

Cochran-Mantel-Haenszel検定の概要とSASでの実装方法

○石本りさ

(イーピーエス株式会社)

Overview of the Cochran-Mantel-Haenszel Test and Its Implementation in SAS

Risa Ishimoto

EPS Corporation

要旨

Cochran-Mantel-Haenszel (CMH) 検定は複数の独立した分割表の層別解析を行うために多くの研究で広く用いられているが、初学者に向けて CMH 検定の解説を行っている文献は少ない。そのため、本稿では初学者が初めて CMH 検定を行う際の理解を助けることを目的とし、CMH 統計量の導出式について記述し、SAS の FREQ Procedure 及び LOGISTIC Procedure を用いた CMH 統計量の算出方法について紹介する。

キーワード : Cochran-Mantel-Haenszel 検定, proc freq, proc logistic

1. 緒言

臨床試験において収集されるデータにはカテゴリカルデータも多く、そのデータを解析する機会も多くある。カテゴリカルデータの解析において基本的な手法の1つには Pearson のカイ二乗検定があり、この検定においてクロス集計表における行変数及び列変数の独立性を調べることができる。しかし、Pearson のカイ二乗検定単独では交絡因子による影響を考慮した評価をすることはできない。CMH 検定では一つのクロス集計表を交絡因子で層別化することにより複数のクロス集計表を作成し、交絡因子の影響を調整した上で行変数及び列変数の関連性を調べることができる。広く用いられている検定手法の一つであるものの、初学者向けに解説を行っている文献は少ない。

そのため、本稿では初学者が初めて CMH 検定を行う際の理解を助けることを目的とし、CMH 統計量の導出式について記述し、SAS の FREQ Procedure 及び LOGISTIC Procedure を用いた CMH 統計量の算出方法について紹介する。さらに CMH 統計量の導出式への理解を深めるために、データステップのみを用いた CMH 統計量の導出についても解説する。

2. CMH 統計量の導出方法

前述したとおり、CMH 検定では一つのクロス集計表を交絡因子で層別化することにより複数のクロス集計表を作成し、層別解析を行う。イメージが捉えやすくなるよう、層化変数に性別および年齢を設定し、CMH 検定を行った際の概略図を Fig2.1 に示した。

