医学データの統計解析の基本

2つの割合の比較

朝倉こう子*・濱崎俊光*

Fundamentals of statistical analysis in biomedical research : Two-sample tests for comparing proportions

1 生物学や医学分野において、個体(例えば、培養細胞、動物あるいは被験者・患者)に対して化合物や医薬品といった処理を与えることにより惹き起こされる反応を観測することは、その物質の特性を評価する重要な役割を果たす。例えば、動物に対してある薬物の用量を反復投与し、その投与により惹き起こされる毒性発現の有無を観測することで、無影響量や無毒性量を評価することができ、これはその薬剤のヒトに対する投与量を決定するための有益な情報を与える。ヒトを対象とする医薬品の臨床試験では、患者に対するその医薬品の推奨用量を決定するための情報を得るために、例えば、医薬品を投与したときの効果発現の有無、疾患増悪の有無や疾患再発の有無、有害事象の発現の有無などの反応が観測される。

医学データの統計解析の基本の第2回(朝倉・濱崎, 2015)では、「2つの平均の比較」として処理を与えたときの個体反応が連続尺度である場合を取り扱ったが、前記のように医学研究においては個体反応が反応の有無として二値尺度で得られることも多い。そこで、第3回では個体反応が二値尺度であるデータに対する代表的な統計手法を2、3概観し、それらのデータ解析における使用上の留意点を述べる。また、事例を通して統計計算ソフトウェアでの実行の仕方を概説する。

2 第2回と同様に、本稿でも最も単純な場合として2つの処理AとBの比較を考える。ここでは、2つの処理の反応割合 p_A と p_B に違いがあるか否かに関心がある。いま、2つの処理AとBをそれぞれ m 例と n 例、合計で $N(=m+n)$ 例の個体に与え、個体の反応の有無を観測する。そして、観測の結果、処理AとBにおいてそれぞれ a 例と c 例の反応個体を観測したとする。このような過程を通して収集されたデータを要約するのに最もよく用いられるのは、表1に示す 2×2 分割表である。この 2×2 分割表

表1 2つの処理に関する反応の有無の要約: 2×2 分割表

処理	反応		計
	あり	なし	
A	a	b	m
B	c	d	n
計	r	s	N

* Koko Asakura, Toshimitsu Hamasaki
Office of Biostatistics and Data Management
Department of Advanced Medical Technology Development
National Cerebral and Cardiovascular Center
国立循環器病研究センター先進医療・治験推進部 DM/ 統計室
〒565-8565 吹田市藤白台5-7-1 E-mail: toshi.hamasaki@ncvc.go.jp

から、2つの処理AとBにおける反応割合 p_A と p_B をそれぞれ $\hat{p}_A=a/m$ と $\hat{p}_B=c/n$ で推定し、これらを比較する。2つの処理の反応割合の比較には、Pearsonのカイ2乗検定(Pearson, 1900)(以降では、「カイ2乗検定」と呼ぶ)や尤度比検定、Fisher-Irwinの直接確率計算(Fisher, 1934; Irwin, 1935)(以降では、「Fisherの直接確率計算」と呼ぶ)が代表的な統計手法であり、実際、これらは実地でよく用いられ、特に計算の容易さから、カイ2乗検定が第一選択肢であることが多い。これらの3つの統計手法のいずれも、2つの処理の反応割合の違いを測る指標として差 $p_A - p_B$ 、相対比 p_A/p_B 、あるいはオッズ比 $p_A/(1-p_A)/(p_B/(1-p_B))$ のいずれを選択しようが、指標の種類に依らず共通に使用することができる(例えば、Fleiss et al.(2003)を参照)。本稿では、主として差 $p_A - p_B$ に関する差の仮説検定を記述する。

