

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

水野 怜奈* 松田 眞一†

E-Mail: matsu@nanzan-u.ac.jp

生存時間解析ではログランク検定で2つの群の生存関数の比較を行うことが多い。生存時間データでは打ち切りが起ることが多く、その場合の例数設計についての比較はあまり行われていない。本論文ではその際に用いられる例数設計法として Freedman, Lakatos, Rubinstein の方法の比較を行う。分布は指数分布に限ったため、Freedman の方法が全体的にはよかったが、途中打ち切りのパターンによっては Lakatos の方法の方が優れている場合があることを突き止めた。

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)

本論文では、Rによるシミュレーションを用いて、ログランク検定 (Armitage and Berry[1] 参照) 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。

先行研究として以下の3つの研究が挙げられる。Lakatos and Lan[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。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が、これも途中打ち切りは考慮していない。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。

2 打ち切りについて

打ち切りとは、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。本論文は、打ち切りありの場合を主に考えている。

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

*南山大学大学院理工学研究科システム数理専攻

†南山大学理工学部システム数理学科

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

の3つのタイプの打ち切りがある。(Kleinbaum and Klein[3] 参照)

本論文では、2番目と3番目の途中打ち切りが発生する状況を想定している。

3 用語と記号について

生存時間解析に関する用語と記号について述べる。以下では統一してイベントは死亡として扱う。

T : 生存時間の確率変数 ($T > 0$)

t : T での興味のあるイベント (死亡) の観察期間

d : $(0, 1)$ 確率変数で、

$d = 1$ はイベント (死亡) が起こる場合

$d = 0$ は打ち切りの場合

$S(t)$: 生存関数 (T が t をこえる確率)

$S_1(t)$: 実験群の生存関数

$S_2(t)$: 対照群の生存関数

本論文ではログランク検定を両側検定として下記の帰無仮説 H_0 と対立仮説 H_1 を対象とする。

$$H_0: S_1(t) = S_2(t)$$

$$H_1: S_1(t) \neq S_2(t)$$

次に、Lakatos, Freedman, Rubinstein の方法の中で使われる記号についてまとめる。

まず、数式で同じ表現として使われる記号についてまとめる。

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

θ : ハザード比

d : 両群の総死亡数

N : 両群での必要例数

w : 両群での脱落率

Lakatos の方法で使われる記号は下記のとおりである。

F : 実験群の生存時間分布

G : 対照群の生存時間分布

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

X_h : 対照群の指標 ($(0, 1)$ 確率変数, 1 : 対照群の死亡, 0 : 対照群以外の死亡)

P_{jik} : i 番目区間の k 番目の死亡直前の j 群のハザード

d_i : i 番目の区間の死亡数

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

P_e : 実験群の累積死亡率

P_c : 対照群の累積死亡率

4 3つの例数設計法

本論文で用いた方法のうち、Lakatos の方法は次節で詳述する。

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。

Rubinstein の方法は Rubinstein et al.[7] を参照のこと。

5 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、ハザード関数と比は一定ではない。Lakatos の方法とは、ノンコンプライアンス、ラグタイムなどが無制限の割合で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)

5.1 Lakatos の方法の求め方

Lakatos[4] を参照して Lakatos の方法での例数設計法を説明する。

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、

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

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。

$H_0: (1-F) = (1-G)$ vs. $H_1: (1-F) \neq (1-G)$ を検定するためのログランク統計量は 2×2 分割表をから計算されるのが普通である (Armitage and Berry[1] 参照) が、Tarone-Ware の重みを使って一般的に表すと、

$$L = \frac{\sum_{h=1}^d W_h \left(X_h - \frac{m_h n_h}{m_h + n_h} \right)}{\left[\sum_{h=1}^d W_h^2 \left(\frac{m_h n_h}{(m_h + n_h)^2} \right) \right]^{\frac{1}{2}}} \quad (1)$$

