

SAS FREQプロシジャによるCochran-Mantel-Haenszel 統計量の使い方

岸本, 淳司
SASインスティテュートジャパン統計解析研究室

<https://doi.org/10.15017/1470162>

出版情報 : 九州大学大型計算機センター広報. 23 (6), pp.663-682, 1990-11-26. 九州大学大型計算機センター
バージョン :
権利関係 :

SAS FREQプロシジャによる Cochran-Mantel-Haenszel 統計量の使い方

岸本 淳司*

はじめに

カテゴリカルデータの解析は、現実の場面からの要求により、その需要が急速に増大している。SASシステムでは、ランダム化に基礎を置くカテゴリカルデータの解析手法群はFREQプロシジャで対応している。SASバージョン5の改訂で、Cochran-Mantel-Haenszel 統計量を求めるためのオプションがFREQプロシジャに追加された。改訂当時(1985年)、その重要性に気づいた人はほとんどいなかったのではなからうか。筆者自身、1988年に W.Stanish (G.Kochの弟子)の講義を受けて、初めてその意味を理解することができた。それはつまり、独立性の検定・ノンパラメトリック検定から生存分析の検定に至るまでの統一的な基本概念なのである。その後、九州大学の柳川先生の著書「離散多変量データの解析(1986)」に、Mantel-Haenszel 流の解析法について見事に述べられていることがわかった。残念ながら同書には些細な計算間違いと定理の間違いが存在するが、それを越えて素晴らしい本であることは間違いない。本稿は、いわば同書の内容をSASで実行するための例題集である。

本稿の内容は、1988年10月に行われたSASユーザー会第2ワーキンググループの発表資料に基づいている。また、1990年6月に九州大学大型計算機センターの招きにより行ったSASによるカテゴリカル解析セミナーの資料の一部にも用いた。今回九大大型計算機センター広報に掲載するにあたって、九大センターでのSAS環境に対応した書換えを行っている。

平成2年9月25日 受理

* SASインスティテュートジャパン 統計解析研究室

1. $r \times k$ 表の解析

1.1 カテゴリーに順序がない $r \times k$ 表の検定

表1.1 は、消化性潰瘍の患者1796例と対照6087例について、血液型の分布を調べたものである（柳川 1986 p.83より）。

表1.1 消化性潰瘍と血液型

	血液型			計
	O	A	B	
対照群	2892	2625	570	6087
潰瘍群	983	679	134	1796

対照-潰瘍の各群の標本数が固定であるから、2個の多項モデルが想定できる。検定すべき帰無仮説は、各群において血液型の分布が等しいことである。総標本数が固定のときに1つの多項分布を考え2要因間の独立性を検定する場合も含め、帰無仮説はオッズ比を用いて次のように統一的に表現できる。

$$H_0: \phi_{ij} = 1, \quad i=2,3,\dots,r, \quad j=2,3,\dots,k$$

対立仮説は帰無仮説が成り立たないことである。

さて、 $r \times k$ 表に想定される確率モデルが何であっても、オッズ比に関する条件付推論は周辺和を固定した多重超幾何分布について考えればよい。同分布より、頻度の期待値と分散共分散が理論的に求まり、条件付カイ自乗統計量を計算することができる。この統計量は、ピアソンのカイ自乗統計量の $(1-1/n)$ 倍に等しく、近似的に自由度 $(r-1) \times (k-1)$ のカイ自乗分布をする。SASでは、この値をCMH (Cochran-Mantel-Haenszel) 統計量のType 3として求めることができる。具体的には、FREQプロシジャのTABLESステートメント中にCMHというオプションをつける。すると3種類のCMH統計量が得られるが、そのうち"GENERAL ASSOCIATION"というラベルのついた値がカテゴリーの順序を考えないときの検定統計量である。