カイ2乗検定は、標本サイズが十分に大きいときにカイ2乗検定統計量の分布が正規分布で近似できるとの性質を利用し、正規分布から計算されるp値(あるいは棄却限界値)に基づき仮説の取捨選択が行われる。しかし、標本サイズが十分に大きくなかったり、反応割合がゼロまたは1に近いといった極端な状況では、カイ2乗検定統計量の分布の正規分布への近似は良好でなく、適切に仮説を取捨選択できないことが知られている。このような問題への対処として、正規分布への近似の改善を意図したYatesの連続修正のカイ2乗検定(Yates, 1934)(以降では、「Yates修正のカイ2乗検定」と呼ぶ)やFisherの直接確率計算を代替として用いることが多くの統計学の教科書で推奨されている。実際のデータ解析では、どの程度まで標本サイズが小さい、どの程度まで反応割合がゼロあるいは1に近ければ、カイ2乗検定の代わりにYates修正のカイ2乗検定やFisherの直接確率計算を用いるべきかといった疑問が生じる。この手法選択の判断基準としてはCochran基準(Cochran, 1954)が一般的であり、後述の統計計算ソフトウェアにおいてもCochran基準にはかり、カイ2乗検定を用いることが妥当でない場合には警告が発せられる。Cochran基準によれば、いずれかのセル頻度の期待度数が5よりも小さい場合にカイ2乗検定の使用を避けることになる。また、Cochran(1954)は、 $N < 20$ であれば最も小さいセル頻度は5よりも小さくなるため、 $N < 20$ の場合はカイ2乗検定を使用することを推奨していない。

ここで、カイ2乗検定、Yates修正のカイ2乗検定およびFisherの直接確率計算の統計手法の性能を心象で把握し、実際のデータ解析の場面での統計手法の選択指針として役だつように、これらの3つの統計手法の仮説検定の第1種の過誤確率の挙動を図1に示す。これは帰無仮説 $H_0: p_A - p_B = 0$ と対立仮説 $H_1: p_A - p_B \neq 0$ の両側検定での有意水準を0.05とする場合の、シミュレーションを通して数値的に評価した実質的な第1種の過誤確率である*1。

図1から、カイ2乗検定では、標本サイズが大きい場合(例えば、 $m=n=200$)、実質的な第1種の過誤確率は事前に定めた有意水準(名目水準)0.05の付近の値であるが、標本サイズが小さい場合(例えば、 $m=n=25$)に、それは名目水準よりも大きい。これは標本サイズが小さいために正規近似が十分に成りたないことによる。第1種の過誤確率が名目水準よりも大きいことは、2つの処理の反応割合に差がないことが真であったとしても、差があるとする対立仮説を採択する確率が高いことを意味する。医薬品の評価においては、第1種の過誤確率が名目水準を超えることは、例えば対照薬と比較し効果の優れていない薬剤を「優れている」と結論づける可能性が高くなるた

*1 実質的な第1種の過誤確率は次のように評価した：処理Aと処理Bについて $p_A = p_B = p$ の2項分布に従う変数として標本サイズ $m=n$ の標本を、標本サイズと反応割合の各組み合わせについて100万個をコンピュータで生成し、これらの生成標本に対して各検定を適用したときに有意水準0.05で有意になった生成標本の個数の割合を求めた。

め、第1種の過誤確率を適切に制御せねばならない(ICH-E9ガイドライン, 1998)。他方、反応割合がゼロあるいは1に近い場合、標本サイズの大きさにかかわらず、カイ2乗検定の第1種の過誤確率は有意水準0.05よりも小さい。これは、2つの処理の反応割合に差がないことが真であるときに対立仮説を選択する確率が小さい、すなわちより保守的な立場で仮説を評価することを意味する。

図1において、Yates修正のカイ2乗検定と Fisher の直接確率計算の検定の実質的な第1種の過誤確率は類似した挙動を示し、標本サイズと反応割合のすべての組み合わせにおいて名目水準よりも大きくならないが、その値は極端に小さい。Yates修正のカイ2乗検定と Fisher の直接確率計算を適用すれば、どのような状況でも常に保守的な立場で仮説を評価することになる。どのような場合でも言い過ぎのリスクを抑制する立場をとれば、常に Yates修正のカイ2乗検定あるいは Fisher の直接確率計算を用いる選択肢もある。しかし、Yates修正のカイ2乗検定と Fisher の直接確率計算は検出力の低い統計手法であり、2つの処理の反応割合に差があることが真であったとしても、その差を見逃してしまう確率が高いことに留意せねばならない。標本サイズが小さければ、カイ2乗検定は第1種の過誤確率が有意水準を超えて大きくなるが、