であり、これはハザード関数の形について仮定していないと気づく。よって、局所的に条件を固定し、式 (1) の近似の期待値は、

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

となる。ログランク検定に対応する期待値はすべての i と k において、 $W_{ik} = 1$ のときに得られるので下記の式となる。

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

また、ログランク統計量は $N(E, 1)$ (平均 E , 分散 1 に従う正規分布) として扱うことで、

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

が得られる。

一方、すべての i, k において、 $\phi_{ik} \equiv \phi_i$ と仮定し、

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

$$\gamma_i = \frac{\phi_i \theta_i}{1 + \phi_i \theta_i} - \frac{\phi_i}{1 + \phi_i}$$

$$\eta_i = \frac{\phi_i}{(1 + \phi_i)^2}$$

とおくと、式 (3) は、

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

となる。

d において、式 (4), (6) を解くと、

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

となり、

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

が得られる。計算に必要な各数値は区間ごとの死亡数、生存数のデータがあれば求めることができる。

よって、累積死亡率を使って

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

が成り立つことから、必要な例数は、

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

と得られる。

本論文での Lakatos の方法における打ち切り補正は、臨床試験ハンドブック [9] にある Freedman の方法の打ち切り補正の場合と同じように全体での脱落率 w を用いて、

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

とする。

6 例数設計方法のプログラミングについて

Lakatos, Freedman, Rubinstein の方法でそれぞれ打ち切りなしの場合と打ち切りありの場合のプログラムを R で作成した。本論文では、打ち切りありの場合を主に行っており、ここでは Lakatos の方法の打ち切りありのプログラミングの説明のみ述べる。

6.1 Lakatos 打ち切りありプログラム

Lakatos の方法のプログラムは Lakatos[4] に掲載されている SAS プログラムを参考にして作成した。引数は x , y , α , β である。 x , y はそれぞれ実験群と対照群の 5×3 行列の表を用いる。これらの表は行が試験期間の区間を表しており、今回は 5 区間とする。列はそれぞれ、区間終了時点の累積脱落数、累積死亡数、生存数である。なお、両群での脱落率 w はそれぞれの群の脱落率の平均を用いている。

7 打ち切りなしの場合のシミュレーション

本論文は、 $\alpha = 0.05$ 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、ログランク検定に対する 3 つの例数設計方法 Lakatos, Freedman, Rubinstein の方法で例数設計を行う。なお、Rubinstein の方法ではフォローアップ期間を 0 としている。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。ここから、データによる結果で必要例数がどんなばらつきとなるかという分布を得ることができる。また、Lakatos の方法が試験期間を区間ごとに分割し、区間ごとの死亡数と生存数を計算することが必要であるので、その行列を作成している。ここで得られた実験群と対照群それぞれの行列を直接 Lakatos のプログラムに入れることによって、必要例数が求められることができる。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、プログラムに入れることによって必要例数を求めることができる。

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 P_1 , P_2 を定義する。 P_1 , P_2 は生存率の平均が両群で 0.5 となる条件の下で、ハザード比が与えられた θ と一致するように求めた。(表 1 参照)

表 1: ハザード比と生存率の対応表

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

7.1 打ち切りなしシミュレーションのプログラム

シミュレーションのプログラムは引数として生存率 P1, P2 が指定できる。手順は以下の通りである。

1. 決められたハザード比により実験群と対照群それぞれの引数 P1, P2 を用いて 5 区間での死亡数, 生存数の行列を指数乱数を用いて作成する。試験期間は 1 年としたので, 各行は順番に 0.2, 0.4, 0.6, 0.8, 1.0 と時刻で 5 区間に分けている。今回は各群の例数を 200 とした。
2. 得られたデータに基づいて 3 つの例数設計法で必要例数を計算する。
3. 手順 1, 2 を 10000 回繰り返す。
4. 元々の理論値と Lakatos の方法のみ生成した行列をすべて合算したものをを用いて必要例数の真値も求める。Lakatos の方法のみ近似であるが, 10000 回繰り返しているのでほぼ収束している。

7.2 Lakatos の方法の結果

$\beta = 0.1, 0.2$ それぞれの $\theta = (1.33, 1.5, 2.0, 3.0)$ のときの平均値, 中央値, 真値の結果について表 2, 表 3 に示す。また, $\beta = 0.1$ のときの必要例数のヒストグラムを図 1 に示す。

$\theta = 1.33$ のとき, 図 1 を見ると分かるように, 真値はほとんど中央値に近い部分にあるが, 外れ値があるため平均値が大きくなっている。これは $\theta = 1.5$ のときも同様のことがいえる。また, $\theta = 2.0, 3.0$ の場合は, 平均値と中央値との差が小さくなっている。全体的にみると, θ が大きくなると中央値が真値に近づいてると分かる。 $\beta = 0.2$ のときも同様の変化があった。

表 2: Lakatos の方法の必要例数 ($\beta=0.1$)

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

表 3: Lakatos の方法の必要例数 ($\beta=0.2$)

