比率モデルにおける対照群との比の統計解析法

2016SS040 松下舞

指導教員：白石高章

1　はじめに

通常よく知られている比率の差の検定法は, 比率の小さい場合に帰無仮説が棄却さないため明確な結論が出ない. 本論では, 比率の比による 2 標本モデルの推定法の理論を用いて多標本モデルにおける理論とオッズ比を用いた理論を述べる. データ解析例では比較的比率の差が少ない医療データを使用し, 比率の比とオッズ比による検定の比較を行う.

2　2 標本モデルの推測法

第 1 標本 $X_1, \cdots, X_{n_1}$ を成功の確率が p_1, の独立な n_1 回のベルヌーイ試行とし, 第 2 標本 $Y_1, \cdots, Y_{n_2}$ を成功の確率 p_2 の独立な n_2 回のベルヌーイ試行とする. さらに, $X_1, \cdots, X_{n_1}$ と $Y_1, \cdots, Y_{n_2}$ は互いに独立とする. p_1, p_2 が 0 または 1 の明白な場合を除き, 以降 $0 < p_i < 1 \quad (i = 1, 2)$ を仮定する. 確率変数 X,Y をそれぞれ

$$\bar{X} \equiv \frac{X_1 + \cdots + X_{n_1}}{n_1} \quad , \qquad \bar{Y} \equiv \frac{Y_1 + \cdots + Y_{n_2}}{n_2}$$

で定義し, p_1, p_2 の点推定量をそれぞれ, $\hat{p}_1 \equiv \bar{X}$, $\hat{p}_2 \equiv \bar{Y}$ とする. $n \equiv n_1 + n_2$ とおき,

$$\lim_{n\to\infty} \frac{n_1}{n} = \lambda \quad (0 < \lambda < 1)$$

を仮定する.

$$T(\boldsymbol{p}) \equiv \sqrt{n}\frac{\{\log(\frac{\hat{p}_1}{\hat{p}_2}) - \log(\frac{p_1}{p_2})\}}{\tilde{\sigma}_n},$$

$$\tilde{\sigma}_n \equiv \sqrt{\frac{nn_2\hat{p}_2(1-\hat{p}_1) + nn_1\hat{p}_1(1-\hat{p}_2)}{\hat{p}_1\hat{p}_2 n_1 n_2}}$$

である. p_1/p_2 に対する $100(1-\alpha)$ の漸近的な信頼区間は

$$\frac{\hat{p}_1}{\hat{p}_2}\exp\left\{-\frac{\tilde{\sigma}_n}{\sqrt{n}}z\left(\frac{\alpha}{2}\right)\right\} < \frac{p_1}{p_2} < \frac{\hat{p}_1}{\hat{p}_2}\exp\left\{\frac{\tilde{\sigma}_n}{\sqrt{n}}z\left(\frac{\alpha}{2}\right)\right\}$$

である. 次に漸近的な検定を行う.

① 帰無仮説 $H_0 : \frac{p_1}{p_2} = 1$　vs.　対立仮説 $H_1 : \frac{p_1}{p_2} \neq 1$

$$\phi(T) = \begin{cases} 1 & (|T| > z\left(\frac{\alpha}{2}\right) \\ 0 & (|T| < z\left(\frac{\alpha}{2}\right) \end{cases}$$

② 帰無仮説 $H_0 : \frac{p_1}{p_2} = 1$　vs.　対立仮説 $H_2 : \frac{p_1}{p_2} > 1$

$$\phi(T) = \begin{cases} 1 & (T > z(\alpha)) \\ 0 & (T < z(\alpha)) \end{cases}$$

③ 帰無仮説 $H_0 : \frac{p_1}{p_2} = 1$　vs.　対立仮説 $H_3 : \frac{p_1}{p_2} < 1$

$$\phi(T) = \begin{cases} 1 & (T < -z(\alpha.) \\ 0 & (T > -z(\alpha)) \end{cases}$$

3　k 標本モデルにおける対照群との相違に関する多重比較

水準 A_iにおける標本の観測値を $(X_{i1}, X_{i2}, \cdots, X_{in_i})$ とし, X_{ij}は成功の確率が p_iのベルヌーイ試行とする. すなわち, $X_{ij} \sim B(1, p_i)$ である. さらにすべての X_{ij}は互いに独立であると仮定する. 第 k 標本を対照標本, 第 1 標本から第 $k-1$ 標本は処理標本とし, 下の表 1 のモデルについて考察する.

