

医学統計学の道標

- EZRによる統計的データ解析の実践 -

下川敏雄

和歌山県立医科大学 医学系研究科 医療データサイエンス講座

和歌山県立医科大学附属病院 臨床研究センター

シエーマ

1. EZRとは何か？
2. EZRにおいて実行できる統計解析 1 : 質的データ(2章)
2. EZRにおいて実行できる統計解析 2 : 量的データ(1章)
3. EZRにおいて実行できる統計解析 3 : 生存時間的データ(3章)



Rとは何か？



EZRとは



自治医科大学附属さいたま医療センター 血液科の神田善伸教授が統計解析環境RのGUI環境であるRコマンダーをカスタマイズしたものがEZRである。

EZRでは、医学統計学における諸種の統計的方法をマウスのみで実行できる。

ダウンロードは、ブラウザで「EZR」と入力すれば、左図のホームページがトップにくる。

<http://www.jichi.ac.jp/saitama-sct/SaitamaHP.files/statmed.html>

Warning!!

MacOSでは、EZRのもととなるRcommanderを動かすために、x11というプラグインが必要になる。そのため、EZRをインストールする前に、

<https://www.xquartz.org/>

から、x11をダウンロードおよびインストールする必要がある。

EZRのインストール方法：Windowsを例に説明

The screenshot shows the EZR website with the following annotations:

- STEP.1: 希望のOSを選択** (Select your preferred OS): A red box highlights the download links for Windows, Mac OS X, and Linux. A red arrow points from this box to the text "STEP.2: 「〇〇はここをクリックしてダウンロード...」をクリック" (Click on the OS name to download...).
- 右側の画面が表示される** (The right screen is displayed): A green arrow points from the right side of the website to the next screenshot.

<http://www.jichi.ac.jp/saitamasct/SaitamaHP.files/statmed.html>
「EZR」を検索エンジンで調べると最初に出てくる

右側の画面が表示される

The screenshot shows the EZR website with the following annotations:

- STEP.2: 「〇〇はここをクリックしてダウンロード...」をクリック** (Click on the OS name to download...): A red box highlights the text "Windows版はここをクリックしてダウンロードしてください(Ver. 1.36 2017/9/1)." (Click here to download the Windows version (Ver. 1.36 2017/9/1)).
- EZRの実行ファイルがダウンロードされる** (The EZR executable file is downloaded): A green arrow points from the download link to the next screenshot.

EZRの実行ファイルがダウンロードされる

The screenshot shows the EZR installation window with the following text:

EZRをインストールします
デスクトップフォルダ
EZR をインストールするフォルダを選択してください。

セットアップは次のフォルダにファイルをインストールします。
もし、EZR を違うフォルダにインストールしたければ「参照」を押して別のフォルダを選択してください。

デスクトップフォルダ
C:\Program Files\EZR 参照...

スペースが必要です: 481.23MB
スペースがあります: 326.19GB

CreateInstall Free
次へ > キャンセル

実行ファイルを実行するとインストールされる

EZRの起動方法

Windowsの場合

EZRインストール後に作成されるEZRのアイコンをダブルクリックするか、メニューからEZRを選択すればよい(通常のアプリケーションと同じ)。

MacOSの場合

MacOSの場合には、Rを起動したうえで、EZRのプラグインを読み込む必要がある。

[STEP.1] Rを起動する(のアイコン)。

[STEP.2] 次のように入力する

```
> library(Rcmdr)
```

[STEP.3] Rcommanderが起動するので、

「ツール」→「Rcmdrプラグインのロード」

としてRcmdrPlugin.EZRを選択し、「再起動しますか？」という問いに「はい」を選ぶとEZRが起動する。

Rコマンドーの動作が遅くなる場合には、「ツール」→「R.appのためのMac OS Xのapp.napの管理」でapp nappの設定をオフに変更する。

EZRのアウトプットと解釈の留意点

出力結果の一例

R コマンド

ファイル 編集 アクティブデータセット 統計解析 グラフと表 ツール ヘルプ 標準メニュー

データセット: Dataset 編集 表示 保存 モデル: <アクティブモデルなし>

Rスクリプト Rマークダウン

```
##### 1標本の t 検定#####  
res <- NULL  
(res <- t.test(Dataset$血清クレアチニン濃度, alternative='two.sided', mu=4.3, conf.level=0.95))  
cat(gettextRcmdr("mean"), " = ", res$estimate, " ", gettextRcmdr("95% CI"), " ", res$conf.int[1], "-", res$conf.int[2], " ", "  
", sep="")  
cat(gettextRcmdr("p.value"), " = ", signif(res$p.value, digits=3), " ", sep="")
```

R スクリプト

EZRではメニューで選ばれた解析方法をRに変換して実行している。ここは、そのRでの実行コマンドが表示されている。

出力

```
> ##### 1標本の t 検定#####  
> res <- NULL  
> (res <- t.test(Dataset$血清クレアチニン濃度, alternative='two.sided', mu=4.3, conf.level=0.95))  
  
One Sample t-test  
data: Dataset$血清クレアチニン濃度  
t = -2.0711, df = 5, p-value = 0.09311  
alternative hypothesis: true mean is not equal to 4.3  
95 percent confidence interval:  
 3.814410 4.352257  
sample estimates:  
mean of x  
 4.083333
```

Rでの実行コマンド (無視してよい)

Rでの結果

Rでの実行コマンド (無視してよい)

EZRでの結果 (Rでの結果の要約)

出力
EZRでの結果の出力

メッセージ

```
[1] メモ: Rコマンドのバージョン 2.4.0. Web Feb 26 12:05:06 2016  
[2] メモ: データセット Dataset には 688 行、10 列あります。  
[3] メモ: データセット Dataset には 6 行、1 列あります。
```

Rでの実行コマンド
(無視してよい)

Rでの結果

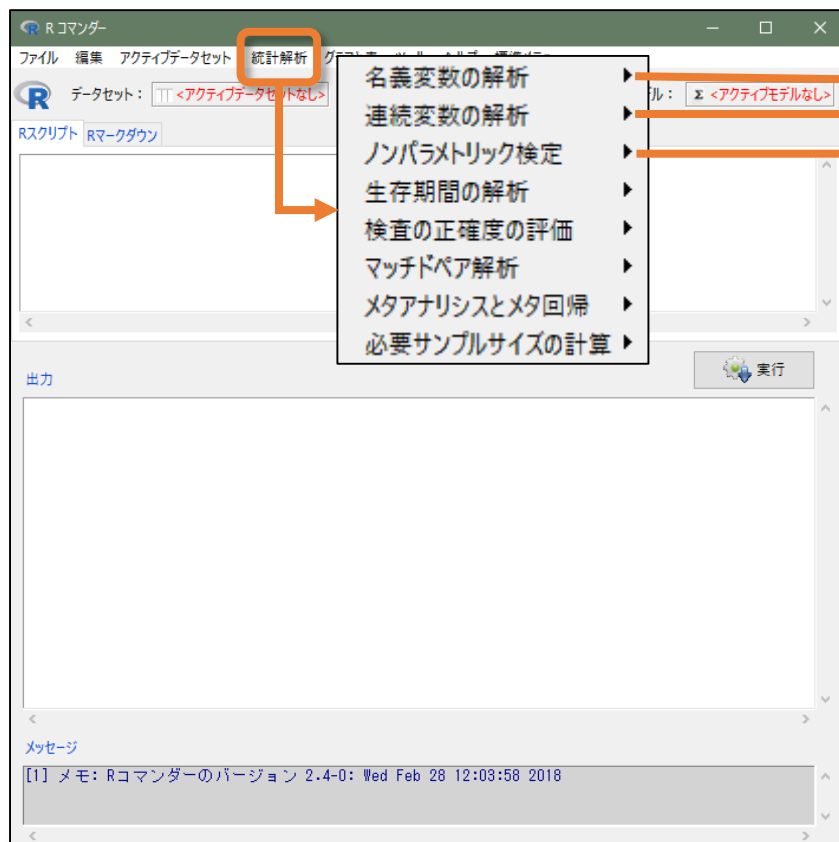
Rでの実行コマンド
(無視してよい)

EZRでの結果
(Rでの結果の要約)

これらを解釈する

EZRの結果は日本語で表示される

EZRにおけるメニュー



頻度分布
比率の信頼区間の計算
1標本の比率の検定
2群の比率の差の信頼区間の計算
2群の比率の比の信頼区間の計算

分割表の直接入力と解析
分割表の作成と群間の比率の比較(Fisherの正確検定)
対応のある比率の比較(二分割表の対称性の検定、McNemar検定)
対応のある3群以上の比率の比較(Cochran Q検定)
比率の傾向の検定(Cochran-Armitage検定)

二値変数に対する多変量解析(ロジスティック回帰)

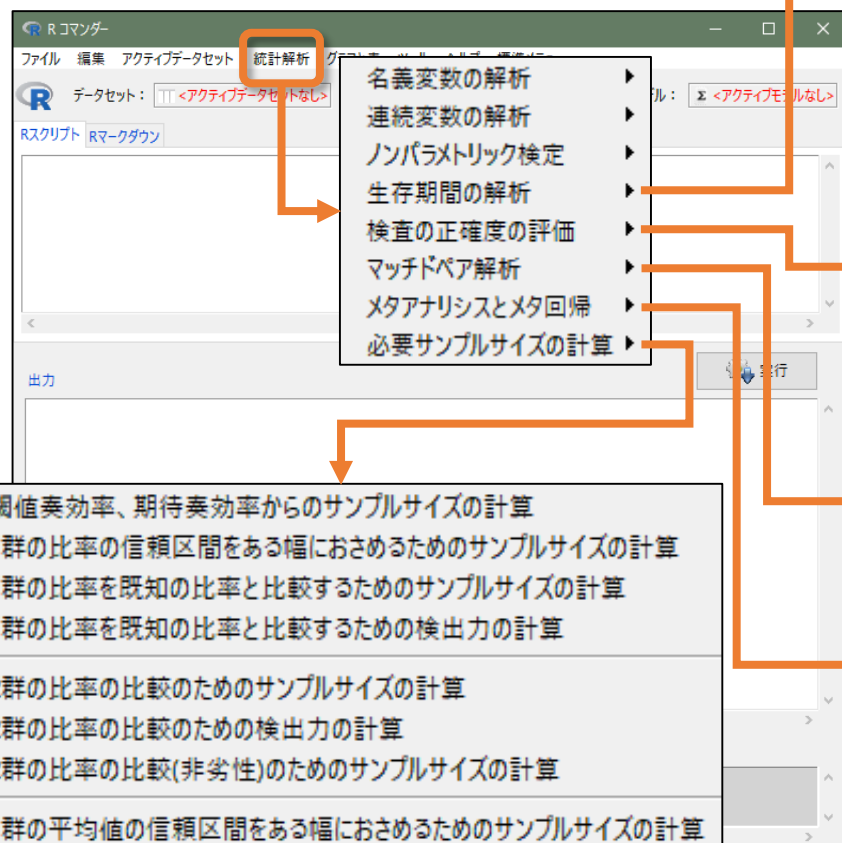
連続変数の要約
外れ値の検定と除外(Smirnov-Grubbs検定)
正規性の検定(Kolmogorov-Smirnov検定)
平均値の信頼区間の計算
1標本の平均値のt検定

2群の等分散性の検定(F検定)
2群間の平均値の比較(t検定)
対応のある2群間の平均値の比較(paired t検定)
3群以上の等分散性の検定(Bartlett検定)
3群以上の間の平均値の比較(一元配置分散分析one-way ANOVA)
対応のある2群以上の間の平均値の比較(反復[経時]測定分散分析)
複数の因子での平均値の比較(多元配置分散分析multi-way ANOVA)
連続変数で補正した2群以上の間の平均値の比較(共分散分析ANCOVA)

相関係数の検定(Pearsonの積率相関係数)
線形回帰(単回帰、重回帰)

2群間の比較(Mann-Whitney U検定)
対応のある2群間の比較(Wilcoxon符号付順位和検定)
3群以上の間の比較(Kruskal-Wallis検定)
対応のある3群以上の間の比較(Friedman検定)
連続変数の傾向の検定(Jonckheere-Terpstra検定)
相関係数の検定(Spearmanの順位相関係数)

EZRにおけるメニュー



閾値奏効率、期待奏効率からのサンプルサイズの計算
 1群の比率の信頼区間をある幅におさめるためのサンプルサイズの計算
 1群の比率を既知の比率と比較するためのサンプルサイズの計算
 1群の比率を既知の比率と比較するための検出力の計算

2群の比率の比較のためのサンプルサイズの計算
 2群の比率の比較のための検出力の計算
 2群の比率の比較(非劣性)のためのサンプルサイズの計算

1群の平均値の信頼区間をある幅におさめるためのサンプルサイズの計算
 2群の平均値の比較のためのサンプルサイズの計算
 2群の平均値の比較のための検出力の計算
 2群の平均の比較(非劣性)のためのサンプルサイズの計算

対応のある2群の平均値の比較のためのサンプルサイズの計算
 対応のある2群の平均値の比較のための検出力の計算

2群の生存曲線の比較のためのサンプルサイズの計算
 2群の生存曲線の比較のための検出力の計算
 2群の生存曲線の比較(非劣性)のためのサンプルサイズの計算

生存曲線の記述と群間の比較(Logrank検定)
 生存期間の傾向の検定(Logrank trend検定)

生存期間に対する多変量解析(Cox比例ハザード回帰)
 時間依存性変数を含む生存期間に対する多変量解析(Cox比例ハザード回帰)

累積発生率(競合イベントを含む)の記述と群間の比較(Gray検定)
 累積発生率(競合イベントを含む)に対する多変量解析(Fine-Gray比例ハザード回帰)

定性検査の診断への正確度の評価
 2つの定性検査の一致度の評価(Kappa係数)
 陽性的中率、陰性的中率の計算

定量検査の診断への正確度の評価(ROC曲線)
 2つのROC曲線のAUCの比較
 質問項目の信頼性の評価(Cronbachの α 信頼性係数)

マッチさせたコントロールの抽出
 マッチさせたサンプルの比率の比較(Mantel-Haenszel検定)
 マッチさせたサンプルの比率の多変量(条件付ロジスティック回帰)
 マッチさせたサンプルの生存率の多変量(層別化比例ハザード回帰)

比率の比較のメタアナリシスとメタ回帰
 平均値の比較のメタアナリシスとメタ回帰
 ハザード比のメタアナリシスとメタ回帰



EZRにおいて実行できる統計解析

名義変数の解析編(2章)

名義変数の解析(2章)

頻度分布
比率の信頼区間の計算
1標本の比率の検定
2群の比率の差の信頼区間の計算
2群の比率の比の信頼区間の計算
分割表の直接入力と解析
分割表の作成と群間の比率の比較(Fisherの正確検定)
対応のある比率の比較(二分割表の対称性の検定、McNemar検定)
対応のある3群以上の比率の比較(Cochran Q検定)
比率の傾向の検定(Cochran-Armitage検定)
二値変数に対する多変量解析(ロジスティック回帰)

頻度分布	DataSet
比率の信頼区間の計算	Direct
1標本の比率の検定	DataSet
2群の比率の差の信頼区間の計算	Direct
2群の比率の比の信頼区間の計算	Direct
分割表の直接入力と解析	Direct
分割表の作成と群間の比率の比較 (Fisherの正確検定)	DataSet
対応のある比率の比較 (二分割表の対称性の検定, McNemar検定)	DataSet
対応のある3群以上の比較(Cochran Q検定)	DataSet
比率の傾向の比較(Cochran-Armitage検定)	DataSet
二値変数に対する多変量解析 (ロジスティック回帰)	DataSet

■ **Direct** : 標本サイズ(サンプル数), 成功回数(イベント数)等を直接入力する.

■ **DataSet** : データ集合を読み込んで解析する.

直接入力(Direct)における様々な統計量の計算



Clinical
Question

大腸内視鏡検査において，NBI検査と白色光検査で1個以上の
鋸歯状病変の検出割合に違いがあるか？

	所見有り	所見無し	計
NBI	204 (51.1%)	195 (48.9%)	399
白色光	158 (39.4%)	243 (60.6%)	401
計	362 (45.2%)	438 (54.2%)	800

群1の総サンプル数(NBIの例数)

399

群1のイベント数(鋸歯状病変例数)

204

群2の総サンプル数(白色光の例数)

401

群2のイベント数(鋸歯状病変例数)

158

直接入力(Direct)における様々な統計量の計算

リスク差の計算方法

頻度分布
比率の信頼区間の計算
1標本の比率の検定
2群の比率の差の信頼区間の計算
2群の比率の比の信頼区間の計算

分割表の直接入力と解析
分割表の作成と群間の比率の比較(Fisherの正確検定)
対応のある比率の比較(二分割表の対称性の検定、McNemar検定)
対応のある3群以上の比率の比較(Cochran Q検定)
比率の傾向の検定(Cochran-Armitage検定)

二値変数に対する多変量解析(ロジスティック回帰)

群1の総サンプル数(NBIの例数)	399
群1のイベント数(鋸歯状病変例数)	204
群2の総サンプル数(白色光の例数)	401
群2のイベント数(鋸歯状病変例数)	158

R 2群の比率の差の信頼区間の計算

群1の総サンプル数	399
群1のイベント数	204
群2の総サンプル数	401
群2のイベント数	158
信頼区間	95

OK キャンセル

> #####2群の比率の差の信頼区間の計算#####

```
> prop.diff.conf(204, 399, 158, 401, 95)
```

```
[1] 比率の差      : 0.117
```

```
[1] 95% 信頼区間      : 0.049 - 0.186
```

NBIと白色光のリスク差(鋸歯状病変の検出割合の差)は、0.117 [0.049, 0.186]である。

EZRでは「群1ー群2」でリスク差が計算される。

直接入力(Direct)における様々な統計量の計算

リスク比

頻度分布
比率の信頼区間の計算
1標本の比率の検定
2群の比率の差の信頼区間の計算
2群の比率の比の信頼区間の計算
分割表の直接入力と解析
分割表の作成と群間の比率の比較(Fisherの正確検定)
対応のある比率の比較(二分割表の対称性の検定、McNemar検定)
対応のある3群以上の比率の比較(Cochran Q検定)
比率の傾向の検定(Cochran-Armitage検定)
二値変数に対する多変量解析(ロジスティック回帰)

群1の総サンプル数(NBIの例数)	399
群1のイベント数(鋸歯状病変例数)	204
群2の総サンプル数(白色光の例数)	401
群2のイベント数(鋸歯状病変例数)	158

R 2群の比率の比の信頼区間の計算

群1の総サンプル数	399
群1のイベント数	204
群2の総サンプル数	401
群2のイベント数	158
信頼区間	95

OK キャンセル

> #####2群の比率の比の信頼区間の計算#####

```
> prop.ratio.conf(204, 399, 158, 401, 95)
```

```
[1] 比率の比      : 1.298
```

```
[1] 95% 信頼区間      : 1.112 - 1.515
```

NBIと白色光のリスク比(NBIの検出割合／白色光の検出割合)は、1.298[1.112, 1.515]である。つまり、NBIは白色光に比べて1.298倍ほど鋸歯状病変の検出能がある。

EZRでは「群1/群2」でリスク比が計算される。

直接入力(Direct)における様々な統計量の計算

オッズ比

頻度分布

比率の信頼区間の計算

1標本の比率の検定

2群の比率の差の信頼区間の計算

2群の比率の比の信頼区間の計算

分割表の直接入力と解析

分割表の作成と群間の比率の比較(Fisherの正確検定)

対応のある比率の比較(二分割表の対称性の検定、McNemar検定)

対応のある3群以上の比率の比較(Cochran Q検定)

比率の傾向の検定(Cochran-Armitage検定)

二値変数に対する多変量解析(ロジスティック回帰)

次のクロス集計表を入力する。

	所見有り	所見無し
NBI	204	195
白色光	158	243

オッズ比を計算するにはチェックが必要

分割表の直接入力と解析

行数: 2
列数: 2

数を入力:

	1	2
1	204	195
2	158	243

パーセントの計算

☐ 行のパーセント
☐ 列のパーセント
☐ 総計のパーセント
☒ パーセント表示無し

仮説検定

☐ 独立性のカイ2乗検定(連続補正有り)
☐ カイ2乗統計量の要素
☐ 期待度数の表示
☒ フィッシャーの正確検定

ヘルプ OK キャンセル

クロス集計表のセルの数を入力
行数: 縦方向
列数: 横方向

パーセントの計算
行: 横方向で100%になるように計算
列: 縦方向で100%になるように計算
総計: 全体で100%になるように計算

直接入力(Direct)における様々な統計量の計算

EZRではオッズ比を計算するためのメニューが用意されていない。一方で、RのFisherの正確検定のアウトプットには、オッズ比及び95%信頼区間のアウトプットが表示されるため、それを用いる。

```
> #####分割表の直接入力と解析#####
```

```
> .Table <- NULL
```

```
> .Table <- matrix(c(204, 195, 158, 243), 2, 2, byrow=TRUE)
```

(中略)

```
> fisher.test(.Table)
```

Fisher's Exact Test for Count Data

data: .Table

p-value = 0.001072

alternative hypothesis: true odds ratio is not equal to 1

95 percent confidence interval:

1.203573 2.151081

sample estimates:

odds ratio

1.607948

オッズ比及び95%信頼区間を表している

(下略)

NBIと白色光のオッズ比(NBIのオッズ／白色光のオッズ)は、1.608[1.203, 2.151]である。つまり、NBIは白色光に比べて1.608倍ほど鋸歯状病変の検出能がある。

EZRでは「群1/群2」でオッズ比が計算される。

リスク差とリスク比の違い



リスク比では1.298倍，オッズ比では1.608倍ほどNBIは白色光に比べて鋸歯状病変の検出能がある。どちらが正しい結果なの？

コホート研究ではリスク比，ケースコントロール研究ではオッズ比を用いると言われているが，近年では，オッズ比で解析することも増えてきている

■ リスク比の欠点(1)：ロジスティック回帰と対応しない

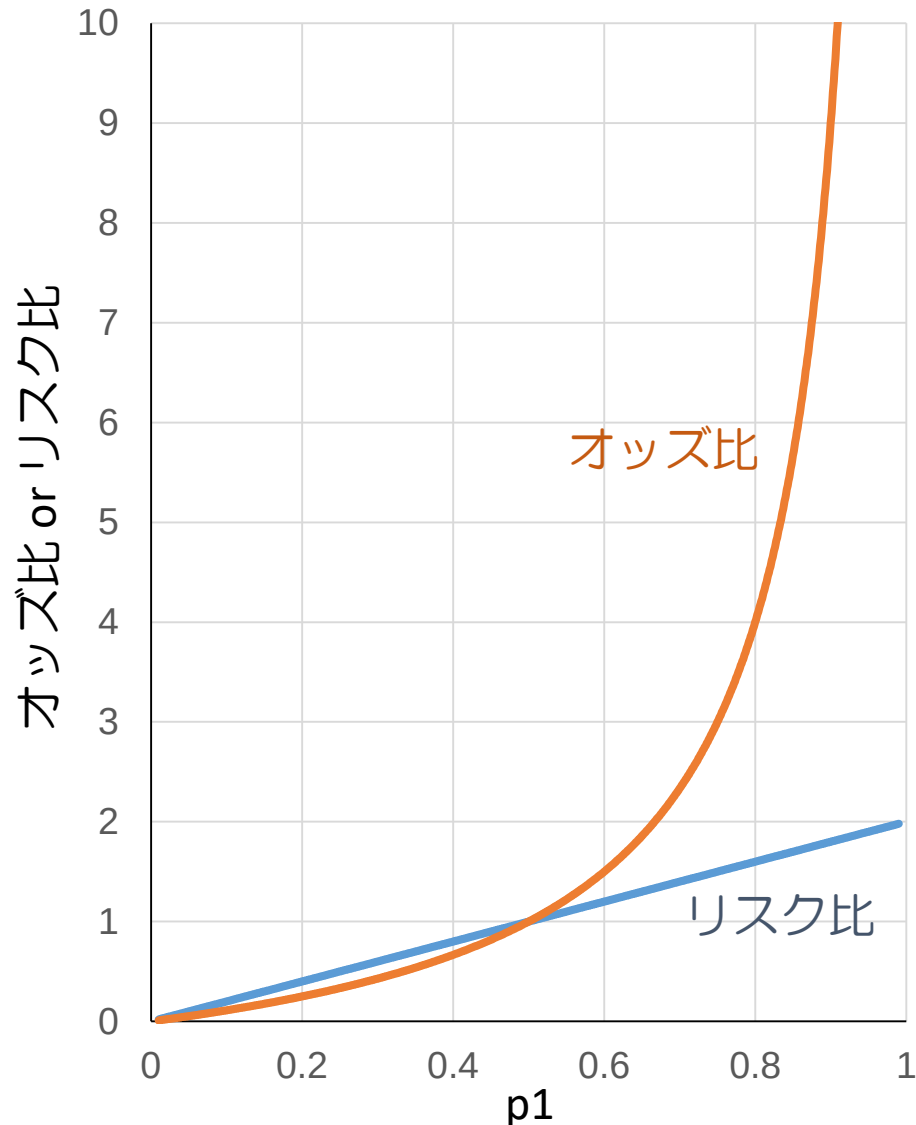
■ リスク比の欠点(2)：上限が分母のリスクで決まる(おまけ1 参照).

リスク比の場合には，0倍～有限倍(分母のリスクで上限が決まる)までしか定義域がないのに対して，オッズ比の場合には，0倍～ ∞ 倍まで定義可能である。

■ リスク比の欠点(3)：ラベル付けの問題(おまけ2参照)

関心のあるイベントではなく，関心のあるイベントの非発現を考えると，「〇〇倍ほど皮膚かぶれになる」といっていたものが「 $1/〇〇$ 倍ほど関心のあるイベントが起きない」と逆数での解釈になるはずだが，リスク比ではこのような数値にはならない。

おまけ1：リスク比の天井効果



p_2 (例：白色光)を0.5(オッズ=1.0)に固定したときの p_1 (例：NBI)に対するオッズ比orリスク比の推移

$p_1=0$ のとき

$$\text{リスク比} = \frac{0}{0.5} = 0$$

$$\text{オッズ比} = \frac{0 / (1 - 0)}{1} = 0$$

$p_1=1$ のとき

$$\text{リスク比} = \frac{1}{0.5} = 2$$

$$\text{オッズ比} = \frac{1 / (1 - 1)}{1} = \infty$$

つまり、リスク比では2倍までしか表すことができないのに対して、オッズ比では ∞ まで表すことができる。

おまけ2：リスク比の逆数の問題

	所見有り		所見無し		計
NBI	204 (51.1%)	Odds=1.045	195 (48.9%)	Odds=0.957	399
白色光	158 (39.4%)	Odds=0.650	243 (60.6%)	Odds=1.538	401
計	362 (45.2%)		438 (54.2%)		800

例：NBIで所見有りのオッズ=0.511/(1-0.511)=1.045

■ リスク比

所見有り： $0.511 / 0.394 = 1.297$ (丸め誤差のためEZRと少し異なる)

所見なし： $0.489 / 0.606 = 0.807$

NBIは白色光に比べて0.807倍所見無しになるので、その逆数 $1/0.807$ が所見有りのリスク比になるはずなのに、その結果は、1.239であるが、**そうはならない**。

■ オッズ比

所見有り： $1.045 / 0.650 = 1.608$

所見なし： $0.957 / 1.538 = 0.622$

NBIは白色光に比べて0.622倍所見無しになるので、その逆数 $1/0.622$ が所見有りのオッズ比になるはずなのに、その結果は、1.608であり、**一致する**。

■ データを読み込むことでできるメニューと統計的方法

頻度分布

比率の信頼区間の計算

1標本の比率の検定

→ カイ2乗検定(母比率の検定), 正確検定

2群の比率の差の信頼区間の計算

2群の比率の比の信頼区間の計算

分割表の直接入力と解析

分割表の作成と群間の比率の比較
(Fisherの正確検定)

→ カイ2乗検定(母比率の差の検定)
Fisherの正確検定

対応のある比率の比較
(二分割表の対称性の検定, McNemar検定)

→ McNemar検定

対応のある3群以上の比較(Cochran Q検定)

→ Cochran Q検定

比率の傾向の比較(Cochran-Armitage検定)

→ Cochran-Armitage検定

二値変数に対する多変量解析
(ロジスティック回帰)

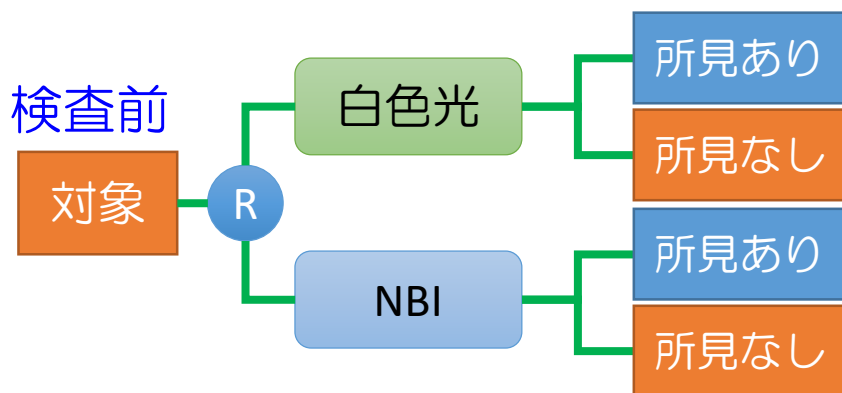
→ ロジスティック回帰分析 (変数選択を含む)

EZRの内容と統計的方法の関係(1/5)



Clinical Question

大腸内視鏡検査において、NBI検査を実施した被験者群と白色光検査を実施した被験者群で無作為化比較試験を実施し、1個以上の鋸歯状病変の検出割合を比較した。



	所見有り	所見無し	計
NBI	204 (51.1%)	195 (48.9%)	399
白色光	158 (39.4%)	243 (60.6%)	401
計	362 (45.2%)	438 (54.2%)	800

統計的方法

検定：カイ2乗検定(標本サイズ多数)
Fisherの正確検定(標本サイズ少数)

要約：各群の所見有りの割合
オッズ比及び95%信頼区間

EZRでの解析手順(2.2.5節)

1. 「分割表の作成と群間の比率の比較(Fisherの正確検定)」を選択する。
2. 「行の選択」で群, 「列の選択」でアウトカムを選択する。
3. 「パーセントの計算」で「行」を選択する(各群のアウトカムの割合が計算できる)。
4. 「仮説検定」で「カイ2乗検定」and/or 「フィッシャーの正確検定」を選択する。

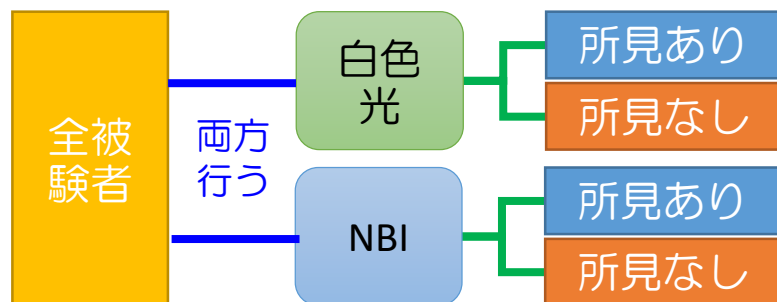
Fisherの正確検定を選択しないとオッズ比及び95%信頼区間が計算できないので注意。

EZRの内容と統計的方法の関係(2/5)



Clinical Question

大腸内視鏡検査において、各被験者に対してNBI検査と白色光検査を実施し、1個以上の鋸歯状病変の検出割合を比較した。



		NBI		合計
		あり	なし	
白色光	あり	85	10	95
	なし	18	61	79
合計		103	71	174

統計的方法

検定：McNemar検定

要約：対応のあるクロス集計表
全体パーセント

EZRでの解析手順(2.4.1節)

- 「対応のある比率の比較(二分割表の対称性の検定、McNemar検定)」を選択して解析すると、**対応のあるクロス集計表**及び**McNemar検定**の結果が表示される。
- 「分割表の直接入力と解析」を選択し「パーセントの計算」で「総計のパーセント」を選択する(**全体パーセント**)。

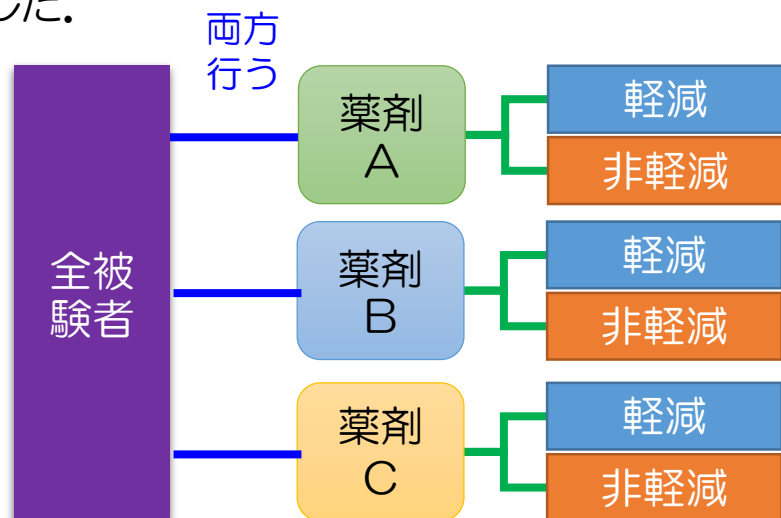
全体パーセントの計算において、検定結果は意味がないのですべてチェックを外すことが推奨される。

EZRの内容と統計的方法の関係(3/5)



Clinical Question

神経障害性疼痛患者に対して、3種類の除痛薬(A,B,C)を投与したときの痛みの軽減の有無を調べるためのチェンジオーバー試験を実施した。



		B	
		あり	なし
A	あり	5	1
	なし	4	2

		C	
		あり	なし
A	あり	2	4
	なし	1	5

		C	
		あり	なし
B	あり	3	6
	なし	0	3

統計的方法

検定：CochranのQ検定，McNemar検定の多重比較

要約：対応のあるクロス集計表
全体パーセント

EZRでの解析手順(2.4.2節)