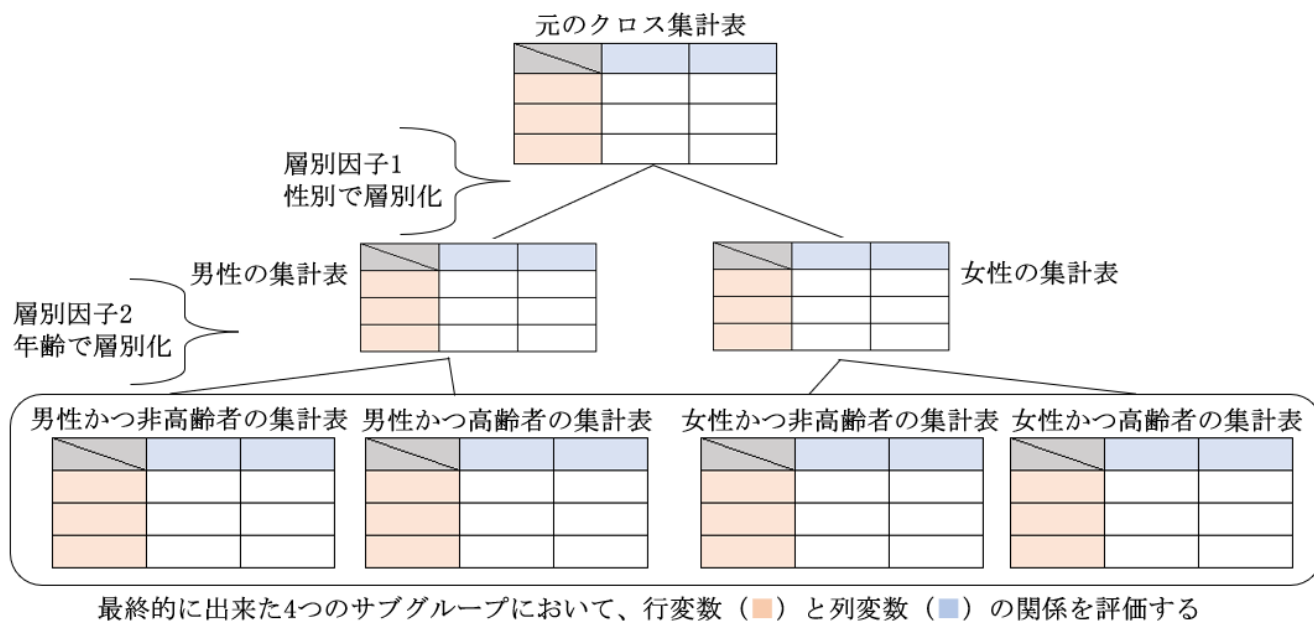


Fig 2.1 性別（2層）・年齢カテゴリ（2層）で CMH 検定を行った際の概略図

今回は 2×2 の分割表を想定し、Table 2.1 を用いて CMH 統計量の導出について最初に解説する。Table 2.1 は一つの分割表に対し層が k 個あると仮定した際の、 i 層目の分割表を示している。

Table 2.1 クロス集計表のサンプル

		Response		Total
		Better	Same	
Treatment	Active	a_i	b_i	m_{1i}
	Placebo	c_i	d_i	m_{2i}
Total		n_{1i}	n_{2i}	N_i

まず、CMH 検定ではオッズ比は全ての層で等質であるという仮定を置く。また、帰無仮説は「共通オッズ比は 1 である」とし、対立仮説は「共通オッズ比は 1 ではない」とする。オッズ比はある群における事象の起こりやすさに対する別の群における事象の起こりやすさであるため、オッズ比が 1 であれば、群と事象の起こりやすさに関連はないと判断することができる。つまり、CMH 検定における帰無仮説は「いずれの層においても行変数及び列変数の間に連関性はない」と言い換えることができる。

この帰無仮説の下で Table 2.1 における a_i は超幾何分布に従うことが知られている。超幾何分布とは、二種

類の性質で構成される集団から、無作為かつ非復元抽出を行ったときに、特定の性質のものがいくつあるかという確率を表す離散型確率分布である。例えば、母集団のサイズを N 、その中で特定の性質の要素数を M 、無作為に抽出される標本サイズを n 、標本中における特定の性質の要素数を a とする。その際、確率 $P(X = a)$ は以下の式で表される。

$$P(X = a) = \frac{{}^M C_k \times {}^{N-M} C_{n-a}}{{}^N C_n}$$

$$\max\{0, n - (N - M)\} \leq a \leq \min\{n, M\}$$

また、この確率変数 X における期待値と分散は下記になる。

$$E(X) = n \times \frac{M}{N} = \frac{nM}{N}$$

$$V(X) = n \times \frac{M(N - M)}{N^2} \times \frac{N - n}{N - 1} = \frac{nM(N - M)(N - n)}{N^2(N - 1)}$$

そのため、Table 2.1 における a_i の期待値と分散は下式にて表すことができる。

$$E(a_i) = \frac{n_{1i}m_{1i}}{N_i}$$

$$V(a_i) = \frac{n_{1i}m_{1i}n_{2i}m_{2i}}{N_i^2(N_i - 1)}$$

CMH 検定統計量は Z 統計量を二乗した値、つまりカイ二乗統計量 χ^2_{MH} で表される。すべての周辺度数を固定した条件のもとで、上記の期待値と分散を用いると算出式は以下となる。

$$\chi^2_{MH} = Z^2 = \frac{[\sum_{i=1}^k (a_i - E_i)]^2}{\sum_{i=1}^k V_i} = \frac{[\sum_{i=1}^k (a_i - n_{1i}m_{1i}/N_i)]^2}{\sum_{i=1}^k \frac{n_{1i}m_{1i}n_{2i}m_{2i}}{N_i^2(N_i - 1)}}$$

こうして導いた上記の検定統計量が自由度 1 のカイ二乗分布に従うことを利用して、CMH 検定では仮説検定を行う。導出式の複雑さから難しく見えるが、カイ二乗統計量の算出式に超幾何分布より導いた期待値と分散を代入したのみと考えるとシンプルである。また、資料によっては次式として表されていることもあるが、分子の表し方が異なるだけで内容としては上記式と同一である。

$$\chi^2_{MH} = \frac{[\sum_{i=1}^k (a_i d_i - c_i b_i)]^2}{\sum_{i=1}^k \frac{n_{1i}m_{1i}n_{2i}m_{2i}}{N_i^2(N_i - 1)}}$$