次ページに上部にSASのコーディング例を示し、それによる出力を下部に枠に囲って示す。以下の例でも同様に示す。CHISQ オプションにより出力された統計表の中のピアソンのカイ自乗値 29.697と (n が大きいので) 非常に近い値 29.694 が求めるCMH統計量である。この値は高度に有意である。ゆえに2つの群で血液型の分布は異なると推論できる。

```

DATA EX1_1;
  INPUT GROUP $ BLOOD $ WT @@;
  CARDS;
CONT 0 2892  CONT A 2625  CONT B 570
CASE 0 983  CASE A 679  CASE B 134
;

PROC FREQ DATA=EX1_1;
  WEIGHT WT;
  TABLES GROUP*BLOOD / CMH CHISQ NOPRINT;
RUN;
    
```

STATISTICS FOR TABLE OF GROUP BY BLOOD

STATISTIC	DF	VALUE	PROB
CHI-SQUARE	2	29.697	0.000
LIKELIHOOD RATIO CHI-SQUARE	2	29.796	0.000
MANTEL-HAENSZEL CHI-SQUARE	1	24.083	0.000
PHI		0.061	
CONTINGENCY COEFFICIENT		0.061	
CRAMER'S V		0.061	

SAMPLE SIZE = 7883

SUMMARY STATISTICS FOR GROUP BY BLOOD

COCHRAN-MANTEL-HAENSZEL STATISTICS (BASED ON TABLE SCORES)

STATISTIC	ALTERNATIVE HYPOTHESIS	DF	VALUE	PROB
1	NONZERO CORRELATION	1	24.083	0.000
2	ROW MEAN SCORES DIFFER	1	24.083	0.000
3	GENERAL ASSOCIATION	2	29.694	0.000

TOTAL SAMPLE SIZE = 7883

1. 2 一方のカテゴリーに順序がある $r \times k$ 表の検定 (1)

表1.2 は、ジョージア州沿岸地域の肺癌の事例対照研究データである (柳川 1986 p.84より)。喫煙のカテゴリーに量的順序があるので、「喫煙量が増加するにつれて肺癌のリスクが増加しているか?」という「用量・反応関係」を調べたい。

表1.2 肺癌の事例対照研究データ

	喫 煙			計
	-	+	++	
対照	105	72	25	203
肺癌	13	56	32	101

対立仮説は、

$$H_1: 1 \leq \phi_{12} \leq \dots \leq \phi_{1k} \quad \text{または} \\ 1 \geq \phi_{12} \geq \dots \geq \phi_{1k}$$

喫煙量のカテゴリーに対応するスコア c_j に、たとえば、 $1, 2, \dots, k$ と自然数を割り当てて、次の関係を満たしていると仮定すると、

$$\ln(\phi_{ij}) = \beta_i(c_j - c_1), \quad i=2, 3, \dots, r, \quad j=2, 3, \dots, k$$

H_1 は次のように表わせる。

$$K_1: \beta_i \neq 0 \quad \text{であるような } i \text{ が少なくとも } 1 \text{ つ存在する } (i=2, 3, \dots, r)$$

このとき条件付カイ自乗統計量は、近似的に自由度 $r-1$ のカイ自乗分布に従い、CMH統計量の Type2として求めることができる (ただし、 $r=2$ の場合には、Type1と同じである)。具体的には、FREQプロシジャのTABLESステートメントで、列変数 (後に指定する方) に順序カテゴリーを表わす変数を指定し、さらにCMHオプションを指定する。3種類出力される統計量の内、“ROW MEAN SCORES DIFFER” というラベルがついている2番目の統計量が一方のカテゴリーに順序があるとしたときのCMH統計量である。この例の場合カイ自乗値は41.486で、高度に有意である。

順序カテゴリーに与えるスコアは、デフォルトでは、カテゴリー変数が文字型の場合は自然数が割り当てられ、数値型の場合はその変数の値となる。このスコアのつけ方を明示的に指示するためには SCORES=TABLE というオプションをTABLESステートメントに追加する。また、順序カテゴリーの値が文字型の場合、カテゴリーの出現順序で大小関係を示すために、PROC FREQ ステートメントにORDER=DATAというオプションをつける。