表 1　k 標本比率モデル

水準	標本	サイズ	データ
処理 1	第 1 標本	n_1	$X_{11}, \cdots, X_{1n_1}$
処理 2	第 2 標本	n_2	$X_{21}, \cdots, X_{2n_2}$
⋮	⋮	⋮	⋮
処理 k-1	第 k-1 標本	n_{k-1}	$X_{k-1\ 1}, \cdots, X_{k-1n_{k-1}}$
対照	第 k 標本	n_k	$X_{k1}, \cdots, X_{kn_k}$

分散の多重比較法の正確な理論を平均の場合と同様に論述することは非常に難しい. この場合, 次のボンフェローニの不等式 (文献 [1]) を使うことで容易に論じることができる. $k-1$ 個の事象 $A_1, A_2, \cdots, A_{k-1}$に対して

第 k 標本の対照標本と第 i 標本の処理標本を比較することを考える.

$$T_i \equiv \frac{\sqrt{n_i + n_k}\{\log\hat{p}_i - \log\hat{p}_k\}}{\tilde{\sigma}_{in}}$$

ただし,

$$\hat{p}_i \equiv \bar{X}_i \equiv \frac{X_{i1} + X_{i2} + \cdots + X_{in_i}}{n_i},$$

$$\tilde{\sigma}_{in} \equiv \sqrt{\frac{(n_i + n_k)n_k\hat{p}_k(1-\hat{p}_i) + (n_i + n_k)n_i\hat{p}_i(1-\hat{p}_k)}{\hat{p}_i\hat{p}_k n_i n_k}}$$

$$T_i(p) \equiv \sqrt{n_i + n_k}\frac{\{\log(\frac{\hat{p}_i}{\hat{p}_k}) - \log(\frac{p_i}{p_k})\}}{\tilde{\sigma}_{in}}$$

とおくと, $\lim_{n\to\infty} P\left(|T_i(\boldsymbol{p})| < z\left(\frac{\alpha}{2(k-1)}\right)\right) = 1 - \frac{\alpha}{k-1}$ であるため, ボンフェローニの不等式を使って, $\left\{\frac{p_i}{p_k}\middle| 1 \leqq i \leqq k-1\right\}$ に対する $100(1-\alpha)$ の同時信頼区間は,

$$\frac{\hat{p}_i}{\hat{p}_k}\exp\left\{-\frac{\tilde{\sigma}_{in}}{\sqrt{n_i+n_k}}z\left(\frac{\alpha}{2(k-1)}\right)\right\} < \frac{p_i}{p_k} < \frac{\hat{p}_i}{\hat{p}_k}\exp\left\{\frac{\tilde{\sigma}_{in}}{\sqrt{n_i+n_k}}z\left(\frac{\alpha}{2(k-1)}\right)\right\} \quad (1)$$

$(1 \leqq i \leqq k-1)$ で与えられる.

一つの比較のための検定は, 帰無仮説 $H_i^r : \frac{p_i}{p_k} = 1$ に対して, 3 種の対立仮説に対する水準 α のボンフェローニの不等式による多重比較検定は, 次で与えられる.

①両側対立仮説　$H_i^{r\pm} : \frac{p_i}{p_k} \neq 1$

$$\phi(T_i) = \begin{cases} 1 & \left(|T_i| > z\left(\frac{\alpha}{2(k-1)}\right)\right) \\ 0 & \left(|T_i| < z\left(\frac{\alpha}{2(k-1)}\right)\right) \end{cases}$$

②片側対立仮説 $H_i^{r+} : \frac{p_i}{p_k} > 1$

$$\phi(T_i) = \begin{cases} 1 & \left(T_i > z\left(\frac{\alpha}{k-1}\right)\right) \\ 0 & \left(T_i < z\left(\frac{\alpha}{k-1}\right)\right) \end{cases}$$

③片側対立仮説 $H_i^{r-} : \frac{p_i}{p_k} < 1$

$$\phi(T_i) = \begin{cases} 1 & \left(T_i < -z\left(\frac{\alpha}{k-1}\right)\right) \\ 0 & \left(T_i > -z\left(\frac{\alpha}{k-1}\right)\right) \end{cases}$$

4　k 標本モデルにおけるオッズ比の検定

事象の両群における確率を p_i(第 i 標本), p_k(第 k 標本) とすれば, オッズ比は

$$\frac{p_i/(1-p_i)}{p_k/(1-p_k)} = \frac{p_i(1-p_k)}{p_k(1-p_i)}$$

と表される.

$$S_i = \frac{\sqrt{n_i+n_k}\left\{\log\left(\frac{\hat{p}_i}{1-\hat{p}_i}\right) - \log\left(\frac{\hat{p}_k}{1-\hat{p}_k}\right)\right\}}{\tilde{\sigma}_{in}},$$

$$S_i(\boldsymbol{p}) \equiv \frac{\sqrt{n_i+n_k}\left\{\log\left(\frac{\hat{p}_i(1-\hat{p}_k)}{\hat{p}_k(1-\hat{p}_i)}\right) - \log\left(\frac{p_i(1-p_k)}{p_k(1-p_i)}\right)\right\}}{\tilde{\sigma}_{in}}$$

とおく.

