

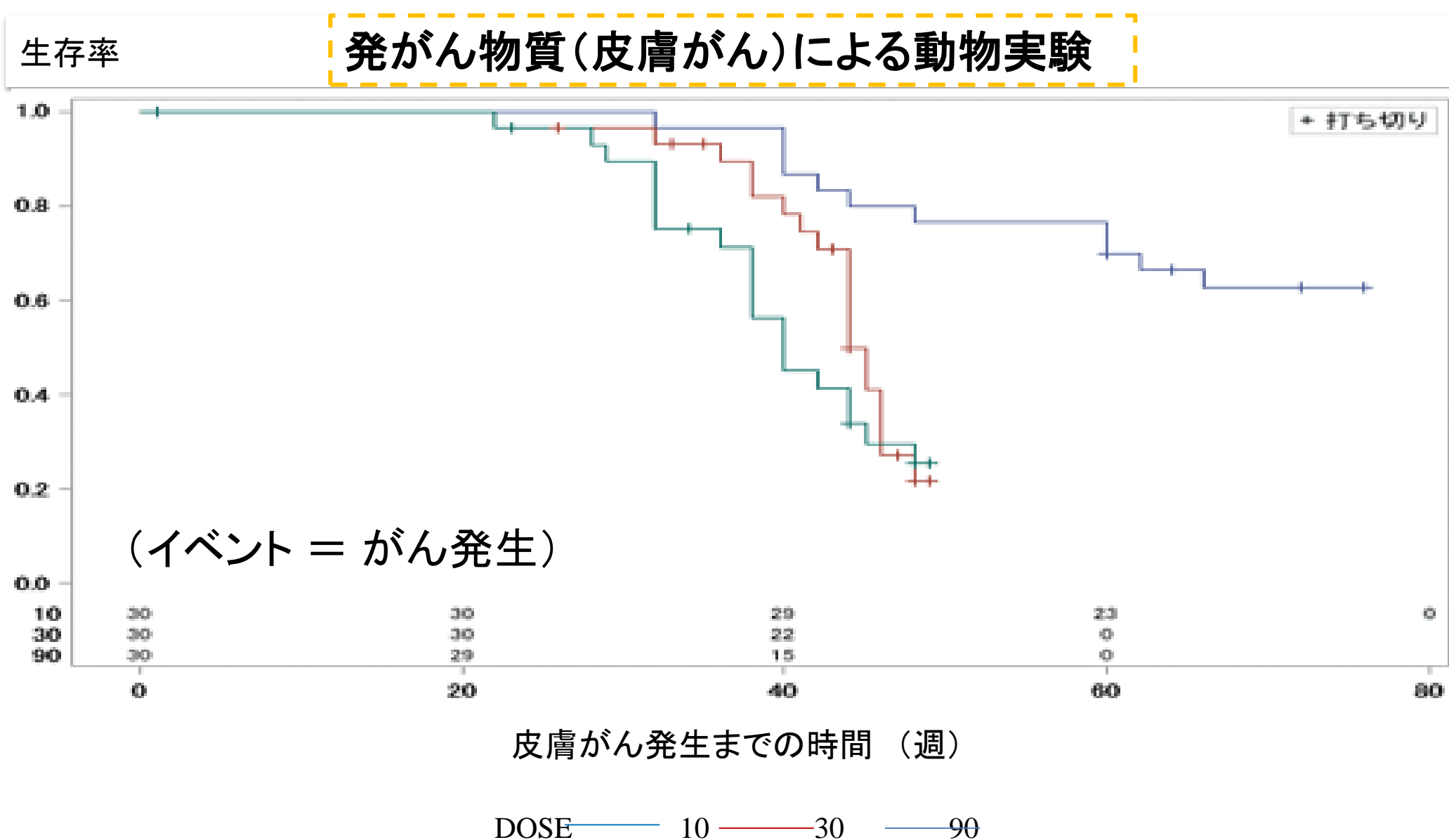
生存時間解析におけるノンパラメトリック検定の多重比較に関する研究

島村 文也 ・ 浜田 知久馬 ・ 佐野 雅隆
東京理科大学大学院 工学研究科 経営工学専攻

背景・目的

生存時間解析における多群比較^[3]