CMH統計量Type2は、通常の分散分析のF値の関数である。従って、次ページのプログラムで示したように、CMH流解析で与えたのと同じスコアを用いた分散分析と同じ結論となる。

```

DATA EX1_2;
  INPUT GROUP $ SMOKE WT @@;
  CARDS;
CONT 1 105   CONT 2 72   CONT 3 26
CASE 1 13   CASE 2 56   CASE 3 32
;

PROC FREQ DATA=EX1_2;
  WEIGHT WT;
  TABLES GROUP*SMOKE / CMH SCORES=TABLE NOPRINT;
RUN;

PROC ANOVA DATA=EX1_2;
  FREQ WT;
  CLASS GROUP;
  MODEL SMOKE = GROUP;
RUN;

```

COCHRAN-MANTEL-HAENSZEL STATISTICS (BASED ON TABLE SCORES)

STATISTIC	ALTERNATIVE HYPOTHESIS	DF	VALUE	PROB
1	NONZERO CORRELATION	1	41.486	0.000
2	ROW MEAN SCORES DIFFER	1	41.486	0.000
3	GENERAL ASSOCIATION	2	45.067	0.000

TOTAL SAMPLE SIZE = 304

ANALYSIS OF VARIANCE PROCEDURE

SOURCE	DF	ANOVA SS	F VALUE	PR > F
GROUP	1	22.47599453	47.91	0.0001

1.3 一方のカテゴリに順序がある $r \times k$ 表の検定 (2)

表1.3 は、ある地区でランダムに選んだ 400人の男子中学生を対象に、他人からのいじめの被害者意識と、他人へのいじめ志向とにかんするアンケートの集計結果である (柳川 1986 p.84)。

表1.3 いじめの調査

		いじめ志向			計
		なし	少しあり	強くあり	
被害者意識	なし	32	21	42	95
	少しあり	38	42	76	156
	強くあり	27	40	82	149
		97	103	200	400

総標本数が固定されているので、1つの多項分布モデルが想定される。しかしオッズ比に関する推論は 1.2節と同じである。

被害者意識のカテゴリに関しては順序関係を見せず、いじめ志向に関しては順序カテゴリとして分析してみる。ここでは、自然数ではなく、いわゆるウィルコクソン・スコアを割り当てることにする。つまり、被害者意識のカテゴリが変わると、いじめ志向に割り当てたスコアの平均値が変化するか、という問いになる。

FREQ プロシジャで CMH 統計量を求めるときに、このウィルコクソン・スコアを割り当てるためには、TABLES ステートメントに SCORES=RANK というオプションを指定する。そして、1.2 節で説明したように、一方のカテゴリのみに順序があるとした場合は、Type 2 の CMH 統計量を見ればよい。

ウィルコクソン・スコアを割り当てた CMH 統計量 Type 2 は、結局クラスカル・ワリス検定と等価になる。次ページの例では、NPAR1WAY プロシジャの出力と比較することによって、この事実を確認することができる。

```

DATA EX1_3;
  INPUT A B WT @@;
  DO N=1 TO WT;
    OUTPUT;
  END;
CARDS;
1 1 32  1 2 21  1 3 42
2 1 38  2 2 42  2 3 76
3 1 27  3 2 40  3 3 82
;

PROC FREQ DATA=EX1_3;
  TABLES A*B / CMH SCORES=RANK NOPRINT;
RUN;

PROC NPARIWAY DATA=EX1_3 WILCOXON;
  CLASS A;
  VAR B;
RUN;

```

COCHRAN-MANTEL-HAENSZEL STATISTICS (BASED ON RANK SCORES)

