

【統計一口メモ 第20話】

＜ピュアレビューの McNemar 検定って？＞

名古屋市立大学大学院医学研究科 非常勤講師 薬学博士 松本一彦

例題：病理医2人による病理標本50枚の肝腫瘍診断で、腫瘍の有無について2人の診断の一致性を検討した。ピュアレビューの結果は一致しているといえるだろうか？

		病理医B		計
		有	無	
病理医A	有	26	7	34
	無	2	15	16
計		27	23	50

§1. いつもの「2x2分割表 χ^2 検定」ではダメ！—オッズ比をしてみる

このような2x2分割表を見たら、多くの場合 χ^2 検定あるいはFisher正確検定を行い、 χ^2 検定ならば $\chi^2=20.46$ $p<0.001$ 、Fisher正確検定ならば $p<0.001$ となり著明な差がある、すなわち、“一致していない”ということになります。

この解析法は正しいでしょうか？この2x2分割表のオッズ比は27.86と極めて大きな数値となります。この検定は病理医によって腫瘍有と無が一致した標本数と一致しなかった標本数の差をみているものであって、一致しなかった数(7と2)が小さな値を示す場合すなわちオッズ比が大きい場合は差が有意となりやすい傾向にあります。

	オッズ
26/7 =	3.71
2/15 =	0.13
	27.86
	オッズ比

したがってこの検定では2人の判定を評価することはできません。

§2. 対応のある2x2分割表—McNemar 検定¹⁾

表のように「有」「無」での2x2分割表の場合は通常の χ^2 検定ではなく、McNemar 検定で行います。McNemar 検定は χ^2 検定やFisher 検定のように2群間の独立性を検定するものではありません。「対応のある2x2データの比率の差」を検定する手法です。

＜McNemar 検定＞

		方法2		計	出現率
		+	-		
方法1	+	a	b	r_1	$p_1=(a+b)/N$
	-	c	d	$N-r_1$	
計		r_2	$N-r_2$	N	
出現率		$p_2=(a+c)/N$			

2つの標本中の(+)である個体の比率は、方法1では $(a+b)/N$ 、方法2では $(a+c)/N$ です。

研究者が興味があるのは、この2つの比率の差です。

$$\text{母比率} = p_1 = (a+b)/N$$

$$\text{母比率} = p_2 = (a+c)/N$$

$$p_1 = (26+7)/50 = 0.660$$

$$p_2 = (26+2)/50 = 0.560$$

$$\text{比率の差} = |p_1 - p_2| = |(b-c)/N| = |(7-2)/50| = 0.1000$$

$$\text{比率の差の標準誤差} = \text{SE}(\hat{p}_1 - \hat{p}_2) = \sqrt{(b+c)/N}$$

上記式に連続修正項 $\pm 1/N$ を加えて近似的に Z 検定を行います。

$$Z = \frac{(\hat{p}_1 - \hat{p}_2) \pm 1/N}{\text{SE}(\hat{p}_1 - \hat{p}_2)} = \frac{(b-c) \pm 1}{\sqrt{(b+c)}} \sim N(0,1) \text{ 分布}$$

帰無仮説は $(b-c)/N$ の期待値が 0 である、すなわち b と c の期待値が等しいということです。Z を 2 乗すると自由度 1 の χ^2 分布となる性質を利用して、帰無仮説が有意水準 α で棄却できる条件は次の式になります。

$$M = \frac{(|b-c|-1)^2}{b+c} > \chi_1^2(\alpha)$$

これは McNemar 検定とよばれています。JMP の Bowker 検定と同じです。

例題では

$$M = \frac{(|7-2|-1)^2}{7+2} = \frac{16}{9} = 1.78 < \chi_1^2(0.05) = 3.87 \quad P = 0.182149$$

となり、病理医 A と B の腫瘍判定は $p=0.182$ で差はみられず、「判定は一致している」となります。ただし、この解法は連続修正の $1/N$ が組み入れられています。JMP および Pharmaco では 2x2 分割表解析には連続修正を入れていないため、McNemar 検定においても入れません。したがって、結果は以下ようになります。

$$M = \frac{|7-2|^2}{7+2} = \frac{25}{9} = 2.78 < \chi_1^2(0.05) = 3.87 \quad P = 0.0954$$

p 値は $p=0.095$ と連続修正を入れたとき ($p=0.182$) より小さくなりましたが、結論は変わりません。

§ 3. さまざまな例題

例題：血清 IgE 抗体の定量検査で、患者 166 名に RAST 法とスクラッチ法を実施し、その感度を比較した。両法で差があるといえるだろうか？¹⁾

	スクラッチ		計	RAST の感度
	+	-		
RAST +	85	2	87	87/166 = 52.4%
RAST -	18	61	79	
計	108	63	166	

スクラッチの感度 103/166 = 62.0%

< Pharmaco 入力形式と解 > 処置 A = RAST 処置 B = スクラッチ

McNemar	P	N	検定法	McNemar	
			【処置A】	【処置B】	【例数】
P	85	2	P	P	85
N	18	61	P	N	18
+ = P			N	P	2
- = N			N	N	61
			P 値	0.0003	**

例題：アセプトロールの狭心症に対する治療効果について二重盲検法でクロスオーバー法を用いて検討した。対照薬としてはアルプリノロールを用い、患者は54名を対象とした。改善度に有意差がみとめられるだろうか？¹⁾

		対照薬		計
		+	-	
アセプトロール	+	25	7	32
	-	1	21	22
	計	26	28	54

P 値 0.0339 *

例題：喀痰の50試料を2つの異なる培地AとBとで各々培養する。2つの培地の結核菌検出能力に差があるだろうか？²⁾

帰無仮説は両培地の効果が等しい(両側検定) 連続修正なし

		培養液 B		計
		+	-	
培養液 A	+	20	12	32
	-	2	16	18
	計	22	28	54

P値 0.008 **

著者の Armitage は連続修正しないことを基本にしていて、わざわざ連続修正した場合の結果も載せている。

※ボクをつぶやき:McNemar 検定は差がないことを期待している統計の世界では数少ない手法といえるのでは？ただ、2x2分割表しか使えないので、「無効、やや有効、有効、著効」のような'axb分割表の場合はカッパー係数を用いる。ボクの知っている限り、カッパー係数を記載している教科書は少ないし、あいまいな範囲を記載しているのみで、いずれも p 値は算出していない。ボクの知っている限りでは JMP がカッパー係数をとりあげ、p値まで算出している。 χ^2 検定でできればPharmacolにも採用するのだが。

1) 丹後俊郎「医学への統計学」第3版 朝倉書店 2013年

2) P.Armitage, G.Berry 「医学研究のための統計的方法」サイエンティスト社 2001年