例数不揃いデータにおける Williams 多重比較検定法の検証

杉本 忠則

1. はじめに

最近、中西らは日本薬理学雑誌において、事前の試験計画で例数を揃えた場合試験途中で欠損例が発生することにより例数不揃いとなっても Williams 多重比較検定法 (Williams 法) では第 1 種の過誤確率が有意水準を上回らないことをシミュレーションにより示し、同検定法が利用できることを報告した(1)。Williams 法については、例数が揃っていることを前提とされてきたため、従来の考え方とは矛盾しているようにも思われる。最近の研究により Williams 法が例数不揃いでも適用可能であることが数理的に証明されたと思われた読者がいるかもしれないが、そのようなことはない。Williams 法は例数が揃った場合にのみ採用できる検定手法であり、例数不揃いデータに対しては第 1 種の過誤確率が有意水準を保てなくなる場合がある。このことは、統計解析の専門書にも書かれているが、どのような場合にどのような不都合があるかまでは説明されていない。中西論文では欠損例の発生については乱数に委ねているため、個別欠損例の結果が分からない。そこで、本論文では欠損例のパターン毎に Williams 法を適用しどのような不都合が起きるかをシミュレーションにより検討した。

2. 検証方法

中西論文では帰無仮説下で発生させたデータに対し Williams 法を適用し、有意と判定された割合が 2 項分布すると仮定している。シミュレーションは複数回行い、各割合が 2 項分布の 95% 信頼区間に対しどのように収まるかを調べることににより検定手法の妥当性を判断している。本論文でもこれに倣い検証したが、詳しくは以下の手順に従う。有意水準については、中西論文と同じく片側 0.025 とした。また、本シミュレーションでは 2 億回の検定が行われるため、最後に 2 億回のシミュレーションデータによる第 1 種の過誤確率の点推定値と 95% 信頼区間も算出した。

手順 1. 対照群を含めて 4 群で、各群例数 (表 1: 電子付録) を設定する。

手順 2. 帰無仮説下のデータとして正規分布 $N(0, 1)$ に従う乱数を発生させ、式 (1) に示す検定統計量 (t_p) を算出する (2)。

$$t_p = \frac{M_1 - \bar{x}_0}{\sqrt{V_E \left(\frac{1}{n_1} + \frac{1}{n_0} \right)}} \quad (1)$$

M_1 : 薬物群 1 の平均値または、薬物群 1 と他群とを併合した場合の平均値

$\bar{x}_0$: 対照群の平均値

n_1 : 薬物群 1 の例数

n_0 : 対照群の例数

V_E : 全体の誤差分散

上記の式の値が、Williams 法の上側 2.5% 点値を超える場合有意と判定する。

手順 3. 手順 2 を 10 万回実施し、有意と判定される割合 (p_1) を求める。なお、 p_1 は 10 万例のシミュレーションデータから求まる第 1 種の過誤確率の点推定値となる。

手順 4. 手順 3 で求めた p_1 が式 (2) で示す範囲に入るかどうかを調べる。

$$0.025 - 1.96 \sqrt{\frac{0.025 \times 0.975}{100000}} \sim 0.025 + 1.96 \sqrt{\frac{0.025 \times 0.975}{100000}} \quad (2)$$

式 (2) は二項分布 $B(0.025, 100000)$ の 95% 信頼区間であるので、手順 2 で有意と判定される割合 (第 1 種の過誤確率) の母平均値が 0.025 であれば、この区間に収まる確率は 95% と期待される。

手順 5. 手順 2~4 を 2000 回繰り返し、 p_1 が式 (2) の区間に入る回数 (p_{1_in})、下限値を下回る回数 (p_{1_lower})、上限を上回る回数 (p_{1_upper})、 p_1 の平均値 (p_{1_mean}) を求める。 p_{1_mean} は 2 億例のシミュレーションデータから求まる第 1 種の過誤確率の点推定値であるため、その 95% 信頼区間を式 (3) より求める。

キーワード: Williams 多重比較検定法, 多重比較, シミュレーション

大日本住友製薬株式会社 研究管理部 (〒564-0053 大阪府吹田市江の木町 33-94) 原稿受理日: 2015 年 1 月 14 日

Title: Verification of Williams multiple comparison in unbalanced data Author: Tadanori Sugimoto

$$\begin{aligned}
 & p1_mean - 1.96 \sqrt{\frac{p1_mean(1-p1_mean)}{200000000}} \sim \\
 & p1_mean + 1.96 \sqrt{\frac{p1_mean(1-p1_mean)}{200000000}} \quad (3)
 \end{aligned}$$