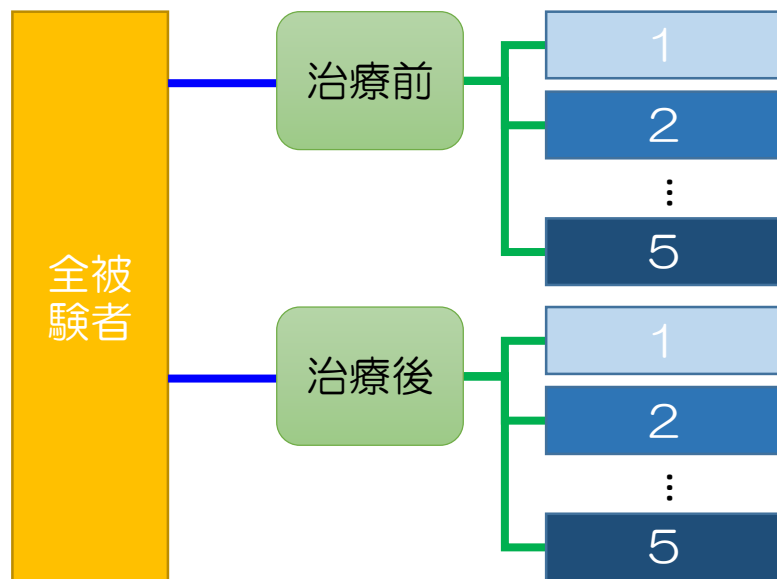
- 「対応のある3群以上の比較(Cochran Q検定)」を選択して解析すると、CochranのQ検定の結果が表示される。
- 「対応のある比率の比較(二分割表の対称性の検定、McNemar検定)」を選択して解析すると、対応のあるクロス集計表及びMcNemar検定の結果が表示される(これをペアワイズに行う)。その後、p値×比較回数を計算するとBonferroniの多重比較の結果になる。

EZRの内容と統計的方法の関係(4/5)



Clinical Question

あるペインクリニックにおいて、新たな治療法の効果が検討された。結果は、治療前後での5段階評価での評点によって実施された。



		治療後					計
		5	4	3	2	1	
治療前	5	8	3	1	0	1	13
	4	9	10	8	1	1	29
	3	3	16	24	3	1	47
	2	2	5	10	9	2	28
	1	0	3	6	9	5	23
計		22	37	49	22	10	140

統計的方法

検定：McNemar検定

要約：対応のあるクロス集計表
全体パーセント

EZRでの解析手順(2.4.1節)

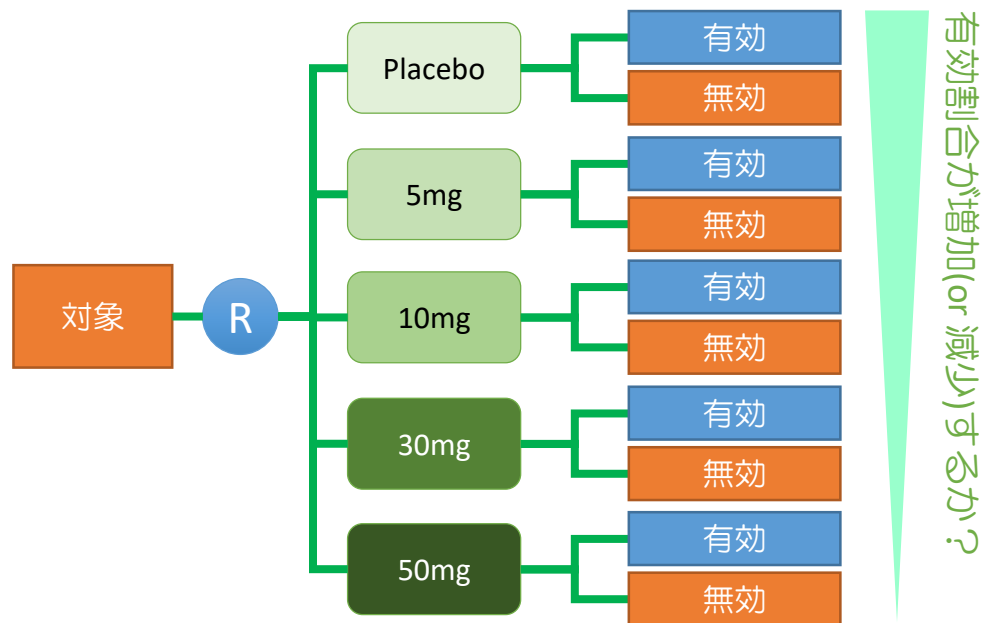
- 「対応のある比率の比較(二分割表の対称性の検定、McNemar検定)」を選択して解析すると、[対応のあるクロス集計表](#)及び[McNemar検定](#)の結果が表示される。
- 「分割表の直接入力と解析」を選択し「パーセントの計算」で「総計のパーセント」を選択する([全体パーセント](#))。

EZRの内容と統計的方法の関係(5/5)

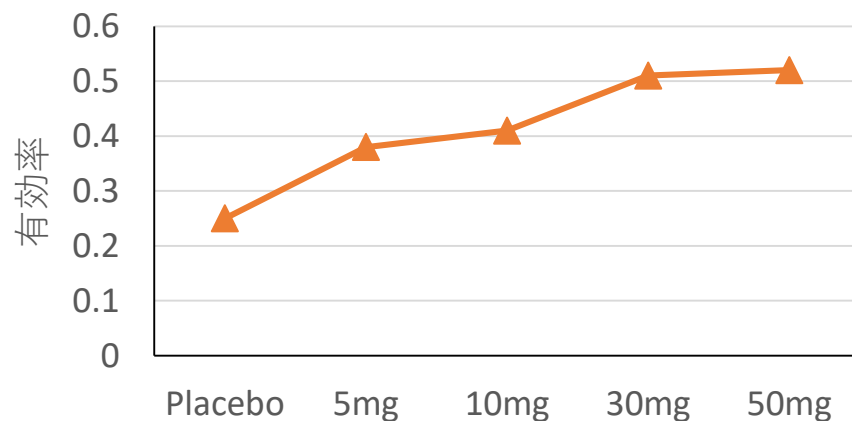


Clinical Question

ある薬剤の用量反応関係を調査するために、Placebo, 5mg, 10mg, 30mg, 50mgの投与群とそのときの効果の有無を調べた。



	Placebo	5mg	10mg	30mg	50mg
有効	79	81	169	318	379
無効	235	132	243	305	204



統計的方法

検定：Cochran-Armitage検定

要約：有効割合

EZRでの解析手順(2.2.2節)

1. 「比率の傾向の比較(Cochran-Armitage検定)」を選択して解析すると、Cochran-Armitage検定の結果が表示される。
2. 「分割表の作成と群間の比率の比較(Fisherの正確検定)」を選択する。
3. 「行の選択」で用量、「列の選択」で有効の有無を選択する。
4. 「パーセントの計算」で「行のパーセント」を選択する(有効割合)。

EZRによるロジスティック回帰分析の留意点

二値変数に対する多変量解析(ロジスティック回帰)

モデル名を入力: GLM.1

変数 (ダブルクリックして式に入れる)

応答
変数1
変数2
変数3

モデル式: + * : / %in% - ^ ()

目的変数 応答 ~ 説明変数 変数1 + 変数2 + 変数3

層別化因子は + strata(####)と入力する ↑ ↑ ↑

☐ 3レベル以上の因子についてその因子全体のP値の計算(Wald検定)

☐ モデル解析用に解析結果をアクティブモデルとして残す

☐ ROC曲線を表示する

☐ 基本的診断プロットを表示する

☐ 傾向スコア変数を自動作成する

☐ AICを用いたステップワイズの変数選択を行う。

☐ BICを用いたステップワイズの変数選択を行う。

☐ P値を用いたステップワイズの変数選択(減少法)を行う。

↓一部のサンプルだけを解析対象にする場合の条件式。例: age>50 & Sex==0 や age<50 | Sex==1

<全ての有効なケース>

ヘルプ リセット OK キャンセル 適用

変数間に「+」を入れる

交互作用を検討す場合には、「*」にする。
例(変数1と変数2の交互作用): 変数1*変数2

変数増減法によるステップワイズ法

AIC(Akaike's Information Criteria)とBIC(Bayesian Information Criteria)では、BICのほうが選択される変数の数が少なくなる傾向にある。

3水準以上の共変量に対する評価

- ☐ 3レベル以上の因子についてその因子全体のP値の計算(Wald検定)
- ☐ モデル解析用に解析結果をアクティブモデルとして残す
- ☐ ROC曲線を表示する
- ☐ 基本的診断プロットを表示する
- ☐ 傾向スコア変数を自動作成する
- ☐ AICを用いたステップワイズの変数選択を行う。
- ☐ BICを用いたステップワイズの変数選択を行う。
- ☐ P値を用いたステップワイズの変数選択(減少法)を行う。

共変量が3水準(A, B, C)が存在する場合, ロジスティック回帰モデルでは, 2個の変数で表現する

変数B 共変量がBならば1, それ以外は0

変数C 共変量がCならば1, それ以外は0

したがって, ロジスティック回帰のモデルは

その他の共変量 + $\beta_B \times$ 変数B + $\beta_C \times$ 変数C

水準Aの場合	0	0
水準Bの場合	1	0
水準Cの場合	0	1

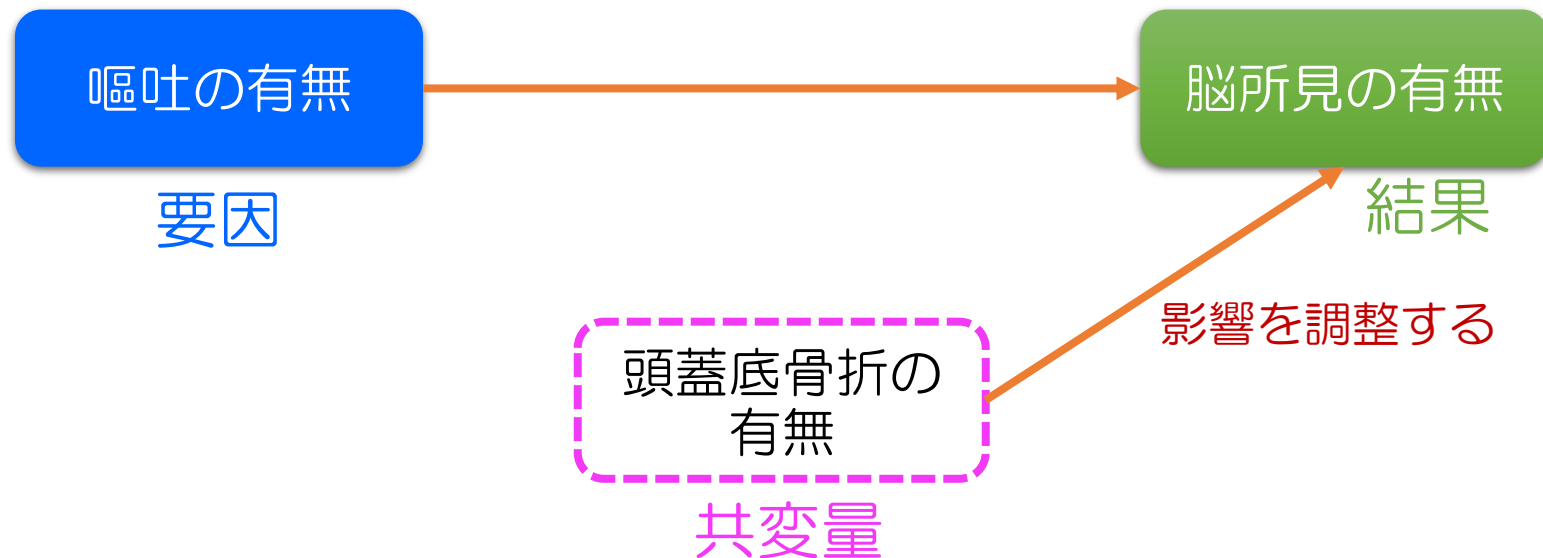
Wald検定での仮説は

帰無仮説 H_0 : この共変量は不要である(変数B, Cの回帰パラメータはともに0である)

対立仮説 H_0 : この共変量は不要である(変数B, Cの回帰パラメータはいずれか or どちらも0でない)

Wald検定とは, 変数B, Cの回帰係数がともに0であるという帰無仮説のもとで, 帰無仮説からの回帰係数の隔たりを測っている.

2.5.2節の頭部外傷データを用いる。このデータは、カナダの3121名の軽度外部外傷患者に対するCTによる脳所見(有:1, 無:2)に対する影響要因を探索している。ここでは、嘔吐の有無が脳所見と関連するか否かを評価する。このとき、頭蓋底骨折の有無が影響を及ぼすことが想定されるため、調整を行う。



1. 無作為化比較試験の主解析において、割付調整因子を共変量にしたMantel-Haentzel検定を用いることが多い。
2. Mantel-Haentzel検定は共変量と要因を説明変数としたロジスティック回帰分析を実施することと同じである。

EZRによる解析

グループ： vomiting：

アウトカム： clinically.important.brain.injury：

共変量： basal.skull.fracture：

嘔吐(無(0), 有(1))

CTによる脳所見(無(0), 有(1))

頭蓋底骨折(無(0)/有(1))

- 名義変数の解析
- 連続変数の解析
- ノンパラメトリック検定
- 生存期間の解析
- 検査の正確度の評価
- マッチドペア解析
- メタアナリシスとメタ回帰
- 必要サンプルサイズの計算

マッチさせたコントロールの抽出

マッチさせたサンプルの比率の比較(Mantel-Haenzel検定)

マッチさせたサンプルの比率の多変量(条件付ロジスティック回帰)

マッチさせたサンプルの生存率の多変量(層別化比例ハザード回帰)

マッチさせたサンプルの比率の比較(Mantel-Haenzel検定)

複数の選択はCtrlキーを押しながらクリック。↓

比較する群の変数(コントロール=0、ケース=1) (1つ選択)

比率を比較する変数(1つ以上選択)

clinically.important.brain.injury
GCS.13
GCS.15.2hours
GCS.decrease
hi
loss.of.consciousness
open.skull.fracture
vomiting
vomiting

age.65
amnesia.before
basal.skull.fracture
clinically.important.brain.injury
GCS.13
GCS.15.2hours
GCS.decrease
high.risk

マッチさせた層を示す変数(1つ選択、通常はpairmatch)

age.65
amnesia.before
basal.skull.fracture
clinically.important.brain.injury
GCS.13
GCS.15.2hours
GCS.decrease
high.risk

カイ2乗検定の連続性補正

☒ Yes
☐ No

ヘルプ リセット OK キャンセル 適用

EZRによる解析

Rパートの結果

```
Mantel-Haenszel chi-squared test with continuity correction
data: Dataset$vomiting and Dataset$clinically.important.brain.injury and
Dataset$basal.skull.fracture
Mantel-Haenszel X-squared = 77.034, df = 1, p-value < 2.2e-16
alternative hypothesis: true common odds ratio is not equal to 1
95 percent confidence interval:
 2.909690 5.590633
sample estimates:
common odds ratio
 4.033238
```

頭蓋底骨折の有無を調整した共通オッズ比及び95%信頼区間
を表している

EZRパートの結果

vomiting=0 vomiting=1 MH. p. value		
clinically.important.brain.injury=0	2631	240
clinically.important.brain.injury=1	182	68

1.68e-18

p 値

特殊な状況でのロジスティック回帰分析

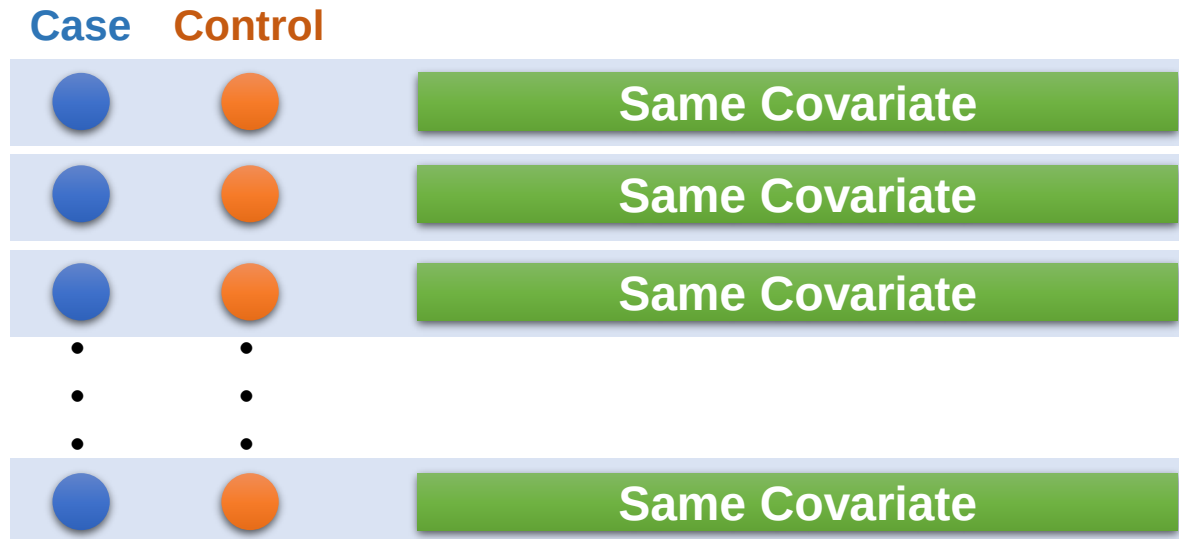
テキストには非掲載

■ 条件付ロジスティック回帰

EZR: 「マッチドペア解析」 → 「マッチさせたサンプルの比率の多変量(条件付ロジスティック回帰)」

- 皮膚科等での無作為化ハーフサイド試験
- 1対1マッチングされたケースコントロール研究

では、共変量が同じで、アウトカムに対応がある。



このような場合に利用できるロジスティック回帰モデルが条件付きロジスティック回帰分析である。



EZRにおいて実行できる統計解析

連続変数の解析編(1章)

統計的方法の取捨選択

単群(単アーム)研究 or 比較研究

単群試験

1標本t検定

比較研究

介入は 同一被験者 or 異なる被験者

同一被験者

異なる被験者

正規性が 仮定できる or できない

正規性が 仮定できる or できない

できる

できない

できる

できない

対応のあるt検定

Wilcoxon符号付
順位検定

等分散性が 仮定で
きる or できない

Wilcoxon検定
(Mann-Whitney検定)

できる

できない

2標本t検定

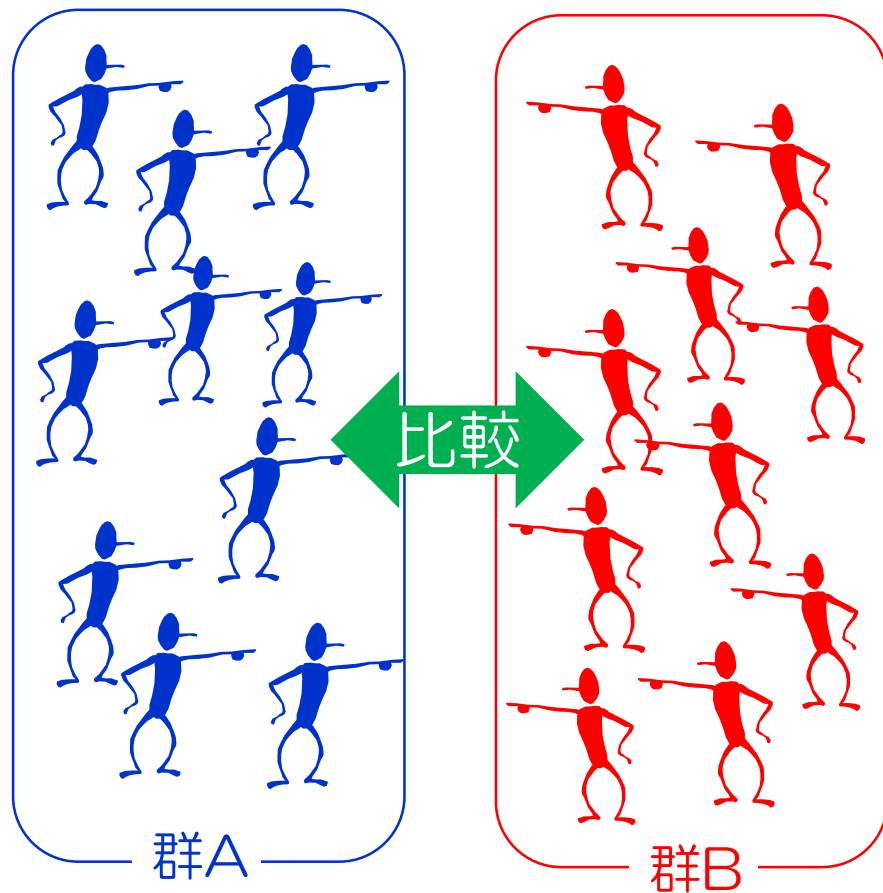
Welch検定

対応のあるデータと対応のないデータ

対応のないデータ(独立2標本)

■ 例：無作為化比較試験

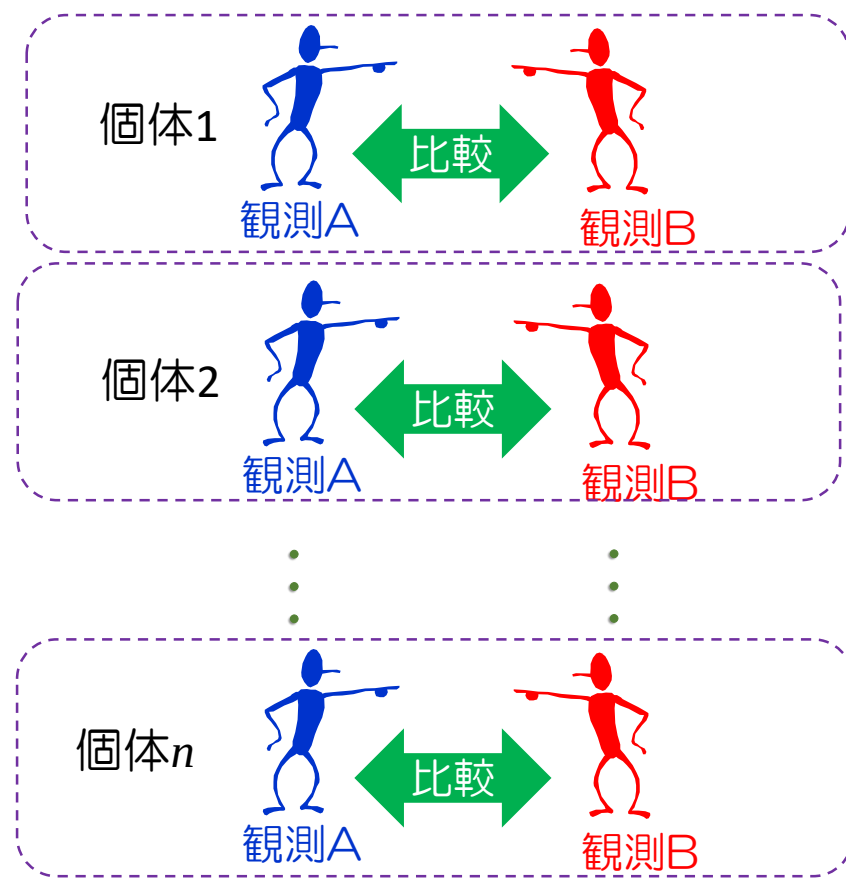
治療Aと治療Bを異なる被験者に実施して有効性を比較する。



対応のある標本

■ 例：クロスオーバー試験

同一被験者に対して、治療Aと治療Bに実施する。



パラメトリック検定とノンパラメトリック検定

臨床試験のデザイン

平均値に基づいて試験デザイン(症例設計)を行うことが多い。そのため、パラメトリック検定を用いて評価することが原則になる。

観察研究 or 臨床試験の主解析以外の評価

アウトカムが著しく正規分布から外れた場合にはノンパラメトリック検定の選択が考えられる。ノンパラメトリック検定を用いる条件には、

- (1) 外れ値が存在する場合
 - (2) アウトカムが著しく歪んでおり正規分布に従わない場合
- が考えられる。

可能な限りパラメトリック検定を用いる理由

- 多変量解析において、重回帰分析を用いることが多いが、アウトカムが正規分布に従うことを仮定しているため(ノンパラメトリック検定でアウトカムを比較したあとで重回帰分析を用いるのは理論的に整合性がとれない。
- 2標本t検定のほうがWilcoxon検定に比べて検出力(群間に違いがあるときに正しく違いがあるを示すことができる確率)が一般的に高い。
- パラメトリック検定では、結果が有意なときに「平均値に差がある」といえるがノンパラメトリック検定の場合には「相対的に違いがある」としかいえず、治療効果などを明示的に表すことができない。

連続変数の解析における留意点：対応があるデータの場合

- 名義変数の解析
- 連続変数の解析
- ノンパラメトリック検定
- 生存期間の解析
- 検査の正確度の評価
- マッチドペア解析
- メタアナリシスとメタ回帰
- 必要サンプルサイズの計算

パラメトリック

- 連続変数の要約
- 外れ値の検定と除外(Smirnov-Grubbs検定)
- 正規性の検定(Kolmogorov-Smirnov検定)
- 平均値の信頼区間の計算
- 1標本のt検定
- 2群の等分散性の検定(F検定)
- 2群間の平均値の比較(t検定)
- 対応のある2群間の平均値の比較(paired t検定)**
- 3群以上の等分散性の検定(Bartlett検定)
- 3群以上の間の平均値の比較(一元配置分散分析one-way ANOVA)
- 対応のある2群以上の間の平均値の比較(反復[経時]測定分散分析)
- 複数の因子での平均値の比較(多元配置分散分析multi-way ANOVA)
- 連続変数で補正した2群以上の間の平均値の比較(共分散分析ANCOVA)
- 相関係数の検定(Pearsonの積率相関係数)
- 線形回帰(単回帰、重回帰)

対応のあるt検定

対応のある2群間の平均値の比較(paired t検定)

第1の変数 (1つ選択) 第2の変数 (1つ選択)

対立仮説

☒ 両側

☐ 差 < 0

☐ 差 > 0

信頼水準

0.95

↓一部のサンプルだけを解析対象にする場合の条件式。例: age > 50 & Sex == 0 や age < 50 | Sex == 1

<全ての有効なケース>

ヘルプ リセット OK キャンセル 適用

「差は(第1の変数)-(第2の変数)」である

ノンパラメトリック

- 2群間の比較(Mann-Whitney U検定)
- 対応のある2群間の比較(Wilcoxon符号付順位和検定)**
- 3群以上の間の平均値の比較(非パラメトリック検定)
- 対応
- 連続変数の平均値の検定(非パラメトリック検定)
- 相関係数の検定(Spearmanの順位相関係数)

Wilcoxon符号付順位和検定

対応のある2群間の比較(Wilcoxon符号付順位和検定)

第1の変数 (1つ選択) 第2の変数 (1つ選択)

対立仮説

☒ 両側

☐ 差 < 0

☐ 差 > 0

検定のタイプ

☒ デフォルト

☐ 正確

☐ 正規近似

☐ 連続修正を用いた正規近似

通常は「正確」を選択したほうがよい

↓一部のサンプルだけを解析対象にする場合の条件式。例: age > 50 & Sex == 0 や age < 50 | Sex == 1

<全ての有効なケース>

ヘルプ リセット OK キャンセル 適用

連続変数の解析における留意点： 独立2標本の場合

- 名義変数の解析
- 連続変数の解析
- ノンパラメトリック検定
- 生存期間の解析
- 検査の正確度の評価
- マッチドペア解析
- メタアナリシスとメタ回帰
- 必要サンプルサイズの計算

パラメトリック

- 連続変数の要約
- 外れ値の検定と除外(Smirnov-Grubbs検定)
- 正規性の検定(Kolmogorov-Smirnov検定)
- 平均値の信頼区間の計算
- 1標本の
- 2標本の平均値の比較(t検定)**
- 2群の等分散性の検定(F検定)
- 2群間の平均値の比較(t検定)**
- 対応のある2群間の平均値の比較(paired t検定)
- 3群以上の等分散性の検定(Bartlett検定)
- 3群以上の間の平均値の比較(一元配置分散分析one-way ANOVA)
- 対応のある2群以上の間の平均値の比較(反復[経時]測定分散分析)
- 複数の因子での平均値の比較(多元配置分散分析multi-way ANOVA)
- 連続変数で補正した2群以上の間の平均値の比較(共分散分析ANCOVA)
- 相関係数の検定(Pearsonの積率相関係数)
- 線形回帰(単回帰、重回帰)

ノンパラメトリック

- 2群間の比較(Mann-Whitney U検定)**
- 対応のある2群間の比較(Wilcoxon符号付順位和検定)
- 2群間の比較(Mann-Whitney検定)**
- 連続変数の傾向の検定(Jonckheere-Terpstra検定)
- 相関係数の検定(Spearmanの順位相関係数)

Wilcoxon検定とMann-Whitney
検定は同じもの

2群間の平均値の比較(t検定)

複数の選択はCtrlキーを押しながらクリック。↓
目的変数(1つ選択) 比較する群(1つ以上選択、ただし2種類の値だけを持つこと)

等分散と考えますか？
はい：2標本 t 検定
いいえ：Welch検定

差： <層別因子が選択されていません>

対立仮説 信頼水準 等分散と考えますか？ グラフ

☒ 両側 ☐ 片側 ☒ はい(t検定) ☐ いいえ(Welch検定) ☐ 箱ひげ ☐ 棒 ☐ 折れ線

☐ 差 < 0 ☐ 差 > 0

↓一部のサンプルだけを解析対象にする場合の条件式。例: age>50 & Sex==0 や age<50 | Sex==1

<全ての有効なケース>

ヘルプ リセット OK キャンセル 適用

2群間の比較(Mann-Whitney U検定)

目的変数(1つ選択) 比較する群(1つ以上選択、ただし2種類の値だけを持つこと)

対立仮説 検定のタイプ

☒ 両側 ☐ 片側 ☒ デフォルト ☐ 正確 ☐ 正規近似 ☐ 連続修正を用いた正規近似

☐ 差 < 0 ☐ 差 > 0

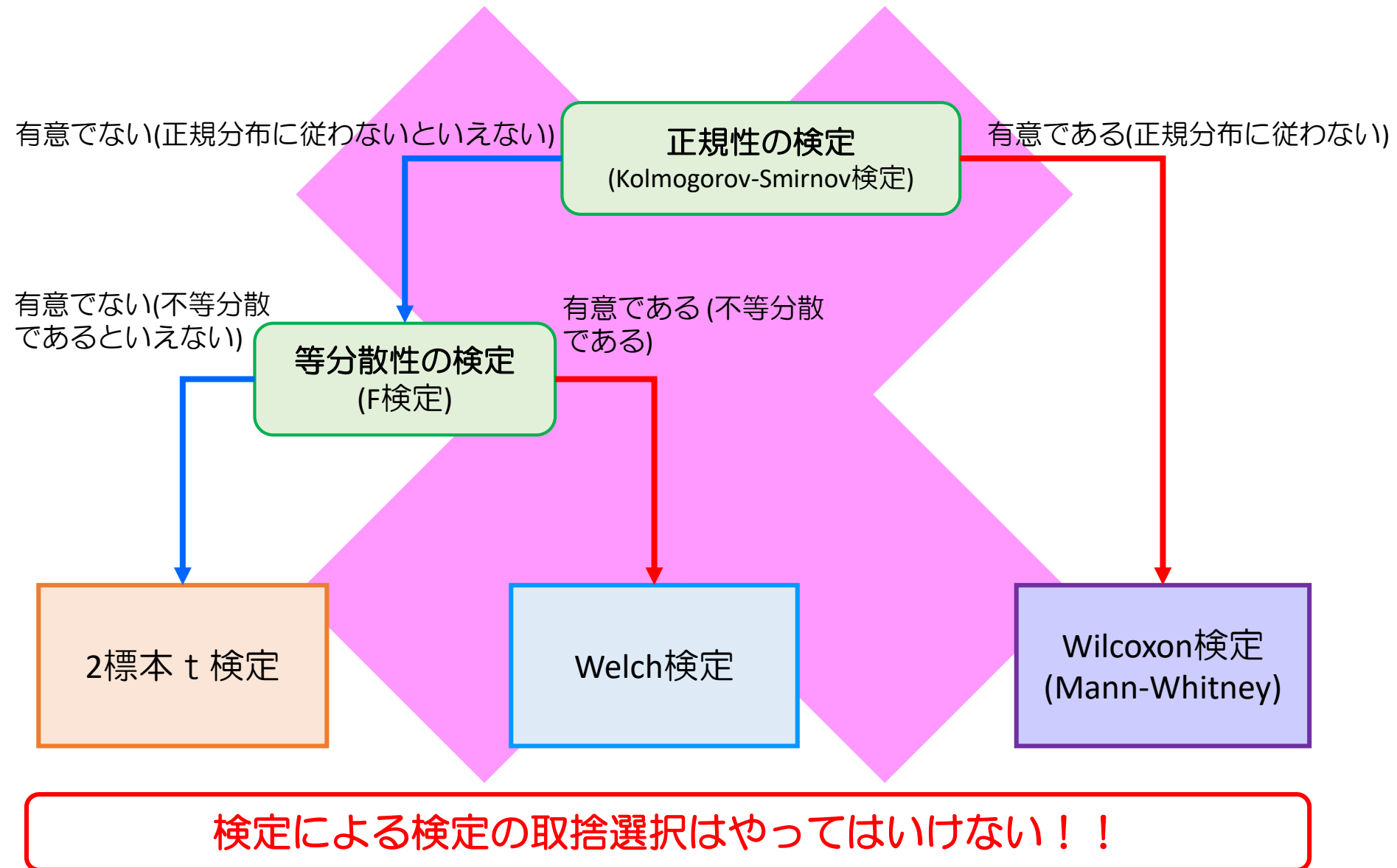
通常は「正確」を選択したほうがよい

↓一部のサンプルだけを解析対象にする場合の条件式。例: age>50 & Sex==0 や age<50 | Sex==1

<全ての有効なケース>

ヘルプ リセット OK キャンセル 適用

連続変数の解析における留意点：独立2標本の場合



連続変数の解析における留意点：独立2標本の場合

データ解析における手順：私案

- ・ヒストグラムが左右対称に布置している
- ・正規プロットが直線におおよそ布置している

正規性のグラフによる点検

ヒストグラム・正規プロット
(正規プロットはEZRではQQプロット)

- ・ヒストグラムが著しく歪んでいる
- ・正規プロットが著しく直線に布置しない

分散(標準偏差)が
大きく異なる

記述統計(分散ある
いは標準偏差)

分散(標準偏差)が
著しく異なる

2標本 t 検定

Wilcoxon検定
(Mann-Whitney)

Welch検定は使わない

Wilcoxon検定
(Mann-Whitney)