θ	平均値	中央値	真値
1.33	359571.6	904.2	901.8
1.5	4095177	445.3	446.5
2.0	181.6	154.6	155.3
3.0	67.7	64.1	64.3

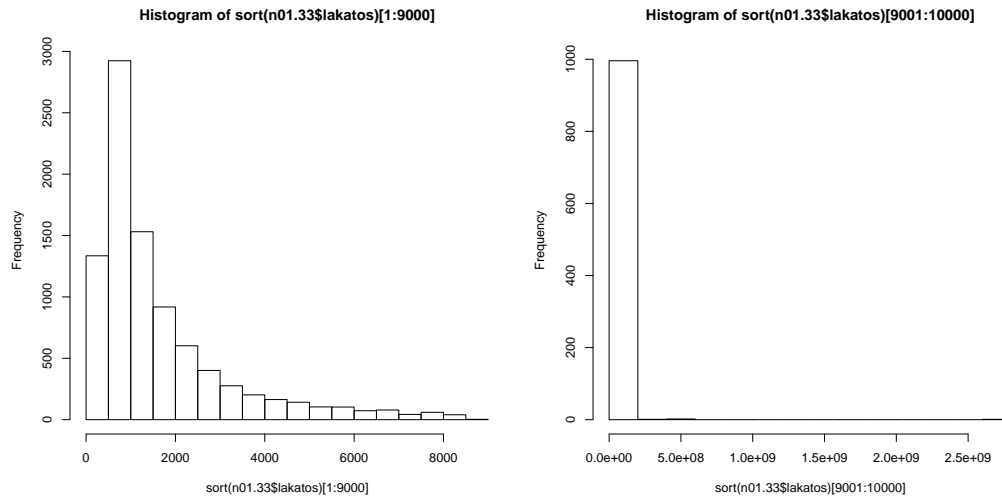


図 1: Lakatos の方法における推定されたサンプルサイズの分布 (左:小さいものから 9000 番目まで, 右:9001 番目以降)

7.3 Freedman の方法の結果

Lakatos の方法と同様に，平均値，中央値，真値を表 4，表 5 に示す。

表 4: Freedman の方法の必要例数 ($\beta = 0.1$)

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 5: Freedman の方法の必要例数 ($\beta = 0.2$)

θ	平均値	中央値	真値
1.33	6737.1	762.0	782.5
1.5	1474.3	391.1	392.4
2.0	165.1	142.1	141.2
3.0	65.7	62.6	62.7

Lakatos の結果と同様， $\theta = 1.33, 1.5$ は真値はほとんど中央値に近い部分にあるが，外れ値があるため平均値が大きくなっている。また，実験群と対照群で生存率が同じになった場合は無限大の値が出たので計算から外しているため，Lakatos の方法よりは平均値が安定しているように見える。 $\theta = 2.0, 3.0$ の場合は，平均値と中央値との差が小さくなっている。全体的にみると， θ が大きくなると中央値が真値に近づいており，Lakatos の方法と似たような減り方をしている。 $\beta = 0.2$ のときも同様の変化があった。

7.4 Rubinstein の方法の結果

同様に，平均値，中央値，真値を表 6，表 7 に示す。

Lakatos, Freedman の方法と同様， $\theta = 1.33, 1.5$ は真値はほとんど中央値に近い部分にあるが，外れ値があるため平均値が大きくなっている。また， $\theta = 1.33$ のときに無限大の値が出たので計算から外している。 $\theta = 2.0, 3.0$ の場合は，平均値と中央値との差が小さくなっている。 $\theta = 2.0, 3.0$ の真値の減り方は小さい。Lakatos, Freedman の方法と同様，中央値が真値に近づいている。 $\beta = 0.2$ のときも同様の変化があった。

7.5 比較

3つの方法の必要例数の真値を比較したのを表 8，表 9 に示す。

まず， $\beta = 0.1$ のときは $\theta = 1.33$ の場合，Rubinstein の方法が最も真値が小さくなっていて，Lakatos の方法が最も大きい。Rubinstein の方法は $\theta = 1.33$ のとき以外は真値が最も大きくなっている。Lakatos と Freedman の方法は θ が大きくなると，似たような減り方をしていて， $\theta = 3.0$ のときは，真値にはほとんど差がない。3つの方法すべて θ が大きくなるにつれて真値は小さくなっている。

表 6: Rubinstein の方法の必要例数 ($\beta = 0.1$)

