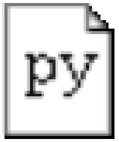
と WinLD

Rで統計解析入門

(16) 群逐次法

WinLD のセットアップ

- 以下のアドレスから圧縮ファイルをダウンロードする
  https://www.biostat.wisc.edu/content/lan-demets-method-statistical-programs-clinical-trials

- WinLD.zip を解凍し，

WinLD.exe をダブルクリックして起動する

中間解析と多重性の問題

- 2 群（例：治療薬と対照薬）の比較を行うことを考える
- 計画された患者さんが全て集まった後に解析を行うが（最終解析），ある程度の患者さんが集まった時点（例：全体の半分程集まった時点）でも解析（中間解析）を行いたいとする
- しかし，何も方策を立てずに 1 回の検定の際に有意水準 α = 5% のまま解析を行ってしまうと多重性の問題が生じる

検定回数	最大の Type I error rate
1	5.0 %
2	9.8 %
3	14.3 %
5	22.6 %
10	40.1 %

※ 最大の Type I error rate = $1 - (1-\alpha)^k$

中間解析と多重性の問題

- 中間解析と最終解析のそれぞれで検定を行う際，全体の Type I error を 5% にするのであれば，1 回の検定の際の有意水準 α は 5% では拙い
- しかし，例えばボンフェローニの方法で有意水準 α を調整すると非効率（中間解析の回数が増えれば，ハードルがどんどん高くなる）
- 今回は，中間解析（1 回とは限らない）と最終解析のそれぞれで検定を行う際の有意水準 α の調整方法等について解説を行う

検定回数	各回の検定で用いる有意水準 α
1	5.00 %
2	2.50 %
3	1.67 %
4	1.25 %
5	1.00 %

※ ボンフェローニの方法で調整

本日のメニュー

1. 情報分数 t ，z-score，B-value
2. Pocock の方法と O'brien-Fleming の方法
3. Lan & DeMets の α 消費関数
4. 中間解析後の推定
   - p 値の調整
   - 信頼区間と点推定値の調整

2 つの治療の比較

- X_i：被験者 i の対照薬の反応，Y_i：被験者 i の被験薬の反応（同じ被験者）
  $D_i = Y_i - X_i \sim N(\delta, \sigma^2)$ とする ⇒ σ^2 は簡単のため既知とし，
  $\delta = 0$ かどうかの検定を行うことを考える
- 被験者が N 例入った時点で試験終了，検定を行う
  ⇒ 検定統計量 Z_N は以下の通り

$$Z_N = \frac{1}{\sqrt{V_N}} \sum_{i=1}^{N} D_i = \frac{S_N}{\sqrt{V_N}}, \quad V_N = \mathrm{var}(S_N) = N\,\mathrm{var}(D_1)$$

- ここで，以下の式変形を行う（ n < N ，n は中間解析時の例数）

$$Z_N = \{S_n + S_N - S_n\} / \sqrt{V_N} = \underbrace{S_n / \sqrt{V_N}}_{1\sim n\text{ までの分}} + \underbrace{(S_N - S_n) / \sqrt{V_N}}_{n+1\sim N\text{ までの分}}$$

- Z_n の第一項目を **B-value** と呼ぶ

var：分散

2 つの治療の比較

$$t = v_n / v_N = \mathrm{var}(S_n) / \mathrm{var}(S_N)$$

- 上記の t を情報分数※と呼ぶ
  - 0 ～ 1 の値をとる（試験開始時：t ＝0，試験終了時：t ＝1）
  - ざっくりと言えば「中間解析を行ったタイミングを表す値」
  - 分散が既知なので，t = n/N = (中間解析時の例数)/(試験終了時の例数) となる
- 上記の t を使って中間での z-score：$S_n/v_n^{1/2}$ を Z(t) と表すと，B-value：B(t) は以下となる

$$B(t) = \frac{S_n}{\sqrt{v_N}} = \sqrt{t}\,Z(t),\ B(1) = Z(1) = \frac{S_N}{\sqrt{v_N}}$$

※ 英語は trial fraction，information fraction など

z-score と B-value の使い分け

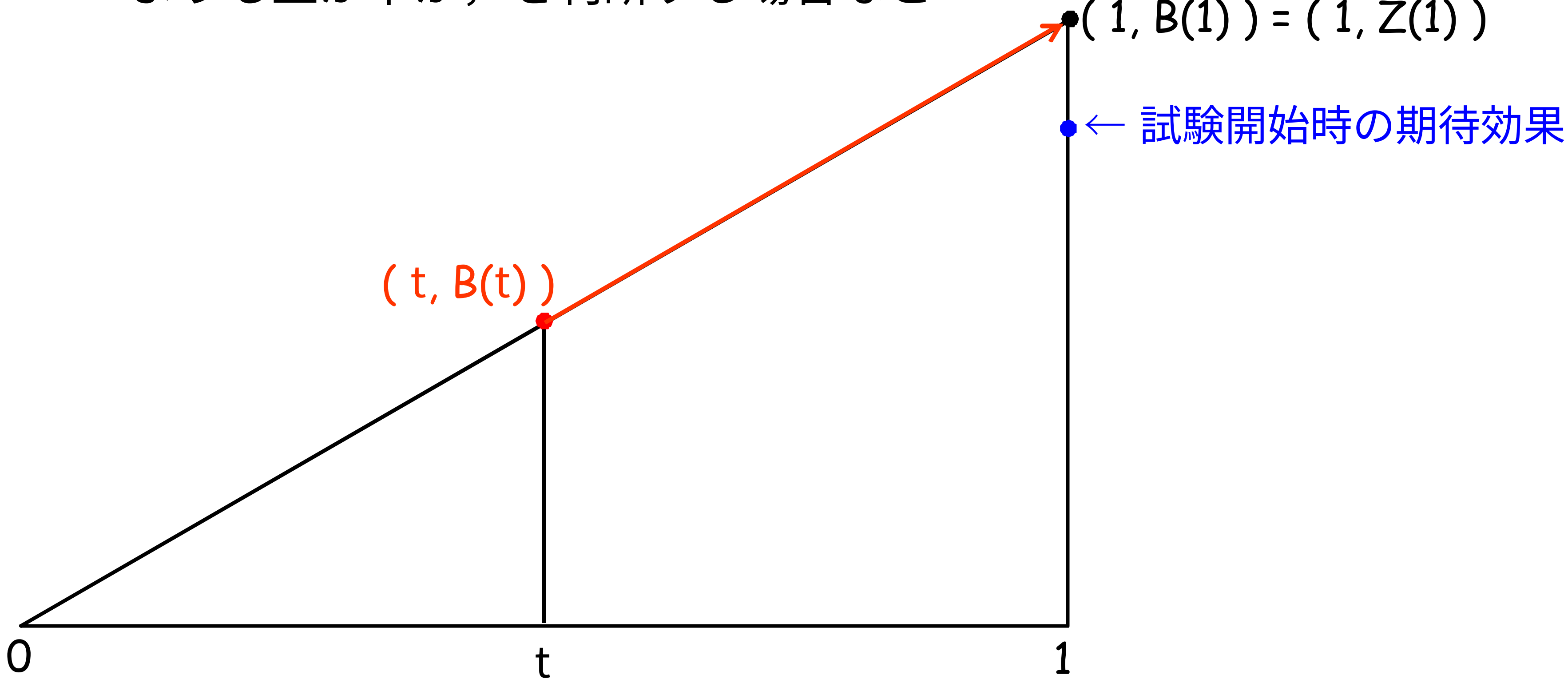
- z-score：棄却域を決める際に用いる（帰無仮説の下で計算する場合）
- B-value：観測された結果が期待される傾向に従っているかを判断する際に用いる（対立仮説の下で計算する場合）
- $Z(t_1), \ldots, Z(t_k)$ と $B(t_1), \ldots, B(t_k)$ はそれぞれ多変量正規分布に従う
- ここで，ドリフトパラメータ $\theta = E[Z(1)] = \delta / \{ var(\hat{\delta}) \}^{1/2}$
（試験終了時の治療群と対照群の標準化した効果の差）を定義すると...

	期待値	共分散	分散
z-score	$E[Z(t)] = \theta t^{1/2}$	$cov[Z(t_1), Z(t_2)] = (t_1/t_2)^{1/2}$	$V[Z(t)] = 1$
B-value	$E[B(t)] = \theta t$	$cov[B(t_1), B(t_2)] = t_1$	$V[B(t)] = t$

$B(t) = t^{1/2} Z(t)$, $t_1 \leqq t_2$

【参考】B-value を用いるメリット

- B-value の平均は θt と「 t の線形関数」になっているので，
  B-value でデータの傾向を表すと解釈が楽になる
  ⇒ 例えば，今のデータから「治療効果」が「期待される効果」
  よりも上か下か，を判断する場合など

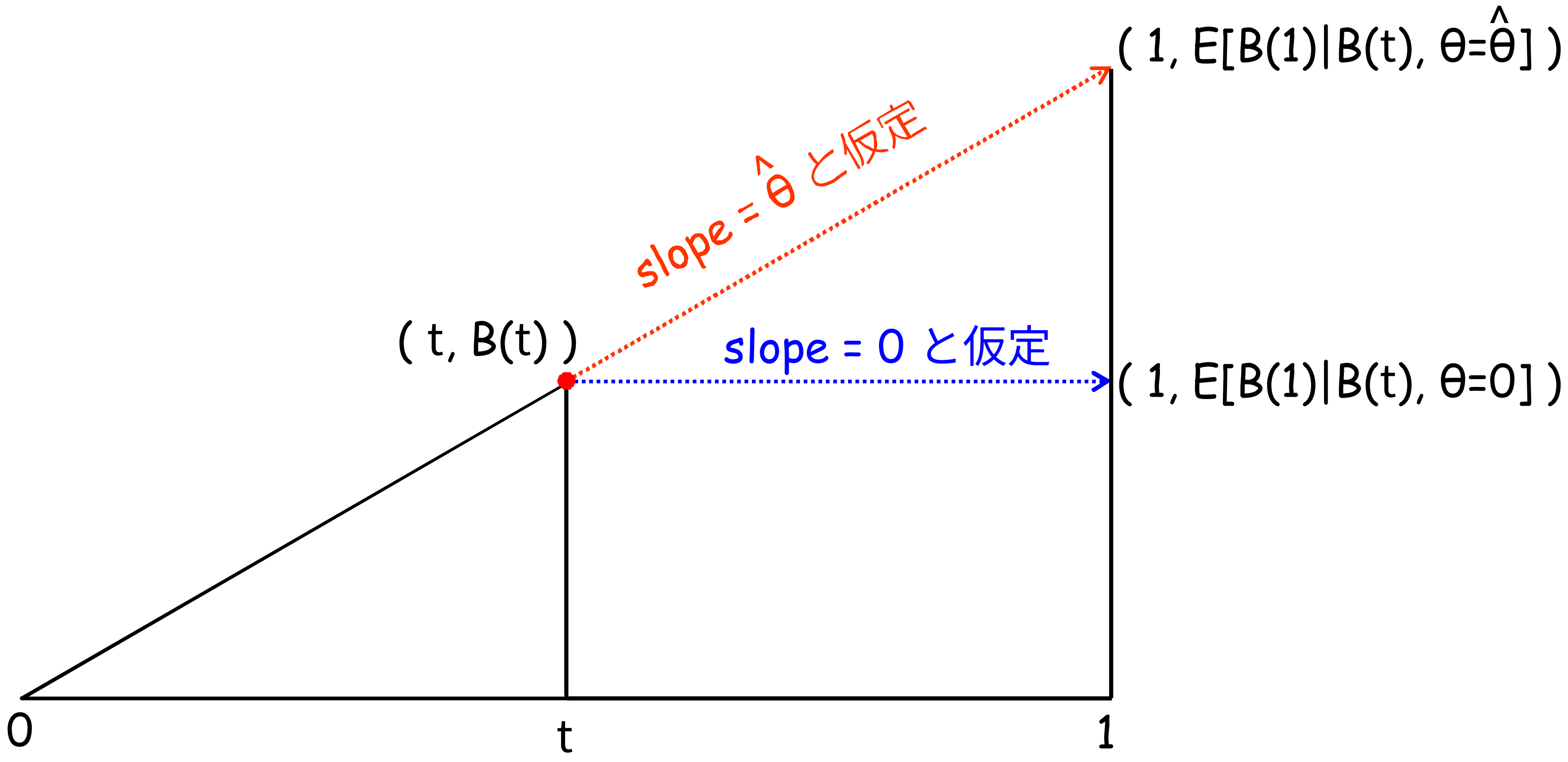


$B(t) \sim N(\theta t, t)$

【参考】B-value を用いるメリット

- $B(t) \sim N(\theta t, t)$，ドリフト $\theta = E[Z(1)] = \delta/\{ var(\hat{\delta}) \}^{1/2}$

  ⇒ t の時点における条件付き検出力を求めるときに B-value は重宝する



Lan et.al.（1988）より引用

例 1：棄却域の算出

- 200 例集積するデザインとし，途中（100 例集まった時点）で中間解析を 1 回，試験終了時に 1 回，合計 2 回の解析を行うと計画
- 100 例集まった時点で解析した結果，z-score が 1.7 であった
  - $t = 100/200 = 0.5$，z-score：$Z(0.5) = 1.7$
  - B-value：$B(0.5) = (0.5)^{1/2}\,Z(0.5) = 0.7 \times 1.7 = 1.2$
  - 中間解析では有意水準 $\alpha = 0.005$（片側），全体（ 2 回の解析）での Type I error を 0.025（片側）にする場合
    - z-score に関する棄却域 c_1（中間解析時），c_2（最終解析時）
      $\Pr\{ Z(0.5) > c_1 \} = 0.005$ かつ $\Pr\{ Z(0.5) > c_1 \text{ or } Z(1.0) > c_2 \} = 0.025$
      を満たす c_1，c_2 を計算すればよい
    - B-value に関する棄却域 a_1（中間解析時），a_2（最終解析時）
      $\Pr\{ B(0.5) > a_1 \} = 0.005$ かつ $\Pr\{ B(0.5) > a_1 \text{ or } B(1.0) > a_2 \} = 0.025$
      を満たす a_1，a_2 を計算すればよい

例 2：2 群の連続変数の比較

- 2 群の連続変数の比較を行う試験を考える
- 各群 200 例集積するデザインとし，途中（各群 100 例集まった時点）で中間解析を行うことを計画している
- 各群 100 例集まった時点で解析した結果，群間差は 0.3，その標準誤差は 0.4 であったとする（分散既知 → 統計量は正規分布に従うとする）
  - $t = 100/200 = 0.5$，$\hat{\delta}(0.5) = 0.3$，$se\{\hat{\delta}(0.5)\} = 0.4$
  - z-score：$Z(0.5) = 0.3/0.4 = 0.75$
  - B-value：$B(0.5) = (0.5)^{1/2} Z(0.5) = 0.70 \times 0.75 = 0.53$
  - 中間解析において有意水準 $\alpha = 0.005$（片側）で検定を行う場合は，$\Pr\{ Z(0.5) > c_1 \} = 0.005$ を満たす c_1 を計算すればよい ⇒ $c_1 = 2.576$ となる
  - $0.75 < 2.576$ なので，試験を続けた

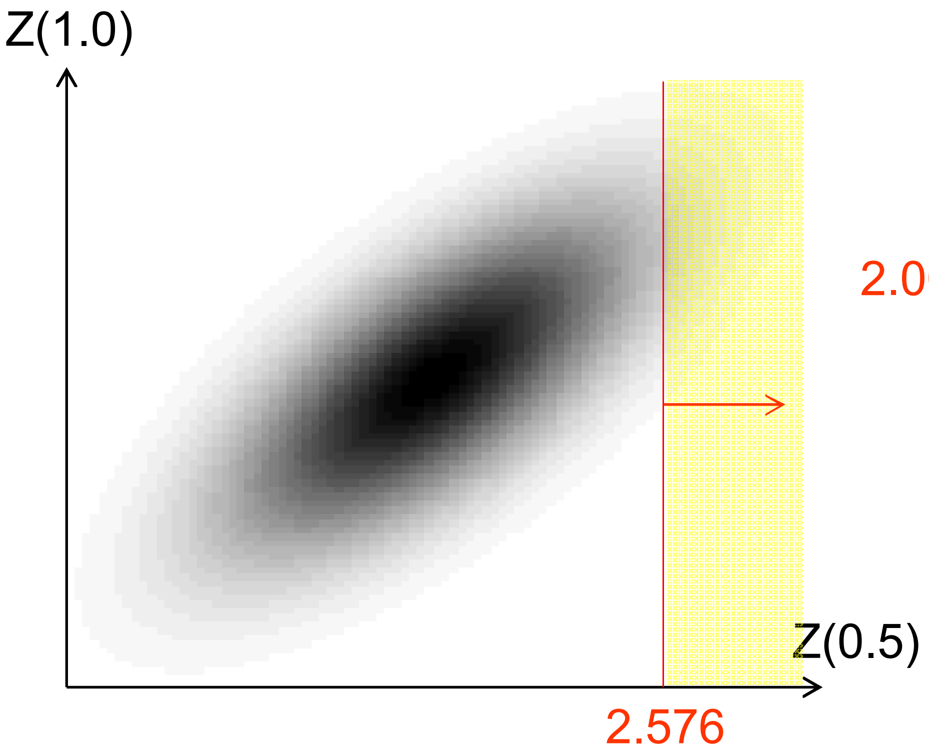
※ se：標準誤差，qnorm(0.005, lower=F) = 2.576

例 2：2 群の連続変数の比較

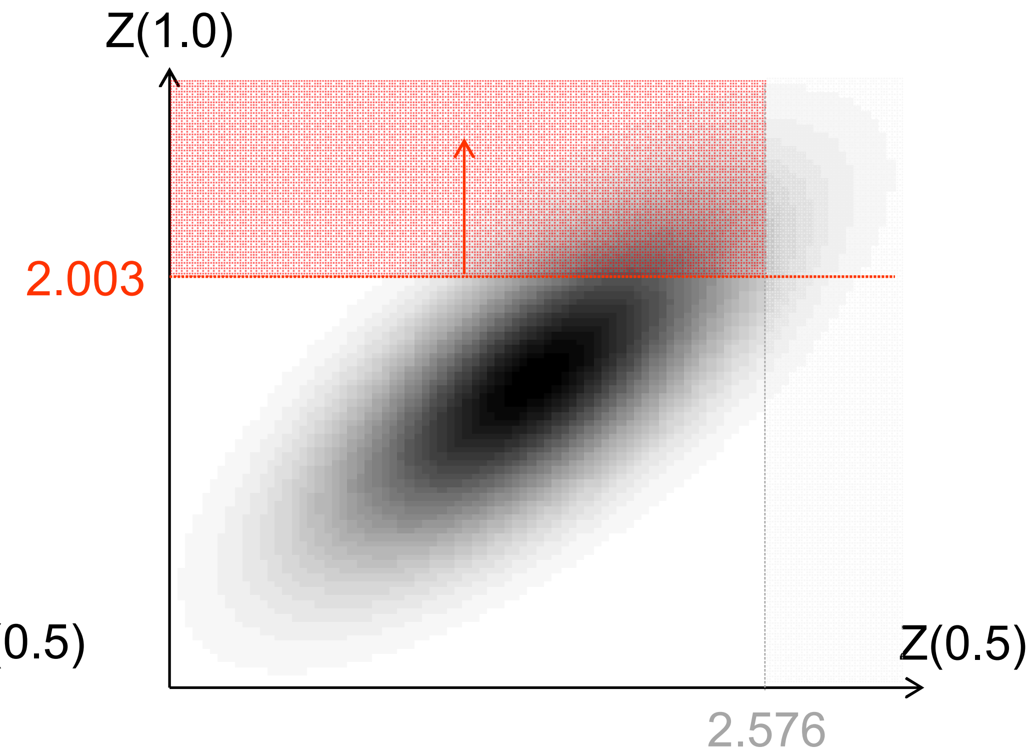
- 試験終了時点における群間差は 0.6，その標準誤差は 0.28 であった
  - $t = 1$，$\hat{\delta}(1.0) = 0.6$，$se\{\hat{\delta}(1.0)\} = 0.28$
  - z-score：$Z(1) = 0.6/0.28 = 2.14$，ドリフト：$\hat{\theta} = E[Z(1)] = 2.14$
  - B-value：$B(1) = (1)^{1/2}\, Z(1) = 2.14$
  - 最終解析において全体の Type I error が 0.025 になるよう検定を行う場合は $\Pr\{ Z(0.5) > 2.576 \text{ or } Z(1.0) > c_2 \} = 0.025$ を満たす c_2 を計算すればよい
    ⇒ $c_2 = 2.003$ となる（ $2.14 > 2.003$ なので有意となった）

