1]。

[1] このマトリックスでは対角線以外のセルに 6 を割り当てている。これは、GENMOD プロシジャーでは、もっとも値の大きなカテゴリーのパラメーターを 0 とすることを考慮したため。標準化の仕方がソフトによって異なるので注意が必要。例えば、Powers & Xie（2000）では、準完全移動モデ

```
DATA ;
SET HOUT ;
/* DEFINITION OF PARAMETERS FOR QUASI-INDEPENCE MODEL */
IF FATHER=1 AND SON =1 THEN QI=1 ;
IF FATHER=1 AND SON >1 THEN QI=6 ;
IF FATHER=2 AND SON <2 THEN QI=6 ;
IF FATHER=2 AND SON =2 THEN QI=2 ;
IF FATHER=2 AND SON >2 THEN QI=6 ;
IF FATHER=3 AND SON <3 THEN QI=6 ;
IF FATHER=3 AND SON =3 THEN QI=3 ;
IF FATHER=3 AND SON >3 THEN QI=6 ;
IF FATHER=4 AND SON <4 THEN QI=6 ;
IF FATHER=4 AND SON =4 THEN QI=4 ;
IF FATHER=4 AND SON >4 THEN QI=6 ;
IF FATHER=5 AND SON <5 THEN QI=6 ;
IF FATHER=5 AND SON =5 THEN QI=5 ;
PROC GENMOD ;
CLASS FATHER SON QI ;
MODEL COUNT=FATHER SON QI /DIST=POI LINK=LOG PREDICTED ;
RUN ;
```

例 7　SAS を用いた準完全移動モデル（2）

d. Symmetry, Quasi-Symmetry モデル

Symmetry モデルでは、対角線を挟んで左下の部分のセル、右下の部分のセルに同じ交互作用を仮定する。レヴェルモデルと同様の表現を用いれば、Symmetry モデルを行うには、下の様な対角線を挟んで対称なマトリックスを適用すればよい。

1	2	3	4	5
2	6	7	8	9
3	7	10	11	12
4	8	11	13	14
5	9	12	14	15

表 8 に Symmetry モデルを適用した結果が示されているが、期待値の周辺度数が観察値と一致していないことに注意しなければならない。これは、実行例をみればわかる様に、Symmetry モデルでは、対角線を挟んで対象な交互作用だけを独立変数とし、行変数、列変数がモデルに入っていないことによる。

ルを対角線以外のセルに 1、対角線上のセルには 2 から 6 を当てはめるマトリックスで表現しているが、これは筆者等がもっとも値の小さいセルを基準カテゴリーとするソフト（GLIM 等）を使用しているためだと思われる。

表 8 Symmetry モデルによる分析結果

	Son's Occupation					
Father's Occupation	Upper Non-manual	Lower Non-manual	Upper Manual	Lower Manual	Farm	Total
Upper Nonmanual	1414.0	622.5	550.0	699.5	224.5	3510.5
Lower Nonmanual	622.5	524.0	451.0	808.5	202.5	2608.5
Upper Manual	550.0	451.0	856.0	1224.0	274.5	3355.5
Lower Manual	699.5	808.5	1224.0	3325.0	924.0	6981.0
Farm	224.5	202.5	274.5	924.0	1832.0	3457.5
Total	3510.5	2608.5	3355.5	6981.0	3457.5	19913.0