■ ヒストグラム：「グラフと表」→「ヒストグラム」

■ 正規プロット：「グラフと表」→「QQプロット」

分散分析と多重比較

パラメトリック

- 名義変数の解析 ▶
- 連続変数の解析 ▶
- ノンパラメトリック検定 ▶
- 生存期間の解析 ▶
- 検査の正確度の評価 ▶
- マッチドペア解析 ▶
- メタアナリシスとメタ回帰 ▶
- 必要サンプルサイズの計算 ▶

連続変数の要約 外れ値の検定と除外(Smirnov-Grubbs検定) 正規性の検定(Kolmogorov-Smirnov検定) 平均値の信頼区間の計算 1標本の平均値のt検定
2群の等分散性の検定(F検定) 2群間の平均値の比較(t検定) 対応のある2群間の平均値の比較(paired t検定) 3群以上の等分散性の検定(Bartlett検定) 3群以上の間の平均値の比較(一元配置分散分析one-way ANOVA) 対応のある2群以上の間の平均値の比較(反復[経時]測定分散分析) 複数の因子での平均値の比較(多元配置分散分析multi-way ANOVA) 連続変数で補正した2群以上の間の平均値の比較(共分散分析ANCOVA)
相関係数の検定(Pearsonの積率相関係数) 線形回帰(単回帰、重回帰)

一元配置の分散分析

繰り返し測定の分散分析

多元配置の分散分析

ノンパラメトリック

2群間の比較(Mann-Whitney U検定) 対応のある2群間の比較(Wilcoxon符号付順位和検定) 3群以上の間の比較(Kruskal-Wallis検定) 対応のある3群以上の間の比較(Friedman検定) 連続変数の傾向の検定(Jonckheere-Terpstra検定) 相関係数の検定(Spearmanの順位相関係数)

Kruskal-Wallis検定

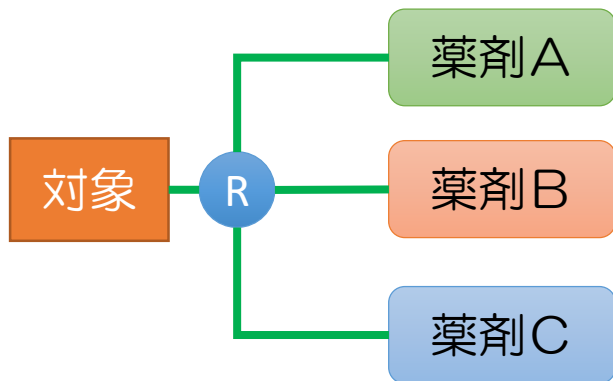
Friedman検定

分散分析の種類(1/4)：多標本データの比較



Clinical Question

神経障害性疼痛患者がランダムに割り付けられ、3種類の除痛薬(A,B,C)のいずれかが投与された。



Data structure

P：被験者

薬剤A

P P P P P P P P P P P

薬剤B

P P P P P P P P P P

薬剤C

P P P P P P P P P P P

3薬のVAS減少量に違いがあるか？

■ パラメトリック検定：一元配置の分散分析

EZR：3群以上の間の平均値の比較(一元配置分散分析one-way ANOVA)

■ ノンパラメトリック検定：Kruskal-Wallis検定

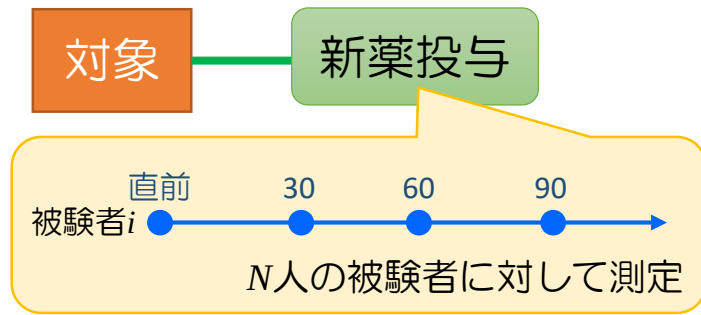
EZR：3群以上の間の比較(Kruskal-Wallis検定)

分散分析の種類(2)：繰り返し測定データの解析(1)



Clinical Question

N 名の神経障害性疼痛患者に対して、新薬の除痛薬を投与し、投与直前、30分後、60分後、90分後のVASの変化を調査した。



Data structure

直前	30分後	60分後	90分後
	Patient.1		
	Patient.2		
	Patient.3		
⋮	⋮	⋮	⋮
	Patient. N		

■ パラメトリック検定：繰り返し測定の分散分析

EZR：対応のある2群以上の間の平均値の比較(反復[経時]測定分散分析)

■ ノンパラメトリック検定：Friedman検定

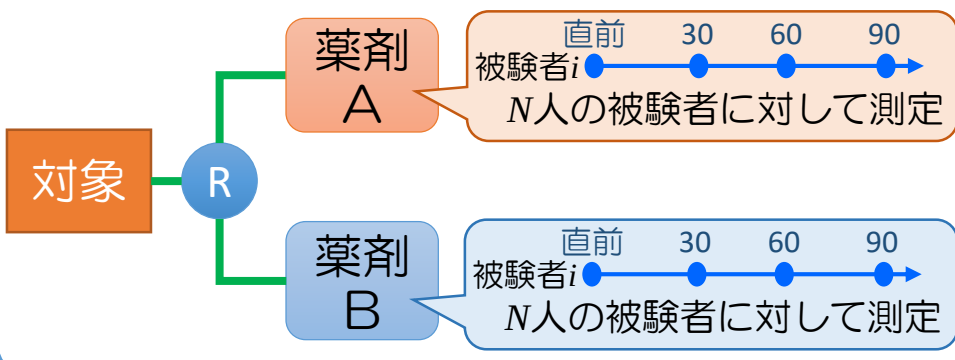
EZR：対応のある3群以上の間の比較(Friedman検定)

分散分析の種類(3)：繰り返し測定データの解析(2)

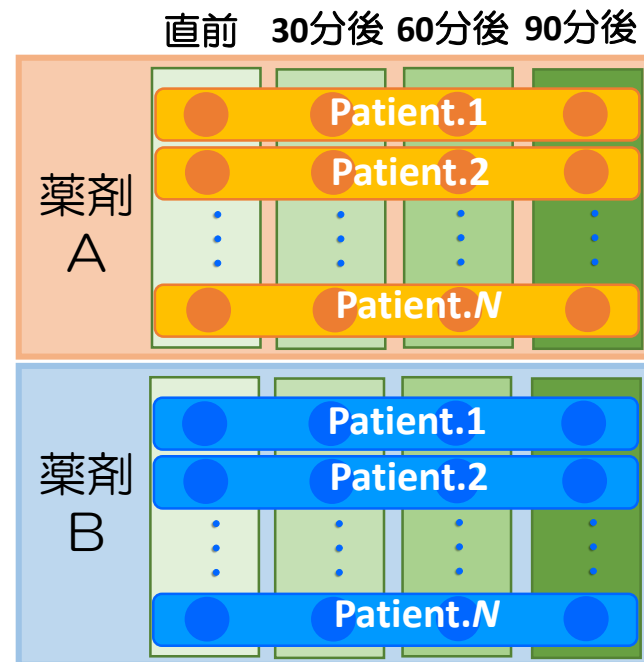


Clinical Question

神経障害性疼痛患者がランダムに割り付けられ、3種類の除痛薬(A,B)のいずれかが投与され、投与直前、30分後、60分後、90分後のVASの変化を調査した。



Data structure



- 2剤のVASに違いがあるか？
- VASに経時的変化があるか？
- (時間×薬剤の交互作用があるか？)

■ パラメトリック検定：繰り返し測定の分散分析

EZR：対応のある2群以上の間の平均値の比較(反復[経時]測定分散分析)

■ ノンパラメトリック検定：存在しない

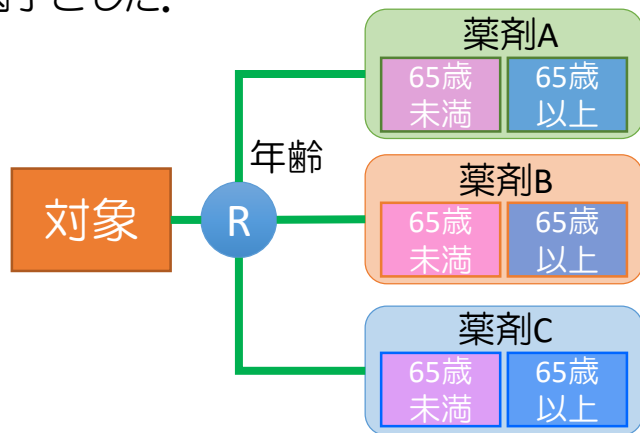
EZR：存在しない

分散分析の種類(4)：多元配置の分散分析



Clinical Question

神経障害性疼痛患者がランダムに割り付けられ、3種類の除痛薬(A,B,C)のいずれかが投与された。この試験では、年齢(65歳以上、65歳未満)を割付調整因子とした。



Data structure

P：被験者

	65歳未満	65歳以上
薬剤A	P P P P P P P P P P	P P P P P P P P P P P
薬剤B	P P P P P P P P P P P P	P P P P P P P P P P P
薬剤C	P P P P P P P P P P	P P P P P P P P P

- 年齢で調整(adjust)したときの3薬のVAS減少量に違いがあるか？
- (年齢×薬剤の交互作用効果があるか？)

■ パラメトリック検定：多元配置の分散分析

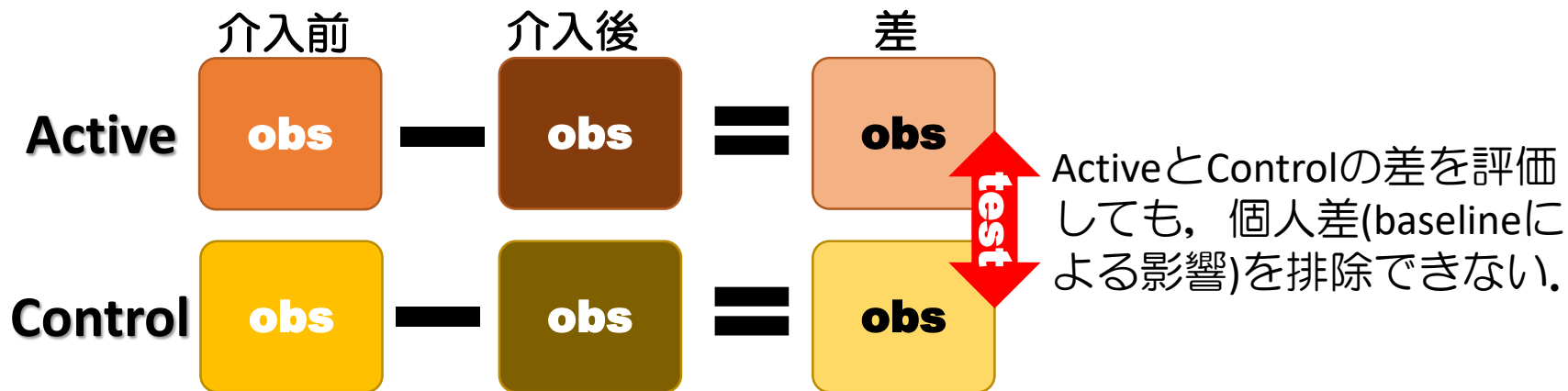
EZR：複数の因子での平均値の比較(多元配置分散分析multi-way ANOVA)

■ ノンパラメトリック検定：存在しない

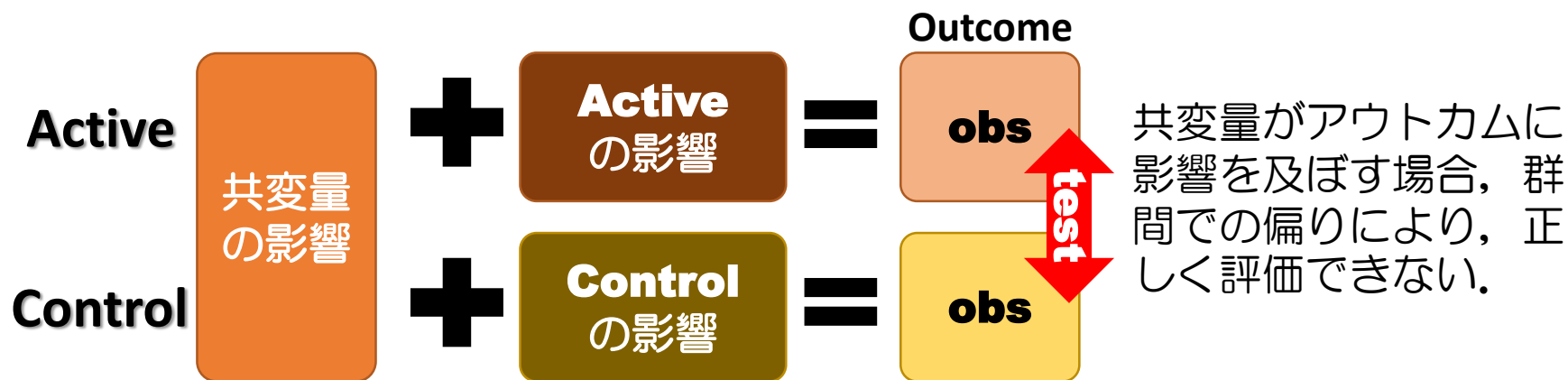
EZR：存在しない

共分散分析の動機

- 介入前後での変化量 or 変化率を比較する



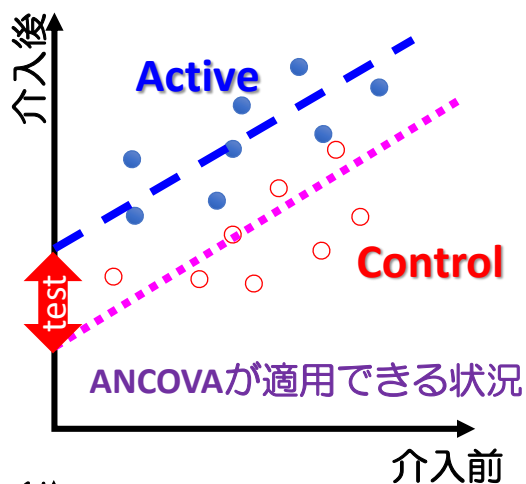
- 介入前の共変量の影響を排除して解析する



共分散分析の適用

共分散分析の手順

- 共分散分析(ANCOVA : ANalysis of COVariance)とは、2群のそれぞれに対して、説明変数Xに介入前、応答変数Yに介入前としたときの回帰直線を引き、その切片の差を比較する方法である。
- ただし、切片を比較するうえにおいて、傾きが有意に異なる場合には評価が不可能になるため、事前に傾きが異なるか否かを検定する必要がある。

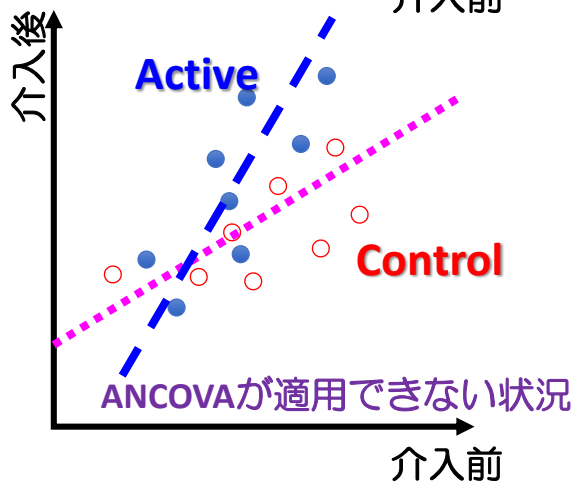


[STEP.1] Active, Controlそれぞれに回帰直線を引く。

[STEP.2] 2種類の検定を考える

[STEP.2-1] 傾きの違い(交互作用)を比較する(介入前値の違いによる介入後値の変化が同じであるか検討する) [Test.1]

[STEP.2-1] 切片を比較する(介入前値の違いによる介入後値の変化が同じであるか検討する) [Test.2]



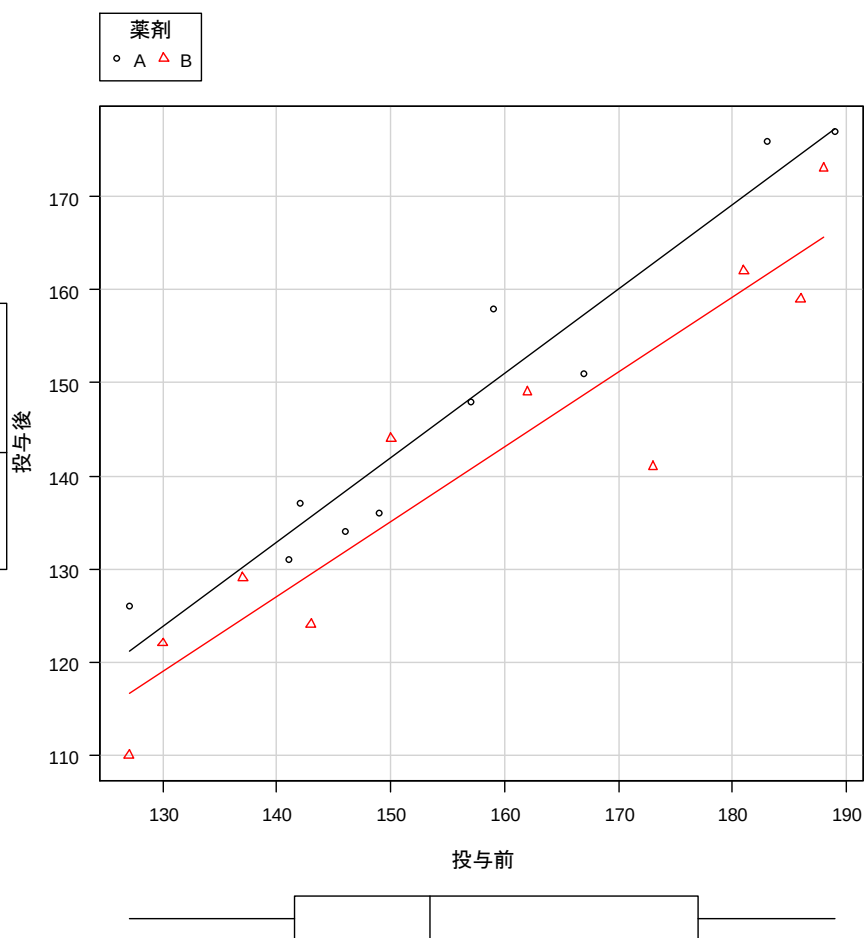
Test.1が有意な場合には治療前後での比較は不可能

Test.2が有意な場合には変化に違いがあると解釈

EZRによるANCOVAの実行

これは、10名ずつランダムに割り付けたうえで、2種類の降圧剤(A,B)のいずれかを投与したときの、投与前後での収縮期血圧の値である。

薬剤A	投与前	159	127	142	146	157	183	149	141	189	167
	投与後	158	126	137	134	148	176	136	131	177	151
薬剤B	投与前	181	162	188	130	127	186	137	173	143	150
	投与後	162	149	173	122	110	159	129	141	124	144



ANCOVAの手順

Test.1 : 傾きの違い(交互作用)を比較する

Test.2 : 切片を比較する

因みに、左図はEZRにおいて次の手順で作成できる

1. 「グラフと表」→「散布図」を選択する。
2. 「x変数(1つ選択)」で「投与前」, 「y変数(1つ選択)」で「投与後」を選択する。
3. 「層別」(左下のボタン)をクリックし, 「薬剤」を選択する。

EZRによるANCOVAの実行

名義変数の解析

連続変数の解析

ノンパラメトリック検定

生存期間の解析

検査の正確度の評価

マッチドペア解析

メタアナリシスとメタ回帰

必要サンプルサイズの計算

連続変数の要約

外れ値の検定と除外(Smirnov-Grubbs検定)

正規性の検定(Kolmogorov-Smirnov検定)

平均値の信頼区間の計算

1標本の平均値のt検定

2群の等分散性の検定(F検定)

2群間の平均値の比較(t検定)

対応のある2群間の平均値の比較(paired t検定)

3群以上の等分散性の検定(Bartlett検定)

3群以上の間の平均値の比較(一元配置分散分析one-way ANOVA)

対応のある2群以上の間の平均値の比較(反復[経時]測定分散分析)

複数の因子での平均値の比較(多元配置分散分析multi-way ANOVA)

連続変数で補正した2群以上の間の平均値の比較(共分散分析ANCOVA)

相関係数の検定(Pearsonの積率相関係数)

線形回帰(単回帰、重回帰)

連続変数で補正した2群以上の間の平均値の比較(共分散分析ANCOVA)

モデル名を入力: AnovaModel.5

目的変数(1つ選択) 比較する群(1つ選択)

投与後 投与後
投与前 投与前
薬剤

補正に用いる連続変数(1つ選択)

投与後
投与前

☐ モデル解析用に解析結果をアクティブモデルとして残す

↓一部のサンプルだけを解析対象にする場合の条件式。例: age>50 & Sex==0 や age<50 | Sex==1

<全ての有効なケース>

ヘルプ リセット OK キャンセル 適用

(上略)

```
> cat(gettextRcmdr("P value for interaction between factor(薬剤) and factor(投与前/投与後)", 0.473, ""))
```

群別変数と共変数の交互作用のP値は 0.473

```
> AnovaModel.5 <- AnovaModel.5 <- lm(投与後 ~ factor(薬剤) + factor(投与前/投与後), data = data, na.action=na.omit)
```

```
> Anova(AnovaModel.5, type="III")
```

Anova Table (Type III tests)

Response: 投与後

	Sum Sq	Df	F value	Pr(>F)
(Intercept)	80.1	1	2.0510	0.17024
factor(薬剤)	283.5	1	7.2634	0.01533 *
投与前	5917.0	1	151.6012	6.782e-10 ***
Residuals	663.5	17		

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

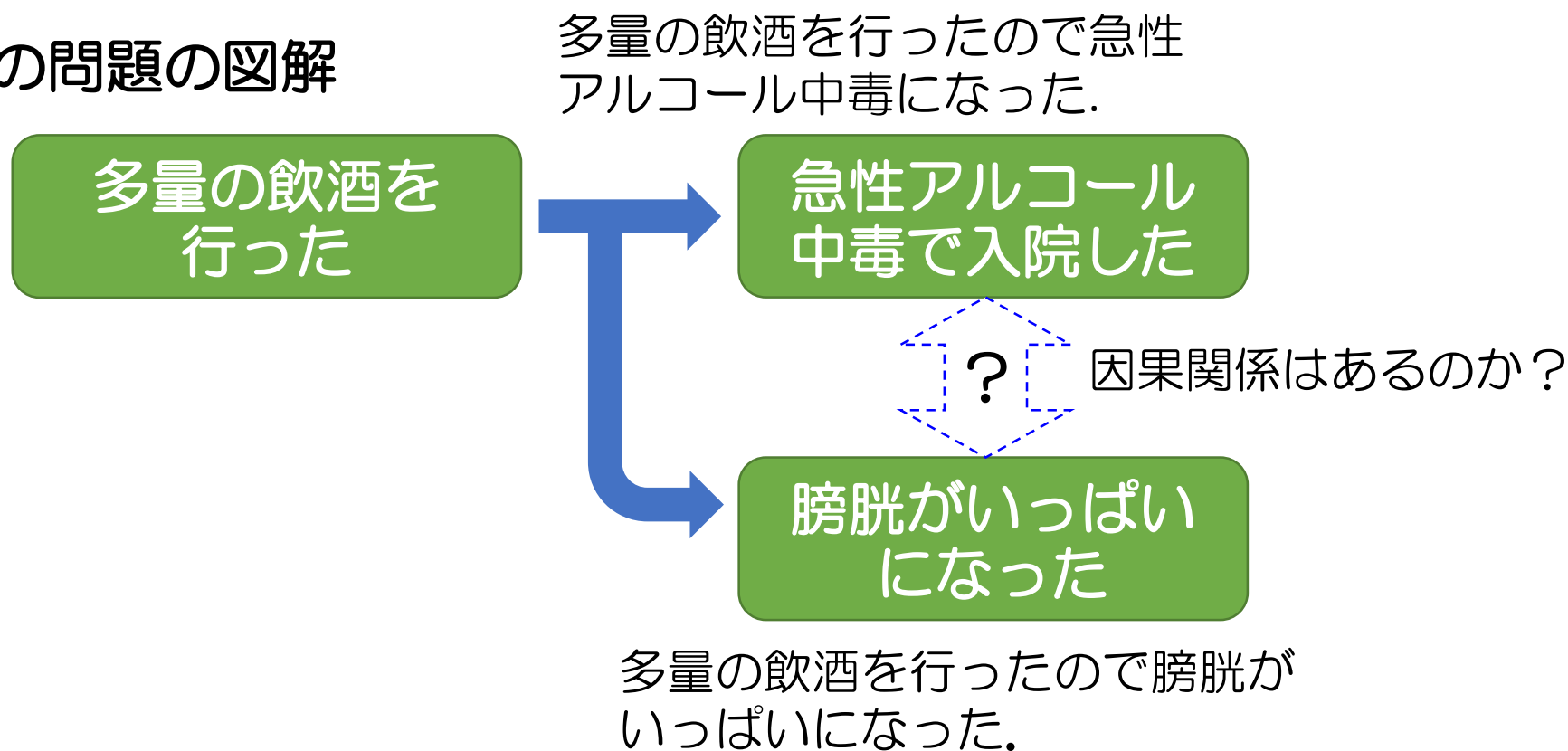
「有意でない」場合に、ANCOVAによる比較可能性が担保できていると判断される

この部分がANOCVAの結果である。p値が0.05未満なので、降圧剤による違いが認められる。

因果(回帰)と相関は違う

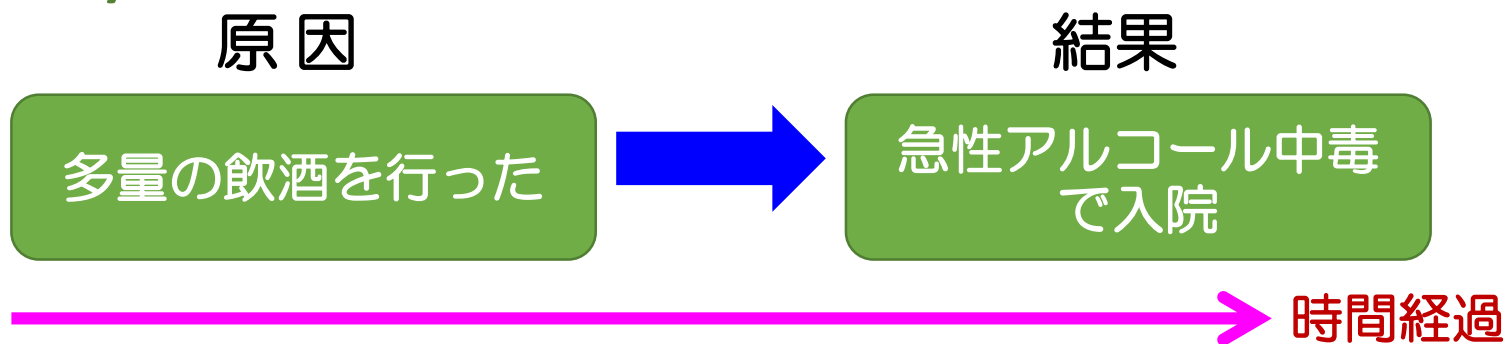
急性アルコール中毒で入院した患者100人すべての膀胱がいっぱいだった。膀胱圧と急性アルコール中毒には因果関係があるといえるか？

この問題の図解



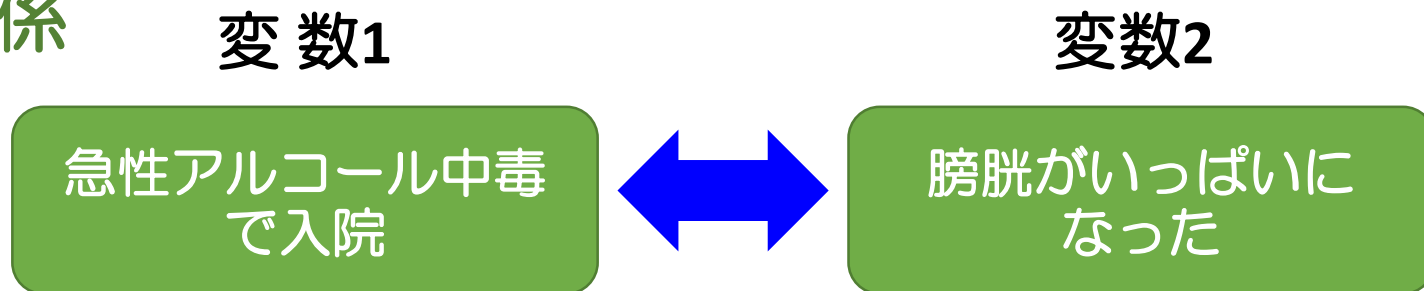
因果(回帰)と相関は違う

因果(回帰)関係



「原因があった(なかった)から結果が起きた(起きなかった)」という関係を因果関係という。

相関関係



2個の変数の関係性のみのことを相関関係という

相関分析

パラメトリック相関

Pearsonの積率相関係数 (通常の相関係数)

連続変数の要約

外れ値の検定と除外(Smirnov-Grubbs検定)

正規性の検定(Kolmogorov-Smirnov検定)

平均値の信頼区間の計算

1標本の平均値のt検定

2群の等分散性の検定(F検定)

2群間の平均値の比較(t検定)

対応のある2群間の平均値の比較(paired t検定)

3群以上の等分散性の検定(Bartlett検定)

3群以上の間の平均値の比較(一元配置分散分析one-way ANOVA)

対応のある2群以上の間の平均値の比較(反復[経時]測定分散分析)

複数の因子での平均値の比較(多元配置分散分析multi-way ANOVA)

連続変数で補正した2群以上の間の平均値の比較(共分散分析ANCOVA)

相関係数の検定(Pearsonの積率相関係数)

線形回帰(単回帰、重回帰)

ノンパラメトリック相関

Spearmanの順位相関係数, Kendallの順位相関係数

2群間の比較(Mann-Whitney U検定)

対応のある2群間の比較(Wilcoxon符号付順位和検定)

3群以上の間の比較(Kruskal-Wallis検定)

対応のある3群以上の間の比較(Friedman検定)

連続変数の傾向の検定(Jonckheere-Terpstra検定)

相関係数の検定(Spearmanの順位相関係数)

順位相関係数にはSpearmanの順位相関係数をもちいることが一般的

1. データに外れ値あるいはいずれか(or 両方)の分布が著しく歪んでいる場合には、ノンパラメトリック相関を用いることが推奨される。
2. 相関係数の表記において、無相関性の検定におけるp値を記載することが多いが、SAMPLガイドラインでは95%信頼区間を表記することが推奨されている。
(検定は標本サイズに依存するため、症例数が多いと低い相関係数でも有意になるため)

重回帰分析

線形回帰(単回帰、重回帰)

モデル名を入力: RegModel.4

複数の選択はCtrlキーを押しながらクリック。↓

目的変数 (1つ選択)

説明変数 (1つ以上選択)

目的変数
応答変数
従属変数

説明変数
独立変数

☐ 3レベル以上の因子についてその因子全体のP値の計算(Wald検定)

☐ モデル解析用に解析結果をアクティブモデルとして残す

☐ 基本的診断プロットを表示する

☐ AICを用いたステップワイズの変数選択を行う。

☐ BICを用いたステップワイズの変数選択を行う。

☐ P値を用いたステップワイズの変数選択(減少法)を行う。

↓一部のサンプルだけを解析対象にする場合の条件式。例: age>50 & Sex==0 や age<50 | Sex==1

<全ての有効なケース>

ヘルプ リセット OK キャンセル 適用

重回帰分析では交互作用を設定できない。

変数増減法によるステップワイズ法

AIC(Akaike's Information Criteria)とBIC(Bayesian Information Criteria)では、BICのほうが選択される変数の数が少なくなる傾向にある。



EZRにおいて実行できる統計解析

生存時間解析編(3章)

生存曲線の推定(Kaplan-Meier曲線)・ログランク検定

生存曲線の記述と解釈(1標本)：3.1.3節

ここでは、North Central Cancer Treatment Groupによって実施された進行肺癌患者に対するデータを用いる。このデータは、228名の進行肺癌患者の全生存期間(日)がとられている。

生存曲線の記述と群間の比較(Logrank検定)

生存期間の傾向の検定(Logrank trend検定)

生存期間に対する多変量解析(Cox比例ハザード回帰)

時間依存性変数を含む生存期間に対する多変量解析(Cox比例ハザード回帰)

累積発生率(競合イベントを含む)の記述と群間の比較(Gray検定)

累積発生率(競合イベントを含む)に対する多変量解析(Fine-Gray比例ハザード回帰)

生存曲線の記述と群間の比較(Logrank検定)

1標本：Kaplan-Meier推定

2標本：ログランク検定(or 一般化Wilcoxon検定)

time

status

生存曲線の記述と群間の比較(Logrank検定)

観察期間の変数(1つ選択) イベント(1)、打ち切り(0)の変数(1つ選択)

↓ 複数の選択はCtrlキーを押しながらクリック。

群別する変数を選択(0～複数選択可) 層別化変数(0～1つ選択)

解析方法 ☒ logrank ☐ Peto-Peto-Wilcoxon

異なる線の区別 ☒ 色分け ☐ 線種 ☐ 線の太さ

凡例の書式 ☒ 右上・縦並び ☐ 下部・横並び ☐ マウス指定・縦並び

X軸の単位 ☒ そのまま ☐ 日を月に変換 ☐ 日を年に変換 ☐ 月を年に変換

Post-hoc検定 (比較する群が1つの場合) ☒ No ☐ Bonferroni ☐ Holm

オプション

☒ 打ち切りをマークで表示する ☐ 95%信頼区間を表示する ☐ 異なる層毎に別の図に表示する ☐ At riskのサンプル数を表示する

生存率を表示するポイント(観察期間の単位が日で1年生存率なら365)

X軸の範囲(左端,右端) 例: 0, 365 Y軸の範囲(下端,上端) 例: 0.8, 1.0

X軸のラベル Y軸のラベル

↓ 一部のサンプルだけを解析対象にする場合の条件式。例: age>50 & Sex==0 や age<50 | Sex==1

ヘルプ リセット OK キャンセル 適用

結果の解釈

生命表を用いた年次生存率の推定

```
Call: survfit(formula = Surv((time/365.25), status == 1) ~ 1, data = Dataset,  
na.action = na.omit, conf.type = "log-log")
```

time	n.risk	n.event	survival	std.err	lower 95% CI	upper 95% CI
5	228	1	0.9956	0.00438	0.9693	0.999
11	227	3	0.9825	0.00869	0.9539	0.993
12	224	1	0.9781	0.00970	0.9481	0.991