また、サンプルサイズが小さい場合、過度に有意な結果が生じるのを防ぐためにイエーツの補正を用いて

連続修正を行う場合がある。連続修正を行った場合の CMH 検定統計量の算出式は、分子から 0.5 を引いた下式となる。この補正をかけると CMH 検定統計量は小さくなるため、算出される p 値は大きくなり、統計的な有意差は生じにくくなる。

$$\chi^2_{MH} = Z^2 = \frac{[\sum_{i=1}^k (a_i - E_i)]^2}{\sum_{i=1}^k V_i} = \frac{[\sum_{i=1}^k (a_i - \frac{n_{1i}m_{1i}}{N_i} - 0.5)]^2}{\sum_{i=1}^k \frac{n_{1i}m_{1i}n_{2i}m_{2i}}{N_i^2(N_i - 1)}}$$

3. SAS による CMH 検定の実行

3.1 FREQ Procedure を用いた CMH 検定の実行方法

最初に FREQ Procedure を用いた CMH 検定の実行方法について解説する。サンプルデータとして、Table 3.1.1 のデータを用いて解説を行う。今回のサンプルデータは FREQ Procedure における CMH 検定の実行方法を解説した SAS Procedure Guide より引用した。^[1]

こちらのサンプルデータによる解析は、Treatment の Response に対する影響を調べることを目的としており、Gender で層別を行っている。

```
/* CODE 3.1 */
data Migraine;
input Gender $ Treatment $ Response $ Count @@;
datalines;
female Active Better 16   female Active Same 11
female Placebo Better 5   female Placebo Same 20
male Active Better 12    male Active Same 16
male Placebo Better 7    male Placebo Same 19
;
run;
```

Table 3.1.1 サンプルデータ

	Gender	Treatment	Response	Count
1	female	Active	Better	16
2	female	Active	Same	11
3	female	Placebo	Better	5
4	female	Placebo	Same	20
5	male	Active	Better	12
6	male	Active	Same	16
7	male	Placebo	Better	7
8	male	Placebo	Same	19

```

/* CODE 3.2 */
ods output cmh = cmh;
Proc freq data = <source dataset>;
    Tables gender * treatment * response / relrisk plots (only) = oddsratioplot (stats) cmh;
    Weight count;
run;

```

上記のように FREQ Procedure において、Tables ステートメント内に行変数、列変数及び層化変数を指定した上で CMH オプションを記述し、Weight ステートメントにて分析変数の数値の重みを設定することで CMH 検定を実行することができる。この時、左端変数が層化を指定し、二つの右端変数が表の行と列を指定することとなるため、変数を設定する際にはその点を注意する必要がある。

上記コードを実行した際、アウトプットに出力される結果が Table 3.1.1～Table 3.1.4 及び Fig 3.1.1 である。まず RELRISK オプションを設定したことにより、Table 3.1.1 内において、層別の分割表におけるオッズ比、相対リスクの値及び Wald 法で求められた 95%信頼限界が計算されている。CMH オプションにより共通オッズ比及び相対リスク推定値も計算されており、それらの結果を示しているのが Table 3.1.2 である。

また PLOTS オプションを設定しているため、gender の各水準及び全体のオッズ比とその 95%信頼限界がプロットされたグラフも出力されており、Fig 3.1.1 がその図を示している。このとき、“plots=oddsratioplot”ではなく“plots(only)=oddsratioplot”と記述することで、オッズ比プロットのみを特定してプロットを出力している。もし前者のコードで実行した場合には、オッズ比プロットに加えて相対リスクプロットも出力される。さらに、“oddsratioplot(stats)”において stats と指定していることで、プロット図内に統計量を表している。プロットをアウトプットに出力するためには、予め ods graphics オプションを on にする必要がある。

Table 3.1.1 層別変数毎のオッズ比及び相対リスク

表 1 : Treatment * Response の統計量
層別変数 : Gender=female

オッズ比と相対リスク			
統計量	値	95% 信頼限界	
オッズ比	5.8182	1.6755	20.2034
相対リスク (列 1)	2.9630	1.2740	6.8913
相対リスク (列 2)	0.5093	0.3103	0.8357

表 2 : Treatment * Response の統計量
層別変数 : Gender=male

オッズ比と相対リスク			
統計量	値	95% 信頼限界	
オッズ比	2.0357	0.6478	6.3975
相対リスク (列 1)	1.5918	0.7413	3.4180
相対リスク (列 2)	0.7820	0.5259	1.1626