	L^2	df	p
Symmetry Model	2804.87	10	＜0.01

```
DATA ;
SET HOUT ;
IF SON=1 AND FATHER=1 THEN SYMM=1 ;
IF SON=1 AND FATHER=2 THEN SYMM=2 ;
IF SON=1 AND FATHER=3 THEN SYMM=3 ;
IF SON=1 AND FATHER=4 THEN SYMM=4 ;
IF SON=1 AND FATHER=5 THEN SYMM=5 ;
IF SON=2 AND FATHER=1 THEN SYMM=2 ;
IF SON=2 AND FATHER=2 THEN SYMM=6 ;
IF SON=2 AND FATHER=3 THEN SYMM=7 ;
IF SON=2 AND FATHER=4 THEN SYMM=8 ;
IF SON=2 AND FATHER=5 THEN SYMM=9 ;
IF SON=3 AND FATHER=1 THEN SYMM=3 ;
IF SON=3 AND FATHER=2 THEN SYMM=7 ;
IF SON=3 AND FATHER=3 THEN SYMM=10;
IF SON=3 AND FATHER=4 THEN SYMM=11;
IF SON=3 AND FATHER=5 THEN SYMM=12;
IF SON=4 AND FATHER=1 THEN SYMM=4 ;
IF SON=4 AND FATHER=2 THEN SYMM=8 ;
IF SON=4 AND FATHER=3 THEN SYMM=11;
IF SON=4 AND FATHER=4 THEN SYMM=13;
IF SON=4 AND FATHER=5 THEN SYMM=14;
IF SON=5 AND FATHER=1 THEN SYMM=5 ;
IF SON=5 AND FATHER=2 THEN SYMM=9 ;
IF SON=5 AND FATHER=3 THEN SYMM=12;
IF SON=5 AND FATHER=4 THEN SYMM=14;
IF SON=5 AND FATHER=5 THEN SYMM=15;
PROC GENMOD ;
CLASS SYMM ;
MODEL COUNT=SYMM /DIST=POI LINK=LOG PREDICTED ;
TITLE "SYMMETRY MODEL" ;
RUN;
```

例 8 SAS による Symmetry モデルの実行例

期待値と観測値の周辺分布が一致しない Symmetry モデルは、たとえデータにフィットしたとしてもあまり現実的ではない。期待値と観測値の周辺分布が一致するモデルは Quasi-Symmetry モデルと呼ばれる。下の例に示されているように、Quasi-Symmetry モデルでは、マトリックス変数に加えて行変数、列変数も独立変数となる。

```
PROC GENMOD ;
CLASS FATHER SON SYMM ;
MODEL COUNT=FATHER SON SYMM /DIST=POI LINK=LOG PREDICTED ;
TITLE "QUASI-SYMMETRY MODEL" ;
RUN ;
```

例 9 SAS による Quasi-Symmetry モデルの実行例

表 9 Quasi-Symmetry モデルの結果

Father's Occupation	Son's Occupation: Upper Non-manual	Lower Non-manual	Upper Manual	Lower Manual	Farm	Total
Upper Nonmanual	1414.0	575.9	289.1	599.0	42.1	2920
Lower Nonmanual	669.1	524.0	264.2	752.3	43.4	2253
Upper Manual	810.9	637.8	856.0	1658.0	123.4	4086
Lower Manual	800.0	864.7	789.2	3325.0	224.1	6003
Farm	406.9	361.6	425.6	1624.0	1832.0	4650
Total	4101	2964	2624	7958	2265	19912

	L^2	df	p
Quasi-Symmetry Model	27.45	6	<0.01

4. 順序変数を仮定するモデル

a. Uniform Association Model

ここで、移動元の階層的地位を独立変数とし、通常の回帰分析と同様に、移動元（ここでは父親の階層的地位）が 1 単位上昇すれば、移動先（息子の階層的地位）がある割合だけ上昇すると考える。すると、ある移動元から移動先のある地位につくものと一つ下の地位につくもののオッヅは次のようになる。

$$\Phi_{ij} = \log(F_{ij} / F_{i,j+1})$$
$$= \lambda_j - \lambda_{j+1} + b_i$$

ここで、b_i は、移動元の階層的地位が 1 単位上昇したときの効果を示す。すなわち、順序変数としての各階層的地位の間の差を同一とし、1 単位の上昇は同じ効果を持つと仮定する。このようなモデルを Uniform Association モデルという。このモデルを通常の対数線形モデルの式で表すと、次のようになる。

```
DATA ;
SET HOUT ;
COV1=SON ;
COV2=FATHER ;
RC=FATHER*SON ;

PROC GENMOD;
CLASS FATHER SON ;
MODEL COUNT=FATHER SON RC/DIST=POI LINK=LOG PREDICTED ;
RUN ;
```