3. 検証プログラム

検証は統計解析ソフトウェア R (Ver. 3.1.1) により行った。今回のシミュレーションを実行した関数 (funcWilliams) のソースコードを Appendix (電子付録) に示す。funcWilliams は、4 群での Williams 法検証のシミュレーションを実施するために作成した関数である。入力引数に、乱数シードと各群の例数を入れて実行すると、乱数シード、手順 2 の実施回数、手順 2~4 の実施回数、p1_in の回数、p1_lower の回数、p1_upper の回数、p1_mean、p1_mean の 95% 信頼区間下限値、p1_mean の 95% 信頼区間上限値の順に出力される。例えば、funcWilliams (1060, 6, 6, 6, 6) を実行すれば、表 1 の No. 1 の群数例数構成でのシミュレーションが実行され結果が出力される。なお、計算量が多いため funcWilliams (1060, 6, 6, 6, 6) の結果が出力されるまで、著者の解析環境 (CPU: Core i5 3320M 2.6 GHz) では 40 時間以上要した。

4. 結果と考察

検証した検定手法が適切なものであれば、第 1 種の過誤確率は事前に設定された有意水準を超えることはない。しかし、今回の検証作業はシミュレーションのため、正規乱数発生により生ずる誤差がある。このため、第 1 種の過誤確率の発生が有意水準を母平均とする二項分布に従うとして、その 95% 信頼区間に収まるかどうかを確認することにより評価した。また、95% 信頼区間に収まらない場合は、95% 信頼区間の下限を下回る数と 95% 信頼区間の上限を上回る数の大小関係を調べることにした。このため、結果を次の 3 種類に分けた。「結果 1」は 95% 信頼区間に全体の約 95% 以上が収まる場合とする。「結果 2」は「結果 1」とは判定されず「95% 信頼区間の下限を下回る回数 < 95% 信頼区間の上限を上回る回数」となる場合とする。「結果 3」は「結果 1」とは判定されず「95% 信頼区間の下限を下回る回数 > 95% 信頼区間の上限を上回る回数」となる場合とする。なお、「結果 1」において、約 95% としているが、これはシミュレーションによる誤差があるためであり、今回の判断基準では 94% とした。検定手法を評価する場合、「結果 1」となれば、「第 1 種の過誤確率 = 有意水準」であるため適切な検定手法と言える。「結果 2」となれば、「第 1 種の過誤確率 > 有意水準」であるため不適切な検定手法である。「結果 3」となれば、「第 1 種の過誤確率 < 有意水準」であるため検定手法として用いても構わないが、保守的

な検定手法である。

シミュレーション結果の詳細は表 1 に示すが、例数を過不足させる群が同じである場合、同じ結果となったため表 1 の結果は表 2 (電子付録) としてまとめることができた。全群で例数が揃っている場合は、以前から知られていた通り第 1 種の過誤確率と有意水準とは一致した。対照群で不足、高用量群で過剰、あるいは中用量群で不足が起きた場合、「第 1 種の過誤確率 > 有意水準」となり、検定手法として不適切であることがわかった。対照群で過剰、高用量群で不足、あるいは中用量群で過剰が起きた場合、「第 1 種の過誤確率 < 有意水準」となり、保守的な検定手法であることがわかった。低用量群では過不足が起きても第 1 種の過誤確率と有意水準とは一致した。以上より、Williams 法は例数が揃っている場合には検定手法として利用できることが確認できた。一方、高々 10 万回での信頼区間での結果ではあるが、数学的厳密さと言う観点からは例数が揃っていない場合には不都合が確認できた。ただし、統計手法として不適切とされた場合でも、第 1 種の過誤確率は有意水準を高々 0.004 超えただけであり、近似的な使用であれば大きな問題にはならない。今回のシミュレーションでは 2 例以上の欠損での検討結果を示していないが、その場合より大きなズレが生じるので注意を払う必要がある (データは示していないが、対照群の 2 例欠損ではより大きなズレになることを確認している)。例数の過不足が生じる場合、第 1 種の過誤確率と有意水準の関係が例数の過剰と不足とで全く逆の関係になることは大変興味深い。今回のシミュレーションに使用したプログラムは Appendix に示しているので、興味のある方はこれを参考に別の群数例数構成でも検討して頂きたい。

Williams 法は検出力が高いことより、これを活用することにより少ない例数で結論を導くことができる。薬理実験では例数を増やすことが難しいことが多いため、Williams 法を採用することを考えている薬理研究者も多いと思われる。しかし、理論的には例数が揃っていることが条件であり、例数不揃いの場合には小さいとはいえ今回示したような問題が残っている。薬理実験の解析に高い精度が求められるかは、場合によるため何とも言えないが、薬理研究者には Williams 法を正しく理解した上で採用して頂きたい。

著者の利益相反：杉本忠則 (大日本住友製薬株式会社)。

文 献

- 1) 中西展大, 他. 日薬理誌. 2014;144:185-191.
- 2) 永田靖, 他. 統計的多重比較法の基礎. サイエンス出版社; 1997. p. 45-52, 176-177.

[表 1, 2 および Appendix は J-STAGE (<https://www.jstage.jst.go.jp/browse/fpj/-char/ja/>) の電子付録に掲載しています。]