※ $t_1 = \mathrm{var}[\hat{\delta}(0.5)]^{-1} / \mathrm{var}[\hat{\delta}(1.0)]^{-1} = \{ se[\hat{\delta}(0.5)] \}^{-2} / \{ se[\hat{\delta}(1.0)] \}^{-2} = 0.5$

例 2：2 群の連続変数の比較



【1 回目の棄却域（黄色部分）】



【2 回目の棄却域（桃色部分）】

- 2 回目の棄却域について，1 回目の棄却域が含まれていない点に注意
  ⇒ Pr{ Z(1.0) > 2.003 } ではなく Pr{ **Z(0.5) ≦ 2.576**, Z(1.0) > 2.003 } が棄却域

例２：棄却点の算出

```
> # t=0.5 のときの棄却点 c_1
> qnorm(1-0.005)
[1] 2.575829

> # t=1.0 のときの棄却点 c_2
> tcur    <- 1.0                              # 今から計算する t
> tprev   <- c(0.5)                           # 今までの t
> cprev   <- c(2.575829)                      # 今までの棄却点
> cumulal <- 0.025                            # 今消費できるα
> other   <- c(tcur, tprev, cprev, cumulal) # 解を求めるためのおまじない
> answer  <- findroot(rename, -7, 7, other) # 解を変数 answer に格納
> print(round(answer,digits=4))              # 表示
[1] 2.0028
```


"Statistical Monitoring of Clinical Trials" 中の関数 findroot() のプログラムを使用

前頁のプログラムでやっていること

- 1 回目（ t = 0.5 ）の棄却点の計算
  - $\Pr\{ Z(0.5) > c_1 \} = 0.005$ を満たす c_1 を導出（ 2 頁前の左図の面積）
  - $z_{1-0.005} = z_{0.995}$ を計算すればよい ⇒ $c_1 = 2.576$

- 2 回目（ t = 1.0 ）の棄却点の計算
  - $\Pr\{ Z(0.5) > 2.576 \text{ or } Z(1.0) > c_2 \} = 0.025$

    $\Leftrightarrow \Pr\{ Z(0.5) > 2.576 \} + \Pr\{ Z(0.5) \leqq 2.576,\ Z(1.0) > c_2 \} = 0.025$

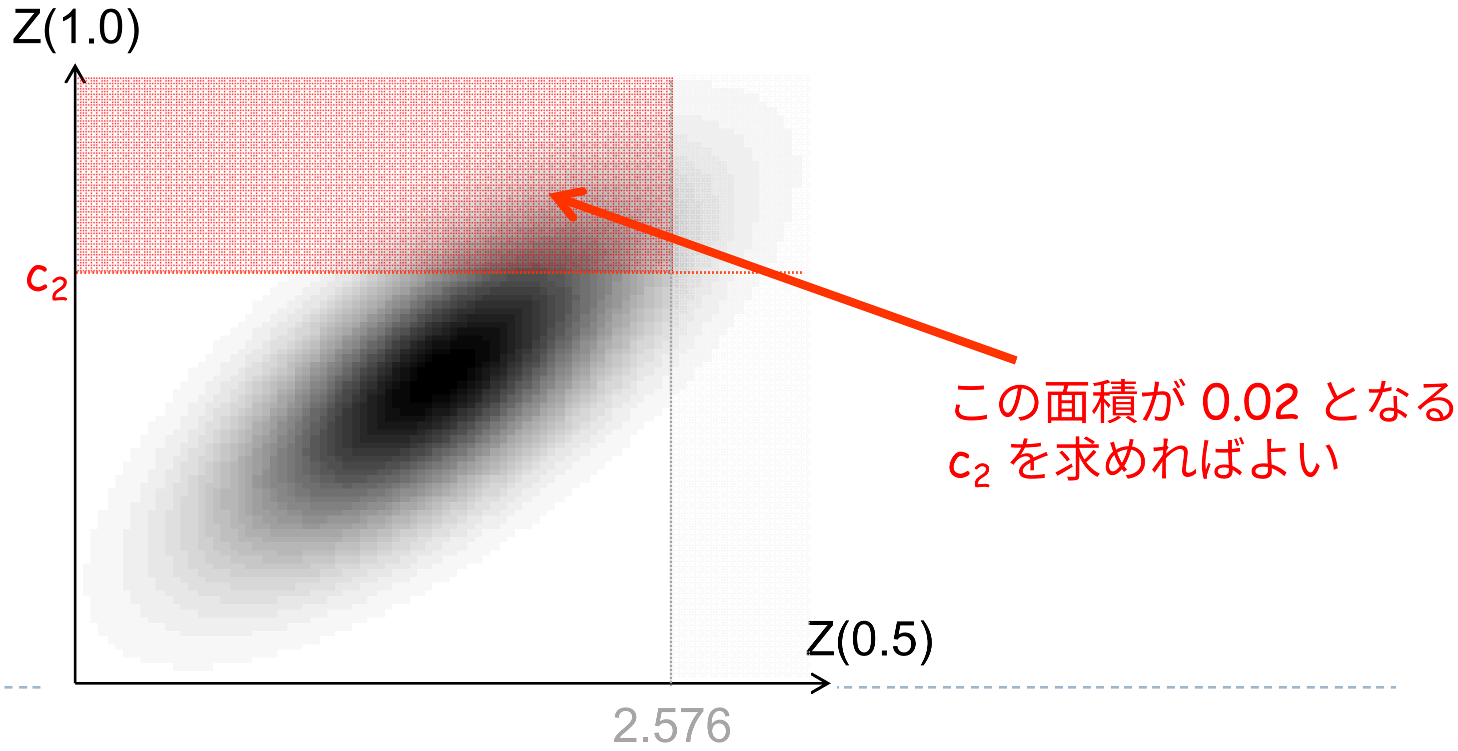
    0.005

    $\Leftrightarrow \Pr\{ Z(0.5) \leqq 2.576,\ Z(1.0) > c_2 \} = 0.02$ を満たす c_2 を導出

    （ 2 頁前の右図の面積）⇒ $c_2 = 2.003$
  - 1 回目の t と棄却点，2 回目の t と今消費できる累積 α（ = 0.025）を使って数値積分（二次元正規分布の面積を算出することで計算）する

前頁のプログラムでやっていること

- $\Pr\{ Z(0.5) \leqq 2.576,\ Z(1.0) > c_2 \} = 0.02$ を満たす c_2 を導出する方法は??
  - { Z(0.5)，Z(1.0) } は（帰無仮説の下では）平均 (0, 0)，分散 (1, 1)，共分散 $(0.5/1.0)^{1/2} = 0.7071$ なる二次元正規分布
  - 分布が決まっているので，数値積分等により c_2 を求めることが出来る



【参考】くり返し信頼区間

- くり返し信頼区間※：各回の検定に対応する多重性を調整した信頼区間
  - 両側信頼区間：$\hat{\delta}(t_i) \pm c_i \times se\{\hat{\delta}(t_i)\}$
  - 片側信頼区間：両側信頼区間の一方を -∞ か ∞ に置き換える

先ほどの「2 群の連続変数の比較の例」では・・・

- 100 例の時点：群間差 0.3，その標準誤差 0.4，棄却点 $c_1 = 2.576$

  ⇒ 両側信頼区間は $0.3 \pm 2.576 \times 0.4 = (-0.73, 1.33)$
- 最終の時点：群間差 0.6，その標準誤差 0.3，棄却点 $c_2 = 2.003$

  ⇒ 片側信頼区間は $0.6 + 2.003 \times 0.3 = (-\infty, 1.20)$

※ くり返し信頼区間は過度に広くなる傾向があるので注意

※ 後で出てくる Stagewise ordering による信頼区間とは別物

以上の概念が使える状況

- 2 群の平均値の差の比較（分散既知の場合）
- 2 群の割合の差の比較
- 2 群の生存割合の比較

※ 情報分数はざっくりと

t = n/N = (中間解析時の例数)/(試験終了時の例数) でよい

- なお，2 群の平均値の差の比較を 2 標本 t 検定で行う場合は，
  若干の変法を用いることで適用できる（後述）

例 3：2 群の生存割合の比較

- 2 群の生存割合の比較を行う試験を考える
- 2 群合計のイベントを 200 例集積するデザインとし，途中（50 例と 100 例集まった時点の 2 回）で中間解析，最終解析を 1 回行うことを計画している
- 合計 3 回の解析において全体の有意水準 α が 0.025（片側）となるように，各回の検定における棄却点 c_1，c_2，c_3 を求める

  ⇒ $\Pr\{ Z(\underline{0.25}) > c_1 \text{ or } Z(\underline{0.5}) > c_2 \text{ or } Z(\underline{1.0}) > c_3 \} = 0.025$ を満たす c_1，c_2，c_3 を求めればよい
- 実際に，50 例と 100 例の時点で検定を行ったとする

  ⇒ ログランク統計量を $Z(t)$，対応する B-value を $B(t)$ とすると
  - 50 例の時点：$t_1 = 0.25$，$Z(0.25) = 1.0$，$B(0.25) = 0.25^{1/2} \times 1.0 = 0.5$
  - 100 例の時点：$t_2 = 0.5$，$Z(0.5) = 1.7$，$B(0.2) = 0.5^{1/2} \times 1.7 = 1.2$
  - いずれの時点においても「試験継続」と判断した

例 3：2 群の生存割合の比較

- しかし，実際の試験終了時のイベント数は 180 例であったとする
- 「正しい」情報分数は t_1 = 50/180 = 0.28，t_2 = 0.56， t_3 = 1 となり，
  Pr{ Z(0.28) > c_1 or Z(0.56) > c_2 or Z(1.0) > c_3 } = 0.025 を満たす
  c_1，c_2 を算出しなおさなければいけないように思うが，その必要はない
  （最終イベント数を 200 例として計算した棄却点のままで問題ない）

※ 最終イベント数を 200 例として求めた情報分数の場合でも，
最終イベント数を 180 例として求めた情報分数の場合でも，
($Z(t_1)$, $Z(t_2)$, $Z(t_3)$) は帰無仮説の下で平均 0，分散 1，共分散 $(t_i/t_j)^{1/2}$ の 3 変量正規分布に従う

⇒ 例えば，(50/200)÷(100/200) = (50/180)÷(100/180) = 50/100 となり，
最終イベント数はキャンセルされるので共分散の値に影響しない

⇒ $Z(t_1)$ の分布や ($Z(t_1)$, $Z(t_2)$) の同時分布の形が変わらないので，ズレは問題にならない

例 3：2 群の生存割合の比較

- しかし，実際の試験終了時のイベント数は 180 例であったとする
- 「正しい」情報分数は t_1 = 50/180 = 0.28，t_2 = 0.56， t_3 = 1
- c_3 は算出しなおす必要あり（試験開始時に求めたものを使っては×）

  ⇒ Pr{ Z(0.28) > c_1 or Z(0.56) > c_2 or Z(1.0) > c_3 } = 0.025 を満たす

  c_3 を算出しなおさなければいけない（ c_1 や c_2 は元の値で計算）

※ ($Z(t_1)$, $Z(t_2)$, $Z(t_3)$) は帰無仮説の下で平均 0，分散 1，共分散 $(t_i/t_j)^{1/2}$ の 3 変量正規分布に従う

⇒ 例えば， 50/200 と計算していたものが， 50/180 と計算される

⇒ 例えば，100/200 と計算していたものが，100/180 と計算されるので，共分散の値に影響し，($Z(t_1)$, $Z(t_2)$, $Z(t_3)$) の同時分布の形が変わる・・・

例 3：2 群の生存割合の比較（計算例）

- イベントが 50 例，100 例，200例（最終解析）の時点で解析を行うデザイン
- 1 回目の有意水準 α を 0.005（片側）とし，
  $\Pr\{ Z(\underline{0.25}) > c_1 \} = 0.005$ を満たす c_1 を求める ⇒ $c_1 = 2.58$
- 2 回目までの有意水準 α を 0.01（片側）とし，
  $\Pr\{ Z(\underline{0.25}) > 2.58 \text{ or } Z(\underline{0.5}) > c_2 \} = 0.01$ を満たす c_2 を求める ⇒ $c_2 = 2.49$
- 合計 3 回の解析における全体の有意水準 α を 0.025（片側）とし，
  $\Pr\{ Z(\underline{0.25}) > 2.58 \text{ or } Z(\underline{0.5}) > 2.49 \text{ or } Z(\underline{1.0}) > c_3 \} = 0.025$ を満たす c_3 を求める ⇒ $c_3 = \underline{2.09}$
- しかし，実際は最終解析時の例数は 180 例
  $\Pr\{ Z(\underline{0.28}) > 2.58 \text{ or } Z(\underline{0.56}) > 2.49 \text{ or } Z(\underline{1.0}) > c_3 \} = 0.025$ を満たす c_3 を求める ⇒ $c_3 = \underline{2.08}$

例３：２群の生存割合の比較（計算例）

```
> # 例 3：t=0.25 のときの棄却点 c_1
> qnorm(1-0.005)
[1] 2.575829
> # 例 3：t=0.5 のときの棄却点 c_2
> tcur    <- 0.5                               # 今から計算する t
> tprev   <- c(0.25)                           # 今までの t
> cprev   <- c(2.58)                           # 今までの棄却点
> cumulal <- 0.01                              # 今消費できる累積α
> other   <- c(tcur, tprev, cprev, cumulal) # 解を求めるためのおまじない
> answer  <- findroot(rename, -7, 7, other) # 解を変数 answer に格納
> print(round(answer,digits=4))              # 表示
[1] 2.489
> # 例 3：t=1 のときの棄却点 c_3
> tcur    <- 1                                 # 今から計算する t
> tprev   <- c(0.25, 0.5)                      # 今までの t
> cprev   <- c(2.58, 2.49)                     # 今までの棄却点
> cumulal <- 0.025                             # 今消費できる累積α
> other   <- c(tcur, tprev, cprev, cumulal) # 解を求めるためのおまじない
> answer  <- findroot(rename, -7, 7, other) # 解を変数 answer に格納
> print(round(answer,digits=4))              # 表示
[1] 2.0895
```

例３：２群の生存割合の比較（計算し直しの例）

```
> # 例 3：t=0.28 のときの棄却点 c_1
> qnorm(1-0.005)
[1] 2.575829
> # 例 3：t=0.56 のときの棄却点 c_2
> tcur    <- 0.56                          # 今から計算する t
> tprev   <- c(0.28)                       # 今までの t
> cprev   <- c(2.58)                       # 今までの棄却点
> cumulal <- 0.01                          # 今消費できる累積α
> other   <- c(tcur, tprev, cprev, cumulal) # 解を求めるためのおまじない
> answer  <- findroot(rename, -7, 7, other) # 解を変数 answer に格納
> print(round(answer,digits=4))            # 表示
[1] 2.489
> # 例 3：t=1 のときの棄却点 c_3
> tcur    <- 1                             # 今から計算する t
> tprev   <- c(0.28, 0.56)                  # 今までの t
> cprev   <- c(2.58, 2.49)                 # 今までの棄却点
> cumulal <- 0.025                         # 今消費できる累積α
> other   <- c(tcur, tprev, cprev, cumulal) # 解を求めるためのおまじない
> answer  <- findroot(rename, -7, 7, other) # 解を変数 answer に格納
> print(round(answer,digits=4))            # 表示
[1] 2.0787
```