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

表 7: Rubinstein の方法の必要例数 ($\beta = 0.2$)

θ	平均値	中央値	真値
1.33	3816.5	814.6	829.4
1.5	1105.8	548.2	549.0
2.0	328.4	308.7	309.0
3.0	203.3	199.8	199.6

表 8: 3つの方法の打ち切りなしにおける必要例数の真値 ($\beta = 0.1$)

θ	1.33	1.5	2.0	3.0
Lakatos	1207.2	597.8	207.9	86.0
Freedman	1047.6	523.3	189.1	84.0
Rubinstein	995.2	650.8	362.3	232.7

表 9: 3つの方法の打ち切りなしにおける必要例数の真値 ($\beta = 0.2$)

θ	1.33	1.5	2.0	3.0
Lakatos	901.8	446.5	155.3	64.3
Freedman	782.5	392.4	141.2	62.7
Rubinstein	829.4	549.0	309.0	199.6

そして、 $\beta = 0.2$ のときは $\theta = 1.33$ の場合、Freedman の方法が最も真値が小さくなっていた。Rubinstein の方法は $\theta = 1.33$ のとき以外は真値が最も大きくなっている。 $\beta = 0.1$ のときと同様に、Lakatos と Freedman の方法は θ が大きくなると、似たような減り方をしている、 $\theta = 3.0$ のときは、真値にはほとんど差がない。また、3つの方法すべて θ が大きくなるにつれて真値は小さくなっている。

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、Lakatos, Freedman, Rubinstein の方法で打ち切りありの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は、実験群と対照群それぞれ区間ごとの脱落率を追加し、打ち切りがあった脱落数を求めることである。脱落は一様乱数で各データに振り分けた。打ち切りがない場合で説明した Lakatos の方法で用いられる行列に、区間ごとの死亡数と生存数に加えて、脱落率による脱落数の列を追加する。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。このシミュレーションでは、1列目から脱落数、死亡数、生存数の順に、区間ごとのそれぞれ実験群と対照群の5×3行列を示している。

引数として脱落率のベクトル w_1 と w_2 が実験群と対照群それぞれで指定できる。これは実験データの値の区間ごとに脱落率を変えることができるということである。全体の脱落率 w は、 w_1 と w_2 それぞれの平均値を足して2で割って求め、今回は各群の例数を200としている。繰り返し回数は10000回である。

以下の4つのパターンで脱落率を区間ごとに变えて、3つの方法についての結果を述べる。本シミュレーションでも必要例数の分布を求めているが、打ち切りのない場合と比べて特筆すべきことはなかったので省略する。

- パターン1(均一型): w_1 が (0.03,0.03,0.03,0.03,0.03), w_2 が (0.03,0.03,0.03,0.03,0.03) のとき
- パターン2(偏り型): w_1 が (0.08,0.08,0.08,0.08,0.08), w_2 が (0.03,0.03,0.03,0.03,0.03) のとき
- パターン3(漸増型): w_1 が (0.01,0.03,0.05,0.07,0.09), w_2 が (0.01,0.03,0.05,0.07,0.09) のとき
- パターン4(ジグザグ型): w_1 が (0.03,0.08,0.03,0.08,0.03), w_2 が (0.03,0.08,0.03,0.08,0.03) のとき

8.1 パターン1(均一型)

パターン1の状況での3つの方法の必要例数の真値を比較したものを表10, 表11に示す。

表10: 3つの方法の打ち切りありにおける必要例数の真値 (パターン1, $\beta = 0.1$)

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11: 3つの方法の打ち切りありにおける必要例数の真値 (パターン1, $\beta = 0.2$)

θ	1.33	1.5	2.0	3.0
Lakatos	918.9	457.2	158.7	65.6
Freedman	806.7	404.5	145.6	64.7
Rubinstein	837.8	552.8	310.2	200.1

8.2 パターン2(偏り型)

パターン2の状況での3つの方法の必要例数の真値を比較したものを表12, 13に示す。

表 12: 3 つの方法の打ち切りありにおける必要例数の真値 (パターン 2, $\beta = 0.1$)

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 13: 3 つの方法の打ち切りありにおける必要例数の真値 (パターン 2, $\beta = 0.2$)

θ	1.33	1.5	2.0	3.0
Lakatos	1372.4	604.0	186.8	72.7
Freedman	828.1	415.2	149.5	66.4
Rubinstein	844.9	555.9	311.2	200.5