Table 3.1.2 共通オッズ比及び相対リスクの推定値

共通オッズ比と相対リスク				
統計量	手法	値	95% 信頼限界	
オッズ比	Mantel-Haenszel	3.3132	1.4456	7.5934
	ロジット	3.2941	1.4182	7.6515
相対リスク (列 1)	Mantel-Haenszel	2.1636	1.2336	3.7948
	ロジット	2.1059	1.1951	3.7108
相対リスク (列 2)	Mantel-Haenszel	0.6420	0.4705	0.8761
	ロジット	0.6613	0.4852	0.9013

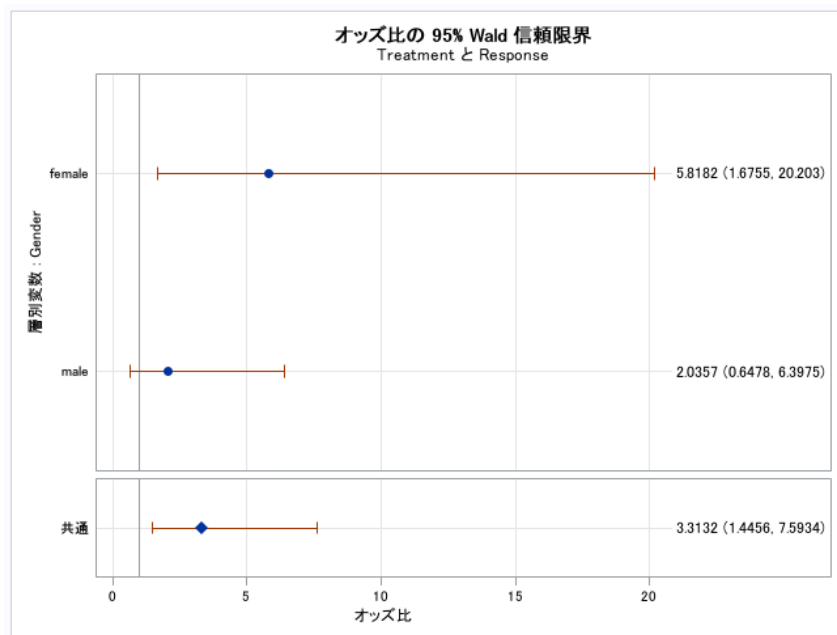


Fig 3.1.1 層別変数毎のオッズ比及び共通オッズ比のプロット図

Table 3.1.3 Cochran-Mantel-Haenszel 統計量の出力画面

Cochran-Mantel-Haenszel 統計量 (テーブルスコアに基づ)				
統計量	対立仮説	自由度	値	p 値
1	相関統計量	1	8.3052	0.0040
2	ANOVA 統計量	1	8.3052	0.0040
3	一般関連統計量	1	8.3052	0.0040

また、Table3.1.3 に示しているのが CMH 統計量の出力画面である。CMH 統計量においては、相関統計量、ANOVA 統計量及び一般関連統計量の三つの結果が表示される。相関統計量は行変数及び列変数が順序尺度である場合の CMH 統計量、ANOVA 統計量は列変数が順序尺度である場合の CMH 統計量、そして一般関連統計量は行変数及び列変数が順序尺度でない場合に用いられる CMH 統計量である。これらの統計量についていずれを使用するかについては、解析する分割表内の行や列が順序を持つかどうかで異なるが、今回のような 2×2 分割表の場合において順序関係は問題にならず、これらの三つの統計量は全て同じ値になる。¹⁾

Table 3.1.4 オッズ比等質性に対する Breslow-Day 検定の出力画面

オッズ比等質性に対する Breslow-Day 検定	
カイ 2 乗値	1.4929
自由度	1
Pr > ChiSq	0.2218

ods output 出力画面に最下段に表示される結果 (Table 3.1.4) がオッズ比等質性に対する Breslow-Day 検定の結果である。CMH 検定では 2 章で解説した通りに、オッズ比は全ての層で等質である、という仮定を置いていた。この仮定に対する検定が Breslow-Day 検定であり、帰無仮説は「各層におけるオッズ比は等しい」というものである。今回の結果では p 値が 0.2218 であり、帰無仮説は棄却されないため、オッズ比は全ての層で等質ではないとはいえないと判断できる。

3.2 LOGISTIC Procedure を用いた正確な CMH 検定の実行方法

ロジスティック回帰分析では 2 値の目的変数における特定の事象が発生する確率を、説明変数を用いて構築されたロジスティック回帰モデルより予測する。カテゴリカルデータの分析に対して用いられ、2 値アウトカムの評価ができるという点は CMH 検定と同様である。このロジスティック回帰分析において、 k 個の独立した 2×2 表の層別解析における周辺度数を固定した場合のスコア検定の結果は、CMH 検定の結果と一致することが報告されている。^[2] これに関する理論的背景については十分なロジスティック回帰分析におけるスコア検定への知識を要するため、詳細に関しては関連する参考文献を参照されたい。^{[3][4]}