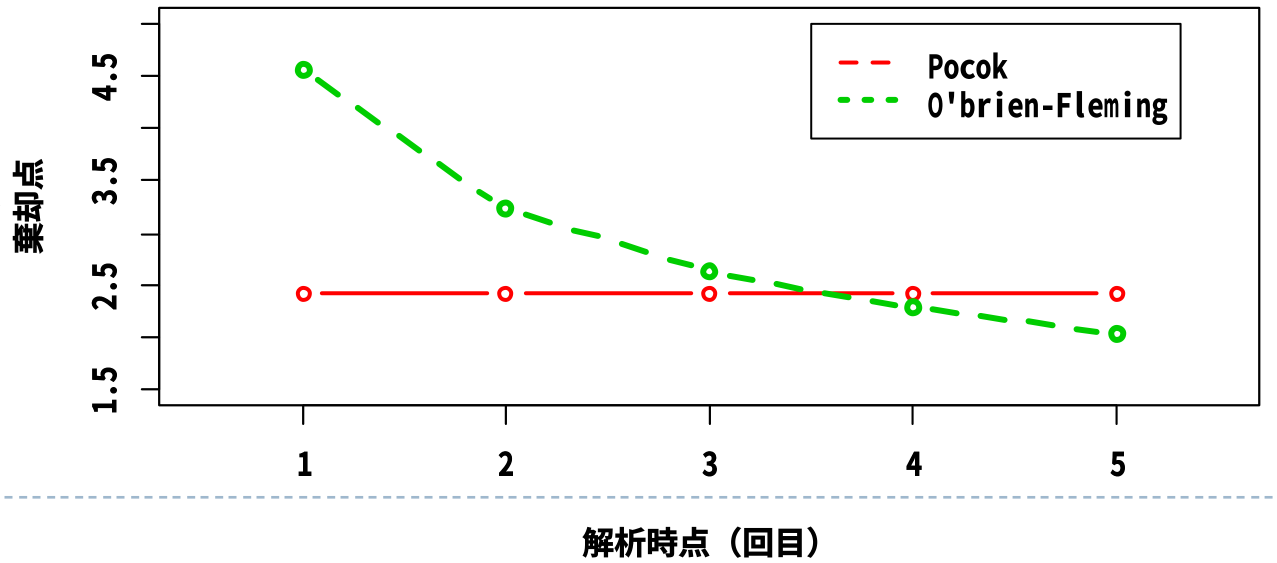
本日のメニュー

1. 情報分数 t，z-score，B-value
2. Pocock の方法と O'brien-Fleming の方法
3. Lan & DeMets の α 消費関数
4. 中間解析後の推定
   - p 値の調整
   - 信頼区間と点推定値の調整

Pocock の方法と O'brien-Fleming の方法

- 帰無仮説の下での $(Z(t_1), \cdot\cdot\cdot, Z(t_k))$ の分布は，平均 0，分散 1，共分散 $(t_i/t_j)^{1/2}$ の多変量正規分布 ⇒ 情報分数の比のみに依存
- 解析の回数 k と全体の Type I error（例：片側検定，0.025）を指定すれば（全体の例数を N，増分は等しく N/k 例ずつ増えると仮定し）各解析時の棄却点を求めることが出来る方法がいくつかある
  - Pocock の方法：z-score に関する棄却点が全て等しくなるように計算
  - O'brien-Fleming の方法：B-value に関する棄却点が全て等しくなるように計算

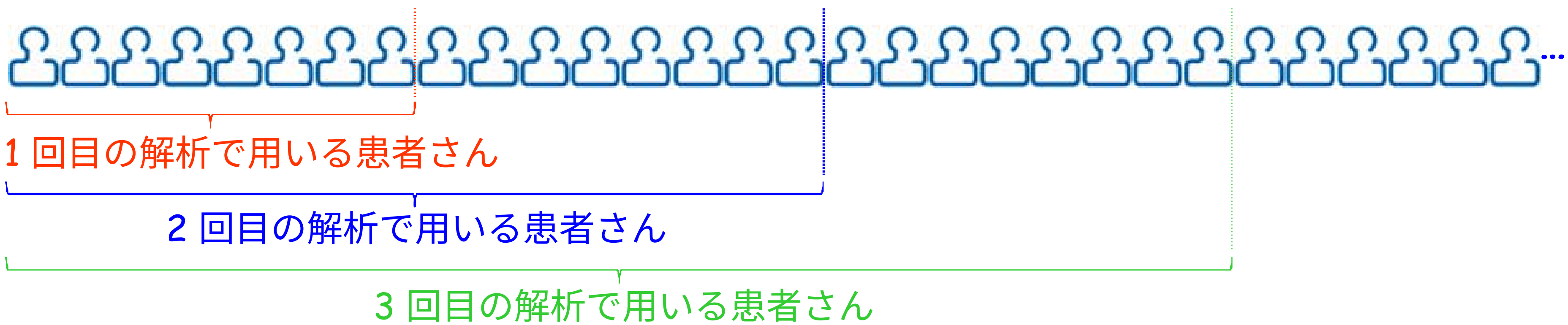
z-score→



前頁のグラフを描くプログラム

```
> x  <- 1:5
> y1 <- rep(2.41, 5)
> y2 <- c(4.56, 3.23, 2.63, 2.28, 2.04)
> plot(x, y1, xlim=c(0.5,5.5), ylim=c(1.5,5), type="b",
+      col=2, lty=1, lwd=2, ann=F)
> par(new=T)
> plot(x, y2, xlim=c(0.5,5.5), ylim=c(1.5,5), type="b",
+      col=3, lty=2, lwd=3, xlab="解析時点（回目）", ylab="棄却点")
> legend(3.5, 5, c("Pocok","O'brien-Fleming"), col=2:3, lty=2:3, lwd=2:3, ncol=1)
```

【参考】患者さんが等しく増えるとは ⇒ 等しく 8 例ずつ増えると仮定すると・・・



Pocock の方法

$$\Pr\{\bigcup_{i=1}^{k} Z(i/k) > c(k)\} = \alpha$$

- α = 0.025（片側）の場合，
  - k = 1：c(k) = 1.96
  - k = 2：c(k) = 2.18 → 1 回目も 2 回目も棄却点は 2.18
  - k = 3：c(k) = 2.29 → 1 回目 ～ 3 回目の棄却点は 2.29
  - k = 4：c(k) = 2.36 → 1 回目 ～ 4 回目の棄却点は 2.39
  - k = 5：c(k) = 2.41 → 1 回目 ～ 5 回目の棄却点は 2.41

Pocock の方法（ k=3 のとき）

$Z(i/k) = Z_i,\ U_i \sim N(0,1),\ \phi(x) = \frac{1}{\sqrt{2\pi}}\exp\{-\frac{x^2}{2}\}$ とすると，

$$\alpha = Pr\left\{\bigcup_{i=1}^{3} Z_i > c\right\} = 1 - Pr\{Z_1 \le c,\ Z_2 \le c,\ Z_3 \le c\}$$

$$= 1 - Pr\left\{U_1 \le c,\ \frac{1}{\sqrt{2}}(U_1 + U_2) \le c,\ \frac{1}{\sqrt{3}}(U_1 + U_2 + U_3) \le c\right\}$$

$$= 1 - \iiint_{\substack{u_1 \le c \\ u_1 + u_2 \le \sqrt{2}c \\ u_1 + u_2 + u_3 \le \sqrt{3}c}} \underbrace{\phi(u_1)\phi(u_2)\phi(u_3)}_{f(u_1, u_2, u_3)}\, du_1 du_2 du_3$$

ここで $V_1 = U_1,\ V_2 = U_1 + U_2,\ V_3 = U_1 + U_2 + U_3$ とおくと

$$\left\{\begin{array}{rcl} U_1 & = & V_1 \\ U_2 & = & V_2 - V_1 \\ U_3 & = & V_3 - V_2 \end{array}\right., \quad \text{ヤコビ行列}\ J = \begin{vmatrix} 1 & 0 & 0 \\ -1 & 1 & 0 \\ 0 & -1 & 1 \end{vmatrix} = 1$$

$$\therefore\ \alpha = 1 - \int_{-\infty}^{\sqrt{3}c}\int_{-\infty}^{\sqrt{2}c}\int_{-\infty}^{c} f(v_1, v_2 - v_1, v_3 - v_2)\, dv_1 dv_2 dv_3$$

Pocock の方法（前頁の積分の数値計算）

```
> install.packages("cubature", dep=T)
> library(cubature)
> g <- function(x) exp(-x[1]^2/2)*exp(-(x[2]-x[1])^2/2)*
                   exp(-(x[3]-x[2])^2/2)/(2*pi)^(3/2)
> x <- 2.289*c(sqrt(1), sqrt(2), sqrt(3))   # Pocock(c=2.289)
> adaptIntegrate(g, rep(-10,3), x)
$integral
[1] 0.97497
```


※ 実際に棄却点を求める際は，積分結果が 0.975 になるまで c の値をいろいろ変えて計算することになる

O'brien-Flemingの方法

$$\Pr\{\bigcup_{i=1}^{k} B(i/k) > a(k)\} = \Pr\{\bigcup_{i=1}^{k} Z(i/k) > c(k)\} = \alpha$$

$$= \Pr\{\bigcup_{i=1}^{k} Z(i/k) > a(k)/t^{1/2}\} = \alpha$$

- $\alpha = 0.025$（片側），$k = 5$，$a(5) = 2.04$ である場合，$Z(t) = B(t)/t^{1/2}$ なる関係から・・・
  - $c(1) = 2.04/(1/5)^{1/2} = 4.56$
  - $c(2) = 2.04/(2/5)^{1/2} = 3.23$
  - $c(3) = 2.04/(3/5)^{1/2} = 2.63$
  - $c(4) = 2.04/(4/5)^{1/2} = 2.28$
  - $c(5) = 2.04/(5/5)^{1/2} = 2.04$

O'brien-Fleming の方法（ k=3，k=5 のとき）

$Z(i/k) = Z_i,\ U_i \sim N(0,1),\ \phi(x) = \frac{1}{\sqrt{2\pi}} \exp\{-\frac{x^2}{2}\}$ とすると，

$$\alpha = Pr\left\{\bigcup_{i=1}^{3} Z_i > \frac{c}{\sqrt{i/3}}\right\} = 1 - Pr\left\{Z_1 \le \sqrt{\frac{3}{1}}c,\ Z_2 \le \sqrt{\frac{3}{2}}c,\ Z_3 \le \sqrt{\frac{3}{3}}c\right\}$$

$$= 1 - Pr\left\{U_1 \le \sqrt{3}c,\ U_1 + U_2 \le \sqrt{3}c,\ U_1 + U_2 + U_3 \le \sqrt{3}c\right\}$$

$$= 1 - \iiint_{\substack{u_1 \le \sqrt{3}c \\ u_1 + u_2 \le \sqrt{3}c \\ u_1 + u_2 + u_3 \le \sqrt{3}c}} \underbrace{\phi(u_1)\phi(u_2)\phi(u_3)}_{f(u_1, u_2, u_3)}\, du_1 du_2 du_3$$

ここで $V_1 = U_1,\ V_2 = U_1 + U_2,\ V_3 = U_1 + U_2 + U_3$ とおくと，Pocock の方法のときと同様の計算により

$$\alpha = 1 - \int_{-\infty}^{\sqrt{3}c} \int_{-\infty}^{\sqrt{3}c} \int_{-\infty}^{\sqrt{3}c} f(v_1, v_2 - v_1, v_3 - v_2)\, dv_1 dv_2 dv_3$$

同様に，$k = 5$ のときは以下となる．

$$\alpha = 1 - \int_{-\infty}^{\sqrt{5}c} \cdots \int_{-\infty}^{\sqrt{5}c} f(v_1, v_2 - v_1, v_3 - v_2, v_4 - v_3, v_5 - v_4)\, dv_1 \cdots dv_5$$

O'brien-Fleming の方法（前頁の積分の数値計算）

```
> library(cubature)
> g <- function(x) exp(-x[1]^2/2)*exp(-(x[2]-x[1])^2/2)*
                   exp(-(x[3]-x[2])^2/2)/(2*pi)^(3/2)
> x <- 2.004*c(sqrt(3), sqrt(3), sqrt(3))   # O'Brien & Fleming
> adaptIntegrate(g, rep(-10,3), x)          # (c=2.004)
$integral
[1] 0.9749977
```

※ 実際に棄却点を求める際は，積分結果が 0.975 になるまで c の値をいろいろ変えて計算することになる

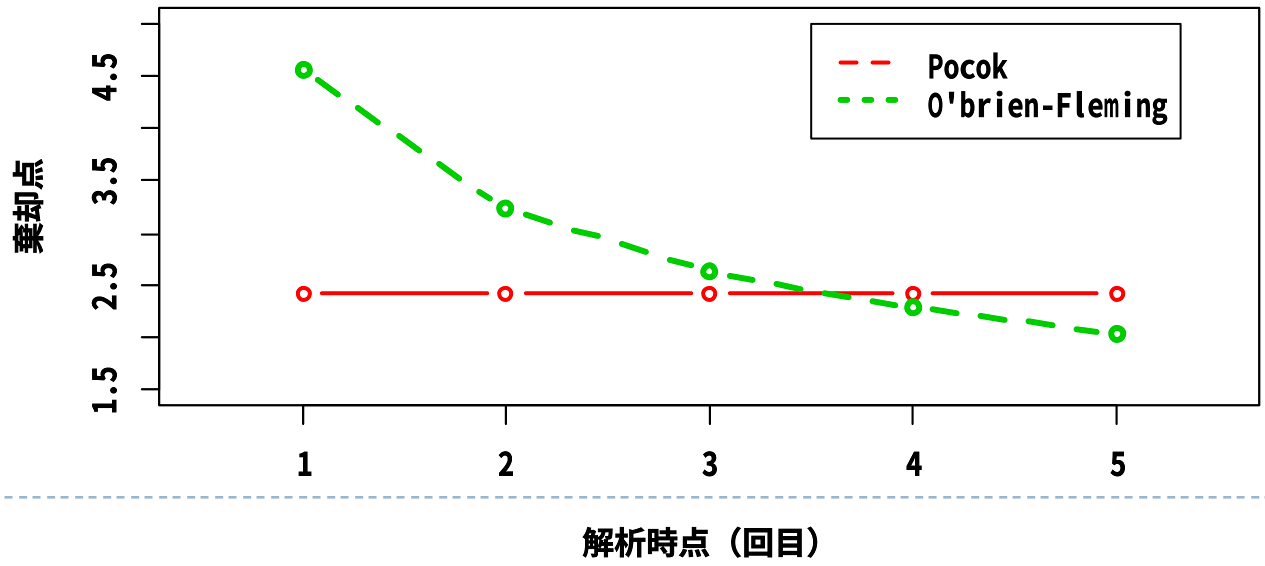
Pocock の方法と O'brien-Fleming の方法

```
> install.packages("gsDesign", dep=T)

> library(gsDesign)

> PO <- gsDesign(k=5, timing=1, sfu="Pocock", alpha=0.025, test.type=1)

> PO$upper$bound    # Pocock の方法（解析回数：5 回）

[1] 2.413180 2.413180 2.413180 2.413180 2.413180


> OF <- gsDesign(k=5, timing=1, sfu="OF", alpha=0.025, test.type=1)

> OF$upper$bound    # O'brien-Fleming の方法（解析回数：5 回）

[1] 4.561743 3.225639 2.633723 2.280871 2.040073
```

Pocock の方法と O'brien-Fleming の方法

- Pocock の方法：
  z-score に関する棄却点が全て等しくなるように計算
  ⇒ 何回目の検定でも同じ棄却限界値を用いる
- O'brien-Fleming の方法：
  B-value に関する棄却点が全て等しくなるように計算
  ⇒ データ数が少ない場合は基準を厳しく（棄却点を大きく）し，
  最終解析では基準を甘く（棄却点を小さく）する

z-score→



本日のメニュー

1. 情報分数 t，z-score，B-value
2. Pocock の方法と O'brien-Fleming の方法
3. **Lan & DeMets の α 消費関数**
4. 中間解析後の推定
   - p 値の調整
   - 信頼区間と点推定値の調整

Lan & DeMets の α 消費関数

- 前節の方法は「患者さんが均等に増える」と仮定した方法
  - 1 回目の解析：各群の例数は　n 例
  - 2 回目の解析：各群の例数は 2n 例
  - 3 回目の解析：各群の例数は 3n 例・・・となることを仮定

⇒ 実際にはきっちり均等に増えることは難しく，ズレることがしばしば・・・

- 患者さんが不均等に増えても各解析時の棄却点を求めるために，Lan & DeMets の α 消費関数 α(t) を導入する
  - α(0) = 0， α(1) = 0.025 を満たす増加関数（片側の場合を想定して 0.025 ）
  - α(t)：情報分数 t までに消費した有意水準を表す

$$\alpha(t_j) = \Pr\{\cup_{i=1}^{j} Z(t_i) > c_i\} \qquad (j = 1, \cdots, k)$$

$$\alpha(t_j) - \alpha(t_{j-1}) = \Pr\{\cap_{i=1}^{j-1} Z(t_i) \le c_i \cap Z(t_j) > c_j\}$$

Lan & DeMets の α 消費関数

- Pocock 型の α 消費関数（片側の場合）：

  $\alpha_{p1}(t) = 0.025 \log\{ 1 - (e-1)t \}$

- Pocock 型の α 消費関数（両側の場合）：

  $\alpha_{p2}(t) = 0.05 \log\{ 1 - (e-1)t \}$

- O'brien-Fleming 型の α 消費関数（片側の場合）：

  $\alpha_{of1}(t) = 2 \{ 1 - \Phi(z_{0.0125}/t^{1/2}) \} = 2 \{ 1 - \Phi(2.2414/t^{1/2}) \}$

- O'brien-Fleming 型の α 消費関数（両側の場合）：

  $\alpha_{of2}(t) = 4 \{ 1 - \Phi(z_{0.0125}/t^{1/2}) \} = 4 \{ 1 - \Phi(2.2414/t^{1/2}) \}$