STATISTIC	ALTERNATIVE HYPOTHESIS	DF	VALUE	PROB
1	NONZERO CORRELATION	1	5.326	0.021
2	ROW MEAN SCORES DIFFER	2	5.350	0.069
3	GENERAL ASSOCIATION	4	7.862	0.097

TOTAL SAMPLE SIZE = 400

KRUSKAL-WALLIS TEST (CHI-SQUARE APPROXIMATION)
CHISQ= 5.35 DF= 2 PROB > CHISQ=0.0689

1. 4 両方のカテゴリーに順序がある $r \times k$ 表の検定

表1.3 のデータに対して、今度は被害者意識の方も順序カテゴリーと考えて解析してみよう。対立仮説は、

$$H_1 : 1 \leq \phi_{12} \leq \dots \leq \phi_{1k} \text{ または } 1 \geq \phi_{12} \geq \dots \geq \phi_{1k}, \quad i=2, 3, \dots, r \\ 1 \leq \phi_{2j} \leq \dots \leq \phi_{rj} \text{ または } 1 \geq \phi_{2j} \geq \dots \geq \phi_{rj}, \quad j=2, 3, \dots, k$$

2つのカテゴリー変数にそれぞれスコア d_1, d_2, \dots, d_r ($d_1 \leq d_2 \leq \dots \leq d_r$) ; c_1, c_2, \dots, c_k ($c_1 \leq c_2 \leq \dots \leq c_k$)を割り付け、次の関係式を仮定する。

$$\ln(\phi_{ij}) = \beta (d_i - d_1)(c_j - c_1)$$

すると H_1 は次のように表わされる。

$$K_1 : \beta = 0$$

このとき条件付カイ自乗統計量は近似的に自由度1のカイ自乗分布に従い、SASではこれをCMH統計量のType1と呼んでいる。SASのFREQプロシジャで出力される3つのCMH統計量のうち、“NONZERO CORRELATION”というラベルがついているのがそれである。SCORES=RANKのオプションを指定すると、行変数・列変数ともにウィルコクソン・スコアを割り当てることになる。この例でのカイ自乗値は、次ページに示すように5.326であり、p値は両側対立仮説に対応して0.021であるから、被害者意識が大きい人ほどいじめ志向が大きいという仮説をサポートする結果である。

CMH統計量Type1は、FREQプロシジャにCHISQオプションを指定したとき出力されるMantel-Haenszel統計量のことである。これは、次ページの出力で確かめられる。また、周辺スコアを与えたときのピアソンの相関係数の自乗 $\times(N-1)$ に一致する。

```
PROC FREQ DATA=EX1_3;
  TABLES A*B / CMH SCORES=RANK CHISQ NOPRINT;
RUN;
```

STATISTICS FOR TABLE OF A BY B

STATISTIC	DF	VALUE	PROB
CHI-SQUARE	4	7.882	0.096
LIKELIHOOD RATIO CHI-SQUARE	4	7.750	0.101
<u>MH CHI-SQUARE (RANK SCORES)</u>	<u>1</u>	<u>5.326</u>	<u>0.021</u>
PHI		0.140	
CONTINGENCY COEFFICIENT		0.139	
CRAMER'S V		0.099	

SAMPLE SIZE = 400

SUMMARY STATISTICS FOR A BY B

COCHRAN-MANTEL-HAENSZEL STATISTICS (BASED ON RANK SCORES)

STATISTIC	ALTERNATIVE HYPOTHESIS	DF	VALUE	PROB
<u>1</u>	<u>NONZERO CORRELATION</u>	<u>1</u>	<u>5.326</u>	<u>0.021</u>
2	ROW MEAN SCORES DIFFER	2	5.350	0.069
3	GENERAL ASSOCIATION	4	7.862	0.097