- 医薬研究では、イベント発生までの時間(生存時間)を評価
- ノンパラメトリック検定を用いて、各群の生存関数を比較



ノンパラメトリック検定の多群比較における問題点

- 検定の多重性^[1]
 - 検定を複数行くと、 α エラーが増大する
 - α エラーを有意水準以下に抑えるために、厳しめに検定を行う
 - 多重比較法が複数存在
 - 生存関数の多群比較における多重比較法の性能評価は十分でない
- 生存時間に対する順位付けの方法の決定
 - Joint-ranking 法**
 - すべての群の生存時間に対して順位付けを行う
 - Separate-ranking 法**
 - 比較する2つの群の生存時間に対してのみ順位付けを行う

e.g., DOSE 10とDOSE 30の各群の生存関数を比較

| DOSE | 生存時間(週) | | | | | |
|------|---------|----|----|----|----|--|
| 10 | 27 | 31 | 32 | 33 | 34 | |
| 30 | 21 | 23 | 26 | 29 | 30 | |
| 90 | 20 | 22 | 24 | 25 | 28 | |

DOSE 90の群の生存時間を考慮

| Joint-ranking 法 | | Separate-ranking 法 | |
|-----------------|---------------|--------------------|------------|
| DOSE | 順位 | DOSE | 順位 |
| 10 | 8 12 13 14 15 | 10 | 4 7 8 9 10 |
| 30 | 2 4 7 10 11 | 30 | 1 2 3 5 6 |
| 90 | 1 3 5 6 9 | 90 | |

生存時間解析において、より性能の高い多群比較を行うための適切な方法を検討

- Joint-ranking 法と Separate-ranking 法の性能評価
- α エラーを有意水準付近以下に抑えつつ、検出力の高い多重比較法の検討

方法

本研究で検討する多重比較法^[1]

Bonferroni 法

$$\alpha' = \frac{\alpha}{h}$$

Sidak 法

$$\alpha' = 1 - (1 - \alpha)^{\frac{1}{h}}$$

Holm 法

$-p$ 値を小さい順に並べてから設定した有意水準 α を調整

$$p_1 \leq p_2 \leq \dots \leq p_h \rightarrow \alpha'_1 = \frac{\alpha}{h}, \alpha'_2 = \frac{\alpha}{h-1}, \dots, \alpha'_h = \frac{\alpha}{1}$$