例 10 SAS による Uniform Association モデルの実行例

$$\log F_{ij} = \mu + \lambda_i^R + \lambda_j^C + b_{ij}$$

SASのGENMODプロシジャーを用いて、Uniform Associationモデルを実行するには下の例にある様に、新しい変数（例ではRC）を作成して独立変数に加えればよい。

Uniform Association モデルを適用した結果は表 10 に示した。ここで、例 11 の出力結果にあるパラメーターについて確認をしておこう。例えば、父職と息子の職業の双方が Upper Nonmanual である場合、すなわち 1 行・1 列のセルについて、パラメータを前出の式に代入すると次のようになる。

$$
\begin{aligned}
\log(F_{11}) &= \mu + \lambda_{i-1}^{R} + \lambda_{j-1}^{C} + b_{ij} \\
&= 0.1123 + 2.6724 + 4.1953 + 0.2693 \\
&= 7.2493
\end{aligned}
$$

EXP(7.2943)　1407.07なので、表にある期待値を得ることができる。

表 10 Uniform Association Model による期待値

Father's Occupation	Son's Occupation: Upper Nonmanual	Lower Nonmanual	Upper Manual	Lower Manual	Farm	Total
Upper Nonmanual	1407.07	647.02	320.36	483.31	62.24	2920
Lower Nonmanual	778.96	468.87	303.89	600.12	101.16	2253
Upper Manual	894.95	705.14	598.23	1546.45	341.23	4086
Lower Manual	734.60	757.64	841.38	2874.06	822.33	6030
Farm	285.42	385.33	560.14	2481.07	938.05	4650
Total	4101	2964	2624	7985	2265	19939

L^2 ＝2280.69　　*df*＝15　P<0.001

Analysis Of Parameter Estimates

Parameter		DF	Estimate	Standard Error	Wald 95% Confidence Limits		Chi-Square	Pr > ChiSq
Intercept		1	0.1123	0.1163	-0.1156	0.3403	0.93	0.3341
Father	1	1	2.6724	0.0595	2.5558	2.7889	2020.18	<.0001
Father	2	1	1.8118	0.0533	1.7073	1.9162	1155.92	<.0001
Father	3	1	1.6813	0.0410	1.6010	1.7617	1680.89	<.0001
Father	4	1	1.2146	0.0273	1.1611	1.2682	1977.86	<.0001
Father	5	0	0.0000	0.0000	0.0000	0.0000	.	.
Son	1	1	4.1953	0.0699	4.0584	4.3323	3604.00	<.0001
Son	2	1	3.1492	0.0610	3.0297	3.2687	2667.55	<.0001
Son	3	1	2.1770	0.0488	2.0813	2.2726	1990.75	<.0001
Son	4	1	2.3189	0.0318	2.2566	2.3812	5323.62	<.0001
Son	5	0	0.0000	0.0000	0.0000	0.0000	.	.
RC		1	0.2693	0.0048	0.2598	0.2787	3141.16	<.0001
Scale		0	1.0000	0.0000	1.0000	1.0000		

例 11　SAS による Uniform Association モデルのアウトプット（部分）

b. Row Effect Model、Column Effect Model、Row Column Effect Model

Uniform association モデルは、各セルに同一の交互作用効果を設定したモデルだが、次に、列または行、及びその両方を考慮して交互作用効果を設定するモデルを取り上げる。これらは、それぞれ、Row Effect Model Column Effect Model Row Column Effect Model と呼ばれる。

最初に Row Effect Model を取り上げる。Row Effect Model は以下のように表現できる。

$$\log F_{ij} = \mu + \lambda_i^R + \lambda_j^C + \tau_i(\nu_j - \bar{\nu})$$

ここで、式の最後の項のうち τ_i は、行効果（Row Effect）を示しており、列の順序変数間の差が行のカテゴリー毎にどのように違うかを示す。$\bar{\nu}$ は基準変数を表し、ANOVA コーディングの時は 0、DUMMY コーディングの時はいずれかのカテゴリーを用いる。下の例では、行の最大値、すなわち、農業を基準変数としている。SAS で Row Effect モデルを実行するには、先に示した”COV1”と行変数である FATHER の交互作用項を独立変数に加えればよい。SPSS の GENLOG または LOGLINEAR コマンドを用いる場合は、変数 COV1 を共変量として定義し、DESIGN 以下に、行変数、列変数に加え、”FATHER BY COV1”という項を加えればよい。