8.3 パターン 3(漸増型)

パターン 3 の状況での 3 つの方法の必要例数の真値を比較したものを表 14, 表 15 に示す。

表 14: 3 つの方法の打ち切りありにおける必要例数の真値 (パターン 3, $\beta = 0.1$)

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 15: 3 つの方法の打ち切りありにおける必要例数の真値 (パターン 3, $\beta = 0.2$)

θ	1.33	1.5	2.0	3.0
Lakatos	929.7	462.5	160.5	66.4
Freedman	823.7	413.0	148.7	66.0
Rubinstein	843.4	555.3	311.0	200.4

8.4 パターン 4(ジグザグ型)

パターン 4 の状況での 3 つの方法の必要例数の真値を比較したものを表 16, 表 17 に示す。

表 16: 3つの方法の打ち切りありにおける必要例数の真値 (パターン 4, $\beta = 0.1$)

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

表 17: 3つの方法の打ち切りありにおける必要例数の真値 (パターン 4, $\beta = 0.2$)

θ	1.33	1.5	2.0	3.0
Lakatos	933.5	464.5	161.1	66.5
Freedman	823.7	413.0	148.7	66.0
Rubinstein	843.4	555.3	311.0	200.4

9 検証

9.1 ログランク検定による3つの方法の検出力を求めるプログラム

決められたハザード比により実験群と対照群それぞれの生存率 P_1 , P_2 と脱落率 w_1 と w_2 を引数としていれ, `size` には, 表 8~17 に示されているシミュレーションで得られた必要例数を 2 で割った値をそれぞれ Lakatos, Freedman, Rubinstein の方法の順番に入れ, 3 つの方法の検出力を同時に求められるプログラムとした。`size` にいれる値は整数とするので, 2 で割ったあとに 1 の位を切り上げる。本論文は 10000 回で繰り返している。

9.2 打ち切りなしの場合の検出力

表 8, 表 9 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。打ち切りなしの場合の検出力は, $\beta = 0.1$ の場合, 表 18 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。 $\beta = 0.2$ の場合のときも, 表 19 より, $\theta = 1.33$ のときは, どの例数設計法も検出力 80% に近いが, $\beta = 0.1$ と比べると, Rubinstein の方が Lakatos よりも近い検出力 80% を取っている。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。Lakatos と Freedman の方法は, 想定した検出力を保とうとしているが, Rubinstein の方法は検出力が 90% 以上となり, やはり保てていない。また, Lakatos の方法は少し過剰に評価されている。打ち切りなしの場合, 今回の実験では Freedman の方法が検出力 90%, 80% に近い値をとっているため, Freedman の方法が優れていると思われる。

表 18: 打ち切りなしの場合の検出力 ($\beta = 0.1$)

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

表 19: 打ち切りなしの場合の検出力 ($\beta = 0.2$)

θ	1.33	1.5	2.0	3.0
Lakatos	0.8639	0.8574	0.8524	0.8716
Freedman	0.8105	0.8097	0.8168	0.861
Rubinstein	0.8327	0.9152	0.9893	0.9998

9.3 打ち切りありの場合の検出力

表 10~17 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。打ち切りありの場合の検出力は、 $\beta = 0.1, 0.2$ の場合どちらも表 20, 表 21, 表 24, 表 25, 表 26, 表 27 より、打ち切りなしと同様、 $\theta = 1.33$ のときは、どの例数設計法も想定した検出力に近いことが分かる。 $\beta = 0.2$ では、Rubinstein の方が Lakatos よりも近い検出力 80% を取っている。Lakatos と Freedman の方法は、想定した検出力を保とうとしているが、Rubinstein の方法は検出力が 90% 以上となり、やはり保てていない。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%, 80% に近い値をとっているが、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ので注意が必要である。脱落率が均一型、漸増型、ジグザグ型の 3 パターンであると、Freedman の方法が優れていると思われる。

表 20: パターン 1(均一型) のときの検出力 ($\beta = 0.1$)

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 21: パターン 1(均一型) のときの検出力 ($\beta = 0.2$)

θ	1.33	1.5	2.0	3.0
Lakatos	0.8507	0.8531	0.8547	0.8595
Freedman	0.8061	0.8118	0.8237	0.8595
Rubinstein	0.8194	0.9102	0.9896	0.9994