(省略)

363	69	2	0.4154	0.03583	0.3448	0.484
-----	----	---	--------	---------	--------	-------

364	67	1	0.4092	0.03582	0.3387	0.478
------------	-----------	----------	---------------	----------------	---------------	--------------

1年生存率(365.25日)

371	65	2	0.3966	0.03581	0.3264	0.466
-----	----	---	--------	---------	--------	-------

(省略)

707	15	1	0.1246	0.02904	0.0748	0.188
-----	----	---	--------	---------	--------	-------

728	14	1	0.1157	0.02830	0.0676	0.178
------------	-----------	----------	---------------	----------------	---------------	--------------

2年生存率(365.25×2=730.5日)

731	13	1	0.1068	0.02749	0.0606	0.168
-----	----	---	--------	---------	--------	-------

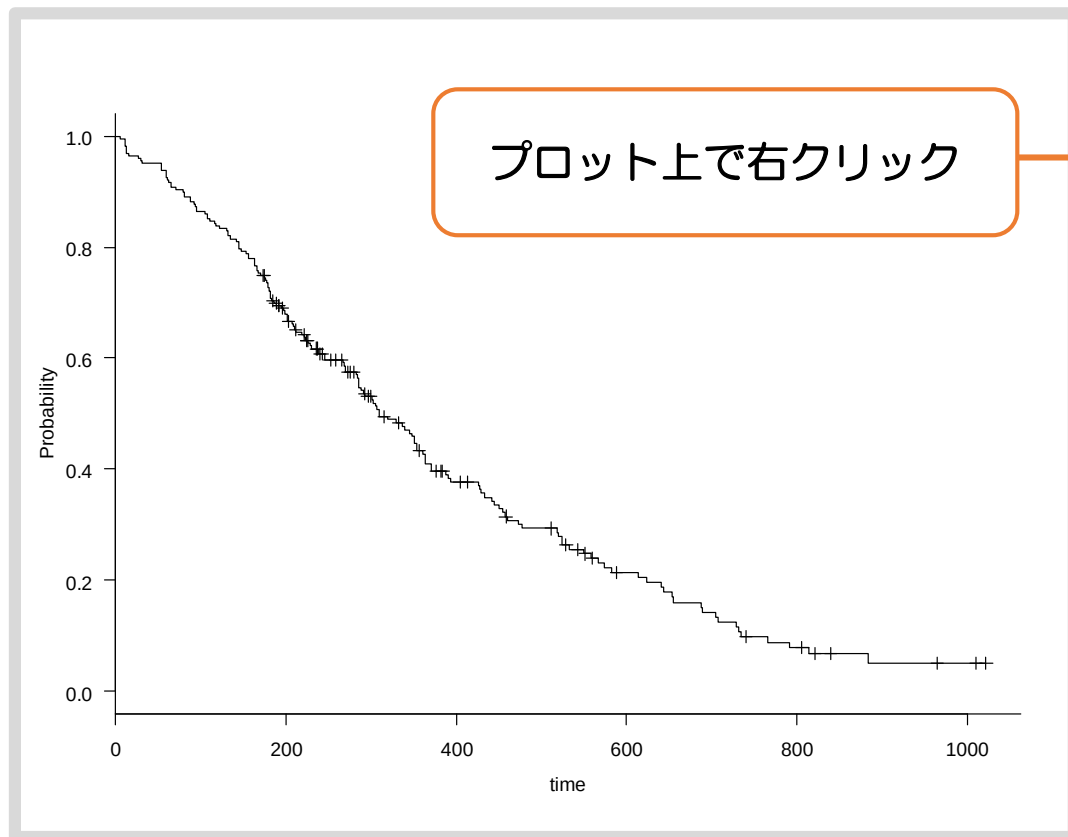
(省略)

中央生存期間(MST)の推定：EZRのアウトプットによる解釈

サンプル数	生存期間中央値	95%信頼区間
-------	---------	---------

1	228	310	284-361
---	-----	-----	---------

Kaplan-Merierプロットの出力結果と保存方法



「メタファイルにコピー」
ベクター形式でコピーできる

「ビットマップにコピー」
画像形式でコピーできる

「メタファイルに保存」
TIFFでコピーできる

「ビットマップにコピー」
BMPでコピーできる

生存曲線の推定(Kaplan-Meier曲線)・ログランク検定

生存曲線の比較：3.2.3節

ここでは、卵巣癌データを用いて生存曲線を比較する。このデータは、26名の卵巣癌患者に対する2種類の抗癌剤(既存薬, 新薬)における全生存期間(日)がとられている。

生存曲線の記述と群間の比較(Logrank検定)

生存期間の傾向の検定(Logrank trend検定)

生存期間に対する多変量解析(Cox比例ハザード回帰)

時間依存性変数を含む生存期間に対する多変量解析(Cox比例ハザード回帰)

累積発生率(競合イベントを含む)の記述と群間の比較(Gray検定)

累積発生率(競合イベントを含む)に対する多変量解析(Fine-Gray比例ハザード回帰)

生存曲線の記述と群間の比較(Logrank検定)

1標本：Kaplan-Meier推定

2標本：ログランク検定(or 一般化Wilcoxon検定)

time

group

検定方法は2種類存在
「logrank」

→ログランク検定

「Peto-Peto-Wilcoxon」

→一般化Wilcoxon検定

生存曲線の記述と群間の比較(Logrank検定)

観察期間の変数(1つ選択) イベント(1)、打ち切り(0)の変数(1つ選択)

group
status
time

group
status
time

↓ 複数の選択はCtrlキーを押しながらクリック。

群別する変数を選択(0~複数選択可) 層別化変数(0~1つ選択)

group
status
time

group
status
time

解析方法

☒ logrank
☐ Peto-Peto-Wilcoxon

異なる線の区別

☒ 色分け
☐ 線種
☐ 線の太さ

凡例の書式

☒ 右上・縦並び
☐ 下部・横並び
☐ マウス指定・縦並び

X軸の単位

☒ そのまま
☐ 日を月に変換
☐ 日を年に変換
☐ 月を年に変換

Post-hoc検定
(比較する群が1つの場合)

☒ No
☐ Bonferroni
☐ Holm

オプション

☒ 打ち切りをマークで表示する ☐ 95%信頼区間を表示する ☐ 異なる層毎に別の図に表示する ☐ At riskのサンプル数を表示する

生存率を表示するポイント(観察期間の単位が日で1年生存率なら365) <none>

X軸の範囲(左端,右端) 例: 0, 365 <auto> Y軸の範囲(下端,上端) 例: 0.8, 1.0 <auto>

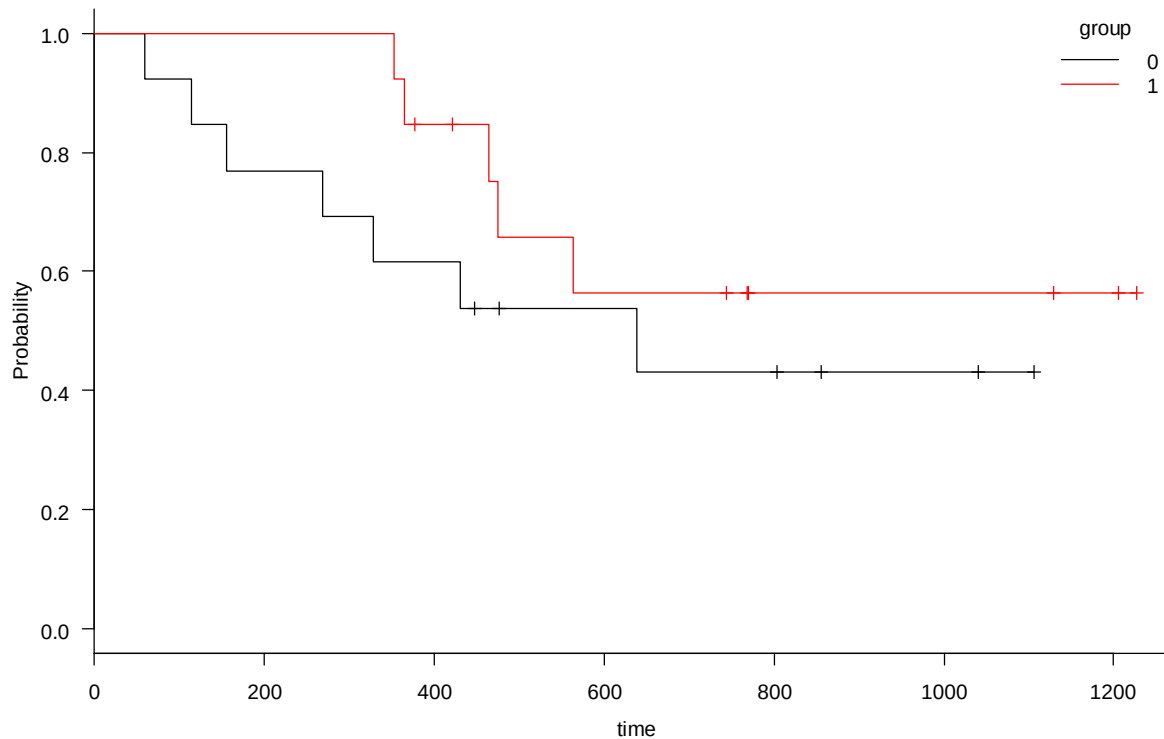
X軸のラベル <auto> Y軸のラベル <auto>

↓ 一部のサンプルだけを解析対象にする場合の条件式。例: age>50 & Sex==0 や age<50 | Sex==1

<全ての有効なケース>

ヘルプ リセット OK キャンセル 適用

結果の解釈



中央生存期間(MST)の推定：EZRのアウトプットによる解釈

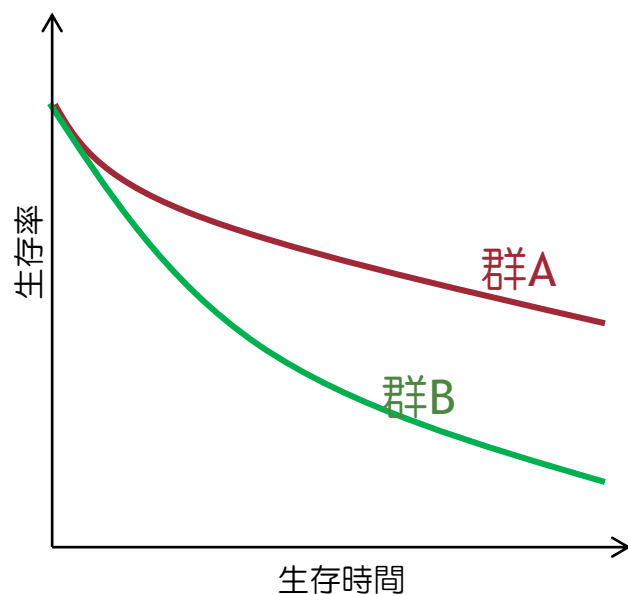
```
> km.summary.table
```

	サンプル数	生存期間中央値	95%信頼区間	P値
group=0	13	638	156-NA	0.303
group=1	13	NA	464-NA	

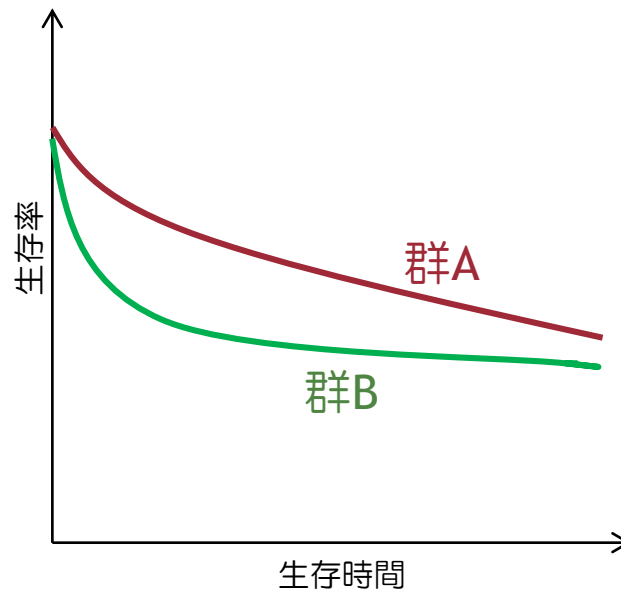
NA：計算できなかった場合に表示される

ログランク検定と一般化Wilcoxon検定の違い

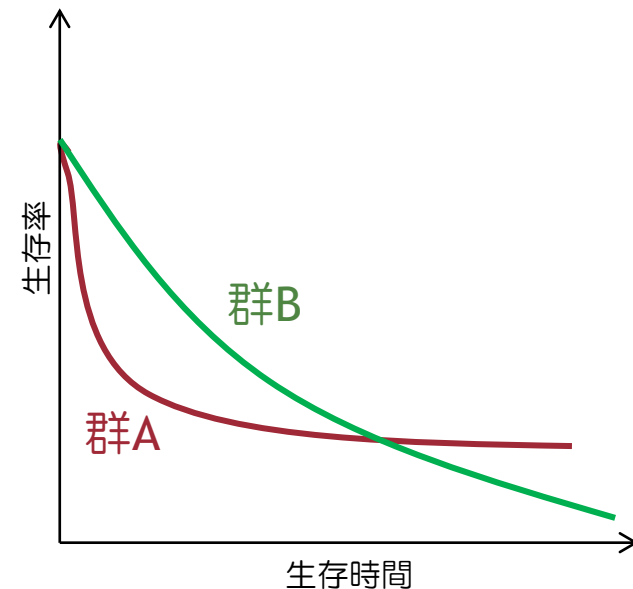
(a) 後期で生存率の差が顕著



(b) 前期で生存率の差が顕著



(c) 生存曲線が交差



(a)の状況では、ログランク検定のほうが一般化Wilcoxon検定よりも検出力が強い(また比例ハザード性の仮定のもとで一様最強力検定となる)。

(b)の状況では、一般化Wilcoxon検定のほうがログランク検定よりも検出力が高い。

(c)の状況では、いずれの方法も検出力が低い(いずれの方法でも、生存時間関数が一様に高いことを仮定するため)。このような場合には、折れ線ハザードモデル(文献の中村：Cox比例ハザードモデルが詳しい)、あるいは比例ハザードモデルによる交互作用の精査が必要である。

比例ハザード・モデル

生存期間に対する多変量解析(Cox比例ハザード回帰)

モデル名を入力:

変数 (ダブルクリックして式に入れる)

生存時間
打ち切り
変数1
変数2
変数3

モデル式:

時間 , イベント ~ 説明変数

層別化因子は + strata(####)と入力する ↑ ↑ ↑

☐ 3レベル以上の因子についてその因子全体のP値の計算(Wald検定)

☐ 比例ハザード性の分析を行う。

☐ マルチンゲール残差をプロットする

☐ ベースラインの生存曲線を示す。

☐ モデル解析用に解析結果をアクティブモデルとして残す

☐ AICを用いたステップワイズの変数選択を行う。

☐ BICを用いたステップワイズの変数選択を行う。

☐ P値を用いたステップワイズの変数選択(減少法)を行う。

↓ 一部のサンプルだけを解析対象にする場合の条件式。例: age>50 & Sex==0 や age<50 | Sex==1

ヘルプ リセット OK キャンセル 適用

変数間に「+」を入れる

交互作用を検討す場合には、「*」にする。
例(変数1と変数2の交互作用): 変数1*変数2

変数増減法によるステップワイズ法

AIC(Akaike's Information Criteria) と BIC(Bayesian Information Criteria)では、BICのほうが選択される変数の数が少なくなる傾向にある。

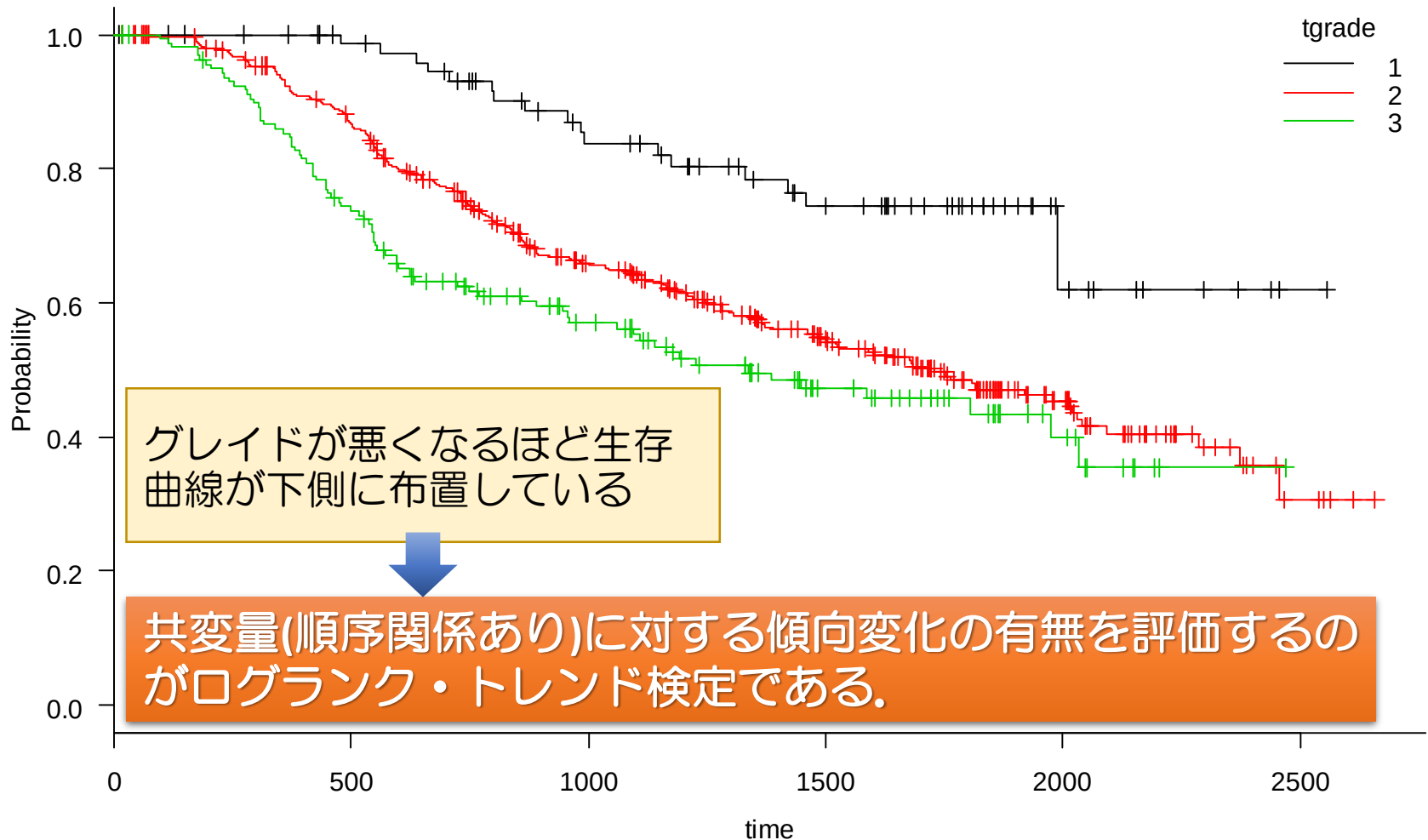
傾向性の検定：ログランク・トレンド検定

テキストには非掲載

ホルモン療法の効果を検討するために、ドイツ乳癌研究グループ(GBSG; German Breast Cancer Study Group)が実施した無作為化比較第Ⅲ相試験の結果である。

このデータは、腫瘍グレード(tgrade)が1から3までとられている。

ここでの関心は、**主要グレードが高くなるにつれて、生存曲線が下側にいく傾向にあるか否かを検定**することにある。



傾向性の検定：ログランク・トレンド検定

テキストには非掲載

The screenshot shows the 'Logrank trend検定' dialog box. It has three main sections for variable selection:

- 観察期間の変数(1つ選択)**: A list of variables including `estrec`, `horTh`, `menostat`, `pnodes`, `progre`, `tgrade`, `time`, and `tsize`. The variable `time` is selected. An annotation box labeled '生存期間 time' points to this selection.
- イベント(1)、打ち切り(0)の変数(1つ選択)**: A list of variables including `age`, `cens`, `estrec`, `horTh`, `menostat`, `pnodes`, `progre`, and `tgrade`. The variable `cens` is selected. An annotation box labeled '打ち切り指標(0: 打ち切り, 1: イベント) cens' points to this selection.
- 比較する群 (1つ選択)**: A list of variables including `estrec`, `horTh`, `menostat`, `pnodes`, `progre`, `tgrade`, `time`, and `tsize`. The variable `tgrade` is selected. An annotation box labeled '順序がある共変量 tgrade' points to this selection.

At the bottom of the dialog, there is a text field for a condition: '<全ての有効なケース>'. Below the text field are buttons for 'ヘルプ', 'リセット', 'OK', 'キャンセル', and '適用'.

(上略)

```
> logrank.trend(res)
```

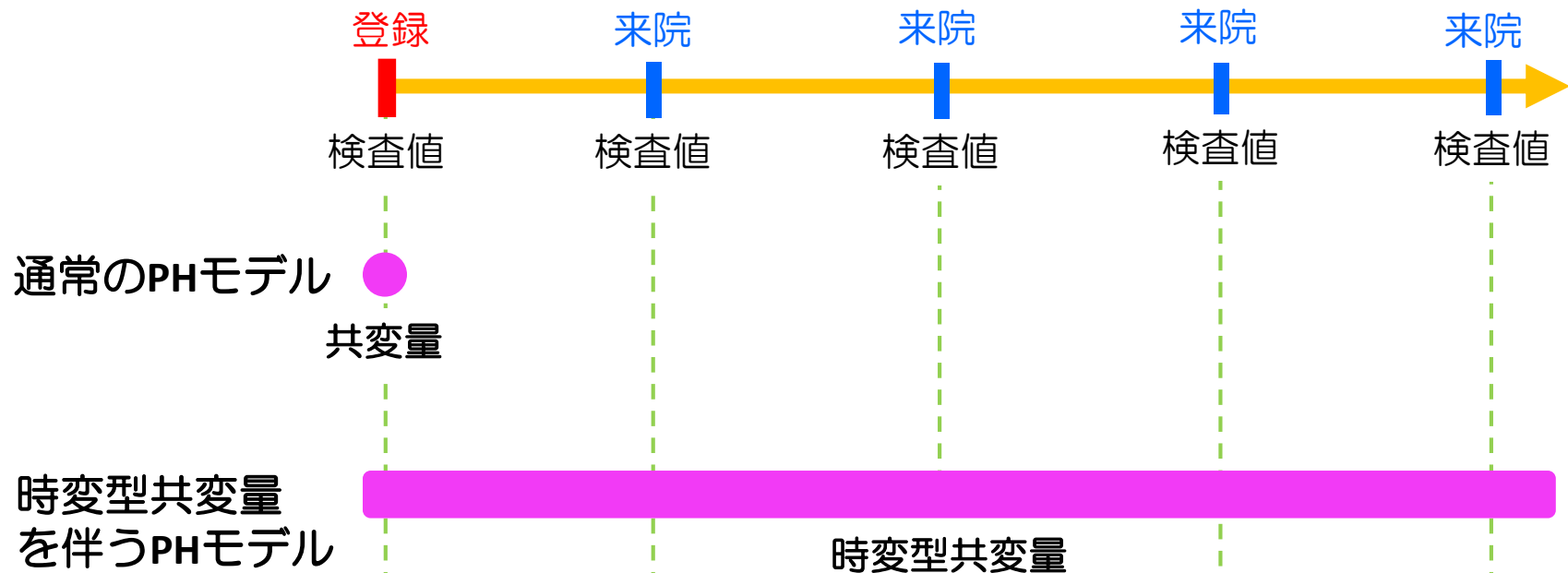
生存期間の傾向の検定 (Logrank trend検定)

カイ二乗値	19.9
自由度	1
P値	8.3e-06

特殊な状況での比例ハザードモデル(1)

■ 時変型共変量を伴う比例ハザードモデル

EZR : 「生存時間の解析」 → 「時間依存性変数を含む多変量解析(Cox比例ハザード回帰)」



1. 臨床試験では、登録時だけでなく、来院時等においてもデータが取得される。
2. 一方で、通常のプロポーションハザードモデルの共変量には登録時の検査データを用いるものの、来院時等でのデータを共変量(多変量解析)に用いない。
3. 時変型共変量を伴う比例ハザードモデルとは時間経過とともに変化する共変量(時変型共変量)を取り扱う。
4. ただし、時変型共変量は解釈が難しいため、あまり用いられない。

特殊な状況での比例ハザードモデル(2)

層別比例ハザードモデル

EZR: 「マッチドペア解析」→「マッチさせたサンプルの生存率の多変(層別化比例ハザード回帰)」

通常の比例ハザード・モデル

$$\text{ハザード} = \text{基線ハザード (BH)} \times \exp \left(\beta_1 x_1 + \beta_2 x_2 + \cdots + \beta_p x_p \right)$$

時間 t のみに依存

共変量: 治療法, 性別, 腫瘍径などのこと

係数1×共変量1+係数2×共変量2+...

共変量 x のみに依存

例えば, 国際共同治験などの場合には, 国や民族差・人種差などが違うため, 基線ハザード自体が異なる可能性がある

層別比例ハザード・モデル

$$\text{ハザード} = \text{BH1, BH2, BH3} \times \exp \left(\beta_1 x_1 + \beta_2 x_2 + \cdots + \beta_p x_p \right)$$

時間 t のみに依存

共変量: 治療法, 性別, 腫瘍径などのこと

係数1×共変量1+係数2×共変量2+...

共変量 x のみに依存

最後を前に余談：EZRの使い勝手の良さ

乳癌に罹患している患者に対するホルモン療法の効果を調査するために、ドイツ乳癌研究グループが8個の共変量

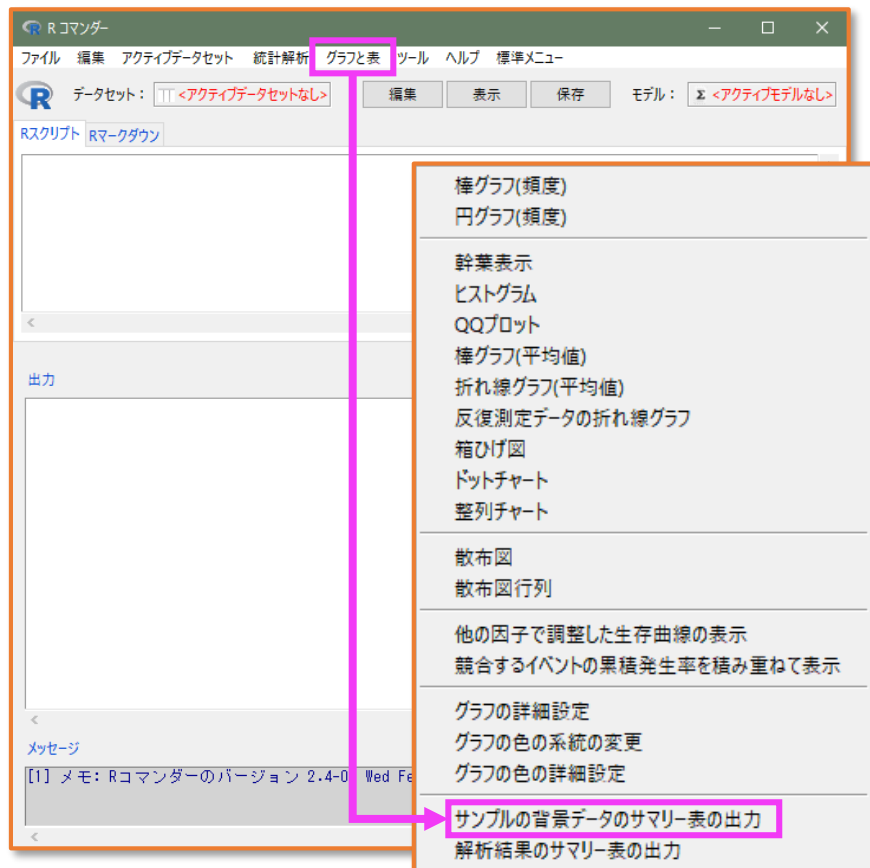
- 年齢(age)
- 閉経の有無(menostat)
- 腫瘍径(tsize)
- 腫瘍グレード(tgrade)
- リンパ節転移個数(pnode)
- プロゲステロン・レセプター(progrec)
- エストロゲン・レセプター(estrec)
- ホルモン療法が行われたか否か(horTh)

とともに生存時間を調査している。

年齢(age)	連続データ
閉経の有無(menostat)	名義データ
腫瘍径(tsize)	連続データ
腫瘍グレード(tgrade)	順序データ
リンパ節転移個数(pnode)	連続データ
プロゲステロン・レセプター(progrec)	連続データ
エストロゲン・レセプター(estrec)	連続データ
ホルモン療法が行われたか否か(horTh)	グループ変数

ホルモン療法の有無で2群に分けられたそれぞれに対して、背景因子の要約の表を作るのは意外に面倒

Summary tableの自動生成



留意点

- 平均値(標準偏差)で要約する場合には、「連続変数(正規分布)」を選択する。
- 「出力先」において結果をcsvファイルに保存する場合には、「csvファイル」を選択する。
- 「四分位範囲」で中央値のバラツキを表示する場合には「四分位変数」を選択する。

Summary tableの自動生成

出力

```
> FinalTable <- FinalTable[which(rownames(FinalTable)!="n"),]
> FinalTable <- rbind(n=row1, FinalTable)
> FinalTable <- rbind(row0, FinalTable)
> row0 <- rep("", length(colnames(FinalTable)))
> row0[3] <- "horTh"
> FinalTable <- rbind(row0, FinalTable)
> finaltable_dataframe_print(FinalTable)
```

Factor	Group	horTh no	yes	p.value
n		440	246	
menostat (%)	Post	209 (47.5)	187 (76.0)	<0.001
	Pre	231 (52.5)	59 (24.0)	
tgrade (%)	1	48 (10.9)	33 (13.4)	0.268
	2	281 (63.9)	163 (66.3)	
	3	111 (25.2)	50 (20.3)	
age		50.00 [21.00, 80.00]	58.00 [32.00, 80.00]	<0.001
estrec		32.00 [0.00, 898.00]	46.00 [0.00, 1144.00]	0.009
pnodes		3.00 [1.00, 51.00]	3.00 [1.00, 36.00]	0.432
progrec		32.00 [0.00, 1600.00]	35.00 [0.00, 2380.00]	0.594
tsize		25.00 [3.00, 120.00]	25.00 [4.00, 100.00]	0.507

> write.table(FinalTable, "clipboard", sep=" ", row.names=F)

度数(割合)

Fisherの正確検定

中央値 最小値, 最大値

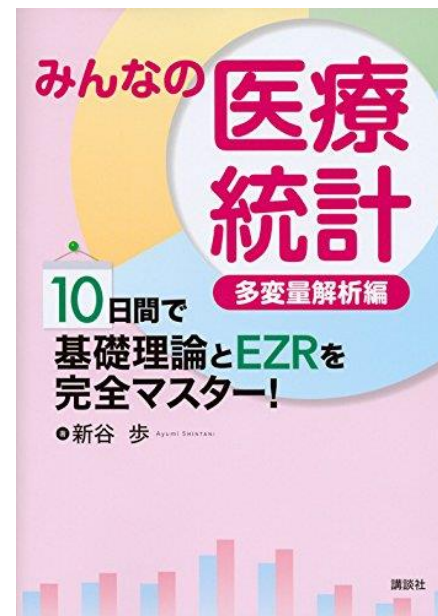
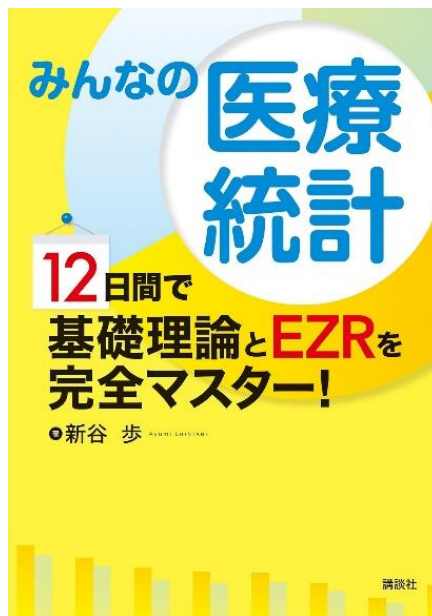
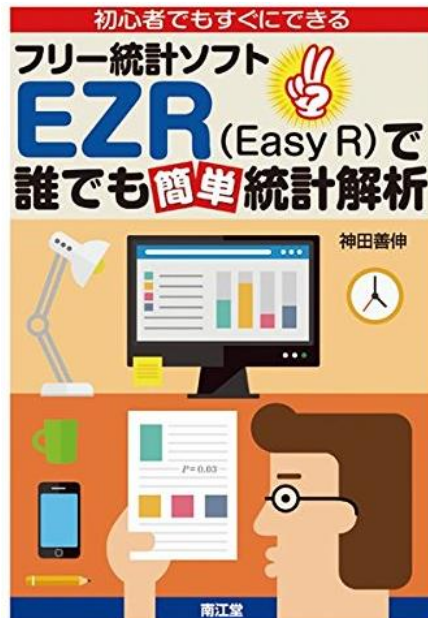
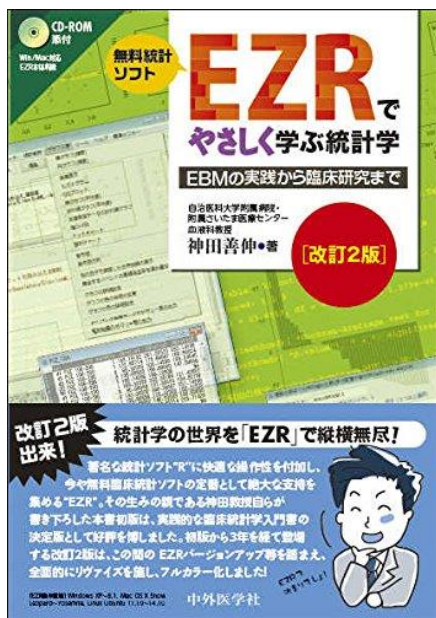
連続変数(正規分布)の場合
平均値 (標準偏差)

Wilcoxon検定

連続変数(正規分布)の場合
2標本 t 検定

因みに、出力先に「クリップボード」(デフォルト)を選択した場合には、Excelで貼り付けを選択すると、結果がペーストされる。

結びに代えて：EZRの書籍紹介



EZRに関する書籍としては、製作者の神田先生、大阪市立大学の新谷歩先生が出版している。いずれの書籍もマニュアルとして利用しやすく、また、医学統計での利用を想定している点も利点の一つである。

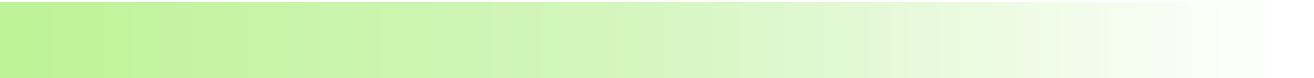
ご清聴ありがとうございます



shimokaw@wakayama-med.ac.jp



症例数の設計



必要症例数の設計とその意味

第1種の過誤(α エラー)と第2種の過誤(β エラー)

	検定により評価	標本サイズで規定
	H_0 が正しい [H_1 : 偽]	H_1 が正しい [H_0 : 偽]
H_0 を棄却 [H_1 を受容]	第1種の過誤 α (H_0 が正しいのに棄却)	検出力 $1-\beta$ (H_0 が正しいので受容)
H_0 を受容 [H_1 を受容]	正しい判断 $1-\alpha$ (H_1 が正しいので棄却)	第2種の過誤 β (H_1 が正しいのに受容)

■ 第2種の過誤が一定水準未満(β 未満)になるように、標本サイズを規定する (本当はpositiveなのに本試験を通じてnegativeと判断する可能性を低くする).