TOTAL SAMPLE SIZE = 400

2. 層別連関表の解析

2.1 層別 $r \times k$ 表の解析

表2.1 は、十二指腸潰瘍の手術法 (A: 0%摘出, B: 25%摘出, C: 50%摘出, D: 75%摘出) 別に、手術直後のダンピング症候群の程度 (無, 軽, 重) を、4つの病院において調べた結果である (柳川 1986 p.113 より)。

表2.1 ダンピング症候群

病院		1			2			3			4		
		無	軽	重	無	軽	重	無	軽	重	無	軽	重
手術法	A	23	7	2	18	6	1	8	6	3	12	9	1
	B	23	10	5	18	6	2	12	4	4	15	3	2
	C	20	13	5	13	13	2	11	6	2	14	8	3
	D	24	10	6	9	15	12	7	7	4	13	6	4

手術法とダンピング症候群との関係を、病院の影響を排除して判定したい。帰無仮説は「全ての層において行変数と列変数との間に連関はない (オッズ比が等しい)」になる。このような層別解析においてCMH統計量は特に活用できる。

CMH統計量の帰無仮説が棄却されると、「いずれかの層で連関がある」ことになる。しかし、各層において連関のパターンが反対になっていると、CMHはそのような連関を検出しにくいので注意が必要である。

層別のCMH統計量をFREQプロシジャで求めるには、TABLESステートメント中に3つ以上の変数を列挙する方法による。すると、最後の2つの変数が行変数と列変数とみなされ、それより前の変数群は層別変数とみなされる。

カテゴリーの順序性を考慮して解析する場合は、各カテゴリーにスコアを割り付けるわけだが、層別解析の場合各層のウィルコクソン・スコアをその層のデータの総数で割ったリジット型のスコア (SCORES=RIDIT) を使うこともある。

次ページのプログラムでは、今までに紹介した3種類のスコアの付け方をすべて試している。出力例にはリジット型スコアのもののみ示す。いずれのスコア法でも、Type 3 → Type 2 → Type 1 と検出の方向性を絞るほどp値が小さくなっているため、摘出の程度を大きくするほどダンピング症候群の程度が大きくなるという相関関係がうかがわれる。

```

DATA EX2_1;
  DO OP='A','B','C','D';
    DO HOSPITAL=1 TO 4;
      DO DUMP='N','L','H';
        INPUT WT @;
        OUTPUT;
      END;
    END;
  END;
CARDS;
23 7 2 18 6 1 8 6 3 12 9 1
23 10 5 18 6 2 12 4 4 15 3 2
20 13 5 13 13 2 11 6 2 14 8 3
24 10 6 9 15 2 7 7 4 13 6 4
;

PROC FREQ DATA=EX2_1 ORDER=DATA;
  WEIGHT WT;
  TABLES HOSPITAL*OP*DUMP / CMH SCORES=TABLE NOPRINT;
  TABLES HOSPITAL*OP*DUMP / CMH SCORES=RANK NOPRINT;
  TABLES HOSPITAL*OP*DUMP / CMH SCORES=RIDIT NOPRINT;
RUN;

```

SUMMARY STATISTICS FOR OP BY DUMP CONTROLLING FOR HOSPITAL				
COCHRAN-MANTEL-HAENZSEL STATISTICS (BASED ON RIDIT SCORES)				
STATISTIC	ALTERNATIVE HYPOTHESIS	DF	VALUE	PROB
1	NONZERO CORRELATION	1	6.915	0.009
2	ROW MEAN SCORES DIFFER	3	7.634	0.054
3	GENERAL ASSOCIATION	6	10.598	0.102
TOTAL SAMPLE SIZE = 417				

2. 2 層別 2×2表における相対リスク

2×2分割表の場合、リスク因子の有無と疾患の有無との関係の指標として相対リスクがよく用いられる。コーホート研究の場合は標本比率を用いて自然に推定できるが、事例対象研究の場合はオッズ比を用いて推定することになる。さて、層別の2×2表においては、各層に共通と仮定される調整済みの相対リスクが興味の対象になる。表2.2は、表1.2のデータをさらに就労状況によって層別したものである（造船所で1年以上働いたことがある(+)、ない(-)、建設工業で1年以上働いたことがある(+)、ない(-)、柳川 1986 p.111 より）。