※「Pocock 型」「O'brien-Fleming 型」と，「方法」ではなく「型」になっているのは，本来の方法を α 消費関数で近似しているため

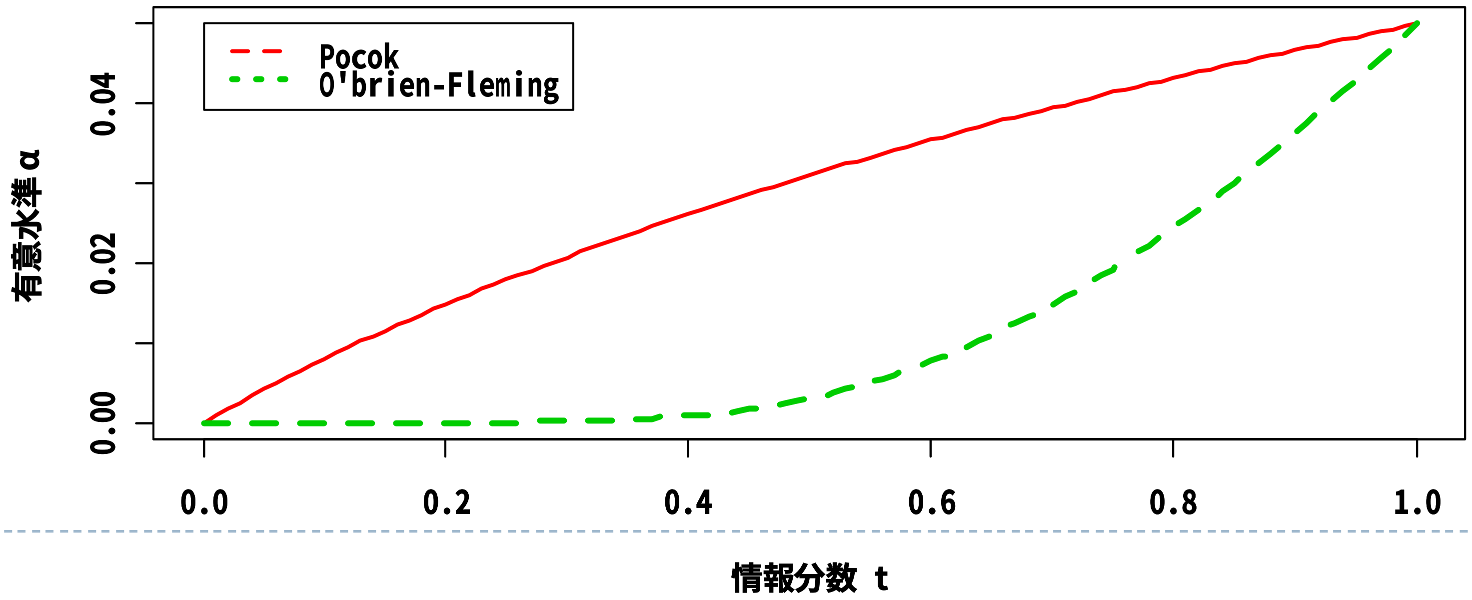
※ 片側：有意水準 2.5%，両側：有意水準 5% とした

Lan & DeMets の α 消費関数

```
> pocock <- function(t,alpha) alpha*log(1+(exp(1)-1)*t)
> pocock(0.5,0.025)  # t=0.5 のときまでに消費したα（片側）
[1] 0.01550286
> pocock(0.5,0.05)   # t=0.5 のときまでに消費したα（両側）
[1] 0.03100573

> of_onesided <- function(t) 2*(1-pnorm(qnorm(0.0125,lower=F)/sqrt(t)))
> of_twosided <- function(t) 4*(1-pnorm(qnorm(0.0125,lower=F)/sqrt(t)))
> of_onesided(0.5)   # t=0.5 のときまでに消費したα（片側）
[1] 0.001525323
> of_twosided(0.5)   # t=0.5 のときまでに消費したα（両側）
[1] 0.003050646

> curve(pocock(x,0.05), xlim=c(0,1), ylim=c(0,0.05), col=2, lty=1, lwd=2, ann=F)
> par(new=T)
> curve(of_twosided,    xlim=c(0,1), ylim=c(0,0.05), col=3, lty=2, lwd=3,
+       xlab="情報分数 t", ylab="有意水準α")
> legend(0, 0.05, c("Pocok","O'brien-Fleming"), col=2:3, lty=2:3, lwd=2:3, ncol=1)
```

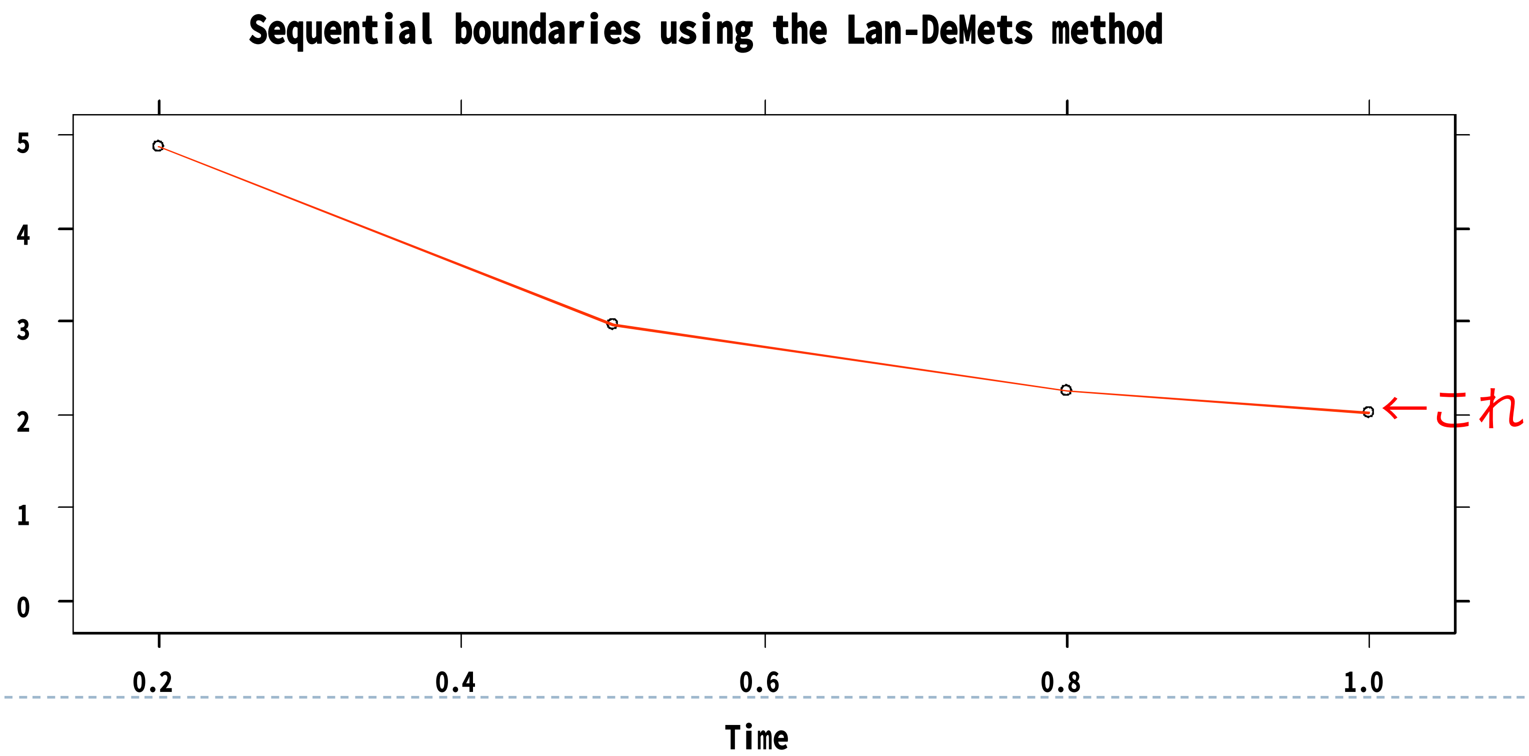
Lan & DeMets の α 消費関数

- Pocock 型：
  z-score に関する棄却点がほぼ等しくなるように計算
  ⇒ 何回目の検定でもほぼ同じ棄却限界値を用いる
- O'brien-Fleming 型：
  B-value に関する棄却点がほぼ等しくなるように計算
  ⇒ データ数が少ない場合は基準を厳しく（棄却点を大きく）し，
  最終解析では基準を甘く（棄却点を小さく）する



例 4：α 消費関数を使った棄却点の算出

- 解析時点：t = 0.2, 0.5, 0.8, 1.0
- 調整方法：O'brien-Fleming 型（片側，α = 0.025）
- 棄却点は c_1 = 4.8769，c_2 = 2.9626，c_3 = 2.266，c_4 = 2.0278



【参考】泥臭く数値計算しても良いが・・・

```
> # t=0.2
> of_onesided(0.2)                            # 今消費できるα
[1] 5.388713e-07
> qnorm(1-of_onesided(0.2))                   # 今使用する棄却点
[1] 4.876885

> # t=0.5
> tcur    <- 0.5                              # 今から計算する t
> tprev   <- c(0.2)                           # 今までの t
> cprev   <- c(4.876885)                      # 今までの棄却点
> cumulal <- of_onesided(tcur)                # 今消費できるα
> other   <- c(tcur, tprev, cprev, cumulal)   # 解を求めるためのおまじない
> answer  <- findroot(rename, -7, 7, other)   # 解を変数 answer に格納
> print(round(answer,digits=4))               # 今使用する棄却点
[1] 2.9628

> # t=0.8
> tcur    <- 0.8                              # 今から計算する t
> tprev   <- c(0.2, 0.5)                      # 今までの t
> cprev   <- c(4.876885, 2.9628)              # 今までの棄却点
> cumulal <- of_onesided(tcur)                # 今消費できるα
> other   <- c(tcur, tprev, cprev, cumulal)   # 解を求めるためのおまじない
> answer  <- findroot(rename, -7, 7, other)   # 解を変数 answer に格納
> print(round(answer,digits=4))               # 今使用する棄却点
[1] 2.2662    ・・・・・・(続く)
```

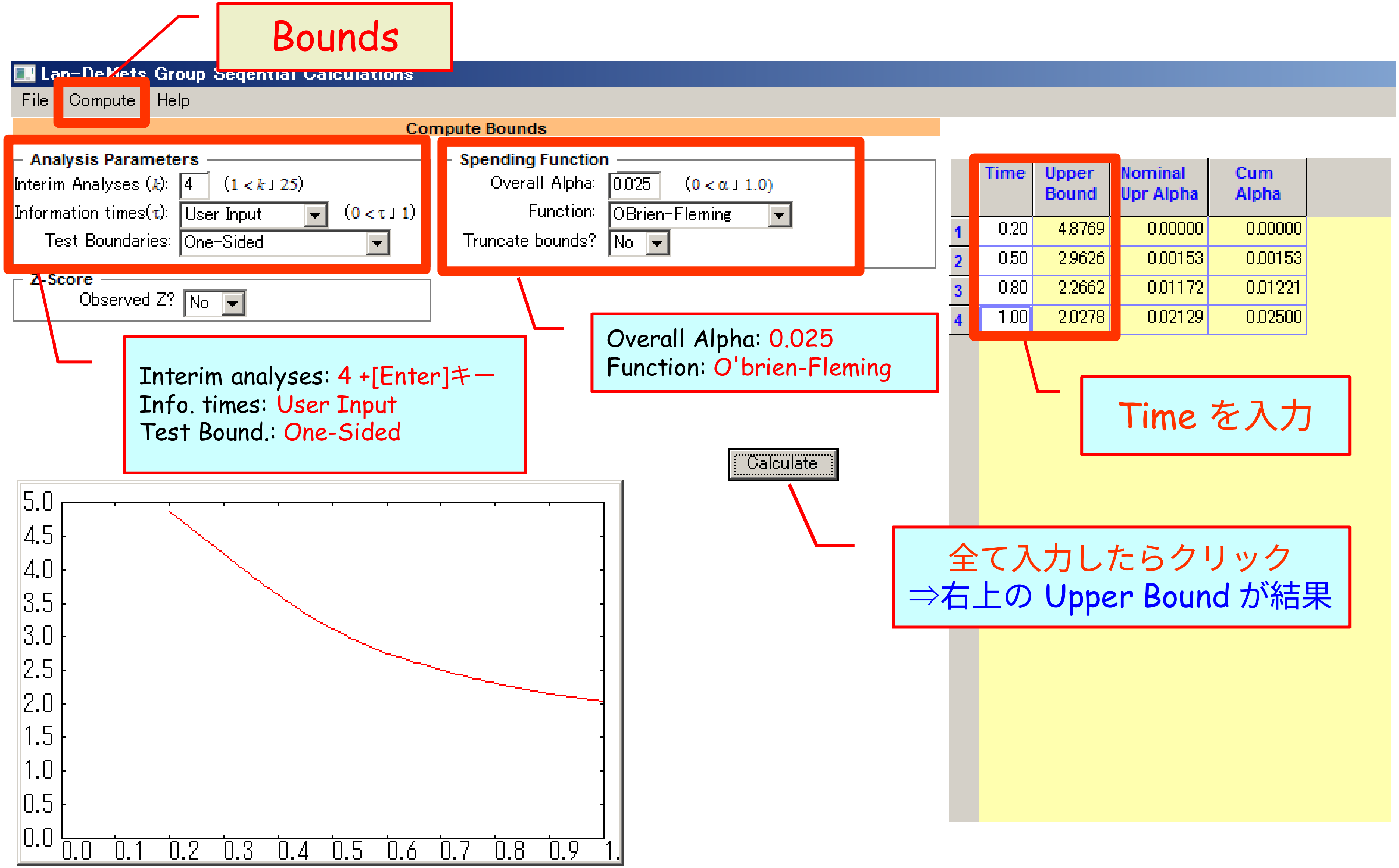
例４：α消費関数を使った棄却点の算出

```
> library(gsDesign)
> t <- c(0.2, 0.5, 0.8)
> OF <- gsDesign(k=4, timing=t, sfu=sfLDOF, alpha=0.025, test.type=1)
> OF$upper$bound
[1] 4.876885 2.962629 2.266195 2.027794

> PO <- gsDesign(k=4, timing=t, sfu=sfLDPocock, alpha=0.025, test.type=1) # 参考
> PO$upper$bound
[1] 2.437977 2.332825 2.324233 2.368653
```

3 頁前のグラフを描くプログラム

```
> install.packages("ldbounds", dep=T)
> library(ldbounds)
> t <- c(0.2, 0.5, 0.8, 1.0)
> # iuse=1:O'Brien Fleming, 2:Pocock, 3:power family, 4:Hwang-Shih-DeCani family
> obf.bd <- bounds(t, iuse=c(1,1), alpha=c(0.025,0.025))   # 両側（←参考までに）
> obf.bd <- bounds(t, iuse=1, alpha=0.025)                 # 片側
> summary(obf.bd)
n =  4
Overall alpha:  0.025
Type: One-Sided Bounds
alpha:  0.025
Spending function: O'Brien-Fleming
Boundaries:
   Time   Upper    Exit pr.   Diff. pr.
1   0.2  4.8769  5.3887e-07  5.3887e-07
2   0.5  2.9626  1.5253e-03  1.5248e-03
3   0.8  2.2662  1.2212e-02  1.0686e-02
4   1.0  2.0278  2.5000e-02  1.2788e-02
> plot(obf.bd)                                             # 前々頁のグラフ
```

WinLD を使った棄却点の算出例

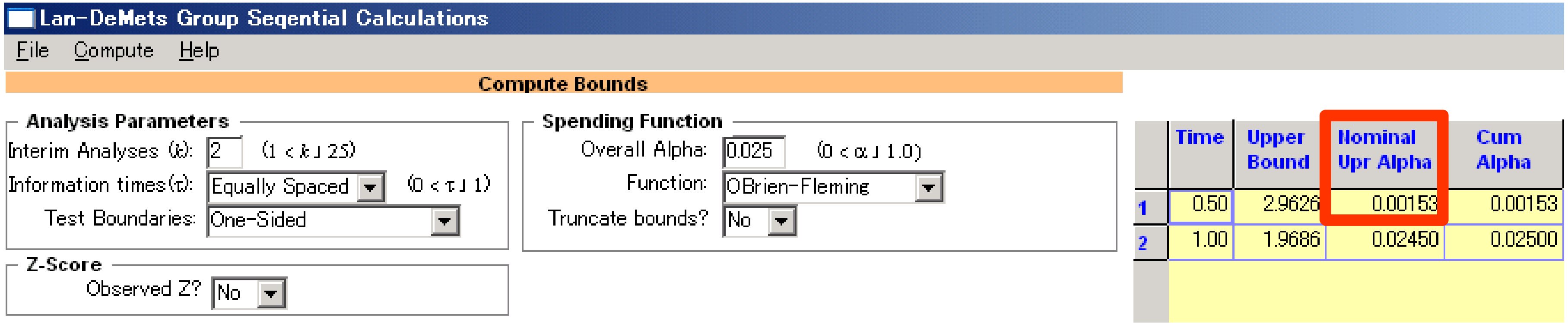


例5：2群の連続変数の比較（再考）

- 2群の連続変数の比較を行う試験を考える
- 各群200例集積するデザイン，途中（各群100例集まった時点）で中間解析を行い，有意水準 α は全体で0.025（片側）と計画している
- 検定は2標本t検定，有意水準 α はO'brien-Fleming型で調整する
- 各群100例集まった時点で解析した結果，t値は1.5であったとする
  - 情報分数 t = 100/200 = 0.5，検定統計量t値：1.5
  - 中間解析における棄却点は2.96，名目上の有意水準は α = 0.0015（片側）となるが，この棄却点（2.96）は正規分布（分散既知の場合の統計量の分布）を仮定したものなので使ってはいけない
  - 有意水準 α = 0.0015（片側）のみを用いて中間解析の検定を行えばよい
    ⇒ これを用いて計算したt統計量の棄却点※は3.0
    ⇒ 1.5 < 3.0なので有意ではない

※ qt(0.0015, df=200-2, lower=F)

例５：２群の連続変数の比較（再考）




```
> t <- 0.5
> gsDesign(k=2, timing=t, sfu=sfLDOF, alpha=0.025, test.type=1)
One-sided group sequential design with
90 % power and 2.5 % Type I Error.
           Sample
            Size
  Analysis Ratio*  Z   Nominal p  Spend
         1  0.502 2.96    0.0015 0.0015
         2  1.003 1.97    0.0245 0.0235
     Total                       0.0250
++ alpha spending: Lan-DeMets O'brien-Fleming approximation spending function
```