■ 統計解析の結果、第1種の過誤が一定水準未満(α 未満)であるか否かを確認する(本当はnegativeなのに本試験結果を通じてpositiveと結論付けるエラーがかなり低いことを確認する).

■ プロトコルにおける記載例

ToGA試験における奏効率は、化学療法群で35%（95%CI: 29-40）、Trastuzumab併用群で47%（95%CI: 42-53）と報告されており、Trastuzumabの上乗せ効果は約12%前後と考えられる。3週1コースのTS-1+CDDP併用療法の奏効率は、2つの臨床試験で48%（95%CI: 33-62）と報告されているが、これらのデータは5-FUベースの術後補助化学療法がほとんど実施されていない症例が対象となっていた。現在はACTS-GCの結果より、本邦における再発症例についてはほぼ全例に前治療としてTS-1が投与されていると考えられる。TS-1術後補助化学療法施行後に再発した胃癌患者に対する化学療法の効果に関するretrospectiveな報告では、再発後の一次治療としてTS-1を含むレジメン群の奏効率が0～6.7%に対し、TS-1を含まないレジメン群では16.3～42.9%であった。今回の対象としてTS-1adjuvant既施行例が含まれることを考慮して、**TS-1+CDDP+Trastuzumab併用療法の閾値奏効率を35%、期待奏効率を50%であると仮定する。**

帰無仮説「TS-1+CDDP+Tmab併用療法の奏効率は35%である」に対して、片側対立仮説「TS-1+CDDP+Tmab併用療法の奏効率は35%を上回る」を有意水準 $\alpha=0.1$ のもとで母比率の検定で評価するとき（片側対立仮説）、検出力 $1-\beta$ が0.8以上になる必要最小被験者数は48例である。そのため、不適格例を考慮して50例とした。

臨床的仮説

文献等から設定する。

試験デザイン：Phase II Study (単アーム)

主要評価項目：奏効割合 (RR: Response Rate)

閾値奏効率：35%， 期待奏効率：50%

統計的仮説

仮説検定(or 信頼区間)から設定する。

第1種の過誤(有意水準)： $\alpha = 0.1$

1－第2種の過誤(検出力)： $1-\beta = 0.8$ (80%)

必要症例数の設計とその意味

H_0 : 奏効率は35%である.

H_1 : 奏効率は35%を上回る.

試験後のはなし

検定により評価

H_0 が正しい [H_1 : 偽]

H_0 を棄却
[H_1 を受容]

第1種の過誤 α
0.10 (片側)

H_0 を受容
[H_1 を受容]

正しい判断 $1-\alpha$

試験後の評価

試験の結果, 奏効率が50%のとき, 有意水準 $\alpha=0.10$ のもとで有意になる.

試験前のはなし

標本サイズで規定

H_1 が正しい [H_0 : 偽]

検出力 $1-\beta$
0.80

第2種の過誤 β

試験計画

必要症例数 = 49

必要症例数の計算(6章)

必要症例数設定のプロセス

STEP.1：主要評価項目(主たるアウトカム)を考える

- 本研究におけるクリニカルクエスチョンに答えられるものか(妥当性)?
- 当該分野においてAgreeされるものか(信頼性)?
- エンドポイントを取得できるか(実施可能性)?
一例えば、抗がん剤治療において標的病変がなければ奏効率が測定できない等

STEP.2：仮説を決定する

- 無作為化比較試験の場合には非劣勢・優越性・同等性を検討する.
- 単アーム試験の場合には閾値を文献等から決定する.
- 本プロトコル治療における主要評価項目での有用性を検討する.
量的変数の場合：共通の分散(標準偏差)を決定する(個体差).
生存時間の場合：受け入れ期間, フォローアップ期間を決定(censoring割合)

STEP.3：有意水準・検出力を決定する

単アーム： $\alpha=0.05$ or 0.10 , $1-\beta = 0.8$
多アーム： $\alpha=0.05$, $1-\beta = 0.80$ or 0.90

試験デザインとアウトカム等の症例設計に必要な情報と対応する検定

	1標本（単アーム試験）		2標本（無作為化比較試験）	
	必要な情報	対応する検定	必要な情報	対応する検定
比率	閾値（比率） 期待値（比率）	母比率の検定 	対照治療での比率 試験治療での比率 （2群の比率 → オッズ比でも可能）	カイ2乗検定 OR Fisherの正確検定 
平均値	閾値（平均値） 期待値（平均値）	母平均の検定 	対照治療での平均値 試験治療での平均値 共通の標準偏差	2標本t検定 
生存曲線	閾値（中央生存期間 OR 年次生存率） 期待値（中央生存期間 OR 年次生存率） （登録期間及び追跡期間）	1標本ログランク検定 OR 信頼区間	対照治療での中央生存期間 OR 年次生存率 試験治療での中央生存期間 OR 年次生存率 （2群の生存率 → ハザード比でも可能） （登録期間及び追跡期間）	ログランク検定 

EZRでは非劣性試験のデザインも可能

	対応のある場合	
	必要な情報	対応する検定
比率	試験治療が有効で対象治療が無効の比率 試験治療が無効で対象治療が有効の比率	McNemar検定 
平均値	被験者毎のアウトカムの差の平均値 被験者毎のアウトカムの差の標準偏差	対応のあるt検定 

EZRでは、1標本における生存曲線の評価における症例数設計ができない。
→ 例えば、SWOGのホームページから計算可能(後ほど説明)

EZRにおける症例数設計の内容

	タイトル	必要な情報	備考
比率	閾値奏効率，期待奏効率からのサンプルサイズの計算	臨：閾値奏効率，期待奏効率 統：有意水準，検出力	3とほぼ同じ (オプションが異なる)
	1群の比率の信頼区間をある幅におさめるためのサンプルサイズの計算	臨：想定する比率，信頼区間の幅 統：信頼係数(confidence level)	
	1群の比率を既知の比率と比較するためのサンプルサイズの計算	臨：既知の比率，想定する比率 統：有意水準，検出力	1とほぼ同じ (オプションが異なる)
	2群の比率の比較のためのサンプルサイズの計算	臨：グループ1の比率，グループ2の比率 統：有意水準，検出力，サンプルサイズの比	
	2群の比率の比較(非劣性)のためのサンプルサイズの計算	臨：各群の比率，臨床的に意味のある差*1 統：有意水準，検出力	
平均	1群の平均値の信頼区間をある幅におさめるためのサンプルサイズの計算	臨：想定する標準偏差，信頼区間の幅 統：信頼係数(confidence level)	信頼区間の幅に平均値は関係ない
	2群の平均値の比較のためのサンプルサイズの計算	臨：2群間の平均値の差，2群共通の標準偏差 統：有意水準，検出力，サンプルサイズの比	
	2群の平均の比較(非劣性)のためのサンプルサイズの計算	臨：平均の差，標準偏差，臨床的に意味のある差*1 統：有意水準，検出力	
	対応のある2群の平均値の比較のためのサンプルサイズの計算	臨：2群間の平均値の差*2，2群共通の標準偏差*2 統：有意水準，検出力	
生存曲線	2群の生存曲線の比較のためのサンプルサイズの計算	臨：登録期間，試験期間*3，年次，各群の生存率 統：有意水準，検出力，サンプルサイズの比	
	2群の生存曲線の比較(非劣性)のためのサンプルサイズの計算	臨：登録期間，試験期間*3，年次，各群の生存率，臨床的に意味のある差*1 統：有意水準，検出力，サンプルサイズの比	

*1：非劣性マージンと呼ばれる。非劣性試験において許容されるアウトカムの範囲を表す。

*2：正しくは，個々の被験者における差の平均値，差の標準偏差を意味する。

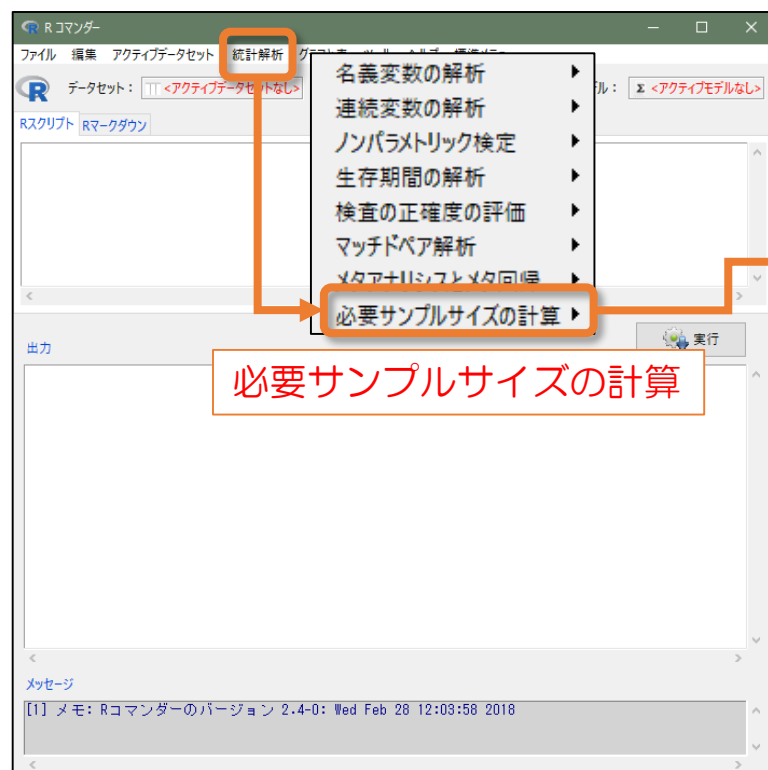
*3：試験期間とは，登録期間＋フォローアップ期間を表している。

症例設計の例示：先ほどのII相試験を例に...

(前略) TS-1+CDDP+Trastuzumab併用療法の閾値奏効率を35%、期待奏効率を50%であると仮定する。

帰無仮説「TS-1+CDDP+Tmab併用療法の奏効率は35%である」に対して、片側対立仮説「TS-1+CDDP+Tmab併用療法の奏効率は35%を上回る」を有意水準 $\alpha=0.1$ のもとで母比率の検定で評価するとき(片側対立仮説)、検出力 $1-\beta$ が0.8以上になる必要最小被験者数(後略)

[STEP.1] メニュー「統計解析」から、以下の手順を実行する



必要サンプルサイズの計算

閾値奏効率、期待奏効率からのサンプルサイズの計算
1群の比率の信頼区間をある幅におさめるためのサンプルサイズの計算
1群の比率を既知の比率と比較するためのサンプルサイズの計算
1群の比率を既知の比率と比較するための検出力の計算

1群の比率を既知の比率と比較するための
サンプルサイズの計算

2群の比率の
2群の比率の
2群の比率の比較(非劣性)のためのサンプルサイズの計算

1群の平均値の信頼区間をある幅におさめるためのサンプルサイズの計算
2群の平均値の比較のためのサンプルサイズの計算
2群の平均値の比較のための検出力の計算
2群の平均の比較(非劣性)のためのサンプルサイズの計算

対応のある2群の平均値の比較のためのサンプルサイズの計算
対応のある2群の平均値の比較のための検出力の計算

2群の生存曲線の比較のためのサンプルサイズの計算
2群の生存曲線の比較のための検出力の計算
2群の生存曲線の比較(非劣性)のためのサンプルサイズの計算

症例設計の例示：先ほどのII相試験を例に...

試験デザイン：Phase II Study (単アーム)

第1種の過誤(有意水準)： $\alpha = 0.1$

主要評価項目：奏効割合 (RR: Response Rate)

1－第2種の過誤(検出力)： $1-\beta = 0.8$ (80%)

閾値奏効率：35%， 期待奏効率：50%

入力前

R 1群の比率を既知の比率と比較するためのサンプルサ... X

既知の比率 (0.0-1.0)	
想定する比率 (0.0-1.0)	
α エラー (0.0-1.0)	0.05
検出力(1- β エラー) (0.0-1.0)	0.80

解析方法

☒ 両側

☐ One-sided

カイ2乗検定の連続性補正

☒ はい(あるいは正確検定)

☐ いいえ

OK キャンセル

入力後

R 1群の比率を既知の比率と比較するためのサンプルサ... X

既知の比率 (0.0-1.0)	0.35
想定する比率 (0.0-1.0)	0.50
α エラー (0.0-1.0)	0.1
検出力(1- β エラー) (0.0-1.0)	0.80

解析方法

☐ 両側

☒ One-sided

カイ2乗検定の連続性補正

☐ はい(あるいは正確検定)

☒ いいえ

OK キャンセル

連続性補正を行うほうが精度が上がるものの、必要症例数が多くなる。

症例設計の例示：先ほどのII相試験を例に...

> #####1群の比率を既知の比率と比較するためのサンプルサイズの計算#####

```
> windows(width=7, height=7); par(lwd=1, las=1, family="sans", cex=1, mgp=c(3.0, 1,  
+ 0))
```

```
> SampleProportionSingleArm(0.35, 0.5, 0.1, 0.80, 1, 0)
```

	仮定
想定する比率	0.35
比較する比率	0.5
α エラー	0.1
	片側検定
検出力	0.8

	計算結果
必要サンプルサイズ	48

単アーム試験における生存曲線に基づく症例数設計

いま、多発性骨髄腫に対する中央全生存期間(MST)が46ヵ月であることが報告されている。今回、新たな治療薬が開発され、MSTが60ヵ月まで延長することが期待されている。登録期間3年、フォローアップ期間5年の片側対立仮説のもとで、有意水準 $\alpha=0.05$ 、検出力 $1-\beta=80\%$ での必要症例数を計算しなさい。

試験デザイン：Phase II Study (単アーム)

第1種の過誤(有意水準)： $\alpha = 0.05$

主要評価項目：全生存期間(OS)

1－第2種の過誤(検出力)： $1-\beta = 0.8$ (80%)

閾値MST：46M, 期待奏効率：60M, 登録期間：36M(3yr.), フォローアップ期間：60M(5yr.)

SWOGのホームページのスクリプトを用いて解析する

<https://stattools.crab.org/>

単アーム試験における生存曲線に基づく症例数設計

CRAB
CANCER RESEARCH
AND BIOSTATISTICS

STATISTICAL TOOLS DESIGN ANALYSIS PROBABILITIES ABOUT US

Statistical Tools

One Arm Survivalを選択

One Arm Survival

One Arm Normal

One Arm Non-Parametric Survival

Two Stage

Two Arm Binomial

Two Arm Survival

Two Arm Normal

Binomial Interaction

Survival Interaction

Survival Noninferiority

Expected Deaths

Frequency Table

Binomial Confidence Interval

Normal

Binomial

Poisson

Chi-Square

SWOG
Leading cancer research. Together.

CRAB
CANCER RESEARCH
AND BIOSTATISTICS

STATISTICAL TOOLS DESIGN ANALYSIS PROBABILITIES ABOUT US

One Arm Survival

One Arm Survival is an interactive program to calculate either estimates of accrual or power for null and alternative median survival rates assuming an exponential distribution. A cube root transformation of the hazard rate is used in the calculations to get good small sample properties. The program gives critical values for either median survival or survival probability.

User Input

片側対立仮説 (1 Sided)を選択

中央生存期間(Median Survival)を選択

Select Calculation, Test type and Parameter of Interest

☒ Sample Size
☐ Power

☒ 1 Sided
☐ 2 Sided

☒ Median Survival
☐ Survival Probability

Select Study and Hypothesis Test Parameters

Accrual Time 36

登録期間 36(M)を入力

Follow-up Time 60

フォローアップ 期間 60(M)を入力

Alpha .05

有意水準0.05を入力

Null Median Survival 46

閾値MSTの46(M)を入力

Alt Median Survival 60

期待MSTの60(M)を入力

Null Survival Prob

Alt Survival Prob

At Time

年次生存割合の場合はこちら側に入力
Null Survival Prob：閾値生存率
Alt Survival Prob：期待生存率
At time：生存期間(1年生存率の場合には1 or 12(どちらでも結果は変わらない))

検出力0.8 (80%)を入力

Power .80

Sample Size 138

Approx Lower Critical Value 0

Approx Upper Critical Value 54.73

Calculate

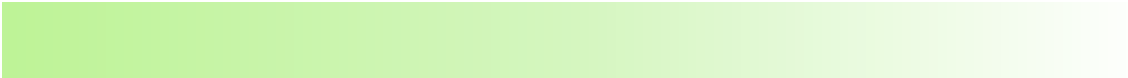
クリックする

Help Document

必要最小症例数(Sample Size)は138例である。



臨床検査データ

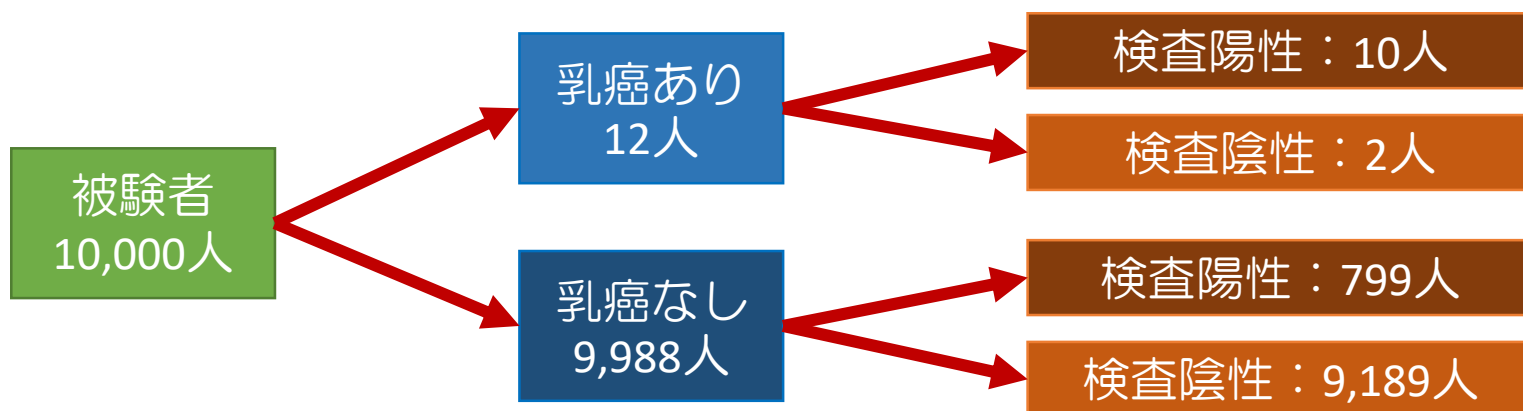


例示：マンモグラフィ検査による乳癌検査の仮想例(新谷, 2015)

乳癌に対するマンモグラフィ検査の予測確度を評価するために、10,000人の検査結果と乳癌の有無を調査した。

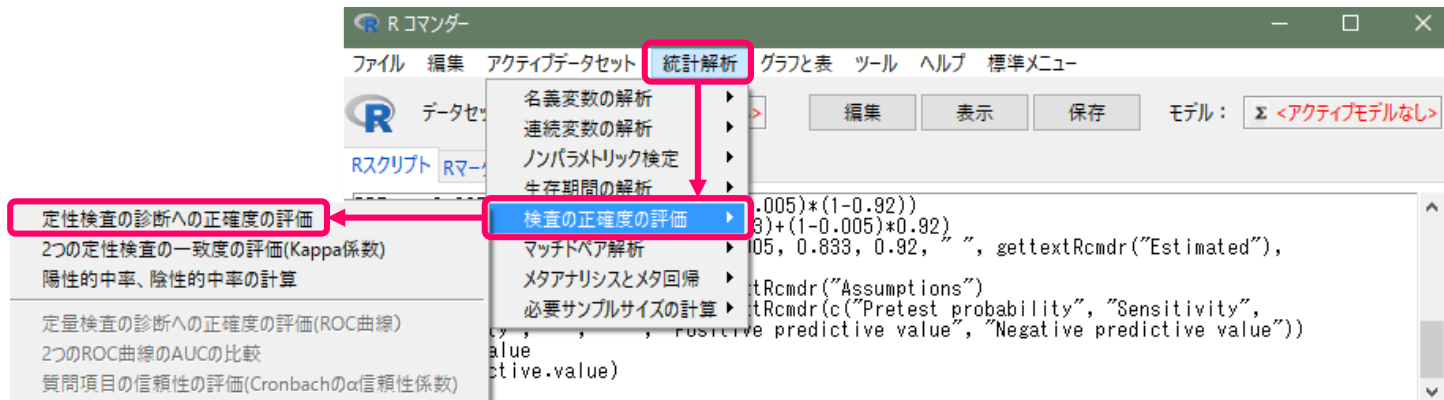
	乳癌あり	乳癌なし	合計
検査陽性	10	799	809
検査陰性	2	9,189	9,191
合計	12	9,988	10,000

乳癌とマンモグラフィ検査の時間的な流れ



EZRによる定性検査データの要約

[STEP.1] 「統計解析」→「検査の正確度の評価」→「定性検査の診断への正確度の評価」を選択



[STEP.2] クロス集計表のデータを入力

定性検査の診断への正確度の評価

サンプル数を入力 疾患陽性 疾患陰性

検査陽性	10	799
検査陰性	2	9189

ヘルプ OK キャンセル



	乳癌あり	乳癌なし
検査陽性	10	799
検査陰性	2	9,189

EZRの結果

```
> remove(predictive.value)

> #####定性検査の診断への正確度の評価#####

> .Table <- matrix(c(10, 799, 2, 9189), 2, 2, byrow=TRUE)

> epi.tests(.Table, conf.level = 0.95)
```

Rのコマンドなので無視

	疾患陽性	疾患陰性	計
検査陽性	10	799	809
検査陰性	2	9189	9191
計	12	9988	10000

アウトプット

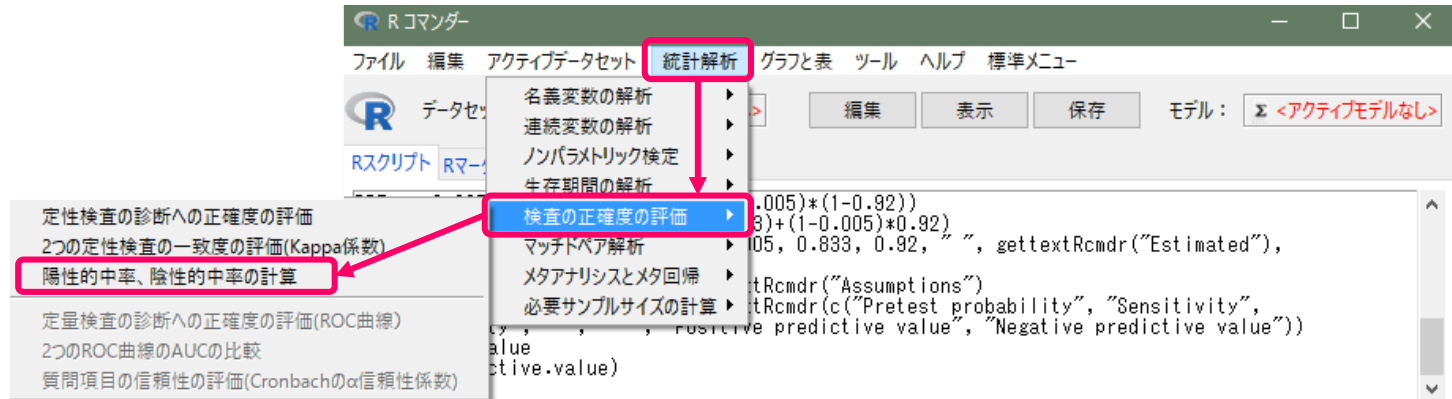
点推定と 95 % 信頼区間

	推定値	信頼区間下限	信頼区間上限
検査の陽性率	0.081	0.076	0.086
真の有病率	0.001	0.001	0.002
感度	0.833	0.516	0.979
特異度	0.920	0.915	0.925
陽性的中率	0.012	0.006	0.023
陰性的中率	1.000	0.999	1.000
診断精度	0.920	0.914	0.925
陽性尤度比	10.417	8.019	13.532
陰性尤度比	0.181	0.051	0.642

EZRによる陽性的中率・陰性的中率の計算(有病率が存在する場合)

有症率を例えば5%と考えた場合の計算方法は以下のとおり

[STEP.1] 「統計解析」→「検査の正確度の評価」→「陽性的中度、陰性的中度の計算」を選択



[STEP.2] テスト前確率(有症率), 感度, 特異度を入力

The dialog box titled '陽性的中率、陰性的中率の計算' (Calculate Positive Predictive Value, Negative Predictive Value) is shown. It contains three input fields: 'テスト前確率(0-1)' (Pre-test probability) with the value 0.05, '感度(0-1)' (Sensitivity) with the value 0.833, and '特異度(0-1)' (Specificity) with the value 0.920. There are 'OK' and 'キャンセル' (Cancel) buttons at the bottom.

テスト前確率(有症率) = 0.05(5%)

感度 = 0.833 (83.3%)

特異度 = 0.920 (92.0%)

EZRの結果

```
> #####陽性的中率、陰性的中率の計算#####
```

```
> windows(width=7, height=7); par(lwd=1, las=1, family="sans",  
cex=1, mgp=c(3.0, 1, 0))
```

```
> predictive.value
```

Rのコマンドなので無視

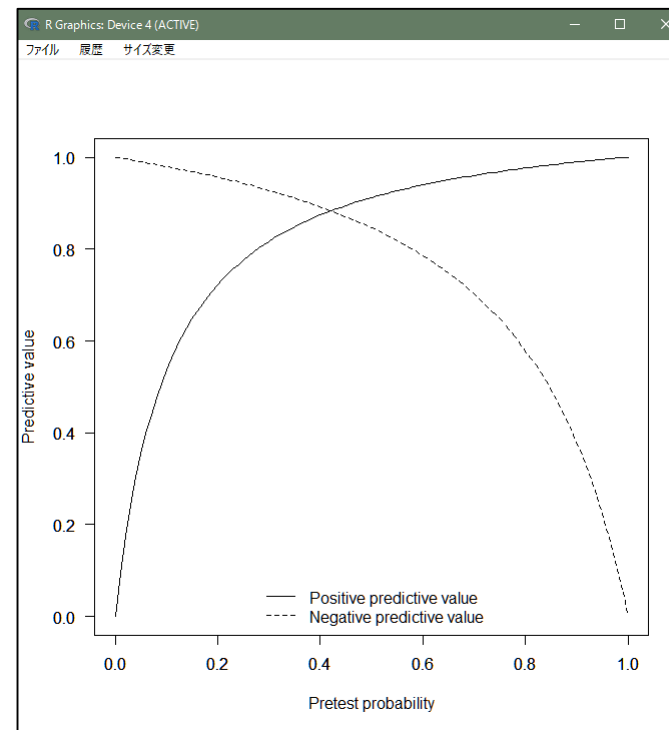
	仮定
テスト前確率 (0-1)	0.05
感度	0.833
特異度	0.92

計算結果

陽性的中率	0.354
陰性的中率	0.991

アウトプット

陽性的中率・陰性的中率の推移のグラフ



「右クリック」→「メタファイルにコピー」で
コピー可能(画質が変化しない)

二つの定性検査の類似性の評価：Kappa係数

2人の病理医による非小細胞肺癌の組織標本75枚の組織学的分類の結果(Gardis, 2009)

		病理医 A		
病理医 B		Grade II	Grade III	合計
	Grade II	41	3	3
	Grade III	4	27	27
	合計	45	30	75

Kappa係数の計算式

観測度数

病理医 A

病理医 B

	Grade II	Grade III	合計
Grade II	41	3	44
Grade III	4	27	31
合計	45	30	75

$$\text{測定者の全一致率} = \frac{41 + 27}{75} = 0.907$$

期待度数

病理医 A

病理医 B

	Grade II	Grade III	合計
Grade II	$\frac{45 \times 44}{75} = 26.4$	$\frac{30 \times 44}{75} = 17.6$	44
Grade III	$\frac{45 \times 31}{75} = 18.6$	$\frac{30 \times 31}{75} = 12.4$	31
合計	45	30	75

$$\text{偶然による一致率} = \frac{26.4 + 12.4}{75} = 0.517$$

測定者間の一致率が偶然によって期待される一致率よりどれくらい大きいのか



$$\text{Kappa係数} = \frac{\text{測定者の全一致率} - \text{偶然による一致率}}{1 - \text{偶然による一致率}}$$



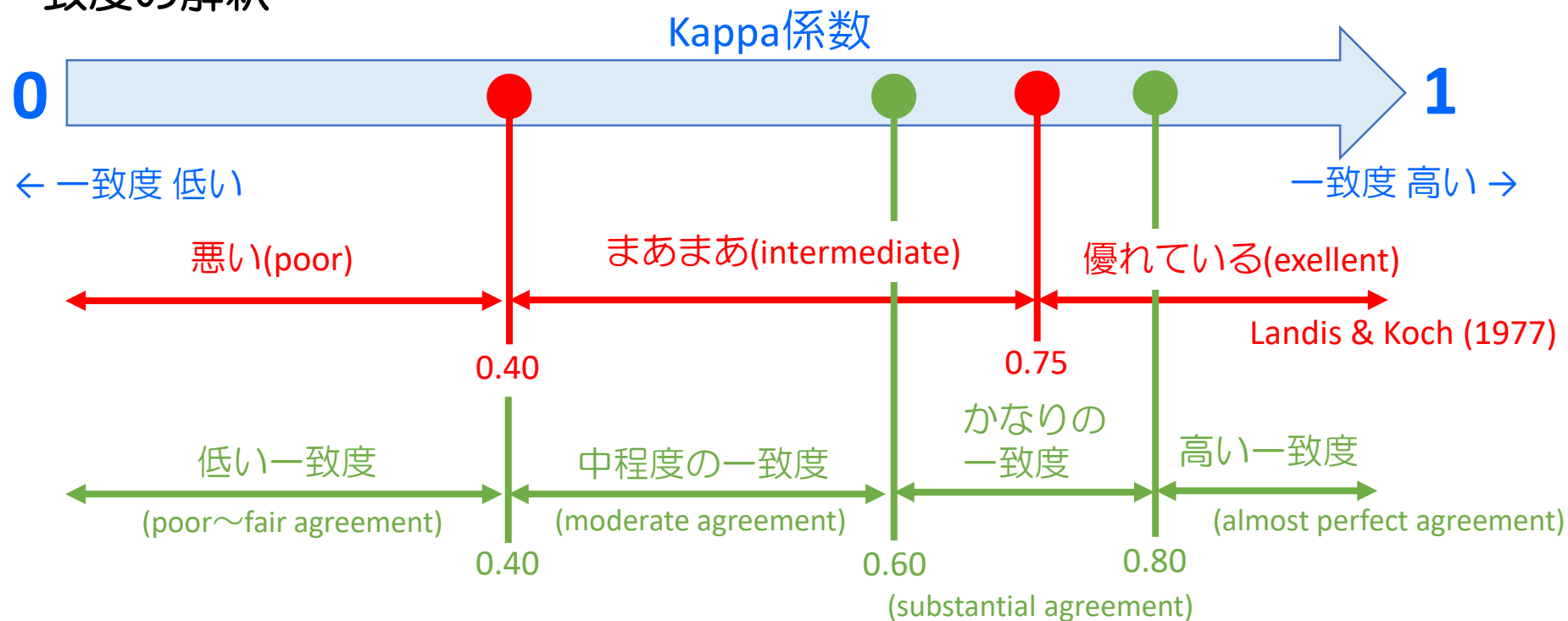
2人の測定者の一致率が偶然による一致を含めずに最大限取り得る値はいくらか

Kappa係数の計算とその結果

$$\text{Kappa係数} = \frac{\text{測定者の全一致率} - \text{偶然による一致率}}{1 - \text{偶然による一致率}}$$

$$\text{Kappa係数} = \frac{0.907 - 0.517}{1.000 - 0.517} = 0.810$$

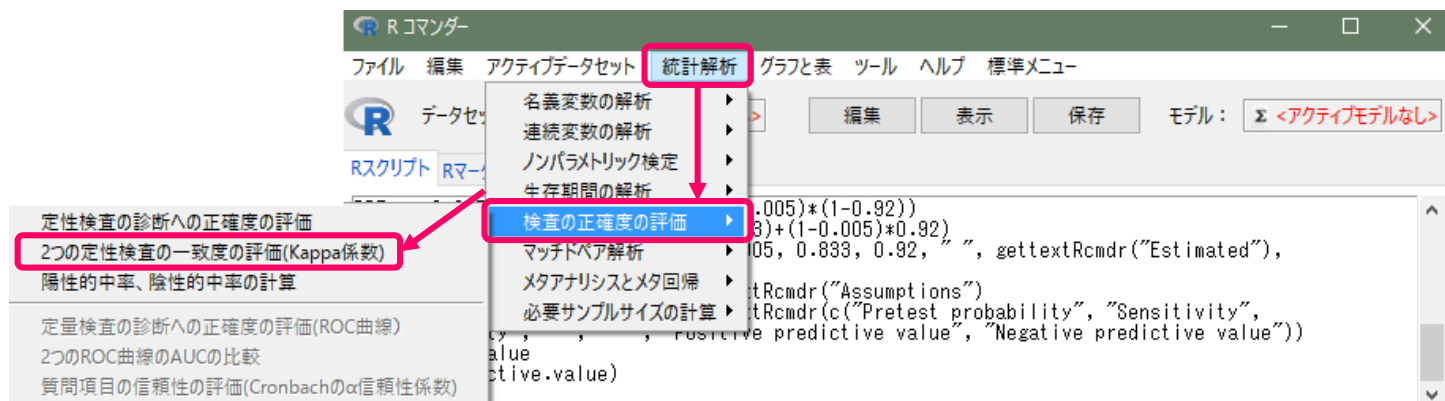
一致度の解釈



事例では高い一致度が示されている

EZRによるKappa係数の計算

[STEP.1] 「統計解析」→「検査の正確度の評価」→「2つの定性検査の一致度の評価(Kappa係数)」を選択



[STEP.2] クロス集計表のデータを入力

2つの定性検査の一致度の評価(Kappa係数)

