

途中打ち切りありの生存時間解析における例数設計法の比較

M2014SS005 水野怜奈

指導教員：松田眞一

1 はじめに

生存時間解析とは、人に対して観察を始めてから、死亡などのイベントが起こるまでの時間をデータとして解析する統計的方法である。時間は、観察をはじめてからイベントが起こるまでの日数などを指し、イベントは死亡、再発など任意の事象である。(Kleinbaum and Klein[3] 参照)

本研究では、Rによってプログラミングを作成して、シミュレーションを行う。ログランク検定に対する3つの例数設計方法 Freedman[2], Rubinstein et al.[7], Lakatos[4] による方法で打ち切りデータにおける例数設計を行い、ログランク検定での検出力を比較し、打ち切りの発生状況によって、どの例数設計法を使用するのが適しているのかを検討する。

1.1 先行研究について

Lakatos[5]では、Rubinstein, Freedman, Lakatosの方法で指数モデル、比例ハザードモデル、非比例ハザードモデルによるシミュレーションを検出力80%で行っていたが、途中脱落による打ち切りについての研究は行われていない。また、高橋[8]では、RubinsteinとFreedmanの方法については、打ち切りがある場合と打ち切りがない場合について検出力の評価とシミュレーションは行われたが、Lakatosの方法については、理論のみで検討がされていなかった。水澤・浜田[6]では、Lakatos, Freedman, Schoenfeldの方法において指数分布、ワイブル分布、区分指数分布の症例数設計法の比較がされていた。

2 生存時間解析について

2.1 打ち切り

打ち切りは、生存時間解析において考慮するもので、生存時間についての情報は持っているが、正確に分からない場合に起きる。本研究は、打ち切りありの場合を主に考えている。(Kleinbaum and Klein[3] 参照)

1. 試験終了時まで、その人にイベントが起こらなかった場合
2. 試験期間の間に、その人が観察不能となった場合
3. 死亡原因が興味のあるイベントでない場合や他の理由により、その人が試験から脱落した場合

3 ログランク検定

ログランク検定は、 Kaplan-Meier 曲線で示される生存データの比較のために使われる検定で、ログランク検定の検定統計量は、群ごとの観察度数 (O) と期待度数 (E)

の差の合計を使う。

$$O_i - E_i = \sum_{f=1}^k (m_{if} - e_{if}) \quad (1)$$

e_{if} : リスクセットにおける各群の割合と2群のイベントの起こった数をかけたもの

m_{if} : i 番目群の f 番目の観察期間における時間区切りのイベントの起こった数

$i = 1, 2$: 群番号

$f = 1, 2, \dots, k$: 観察期間における時間区切りの番号

2群のログランク検定の場合、検定統計量は、一方の群の「観察度数 - 期待度数」の合計の2乗を「観察度数 - 期待度数」の合計の分散で割ることで求められる。 $i = 1, 2$ で、

$$\frac{(O_i - E_i)^2}{\text{Var}(O_i - E_i)} \quad (2)$$

となる。(Kleinbaum and Klein[3] 参照)

4 Lakatos の方法

より一般的な条件で、指数モデルに基づいた通常の検定は一定のハザード比では適用されず、ハザード関数とは一定ではない。よって、ノンコンプライアンス、ラグタイムなど無制限の割合がある状態でログランク検定による生存曲線の比較のための例数設計を推定する方法である。(Lakatos[4] 参照)