【参考】実際の例数が計画時の例数とズレた場合

- 2 群の生存割合の比較を行う試験を考え，2 群合計のイベントを 200 例集積するデザインとし，中間解析を 2 回，最終解析を 1 回行うことを計画する
- 全体の有意水準 α：0.025（片側），α 消費関数：$\alpha(t) = 0.025\ t^{1.5}$
- 1 回目と 2 回目の解析時（中間解析時）の例数は 50 例と 100 例であった
  ⇒ t = 0.25，0.5，$\alpha(t)$ = 0.003125, 0.008838 となり，
  $c_1 = 2.73$，$c_2 = 2.47$ と計算される
- 最終的な例数が 180 例となった場合（計画よりも減った場合），例 3 で紹介したとおり c_1 と c_2 は変更しなくてもよく，c_3 だけを計算し直せば良い
  ⇒ $c_3 = 2.064$ と計算される

【参考】計算例

```
> # t=0.25, 0.5 のときまでに消費出来るα
> f <- function(t) 0.025*t^1.5
> f(1/4); f(1/2)
[1] 0.003125 ; 0.008838835

> # t=0.25 のときの棄却点 c_1
> qnorm(1-0.003125)
[1] 2.734369

> # t=0.5 のときの棄却点 c_2
> tcur    <- 0.5                             # 今から計算する t
> tprev   <- c(0.25)                         # 今までの t
> cprev   <- c(2.734369)                     # 今までの棄却点
> cumulal <- 0.008838835                     # 今消費できる累積α
> other   <- c(tcur, tprev, cprev, cumulal)  # 解を求めるためのおまじない
> answer  <- findroot(rename, -7, 7, other)  # 解を変数 answer に格納
> print(round(answer,digits=4))              # 表示
[1] 2.471

> # t=1 のときの棄却点 c_3
> tcur    <- 1                               # 今から計算する t
> tprev   <- c(0.25, 0.5)                    # 今までの t
> cprev   <- c(2.734369, 2.471)              # 今までの棄却点
> cumulal <- 0.025                           # 今消費できる累積α
> other   <- c(tcur, tprev, cprev, cumulal)  # 解を求めるためのおまじない
> answer  <- findroot(rename, -7, 7, other)  # 解を変数 answer に格納
> print(round(answer,digits=4))              # 表示
[1] 2.064
```

【参考】実際の例数が計画時の例数とズレた場合

- 2 群の生存割合の比較を行う試験を考え，2 群合計のイベントを 200 例集積するデザインとし，中間解析を 2 回，最終解析を 1 回行うことを計画する
- 全体の有意水準 α：0.025（片側），α 消費関数：$\alpha(t) = 0.025\, t^{1.5}$
- 1 回目の解析時（中間解析時）の例数は 100 例であった
  ⇒ t = 0.5，α(t) = 0.008838 となり，$c_1 = 2.3723$ と計算される
- 最終的な例数が 250 例と見込まれる場合（計画よりも増える場合），本当は t = 0.4 だが実際は t = 0.5 として計算しているため α を消費し過ぎていることになる・・・（要調整）
- 以降の解析では，下式の「調整された α 消費関数」を使う

$$\alpha'_*(t) = \alpha_i + \left(\frac{\alpha - \alpha_i}{\alpha - \alpha_*(t_i)}\right)\{\alpha_*(t) - \alpha_*(t_i)\}$$

α：全体の α，t_i：現在の情報分数，α_i：現時点までに消費した α，$\alpha_*(t)$：元の α 消費関数

【参考】実際の例数が計画時の例数とズレた場合

- 最初の 2 回の解析時（中間解析時）の例数は 50 例と 100 例であった
  ⇒ t = 0.25，0.5，α(t) = 0.003125, 0.008838 となった
- 最終的な例数が 250 例と見込まれる場合（計画よりも増える場合），
  本当は t = 0.2，0.4 だが実際は t = 0.25，0.5 として計算しているため
  以降の解析では，下式の「調整された α 消費関数」を使う必要がある

$$\alpha'_*(t) = 0.008838 + \left(\frac{0.025 - 0.008838}{0.025 - 0.025(0.4)^{1.5}} \right) \{0.025\, t^{1.5} - 0.025(0.4)^{1.5}\}$$

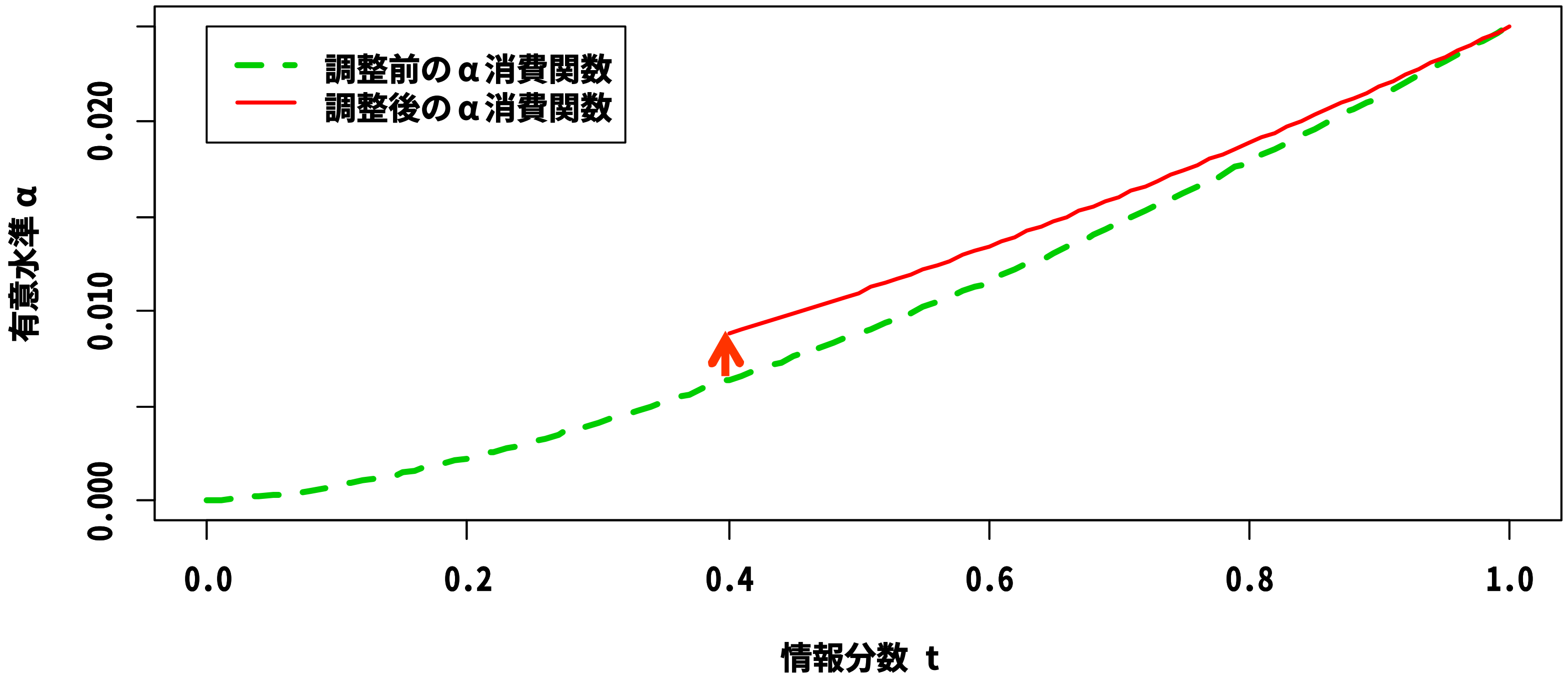
$$= 0.003364 + 0.021635\, t^{1.5}$$

- 次に 150 例の時点（ t = 0.6 ）で解析する場合の α は，以下となる

$$\alpha'_*(0.6) = 0.01341985$$

- これを用いて（ t = 0.4,0.6 として）計算し，c_2 = 2.4183 と計算される

【参考】実際の例数が計画時の例数とズレた場合



- 1 回目の解析後（ 100 例）に試験終了時の例数と t を（ 0.5 から 0.4 に）修正
- α 消費関数を調整

【参考】計算例

```
> # t=0.5 のときの棄却点 c_1
> qnorm(1-0.008838835)
[1] 2.372301
> # 調整したα消費関数
> g <- function(t) 0.003364+0.021635*t^1.5
> g(0.6)
[1] 0.01341904
> # t=0.6 のときの棄却点 c_2
> tcur    <- 0.6                                # 今から計算する t
> tprev   <- c(0.4)                             # 今までの t
> cprev   <- c(2.372301)                        # 今までの棄却点
> cumulal <- 0.01341904                         # 今消費できる累積α
> other   <- c(tcur, tprev, cprev, cumulal)     # 解を求めるためのおまじない
> answer  <- findroot(rename, -7, 7, other)     # 解を変数 answer に格納
> print(round(answer,digits=4))                 # 表示
[1] 2.4183
> # グラフ
> f <- function(t) 0.025*t^1.5
> g <- function(t) ifelse(t<0.4, NA, 0.003364+0.021635*t^1.5)
> curve(f,xlim=c(0,1),ylim=c(0,0.025),col=3,lty=2,lwd=3,ann=F)
> par(new=T)
> curve(g,xlim=c(0,1),ylim=c(0,0.025),col=2,lty=1,lwd=2,xlab="情報分数 t",ylab="有意水準α")
> legend(0,0.025,c("調整前のα消費関数","調整後のα消費関数"),col=3:2,lty=2:1,lwd=3:2, ncol=1)
```

α 消費関数のまとめと補足

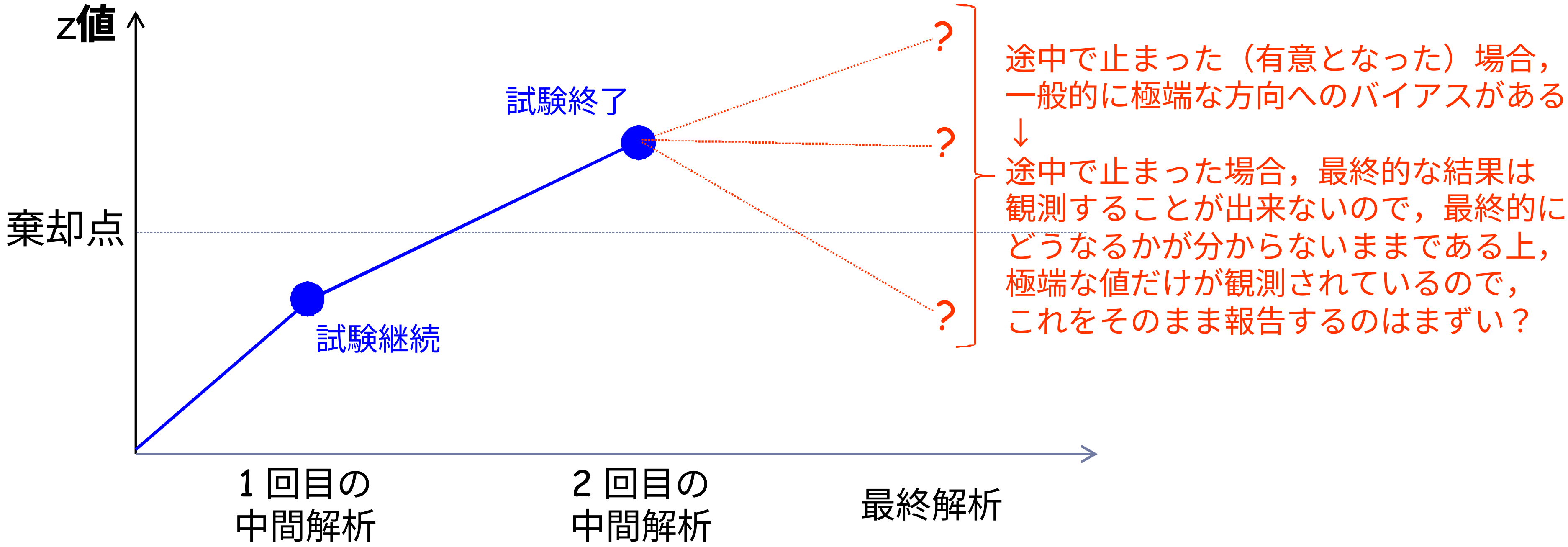
- 患者さんが不均等に増えても各解析時の棄却点を求めることが出来る ⇒ 将来の患者さんの集積状況に柔軟に対応できる
  - 有名なのは Pocock 型や O'brien-Fleming 型
  - 「$\alpha(0) = 0$，$\alpha(1) = 0.025$ を満たす増加関数」であれば何でも良い ⇒ $\alpha(t) = 0.025\,t$（線形関数）とか $\alpha(t) = 0.025\,t^{1.5}$ など
- いつ解析しても良いのだが，あまりにも例数が少ない状況下で中間解析を行うと危険（漸近的に正規分布に従うだけ）なので，ある程度患者さんが集まった段階で行うのが得策
- 数式上は，試験開始時に解析の回数を固定しなくてもよい（途中で変えても良い ⇒ 別の問題は発生するかもしれないが・・・）・・・が，中間解析の結果（$Z_1, Z_2, \ldots$）から解析時期や回数を変更するのはマズい（Type I error があふれる可能性）
- 例数やイベント数が計画時よりも増えた場合は α 消費関数を調整する

本日のメニュー

1. 情報分数 t，z-score，B-value
2. Pocock の方法と O'brien-Fleming の方法
3. Lan & DeMets の α 消費関数
4. 中間解析後の推定
   - p 値の調整
   - 信頼区間と点推定値の調整

中間解析を伴う試験終了後・・・

- 中間解析を伴う試験の報告書は（普通の臨床試験のように）無調整の点推定値，信頼区間，p 値が報告されることがほとんど（9 割程度）
- 一方で，中間解析を伴う試験において，無調整の結果は過大推定になるという話もある※ ⇒ かといって調整方法の決定版はない・・・

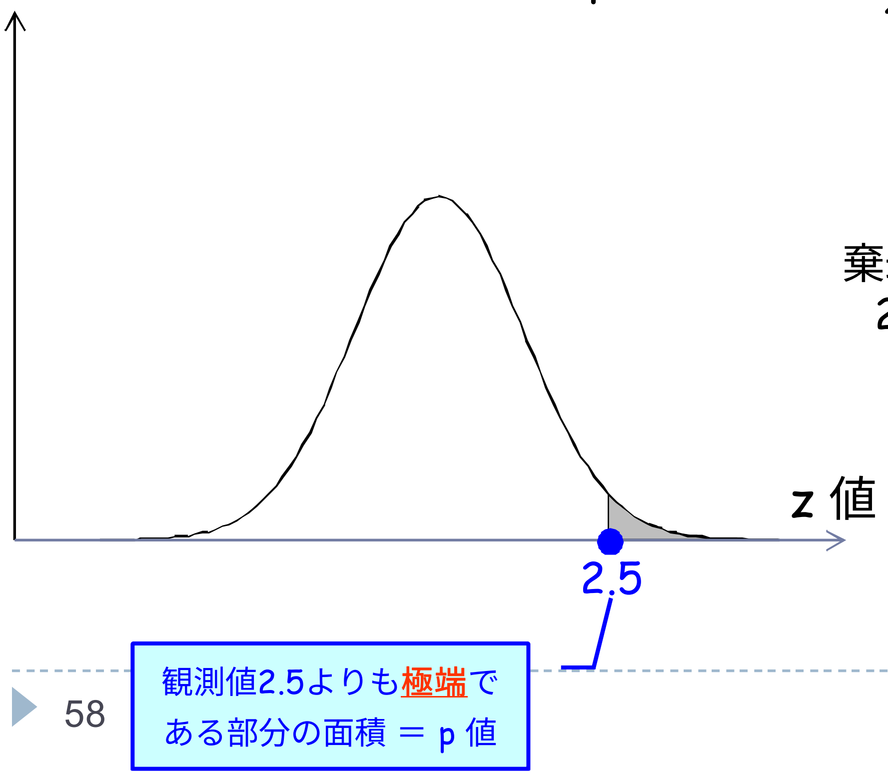


※ 有効中止の場合 ⇒ 無効中止の場合は過小推定になるという話あり

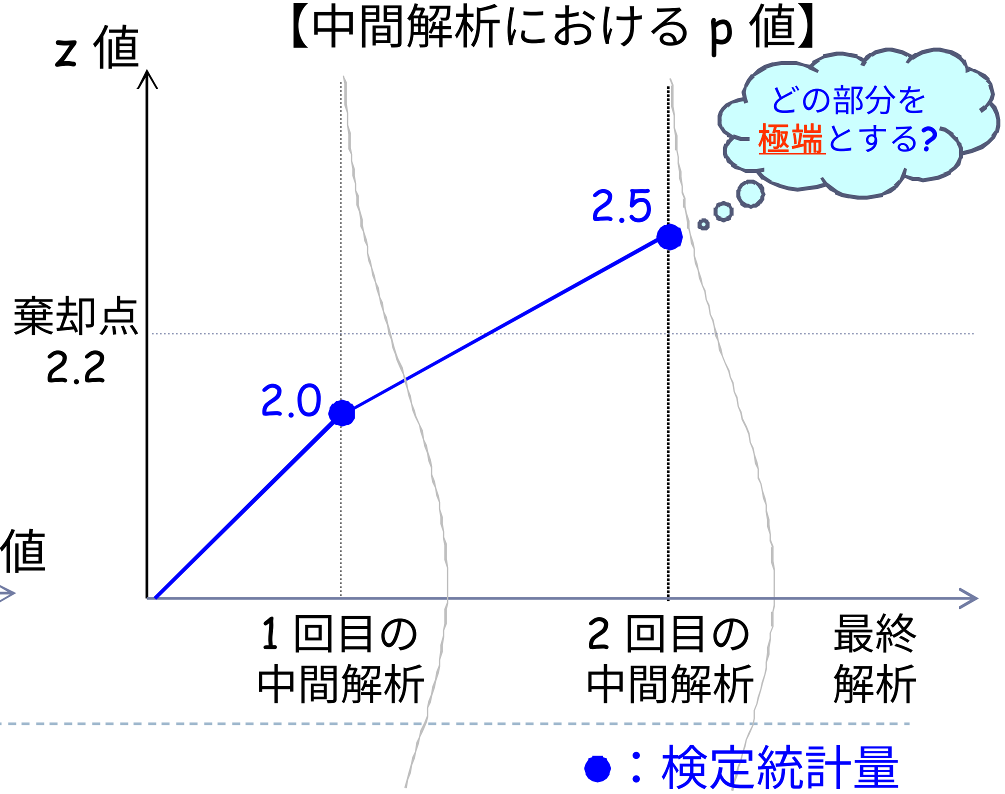
p 値について

- p 値：「帰無仮説下で観測された値よりも極端なこと」が起きる確率
- 中間解析を行わない通常の検定では上記の定義で問題は起きないが，中間解析を行った場合は「極端」の定義が難しい

