多標本ノンパラメトリックモデルにおける対照群との平均比の統計解析法

2017SS003 青原桃香

指導教員:白石高章

1　はじめに

2 標本のデータ解析において, 平均の差の推測論を考えることが多い. 本論では, 多標本の平均の比の推測論について考察する. はじめに信頼区間を求め, 検定方式を与える. これを基に解析手法の C 言語プログラムを作成し, 全国のコンビニの売上データを使って, 検定を行うことで平均の比を分析する.

2　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, $X_{ij} \sim N(\mu_i, \sigma_i^2)$ とする. 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　標本ノンパラメトリックモデル

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}	$X_{k-1\ 1}, \cdots, X_{k-1 n_{k-1}}$
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 (白石 [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

$$P\left(\bigcup_{i=1}^{k-1} A_i\right) \leqq \sum_{i=1}^{k-1} P(A_i) \qquad (2)$$

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.

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

ただし,

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

$$\tilde{\eta}_{in} \equiv \sqrt{\frac{\tilde{\sigma}_i^2}{\hat{\mu}_i^2(\frac{n_i}{n_i+n_k})} + \frac{\tilde{\sigma}_k^2}{\hat{\mu}_k^2(\frac{n_k}{n_i+n_k})}},$$

$$\tilde{\sigma}_i^2 \equiv \frac{1}{n_i-1}\sum_{j=1}^{n_i}(X_{ij}-\bar{X}_{i\cdot})^2,$$

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

とおくと,

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

である。

また, $1 \leqq i \leqq k-1$ を満たすすべての i に対して, μ_i/μ_k の区間推定に興味があるものとする. 定数 α $(0<\alpha<1)$ をはじめに決める. 任意の i に対して I_i を区間とする.

$P(1 \leqq i \leqq k-1$ を満たすすべての i に対して, $\mu_i/\mu_k \in I_i) \geqq 1-\alpha$ となるならば, $\mu_i/\mu_k \in I_i$ $(1 \leqq i \leqq k-1)$ を, $\{\mu_i/\mu_k \mid 1 \leqq i \leqq k-1\}$ に対する信頼係数 $1-\alpha$ の同時信頼区間とよんでいる.

事象 B_i を

$$B_i \equiv \left\{|T_i(\boldsymbol{\mu})| < z\left(\frac{\alpha}{2(k-1)}\right)\right\}$$

とする.

(2) 式で A_i を B_i^c とおくと,

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

ここで,

$$\lim_{n\to\infty} P\Bigg(1 \leqq i \leqq k-1 \text{ を満たすすべての } i \text{ に対して, } |T_i(\boldsymbol{\mu})| < z\left(\frac{\alpha}{2(k-1)}\right)\Bigg)$$

$$= \lim_{n\to\infty} P\left(\bigcap_{i=1}^{k-1} B_i\right) = \lim_{n\to\infty}\left\{1 - P\left(\bigcup_{i=1}^{k-1} B_i^c\right)\right\}$$

$$\geqq \lim_{n\to\infty}\left\{1 - \sum_{i=1}^{k-1} P(B_i^c)\right\} = 1 - \sum_{i=1}^{k-1}\lim_{n\to\infty} P(B_i^c)$$

$$\lim_{n\to\infty} P(B_i^c) = \frac{\alpha}{k-1}$$

$$(\text{与式}) = 1-\alpha$$

よって, ボンフェローニの不等式を使って,

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

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

$\{\mu_i/\mu_k | 1 \leqq i \leqq k-1\}$ に対する $100(1-\alpha)\%$ の漸近的な同時信頼区間は,

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

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

$(1 \leqq i \leqq k-1)$ で与えられる.
一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　$H_i^{A\pm} : \mu_i/\mu_k \neq 1$

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

2. 片側対立仮説　$H_i^{A+} : \mu_i/\mu_k > 1$

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

3. 片側対立仮説　$H_i^{A-} : \mu_i/\mu_k < 1$

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

3　C 言語によるデータ解析

3.1　プログラム

これまでに述べた検定方式の C 言語プログラムによるプログラムを作成した.

3.2　データ

表 2 のデータは, コンビニエンスストア 統計調査月報 : コンビニエンスストア統計時系列データ [3] よりとってきたものである.
また, 第 1 標本は 2008 年の売上データ, 第 2 標本は 2010 年の売上データ, 第 3 標本は 2012 年の売上データ, 第 4 標本は 2014 年の売上データ, 第 5 標本は 2016 年の売上データ, 第 6 標本は 2018 年の売上データ, 第 7 標本は 2020 年の売上データであり, 表 2 は, 2008 年, 2016 年, 2020 年のデータをまとめたものである.

表 2　2008 年 2016 年 2020 年の月別売上データ

対象期間	2008 年	2016 年	2020 年
1 月	574,966	817,625	885,710
2 月	556,317	778,997	849,064
3 月	622,136	864,583	877,500
4 月	605,461	848,314	817,030
5 月	648,846	886,090	849,706
6 月	648,709	872,334	879,287
7 月	745,546	963,260	907,928
8 月	734,211	951,051	947,930
9 月	674,226	874,399	905,998
10 月	686,766	902,027	914,163
11 月	657,758	843,496	
12 月	702,129	907,873	

3.3　データ解析の結果

プログラムを使って, 全国のコンビニの売上データを解析した.
第 1 標本 (2008 年の売上データ) と第 7 標本 (2020 年の売上データ)
平均 ：654755.916667
分散 ：3152924317.743056
μ_i/μ_k の推定値：0.741151
検定統計量 ：-10.729544
μ_i/μ_k に対する信頼係数 $1-\alpha$ の同時信頼区間:
$0.688522 < \mu_i/\mu_k < 0.797802$
両側検定量 ：10.729544
両側検定結果 ：棄却された
片側検定 $(+)$ 結果 ：棄却しない
片側検定 $(-)$ 結果 ：棄却された
両側検定, 片側検定の両方が棄却された.
第 5 標本 (2016 年の売上データ) と第 7 標本 (2020 年の売上データ)
平均 ：875587.416667
分散 ：2515488862.576389
μ_i/μ_k の推定値：0.991121
検定統計量 ：-0.425198
μ_i/μ_k に対する信頼係数 $1-\alpha$ の同時信頼区間：
$0.937762 < \mu_i/\mu_k < 1.047515$
両側検定量 ：0.425198
両側検定結果 ：棄却しない
片側検定 $(+)$ 結果 ：棄却しない
片側検定 $(-)$ 結果 ：棄却しない
両側検定, 片側検定の両方とも棄却されなかった.

2008 年は棄却され, 2016 年は棄却されなかったので, 2016 年の売上の方が多いといえる. 2008 年の信頼区間は 2016 年より狭い範囲に位置している.
また, 2008 年, 2010 年, 2012 年, 2014 年では, 片側検定, 両側検定の両方が棄却され, 20016 年, 2018 年では, 両方とも棄却されなかった.
詳しい検定結果は本論に記載した.

4　おわりに

本論文では, ノンパラメトリックモデルにおける対照群との平均比の統計解析法を提案した. 信頼区間を求め, 検定を行うプログラムを C 言語で作成し, 実際のデータを用いて結果の解析を行った.

参考文献

[1] 白石高章：『統計科学の基礎』日本評論社, 東京,2012.

[2] コンビニエンスストア 統計調査月報：コンビニエンスストア統計時系列データ
https://www.jfa-fc.or.jp/folder/1/img/20190121150250.pdf
https://www.jfa-fc.or.jp/folder/1/img/20200120120214.pdf