図1 帰無仮説 $H_0: p_A - p_B = 0$ と対立仮説 $H_1: p_A - p_B \neq 0$ の両側検定におけるカイ2乗検定、Yates修正のカイ2乗検定および Fisher の直接確率計算の検定の実質的な第1種の過誤確率の挙動

Yates 修正のカイ2乗検定や Fisher の直接確率計算に比較すれば、カイ2乗検定は概して検出力の高い統計手法である。

これまでに統計家により示唆されてきた2×2分割表の解析の方法の選択の留意点(例えば、Campbell(2007)や Lydersen et al.(2009))を、前記の図1の結果や筆者らの限られた経験とあわせると、統計手法の選択の指針を次のように示すことができる。

- ・ Yates 修正のカイ2乗検定や Fisher の直接確率計算と比較すれば、標本サイズが十分に大きければ(例えば、図1からは標本サイズは1群あたり200)、カイ2乗検定の第1種の過誤確率は概して名目水準に近く、カイ2乗検定を用いることに何の問題もなさそうである。
- ・ 標本サイズが小さい場合に、カイ2乗検定の第1種の過誤確率は名目水準よりも大きくなることもある。Cochran 規準を1つの目安として、できればシミュレーションを用いてカイ2乗検定の第1種の過誤確率の挙動を評価し、カイ2乗検定を用いるかどうかを決定するのがよさそうである。なお、反応割合がゼロや1といった極端な値をとる場合に、カイ2乗検定を適用しても第1種の過誤確率が名目水準よりも大きくなることはなく、主として標本サイズで判断すればよさそうである。
- ・ 反応割合がどのような値であれ、また標本サイズが小さかったとしても、Yates 修正のカイ2乗検定の第1種の過誤確率は名目水準よりも大きくなることはない。しかし、場合によって第1種の過誤確率の大きさは名目水準の半分以下となることがあり、Yates 修正のカイ2乗検定は相当に保守的な統計手法である。
- ・ Yates 修正のカイ2乗検定は Fisher の直接確率計算の代用とされることがあるが、現在、大抵の統計計算ソフトウェアには Fisher の直接確率計算の機能が含まれていることから、Yates 修正のカイ2乗検定を用いるご利益はほとんどなさそうである(実際、Campbell(2007)や Lydersen et al.(2009)では、Yates 修正のカイ2乗検定の使用を推奨していない)。
- ・ Yates 修正のカイ2乗検定と同様に、反応割合がどのような値であれ、また標本サイズが小さかったとしても、Fisher の直接確率計算の第1種の過誤確率は名目水準よりも大きくなることはないが、場合によって第1種の過誤確率の大きさは名目水準の半分以下となり、Fisher の直接確率計算は相当に保守的な統計手法である。また、カイ2乗検定に比較して、Fisher の直接確率計算の検出力は低い。標本サイズの大きさで、カイ2乗検定の第1種の過誤確率が名目水準を超えることが予想されそうな状況を除いて、常用的に Fisher の直接確率計算を使用することは避けたほうがよさそうである。

以上のように、カイ2乗検定や Fisher の直接確率計算はどのような状況でも適用できる万能な統計手法ではない。いずれの手法にも一長一短があり、データ解析者は、よりよい判断をくだすために、適用する手法の特徴を十分に理解し、データの様相をよく観察せねばならない。

③ 先に概観したカイ2乗検定、Yates 修正のカイ2乗検定、Fisher の直接確率計算の3つの統計手法について、事例を通して統計計算ソフトウェアで実行する際の留意点と出力される結果を概説する。第2回と同様に、SAS[®] University Edition、R、エクセル統計、JMP[®] の4つの統計計算ソフトウェアを使用し、データ“vitcap2”を