比率の比の場合 (1) と同様に, $\left\{\frac{p_i(1-p_k)}{p_k(1-p_i)} \middle| 1 \leqq i \leqq k-1\right\}$ に対する $100(1-\frac{\alpha}{k-1})$ の漸近的な信頼区間は

$$\frac{\hat{p}_i(1-\hat{p}_k)}{\hat{p}_k(1-\hat{p}_i)}\exp\left\{-\frac{\tilde{\sigma}_{in}}{\sqrt{n_i+n_k}}z\left(\frac{\alpha}{2(k-1)}\right)\right\} < \frac{p_i(1-p_k)}{p_k(1-p_i)} < \frac{\hat{p}_i(1-\hat{p}_k)}{\hat{p}_k(1-\hat{p}_i)}\exp\left\{\frac{\tilde{\sigma}_{in}}{\sqrt{n_i+n_k}}z\left(\frac{\alpha}{2(k-1)}\right)\right\}$$

$(1 \leqq i \leqq k-1)$ で与えられる. 帰無仮説 $H_i^o : \frac{p_i(1-p_k)}{p_k/(1-p_i)} = 1$ に対して, 3 種の対立仮説に対する水準 α のボンフェローニの不等式による多重比較検定は, 次で与えられる.

①両側対立仮説　$H_i^{o\pm} : \frac{p_i(1-p_k)}{p_k(1-p_i)} \neq 1$

$$\phi(S_i) = \begin{cases} 1 & \left(|S_i| > z\left(\frac{\alpha}{2(k-1)}\right)\right) \\ 0 & \left(|S_i| < z\left(\frac{\alpha}{2(k-1)}\right)\right) \end{cases}$$

②片側対立仮説 $H_i^{o+} : \frac{p_i(1-p_k)}{p_k(1-p_i)} > 1$

$$\phi(S_i) = \begin{cases} 1 & \left(S_i > z\left(\frac{\alpha}{k-1}\right)\right) \\ 0 & \left(S_i < z\left(\frac{\alpha}{k-1}\right)\right) \end{cases}$$

③片側対立仮説 $H_i^{o-} : \frac{p_i(1-p_k)}{p_k(1-p_i)} < 1$

$$\phi(S_i) = \begin{cases} 1 & \left(S_i < -z\left(\frac{\alpha}{k-1}\right)\right) \\ 0 & \left(S_i > -z\left(\frac{\alpha}{k-1}\right)\right) \end{cases}$$

5　C言語プログラムの解説とデータ解析

5.1　プログラムの解説

C 言語プログラムにより, これまでに述べた比率の比による検定結果, 及びオッズ比による検定結果を水準 $\alpha = 0.05$ として作成した.

5.2　全国のがん罹患数のデータ

文献 [2] より, 全国の癌罹患数と罹患率のデータのうち, 日本の主要 10 都市のある都道府県を標本とし, 愛知県を対照群とした. ただし, 同一の都道府県でも男女で罹患率に差がある場合があったため男女別の解析を行い, 各標本と対照群において罹患率に差があるかどうかを考察した. 使用した各都道府県のデータは以下に記載する.

表 2　主要 10 都道府県の人口と罹患数データ

都道府県	男性人口	女性人口	男性罹患数	女性罹患数
北海道	2552868	2848342	27095	21471
東京都	6621602	6793747	52321	42331
神奈川県	4567791	4568360	37682	28394
京都府	1235170	1339672	12167	9418
奈良県	660270	727548	6597	4773
大阪府	4288494	4577008	39628	29694
兵庫県	2697830	2923257	25850	19469
広島県	1386430	1476781	12880	9723
沖縄県	720548	740683	4408	3757
愛知県	3762016	3747620	28363	20711

5.3　実行結果

男女ともに比率の比とオッズ比のいずれも水準 5% の両側検定では帰無仮説はすべて棄却することができた. また, 対照群である愛知県と比較したときの 95% 信頼区間は沖縄県は 1.0 より小さく沖縄県以外は 1.0 より大きいことから, 沖縄県は愛知県より罹患率が低く, その他は愛知県より罹患率が高いという結論を得た.

6　終わりに

本論では多標本モデルにおける比率の比による推測法を提案してきた. また多標本モデルにおけるオッズ比による推測法を同時に論じ, どちらが優れているかについても考察してきた. C 言語プログラムによりデータ解析を行った結果, どちらも比率の差より優れていることが分かったが, 2 つの推測法に優劣がつくほどの結果はなかった.

参考文献

[1] 白石高章, 杉浦洋『多重比較法の理論と数値計算』共立出版, 東京,2018

[2] 厚生労働省健康局がん・疾病対策課：「平成 28 年 全国がん登録 罹患数・罹患率」
https://www.mhlw.go.jp/content/10900000/000553552.pdf