Scheffe 法

$-$ 自由度を $r - 1$ とした χ^2 分布を仮定することで調整

$$\text{non-adjusted} : p = \Pr(\chi^2_1 > \chi^2_{ij})$$

$$\text{Scheffe} : p = \Pr(\chi^2_{r-1} > \chi^2_{ij})$$

Tukey 法

$-\chi^2$ 分布の代わりに、「student 化された範囲の q 分布」を用いて調整

μ : 母数
 α : 設定する有意水準
 h : 試験全体での仮説数
 α' : 各比較に用いる有意水準
 r : 群の数
 i, j : 群

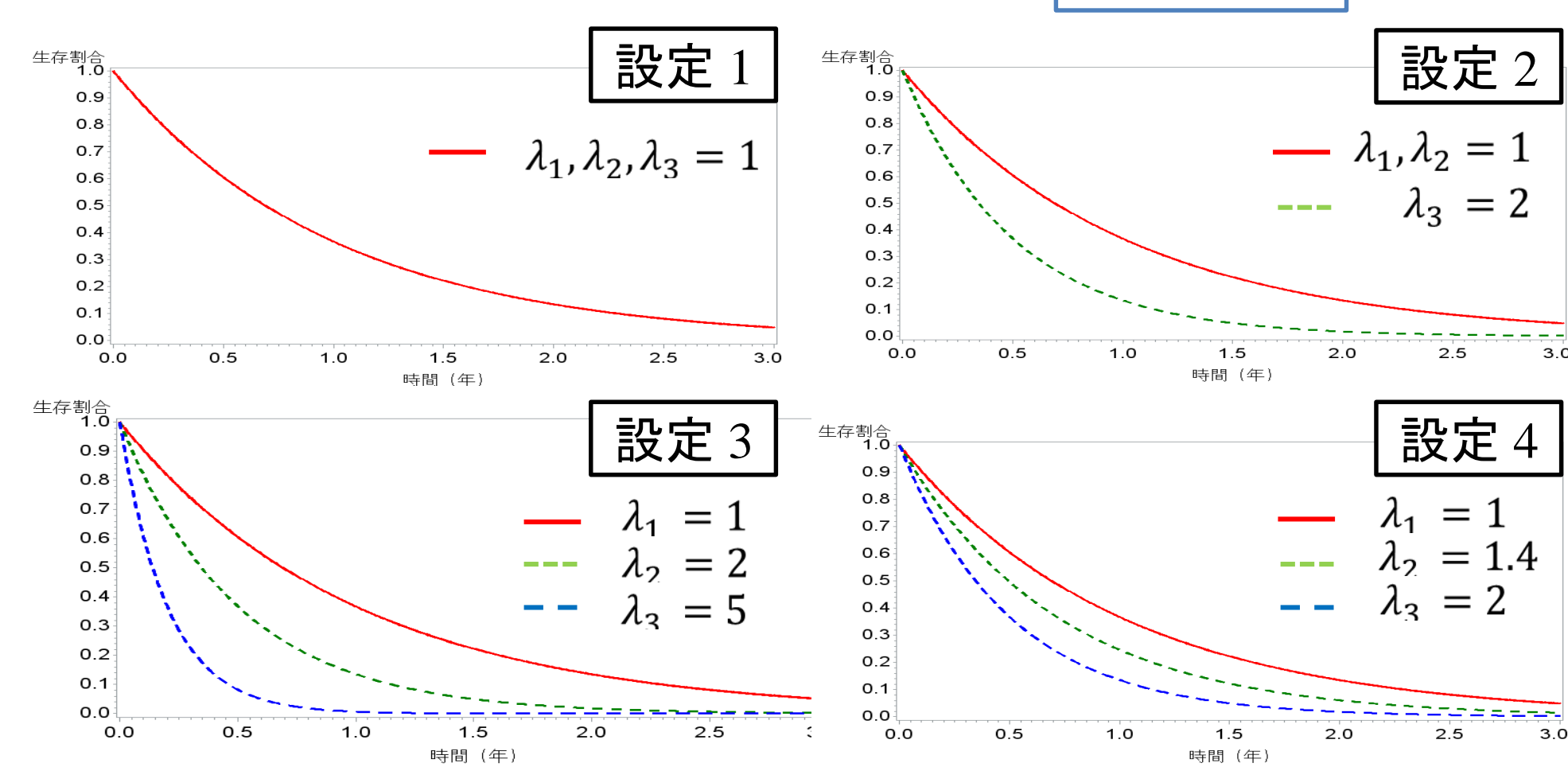
シミュレーション条件

- 生存時間分布: 指数分布
- 各群の症例数: 50 例 (3 群比較)
- 検定方法: ログランク検定
- 検討した方法 (計 10 通り)
 - Joint-ranking 法
 - Separate-ranking 法
 - Bonferroni 法
 - Sidak 法
 - Holm 法
 - Scheffe 法
 - Tukey 法
- シミュレーション回数: 10,000 回
- 有意水準: 両側 5%
- 評価指標: α エラー, 総対検出力, 最大差検出力, 最小差検出力

生存関数におけるハザードの設定^[2]

$$S_i(t) = \exp(-\lambda_i t) \quad [i = 1, 2, 3]$$

i : 群
 λ : ハザード



結果・考察

| 多重比較法 | α エラー (両側 5%) | | | |
|------------|----------------------|----------|--------|----------|
| | 設定 1 | | 設定 2 | |
| | Joint | Separate | Joint | Separate |
| Bonferroni | 0.0465 | 0.0470 | 0.0179 | 0.0171 |
| Sidak | 0.0471 | 0.0479 | 0.0182 | 0.0178 |
| Holm | 0.0465 | 0.0470 | 0.0356 | 0.0388 |
| Scheffe | 0.0394 | 0.0409 | 0.0162 | 0.0147 |
| Tukey | 0.0529 | 0.0537 | 0.0214 | 0.0200 |
| 調整なし | 0.1239 | 0.1267 | 0.0536 | 0.0504 |