*2 Rの ISwRパッケージに組み込まれている、カドミウム産業における雇用者についての曝露状態と年齢および肺活量のデータである。

解析する*2。いま、「10年以上曝露されている群(group=1)」と「曝露されていない群(group=3)」との間で、肺活量が3.7未満である雇用者の割合に差があるか否かに関心があるとする(ここでは例示のために便宜的に3.7を閾値とするが、これは医学的な根拠に基づく値でない)。そこで、連続尺度で得られている肺活量“vital.capacity”(ここでは、“vital”とする)について、3.7未満であれば1、3.7以上であれば0の値をとる二値尺度へ変換し、変換後のデータを“cvital”とする(図2)。「10年以上曝露されている群」と「曝露されていない群」における肺活量が3.7未満である雇用者の割合をそれぞれ p_1 と p_3 とすれば、カイ2乗検定で検討したい帰無仮説と対立仮説(両側対立仮説)は $H_0: p_1 - p_3 = 0$ 、 $H_1: p_1 - p_3 \neq 0$ と記述される。以降では、それぞれのソフトウェアにおける検定の実行と出力される結果について個々に述べる。

① SAS® University Edition

FREQ プロシジャを用い、TABLES ステートメントの中で群を示す変数(群変数)“group”と応答(反応)変数“cvital”を指定し、カイ2乗検定の実行を chisq オプションで、Fisher の直接確率計算の実行を exact オプションで指定する。割合の差の推定

	group	age	vital	cvital					
1	1	39	4.62	0	29	3	35	4.24	0
2	1	40	5.29	0	30	3	41	3.88	0
3	1	41	5.52	0	31	3	38	4.85	0
4	1	41	3.71	0	32	3	41	4.79	0
5	1	45	4.02	0	33	3	36	4.36	0
6	1	49	5.09	0	34	3	36	4.02	0
7	1	52	2.7	1	35	3	41	3.77	0
8	1	47	4.31	0	36	3	41	4.22	0
9	1	61	2.7	1	37	3	37	4.94	0
10	1	65	3.03	1	38	3	42	4.04	0
11	1	58	2.73	1	39	3	39	4.51	0
12	1	59	3.67	1	40	3	41	4.06	0
13	3	27	5.29	0	41	3	43	4.02	0
14	3	25	3.67	1	42	3	41	4.99	0
15	3	24	5.82	0	43	3	48	3.86	0
16	3	32	4.77	0	44	3	47	4.68	0
17	3	23	5.71	0	45	3	53	4.74	0
18	3	25	4.47	0	46	3	49	3.76	0
19	3	32	4.55	0	47	3	54	3.98	0
20	3	18	4.61	0	48	3	48	5	0
21	3	19	5.86	0	49	3	49	3.31	1
22	3	26	5.2	0	50	3	47	3.11	1
23	3	33	4.44	0	51	3	52	4.76	0
24	3	27	5.52	0	52	3	58	3.95	0
25	3	33	4.97	0	53	3	62	4.6	0
26	3	25	4.99	0	54	3	65	4.83	0
27	3	42	4.89	0	55	3	62	3.18	1
28	3	35	4.09	0	56	3	59	3.03	1

図2 変数“cvital”を新たに追加したテストデータ“vitcap2”

SAS システム

FREQ プロシジャ

度数

表 : group * cvital			
group	cvital		
	0	1	合計
1	7	5	12
3	39	5	44
合計	46	10	56

group * cvital の統計量

統計量	自由度	値	p 値
カイ 2 乗値	1	5.9025	0.0151
尤度比カイ 2 乗値	1	5.0955	0.0240
連続性補正カイ 2 乗値	1	4.0174	0.0450
Mantel-Haenszel のカイ 2 乗値	1	5.7971	0.0161
ファイ係数		-0.3247	
一致係数		0.3088	
Cramer の V 統計量		-0.3247	
WARNING: セルの25%において、期待度数が5より小さくなっています。 カイ2乗検定は妥当な検定でないと思われます。			

Fisher の正確検定

セル (1,1) 度数 (F)	7
左側 Pr <= F	0.0280
右側 Pr >= F	0.9962
表の確率 (P)	0.0242
両側 Pr <= P	0.0280

