ワイブルモデルにおけるダネット型多重比較法

2010SE102　鬼頭広大

指導教員：白石高章

1　はじめに

本研究では 2 つのパラメータ α，β をもつワイブル分布に従う多群モデルにおける対照群との多重比較検定について考察する．特に，$\alpha = 1$ のときはワイブル分布は指数分布となる．指数分布モデルにおける多重比較法については白石 [1] で与えられている．ここでは，α を α_0 と与え β についての統計的推論を行う．

2　ワイブル分布

密度関数と分布関数は $x > 0$ に対して，
分布関数 $F(x) = 1 - exp\{-(x/\beta)^{\alpha_0}\}$ と与えられ，微分すると密度関数 $f(x) = (\alpha_0/\beta)(x/\beta)^{\alpha_0 - 1} exp\{-(x/\beta)^{\alpha_0}\}$ が求まる．
$\gamma(i) \equiv \Gamma\left(1 + \frac{i}{\alpha_0}\right)$　$(i = 1,\ 2)$ で定義する．ワイブル分布の平均と分散はガンマ関数を用いて
$E(X) = \beta\gamma(1)$，$Var(X) = \beta^2\left(\gamma(2) - \gamma^2(1)\right)$ と表すことができる．

3　多群ワイブルモデル

以下の表は本研究のモデルを白石 [1] を参照し，設定したものである．

表 1　k 群ワイブルモデル

水準	群	データ	平均	分布
対照	第 1 群	$X_{11}, \cdots, X_{1n_1}$	$\beta_1\gamma(1)$	$We(\alpha_0, \beta_1)$
処理 1	第 2 群	$X_{21}, \cdots, X_{2n_2}$	$\beta_2\gamma(1)$	$We(\alpha_0, \beta_2)$
⋮	⋮	⋮　⋮　⋮	⋮	⋮
処理 $k-1$	第 k 群	$X_{k1}, \cdots, X_{kn_k}$	$\beta_k\gamma(1)$	$We(\alpha_0, \beta_k)$

総標本サイズ：$n \equiv \quad n_1 + \cdots + n_k$(すべての観測値の個数)

$\beta_1, \ldots, \beta_k$はすべて未知とする

第 1 群の対照群と第 i 群の処理群を比較することを考える．1 つの比較のための検定は，

$$帰無仮説 \quad H_i: \quad \beta_i = \beta_1$$

とする．帰無仮説のファミリーは，

$$H_D \equiv \{H_2, \ldots, H_k\} = \{H_i | 2 \leq i \leq k\} \quad である．$$

4　分散安定化変換による統計量の漸近理論

漸近理論を述べるために，

$$(条件 1) \qquad \lim_{n \to \infty} \frac{n_i}{n} = \lambda_i > 0 \qquad (1 \leq i \leq k)$$

を仮定する．
定理を証明するために必要な式を補題として示す．

【補題 1】$X_1, X_2, \ldots, X_n \sim We(\alpha_0, \beta)$ としたとき，次の (1),(2) が成り立つ．

$$\sqrt{n}(\bar{X}_n - \beta\gamma(1)) \xrightarrow{L} N\left(0, \beta^2\{\gamma(2) - \gamma^2(1)\}\right) \tag{1}$$

$$\sqrt{n}\{g(\bar{X}_n) - g(\beta\gamma(1))\} \xrightarrow{L} N(0, \{g'(\beta\gamma(1))\}^2[\beta^2\{\gamma(2) - \gamma^2(1)\}]) \tag{2}$$

□

補題 1 の証明は (1) は中心極限定理を適用すると求めることができる．
(2) はデルタ法を適用することで求めることができる．
しかし (2) の漸近分布の分散は未知母数 β に依存しているため対数変換をすると，

$$\sqrt{n}\{\log(\bar{X}_n) - \log(\beta\gamma(1))\} \xrightarrow{L} N\left(0, \frac{\gamma(2)}{\gamma^2(1)} - 1\right) \tag{3}$$

となる．

【補題 2】　次の式が成り立つことを示す．

$$\sqrt{\frac{\gamma^2(1)}{\gamma(2) - \gamma^2(1)}}\sqrt{n}\{\log(\bar{X}_i.) - \log(\beta_i\gamma(1))\} \xrightarrow{L} Y_i \sim N\left(0, \frac{1}{\lambda_i}\right) \tag{4}$$