今回はまず実際に 3.1 節と同様のサンプルデータを用いて LOGISTIC Procedure でロジスティック回帰分析を行い、CMH 検定の結果と一致することを確認した。CODE 3.3 が該当する LOGISTIC Procedure の実行コードである。着目している "gender" という層別変数を strata ステートメントで指定しプロシジャを実行すると、スコア検定のカイ二乗統計量は CMH 統計量と同様の "8.3052" と一致していることがわかる。ただし、LOGISTIC Procedure で出力された共通オッズ比については FREQ Procedure で出力した共通オッズ比とは一致しないため、一致するのはカイ二乗統計量と p 値のみであるという点は注意が必要である。

```

/* CODE 3.3 */
Proc logistic data=Migraine;
  Class treatment / param=ref;
  Freq count;
  Model response(Event='Better')=treatment;
  Strata gender;
  Exact treatment;
run;

```

Table 3.2.1 ロジスティック回帰モデルにおけるスコア検定統計量及び CMH 検定統計量の一致

包括的帰無仮説: BETA=0 の検定				Cochran-Mantel-Haenszel 統計量 (テーブルスコアに基づく)				
検定	カイ 2 乗値	自由度	Pr > ChiSq	統計量	対立仮説	自由度	値	p 値
尤度比	8.4831	1	0.0036	1	相関統計量	1	8.3052	0.0040
スコア	8.3052	1	0.0040	2	ANOVA 統計量	1	8.3052	0.0040
Wald	7.9947	1	0.0047	3	一般関連統計量	1	8.3052	0.0040

また、一般的に CMH 検定を行うにあたっては各層における各セルの期待度数が 5 以上となるような大規模なサンプルサイズを必要とする。このような場合において、LOGISTIC Procedure を用いて正確な検定を実行すれば、少数のサンプルサイズでも層別解析を行うことが可能である。^[5]

具体的にはロジスティック回帰分析におけるスコア検定をデータへと適用し、結果として CMH 検定の結果を導けることを利用すればよい。CODE 3.3 に記載しているように exact ステートメントを使用することで、Table 3.2.2 に示した正確な検定結果が出力される。

Table 3.2.2 exact ステートメントを指定した際の出力結果

条件付き正確確率検定				
効果	検定	統計量	p 値	
			正確	中間
Treatment	スコア	8.3052	0.0051	0.0039
	確率	0.00242	0.0051	0.0039

4. Procedure を用いない CMH 統計量の導出方法

今回は CMH 統計量導出式の理解を深めるためにも、FREQ Procedure や LOGISTIC Procedure を用いずにデータステップの中で CMH 統計量を導出することも試みた。用いるサンプルデータは同様の Table 3.1.1 である。また、SORT Procedure 及び TRANSPOSE Procedure については導出までのデータセットの変形のために用いている。

```
/* CODE 4.1 */
data Migraine;
  input Gender $ Treatment $ Response $ Count @@;
  datalines;
  female Active Better 16   female Active Same 11
  female Placebo Better 5   female Placebo Same 20
  male Active Better 12    male Active Same 16
  male Placebo Better 7    male Placebo Same 19
  ;
run;

proc sort data = Migraine;
  by Gender;
run;

proc transpose data = Migraine out = wk1 delimiter = _;
  var Count;
  by Gender;
  id treatment response;
run;
```

Table 4.1 CODE 4.1 の出力結果

	Gender	_NAME_	Active_Better	Active_Same	Placebo_Better	Placebo_Same
1	female	Count	16	11	5	20
2	male	Count	12	16	7	19

まずサンプルデータを作成し、後における計算のし易さを考慮し TRANSPOSE Procedure で数値データを横並びにしたものが Table 4.1 のデータとなる。(CODE 4.1)

その次に各群における CMH 統計量導出式内の分母と分子の値を算出する。導出式への当てはめが理解しやすくなるよう、それぞれのデータを Table 2.1 における $a_i \sim d_i$ までの記号で表した。その後に行行列の合計値と、全データの合計値を算出している。ここまでの整えば各層における導出式内の分母と分子が計算でき、Denom 変数が分母、Numer 変数が分子を表している。(CODE 4.2)