表2.2 肺癌事例対象研究データ(2)

層1: 喫煙(-)、建設(-)		層2: 喫煙(-)、建設(+)		層3: 喫煙(+)、建設(-)		
	造船		造船		造船	
	+	-	+	-	+	-
肺癌	1	4	3	5	6	25
対照	10	68	5	22	5	27

層4: 喫煙(+)、建設(+)		層5: 喫煙(++), 建設(+)		層6: 喫煙(++), 建設(+)		
	造船		造船		造船	
	+	-	+	-	+	-
肺癌	8	17	1	23	1	7
対照	7	23	1	20	0	5

喫煙と建設業就労を交絡因子と考え、層別して影響を排除し肺癌と造船業就労との関係（相対リスク）を推定したい。この調整済み相対リスクの計算には2つの方法がある。

Mantel-Haenszel 推定量

- ゼロの度数があっても問題なく計算できる
- 信頼区間を計算するよい方法が（FRE Qプロシジャ作成当時）知られていなかったので、test-basedの区間（分散推定に偏連関はないと仮定する）を求める

ロジット推定量

- ゼロの度数があると計算できないので、その場合は全セルに 0.5を足す
- 信頼区間が正確に定義できる

各層におけるオッズ比の等質性を検定するには、Breslow-Day 統計量が用いられる。層別2×2表にCMHオプションを適用すると、研究法・推定法別の相対リスク推定値とその信頼区間、さらに Breslow-Day統計量が得られる。相対リスクの推定値を求めるためには、リスク因子有、疾患有の情報が頻度表の左上にくるようにしておく必要がある。そのためには ORDER=DATA オプションが有効である。また、信頼区間の有意水準を調整には ALPHA= オプションが用いられる。

このデータの場合、Breslow-Day 検定の結果より、調整済みオッズ比を計算する根拠があることがわかり、その値は約1.6であるが、下側信頼限界が1を越えて小さいので、造船業就労と肺癌との関係については明確ではない。

```

DATA EX2_2;
  DO GROUP='CASE', 'CONT';
    DO STRATA=1 TO 6;
      DO RISK='+', '-';
        INPUT WT @;
        OUTPUT;
      END;
    END;
  END;
CARDS;
  1 4 3 5 6 25 8 17 1 23 1 7
  10 68 5 22 5 27 7 23 1 20 0 5
;

PROC FREQ DATA=EX2_2 ORDER=DATA;
  WEIGHT WT;
  TABLES STRATA*GROUP*RISK / CMH NOPRINT;
RUN;

```

ESTIMATES OF THE COMMON RELATIVE RISK (ROW1/ROW2)				
95%				
TYPE OF STUDY	METHOD	VALUE	CONFIDENCE BOUNDS	
CASE-CONTROL	MANTEL-HAENSZEL	1.614	0.792	3.291
(ODDS RATIO)	LOGIT *	1.587	0.786	3.203
COHORT	MANTEL-HAENSZEL	1.454	0.833	2.538
(COL1 RISK)	LOGIT *	1.455	0.848	2.496
COHORT	MANTEL-HAENSZEL	0.923	0.819	1.040
(COL2 RISK)	LOGIT	0.959	0.873	1.053

THE CONFIDENCE BOUNDS FOR THE M-H ESTIMATES ARE TEST-BASED.
 * DENOTES THAT THE LOGIT ESTIMATORS USE A CORRECTION
 OF 0.5 IN EVERY CELL OF THOSE TABLES THAT CONTAIN A ZERO.

