# 多標本正規分布モデルにおける線形制約がある場合の検定法

2013SE239 矢田　眞夕

指導教員：白石　高章

## 1　はじめに

ノンパラメトリック法における平均に線形制約がある場合の検定法として，ウィルコクソン型の線形順位検定法がよく知られている．本論文では，多標本正規分布モデルにおける線形統計量に基づく検定法について考察する．

## 2　多標本正規分布モデルにおける線形統計量による検定法

ある要因 $A$ があり，$k$ 個の水準 $A_1, \cdots, A_k$ を考える．水準 $A_i$ における標本の観測値 $(X_{i1}, X_{i2}, \cdots, X_{in_i})$ を第 $i$ 標本とし，$X_{ij} \sim N(\mu_i, \sigma^2)$ $(j = 1, \cdots, n\,;\ i = 1, \cdots, k)$ とする．さらに，全ての $X_{ij}$ は互いに独立であるとする．
$\mu_i = \mu + c_i\Delta \quad (c_1 \leqq c_2 \leqq \cdots \leqq c_k)$(少なくとも 1 つの $\leqq$ は $\lneqq$ である）とし帰無仮説 $H_o : \Delta = 0 \quad (\mu_1 = \cdots = \mu_k)$ vs. 対立仮説 $H_A : \Delta > 0 \quad (\mu_1 \leqq \mu_2 \leqq \cdots \leqq \mu_k)$(少なくとも 1 つの $\leqq$ は $\lneqq$ である) を考える．
$R_{ij}$ を $X_{11}, \cdots, X_{kn_k}$ の中での $X_{ij}$ の順位とする．このとき順位統計量 $S$ を，

$$S = \sum_{i=1}^{k} c_i \sum_{j=1}^{n_i} R_{ij}$$

で定義する．文献 [3] より，棄却領域 $(S \geqq C_u)$ は，$X_{ij}$ がロジスティック分布にしたがっているとき局所最強力順位検定である．文献 [2] の (4.2)，(4.3) より

$$\bar{X}_{i\cdot} = \frac{1}{n_i}\sum_{j=1}^{n_i} X_{ij}, \quad \bar{X}_{\cdot\cdot} = \frac{1}{n}\sum_{i=1}^{k}\sum_{j=1}^{n_i} X_{ij}$$

とする．
統計量 S において $R_{ij}$ を $X_{ij} - \bar{X}_{\cdot\cdot}$ と置き換えたものを

$$T_1 = \sum_{i=1}^{k} c_i \sum_{j=1}^{n_i} \left(X_{ij} - \bar{X}_{\cdot\cdot}\right), \quad T_2 = \sum_{i=1}^{k} (c_i - \bar{c}) \sum_{j=1}^{n_i} X_{ij}$$

ただし，$\bar{c} = \frac{1}{n}\sum_{i=1}^{k} n_i c_i$ とおく，このとき次の命題を得る．

命題 1．$T_1 = T_2$ が成り立つ．

(証明)　$TS = \sum_{i=1}^{k}(c_i - \bar{c})\sum_{j=1}^{n_i}(X_{ij} - \bar{X}_{\cdot\cdot})$ とおく

$$TS - T_2 = \sum_{i=1}^{k}(c_i - \bar{c})\sum_{j=1}^{n_i}\bar{X}_{\cdot\cdot} = 0$$

$$TS - T_1 = -\sum_{i=1}^{k}\bar{c}\sum_{j=1}^{n_i}(X_{ij} - \bar{X}_{\cdot\cdot}) = 0$$

となることがわかり，$T_1 = T_2$ が示された．　□

$E(\bar{X}_{\cdot\cdot})$ と $E(T_1)$ を求めると

$$E(\bar{X}_{\cdot\cdot}) = \frac{1}{n}\sum_{i=1}^{k} n_i\mu_i$$

$$E(T_1) = \Delta\sum_{i=1}^{k} c_i n_i\left(c_i - \frac{1}{n}\sum_{i=1}^{k} n_i c_i\right)$$

である．
$H_0$ のもとで，$\Delta = 0$ のとき $E_0(T_1) = 0$ である．
$Y_i = (c_i - \bar{c})\sum_{j=1}^{n_i} X_{ij}$ とおくと

$$V(T_1) = \sum_{i=1}^{k}(c_i - \bar{c})^2 n_i\sigma^2$$

である．$H_0$ の下で
$T_1 \sim N(0, \sigma^2\sum_{i=1}^{k}(c_i - \bar{c})^2 n_i)$ となる．

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

とおいたとき, $T_1$ と $\tilde{\sigma}^2$ が互いに独立を示す．一般性を失うことなく
$\mu_i = 0$ $(i = 1, \cdots, k)$ とする．

$\mathrm{Cov}(\bar{X}_{i'\cdot} - \bar{X}_{\cdot\cdot}, X_{ij} - \bar{X}_{i\cdot})$ これを計算すると

$$i = i' \text{ のとき 与式} = \frac{\sigma^2}{n_i} - \frac{\sigma^2}{n_i} - \frac{\sigma^2}{n} + \frac{\sigma^2}{n} = 0$$

$$i \neq i' \text{ のとき 与式} = 0 - 0 - \frac{\sigma^2}{n} + \frac{\sigma^2}{n} = 0$$