同様に、Column Effect モデルは、上の式の行と列を入れ替えた

$$\log F_{ij} = \mu + \lambda_i^R + \lambda_j^C + \tau_j(\nu_i - \bar{\nu})$$

という式で表され、”SON BY COV2"という項を加えれば実行できる。

Row Column Effect モデルは、行と列のいずれも順序変数の時に用いられ次の式に示される。

$$\log F_{ij} = \mu + \lambda_i^R + \lambda_j^C + \tau_{ij}(\nu_i - \bar{\nu})(\nu_j - \bar{\nu})$$

```
PROC GENMOD;
  CLASS FATHER SON ;
  MODEL COUNT=FATHER SON FATHER*COV1 /DIST=POI LINK=LOG PREDICTED ;
  TITLE 'ROW EFFECT MODEL'
PROC GENMOD;
  CLASS FATHER SON ;
  MODEL COUNT=FATHER SON SON*COV2 /DIST=POI LINK=LOG PREDICTED ;
  TITLE 'COLUMN EFFECT MODEL'
PROC GENMOD;
  CLASS FATHER SON ;
  MODEL COUNT=FATHER SON FATHER*RC SON*RC/DIST=POI LINK=LOG PREDICTED ;
  TITLE 'ROW-COLUMN EFFECT MODEL' ;
RUN ;
```

例12　SASを用いたROW EFFECT MODEL、COLUMN EFFECT MODEL、ROW-COLUMN EFFECT MODELの実行例

```
*ROW EFFECT MODEL.
GENLOG FATHER SON WITH COV1 COV2 RC
  /PRINT FREQ ESTIM
  /PLOT NONE
  /DESIGN FATHER SON FATHER BY COV1.
*COLUMN EFFECT MODEL.
GENLOG FATHER SON WITH COV1 COV2 RC
  /PRINT FREQ ESTIM
  /PLOT NONE
  /DESIGN FATHER SON SON BY COV2.
*ROW-COLUMN EFFECT MODEL.
GENLOG FATHER SON WITH COV1 COV2 RC
  /PRINT FREQ ESTIM
  /PLOT NONE
  /DESIGN FATHER SON FATHER BY RC SON BY RC.
```

例 13　SPSS を用いた ROW EFFECT MODEL、COLUMN EFFECT MODEL、ROW-COLUMN EFFECT MODEL の実行例

表12 Row Effect、Column Effect、Row Column Effectモデルの結果

	Row Effect Model					
	Son's Occupation					
Father's Occupation	Upper Non-manual	Lower Non-manual	Upper Manual	Lower Manual	Farm	Total
Upper Nonmanual	1330.2	643.2	339.1	537.5	70.0	2920
Lower Nonmanual	763.1	464.7	308.4	615.8	101.0	2253
Upper Manual	912.4	710.3	602.7	1538.1	322.5	4086
Lower Manual	909.6	853.7	873.3	2687.2	679.3	6003
Farm	185.8	292.1	500.5	2579.4	1092.2	4650
Total	4101	2964	2624	7958	2265	19912
	Column Effect Model					
	Son's Occupation					
Father's Occupation	Upper Non-manual	Lower Non-manual	Upper Manual	Lower Manual	Farm	Total
Upper Nonmanual	1310.7	568.7	350	688.9	1.7	2920
Lower Nonmanual	750.6	434.6	316.4	743.5	7.8	2252.9
Upper Manual	930.9	719.2	619.4	1737.6	78.9	4086
Lower Manual	823.4	848.9	865	2896.5	569.1	6002.9
Farm	285.3	392.5	473.2	1891.4	1607.5	4649.9
Total	4101	2963.9	2624	7957.9	2265	19911.7
	Row-Column Effect Model					
	Son's Occupation					
Father's Occupation	Upper Non-manual	Lower Non-manual	Upper Manual	Lower Manual	Farm	Total
Upper Nonmanual	1348.0	579.0	345.3	646.2	1.4	2920
Lower Nonmanual	746.5	437.4	319.1	742.5	7.5	2253
Upper Manual	842.2	686.7	624.9	1845.2	86.9	4086
Lower Manual	864.0	864.1	864.7	2855.1	555.0	6003
Farm	300.3	396.7	470.0	1868.9	1614.1	4650
Total	4101	2964	2624	7958	2265	19912