BRESLOW-DAY TEST FOR HOMOGENEITY OF THE ODDS RATIOS
 CHI-SQUARE = 1.047 DF = 5 PROB = 0.959

3. 層別ノンパラメトリック解析

3.1 Friedmanの検定

4人の生徒に4日にわたって同じテストをしたところ表3.1の結果が得られた。日による差はあるか(岩原 1957 p.228 より)。

表3.1 4人に同一のテストを4日行った結果

	第1日	第2日	第3日	第4日
1	80	112	90	102
生 2	126	131	125	130
徒 3	73	73	70	89
4	100	110	123	122

得点は順位にしか意味がないとする。とすると、この問題には、生徒のブロックにより対応づけられた Friedman の検定が適用できる。

さて、Friedmanのカイ自乗は、ブロック要因を層別としてウィルコクソン・スコアを割り当てた Type2の CMH統計量にほかならない。従って、今までの知識により、FREQプロシジャを使って簡単に求めることができる。

このデータの場合、カイ自乗値は $3.9(p=0.272)$ であり、日の間に有意差は認められない。

```

DATA EX3_1;
  DO BLOCK=1 TO 4;
    DO DAY=1 TO 4;
      INPUT TEST 0;
      OUTPUT;
    END;
  END;
  CARDS;
80 112 90 102
126 131 125 130
73 72 70 89
100 110 123 122
;

PROC FREQ DATA=EX3_1;
  TABLES BLOCK*DAY*TEST / CMH SCORES=RANK NOPRINT;
RUN;

```

SUMMARY STATISTICS FOR DAY BY TEST
CONTROLLING FOR BLOCK

COCHRAN-MANTEL-HAENSZEL STATISTICS (BASED ON RANK SCORES)

STATISTIC	ALTERNATIVE HYPOTHESIS	DF	VALUE	PROB
1	NONZERO CORRELATION	1	1.470	0.225
2	ROW MEAN SCORES DIFFER	3	3.900	0.272
3	GENERAL ASSOCIATION	45		

1 OR MORE STATISTICS NOT COMPUTED--SINGULAR COVARIANCE MATRIX

TOTAL SAMPLE SIZE = 16

3.2 McNemarの検定

層別の極端な例として、1対1のマッチドペアの解析がある。たとえば、リハビリ前後の体力テストの結果が表3.2aのように与えられていたとする。

表3.2a リハビリ前後の体力テスト

		リハビリ後	
		OK	NG
リハビリ前	OK	29	9
	NG	23	19

ここで関心があるのは、リハビリ前後のテストの結果の連関ではない。リハビリと体力テストとの連関である。従って、基本的には次の形式の連関表が興味の対象になる。

表3.2b 基本的連関表

		体力テスト	
		OK	NG
リハビリ	前		
	後		

表3.2aのデータは、表3.2bの形式で次のことを表わす。

表3.2c データの再表現

1	OK	NG	2	OK	NG	3	OK	NG	4	OK	NG
前	1	0	前	1	0	前	0	1	前	0	1
後	1	0	後	0	1	後	1	0	後	0	1
×29名			×9名			×23名			×19名		

2つの測定値を持つ小分割表は1人の被験者のデータであるから、これをブロック要因と考え層別したCMH統計量を求めると、これは McNemarのカイ自乗統計量に他ならない(ただし、連続性補正はしていない)。表3.2cで、1と4の型のデータはCMH統計量に影響を与えないので省略してもよい。具体的には、次のようなデータを用意する。

表3.2d CMH用データ

リハビリ	体力テスト	被験者
前	OK	1
後	NG	1
前	OK	2
後	NG	2

```

DATA EX3_2;
  DO SS=1 TO 9;
    REHA='BEFORE'; TEST='OK'; OUTPUT;
    REHA='AFTER '; TEST='NG'; OUTPUT;
  END;
  DO SS=10 TO 32;
    REHA='BEFORE'; TEST='NG'; OUTPUT;
    REHA='AFTER '; TEST='OK'; OUTPUT;
  END;
RUN;

PROC FREQ DATA=EX3_2;
  TABLES SS*REHA*TEST / CMH NOPRINT;
RUN;