列 1 リスクの推定値

	リスク	ASE	(漸近) 95% 信頼区間		(正確) 95% 信頼限界	
行 1	0.5833	0.1423	0.3044	0.8623	0.2767	0.8483
行 2	0.8864	0.0478	0.7926	0.9801	0.7544	0.9621
合計	0.8214	0.0512	0.7211	0.9217	0.6960	0.9109
差	-0.3030	0.1501	-0.5973	-0.0087		
行 1 - 行 2 の差						

列 2 リスクの推定値

	リスク	ASE	(漸近) 95% 信頼区間		(正確) 95% 信頼限界	
行 1	0.4167	0.1423	0.1377	0.6956	0.1517	0.7233
行 2	0.1136	0.0478	0.0199	0.2074	0.0379	0.2456
合計	0.1786	0.0512	0.0783	0.2789	0.0891	0.3040
差	0.3030	0.1501	0.0087	0.5973		
行 1 - 行 2 の差						

標本サイズ = 56

図3 SAS® の FREQ プロシジャ (chisq オプション、riskdiff オプション) によるカイ2乗検定、Yates 修正のカイ2乗検定および Fisher の直接確率計算の結果

値と信頼区間を出力する場合には、さらに riskdiff オプションで指定する。これを実行すれば、図3に示すように、分割表、カイ2乗検定を含む諸種の検定の p 値、割合の差の推定値と信頼区間などが一緒に出力される。ちなみに、カイ2乗検定と Yates 修正のカイ2乗検定の p 値はそれぞれ0.0151と0.0450で、Fisher の直接確率計算の p 値は両側検定、片側検定ともに0.0280である。Yates 修正のカイ2乗検定と Fisher の直接確率計算の p 値は、カイ2乗検定の p 値に比較して大きいことが確認できる。また、(カイ2乗検定に基づく、すなわち Yates 修正を含まない)割合の差($p_1 - p_3$)の推定値は0.3030、95%信頼区間は[0.0087, 0.5973]と計算されている。このとき、riskdiff オプションに代わり riskdiffc オプションを指定すれば、Yates 修正のカイ2乗検定に基づく割合の差の信頼区間が出力される。SAS[®] では特に指定しなくとも、列1(“cvital”がゼロ)と列2(“cvital”が1)のそれぞれを基準とした結果(割合の差の推定値や信頼区間など)が出力される。またカイ2乗検定の適用の妥当性が疑われるとき、SAS[®] では図のように「セルの25%において、期待度数が5より小さくなっています。カイ2乗検定は妥当な検定でないと思われます」との警告が表示される。なお、Fisher の直接確率計算の両側検定の p 値は、分割表において列と行の合計を固定したもとでとりうるすべての組み合わせのうち、観測された表の得られる確率以下となる確率をすべて加えた値であり、R、エクセル統計、JMP[®] でも算出方法は同様である。

②R

“vitcap2”のようにデータが個体毎に得られている場合、R ではまず table 関数や xtabs 関数を用いて分割表を作成し、その後、prop.test 関数を用いてカイ2乗検定を行う。例えば table 関数では群変数と応答変数のデータをベクトルで指定する。prop.test 関数では、作成した表(オブジェクト)のみを指定すればよい。R では、規定手法として Yates 修正のカイ2乗検定が適用される。図4に示すように、SAS[®] の実行結果と一致し、Yates 修正のカイ2乗検定の p 値は0.0450と計算される。95%信頼区間は[-0.0443, 0.6503]である。R では分割表を作成する際に特に指定しなければ、“cvital”がゼロである頻度を1列目とする表が作成され、prop.test 関数の実行結果として、分割表の1列目を基準とした割合の差の推定値と信頼区間が出力される。したがっ

```

2-sample test for equality of proportions with continuity
correction

data:  table
X-squared = 4.0174, df = 1, p-value = 0.04503
alternative hypothesis: two.sided
95 percent confidence interval:
 -0.04428056  0.65034117
sample estimates:
  prop 1      prop 2 
0.4166667 0.1136364 

警告メッセージ:
In prop.test(table) : カイ自乗近似は不正確かもしれません