【通常の検定における p 値】



【中間解析における p 値】



p 値について

- 「極端」とするルールを決めるのは難しい
  - 「$t = 0.2$，$z = 2.0$」と「$t = 0.5$，$z = 2.5$」はどちらが「極端」？
    (t, z) の両方を考慮して大小関係のルールを決める必要がある
    ⇒ (t, z) は情報分数 t のときの z-score であることを表す
- 大小関係のルール ⇒「$(t_2, z_2) \succeq (t_1, z_1)$」は左の方が極端という意味
  - z-score ordering：情報分数 t に関係なく z-score が大きい方が極端
    $(t_2, z_2) \succeq (t_1, z_1) \Leftrightarrow z_2 \geqq z_1$
  - B-value ordering：B-value が大きい方が極端
    $(t_2, B(t_2)) \succeq (t_1, B(t_1)) \Leftrightarrow t_2^{1/2} z_2 \geqq t_1^{1/2} z_1$
  - MLE ordering：省略
  - Stagewise ordering：「情報分数が小さい」又は「情報分数は等しいが z-score が大きい」方が極端〔おすすめ〕
    $(t_2, z_2) \succeq (t_1, z_1) \Leftrightarrow$ 「$t_2 < t_1$」or「$t_2 = t_1$，$z_2 \geqq z_1$」

p 値について

- 例 6：「3 回のうち 2 回目で中止，(t, z) は (0.2, 2.0) と (0.5, 2.5)，棄却点は各回とも 2.2 」の場合
  - z-score ordering：Pr{ (t, Z) $\succeq$ (0.5,2.5) }
    p = Pr{ Z(0.2) ＞ 2.2 }
    + Pr{ Z(0.2) ≦ 2.2, Z(0.5) ＞ 2.5}
    + Pr{ Z(0.2) ≦ 2.2, Z(0.5) ≦ 2.2, Z(1.0) ＞ 2.5 }
    ※ p 値が将来の解析時期（実施していない結果）に影響されてしまう・・・
  - B-value ordering ：Pr{ B(t) ＞ $0.5^{1/2}$×2.5) } = Pr{ B(t) ＞ 1.8 }
    p = Pr{ B(0.2) ＞ 1.8 }
    + Pr{ B(0.2) ≦ 1.0, B(0.5) ＞ 1.8}
    + Pr{ B(0.2) ≦ 1.0, B(0.5) ≦ 1.6, B(1.0) ＞ 1.8 }
    ※ p 値が将来の解析時期（実施していない結果）に影響されてしまう・・・
  - MLE ordering も同様の問題が起きる
  - しかし，Stagewise ordering の場合はこのようなことは起きない
    （ p 値は将来の解析時期に影響されない）

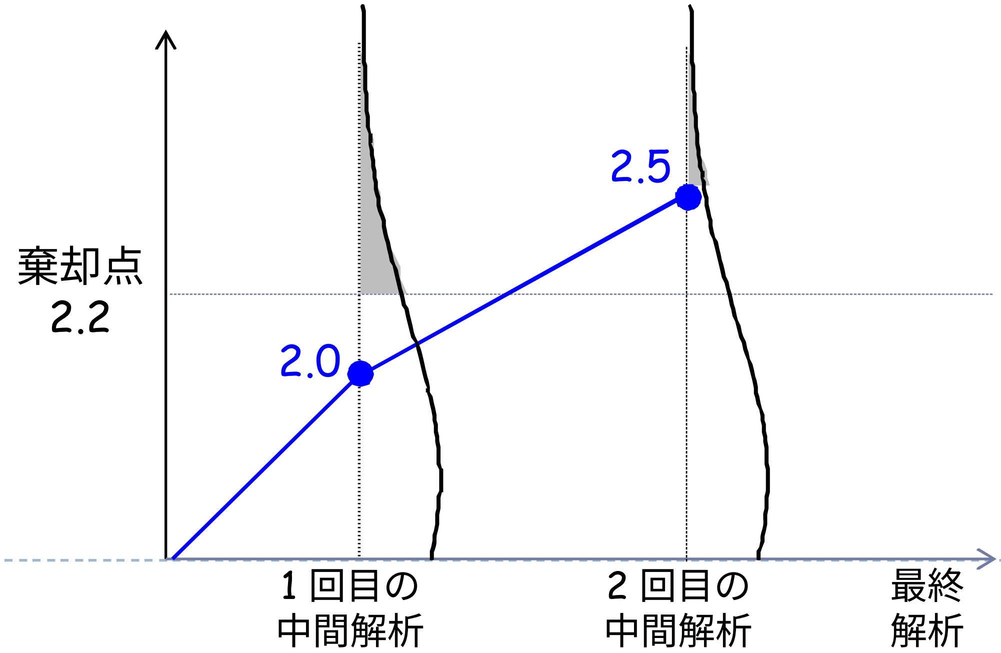
$B(t) = t^{1/2}Z(t)$, $1.0 = 0.2^{1/2}\times 2.2$, $1.6 = 0.5^{1/2}\times 2.2$

Stagewise ordering による p 値の調整

- 例 6：「3 回のうち 2 回目（ j = 2 ）で中止，(t, z) は (0.2, 2.0) と (0.5, 2.5)，棄却点は各回とも 2.2 」の場合
- Stagewise ordering：$p = \Pr\{\bigcup_{i=1}^{j-1} Z(t_i) > c_i\} + \Pr\{Z(t_j) > z_j\}$

  p = Pr{ Z(0.2) ＞ 2.2 }

  + Pr{ Z(0.2) ≦ 2.2, Z(0.5) ＞ 2.5} = 0.01825（1.8%）
- p 値が将来の解析時期（実施していない結果）に影響されない！

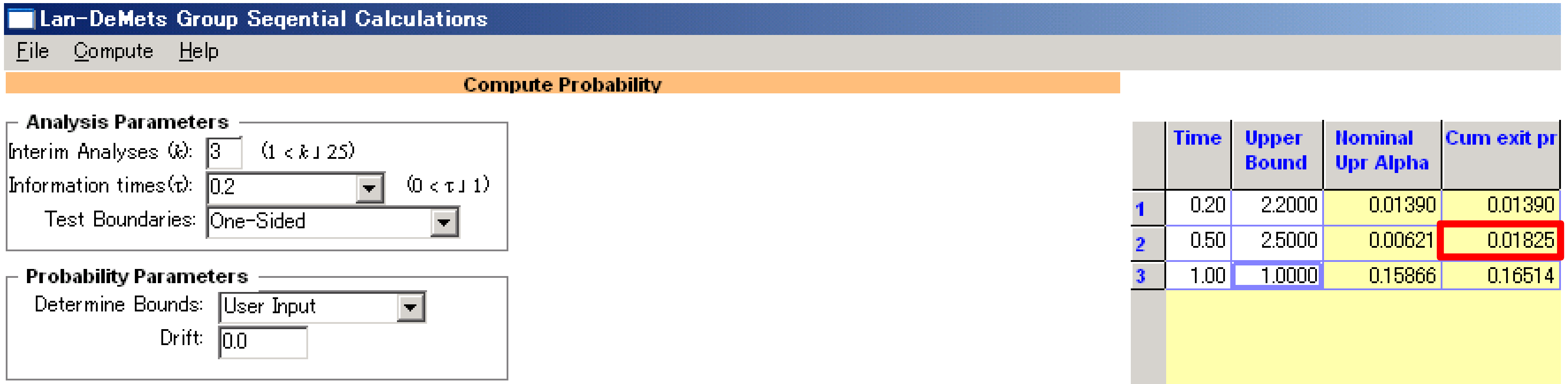


Stagewise ordering による p 値の調整

- 例 6：「3 回のうち 2 回目（j = 2）で中止，(t, z) は (0.2, 2.0) と (0.5, 2.5)，棄却点は各回とも 2.2 」の場合
- Stagewise ordering：$p = \Pr\{\bigcup_{i=1}^{j-1} Z(t_i) > c_i\} + \Pr\{Z(t_j) > z_j\}$

  p = Pr{ Z(0.2) ＞ 2.2 }

  + Pr{ Z(0.2) ≦ 2.2, Z(0.5) ＞ 2.5}
- p 値が将来の解析時期（実施していない結果）に影響されない！



	Time	Upper Bound	Nominal Upr Alpha	Cum exit pr
1	0.20	2.2000	0.01390	0.01390
2	0.50	2.5000	0.00621	0.01825
3	1.00	1.0000	0.15866	0.16514

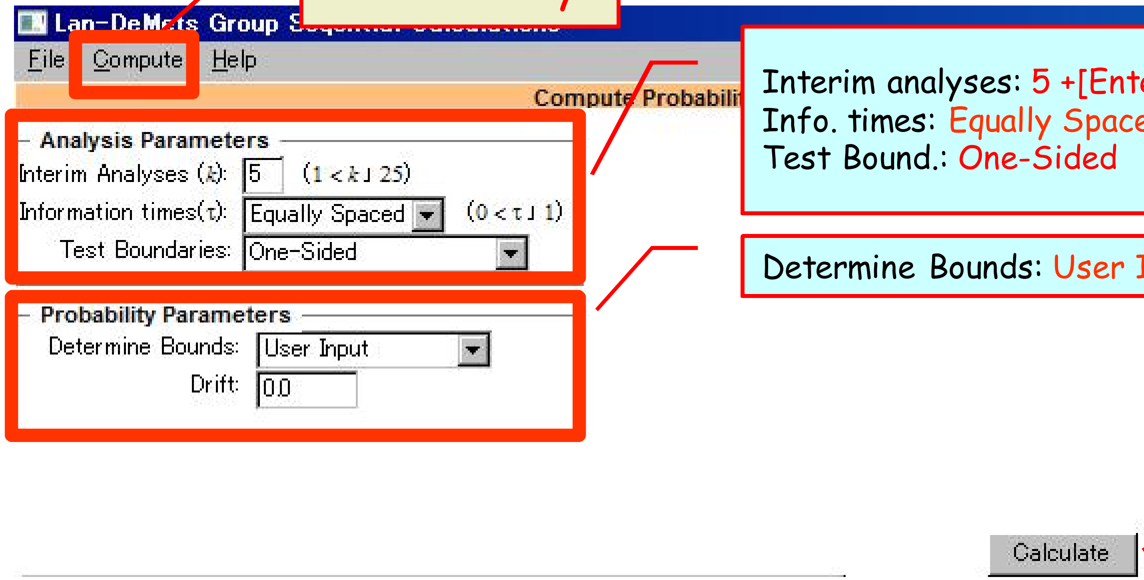
例 7：p 値について

Probability

Interim analyses: 5 +[Enter]キー
Info. times: Equally Spaced
Test Bound.: One-Sided

Determine Bounds: User Input

Upper Bound を入力



	Time	Upper Bound	Nominal Upr Alpha	Cum exit pr
1	0.20	4.5600	0.00000	0.00000
2	0.40	3.2300	0.00062	0.00062
3	0.60	2.9400	0.00164	0.00199
4	0.80	1.0000	0.15866	0.15867
5	1.00	1.0000	0.15866	0.20302

4回目以降は適当な値

全て入力したらクリック
⇒右上の Upper Bound が結果

- $t_1 = 0.2$, $t_2 = 0.4$, $t_3 = 0.6$, $t_4 = 0.8$, $t_5 = 1.0$
- 各回における棄却点：4.56, 3.23, 2.63, 2.28, 2.04 (通常の O'Brien Fleming 法)
- 3 回目で z-score が 2.94 となったので試験中止

  ⇒ Stagewise Ordering による p 値は 0.00199 (0.2%)

信頼区間について〔中間解析でない通常の場合〕

- X_i：被験者 i の対照薬の反応，Y_i：被験者 i の被験薬の反応（同じ被験者）
  $D_i = Y_i - X_i \sim N(\delta, \sigma^2)$ とする ⇒ σ^2 は簡単のため既知とする
- 両側 95% 信頼区間を $(\delta_L, \delta_U) = (\hat{\delta} - 1.96se(\hat{\delta}), \hat{\delta} + 1.96se(\hat{\delta}))$ とする
- 観測された z-score を z_{obs}，$Z \sim N(\delta_L/se(\hat{\delta}), 1)$ とすると，信頼下限 δ_L は $\Pr\{Z \geqq z_{obs}\} = \alpha/2 = 0.025$ となるように（δ_L を）決めた値

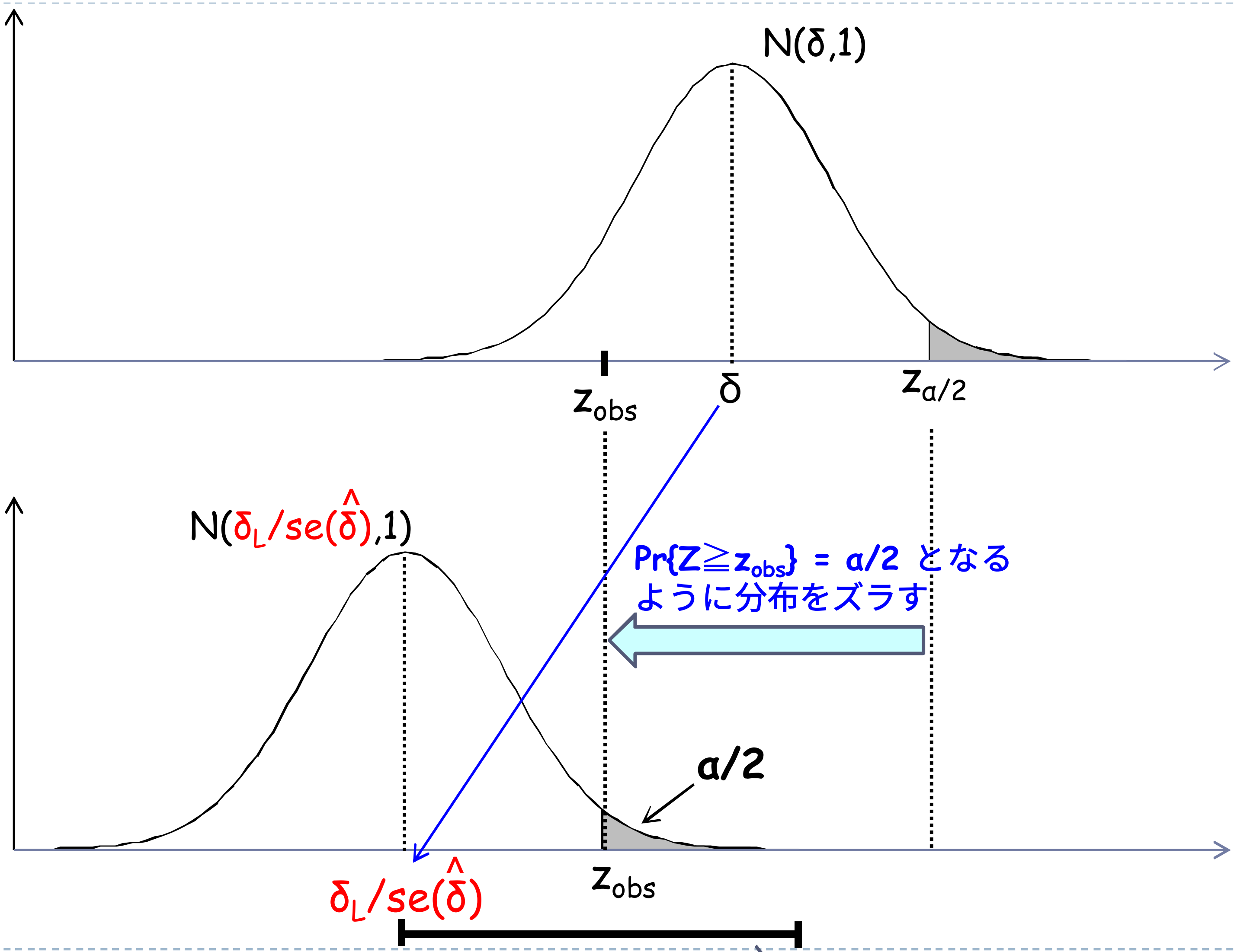
$$\Pr\{Z \geq z_{obs}\} = \Pr\left\{Z - \frac{\delta_L}{se(\hat{\delta})} \geq z_{obs} - \frac{\delta_L}{se(\hat{\delta})}\right\}$$

$$= \Pr\left\{\underbrace{Z - \frac{\delta_L}{se(\hat{\delta})}}_{\sim N(0,1)} \geq \frac{\hat{\delta} - \delta_L}{se(\hat{\delta})}\right\} = \alpha / 2$$

$$\frac{\hat{\delta} - \delta_L}{se(\hat{\delta})} = z_{1-\alpha/2} \Leftrightarrow \delta_L = \hat{\delta} - z_{1-\alpha/2} \cdot se(\hat{\delta}) = \underbrace{\hat{\delta} - 1.96 \cdot se(\hat{\delta})}_{\text{信頼区間の下限の式}}$$