```

SUMMARY STATISTICS FOR REHA BY TEST
CONTROLLING FOR SS

COCHRAN-MANTEL-HAENSZEL STATISTICS (BASED ON TABLE SCORES)

STATISTIC	ALTERNATIVE HYPOTHESIS	DF	VALUE	PROB
1	NONZERO CORRELATION	1	6.125	0.013
2	ROW MEAN SCORES DIFFER	1	6.125	0.013
3	GENERAL ASSOCIATION	1	6.125	0.013

3.3 Stuartの検定

McNemar の検定を3×2表以上に拡張すると Stuart の検定になる。たとえば、十二指腸潰瘍の診断のためのレ線フィルム 100枚を2施設A, Bにランダムな順で渡して、その出来具合を上中下の3段階に分類してもらったとする(佐久間 1977 p.126 より)。

表3.3a レ線フィルムの評価

A\B	上	中	下
上	19	17	7
中	7	26	5
下	3	12	4

ここで関心があるのは、2施設間の評価の連関ではなく、施設と評価との間の連関である。そして1枚1枚のレ線フィルムがブロックを構成している。ゆえに 3.2節と全く同様に考え、

表3.3b データの再表現(2)

	上	中	下
A	1	0	0
B	1	0	0

のフィルムが19枚

	上	中	下
A	1	0	0
B	0	1	0

のフィルムが17枚

.....

レ線フィルムで層別した施設×評価のCMH統計量のType 3を求めると、それがStuartの検定で用いるカイ自乗統計量となる。表3.3aの対角要素はやはりCMH統計量に影響を与えない。

表3.3c CMH用データ(2)

施設	評価	フィルム
A	上	1
B	上	1
A	上	2
B	上	2
•	•	•
•	•	•

次ページのプログラムは表3.3aのデータから表3.3cの形式のデータを作成するために、やや技巧的なコーディングを行っている。カイ自乗値は 7.87(p=0.02) なので、施設と評価の間には連関があるといえそうである。

```

DATA EX3_3;
  RETAIN N 1;
  INPUT A $ B $ WT @@;
  DO BLOCK=N TO N+WT-1;
    HOSPITAL='A'; EVAL=A; OUTPUT;
    HOSPITAL='B'; EVAL=B; OUTPUT;
  END;
  N=BLOCK;
  CARDS;
+ + 19 + +- 17 + - 7
+- + 7 +- +- 26 +- - 5
- + 3 - +- 12 - - 4
;

PROC FREQ DATA=EX3_3;
  TABLES BLOCK*HOSPITAL*EVAL / CMH NOPRINT;
RUN;

```

SUMMARY STATISTICS FOR HOSPITAL BY EVAL
CONTROLLING FOR BLOCK

COCHRAN-MANTEL-HAENSZEL STATISTICS (BASED ON TABLE SCORES)

STATISTIC	ALTERNATIVE HYPOTHESIS	DF	VALUE	PROB
1	NONZERO CORRELATION	1	1.494	0.222
2	ROW MEAN SCORES DIFFER	1	1.494	0.222
3	GENERAL ASSOCIATION	2	7.870	0.020

TOTAL SAMPLE SIZE = 200

おわりに

本稿の理論的背景について勉強されたい方は、ぜひ柳川先生の著書（文献4.）を参照していただきたい。最後に、本稿のセンター広報への掲載を勧めてくださった九大大型計算機センターの桜井先生に謝意を表す。

参考文献

1. 岩原新九郎 1957 教育と心理のための推計学 日本文化科学社
2. 佐久間 昭 1977 薬効評価 I 東京大学出版会
3. Stanish, W.M. 1988 Categorical Data Analysis Course Notes, SAS Institute, Cary: NC.
4. 柳川 堯 1986 離散多変量データの解析 共立出版