設定 1 に関して、Tukey 法が多少有意水準を上回る
 $-\alpha$ エラーを有意水準以下に抑えられないログランク検定の性質が原因^[4]

| 多重比較法 | 総対検出力 | | | | | |
|------------|--------|----------|--------|----------|--------|----------|
| | 設定 2 | | 設定 3 | | 設定 4 | |
| | Joint | Separate | Joint | Separate | Joint | Separate |
| Bonferroni | 0.6521 | 0.7381 | 0.6585 | 0.8172 | 0.0040 | 0.0196 |
| Sidak | 0.6546 | 0.7402 | 0.6614 | 0.8184 | 0.0040 | 0.0199 |
| Holm | 0.7213 | 0.7919 | 0.8622 | 0.9190 | 0.0342 | 0.0814 |
| Scheffe | 0.6755 | 0.7168 | 0.6258 | 0.7993 | 0.0028 | 0.0162 |
| Tukey | 0.6273 | 0.7563 | 0.6906 | 0.8325 | 0.0048 | 0.0235 |
| 調整なし | 0.8247 | 0.8682 | 0.8625 | 0.9190 | 0.0437 | 0.0911 |

Separate-ranking 法と Holm 法を用いた場合が最も高い

| 多重比較法 | 最大差検出力 | | | | | |
|------------|--------|----------|--------|----------|--------|----------|
| | 設定 2 | | 設定 3 | | 設定 4 | |
| | Joint | Separate | Joint | Separate | Joint | Separate |
| Bonferroni | 0.7953 | 0.8398 | 1.0000 | 1.0000 | 0.8437 | 0.8398 |
| Sidak | 0.7977 | 0.8417 | 1.0000 | 1.0000 | 0.8452 | 0.8417 |
| Holm | 0.8294 | 0.8660 | 1.0000 | 1.0000 | 0.8498 | 0.8466 |
| Scheffe | 0.7773 | 0.8247 | 1.0000 | 1.0000 | 0.8310 | 0.8247 |
| Tukey | 0.8123 | 0.8539 | 1.0000 | 1.0000 | 0.8542 | 0.8539 |
| 調整なし | 0.9205 | 0.9270 | 1.0000 | 1.0000 | 0.9292 | 0.9270 |

設定 4 に関しては、Tukey 法が最も検出力が良いが他の手法と大きな差はない

| 多重比較法 | 最小差検出力 | | | | | |
|------------|--------|----------|--------|----------|--------|----------|
| | 設定 2 | | 設定 3 | | 設定 4 | |
| | Joint | Separate | Joint | Separate | Joint | Separate |
| Bonferroni | 0.7953 | 0.8398 | 0.8068 | 0.8382 | 0.2452 | 0.2305 |
| Sidak | 0.7977 | 0.8417 | 0.8076 | 0.8393 | 0.2471 | 0.2325 |
| Holm | 0.8294 | 0.8660 | 0.9098 | 0.9250 | 0.3021 | 0.3034 |
| Scheffe | 0.7773 | 0.8247 | 0.7923 | 0.8232 | 0.2299 | 0.2147 |
| Tukey | 0.8123 | 0.8539 | 0.8210 | 0.8505 | 0.2607 | 0.2467 |
| 調整なし | 0.9205 | 0.9270 | 0.9101 | 0.9256 | 0.3965 | 0.3844 |

設定 4 に関しては、Joint-ranking 法が全体的に 1~2% 程上回ったが、Holm 法においては差はない

まとめ

生存時間に対する順位付けの方法に関して

- α エラーは各手法で名義水準以下に保たれている
- 全体的な検出力は Separate-ranking 法の方が高い
- 設定 4 の最大差・最小差検出力に関してのみ Joint-ranking 法が高い