サンプル数を入力 検査2陽性 検査2陰性

検査1陽性	41	3
検査1陰性	4	27

ヘルプ OK キャンセル

病理科 A		
	Grade II	Grade III
Grade II	41	3
Grade III	4	27

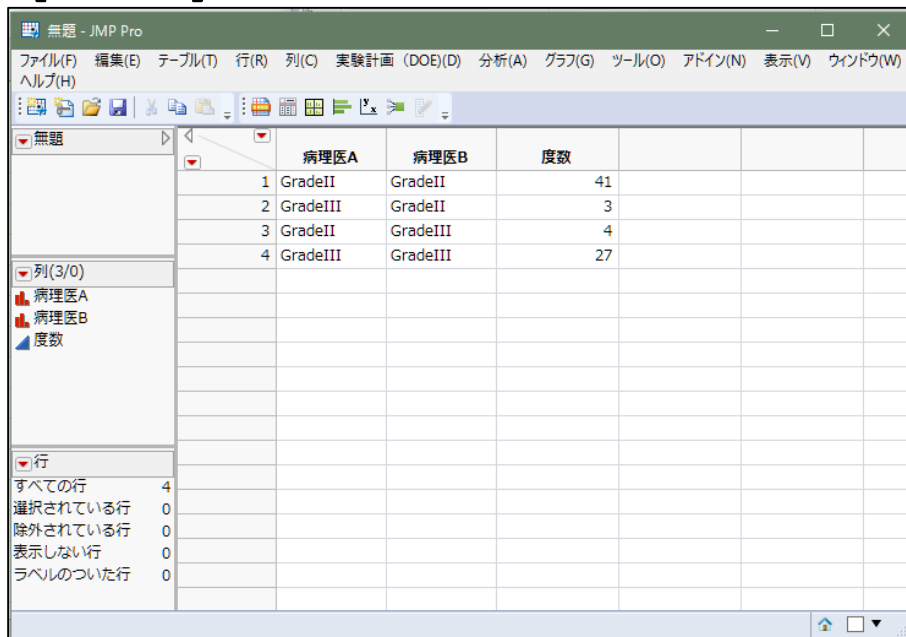
← 病理科 B

アウトプット

```
> res[1]
$kappa
      点推定値 信頼区間下限 信頼区間上限
1 0.8066298    0.6702305    0.9430292
```

EZRによるKappa係数の計算

[STEP.1] 次のように入力



無題 - JMP Pro

	病理医A	病理医B	度数
1	GradeII	GradeII	41
2	GradeIII	GradeII	3
3	GradeII	GradeIII	4
4	GradeIII	GradeIII	27

列(3/0)
病理医A
病理医B
度数

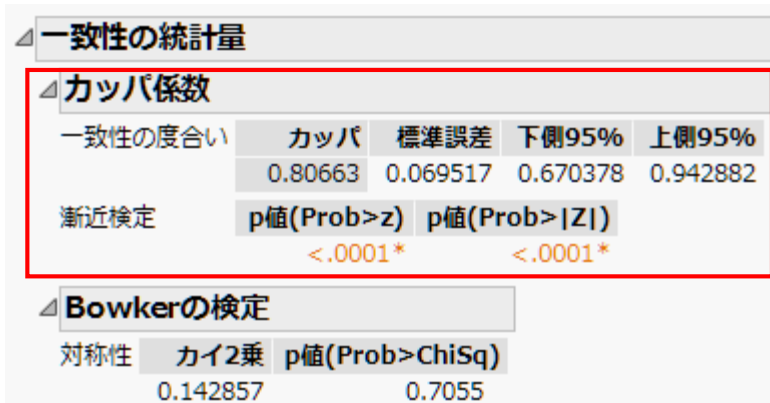
行
すべての行 4
選択されている行 0
除外されている行 0
表示しない行 0
ラベルのついた行 0

[STEP.2] 「分析」→「二変量の関係」

Y, 目的変数：病理医A, X, 説明変数：病理医B
度数：度数



[STEP.3] 「▼」から「一致性の統計量」を選択



一致性の統計量

Kappa係数

一致性の度合い	カッパ	標準誤差	下側95%	上側95%
	0.80663	0.069517	0.670378	0.942882

漸近検定

p値(Prob>z)	p値(Prob> Z)
<.0001*	<.0001*

Bowkerの検定

対称性	カイ2乗	p値(Prob>ChiSq)
	0.142857	0.7055

アウトプット

漸近検定は、
帰無仮説 H_0 「Kappa係数は0である」
対立仮説 H_1 「Kappa係数は0でない」
を検定している。

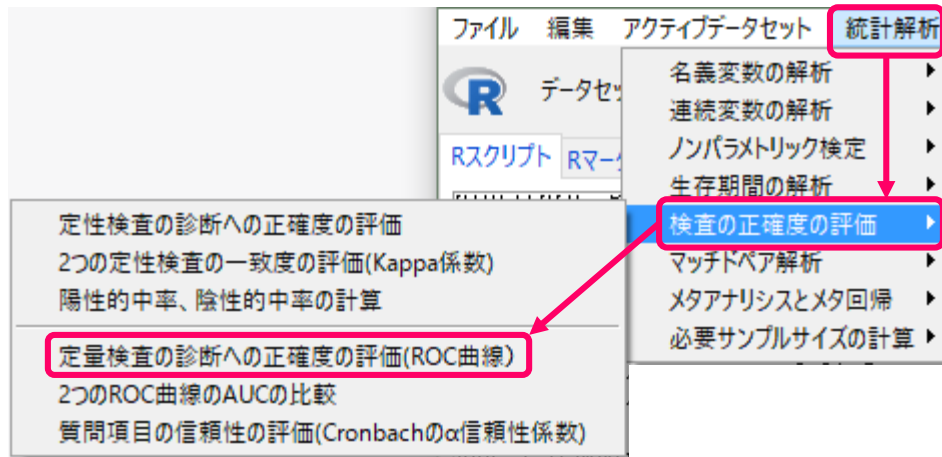
事例：頭部外傷症データ (Zhou, et al., 2002)

頭部外傷症の重篤度を識別するために、CK-BB(クレアチン・キナーゼBB)が有効か否かを判定している。ここに、重篤度は、重度および非重度の2値とする。Zhou *et al.*(2002)は、CK-BBが成人での重篤度の判定だけでなく、未成人においても利用可能であるか否かを検討している。

重症群					非重症群		
140	740	543	490	523	136	60	46
1087	126	913	156	76	286	17	
230	153	230	356	303	281	27	
183	283	463	350	353	23	126	
1256	90	60	323	206	200	100	
700	303	509	1560		146	253	
16	193	576	120		220	70	
800	76	671	216		96	40	
253	1370	80	443		100	6	

EZRによるROC曲線の描写

[STEP.1] 「統計解析」→「検査の正確度の評価」→「定量検査の診断への正確度の評価(ROC曲線)」を選択



[STEP.2] 以下を設定してOKボタンを押す

The dialog box '定量検査の診断への正確度の評価(ROC曲線)' (Evaluation of Diagnostic Accuracy for Quantitative Tests (ROC Curve)) is shown. The '結果(値が0か1の項目を1つ選択)' (Result (Select one item with value 0 or 1)) dropdown is set to 'グループ' (Group). The '予測に用いる値(1つ選択)' (Value used for prediction (Select one)) dropdown is set to '観測値' (Observed value). The '陽性の判定基準' (Positive decision criterion) is set to '自動' (Automatic). The 'ベストの閾値の判定基準' (Criterion for the best threshold) is set to '感度と特異度の和が最大になる閾値' (Threshold that maximizes the sum of sensitivity and specificity). The '偽陽性に対する偽陰性の重み付け' (Weighting for false positive vs false negative) is set to 1. The '有病率(0と1の間の値)' (Prevalence (Value between 0 and 1)) is set to 0.5. The 'ヘルプ' (Help) button is highlighted.

結果：グループ変数
予測に用いる値：検査値

最適なカットオフ値の選定

EZRの結果

> ### ↑ 予測値が閾値「以上」を陽性と判定した場合の感度、特異度を示す。

> ROC

Call:

```
roc.formula(formula = グループ ~ 観測値, data = Dataset, ci = TRUE, direction = "auto")
```

Data: 観測値 in 19 controls (グループ not-serious) < 41 cases (グループ serious).

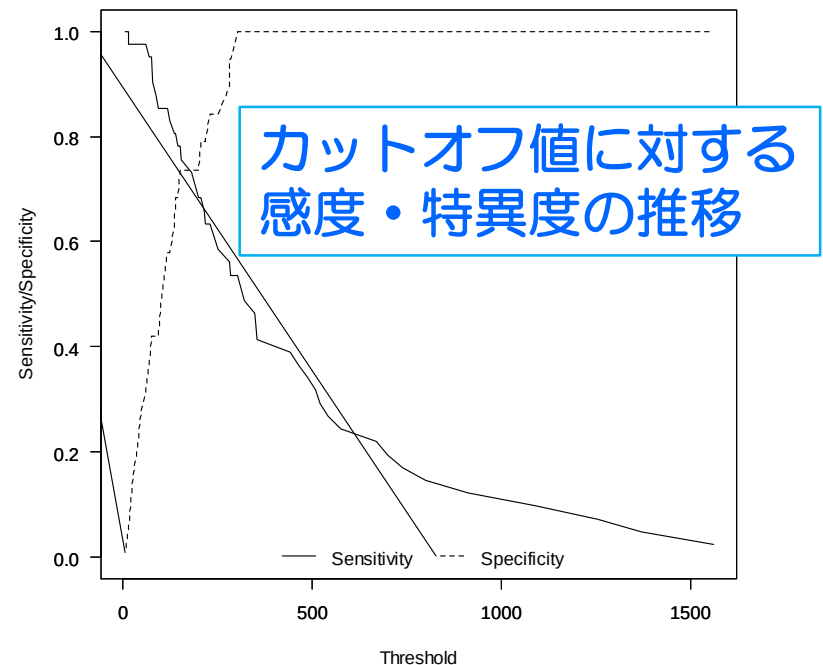
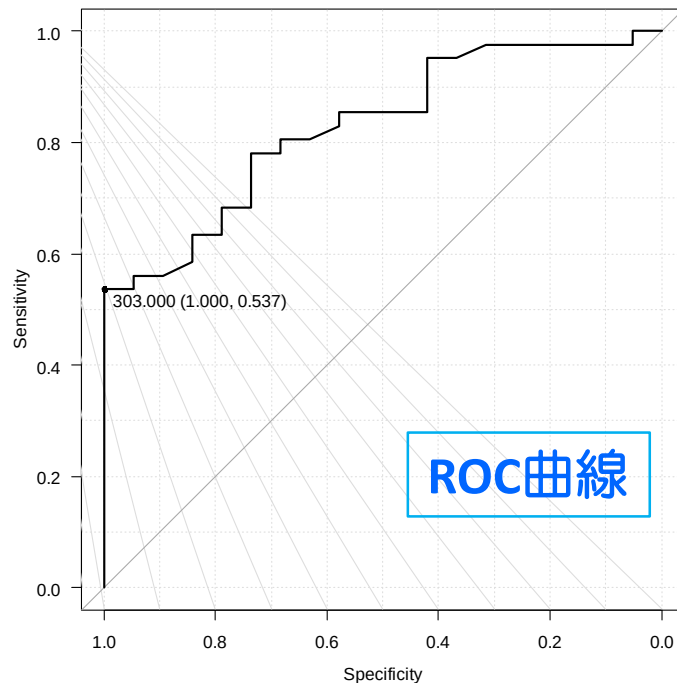
Area under the curve: 0.8286

95% CI: 0.7257-0.9316 (DeLong)

```
> cat(gettextRcmdr("Area under the curve"), signif(ROC$auc[1], digits=3),  
+     gettextRcmdr("95% CI"), signif(ROC$ci[1], digits=3), "-", signif(ROC$ci[3], digits=3), "  
+ ")
```

曲線下面積 0.829 95%信頼区間 0.726 - 0.932

ROC曲線の曲線下面積



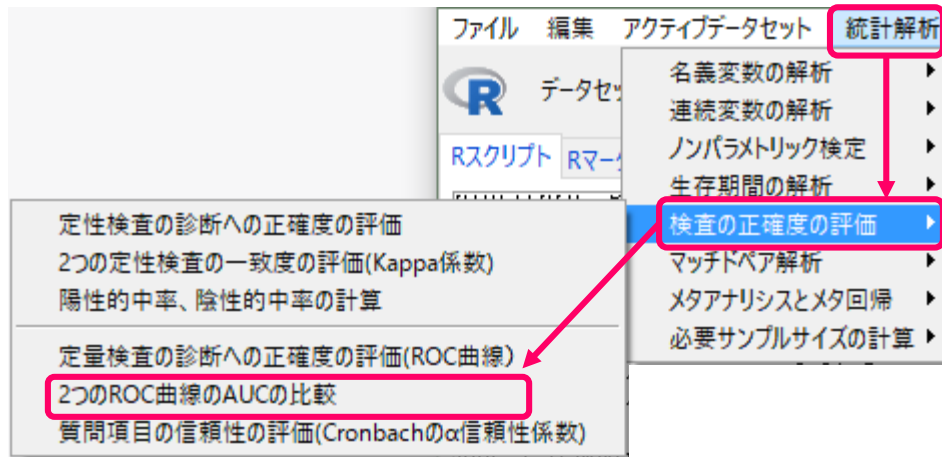
事例検討：仮想データ

これは、ある疾患に対する2種類の検査方法に関するデータである。二つの検査方法の診断能に違いがあるか (対応のあるデータ)

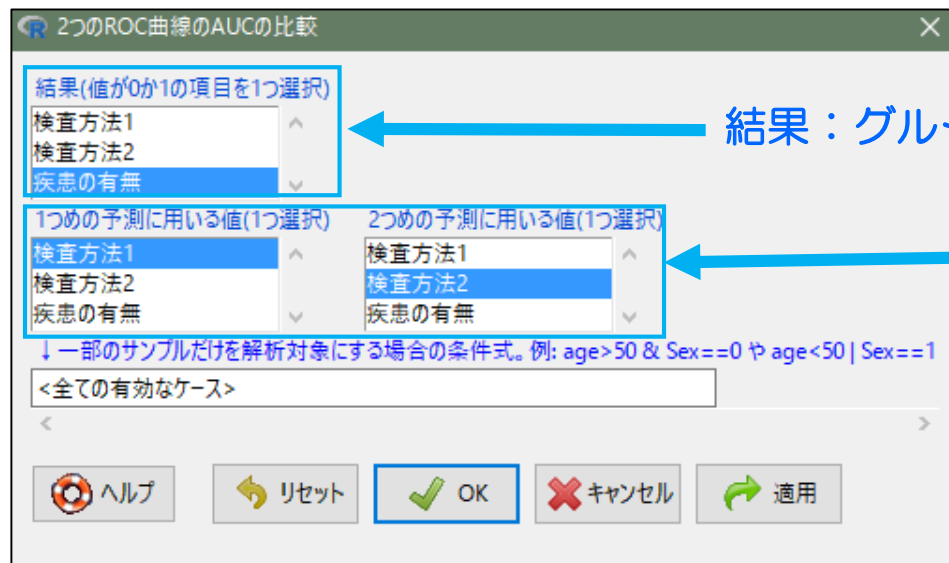
疾患の有無	検査方法1	検査方法2	疾患の有無	検査方法1	検査方法2
あり	59	95	なし	63	93
あり	61	123	なし	64	99
あり	55	74	なし	65	119
あり	66	145	なし	64	92
あり	52	64	なし	68	112
なし	60	84	なし	64	99
なし	61	128	なし	69	113
なし	51	79	あり	62	92
あり	60	112	あり	64	112
あり	61	107	なし	67	128
あり	56	67	なし	65	111
なし	65	98	なし	66	105
なし	63	105	なし	62	104
なし	58	95	なし	66	106
なし	59	79	あり	65	112
あり	61	81	あり	60	115
あり	62	91	なし	68	128
あり	65	142	あり	62	116
あり	63	84	なし	68	134
あり	62	85	なし	70	172

EZRによるROC曲線のAUCの比較

[STEP.1] 「統計解析」→「検査の正確度の評価」→「2つのROC曲線のAUCの比較」を選択



[STEP.2] 以下を設定してOKボタンを押す



EZRでの結果

```
> #####2つのROC曲線のAUCの比較#####  
> ROC1 <- roc(疾患の有無~検査方法1, data=Dataset, ci=TRUE)  
> ROC2 <- roc(疾患の有無~検査方法2, data=Dataset, ci=TRUE)  
> windows(width=7, height=7); par(lwd=1, las=1, family="sans", cex=1, mgp=c(3.0, 1, 0))  
> plot(ROC1, lty=1)  
> plot(ROC2, lty=2, add=TRUE)  
> legend("bottomright", c("検査方法1", "検査方法2"), lty=1:2, box.lty=0)  
> (res <- roc.test(疾患の有無~検査方法1+検査方法2, data=Dataset))
```

DeLong's test for two correlated ROC curves

data: 検査方法1 and 検査方法2 in Dataset by 疾患の有無（あり, なし）

Z = 1.9255, p-value = 0.05417

alternative hypothesis: true difference in AUC is not equal to 0

sample estimates:

AUC of roc1	AUC of roc2
0.7361111	0.5782828

```
> roc.table <- signif(cbind(res$estimate, res$p.value), digits=3)  
> rownames(roc.table) <- c("検査方法1", "検査方法2")  
> colnames(roc.table) <- gettextRcmdr(c("Area under the curve", "p.value"))  
> roc.table[2,2] <- ""  
> data.frame(roc.table)
```

	曲線下面積	P値
検査方法1	0.736	0.0542
検査方法2	0.578	

AUCの差の検定の結果



医学統計学 Q&A



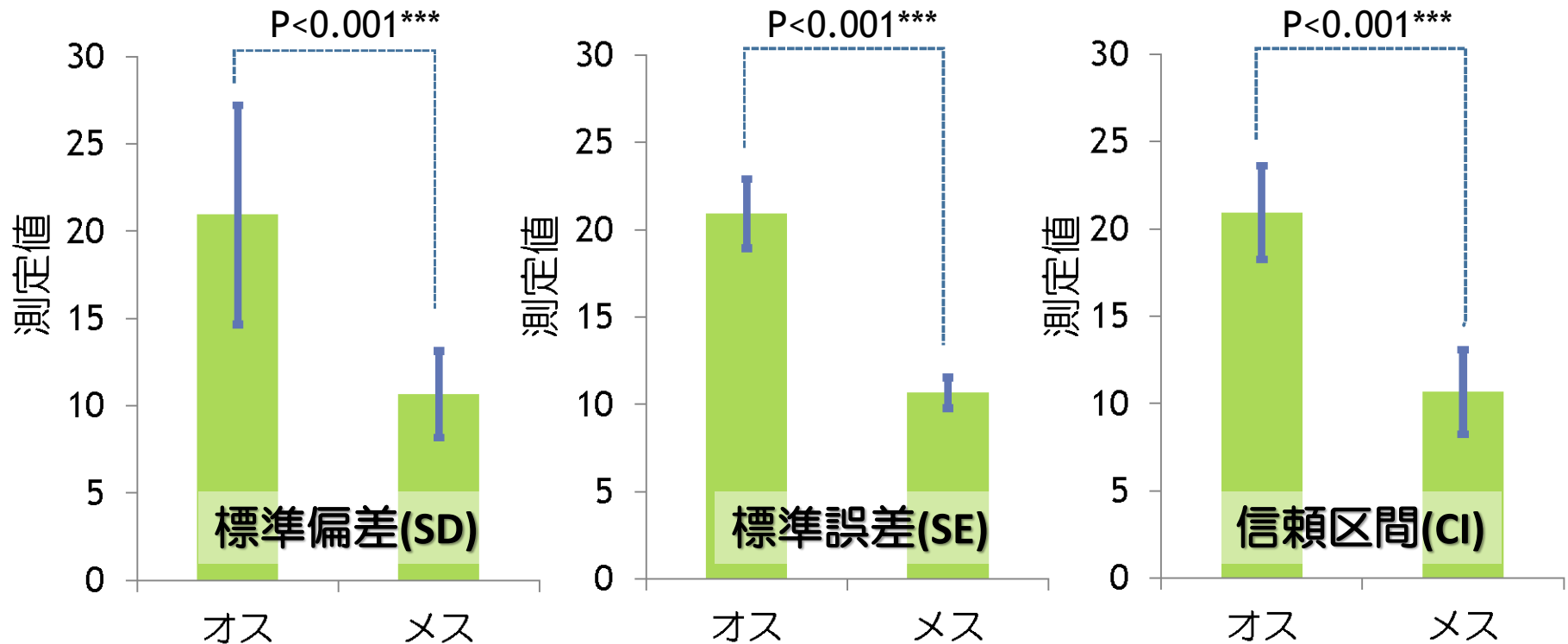
医学統計学Q & A

- **QA1**：平均値に対する要約指標に標準偏差，標準誤差，95%CIのどれを用いる？
- **QA2**：仮説検定って何をやっているの？
- **QA3**：有意でないときの解釈は？ 片側対立仮説が有意で，両側対立仮説が有意でない場合はどうすばよい(対立仮説の使い分け)？
- **QA4**：多変量解析における変数選択の注意点は？
- **QA5**：対応のあるクロス集計表における「一致性」と「差異」を見ることが統計的方法の違いは？
- **QA6**：多重比較って何？

統計学の基礎におけるQ&A(1)：連続データのバラツキの意味



連続データにおいて、標準偏差、標準誤差、信頼区間のどれを平均値に対するバラツキの尺度に用いる？



雄雌のラットの腸内から採取した標本からある測定値を測定したデータ
(Raisman & Field, 1971)

統計学の基礎におけるQ&A(1)：連続データのバラツキの意味

標準偏差(SD)の利用

- 観測値のばらつきを表すのに用いる.
- 例えば, 薬剤の有効性を示すとき, その一般可能性が重要になることがある. SDを用いることは, 個体間でその効果に大きな違いがないことをしめすことができる.

標準誤差(SE)の利用

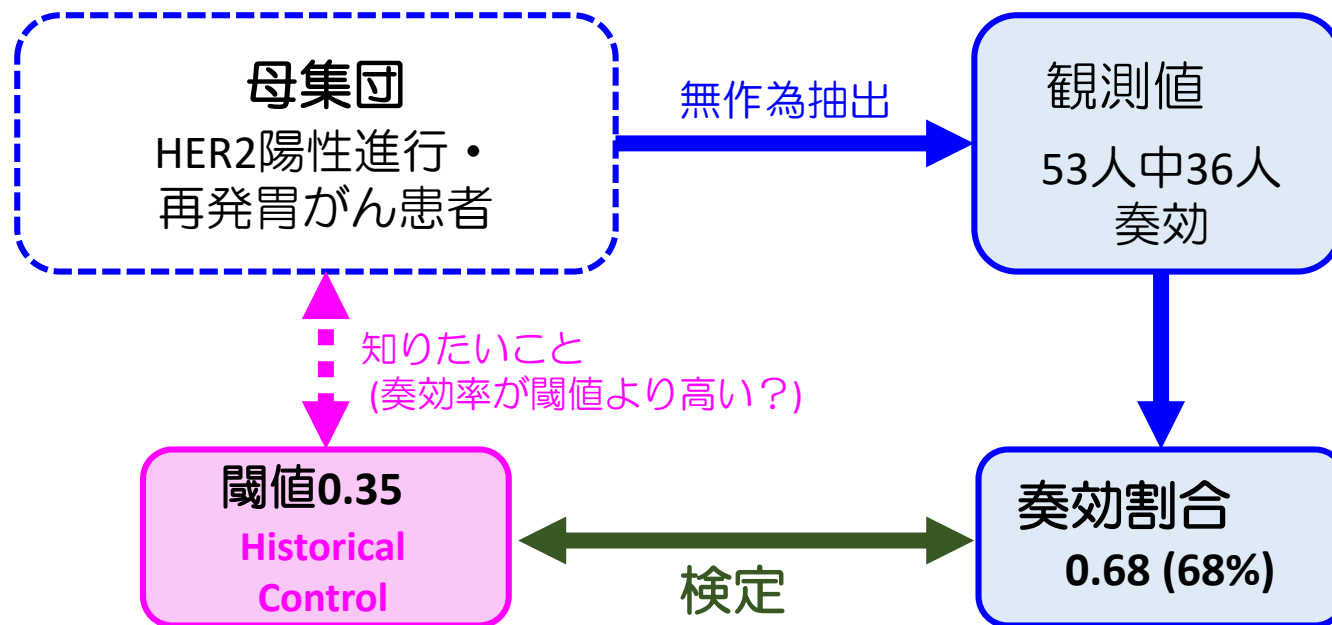
- 平均値の信頼性を表すのに用いる.
- 例えば, ラットに対する毒性試験などにおいて, 平均的な毒性の大きさを知ること考えるとする. SEを用いることは, 平均毒性の信頼性を示すのに用いる.

信頼区間の利用

- 真のパラメータに対する区間として構成される.
- 用途は標準誤差に類似しているが, エンドポイントの要約に用いることが多い.

統計学の基礎におけるQ&A(2)：仮説検定の意味

HER2陽性の進行・再発胃癌患者に対するフッ化ピリミジン系抗がん剤とCDDPを併用したレジメンでの奏効率が35%であると報告されている。新しい臨床試験では、S1+CDDP+Tmabの併用療法による新たなレジメンの有効性を単アーム試験で実施した。その結果、53例中36例の奏効例が認められた(奏効割合=68%) (Kurokawa et al., *BJC*, 2014)



帰無仮説 H_0 (本当は言いたいことと反対の仮説)

「S1+CDDP+Tmabの併用療法による奏効率は35%である(上回らない)」

対立仮説 H_1 (本当は言いたいこと)

「S1+CDDP+Tmabの併用療法による奏効率は35%を上回る」

統計学の基礎におけるQ&A(2)：仮説検定の意味

帰無仮説 H_0 (本当は言いたいことと反対の仮説)

「S1+CDDP+Tmabの併用療法による奏効率は35%である(上回らない)」

■ これが正しいと仮定したとき、今回の試験結果が得られる確率 → **p値**

p値が小さいとは、「**帰無仮説が正しいときに、このような試験結果が得られる確率は非常に小さい**」ことを意味する。

上記を言いかえると、「帰無仮説と逆の仮説である対立仮説が正しい」と結論付けることになる。

このような判断を行うには、「p値がどの程度小さくなければいけないか(閾値)」を予め決めなければならない。

■ 帰無仮説が誤っている(有意である)かどうかを判断するための閾値

→ **有意水準 α**

p値<0.001であることから、高度に有意だった。すなわち、本治療レジメンは、HER2陽性の進行・再発胃癌患者に対して、非常に**有望**であることが証明された。

統計学の基礎におけるQ&A(3)：検定で有意でないときの解釈

帰無仮説を受容したからといって、帰無仮説を支持するということではない。つまり、帰無仮説を支持することはできない。

仮説検定において「帰無仮説を棄却しない」ということは、帰無仮説を棄却する根拠がないというだけであって、積極的に帰無仮説が正しいということを主張するものではない。



同等性試験は？ 同じと言っているのではないの？

ActiveがControlに比べてこれだけ
違いがあれば優れていれと判断する

優越性

ActiveがControlに比べてこれだけ
違う程度であれば劣らないと判
断する(非劣性マージン)

非劣性

検定 ←

→ 検定

同等性

ActiveとControlが同等と判断さ
れる

Control better

0

Active better

効果量

統計学の基礎におけるQ&A(3)：片側検定と両側検定

病院Cに通院する、腎機能障害の患者6名の血清クレアチニン濃度(mg/dl)を測定したところ

4.0 3.9 3.8 4.0 4.4 4.2

という観測値が得られた。これに対して、病院Bにおける、同じ腎機能障害の血清クレアチニン濃度の平均値は4.3(mg/dl)であった。病院Aと病院Bを受診した患者層が異なるといえるか。

■病院Cの標本平均： $\bar{x} = 4.05$

■病院Cの標準偏差： $s = 0.217$

両側対立仮説におけるp値が0.0931なので、有意水準 $\alpha=0.05$ よりも大きい。つまり、帰無仮説が受容された(有意でなかった)。

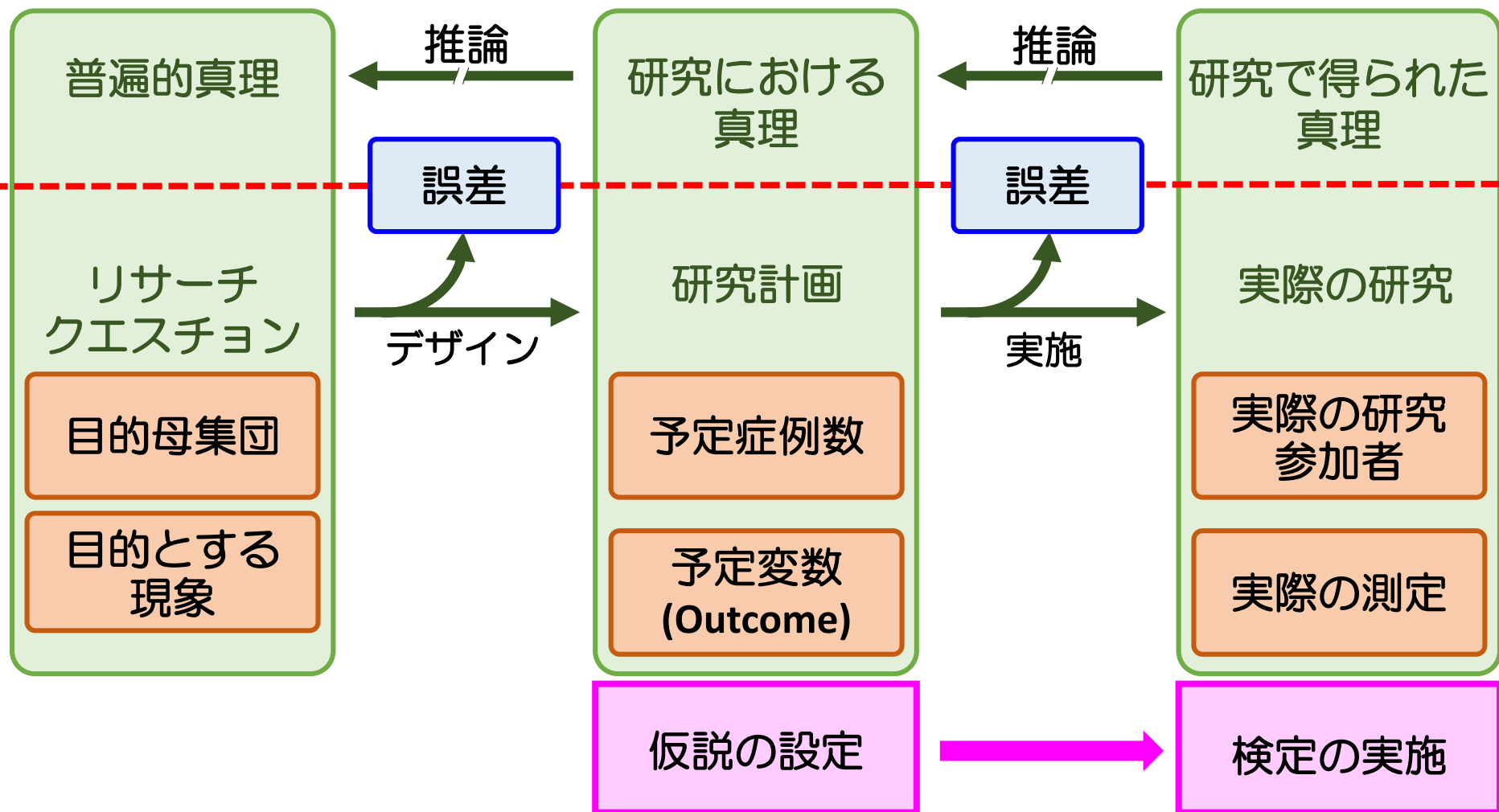
したがって、病院Cの血清クレアチニン濃度は病院Bの4.3(mg/dl)と異なるとはいえなかった。

一方で、「病院Cの血清クレアチニン濃度は4.3(mg/dl)よりも低い」という片側対立仮説では有意だった。

両側対立仮説ではなく、片側対立仮説の結果を利用してもよいか？

統計学の基礎におけるQ&A(3)：片側検定と両側検定

対立仮説は事前に設定しなければならない



両側・片側の選択は、研究計画で行われるため、実際の研究において、それを変更することは許されない。

統計学の基礎におけるQ&A(4)：多変量解析の基本

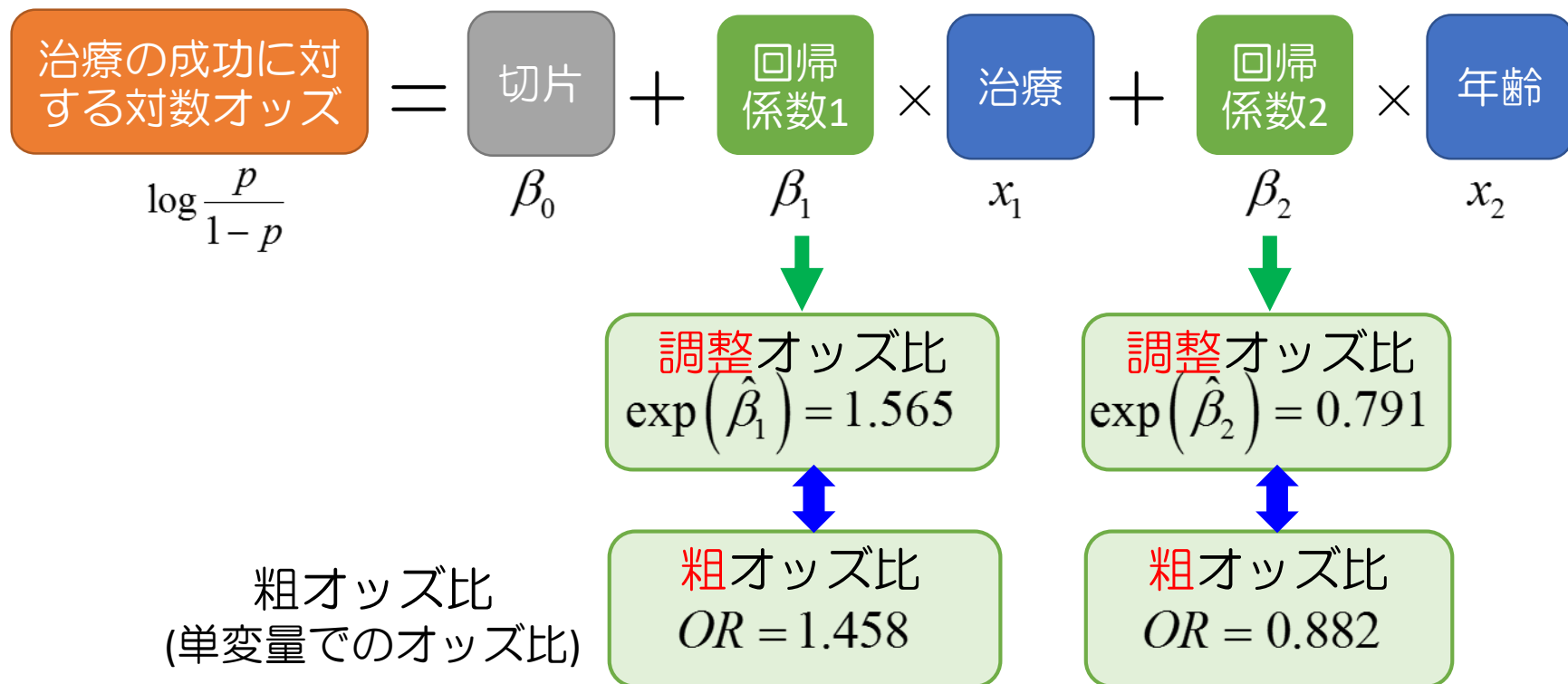
下の表は、ある疾患に対する治療法(既存治療：0，新規治療：1)と年齢(65歳未満：0，65歳以上：1)とその成功の有無(成功：1，失敗：0)に対する2個のクロス集計表である。

年齢	治療	成功/失敗		
		成功(1)	失敗(0)	合計
65歳以上(1)	新規(1)	444	567	1011
	既存(0)	56	150	206
	合計	500	717	1217
65歳未満(0)	新規(1)	286	324	610
	既存(0)	171	254	425
	合計	457	578	1035

(多重)ロジスティック回帰分析のモデルは

$$\begin{array}{ccccccc} \text{治療の成功に対する対数オッズ} & = & \text{切片} & + & \text{回帰係数1} \times \text{治療} & + & \text{回帰係数2} \times \text{年齢} \\ \log \frac{p}{1-p} & & \beta_0 & & \beta_1 & & \beta_2 \\ & & & & x_1 & & x_2 \\ & & & & \begin{array}{l} \text{既存治療：0} \\ \text{新規治療：1} \end{array} & & \begin{array}{l} \text{65歳未満：0} \\ \text{65歳以上：1} \end{array} \end{array}$$

統計学の基礎におけるQ&A(4)：多変量解析の基本



粗オッズ比は、他の要因(共変量)を何も考慮しない。これに対して、調整オッズ比では、

- 治療のオッズ比では年齢による違いを考慮
 - 年齢のオッズ比では治療による違いを考慮
- してオッズ比を計算している。

統計学の基礎におけるQ&A(4)：変数選択の種類

変数増加法 (前進ステップワイズ法)

説明
変数

1個



2個



3個



1. 変数が1個のモデルから始める.
2. 追加する変数の回帰係数が0であるという帰無仮説のもとで検定を行い, 最も検定統計量大きい変数を追加する.
3. 2を検定が有意でなくなるまで繰り返す.