Z_α : 有意水準 α に対応する $N(0, 1)$ の分位点

Z_β : 検出力 $1 - \beta$ に対応する $N(0, 1)$ の分位点

m_h : h 番目の死亡直前の実験群のリスク集合の大きさ

n_h : h 番目の死亡直前の対照群のリスク集合の大きさ

W_h : Tarone-Ware の重み

P_{jik} : i 番目区間の k 番目の死亡直前の j 群のハザード

ϕ_{ik} : i 番目区間の k 番目の死亡直前の2群のリスク集合の大きさの比

θ_{ik} : i 番目区間の k 番目の死亡直前の2群のハザード比

まず、 B 等区間で試験期間を分割し、 i 番目の区間での死亡数が d_i である場合、両群の総死亡数 d は、

$$d = \sum_{i=1}^B d_i$$

となり、また、

$$\phi_{ik} = \frac{m_h}{n_h}, \theta_{ik} = \frac{P_{1ik}}{P_{2ik}}$$

と定義し、 i 番目区間での死亡数が全体の死亡数のどれだけの割合になるかを

$$\rho_i = \frac{d_i}{d}$$

とする。

ログランク統計量である Tarone-Ware 統計量を使い、その期待値は、ハザード関数の形について仮定していないので、局所的に条件を固定し、近似値は、

$$E = \frac{\sum_{i=1}^B \sum_{k=1}^{d_i} W_{ik} \left(\frac{\phi_{ik}\theta_{ik}}{1+\phi_{ik}\theta_{ik}} - \frac{\phi_{ik}}{1+\phi_{ik}} \right)}{\left[\sum_{i=1}^B \sum_{k=1}^{d_i} \frac{W_{ik}^2 \phi_{ik}}{(1+\phi_{ik})^2} \right]^{\frac{1}{2}}} \quad (3)$$

となり、すべての i と k において、 $W_{ik} = 1$ のときに得られる。また、この統計量は $N(E, 1)$ として扱うことで、

$$E = Z_{\alpha/2} + Z_{\beta} \quad (4)$$

をもつ。

すべての i, k において、 $\phi_{ik} \equiv \phi_i$ と $W_{ik} \equiv W_i$ と仮定し、

$$e(D) = \frac{\sum_{i=1}^B W_i \rho_i \gamma_i}{\left(\sum_{i=1}^B W_i^2 \rho_i \eta_i \right)^{\frac{1}{2}}} \quad (5)$$

$$\gamma_i = \frac{\phi_i \theta_i}{1 + \phi_i \theta_i} - \frac{\phi_i}{1 + \phi_i}, \quad \eta_i = \frac{\phi_i}{(1 + \phi_i)^2}$$

とおくと、式 (3) は、

$$E = e(D) \sqrt{d} \quad (6)$$

となる。 d において、式 (4)、(6) を解くと、

$$\sqrt{d} = \frac{(Z_{\alpha/2} + Z_{\beta})}{e(D)}$$

となり、

$$d = \frac{[Z_{\alpha/2} (\sum_{i=1}^B W_i^2 \rho_i \eta_i)^{\frac{1}{2}} + Z_{\beta} (\sum_{i=1}^B W_i^2 \rho_i \eta_i)^{\frac{1}{2}}]^2}{(\sum_{i=1}^B W_i \rho_i \gamma_i)^2} \quad (7)$$

が与えられ、必要例数 N は、

$$N = \frac{2d}{(P_c + P_e)} \quad (8)$$

のように、実験群と対照群の累積死亡率 P_e 、 P_c を使って得られ、打ち切りがある場合は Freedman の方法の同様に脱落率 w を用いて、

$$N = \frac{2d}{(P_c + P_e)} \cdot \frac{1}{1 - w} \quad (9)$$

とする。

5 Freedman の方法

Freedman の方法は、Freedman[2] および臨床試験ハンドブック [9] を参照のこと。

6 Rubinstein の方法

与えられた有意水準、死亡ハザード率、発生率、追跡期間とフォローアップによる損失において、期待される統計的検出力を保証するために必要とされる臨床試験の長さを計算するための方法である。(Rubinstein et al.[7] 参照)

必要とされる臨床試験の長さ T 年の間、患者はポアソン過程に従い、単位時間あたりの患者数 n が参加することによって、必要例数 N は、

$$N = nT \quad (10)$$

と求めることができる。

7 打ち切りなしの場合のシミュレーション

$\alpha = 0.05$ と $\beta = 0.1, 0.2$ と設定し、 $\theta = (1.33, 1.5, 2.0, 3.0)$ 、試験期間は 1 年という条件でシミュレーションを行い、Lakatos, Freedman, Rubinstein の方法で例数設計を行う。このシミュレーションは、必要例数を求めるために真値となる生存率の引数を与えて、ランダムに生存時間データを与えて必要例数を計算し繰り返す。よって、データによる結果がどのような分布となるかを得ることができる。また、繰り返す際に、Lakatos の方法では試験期間を分割し、区間ごとの死亡数と生存数が必要であるので、行列を作成している。Freedman と Rubinstein の方法は、行列から得られる総生存数から生存率を求めている。