□

補題 2 の証明は (条件 1) の下で，補題 1，中心極限定理，スラツキーの定理を適用することによって示すことができる．

$$T_i \equiv \frac{\sqrt{\frac{\gamma^2(1)}{\gamma(2) - \gamma^2(1)}}\{\log(\bar{X}_i.) - \log(\bar{X}_1.)\}}{\sqrt{\frac{1}{n_i} + \frac{1}{n_1}}} \tag{5}$$

とする．このとき次の定理を得る．

【定理 3】　(条件 1) の下で

$$\lim_{n \to \infty} P_0\left(\max_{2 \leq i \leq k} |T_i| \leq t\right) = B_1(t), \tag{6}$$

$$\lim_{n \to \infty} P_0\left(\max_{2 \leq i \leq k} T_i \leq t\right) = B_2(t) \tag{7}$$

が成り立つ．ただし

$$B_1(t) \equiv \int_{-\infty}^{\infty} \prod_{i=2}^{k} \left\{\Phi\left(\sqrt{\frac{\lambda_i}{\lambda_1}} \cdot x + \sqrt{\frac{\lambda_i + \lambda_1}{\lambda_1}} \cdot t\right) - \Phi\left(\sqrt{\frac{\lambda_i}{\lambda_1}} \cdot x - \sqrt{\frac{\lambda_i + \lambda_1}{\lambda_1}} \cdot t\right)\right\} d\Phi(x),$$

$$B_2(t) \equiv \int_{-\infty}^{\infty} \prod_{i=2}^{k} \Phi\left(\sqrt{\frac{\lambda_i}{\lambda_1}} \cdot x + \sqrt{\frac{\lambda_i + \lambda_1}{\lambda_1}} \cdot t\right) d\Phi(x)$$

とする．

[証明]　補題 2 より，

$$\lim_{n\to\infty} P_0\left(\max_{2\le i\le k} |T_i| \le t\right) = P_0\left(\max_{2\le i\le k} \frac{|Y_i - Y_1|}{\sqrt{\frac{1}{\lambda_i}+\frac{1}{\lambda_1}}} \le t\right)$$

を得る．そこで，$\sigma_i^2 = 1/\lambda_i, \sigma_1^2 = 1/\lambda_1 \quad (i = 2,\ldots,k)$ とし白石 [2] 定理 A.7 の (A.4) を適用すると，

$$\lim_{n\to\infty} P_0\left(\max_{2\le i\le k} |T_i| \le t\right) = B_1(t)$$

が導かれ，(6) を得る．同様に (7) も白石 [2] の定理 A.7 の (A.5) を適用すると導くことができる．□

そこで α を与え，

$B_1(t) = 1 - \alpha$ を満たす t の解を $b_1(k, \lambda_1, \ldots, \lambda_k; \alpha)$

$B_2(t) = 1 - \alpha$ を満たす t の解を $b_2(k, \lambda_1, \ldots, \lambda_k; \alpha)$

とし，$b_1(k, \lambda_1, \ldots, \lambda_k; \alpha)$ を $b_1(\alpha)$，$b_2(k, \lambda_1, \ldots, \lambda_k; \alpha)$ を $b_2(\alpha)$ と表記する．このとき定理 3 より，次の漸近的な多重比較法が導かれる．

5　対数変換を使った漸近的な多重比較検定法

平均母数の制約に応じて，水準 α の漸近的な多重比較検定は次の (1) から (3) で与えられる．

(1)　両側検定:帰無仮説 H_i vs. 対立仮説 $H_i^{A\pm}$:$\beta_i \neq \beta_1 (i = 2,\ldots,k)$ のとき，
$|T_i| > b_1(\alpha)$ となる i に対して H_i を棄却し，対立仮説 $H_i^{A\pm}$ を受け入れ，$\beta_i \neq \beta_1$ と判定する．

(2)　片側検定:帰無仮説 H_i vs. 対立仮説 H_i^{A+} : $\beta_i > \beta_1 (i = 2,\ldots,k)$(制約$\beta_2,\ldots,\beta_k \ge \beta_1$) がつけられるとき，$T_i > b_2(\alpha)$ となる i に対して H_i を棄却し，対立仮説 H_i^{A+} を受け入れ $\beta_i > \beta_1$ と判定する．

