ポアソンモデルにおける対照群との平均比の統計解析法

2017SS018 井上博貴

指導教員：白石 高章

1 はじめに

稀に起こる現象の回数はポアソン分布に従う．ポアソン分布は平均と分散は同じであり, 多重比較検定をする際よく行われる方法はボンフェローニの不等式である．しかしこの方法による多重比較検定法は検出力が低いため明確な結果が出ないという欠点がある．そこで改良型としてホルムの方法の理論の手順を述べ, 帰無仮説が棄却されるかを考察する．本論のデータ解析では医療データを使用しボンフェローニの方法とホルムの方法について考察していく．

2 2 標本モデルの推測法

第 1 群 $X_1,\cdots,X_{n_1}$をポアソン分布 $\mathcal{P}_o(\mu_1)$ からの無作為標本とし, 第 2 群 $Y_1,\cdots,Y_{n_2}$をポアソン分布 $\mathcal{P}_o(\mu_2)$ からの無作為標本とする． 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．
確率変数 W_1, W_2 をそれぞれ

$$W_1 \equiv X_1+\cdots+X_{n_1},\quad W_2 \equiv Y_1+\cdots+Y_{n_2}$$

とおく．このとき, μ_iの点推定量は,

$$\hat{\mu}_i = \frac{W_i}{n_i}\quad (i=1,2)$$

$(i=1,2)$ で与えられる.$n\equiv n_1+n_2$とおき,

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

を仮定する．
白石 [1] の定理 3.32(2), 定理 3.35, 系 3.6 を使って,

$$T(\boldsymbol{\mu}) \equiv \sqrt{n}\frac{\left\{\log\left(\frac{\hat{\mu}_1}{\hat{\mu}_2}\right)-\log\left(\frac{\mu_1}{\mu_2}\right)\right\}}{\tilde{\sigma}_n}$$

$$\tilde{\sigma}_n \equiv \sqrt{\frac{n}{\hat{\mu}_1 n_1}+\frac{n}{\hat{\mu}_2 n_2}}$$

とおく．
μ_1/μ_2に対する信頼係数$1-\alpha$の漸近的な信頼区間は

$$\frac{\hat{\mu}_1}{\hat{\mu}_2}\exp\left\{-\frac{\tilde{\sigma}_n}{\sqrt{n}}z\left(\frac{\alpha}{2}\right)\right\} < \frac{\mu_1}{\mu_2} < \frac{\hat{\mu}_1}{\hat{\mu}_2}\exp\left\{\frac{\tilde{\sigma}_n}{\sqrt{n}}z\left(\frac{\alpha}{2}\right)\right\}$$

である．次に漸近的な検定を行う．
① 帰無仮説 $H_0:\mu_1/\mu_2=1$ vs. 対立仮説 $H_1:\mu_1/\mu_2\neq 1$

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

② 帰無仮説 $H_0:\mu_1/\mu_2=1$ vs. 対立仮説 $H_2:\mu_1/\mu_2>1$

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

③ 帰無仮説 $H_0:\mu_1/\mu_2=1$ vs. 対立仮説 $H_3:\mu_1/\mu_2<1$

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

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 $(X_{i1},\cdots,X_{in_i})$ とし, X_{ij} は平均 μ_i のポアソン分布 $\mathcal{P}_o(\mu_i)$ に従うものとする．さらに, 全ての X_{ij} は互いに独立であるとする．すなわち, $X_{ij}\sim\mathcal{P}_o(\mu_i)\quad(j=1,\cdots,n_i,\ i=1,\cdots,k)$ である．
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．
ボンフェローニの不等式 (白石 [2]) を使うことで容易に論じることができる．

表 1 k 標本ポアソンモデル

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	$X_{21},\cdots,X_{2n_2}$
$\vdots$	$\vdots$	$\vdots$	$\vdots$
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	$X_{k1},\cdots,X_{kn_k}$

$$P\left(\bigcup_{i=1}^{k-1}A_i\right) \leqq \sum_{i=1}^{k-1}P(A_i)$$

が成り立つ．
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．

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

ただし,

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

$$\tilde{\sigma}_{in} \equiv \sqrt{\frac{n_i+n_k}{\hat{\mu}_i n_i}+\frac{n_i+n_k}{\hat{\mu}_k n_k}}$$

$$T_i(\boldsymbol{\mu}) \equiv \sqrt{n_i+n_k}\frac{\left\{\log\left(\frac{\hat{\mu}_i}{\hat{\mu}_k}\right)-\log\left(\frac{\mu_i}{\mu_k}\right)\right\}}{\tilde{\sigma}_{in}}$$

とおくと,

$$\lim_{n\to\infty}P\left(|T_i(\boldsymbol{\mu})|<z\left(\frac{\alpha}{2(k-1)}\right)\right)=1-\frac{\alpha}{k-1}$$

であるため, ボンフェローニの不等式を使って,
$\left\{\mu_i/\mu_k\middle|1\leqq i\leqq k-1\right\}$ に対する信頼係数 $1-\alpha$ の漸近的

な同時信頼区間は,

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

で与えられる.