$\alpha = 0.05$

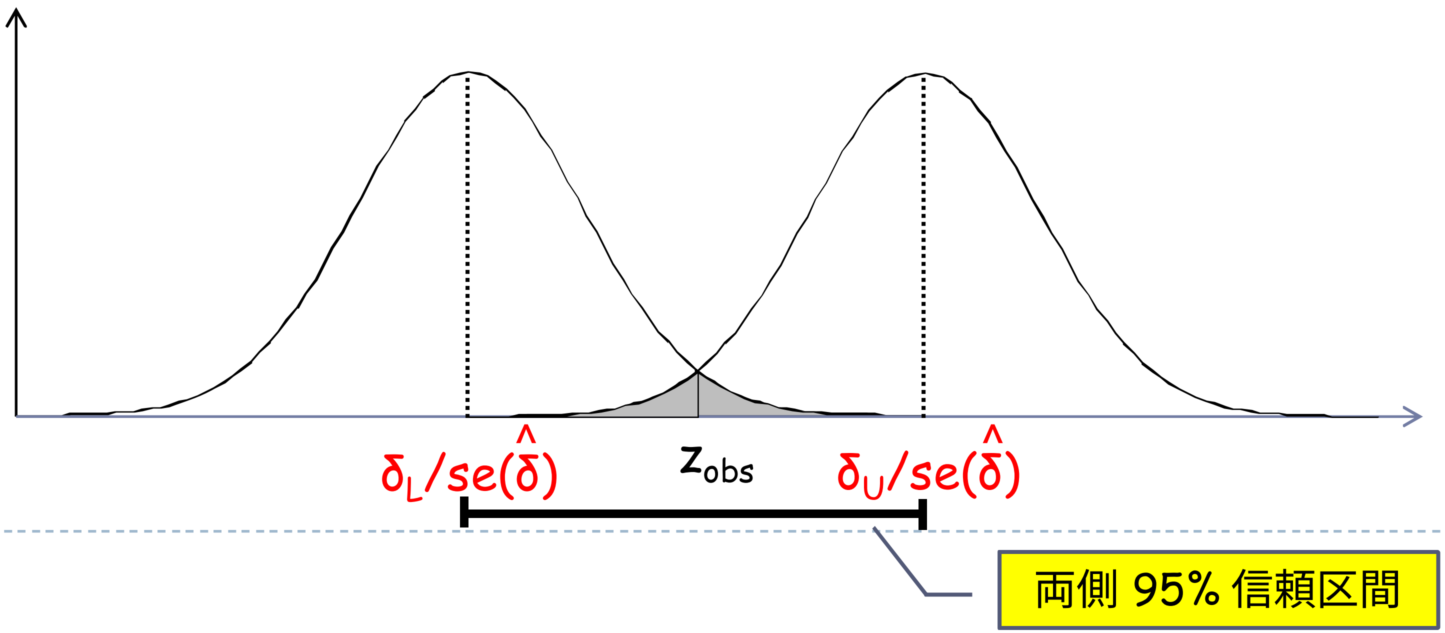
信頼区間について〔中間解析でない通常の場合〕



両側 95% 信頼区間

信頼区間について〔中間解析でない通常の場合〕

- X_i：対照群の反応，Y_i：治療群の反応，$D_i = Y_i - X_i \sim N(\delta, \sigma^2)$ とする
  ⇒ σ² は簡単のため既知とする
- 両側 95% 信頼区間を $(\delta_L, \delta_U) = (\hat{\delta} - 1.96se(\hat{\delta}), \hat{\delta} + 1.96se(\hat{\delta}))$ とする
- 観測された z-score を z_{obs}，$Z \sim N(\delta_U/se(\hat{\delta}), 1)$ とすると，信頼上限 δ_U は $Pr\{Z \leqq z_{obs}\} = 0.025$ となるように（δ_U を）決めた値となる



信頼区間について〔中間解析でない通常の場合〕

- この方法で δ_L と δ_U を見つけるのは結構面倒 ⇒ 以下は計算の一例
- $\hat{\delta} = 3$，$se(\hat{\delta}) = 1$，$z_{obs} = 3$ の場合，[-7, 7] の範囲で δ_L を見つける
  - $\delta_L = -7$ とすると：$Z \sim N(-7, 1)$，$\Pr\{Z \geqq 3\} = 1\text{-}\Phi(10) = 0.000\ldots$
  - $\delta_L = 0$ とすると：$Z \sim N(0, 1)$，$\Pr\{Z \geqq 3\} = 1\text{-}\Phi(3) = 0.001$
  - $\delta_L = 3.5$ とすると：$Z \sim N(3.5, 1)$，$\Pr\{Z \geqq 3\} = 1\text{-}\Phi(-0.5) = 0.308$
  - $\delta_L = 1.7$ とすると：$Z \sim N(1.7, 1)$，$\Pr\{Z \geqq 3\} = 1\text{-}\Phi(1.3) = 0.096$
  - $\delta_L = 0.8$ とすると：$Z \sim N(0.8, 1)$，$\Pr\{Z \geqq 3\} = 1\text{-}\Phi(2.2) = 0.014$
  - $\delta_L = 1.0$ とすると：$Z \sim N(1.0, 1)$，$\Pr\{Z \geqq 3\} = 1\text{-}\Phi(2) = 0.023$
  - $\delta_L = 1.04$ とすると：$Z \sim N(1.04, 1)$，$\Pr\{Z \geqq 3\} = 1\text{-}\Phi(1.96) = 0.025$
- 以上の計算より $\delta_L = 1.04$，同様に計算して $\delta_U = 4.96$・・・と，
  δ_L や δ_U の値を少しずつ変えて解を探す(= grid search) ことになる

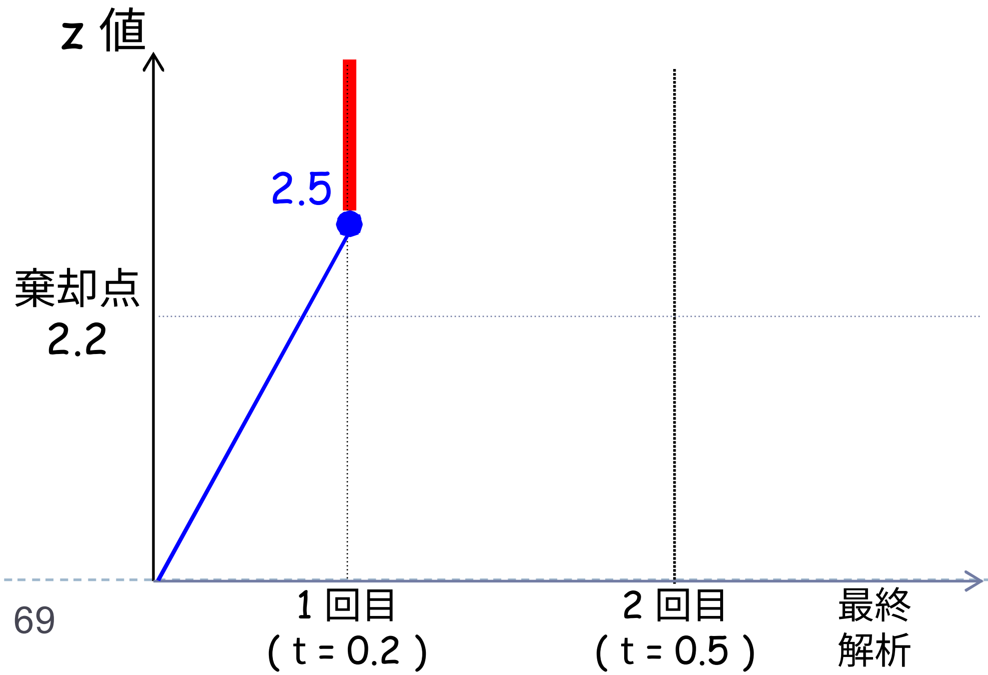
$\delta_L = 3 - 1.96 \cdot 1 = 1.04$

Stagewise ordering による信頼区間と点推定値の調整

- 信頼区間の下限：$\Pr\{ (t, Z) \succeq (t_{obs} , Z_{obs}) \} = 0.025$ を満たす δ_L
- 信頼区間の上限：$\Pr\{ (t, Z) \succeq (t_{obs} , Z_{obs}) \} = 0.975$ を満たす δ_U
- 点推定値：$\Pr\{ (t, Z) \succeq (t_{obs} , Z_{obs}) \} = 0.5$ を満たす δ_{mid}
- ここで $\theta = \delta / se(\hat{\delta})$ とおくと，上記は以下のように書き換わる
  - 信頼区間の下限：$\Pr\{ (t, Z) \succeq (t_{obs} , Z_{obs}) \} = 0.025$ を満たす θ_L
  - 信頼区間の上限：$\Pr\{ (t, Z) \succeq (t_{obs} , Z_{obs}) \} = 0.975$ を満たす θ_U
  - 点推定値：$\Pr\{ (t, Z) \succeq (t_{obs} , Z_{obs}) \} = 0.5$ を満たす θ_{mid}

Stagewise ordering による信頼区間と点推定値の調整

- 信頼区間の下限：$\Pr\{ (t, Z) \succeq (t_{obs}, Z_{obs}) \} = 0.025$ を満たす θ_L
- 信頼区間の上限：$\Pr\{ (t, Z) \succeq (t_{obs}, Z_{obs}) \} = 0.975$ を満たす θ_U
- 点推定値：$\Pr\{ (t, Z) \succeq (t_{obs}, Z_{obs}) \} = 0.5$ を満たす θ_{mid}



$\Pr\{ Z(0.2) > 2.5 \} = 0.025$

となる $\theta_L = \delta_L / se(\hat{\delta})$ を探す

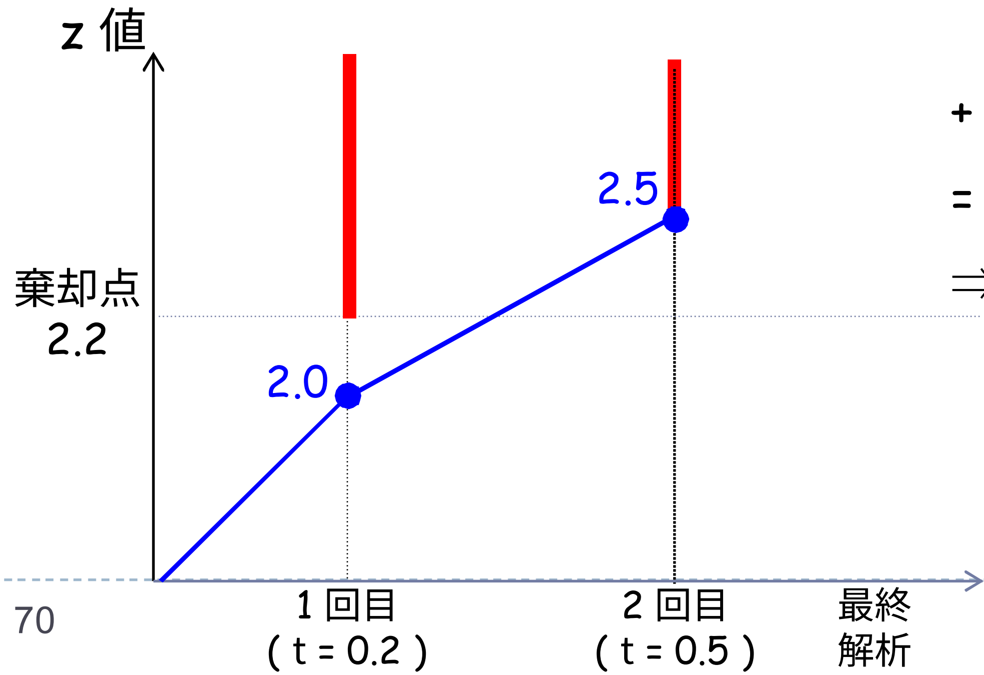
$\Rightarrow$ $Z(0.2)$ は $H_0：\theta = \theta_L$ の下での

分布となり $N(0.2^{1/2}\theta_L, 1)$

●：検定統計量

Stagewise ordering による信頼区間と点推定値の調整

- 信頼区間の下限：$\Pr\{ (t, Z) \succeq (t_{obs}, Z_{obs}) \} = 0.025$ を満たす θ_L
- 信頼区間の上限：$\Pr\{ (t, Z) \succeq (t_{obs}, Z_{obs}) \} = 0.975$ を満たす θ_U
- 点推定値：$\Pr\{ (t, Z) \succeq (t_{obs}, Z_{obs}) \} = 0.5$ を満たす θ_{mid}



$\Pr\{ Z(0.2) > 2.5 \}$

$+ \Pr\{ Z(0.2) \leqq 2.2, Z(0.5) > 2.5 \}$

$= 0.025$ となる $\theta_L = \delta_L / se(\hat{\delta})$ を探す

$\Rightarrow (Z(0.2), Z(0.5))$ は $H_0 : \theta = \theta_L$ の下での二次元正規分布：

平均は $(0.2^{1/2}\,\theta_L, 0.5^{1/2}\,\theta_L)$，

分散 1，共分散 $(0.2/0.5)^{1/2} = 0.63$

●：検定統計量

WinLD で使用する Stagewise ordering による調整

- 信頼区間の下限：$\Pr\{ (t, Z) \succeq (t_{obs} , Z_{obs}) \} = 0.025$ を満たす θ_L
- 信頼区間の上限：$\Pr\{ (t, Z) \succeq (t_{obs} , Z_{obs}) \} = 0.975$ を満たす θ_U
- 点推定値：$\Pr\{ (t, Z) \succeq (t_{obs} , Z_{obs}) \} = 0.5$ を満たす θ_{mid}

$\Rightarrow$ $\theta = \delta / se(\hat{\delta})$ に対する解を求めるので，計算結果に $se(\hat{\delta})$ を掛け算する必要がある点に注意

例 8：信頼区間（≠ repeated confidence interval）

- t_1 = 0.35, t_2 = 0.65, t_3 = 1.0，2 回目で中止
- 棄却限界値：（3.5521, 2.5581, 1.9893）
- 2 回目の z = 2.8，推定値の標準誤差 = 0.083

★ 信頼区間の下限：$\Pr\{ (t, Z) \succeq (t_{obs}, Z_{obs}) \} = 0.025$ を満たす θ_L

- $\Pr\{ Z(0.35) > 3.5521 \} + \Pr\{ Z(0.35) \leqq 3.5521, Z(0.65) > 2.8\} = 0.025$ となる θ_L を探す
- (Z(0.35) , Z(0.65)) は平均がそれぞれ $0.35^{1/2}\theta_L$， $0.65^{1/2}\theta_L$，分散が 1，共分散が $(0.35/0.65)^{1/2} = 0.73$ の二次元正規分布
- θ_L をいろいろ変えて二次元正規分布から確率を計算し，grid-search する

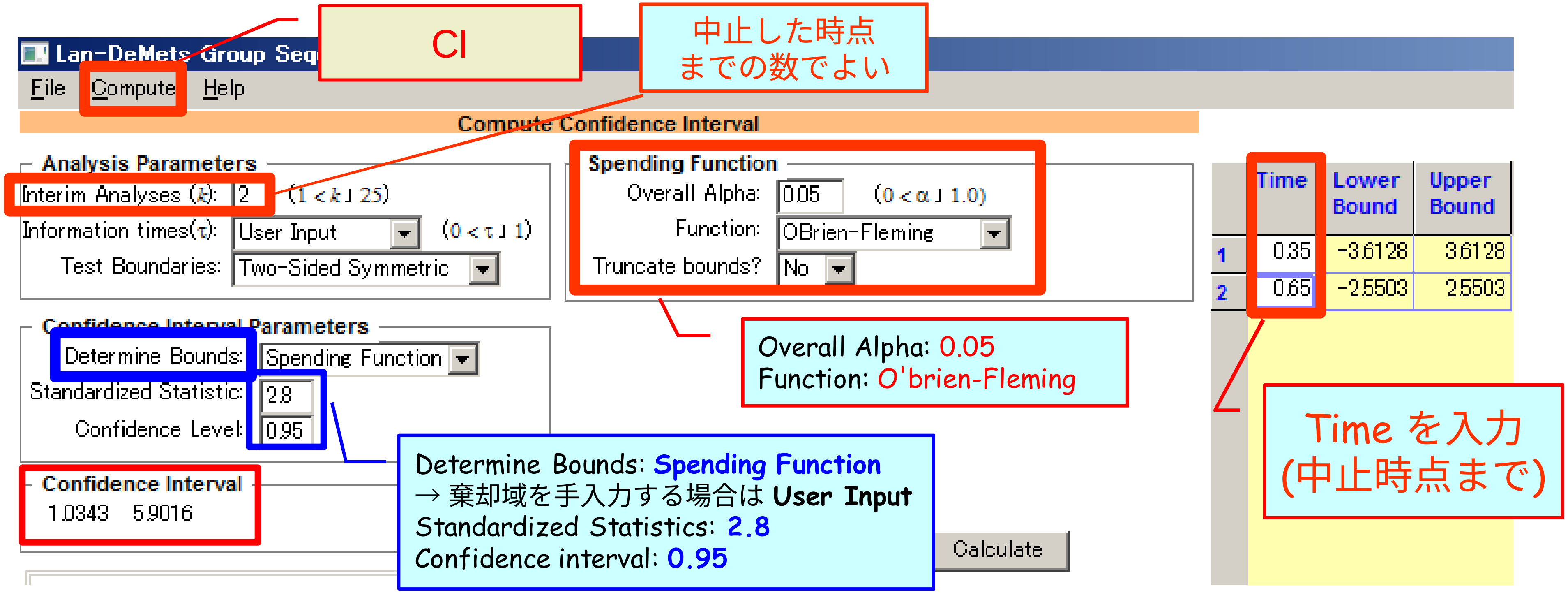
例： $\theta_L = 1.0$ の場合，平均 $(0.35^{1/2}, 0.65^{1/2}) = (0.6, 0.8)$，分散 1，共分散 0.73 の正規分布から $\Pr\{ Z(0.35) > 3.5521 \} + \Pr\{ Z(0.35) \leqq 3.5521, Z(0.65) > 2.8 \}$ を計算する

- 信頼区間の上限と点推定値：★式の右辺をそれぞれ 0.975，0.5 にして計算
  ⇒ ちなみに 95%信頼区間は [1.0343×0.083, 5.9016×0.083] = (0.085, 0.489)
  点推定値は 3.9554×0.083 = 0.328 となる