よって $\mathrm{Cov}(\bar{X}_{i'\cdot} - \bar{X}_{\cdot\cdot}, X_{ij} - \bar{X}_{i\cdot}) = 0$ となり，$\bar{X}_{i'\cdot} - \bar{X}_{\cdot\cdot}$ と $X_{ij} - \bar{X}_{i\cdot}$ が互いに独立
さらに，

$$E_0\left(\frac{T_1}{\sqrt{V_0(T_1)}}\right) = \frac{1}{\sqrt{V_0(T_1)}}E_0(T_1) = 0$$

$$V_0\left(\frac{T_1}{\sqrt{V_0(T_1)}}\right) = \frac{1}{V_0(T_1)}V_0(T_1) = 1$$

となる．
$H_0$ の下で

$$W \equiv \frac{T_1}{\sqrt{V_0(T_1)}} \sim N(0,1)$$

が成り立つ，また

$$Z \equiv \frac{1}{\sigma^2}\sum_{i=1}^{k}\sum_{j=1}^{n_i}\left(X_{ij} - \bar{X}_{i\cdot}\right)^2 \sim \chi^2_{n-k}$$

であるので，文献 [1] の定理 3.21 より，$H_0$ の下で，

$$\frac{W}{\sqrt{\frac{Z}{n-k}}} \sim t_{n-k}$$

が成り立つ．すなわち，

$$T_{\boldsymbol{c}} \equiv \frac{\sum_{i=1}^{k}\left\{(c_i-\bar{c})\sum_{j=1}^{n_i}X_{ij}\right\}}{\sqrt{\tilde{\sigma}^2\sum_{i=1}^{k}(c_i-\bar{c})^2 n_i}},$$

$$T_k \equiv \frac{\sum_{i=1}^{k}\left\{\left(i-\frac{1}{n}\sum_{j=1}^{k}jn_j\right)\sum_{j=1}^{n_i}X_{ij}\right\}}{\sqrt{\tilde{\sigma}^2\sum_{i=1}^{k}n_i\left(i-\frac{1}{n}\sum_{j=1}^{k}jn_j\right)^2}}$$

とおくと，$H_0$ の下で，$T_{\boldsymbol{c}} \sim t_{n-k}$，$T_k \sim t_{n-k}$ である．ただし，$\tilde{\sigma}^2$ は (1) で定義されたものとする．

(i)　$c_1, \cdots, c_k$ が既知のとき
$T_{\mathrm{C}}$ を，検定統計量として，自由度 $n-k$ の $t$ 分布の上側 $100\alpha\%$ 点を $t(n-k;\alpha)$ とすると，帰無仮説 $H_o$ vs. 対立仮説 $H_A$ における水準 $\alpha$ の検定は，

$$\begin{cases} T_{\boldsymbol{c}} > t(n-k;\alpha) \text{ のとき } H_0 \text{ を棄却する} \\ T_{\boldsymbol{c}} < t(n-k;\alpha) \text{ のとき } H_0 \text{ を棄却しない} \end{cases}$$

である．

(ii)　$c_1, \cdots, c_k$ が未知のとき
$T_k$ を，検定統計量として，

$$\begin{cases} T_k > t(n-k;\alpha) \text{ のとき } H_0 \text{ を棄却する} \\ T_k < t(n-k;\alpha) \text{ のとき } H_0 \text{ を棄却しない} \end{cases}$$

である．

## 3　C 言語によるプログラム解説

多標本正規分布モデルにおける線形制約がある場合の検定法による検定結果を出力するプログラムを C 言語によりそれぞれ作成した．ただし，上側 $100\alpha\%$ 点を求めるために，文献 [4] を参考にした．ページの都合上，詳細なプログラムについては，本稿に記載した．

```
int main(void){
    yomikomi();
    keisan1();keisan2();keisan3();keisan4();
    keisan5();keisan6();keisan7();
    output();
    return 0;
}
```

1. yomikomi 関数により，データをテキストファイルから読み込み，標本数, サイズを計算かつ表示する，$c_i(c_1 \leqq c_2 \leqq \cdots \leqq c_k)$ を入力する．
2. keisan1 関数により，$\bar{c}$ の値を計算する．
3. keisan2 関数により，$X_{ij}$ の和の値を計算する．
4. keisan3 関数により，$\sum_{i=1}^{k}\left\{(c_i-\bar{c})\sum_{j=1}^{n_i}X_{ij}\right\}$ の値を計算する．
5. keisan4 関数により，$\sqrt{\sum_{i=1}^{k}(c_i-\bar{c})^2 n_i}$ の値を計算する．
6. keisan6 関数により，$\tilde{\sigma}^2$ の値を計算する．
7. keisan7 関数により，$T_{\mathrm{C}}$ の値を計算する．
8. output 関数により，検定結果を出力する．

### 3.1　愛知県の年平均気温データ

愛知県の年平均気温が上昇しているかどうかを調べるにあたり，1971 年から 2016 年のデータ (文献 [5]) を使用した．このとき，$1971+5(i-1)$ 年の日本の年平均気温 $(i=1,\cdots,10)$ について考察する．気象庁のデータをもとに愛知県の 11 ヶ所の気温データを集めた．統計量 $T_k$ を使って解析する．

### 3.2　解析結果

$T_k = 1.728348$ となり，$\alpha = 0.05$ では帰無仮説 $H_0$ は棄却された．

### 3.3　考察

1971 年から 2016 年の 45 年間で日本の年平均気温が上昇していることが確認できた．

## 4　おわりに

本論では，多標本正規分布モデルにおける線形制約がある場合の検定法を考察した．そして検定を行うための C 言語プログラムを作成し，実際のデータを用いることによってより理解を深めることができた．

## 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] 白石高章 :『多群連続モデルにおける多重比較法』. 共立出版株式会社，東京，2011.

[3] Hajek，J. , Sidak，Z.　and Sen，P. K.　: *Theory of Rank Tests*. Elsevier Inc, New York, 1984.

[4] 早川由宏，白石高章 : Fortan とC言語による統計プログラミングの基礎，Mathematica の使い方. 研究ノート.（2015 年 2 月）.

[5] 国土交通省 : 気象庁，
http://www.jma.go.jp/jma/index.html.