1 つの比較のための検定は, 帰無仮説 $H_i^A : \mu_i/\mu_k = 1$ 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 $H_i^{A\pm} : \mu_i/\mu_k \neq 1$

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

② 片側対立仮説 $H_i^{A+} : \mu_i/\mu_k > 1$

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

③ 片側対立仮説 $H_i^{A-} : \mu_i/\mu_k < 1$

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

ボンフェローニの不等式による多重比較検定法は検出力が低いため使いたくない手法である. そこで改良した方法として, ホルムの方法 (白石 [2]) に基づいて次の検定が与えられる.

第 k 標本は対照群で第 1 標本から第 $k-1$ 標本は処理群である場合を考える.

T_i の実現値を t_i とする.

$\mu_1 = \cdots = \mu_k$ の下で,

$$p_i \equiv \begin{cases} 2\int_{|t_i|}^{\infty}\varphi(x)dx & (\text{①のとき}) \\ \int_{t_i}^{\infty}\varphi(x)dx & (\text{②のとき}) \\ \int_{-\infty}^{t_i}\varphi(x)dx & (\text{③のとき}) \end{cases}$$

$k-1$ 個の p 値 $p_1, \cdots, p_{k-1}$ を小さい方から並べたものを $p_{(1)} \leqq \cdots \leqq p_{(k-1)}$ とし, 対応する帰無仮説を $H_{0(1)}, \cdots, H_{0(k-1)}$ とする. $H_{0(i)} \in \mathcal{H}_{k-1}$ である. ある i が存在して, $1 \leqq j \leqq i$ となるすべての整数 j に対して, $p_{(j)} \leqq \alpha/(k-j)$ ならば, $H_{0(i)}$ を棄却する.

4 C 言語プログラムの解説とデータ解析

4.1 C 言語プログラムの解説

C 言語プログラムにより, これまでに述べたポアソン比による検定結果を水準 $\alpha = 0.05$ として作成した.

4.2 全国のがん死亡者数データ

今回は国立がん研究センターがん情報サービスの都道府県別がん死亡データ (1995 年〜2019 年)[3] より, 2019 年の全国のがん死亡者数のデータのうち, 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. 各標本と対照群において死亡者数に

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 都道府県の人口とがん死亡者数データ

都道府県	男性人口	女性人口	男性がん死亡者数	女性がん死亡者数
北海道	2505804	2798609	11056	8369
埼玉県	3657838	3681231	12071	7720
東京	6854976	7087880	19833	14249
神奈川県	4585415	4614751	14275	9699
京都府	1234456	1348684	4421	3248
大阪府	4235996	4587457	15727	10711
兵庫県	2603761	2858386	9662	6832
広島県	1362763	1445224	4828	3464
福岡県	2418593	2691520	8759	6946
愛知県	3781399	3772474	11698	7851

4.3 実行結果

男性のがん死亡者数でのボンフェローニ法の推測では, 東京と神奈川県以外は棄却することができた. さらにそこからホルムの方法より, 東京を棄却することができた. 女性のがん死亡者数でのボンフェローニ法の推測では, 埼玉県, 東京, 神奈川県以外は棄却することができ, ホルムの方法より, 東京を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男性のがん死亡者数の第 3 標本 (東京) の信頼区間は 1.0 より小さいことから, 東京は愛知県より死亡率が低く, その他は愛知県より死亡率が高いという結論を得た.

5 おわりに

本研究では多標本のポアソンモデルにおける多重比較検定法を提案してきた. また, ボンフェローニの方法とホルムの方法でどれほど検出力を高めることができるのか考察した. C 言語プログラムにより実際にデータを用いて解析を行った結果, ホルムの方法の検定を行うことにより, ボンフェローニの方法の欠点を補うことができることがわかった.

参考文献

[1] 白石高章『統計科学の基礎ーデータと確率の結びつきがよくわかる数理』日本評論社, 東京, 2016

[2] 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

[3] 国立がん研究センターがん情報サービス:「全国がん死亡データ (1958 年〜2019 年)」 https://ganjoho.jp/reg_stat/statistics/dl/index.html 2020/11/25 閲覧