しかし、表 22, 表 23 だけは、先に述べた結果とは異なった。 $\theta = 1.33$ のとき、想定した検出力に近い値をどの例数設計法もとっていたが、Freedman, Rubinstein の方法はかなり離れていると分かる。その点、Lakatos の方法が検出力 90%, 80% 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

表 22: パターン 2(偏り型) のときの検出力 ($\beta = 0.1$)

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 23: パターン 2(偏り型) のときの検出力 ($\beta = 0.2$)

θ	1.33	1.5	2.0	3.0
Lakatos	0.8447	0.8428	0.8443	0.857
Freedman	0.6372	0.6921	0.7612	0.8249
Rubinstein	0.6428	0.8071	0.9715	0.9991

表 24: パターン 3(漸増型) のときの検出力 ($\beta = 0.1$)

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 25: パターン 3(漸増型) のときの検出力 ($\beta = 0.2$)

θ	1.33	1.5	2.0	3.0
Lakatos	0.8532	0.8536	0.8561	0.8685
Freedman	0.8115	0.8138	0.8301	0.8685
Rubinstein	0.8191	0.905	0.9887	0.9995

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が保てていない。表 12, 表 13 とそれ以外の脱落率別の必要例数を比較してみると、Freedman と Rubinstein の方法は、あまり脱落率に関して必要例数の差がないことが分かる。その点、Lakatos の方法はどの値も高くなっている上にパターン 2 ではさらに増やしていることが分かる。そのため、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% と 80% を得られた。

今回の実験では、打ち切りがある場合、Freedman の方法は想定した検出力に近い値をとっているが、 θ と脱落率に変化しても Lakatos の方法のみ想定した検出力より大きくなっている。脱落率に応じて計算する際には、Lakatos の方法は区間ごとに計算するため、適切な例数を求めることができると考えられる。

10 考察

まず、必要例数について述べる。本論文は、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ありの場合も $\theta = 1.33, 1.5$ のとき、どの例数設計法に関しても外

表 26: パターン 4(ジグザグ型) のときの検出力 ($\beta = 0.1$)

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

表 27: パターン 4(ジグザグ型) のときの検出力 ($\beta = 0.2$)

θ	1.33	1.5	2.0	3.0
Lakatos	0.8521	0.858	0.8531	0.8636
Freedman	0.8048	0.8114	0.8227	0.8636
Rubinstein	0.8132	0.9093	0.9879	0.9995

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。想定している状況が正しくても各群 200 例という予備試験結果をもってしてもまれに大きな例数が必要になることが分かった。3 割ぐらゐは真値と比べて倍以上の例数が必要となり、何倍にもなることも珍しくはない。通常の予備試験ではもっと例数が少ない場合が多いと思われるのでこれは注意を要することである。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、それと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。本論文では、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% と 80% を得ている。本論文は、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。また、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う。本論文での設定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らなかったと考えられるので、この方法が決して使えない方法であるということではないことに注意する。

11 おわりに

Lakatos の方法の必要例数を求めるためのプログラムを作成に多く時間を割いてしまったため、本論文では、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、Lakatos の方法は予備試験の結果に左右

されると考えられるので他の区分や例数での予備試験でどのような変化があるか調べてみたい。

参考文献

- [1] Armitage, P., Berry, G. (椿美智子・椿広計 訳) (2001), 『医学研究のための統計的方法』, サイエнтиスト社.
- [2] Freedman, L. S. (1982): Tables of the number of patients required in clinical trials using the log-rank test, *Statistics in Medicine*, **1**, 121-129.
- [3] Kleinbaum, D. G. and Klein, M. (神田英一郎・藤井朋子 訳) (2015), 『エモリー大学クラインバウム教授の生存時間解析』, サイエнтиスト社.
- [4] Lakatos, E. (1988): Sample sizes based on the log-lank statistic in complex clinical trials, *Biometrics*, **44**, 229-241.
- [5] Lakatos, E., Lan, G. K. K. (1992): A comparison of sample size methods for the logrank statistic, *Statistics in Medicine*, **11**, 179-191.
- [6] 水澤純基・浜田知久馬 (2008): 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.
- [7] Rubinstein, L. V., Gail, M. H., and Santner, T. J. (1981): Planning the duration of a comparative clinical trial with loss to follow-up and a period of continued observaton, *Journal of Chronic Diseases*, **34**, 469-479.
- [8] 高橋友弥 (2010): 『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.
- [9] 丹後俊郎・上坂浩之 編 (2006): 『臨床試験ハンドブック -デザインと統計解析-』, 朝倉書店.