例８：信頼区間（乱数生成による計算）

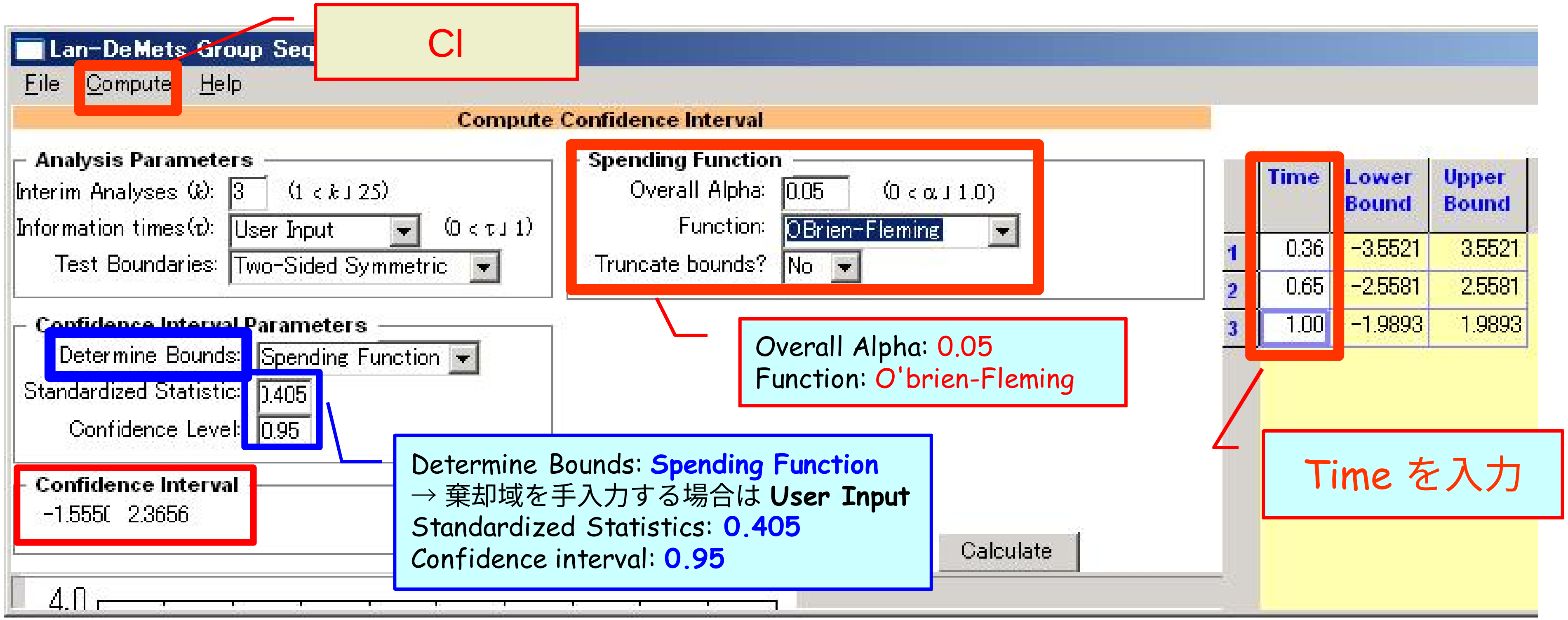
```
> library(MASS)
> V <- matrix(c(1, 0.73380, 0.73380, 1), 2, 2)
>
> # 信頼下限用　θ=1.0343 として計算
> M <- c(0.61190, 0.83388)
> Z <- mvrnorm(n=1000000, M, V)
> result <- ifelse( (Z[,1]>3.5521)|(Z[,1]<3.5521 & Z[,2]>2.8), 1, 0)
> mean(result)
[1] 0.025067

> # 点推定値用　θ=3.9954 として計算
> M <- c(2.05234, 2.79688)
> Z <- mvrnorm(n=1000000, M, V)
> result <- ifelse( (Z[,1]>3.5521)|(Z[,1]<3.5521 & Z[,2]>2.8), 1, 0)
> mean(result)
[1] 0.49956

> # 信頼上限用　θ=5.9016 として計算
> M <- c(3.49143, 4.75802)
> Z <- mvrnorm(n=1000000, M, V)
> result <- ifelse( (Z[,1]>3.5521)|(Z[,1]<3.5521 & Z[,2]>2.8), 1, 0)
> mean(result)
[1] 0.974883
```

例 8：信頼区間（WinLD による計算）



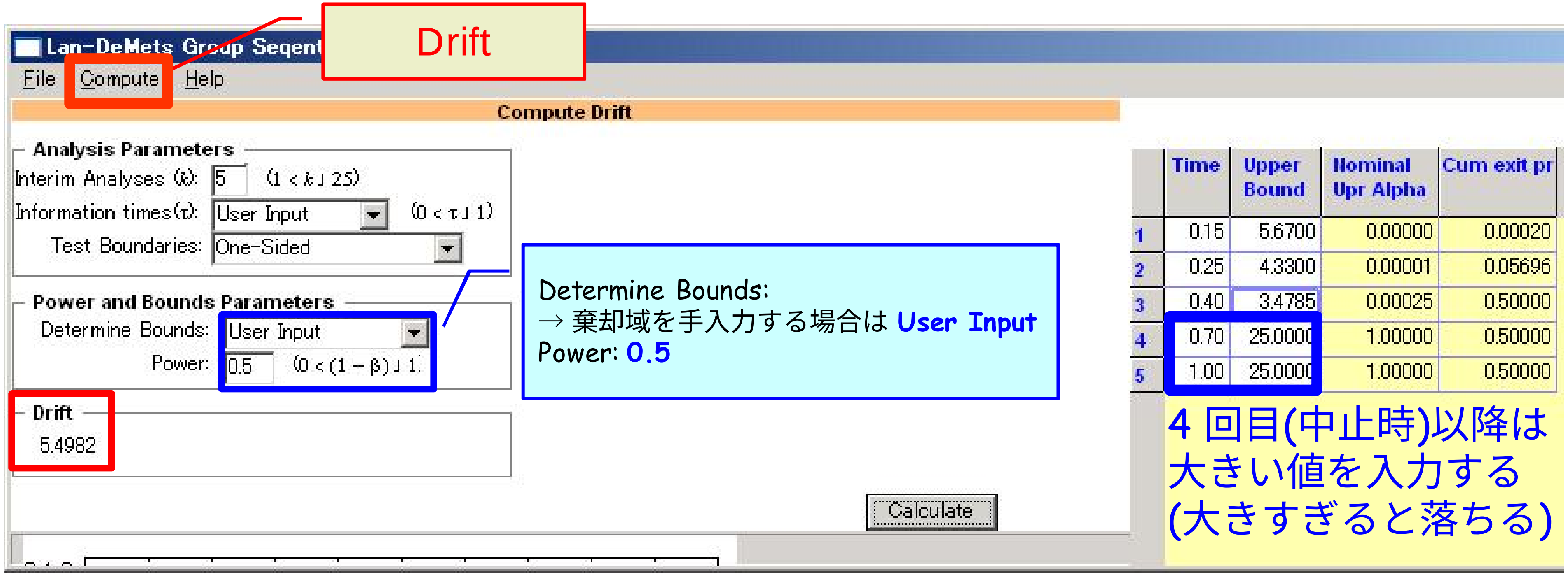
- t_1 = 0.35, t_2 = 0.65, t_3 = 1.0，2 回目で中止
- 棄却限界値： O'brien-Fleming 型（3.6128, 2.5503, ...）
- 2 回目の z = 2.8，推定値の標準誤差 = 0.083
  ⇒ 95%信頼区間は [1.0343×0.083, 5.9016×0.083] = (0.085, 0.489)

例 9：信頼区間（≠ repeated confidence interval）



- t_1 = 0.36, t_2 = 0.65, t_3 = 1.0，最後まで解析した（途中で中止せず）
- 棄却限界値： O'brien-Fleming 型（3.5521, 2.5581, 1.9893）
- 3 回目の z = 0.405，推定値の標準誤差 = 0.046
  ⇒ 95%信頼区間は [-1.5550×0.046, 2.3656×0.046] = (-0.071, 0.108)

例 10：点推定値（Median Unviased Estimator）



- $t_1 = 0.15$, $t_2 = 0.25$, $t_3 = 0.4$, $t_4 = 0.7$, $t_5 = 1.0$，3 回目の解析で中止
- 棄却限界値：5.67, 4.33, 3.36, 2.44, 2.00
- 3回目の $z = 3.4785$，推定値の標準誤差 = 0.06

  ⇒ 点推定値は 5.4982×0.06 = 0.330

Stagewise ordering に関するまとめ

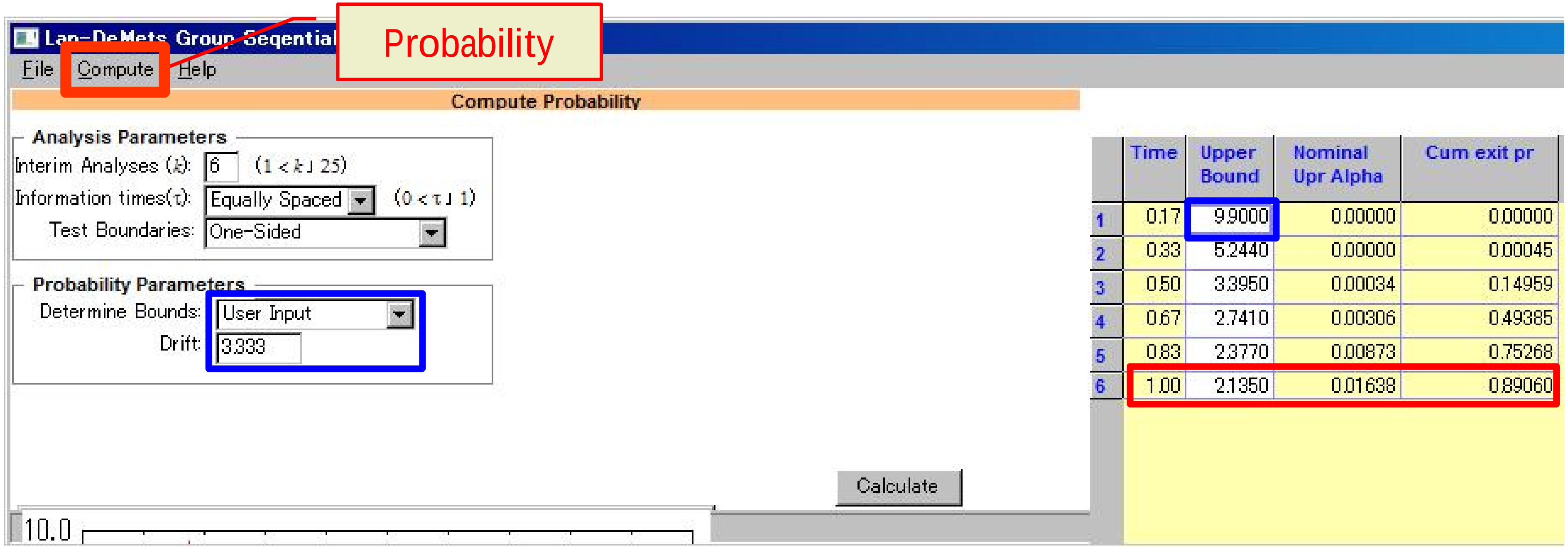
- Stagewise ordering で求めた p 値
  - 一様分布に従う
  - 棄却域と対応している（p 値 ＜ α ⇔ 統計量は棄却点を越える）
  - 将来の解析時期や解析回数に影響されない
  - 1 回目の解析で中止となった場合，通常の p 値と一致する
- Stagewise ordering で求めた信頼区間
  - p 値と対応している（p 値 ＜ α ⇔ $\theta_L > 0$ 又は $\theta_U < 0$ ）
  - 将来の解析時期や解析回数に影響されない
  - 1 回目の解析で中止となった場合，通常の信頼区間と一致する
  - くり返し信頼区間（repeated confidence interval）とは区別すること
- Stagewise ordering で求めた点推定値
  - 「Stagewise ordering で求めた信頼区間の真ん中」という方法※もある

※ mid-point estimator

本日のメニュー

1. イントロ
2. 情報分数 t，z-score，B-value
3. Pocock の方法と O'brien-Fleming の方法
4. Lan & DeMets の α 消費関数
5. 中間解析後の推定
   - p 値の調整
   - 信頼区間と点推定値の調整

【メモ】検出力



- $t_1 = 0.17$, $t_2 = 0.33$, $t_3 = 0.50$, $t_4 = 0.67$, $t_5 = 0.83$, $t_6 = 1.0$
- 棄却限界値：5.029, 3.556, 2.903, 2.514, 2.249, 2.053
- 1 回目の t 値 = 59.035（→ 9.9 で十分代用可）
- ドリフト $\theta = \delta/(2\sigma^2/18)^{1/2} = 3.333$ （$\delta = 2$, $\sigma = 1.8$）⇒ <u>検出力 89%</u>

【メモ】条件付検出力

- 条件付検出力：B(t) = b を与えた下での B(1) > $z_{\alpha/2}$ となる条件付確率
  - ドリフトパラメータ θ：試験終了時に期待される z-score
  - E[B(1) - B(t)] = θ(1 - t)，V[B(1) - B(t)] = 1 - t
  - E_θ[B(1) | B(t)=b] = b + θ(1 - t)，V_θ[B(1) | B(t)=b] = 1 - t
- 条件付検出力 $CP_\theta(t) = 1 - \Phi\left\{\dfrac{z_{\alpha/2} - E_\theta[B(1) \mid B(t) = b]}{\sqrt{1-t}}\right\}$
- 実際に検出力を求める際は，θ に何らかの仮定を置く

① 試験計画時（ 例数設計時）に用いた情報から求めた θ

② θ = 0　（悲観的な仮定）

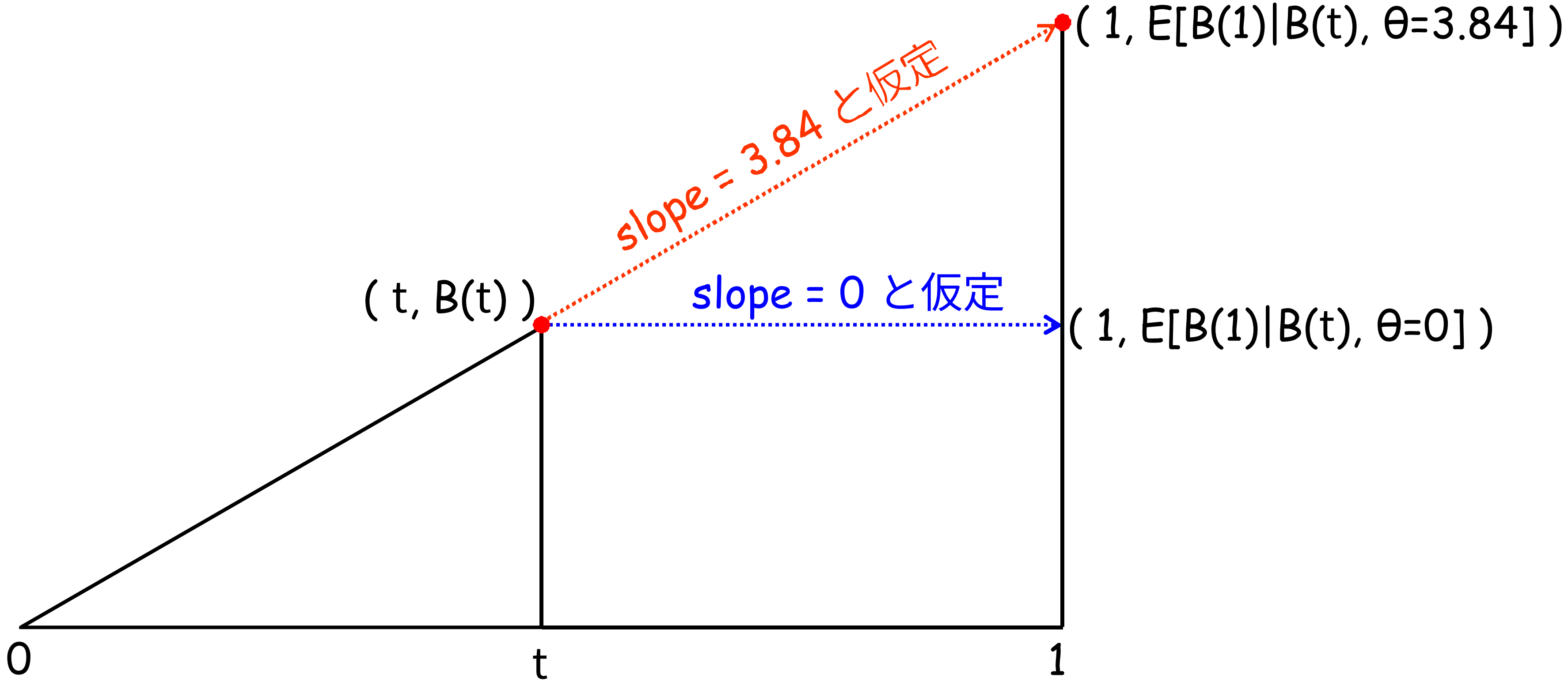
③ $\hat{\theta}$ = B(t)/t　（現時点の傾向）

【メモ】条件付検出力

- t = 0.75，B(0.75) = 0.5 とする

① slope = 試験開始時の想定： E[B(t)]=3.84t とすると，E[B(1) | B(0.75)=0.5]=1.46，V[B(1) | B(0.75)=0.5]=0.25，$CP_{3.84} = 1 - \Phi\{(1.96-1.46)/0.25^{1/2}\} = 1 - \Phi(1) = 0.16$ (α=0.05)

② slope = 0 と仮定すると， $CP_0 = 1 - \Phi\{(1.96-0)/0.25^{1/2}\} \quad 0$



Lan et.al.（1988）より引用

【メモ】参考文献

- Statistical Monitoring of Clinical Trials
  (Michael A. Proschan et. al., Springer)
- The B-Value: A Tool for Monitoring Data
  (K.K.Grodon Lan et.al., Biometrics, 1988)
- Multiple Comparisons Using R (Frank Bretz et. al., CRC press)
- The R Tips 第 2 版(オーム社)
- R 流!イメージで理解する統計処理入門(カットシステム)

終