```

図4 R の chisq.test 関数による Yates 修正のカイ2乗検定の結果

て、関心のある割合(ここでは“cvital”が1である割合)に対応する頻度を1列目として分割表を作成する必要がある(例えば factor 関数を用いて指定しておく)。またオプションで「correct=FALSE」と指定すればカイ2乗検定の結果が、図5のように出力される。SAS[®]の結果と同様に、カイ2乗検定のp値は0.01512、割合の差の95%信頼区間は[0.0087, 0.5973]と求められる。Rでは、いずれかのセルの期待度数が5未満の場合に「カイ自乗近似は不正確かもしれません」とカイ2乗検定の使用に関する警告が表示される。なお、Yates修正のカイ2乗検定の結果にみられるように、検定の結果は有意水準0.05で有意であるが、95%信頼区間はゼロを含み、検定と信頼区間の結果に矛盾が認められる。これは、検定統計量の計算では帰無仮説 $p_1=p_3$ のもとで併合分散を用いるのに対して、信頼区間の算出では $p_1 \neq p_3$ のもとで非併合分散を用いるため、検定と信頼区間の結果が一致しないことが起こりうる。なお、chisq.test関数を用いても同じ結果が得られるが、prop.test関数で出力される割合の差の推定値や信頼区間は出力されない。

Fisherの直接確率計算を実行するためには、fisher.test関数を用いる。カイ2乗検定のprop.test関数やchisq.test関数の場合と同様に、まずtable関数やxtabs

```

2-sample test for equality of proportions without continuity
correction

data:  table
X-squared = 5.9025, df = 1, p-value = 0.01512
alternative hypothesis: two.sided
95 percent confidence interval:
 0.008749743 0.597310864
sample estimates:
 prop 1      prop 2 
0.4166667 0.1136364

警告メッセージ:
In prop.test(table, correct = F) : カイ自乗近似は不正確かもしれません

```

図5 Rのchisq.test関数によるカイ2乗検定の結果

```

Fisher's Exact Test for Count Data

data:  table
p-value = 0.02799
alternative hypothesis: true odds ratio is not equal to 1
95 percent confidence interval:
 0.9634275 30.8801334
sample estimates:
 odds ratio 
5.345953

```

図6 Rのfisher.test関数によるFisherの直接確率計算の結果

関数で分割表を作成し、fisher.test関数ではその分割表を指定する。出力された結果(図6)から、p値はSAS®の実行結果と同様に0.0280である。Rでは規定値で両側検定のp値が計算されるため、片側検定のp値を求めたい場合には、オプションで例えば「alternative="greater"」と指定する。fisher.test関数ではカイ2乗検定(prop.test関数)での割合の差に代わり、オッズ比の推定値と信頼区間が出力される(prop.test関数の場合と同様に、何に対する比になっているかに注意する)。

③エクセル統計

「集計表の作成と分析」の中の「クロス集計表の作成と分析」を用いる。出てくるウィンドウで群変数を「表頭」に、応答変数を「表側」に指定し、次のウィンドウの「全般」タブで「Pearson のカイ二乗」と「Yates の補正」を、「2×2」タブで「Fisher の直接確率」を一緒に選択すれば、図7のように分割表、カイ2乗検定、Yates修正のカイ2乗検定およびFisherの直接確率計算の結果が出力される。カイ2乗検定のp値は、SAS®とRの結果と一致し、0.0151と得られる。先に示したように、SAS®やRではカイ2乗検定の使用に関する警告がされるが、エクセル統計ではそのような警告は表示されない。

	A	B	C	D	E	F	G
1	クロス集計表の作成と分析						
2							
3	クロス集計表						
4			group				
5			1	3	合 計		
6	cvital	0	7	39	46		
7		1	5	5	10		
8		合 計	12	44	56		
9							
10	期待度数						
11			group				
12			1	3			
13	cvital	0	9.857	36.143			
14		1	2.143	7.857			
15				太字:1未満 赤字:5未満			
16							
17	カイ二乗検定				**: 1% 有意 *: 5% 有意		
18	手 法	カイ二乗値	自由度	P 値	判 定		
19	Pearson	5.9025	1	0.0151	*		
20	Yatesの補正	4.0174	1	0.045	*		
21							
22	2行×2列のクロス集計表から得られる統計量						
23	手 法	計算方法	値				
24	Fisherの直接確率	両側P値	0.028				
25		片側P値	0.028				
26							