(3)　片側検定:帰無仮説 H_i vs. 対立仮説 H_i^{A-}:$\beta_i < \beta_1 (i = 2,\ldots,k)$(制約$\beta_2,\ldots,\beta_k \le \beta_1$) がつけられるとき，$-T_i > b_2(\alpha)$ となる i に対して H_i を棄却し，対立仮説 H_i^{A-} を受け入れ $\beta_i < \beta_1$ と判定する．

6　対数変換を使った漸近的な同時信頼区間

$\log(\beta_i) - \log(\beta_1)$ に対する信頼係数 $1 - \alpha$ の同時信頼区間は，次の (1) から (3) によって与えられる．ただし $(i = 2,\ldots,k)$ とする．

(1)　両側信頼区間:

$$\log(\bar{X}_{i\cdot}) - \log(\bar{X}_{1\cdot}) - \frac{b_1(\alpha)\sqrt{\gamma(2) - \gamma^2(1)}\sqrt{\frac{1}{n_i}+\frac{1}{n_1}}}{\sqrt{\gamma^2(1)}}$$

$$< \log(\beta_i) - \log(\beta_1) < \log(\bar{X}_{i\cdot}) - \log(\bar{X}_{1\cdot}) + \frac{b_1(\alpha)\sqrt{\gamma(2) - \gamma^2(1)}\sqrt{\frac{1}{n_i}+\frac{1}{n_1}}}{\sqrt{\gamma^2(1)}}$$

(2)　上側信頼区間 (制約 $\beta_2,\ldots,\beta_k \ge \beta_1$ がつけられるとき):

$$\log(\bar{X}_{i\cdot}) - \log(\bar{X}_{1\cdot}) - \frac{b_2(\alpha)\sqrt{\gamma(2) - \gamma^2(1)}\sqrt{\frac{1}{n_i}+\frac{1}{n_1}}}{\sqrt{\gamma^2(1)}} < \log(\beta_i) - \log(\beta_1) < \infty$$

(3)　下側信頼区間 (制約 $\beta_2,\ldots,\beta_k \le \beta_1$ がつけられるとき):

$$-\infty < \log(\beta_i) - \log(\beta_1) < \log(\bar{X}_{i\cdot}) - \log(\bar{X}_{1\cdot}) + \frac{b_2(\alpha)\sqrt{\gamma(2) - \gamma^2(1)}\sqrt{\frac{1}{n_i}+\frac{1}{n_1}}}{\sqrt{\gamma^2(1)}}$$

7　ダネット型による実行プログラム

大橋 [3] を下に作成した主プログラムの流れを示す．

1. hyou1();, hyou2(); にて予め，$b1(k, 1/k, ..., 1/k; \alpha)$, $b2(k, 1/k, ..., 1/k; \alpha)$ の値を設定した配列を用意する．
2. input(); にて群の個数 k, ワイブル分布のパラメータ α_0, 有意水準 α, 信頼区間 c の値を入力する．
3. hantei(); にて k, α_0, α, c の値が検定が行える範囲内か確認し，output(); にて，値に沿った $b1(k, 1/k, ..., 1/k; \alpha)$, $b2(k, 1/k, ..., 1/k; \alpha)$ の値を出力する．
4. input2(); にて各群のサイズとそれぞれの観測値を入力する．
5. keisan(); にて, T_i を計算する．
6. kekka(); にて検定結果を出力する．

8　おわりに

C 言語によるプログラムをつくり Website[4] から地震のデータをとり，解析をした結果，東日本大震災直後は震度 3 以上の地震の回数が非常に多く，地震の間隔がとても短く，他の月と比べ特徴のある月だとわかった．

参考文献

[1] 白石高章 2013.『多群指数モデルにおける平均パラメータの多重比較法』．計算生物学，34．1-20.

[2] 白石高章 2011.『多群連続モデルにおける多重比較法 パラメトリック，ノンパラメトリックの数理統計』．共立出版株式会社

[3] 大橋一之.『すべての平均相違の多重比較統計量の分布論』．2012 年度南山大学情報理工学部情報システム数理学科卒業論文，2014 年 1 月．

[4] 日本気象協会 tenki.jp
http://bousai.tenki.jp/bousai/earthquake/