	L^2	df	p
Row Effect Model	2080.2	12	＜0.01
Column Effect Model	903.8	12	＜0.01
Row-Column Effect Model	877.8	9	＜0.01

c. 意識調査データ分析での順序変数の対数線形分析

表13人種と男女間役割意識（Ishii-Kuntz (1994), p.20）

Ethnicity	Gender Ideology			
	Traditional	Moderate	Liberal	Total
Black	73	67	77	217
White	464	510	443	1417
Hispanic	106	47	23	176
Total	643	624	543	1810

	L^2	*df*	*P*
Independent Model	57.64	4	.000
Row Effect	2.70	2	.259

順序変数を考慮したモデルは、意識調査データにも適用できる。上に示した例は、人種別に男女の役割についてどのような意識を持っているかを集計した結果である。行変数の人種は名目変数、列変数の意識はTraditional、Moderate、Liberalの三段階に分けられた順序変数である。行効果モデルを適用してみると、自由度2で有意水準が0.259となり、データに適合するモデルを得ることができた。

意識調査の分析における関心は、意識がどのように規定されるかだが、Ishii-Kuntz（1994）によれば、行効果のパラメーターは、黒人では0.292、白人では0.241、ヒスパニックでは-0.533という結果が得られている[2]。すなわち、黒人がもっともリベラルな意識を持ち、白人がそれに続き、ヒスパニックはもっとも保守的な意識を持っている。重要な点は、これらのパラメーターによって人種間での意識の差が数値化されることである。例えば、黒人とヒスパニックのパラメーターの差は0.825だが、これを指数化すると、EXP(0.825)=2.282となる。すなわち、ヒスパニックと比べて黒人の方が2.282倍の割合で、よりリベラルな意識を持っていることが示された。これに比べ、黒人と白人の差はEXP(0.292-0.241)=1.052で、わずかな差である。

表14では、もう一つの順序変数、信仰心が導入されている。

表14　人種、信仰心、性別役割意識の相互関連（Ishii-Kuntz 1994 p.37）

Ethnicity	Religiosity	Gender Ideology			
		Traditional	Moderate	Liberal	Total
Black	Religious	58	45	49	152
	Moderately	11	17	21	49
	Not Religious	3	4	7	14
White	Religious	317	242	145	704
	Moderately	105	157	148	410
	Not Religious	41	109	150	300
Hispanic	Religious	83	24	8	115
	Moderately	16	17	13	46
	Not Religious	7	6	2	15
Total		641	621	543	1,805