変数減少法 (後退ステップワイズ法)

説明
変数

p 個



$p-1$ 個



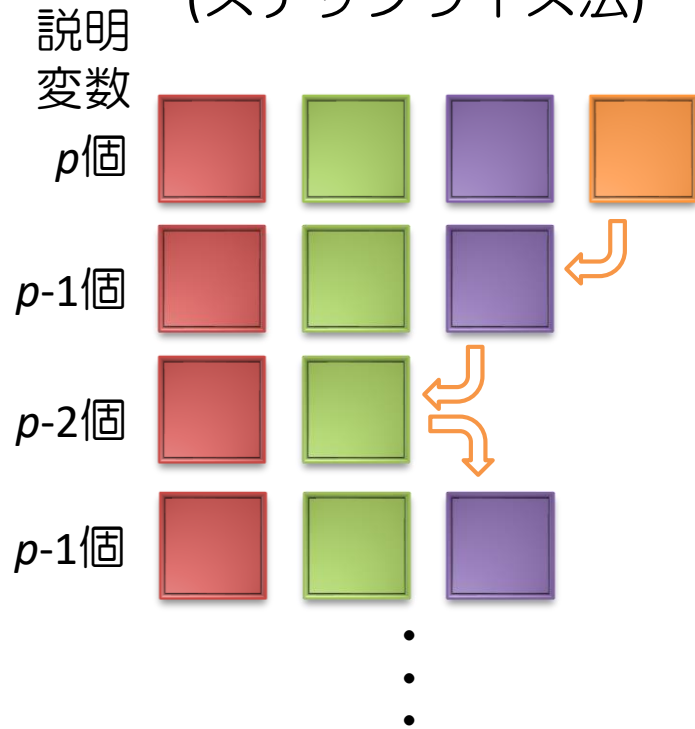
$p-2$ 個



1. すべての変数が含まれるモデルから始める.
2. 減少させる変数の回帰係数が0であるという帰無仮説のもとで検定を行い, 最も検定統計量小さい変数をモデルから削除する.
3. 2の検定がすべての変数で有意になるまで繰り返す.

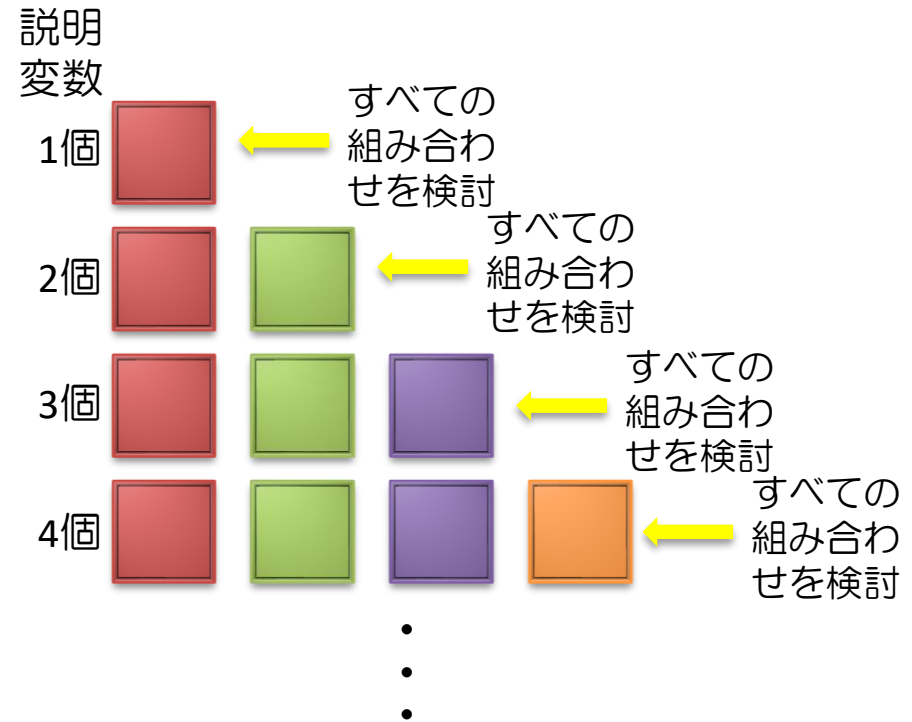
統計学の基礎におけるQ&A(4)：変数選択の種類

変数増減法 (ステップワイズ法)



変数減少法からスタートするが、変数増減法では変数が削除された場合にその変数を考慮しなかったが、変数増減法では変数の削除と削除した変数の追加の両方を検討しながら各ステップを進める。

総当り法



説明変数のすべてのパターンを考え、最適な変数を選択する。

○ 評価基準において必ず最適な変数を選択できる。

× 計算時間がかかる。

統計学の基礎におけるQ&A(4)：変数選択の要件

(1) 評価したい要因は変数選択に強制的に加える

ランダム化比較試験の結果を評価する場合、治療群を表す共変量を含まなければ意味をもたない。つまり、このような場合には、背景因子などの他の共変量を調整したうえで治療群(評価変数)を調べることに意義がある。

(2) 変数増加法の落とし穴

標本サイズが小さい場合に、変数増加法を用いて変数選択を行う場合、結果の解釈が困難なモデルを選択することがしばしばある。また、本当は必要な共変量に取り込まれる前に変数選択が終了する場合がある。

(3) 多数の共変量(項目)がある場合の留意点

医学系研究では、多数の調査項目(共変量)を評価に用いることは少なくない。このような場合には、全ての共変量を用いて変数選択を行うのではなく、事前スクリーニングを行うことが推奨される。事前スクリーニングでは、共変量毎に単変量解析(1個の共変量による回帰モデルを推定する)を実施し、その回帰係数に対する検定(回帰係数が0であるか否かを評価する検定)のp値や回帰係数(オッズ比、ハザード比)を用いる。

統計学の基礎におけるQ&A(4)：変数選択の要件

(4) 欠測が多い共変量(項目)には注意が必要である

多変量解析では、共変量のなかで1個でも欠測があれば、その被験者を削除しなければならない。そのため、欠測が多い共変量をモデルに含めると、多くの被験者を削除することになる。また、観測方法が煩雑な場合には、欠測が多くなる傾向にある。そのため、このような共変量は、予め変数選択の候補から覗いておくことが望ましい。

(5) 可能であれば総当たり法を用いる

変数増加法や変数減少法が必ずしも最適なモデルに到達するとは限らない。最適なモデルを選択できる唯一の方法は、すべての候補モデルを評価する総当たり法のみである。共変量の数が10個の場合、候補となるモデルの数は1,023個である。最近のコンピューターであれば実現不可能な数ではない(共変量の数が20個の場合には1,048,575個となり、不可能に近い数値となる)。そのため、臨床的知見あるいは、事前スクリーニングなどを用いて変数選択に用いる共変量を可能な限り少なくし、そのもとで、総当たり法によって変数選択を実施することが考えられる。

統計学の基礎におけるQ&A(5)：一貫性と違いを見ること

65歳以上の高齢者を対象に、転倒予防訓練と運動機能の低下の有無に関する研究が実施された。この研究では、200人の被験者に対して、転倒予防訓練前に運動機能検査を行い、3カ月の転倒予防訓練後に同様の検査を実施している(柳川他,2011)。

		歩行訓練後		合計
		低下あり	低下なし	
歩行訓練前	低下あり	40 (20.0%)	60 (30.0%)	100
	低下なし	30 (15.0%)	70 (35.0%)	100
合計		70	130	200

- (1) 訓練前，訓練後ともに運動機能が低下している割合は20%
- (2) 訓練前は運動機能が低下していたが，訓練後に低下が認められなかった割合は30%
- (3) 訓練前は運動機能の低下が認められなかったが，訓練後に低下していた割合は15%
- (4) 訓練前，訓練後ともに運動機能の低下が認められなかった割合は35%

一貫性を見るということと違いを見るということ

一貫性を見る場合

例(2種類の定性検査の一貫性評価)

2つの検査を同じ被験者に行ったばあいに、検査結果が一致するか否かを評価する研究

		検査B	
		陽性	陰性
検査A	陽性	(a)	(b)
	陰性	(c)	(d)

(a) 検査Aおよび検査Bともに陽性、(d) 検査Aおよび検査Bともに陰性の度数(被験者数)が診断結果が検査A、検査Bで診断結果が異なる(b)(c)よりも大きくなることが期待される

一致したか否かを検討する
(Kappa係数)

変化(違い)を見る場合

例(手術による不安感に関する調査)

医師からの手術に関する説明前に、手術に対する不安感(あり、なし)を質問したうえで、説明後に同じ質問を行う研究

		説明後	
		あり	なし
説明前	あり	(a)	(b)
	なし	(c)	(d)

説明前に不安ありだった患者が説明後に不安なしに変化することが期待される。

変化したか否かを評価する
(対応がある場合のオッズ比)

統計学の基礎におけるQ&A(6)：多重比較って何？

ここでは、3種類のコレステロール治療薬(A,B,C)を服用したときの血中コレステロール濃度を比較している。

すべての組み合わせでの比較を行うと、3回検定を行うことになる。これによって生じる、「下手な鉄砲も数打てば当たる」では評価にならない



A



B



C

3種類のコレステロール治療の比較

A vs. B, A vs. C, B vs. C (3回の比較)

有意水準 $\alpha=0.05$ とは、違いがないのに(H_0 が正しい)のに、違いがある(H_1 が正しい)と誤ってしまう確率を表している。

→ つまり、コレステロール治療の効果に違いがないのに、20回検定すると、1回は有意差が出てしまう(H_1 と判断してしまう)。

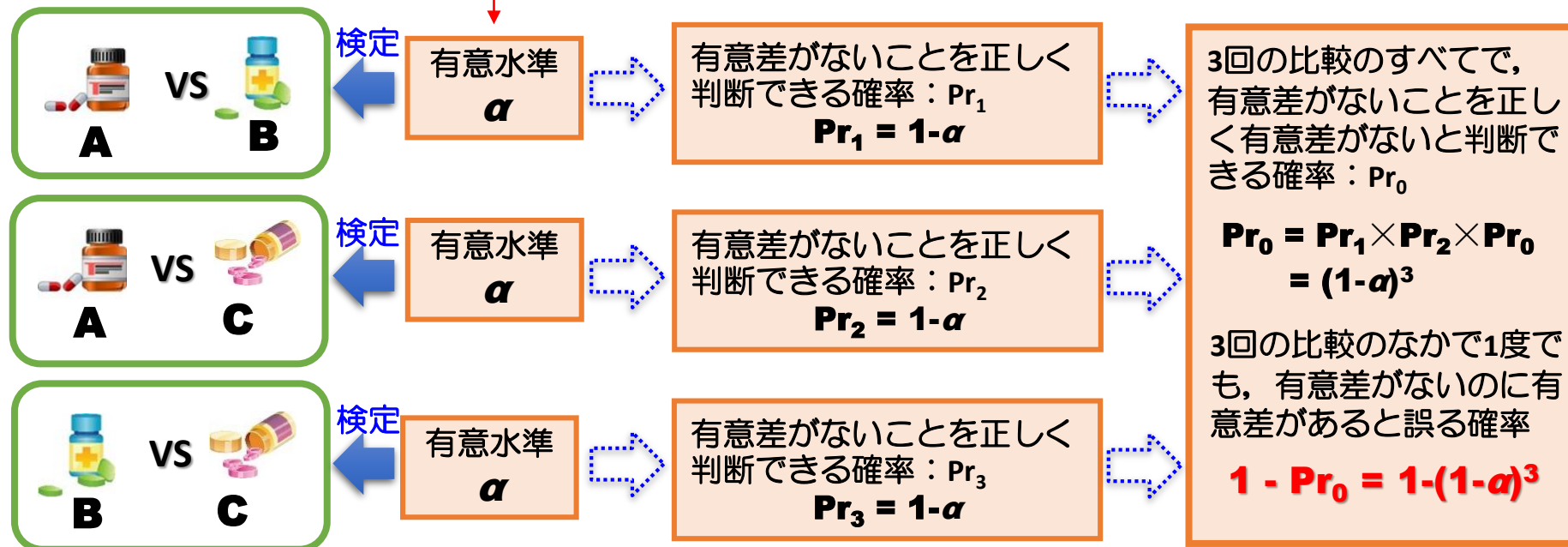
→ 3回比較するということは、その可能性が増しているといえる。

これを多重性という

繰り返し検定を行うことと第1種の過誤の増大の関係

いま、3種類の薬剤の有効性を考える。

有意水準とは「有意差がないのに有意差があると誤る確率(第1種の過誤)」をあらわす



1回の検定の有意水準 α を

$\alpha=0.05$ (1回の検定において有意差がないのに有意差があると誤る確率を0.05とする)

とし、3回の検定を繰り返した場合、3回の検定の中で1度でも有意差がないのに有意差があると誤る確率は、

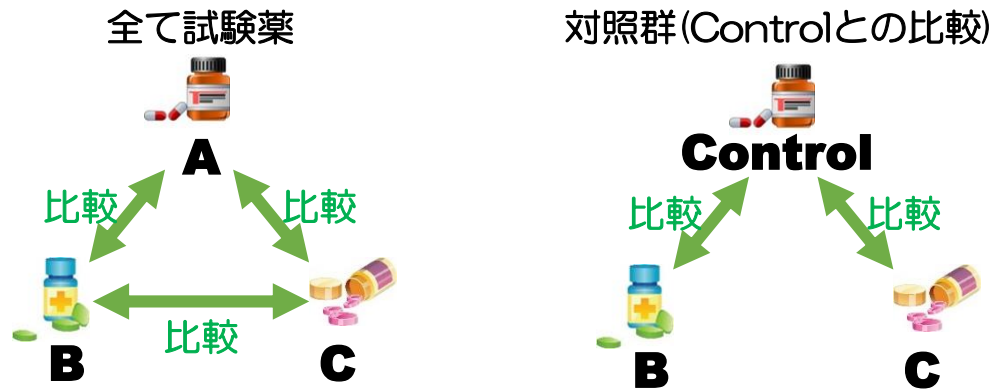
$$1 - (1 - \alpha)^3 = 1 - (1 - 0.05)^3 = 1 - (1 - 0.05)^3 = 0.142$$

となり、誤りの確率が増大してしまう。

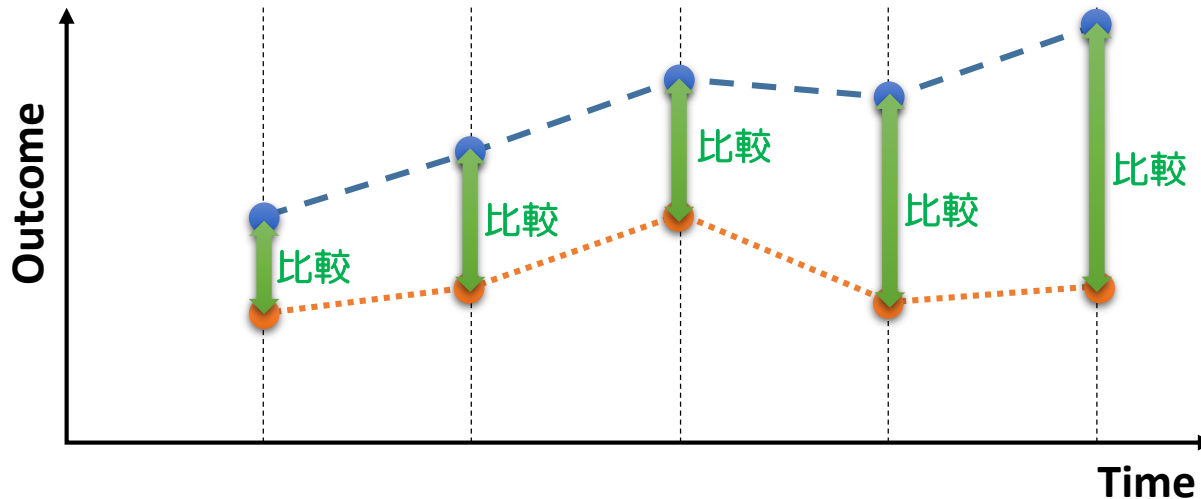
多重比較の要件

多重比較はどのような場面で用いるべきか？

■ 3群以上での比較の場面における対比較



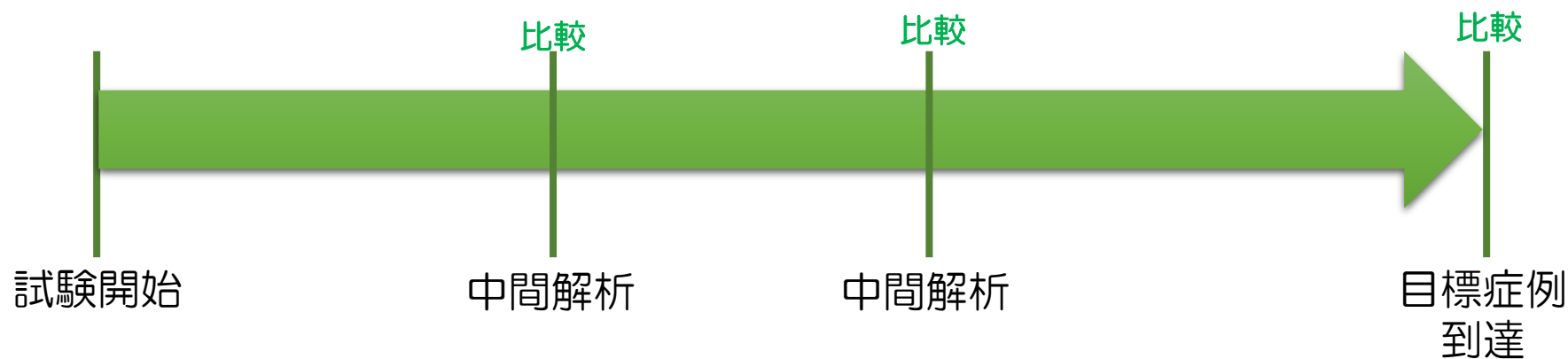
■ 経時測定データにおいて、時点ごとに群間を比較する(下図なら5回比較)



多重比較の要件

多重比較はどのような場面で用いるべきか？

■ 臨床試験における中間解析



- 中間解析とは、無作為化比較試験において、目標登録症例に到達する前に、中間評価を行い、早期に試験を中止するか否かを検討する。
- 上図では、2回の中間解析を実施するため、合計3回の比較を行う。
- 1例が終了する毎に中間解析を行ってしまうと、「下手な鉄砲も数打てば当たる」状況となるため、多重比較を行わなければならない。

多重比較の要件

多重比較は分散分析後に行うべきか？

P84～85にかけて、「複数群間を一気に比較する分散分析で有意差が出たときのみ・・・対比較を行います」と記載されているが、多重比較は分散分析後でないと行えないのか？

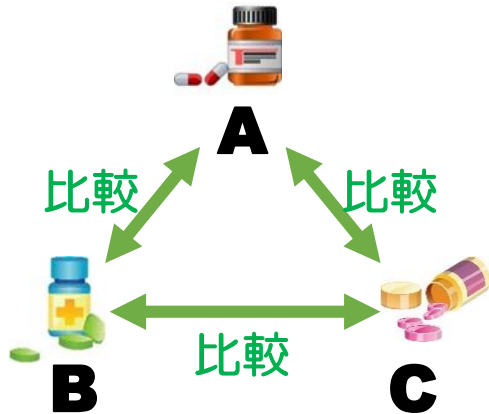
- Tukeyの方法は、分散分析の結果に基づいて多重比較を行うため、分散分析は必須である。
- Bonferroniの方法、Holmの方法、Dunnettの方法は、分散分析の結果を用いるわけではないので、分散分析が必須ではない。

多重比較の適用場面は？

- Bonferroniの方法、Holmの方法は、量的尺度以外のデータにも適用可能である。
- Bonferroniの方法は、あらゆる場面で適用できるが最も有意になりにくくなる多重比較法である。
- Dunnettの方法は、対照(Control)群との比較を伴う多重比較において、最良の多重比較法である。

多重比較の方法：Bonfferoniの方法

Bonfferoniの方法



A群, B群, C群の対比較の場合で全体での有意水準が α の場合には

- A群 対 B群 → 有意水準 $\alpha/3$ と比較(or p値を3倍)
- A群 対 C群 → 有意水準 $\alpha/3$ と比較(or p値を3倍)
- B群 対 C群 → 有意水準 $\alpha/3$ と比較(or p値を3倍)

有意水準 α を比較回数で割る(or p値を比較回数で掛ける)方法がBonfferoniの方法である。多重比較が簡単なため、最も用いられる方法の一つである。EZRでは

- ・パラメトリック検定：2標本 t 検定の結果を調整
- ・ノンパラメトリック検定：Wilcoxon検定の結果を調整

多重比較の方法：Holmの方法

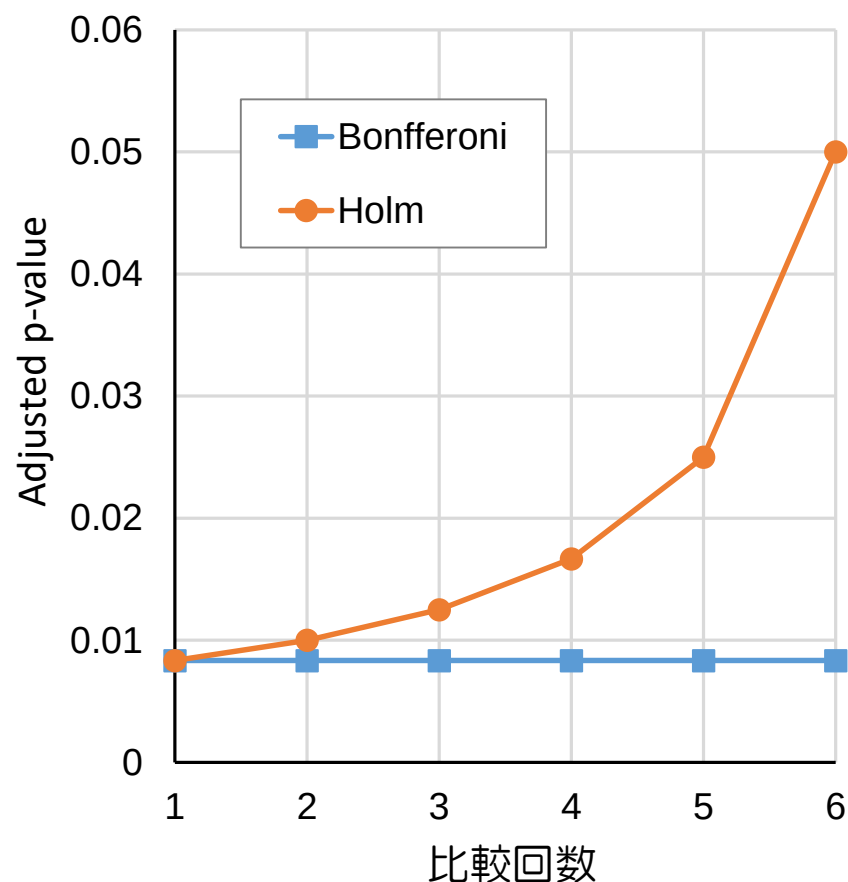
Holmの方法 (パラメトリック：2標本 t 検定, ノンパラメトリック検定：Wilcoxon検定を調整)

最小のp値から並べ替え，シーケンシャルに比較する方法。

いま，6回の比較のp値が次のように与えられているとする。

0.001 0.006 0.011 0.012 0.032 0.045

いま，6回の比較のp値が次のように与えられているとする。

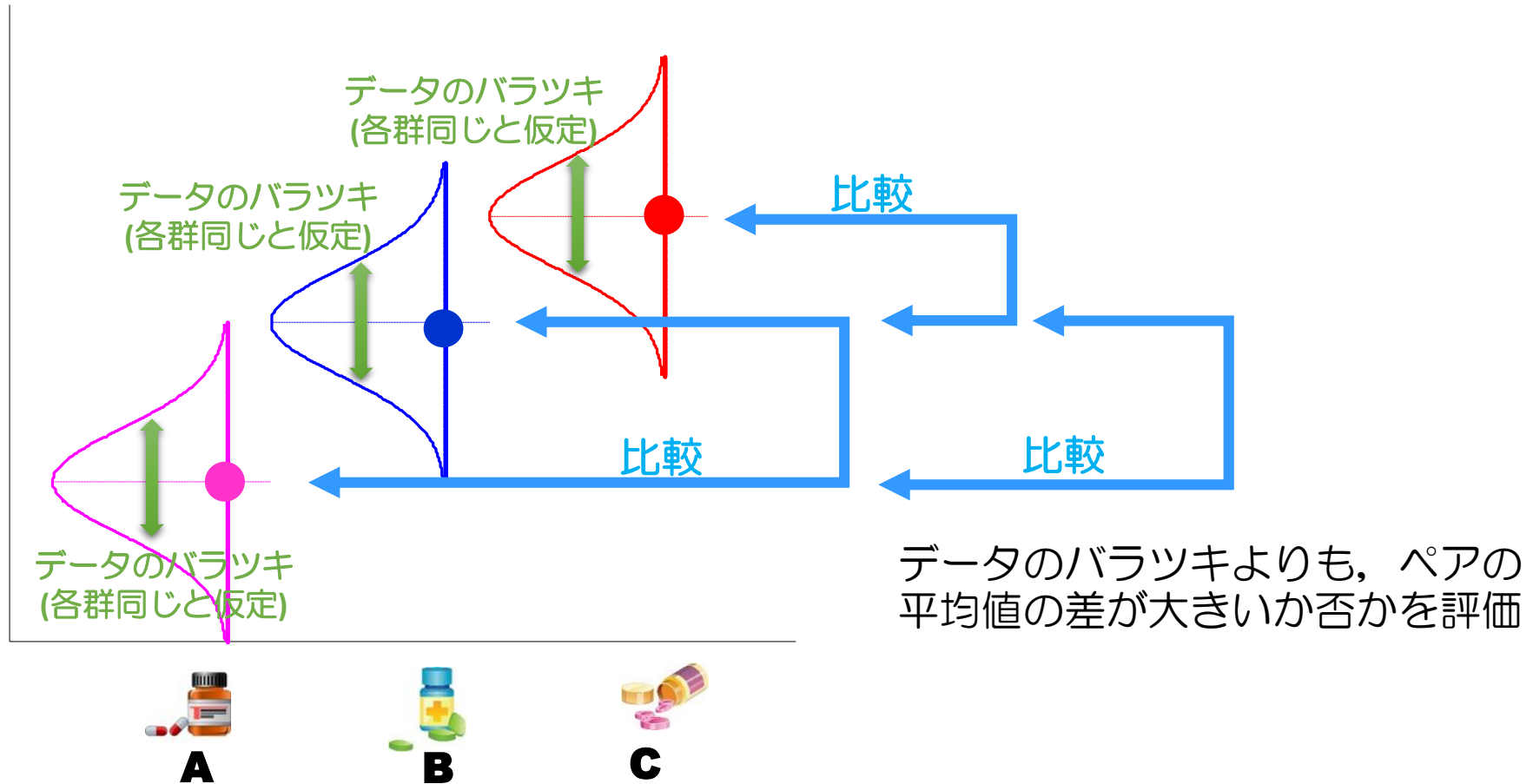


順位	p値	Bonferroni法		Holm法	
		比較α	判定	比較α	判定
1	0.001	0.05/6= 0.008	有意	0.05/6= 0.008	有意
2	0.006	0.05/6= 0.008	有意	0.05/5= 0.010	有意
3	0.011	0.05/6= 0.008	非有意	0.05/4= 0.013	有意
4	0.012	0.05/6= 0.008	非有意	0.05/3= 0.017	有意
5	0.032	0.05/6= 0.008	非有意	0.05/2= 0.025	非有意
6	0.045	0.05/6= 0.008	非有意	0.05/1= 0.050	非有意

以降はすべて有意でない

多重比較の方法：Tukeyの方法

Tukeyの方法

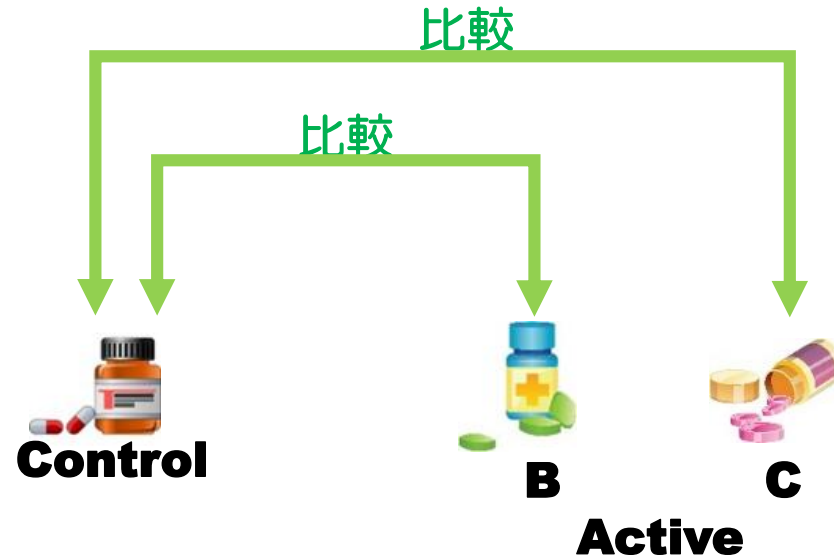


一元配置分散分析において有意だった場合には、Tukeyの方法でもいずれかの対比較において有意になっている。

ノンパラメトリック検定の場合には、順位に基づいてTukeyの方法と同様の検定を行うSteel-Dwass検定がある(Kruskal-Wallis検定に対応)。

多重比較の方法：Dunnettの方法

Dunnettの方法



Controlとの比較のみを実施する多重比較がDunnettの方法である。ノンパラメトリック検定の場合には、Steel検定がある。

Dunnettの方法では、

- 正規分布に従うこと
- すべてのグループ(群)のあいだで、分散が等しいことが仮定される。



その他の疑問



■ その他の疑問

BKUP.1：生存曲線からわかる知られざる解釈

BKUP.2：無作為化比較試験とITT(Intent-To-Treat)の関係(いつもITTが正しいわけではない).