まず、シミュレーションで用いる θ における実験群と対照群それぞれの生存率 $P1, P2$ を定義する。(表 1 参照)

表 1

| θ | $P1$ | $P2$ |
|----------|--------|--------|
| 1.33 | 0.4507 | 0.5492 |
| 1.5 | 0.4301 | 0.5698 |
| 2.0 | 0.3819 | 0.6180 |
| 3.0 | 0.3176 | 0.6823 |

$\beta = 0.1$ のときの 3 つの方法の $\theta = (1.33, 1.5, 2.0, 3.0)$ のときのそれぞれ平均値、中央値、真値の結果を表 2, 表 3, 表 4 に示す。

表 2 Lakatos の方法の結果

| θ | 平均値 | 中央値 | 真値 |
|----------|----------|--------|--------|
| 1.33 | 481364.4 | 1210.5 | 1207.2 |
| 1.5 | 5482280 | 596.1 | 597.8 |
| 2.0 | 243.1 | 207.0 | 207.8 |
| 3.0 | 90.6 | 85.8 | 86.0 |

8 打ち切りありの場合のシミュレーション

打ち切りなしの場合と同様の条件でシミュレーションを行い、3 つの方法で打ち切りデータにおける例数設計を行う。打ち切りなしの場合のシミュレーションと異なる点

表3 Freedmanの方法の結果

| θ | 平均値 | 中央値 | 真値 |
|----------|--------|--------|--------|
| 1.33 | 9019.1 | 1020.1 | 1047.6 |
| 1.5 | 1973.7 | 523.6 | 523.3 |
| 2.0 | 221.0 | 190.3 | 189.1 |
| 3.0 | 88.0 | 83.8 | 84.0 |

表4 Rubinsteinの方法の結果

| θ | 平均値 | 中央値 | 真値 |
|----------|--------|-------|-------|
| 1.33 | 4985.9 | 977.0 | 995.2 |
| 1.5 | 1387.2 | 649.6 | 650.8 |
| 2.0 | 385.8 | 362.1 | 362.3 |
| 3.0 | 237.1 | 232.9 | 232.7 |

は, Lakatosの方法で用いられる行列で, 区間ごとの死亡数と生存数に加えて, 脱落率による打ち切りありの列を追加し, 打ち切りがあった数を求める. その後の計算は, 打ち切りなしと同様にそれぞれの例数設計法のプログラミング内によって必要例数を求めることができる.

脱落率を区間ごとに変えて, 3つの方法についての結果を述べる. w1は実験群の脱落率, w2は対照群の脱落率とする.

- パターン1(均一型):
w1とw2が(0.03,0.03,0.03,0.03,0.03)のとき
- パターン2(偏り型):
w1が(0.08,0.08,0.08,0.08,0.08),
w2が(0.03,0.03,0.03,0.03,0.03)のとき
- パターン3(漸増型):
w1とw2が(0.01,0.03,0.05,0.07,0.09)のとき
- パターン4(ジグザグ型):
w1とw2が(0.03,0.08,0.03,0.08,0.03)のとき

$\beta = 0.1$ のときの3つの方法の必要例数の真値を比較したものを表5, 表6, 表7, 表8に示す.

表5 パターン1(均一型)のときの必要例数の真値

| θ | 1.33 | 1.5 | 2.0 | 3.0 |
|------------|--------|-------|-------|-------|
| Lakatos | 1230.2 | 612.1 | 212.5 | 87.8 |
| Freedman | 1080.0 | 541.6 | 194.9 | 86.6 |
| Rubinstein | 1007.1 | 656.1 | 363.9 | 233.4 |

表6 パターン2(偏り型)のときの必要例数の真値

| θ | 1.33 | 1.5 | 2.0 | 3.0 |
|------------|--------|-------|-------|-------|
| Lakatos | 1837.3 | 808.6 | 250.1 | 97.3 |
| Freedman | 1108.6 | 555.9 | 200.1 | 88.9 |
| Rubinstein | 1017.3 | 660.5 | 365.3 | 234.0 |