[2] このパラメーターは、パラメーターをすべて足すと 0 になる様に標準化する ANOVA コーディングを用いて算出されたと思われる。SAS の GENMOD プロシジャー、SPSS の GEMLOG コマンド等ではもっとも値の大きいカテゴリーを基準変数とする。この場合のパラメーターは、黒人:0.8257、白人:0.7257、ヒスパニック:0.000 となる。どちらの場合も相対的な距離は変わらない。

```
DATA LIST FREE /ETHNO RELIGION GENDER FREQ.
WEIGHT BY FREQ.
BEGIN DATA
   1   1   1    58   1   1   2    45   1   1   3    49
   1   2   1    11   1   2   2    17   1   2   3    21
   1   3   1     3   1   3   2     4   1   3   3     7
   2   1   1   317   2   1   2   242   2   1   3   145
   2   2   1   105   2   2   2   157   2   2   3   148
   2   3   1    41   2   3   2   109   2   3   3   150
   3   1   1    83   3   1   2    24   3   1   3     8
   3   2   1    16   3   2   2    17   3   2   3    13
   3   3   1     7   3   3   2     6   3   3   3     2
END DATA.
COMPUTE COV1=ETHNO.
COMPUTE COV2=RELIGION.
COMPUTE COV3=GENDER.
COMPUTE ER=COV1*COV2.
COMPUTE RG=COV2*COV3.
COMPUTE EG=COV1*COV3.
EXECUTE.

LOGLINEAR ETHNO(1,3) RELIGION(1,3) GENDER(1,3) WITH COV1 TO COV3 ER RG EG
  /PRINT FREQ ESTIM
  /PLOT NONE
  /DESIGN ETHNO RELIGION GENDER.
LOGLINEAR ETHNO(1,3) RELIGION(1,3) GENDER(1,3) WITH COV1 TO COV3 ER RG EG
  /PRINT FREQ ESTIM
  /PLOT NONE
  /DESIGN ETHNO RELIGION GENDER RG.
LOGLINEAR ETHNO(1,3) RELIGION(1,3) GENDER(1,3) WITH COV1 TO COV3 ER RG EG
  /PRINT FREQ ESTIM
  /PLOT NONE
  /DESIGN ETHNO RELIGION GENDER ETHNO BY COV3.
LOGLINEAR ETHNO(1,3) RELIGION(1,3) GENDER(1,3) WITH COV1 TO COV3 ER RG EG
  /PRINT FREQ ESTIM
  /PLOT NONE
  /DESIGN ETHNO RELIGION GENDER ETHNO BY COV3 RG.
LOGLINEAR ETHNO(1,3) RELIGION(1,3) GENDER(1,3) WITH COV1 TO COV3 ER RG EG
  /PRINT FREQ ESTIM
  /PLOT NONE
  /DESIGN ETHNO RELIGION GENDER ETHNO BY COV2 RG.
LOGLINEAR ETHNO(1,3) RELIGION(1,3) GENDER(1,3) WITH COV1 TO COV3 ER RG EG
  /PRINT FREQ ESTIM
  /PLOT NONE
  /DESIGN ETHNO RELIGION GENDER ETHNO BY COV2 ETHNO BY COV3.
LOGLINEAR ETHNO(1,3) RELIGION(1,3) GENDER(1,3) WITH COV1 TO COV3 ER RG EG
  /PRINT FREQ ESTIM
  /PLOT NONE
  /DESIGN ETHNO RELIGION GENDER ETHNO BY COV2 ETHNO BY COV3 RG.
```

例 14 SPSS を用いた 3 次元の順序カテゴリーの対数線形分析例

表15人種、信仰、性役割分業に関する独立モデル、
連関モデルの結果（Ishii-Kuntz 1994 p.39）

Model	Fitted Marginals	L^2	df	prob
1	[E][R][G]	288.46	20	.000
2	[E][RG]	129.40	19	.000
3	[R][EG]	233.25	18	.000
4	[G][ER]	228.87	18	.000
5	[EG][RG]	74.19	17	.001
6	[ER][RG]	69.80	17	.001
7	[ER][EG]	173.66	16	.000
8	[ER][EG][RG]	20.44	15	.156

Partial Association Modelの定義式

$$\log F_{ijk} = \mu + \lambda_i^E + \lambda_j^R + \lambda_k^G + \tau_i^{ER}(\nu_j - \bar{\nu}) + \tau_i^{EG}(w_k - \bar{w}) + \beta^{RG}(\nu_j - \bar{\nu})(w_k - \bar{w})$$