多重比較法に関して

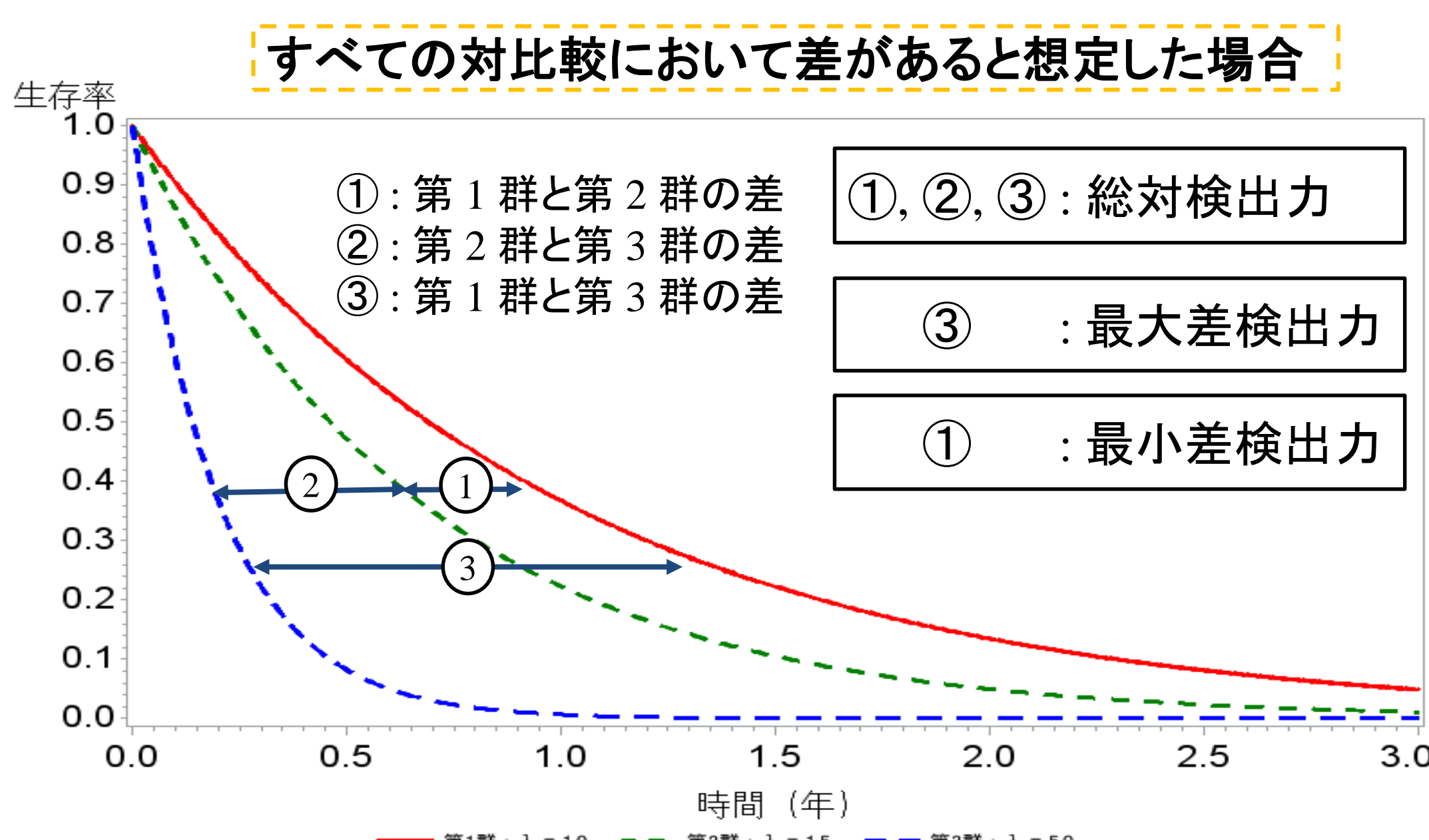
- Scheffe 法は保守的(他の手法も問題なし)
- 検出力
- Holm > Tukey > Bonferroni \cong Sidak > Scheffe

推奨する手法

Separate-ranking 法
Holm 法

参考文献

- 永田靖, 吉田道弘. 統計的多重比較法の基礎. サイエティスト社. 2001.
- 永田靖, 松田真一. 多重比較における新たな検出力の提案と各手法の特徴比較. 応用統計学. 19. 93-113. 1991.
- 大橋靖雄, 浜田知久馬. 生存時間解析 SAS による生物統計. 東京大学出版会. 1995.
- A. M. Kellner, D. Chmelevsky (1983). Small-Sample Properties of Censored-Data Rank Tests. Biometrics. 39. 675-682.



すべての対比較において差があると想定した場合

- ①: 第1群と第2群の差
- ②: 第2群と第3群の差
- ③: 第1群と第3群の差

①, ②, ③: 総対検出力

③: 最大差検出力

①: 最小差検出力

第1群: $\lambda = 1.0$ 第2群: $\lambda = 1.5$ 第3群: $\lambda = 5.0$