表7 パターン3(漸増型)のときの必要例数の真値

| θ | 1.33 | 1.5 | 2.0 | 3.0 |
|------------|--------|-------|-------|-------|
| Lakatos | 1244.6 | 619.2 | 214.9 | 88.8 |
| Freedman | 1102.7 | 553.0 | 199.0 | 88.4 |
| Rubinstein | 1015.2 | 659.6 | 365.0 | 233.9 |

表8 パターン4(ジグザグ型)のときの必要例数の真値

| θ | 1.33 | 1.5 | 2.0 | 3.0 |
|------------|--------|-------|-------|-------|
| Lakatos | 1249.7 | 621.8 | 215.7 | 89.0 |
| Freedman | 1102.7 | 553.0 | 199.0 | 88.4 |
| Rubinstein | 1015.2 | 659.6 | 365.0 | 233.9 |

9 検証

9.1 打ち切りなしの場合の検出力

表2, 表3, 表4に示した必要例数を用いて, シミュレーションより得られた検出力を示している. $\beta = 0.1$ のとき, 表9より, $\theta = 1.33$ のときは, どの例数設計法も検出力90%に近いことが分かる. しかし, どの例数設計法も $\theta = 1.33$ より大きくなると, 想定した検出力よりも離れていることが分かる. Rubinsteinの方法は検出力90%が保てていない. また, Lakatosの方法は少し過剰に評価されている. Freedmanの方法が検出力90%に近い値をとっているため, 打ち切りなしの場合, 今回の実験ではFreedmanの方法が優れていると思われる.

表9 打ち切りなしの場合の検出力

| θ | 1.33 | 1.5 | 2.0 | 3.0 |
|------------|--------|--------|--------|---------|
| Lakatos | 0.9397 | 0.9408 | 0.9357 | 0.94449 |
| Freedman | 0.903 | 0.906 | 0.9166 | 0.9411 |
| Rubinstein | 0.8876 | 0.9582 | 0.996 | 0.9999 |

9.2 打ち切りありの場合の検出力

表5, 表6, 表7, 表8に示した必要例数を用いて, シミュレーションより得られた検出力を示している. $\beta = 0.1$ のとき, 表10, 表12, 表13より, 打ち切りなしと同様, $\theta = 1.33$ のときは, どの例数設計法も検出力90%に近いことが分かる. また, Lakatosの方法は少し過剰に評価されており, Rubinsteinの方法に関しては検出力90%が保てていない. パターン1, パターン3, パターン4のとき, Freedmanの方法が優れていると思われる.

しかし, 表11だけは, 先に述べた結果とは異なった. Freedman, Rubinsteinの方法は想定した検出力より, かなり離れていると分かる. その点, Lakatosの方法が検出力90%に近い値をとり, 安定していた. しかし, $\theta = 1.33$ より大きくなると, 想定した検出力よりも離れているので注意が必要である. Freedmanの方法は, $\theta = 1.33$ より大きくなると, 想定した検出力に近くなり, Rubinsteinの

方法に関しては検出力 90% が保てていない。実験群だけ偏って大きくした脱落率にも Lakatos の方法は安定して検出力 90% を得られた。

表 10 パターン 1(均一型) のときの検出力

| θ | 1.33 | 1.5 | 2.0 | 3.0 |
|------------|--------|--------|--------|--------|
| Lakatos | 0.9353 | 0.9329 | 0.9357 | 0.9388 |
| Freedman | 0.9008 | 0.9017 | 0.9154 | 0.9388 |
| Rubinstein | 0.88 | 0.9452 | 0.9952 | 0.9999 |

表 11 パターン 2(偏り型) のときの検出力

| θ | 1.33 | 1.5 | 2.0 | 3.0 |
|------------|--------|--------|--------|--------|
| Lakatos | 0.927 | 0.9301 | 0.9355 | 0.9399 |
| Freedman | 0.7575 | 0.8187 | 0.8772 | 0.9166 |
| Rubinstein | 0.7202 | 0.8763 | 0.9869 | 0.9997 |

表 12 パターン 3(漸増型) のときの検出力

| θ | 1.33 | 1.5 | 2.0 | 3.0 |
|------------|--------|--------|--------|--------|
| Lakatos | 0.9332 | 0.9365 | 0.9403 | 0.9387 |
| Freedman | 0.9058 | 0.9075 | 0.9255 | 0.9387 |
| Rubinstein | 0.8803 | 0.9513 | 0.9959 | 0.9998 |