表16 Partial Association Modelの結果（Ishii-kuntz 1994 p.41）

パラメーター	パラメーターの内容	τ or β	EXP(τ) or EXP(β)
τ_{1j}^{ER}	The odds of not being religious (as opposed to moderately religious), or the odds of being moderately religious (as opposed to religious), for blacks.	-0.412	0.662
τ_{2j}^{ER}	The odds of not being religious (as opposed to moderately religious), or the odds of being moderately religious (as opposed to religious), for whites.	0.368	1.445
τ_{3j}^{ER}	The odds of not being religious (as opposed to moderately religious), or the odds of being moderately religious (as opposed to religious), for Hispanics.	0.044	1.045
τ_{1k}^{EG}	The odds of holding liberal (as opposed to moderately liberal), or the odds of holding moderately liberal (as opposed to traditional) attitudes, for blacks.	0.391	1.478
τ_{2k}^{EG}	The odds of holding liberal (as opposed to moderately liberal), or the odds of holding moderately liberal (as opposed to traditional) attitudes, for whites	0.148	1.160
τ_{3k}^{ER}	The odds of holding liberal (as opposed to moderately liberal), or the odds of holding moderately liberal (as opposed to traditional) attitudes, for Hispanics.	-0.539	0.583
β^{RG}	The odds of holding liberal (as opposed to moderately liberal), or the odds of holding moderately liberal (as opposed to traditional) attitudes for those with lower level of religiosity.	0.503	1.654

5. 対数線型モデルにおける注意

① セル内の度数

従来のカイ二乗検定を行う際と同様､全体のサンプル数が少なすぎると統計的信頼性が落ちる。特に、対数線型モデルの場合、3重以上のクロス表の分析に力を発揮するが、変数が大きくなるにつれ、急速にセル内の度数は減少する。通常、それぞれのセル内の度数が5未満になってはならないといわれている。

また、度数がゼロのセルがある場合には注意が必要（log0は定義できない）である。通常の統計ソフトは度数がゼロのセルがある場合､非常に小さな値で置き換えて計算するが､あらかじめ、すべてのセルに0.5を加えるという方法がよく用いられる。

② トポロジカルモデルの非決定性

対数線型モデルは、研究者の関心にしたがって様々なモデルをたてることができることが特徴だが、時には、現実的、理論的な意味のないモデルがデータに適合してしまうことがある。特に、トポロジカルモデルの場合、統計的な適合度が同等で、全く異なったレヴェル設定がいくつも存在することが多い。この場合、どれが「正しい」モデルであるかを決定する基準はなく、トポロジカルモデルの非決定性（Indeterminacy）と呼ばれている。

Hout（1983）では、FHモデルとは異なるマトリックスを適用して、FHモデルと同じ適合度、期待値が得られる例が紹介されている。また、鹿又（1992）は、今田（1989）が、1955年から1985年までの4回のSSM調査をトポロジカルモデルを用いて分析し、世代間の移動に変化は見いだせないとしたのに対し、適合度がほぼ同じで、今田の結論とは異なった解釈ができるマトリックスを提示している。

対数線形分析のためのソフトウェア

① SPSS　SPSSには、他にHILOGLINEARというコマンドがあるが、このコマンドでは共変量を定義できない。
② SAS　SASにはGENMODプロシジャーの他に、CATMODプロシジャーで対数線形分析を行える。この場合、基準変数は数値の最も小さいカテゴリーに固定されている。操作性は、GENMODプロシジャーより複雑。
③ GLIM　Generalized Linear Interactive Modelingの略。英国王立統計協会が開発。多く使用されているが、日本では手に入りにくい（？）。
④ LEM　オランダのJ. Vermuntが開発。下記のWebページでプログラム本体、PDF形式のマニュアル、プログラム例などを無料でダウンロードできる（但し、すべて英語）。
http://www.kub.nl/faculteiten/fsw/organisatie/departementen/mto/software2.html

参考文献（ABC順）

Agresti, Alan 1990 *Categorical Data Analysis.* New York, NY: Wiley.
1996 *An Introduction to Categorical Data Analysis.* New York, NY: Wiley
Featherman, David and Robert M. Hauser 1978 *Opportunity and Change.* New York, NY: Academic Press
Hout, Michael 1983 *Mobility Tables.* Beverly Hills, CA: Sage
今田高俊1989『社会階層と政治』東京大学出版会
Ishi-Kuntz Masako 1994 *Ordinal Log-Linear Models* Beverly Hills, CA: Sage
鹿又伸夫1992「階層・移動研究の袋小路と活路」理論と方法 vol. 7, No. 1 pp. 1-18
Powers, Daniel A. and Yu Xie 2000 *Statistical Methods for Categorical Data Analysis.* San Diego, CA: Academic Press