Table 4.2 CODE 4.2 の出力結果

	Gender	a	b	c	d	N
1	female	18	11	5	20	52
2	male	12	16	7	19	54

m1	m2	n1	n2	Denom	Numer
27	25	21	31	3.1864557953	5.0961538462
28	26	19	35	3.1324895825	2.1481481481

```

/* CODE 4.2 */
data wk2;
  set wk1;

  a = Active_Better; b = Active_same;
  c = Placebo_Better; d = Placebo_same;

  N = sum(of Active_ : Placebo_);
  m1 = sum(of a b); m2 = sum(of c d);
  n1 = sum(of a c); n2 = sum(of b d);

  Denom = divide(n1*m1*n2*m2, (N**3-N**2));
  Numer = a - (divide(n1*m1, N));
run;

```

次に各群において算出した分母と分子の総和を導くために、再度 TRANSPOSE Procedure を用いてデータの形を整えていく。Total 変数は female 群と male 群の分母と分子の各総和を表している。(CODE 4.3)

```

/* CODE 4.3 */
proc transpose data = wk2 out = wk3;
  var Numer Denom;
  id Gender;
run;

data wk4;
  set wk3;
  Total = female + male;
run;

proc transpose data = wk4 out = wk5;
  var Total;
run;

```

```

/* CODE 4.4 */
data wk6;
  set wk5;

  CMH = divide(Numer**2, Denom);
  prob = 1 - probchi(CMH, 1);
run;

```

VIEWTABLE: Work.Wk6					
	NAME	Denom	Numer	CMH	prob
1	Total	6.3189453779	7.2443019943	8.3051693355	0.0039532396



Cochran-Mantel-Haenszel 統計量 (テーブルスコアに基づく)				
統計量	対立仮説	自由度	値	p 値
1	相関統計量	1	8.3052	0.0040
2	ANOVA 統計量	1	8.3052	0.0040
3	一般関連統計量	1	8.3052	0.0040

Fig 4.1 データステップで導出した CMH 統計量と FREQ Procedure における出力画面

最終的に二乗した Numer 変数を Denom 変数で割ることにより、CMH 検定統計量を導いた。(CODE 4.4) CMH 統計量は 8.30516...であり、p 値は 0.003952...と計算されている。この結果は Fig 4.1 で示した通りに FREQ Procedure で求めた結果と一致しており、データステップでの再現ができた。

5. まとめ

今回は初学者向けに CMH 検定統計量の導出方法及び SAS での実行方法について解説した。CMH 検定統計量の導出式は一見複雑ではあるが、超幾何分布における期待値と分散の導出方法及びカイ二乗統計量の導出方法を理解していれば、データステップでの再現も可能な比較的シンプルな検定統計量である。本稿の内容を通じて CMH 検定の理解が深まり、実際のデータ解析に役立てれば幸いである。

また今回は 2×2 の分割表での解説のみ行ったが、 $r \times c$ ($r > 2, c > 2$) の分割表においては Table 3.1.3 に示した統計量のうち、順序尺度の有無を考慮し、どれが適切かを検討する必要がある。今後の著者の課題として、多元分割表における CMH 検定についての理解を深めていきたい。

6. 参考文献

- [1] SAS Institute Inc., Cary, NC. Base SAS® 9.4 Procedures Guide: Statistical Procedures, Fourth Edition (2015)
- [2] Day, N.E. and Byar, D.P. Testing hypotheses in case-control studies-equivalence of Mantel-Haenszel statistics and logit score tests. *Biometrics*, 35, 620-630. (1979)
- [3] 丹後 俊郎, 山岡 和枝, 高木 晴良. 新版 ロジスティック回帰分析 -SAS を利用した統計解析の実際-. 朝倉書店. (2013)
- [4] David W. Hosmer, Jr, Stanley Lemeshow, Rodney X. Sturdivant (著), 宮岡 悦良, 早川 有, 川崎 洋平, 下川 朝有 (訳). データ解析のためのロジスティック回帰モデル. 共立出版. (2017)
- [5] SAS Support, "Usage Note 32711: Computing an exact CMH test", <https://support.sas.com/kb/32/711.html>, Accessed July 28, 2024
- [6] John M. Lachin (著), 宮岡 悦良, 遠藤 輝, 黒沢 健, 下川 朝有, 寒水 孝司 (訳). 医薬データのための統計解析 原著第2版. 共立出版. (2020)
- [7] 折村 奈美. データステップを用いた統計検定の再現. SAS ユーザー総会 2024. (2024)