表 13 パターン 4(ジグザグ型) のときの検出力

| θ | 1.33 | 1.5 | 2.0 | 3.0 |
|------------|--------|--------|--------|--------|
| Lakatos | 0.935 | 0.9344 | 0.9373 | 0.9389 |
| Freedman | 0.9041 | 0.9063 | 0.9217 | 0.9389 |
| Rubinstein | 0.8786 | 0.9466 | 0.9952 | 0.9999 |

10 考察

まず、必要例数について述べる。本研究は、指数分布の下で例数設計を行った。打ち切りなしの場合と打ち切りがある場合も $\theta = 1.33, 1.5$ のとき、どの例数設計法に関しても外れ値があるため、中央値と真値に比べて平均値が大きくなっていることが分かった。また、Freedman と Rubinstein の方法は、脱落率を変化させても必要例数の真値にはあまり差がなかった。Lakatos の方法も真値はパターン 2 以外は差がなかった。これは全体の脱落率が 0.1 より小さいので、関係があるのではないかと考えられる。

次に検出力について述べる。本研究では、打ち切りなしの場合では Freedman の方法が想定した検出力に 1 番近い値を取っているので優れていると考えられる。打ち切りがある場合も Freedman の方法が想定した検出力に近い値を取っているが、Lakatos の方法のみが θ と脱落率を変えたどのパターンでも検出力 90% を得ている。本研究は、指数分布の下で区間ごとに脱落率を変化させているため、打

ち切りがあった人を区間ごとに計算して求めている。よって、打ち切りの発生状況においては、Lakatos の方法が脱落率に応じた適切な例数を求めることができるため、優れているのではないかと考えられる。また、Rubinstein の方法は臨床試験の長さを計算する方法であり、直接例数を求める例数設計法ではない。Freedman と Lakatos の方法は、患者の試験期間の最初から打ち切りがあるまでの時間を観察できる。しかし、Rubinstein の方法は試験期間内であれば、途中から参加することができるので、途中から参加した患者に関しては、その時点から観察しているため必要例数が増えてしまい、区間ごとの脱落率によって計算している本研究では、うまく検出力が保たれておらず、今回はあまり精度は高くないと考えられる。

11 おわりに

Lakatos の方法の必要例数を求めるプログラムの作成に多く時間を割いてしまったため、本研究では、指数分布でしか 3 つの例数設計法を比較することができなかった。今後、他の分布についても実験を行いたい。また、本研究以外の例数設計法との比較を行いたいと思った。

参考文献

- [1] Armitage, P., Berry, G. (椿美智子・椿広計 訳), 医学研究のための統計的方法, サイエンティスト社, 2001.
- [2] Freedman, L.S. 'Tables of the number of patients required in clinical trials using the log-rank test', *Statistics in Medicine*, **1**, 121-129, 1982.
- [3] Kleinbaum, D. G. and Klein, M. (神田英一郎・藤井朋子 訳), エモリー大学クライバウム教授の生存時間解析, サイエンティスト社, 2015.
- [4] Lakatos, E. 'Sample sizes based on the log-lank statistic in complex clinical trials', *Biometrics*, **44**, 229-241, 1998.
- [5] Lakatos, E., Gordon Lan, K.K. 'A comparison of sample size methods for the logrank statistic', *Statistics in Medicine*, **11**, 179-191, 1992.
- [6] 水澤純基・浜田知久馬, 生存時間解析における症例数設計方法の性能比較, SAS ユーザー総会 論文集, 19-20, 2008.
- [7] Rubinstein, L.V., Gail, M.H., and Santner, T.J. 'Planning the duration of a comparative clinical trial with loss to follow-up and a period of continued observation', *Journal of Chronic Diseases*, **34**, 469-479, 1981.
- [8] 高橋友弥, 2 つの生存関数の比較に必要なサンプルサイズ, 南山大学大学院数理情報研究科修士論文, 2010.
- [9] 丹後俊郎・上坂浩之 編, 臨床試験ハンドブック - デザインと統計解析 -, 朝倉書店, 2006.