図7 エクセル統計の「クロス集計表の作成と分析」によるカイ2乗検定、Yates修正のカイ2乗検定およびFisherの直接確率計算の結果

④JMP®

「分析」の「二変量の関係」を用いる。XとYに群変数と応答変数を指定する(いずれも名義変数として指定する)。結果のウィンドウには分割表と検定の結果が表示され、カイ2乗検定のp値は、他のソフトウェアによる結果と一致し、0.0151と得られ

groupとcvitalの分割表に対する分析

モザイク図

分割表

		cvital		
		0	1	
group	度数			
	全体%			
	列%			
	行%			
	1	7 12.50 15.22 58.33	5 8.93 50.00 41.67	12 21.43
	3	39 69.64 84.78 88.64	5 8.93 50.00 11.36	44 78.57
		46 82.14	10 17.86	56

検定

N	自由度	(-1)*対数尤度	R2乗(U)
56	1	2.5477702	0.0970

検定	カイ2乗	p値(Prob>ChiSq)
尤度比	5.096	0.0240*
Pearson	5.903	0.0151*

Fisherの正確検定	p値	対立仮説
左片側検定	0.0280*	group=1の場合の確率(cvital=1)は、group=3の場合よりも大きい
右片側検定	0.9962	group=3の場合の確率(cvital=1)は、group=1の場合よりも大きい
両側検定	0.0280*	「cvital=1」である確率は、「group」の水準間で異なる

図8 JMP®の「分割表に対する分析」によるカイ2乗検定とFisherの直接確率計算の結果

る(図8)。JMP®では規定手法としてカイ2乗検定とFisherの直接確率計算の結果が出力される。このとき事前に「環境設定」の「プラットフォーム」中で指定すれば(「分割表」の「リスク差」にチェックを入れる)、割合の差の推定値と信頼区間が併せて出力される。カイ2乗検定の使用に関する警告は、エクセル統計と同様にJMP®でも表示されない。また、JMP®では、Yates修正のカイ2乗検定は実行されない。

文献

- 1) Campbell, I. (2007). Chi-squared and Fisher-Irwin tests of two-by-two tables with small sample recommendations. *Statistics in Medicine* 26,3661-3675.
- 2) Fisher, R.A. (1934). *Statistical Methods for Research Works*. 5th edition. Edinburgh: Oliver and Boyd.
- 3) Fleiss, J.L., Levin, B., Paik, M.C. (2003). *Statistical Methods for Rates and Proportions*, 3rd Edition. New York: John Wiley & Sons.
- 4) Cochran, W.G. (1952). The χ^2 test of goodness of fit. *Annals of Mathematical Statistics* 23,315-345.
- 5) Irwin, J.O. (1935). Tests of significance for difference between percentages based on small numbers. *Metron* 12, 83-94.
- 6) International Conference on Harmonisation E9 Expert Working Group (1998). E9: Statistical Principles for Clinical Trials. International Conference on Harmonisation.
- 7) Lydersen, S., Fagerland, M.W., Laake, P. (2009). Recommended test for association in 2×2 tables. *Statistics in Medicine* 28,1159-1175.
- 8) Person, K. (1900). On the criterion that a given system of deviations from the probable in the case of a correlated system of variable is such that it can be reasonably supposed to have arisen from random sampling. *Philosophical Magazine Series* 5,50,157-172.
- 9) Yates, F. (1934). Contingency tables involving small numbers and the χ^2 test. Supplement to the *Journal of the Royal Statistical Society* 1,217-235.
- 10) 朝倉こう子・濱崎俊光(2015). 2つの平均の比較. *Drug Delivery System* 30,149-158.