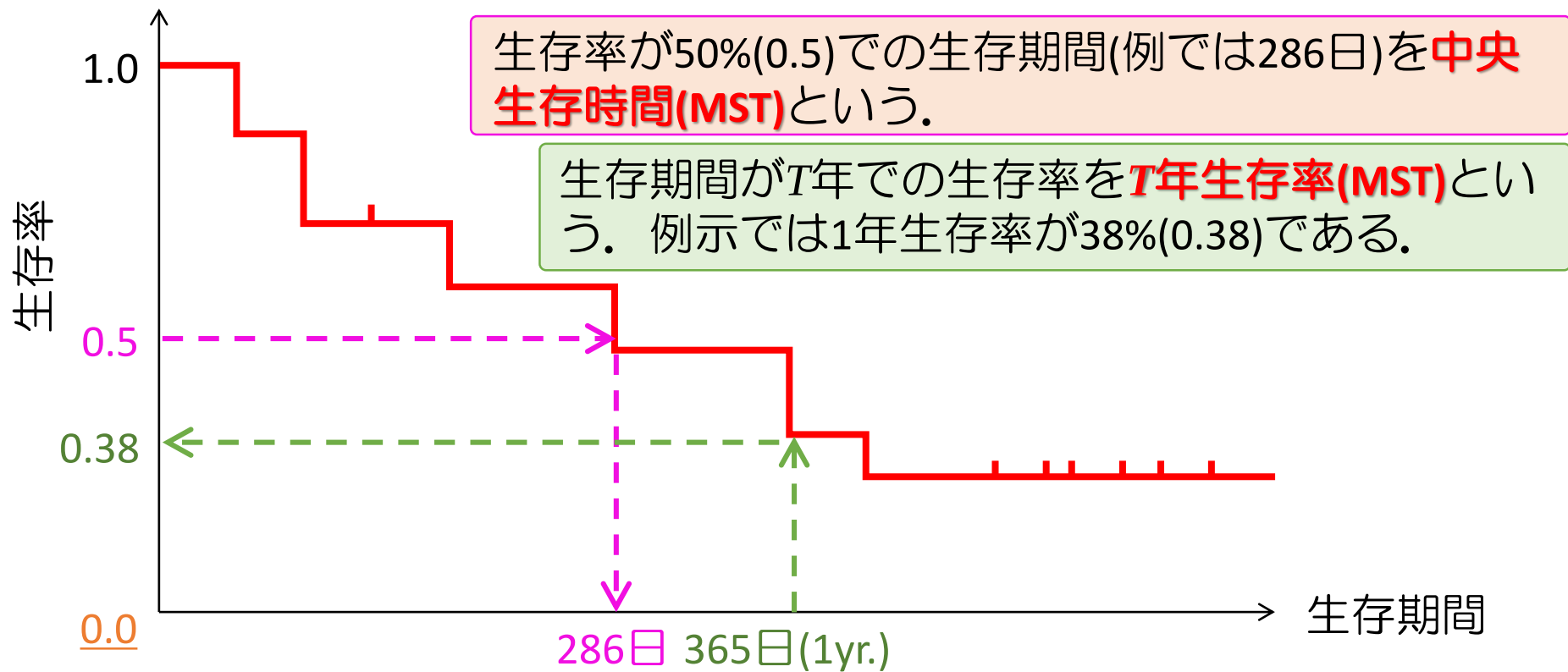
BKUP.3：共変量が量的尺度の場合の2値化の留意点

BKUP.4：交互作用とはなにか？

BKUP.5：打ち切り(censoring)の情報を活用する：競合リスク

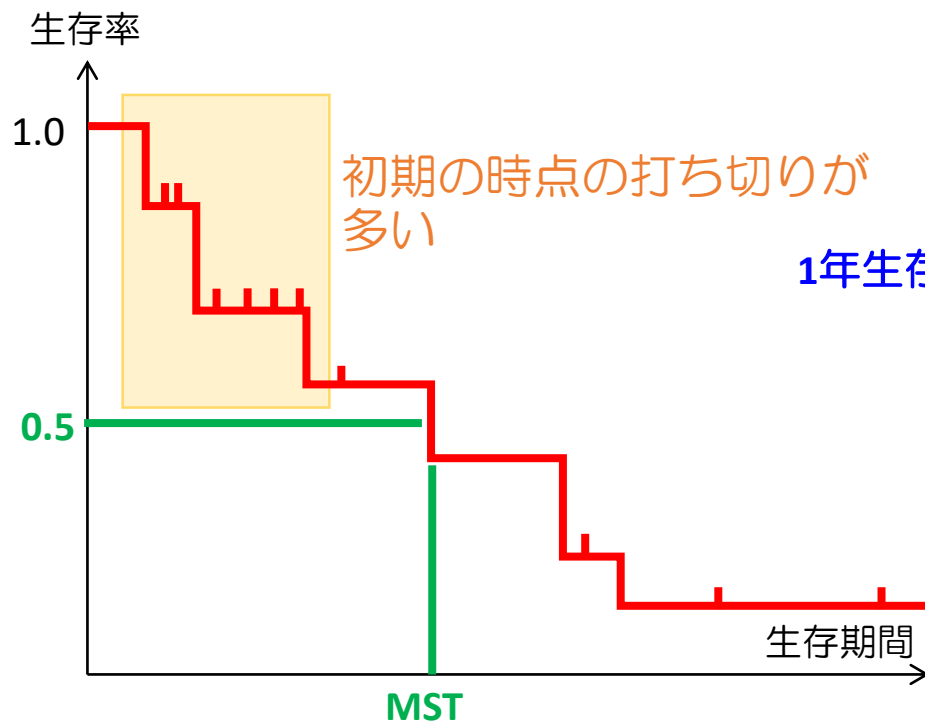
BKUP.6：傾向スコア分析とはなにか？

BKUP.1：生存曲線からわかる知られざる解釈



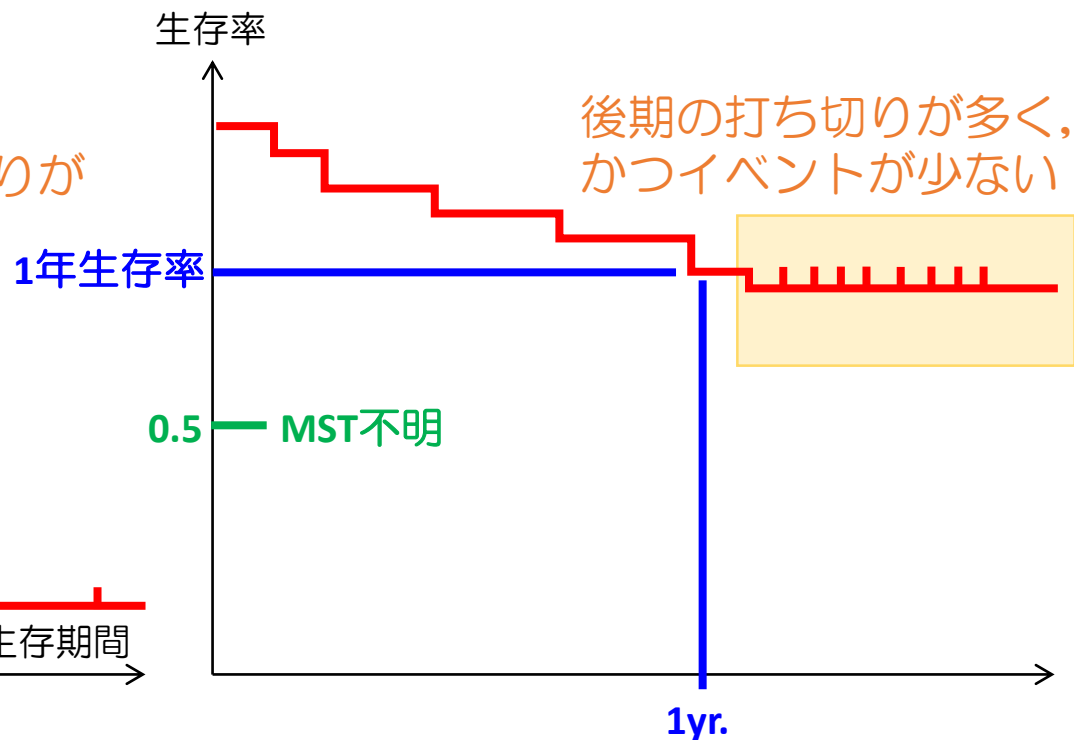
ここで注意点は、365日のところで、矩形の上側を見るのか下側を見るのかという点である。Kaplan-Meier曲線の矩形は上側ではなく下側の生存率を見なければならない。

BKUP.1：生存曲線からわかる知られざる解釈



初期の時点で打ち切りが多いということは、きっちりとしたフォローアップがなされていない、あるいは、研究実施体制に問題があると判断される可能性がある。

また、この場合のMSTは不完全である。



後期の時点で打ち切りが多く、かつフォローアップ期間中にイベントが少ない理由：

- ① 対象疾患の予後が良好な疾患の場合
- ② フォローアップ期間が十分でない場合がある。

この場合には、MSTは計算できないため、年次生存率のみを計算

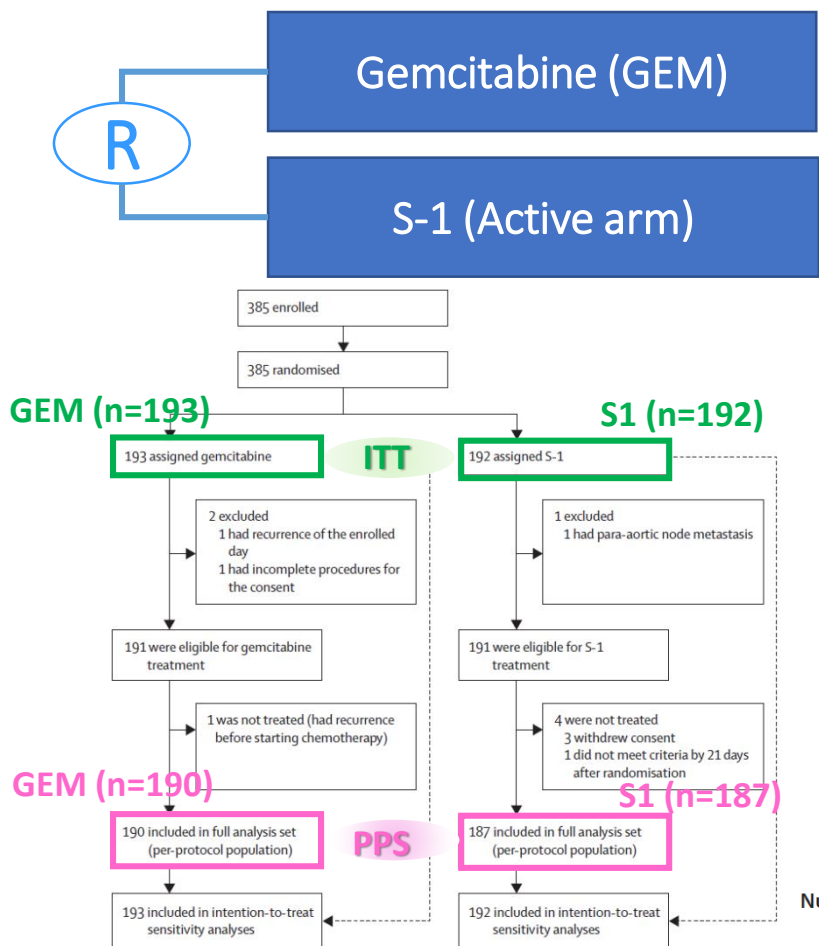
RCTとITTの関係：いつもITTが正しいわけではない

■ JASPAC 01 Study (Uesaka et al., Lancet, 388, 248-257, 2016)

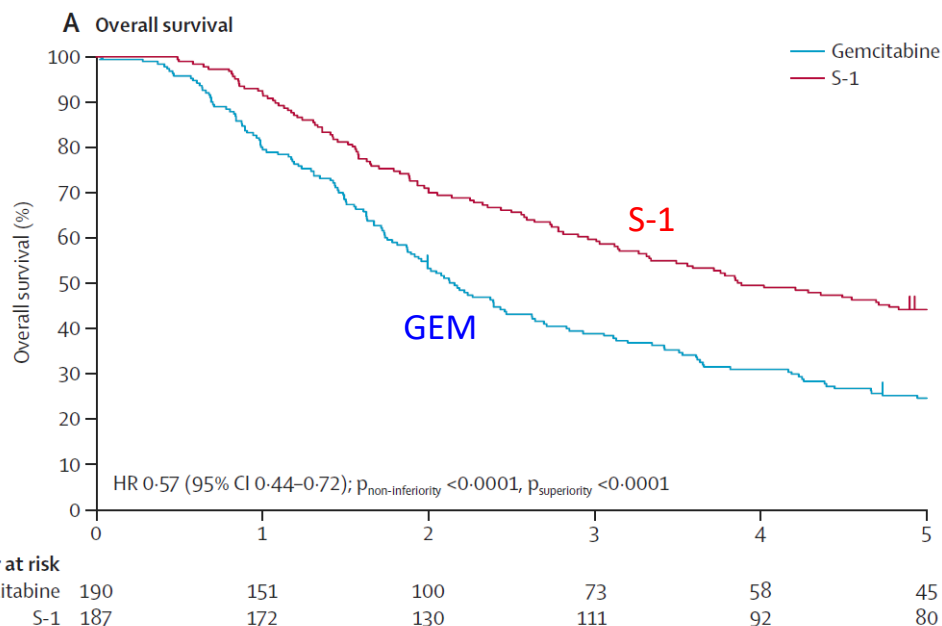
治癒切除された浸潤性膵管癌患者に対する
GEM vs. S-1の無作為化比較非劣勢試験

Primary endpoint :

PPSにおける全生存期間(OS)

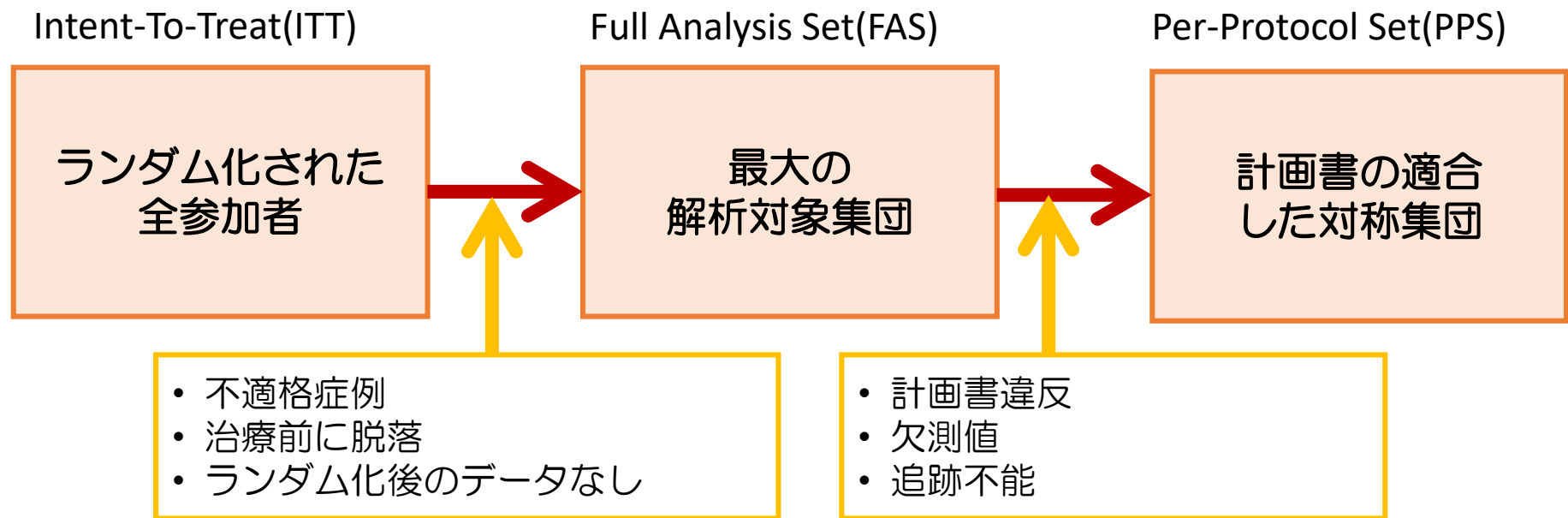


GEM群の3年生存割合を36%と仮定し、S-1群のGEM群に対するハザード比を0.87と見込み、片側 $\alpha=0.025$ 、検出力80%（非劣性マージン1.25）で240イベントが必要である。登録期間3年・観察期間2年とした場合に、必要症例数は360例と計算されている。



無作為化比較試験では、ITT症例で解析することが一般的では？

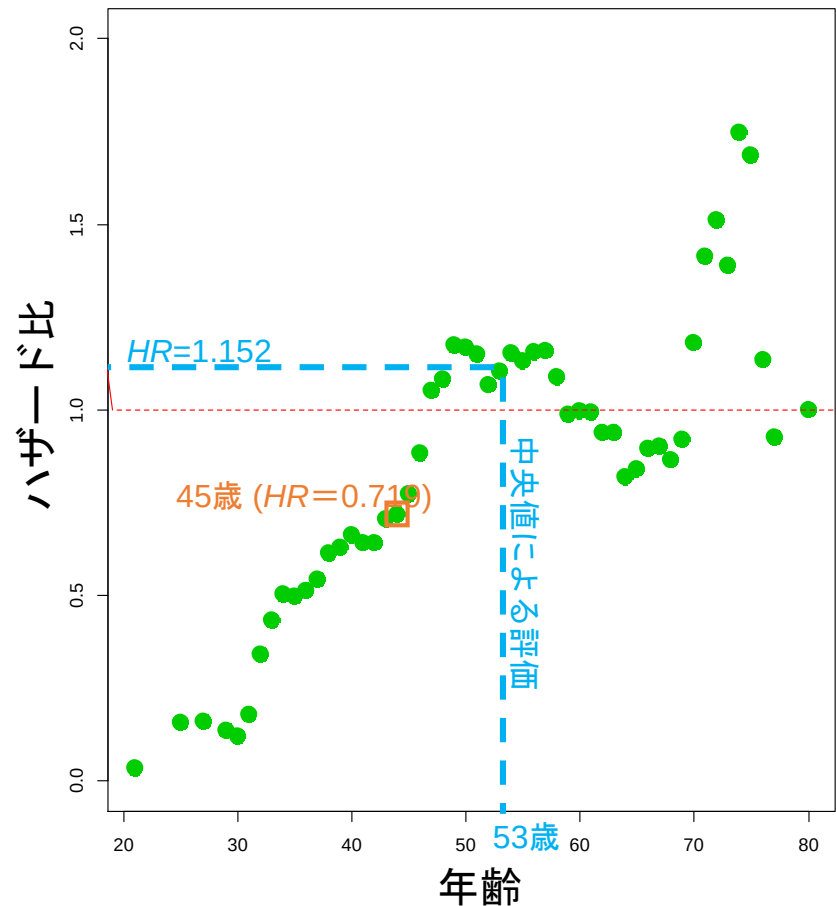
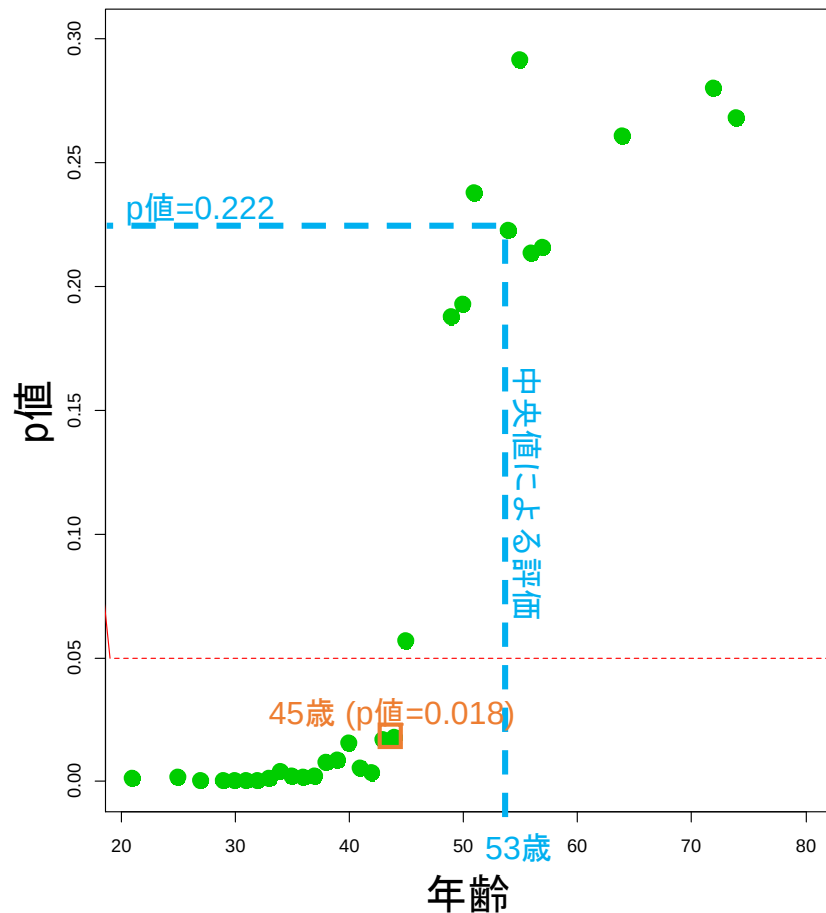
RCTとITTの関係：いつもITTが正しいわけではない



- 選択バイアスの観点から、原則的には、ITT症例(or FAS)での解析が行われる(ICH-E9ガイドラインにも記載).
- 純粋に治療法(あるいは薬剤)の評価を行う場合には、PPSを主解析とする場合がある。例えば、安全性が主要評価項目の場合である。
- また、非劣性試験では、研究の質が悪いと群間差が小さくなることや、プロトコルのアドヒアランスを評価するために、PPSで実施することがある。JASPAC01において、PPSを利用したのはそのため。

BKUP.3：共変量が量的尺度の場合の2値化の留意点

ドイツ乳がん研究グループによって実施されたランダム化比較試験のデータ(Schumacher et al., 1994)



変量の2値化において用いられることが多い中央値(53歳以上, 53歳未満)をカットオフ値にした場合, p値は0.222である(有意でなかった). 一方で, 45歳(45歳以上, 45歳未満)をカットオフ値にした場合, p値は0.018である(有意だった).

BKUP.3：共変量が量的尺度の場合の2値化の留意点

45歳(<45/≥45)の場合

高年齢群(45歳以上)のほうが低年齢群(45歳未満)よりも死亡リスクが低い
(HR=0.719)

解釈が逆



53歳(<53/≥53)の場合

高年齢群(53歳以上)のほうが低年齢群(53歳未満)よりも死亡リスクが低い
(HR=1.152)

カットオフ値の選定に関するRule of Thumb

- (1) 中央値を用いる
- (2) 臨床的にリーズナブルなカットオフ値を選定する

臨床的にリーズナブルとは、今回の事例の場合には、例えば疫学的調査などから得られた乳癌患者の平均年齢を用いたり、あるいは、閉経年齢を用いることを意味する。いずれにしても、カットオフ値選定におけるゴールド・スタンダードは存在しないため、試行的に決定する必要がある。

BKUP.4：交互作用とはなにか？

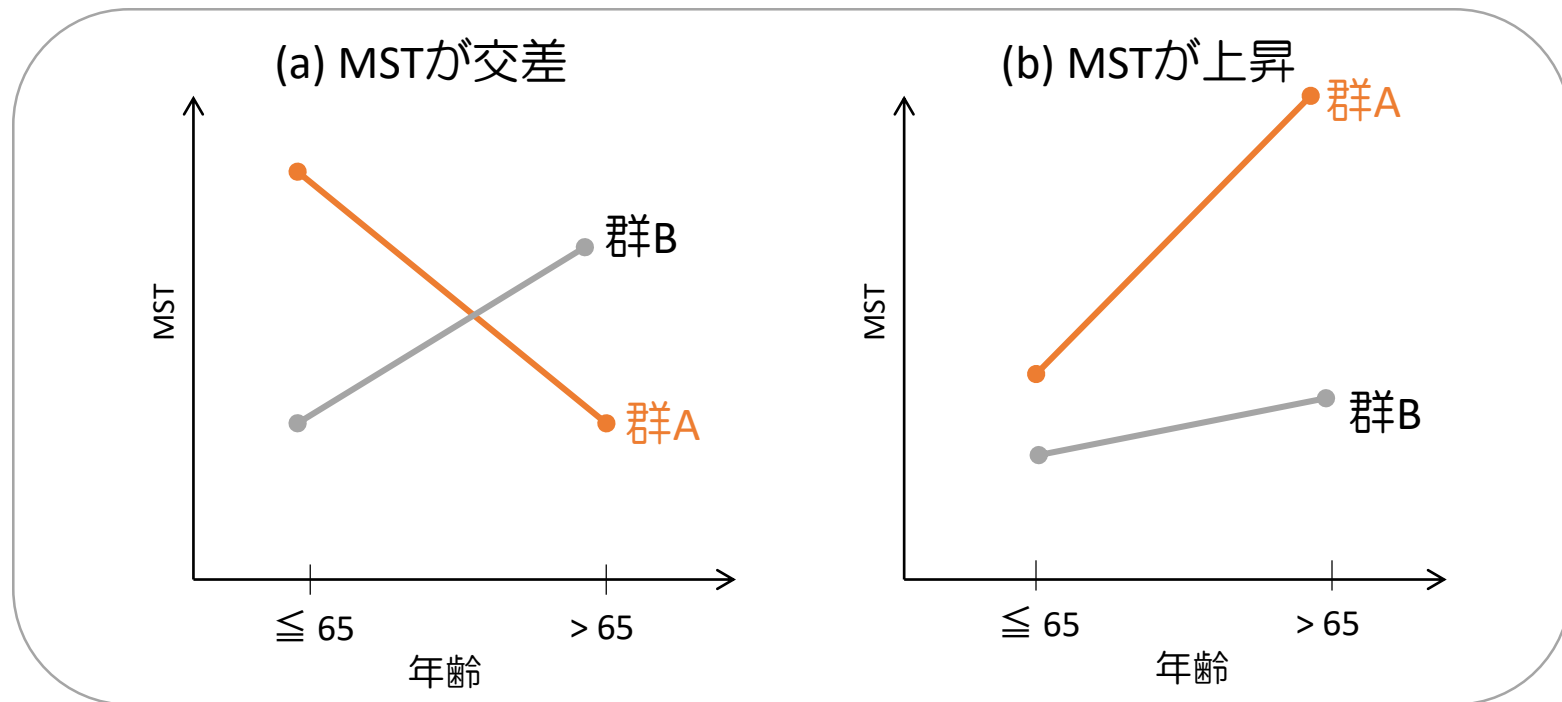
交絡	交互作用
- 重要な予後因子が群間で不均衡 - 試験計画そのものが問題である(割付因子として組み込む必要性) - 交絡がないということは試験計画の妥当性がいえる。	- 薬剤の効果が、サブグループ間で異なる。 - 薬剤の適応患者像に関する制限につながる。 - 交互作用がないということは治療の一般可能性のあることがいえる。

「浜田知久馬(1999). 学会・論文発表のための統計学, 新興交易医書出版部」より引用・改編

- (1) ICH E5ガイドライン「臨床試験の統計的原則」では、多施設共同試験では、薬剤と施設の交互作用を検討することが要求される。これは、薬剤の一般可能性を証明するためである。
- (2) 交互作用は、回帰モデル(生存時間解析では、比例ハザードモデル)に治療(薬剤)のダミー変数と予後因子の積による項を追加し、その回帰係数の有意性により検出する。一方で、交互作用の構成については、Cox(1984)によって詳細に吟味されている。

BKUP.4：交互作用とはなにか？

交互作用の例：下記は抗がん剤Aと抗がん剤Bでの中央生存時間(MST)をプロットしたものである。



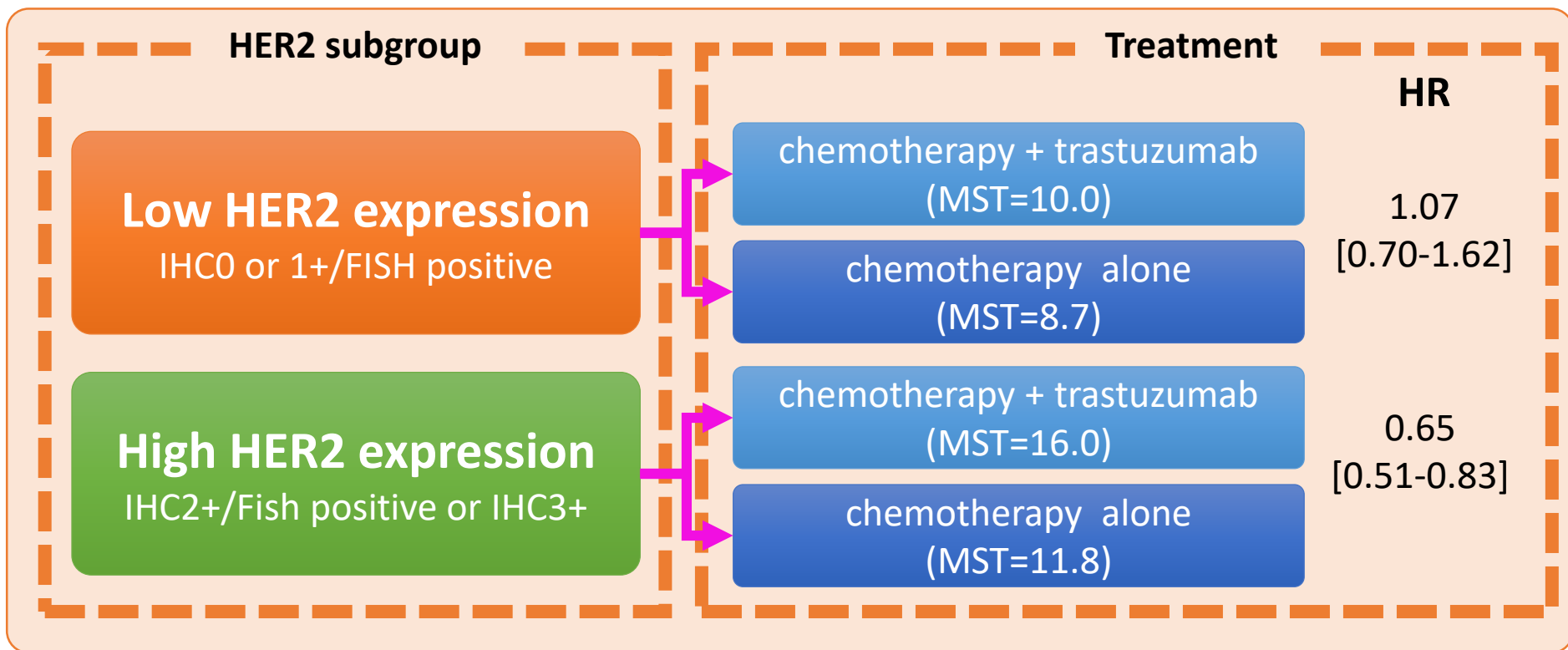
何らかの予後因子によって、抗がん剤の効果が逆転したり、相乗効果によって、より効果が上昇するような現象を交互作用 (interaction) という。

BKUP.4：交互作用とはなにか？

ToGA study(Bang et al.: *The Lancet*, 376, 687-697, 2010)の一文

There was evidence of a **significant interaction test** ($p=0.036$) between treatment and the two HER2 subgroups (high HER2 expression vs low HER2 expression).

切除不能な局所進行性、再発 and/or 転移性のHER2陽性の患者に対する化学療法単独群（5-FU+CDDP (FP療法)あるいはCapecitabine +CDDP(XP療法)）と化学療法+Trastuzumab併用群を比較する国際共同無作為化非盲検試験



Low HER2ではハザード比(HR)が有意でないが, High HER2では有意に chemotherapy + trastuzumabのハザード比が低い.

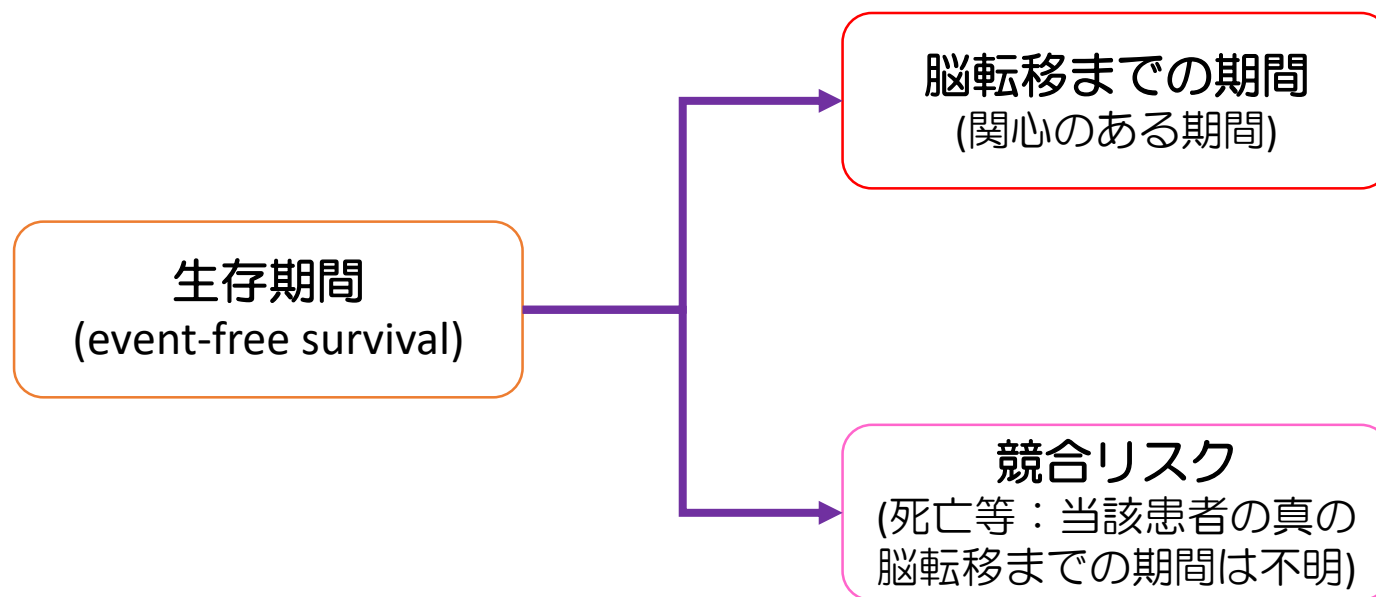
BKUP.5：打ち切り(censoring)の情報を活用する：競合リスク

例：小細胞肺癌患者に対する予防的頭蓋照射の有効性・安全性に対する無作為化比較第III相試験

- 小細胞肺癌では、脳転移が問題になる。
- 脳転移は予後に重大な影響を及ぼす。

脳に予防的頭蓋照射を行うことの有効性・安全性を評価する。

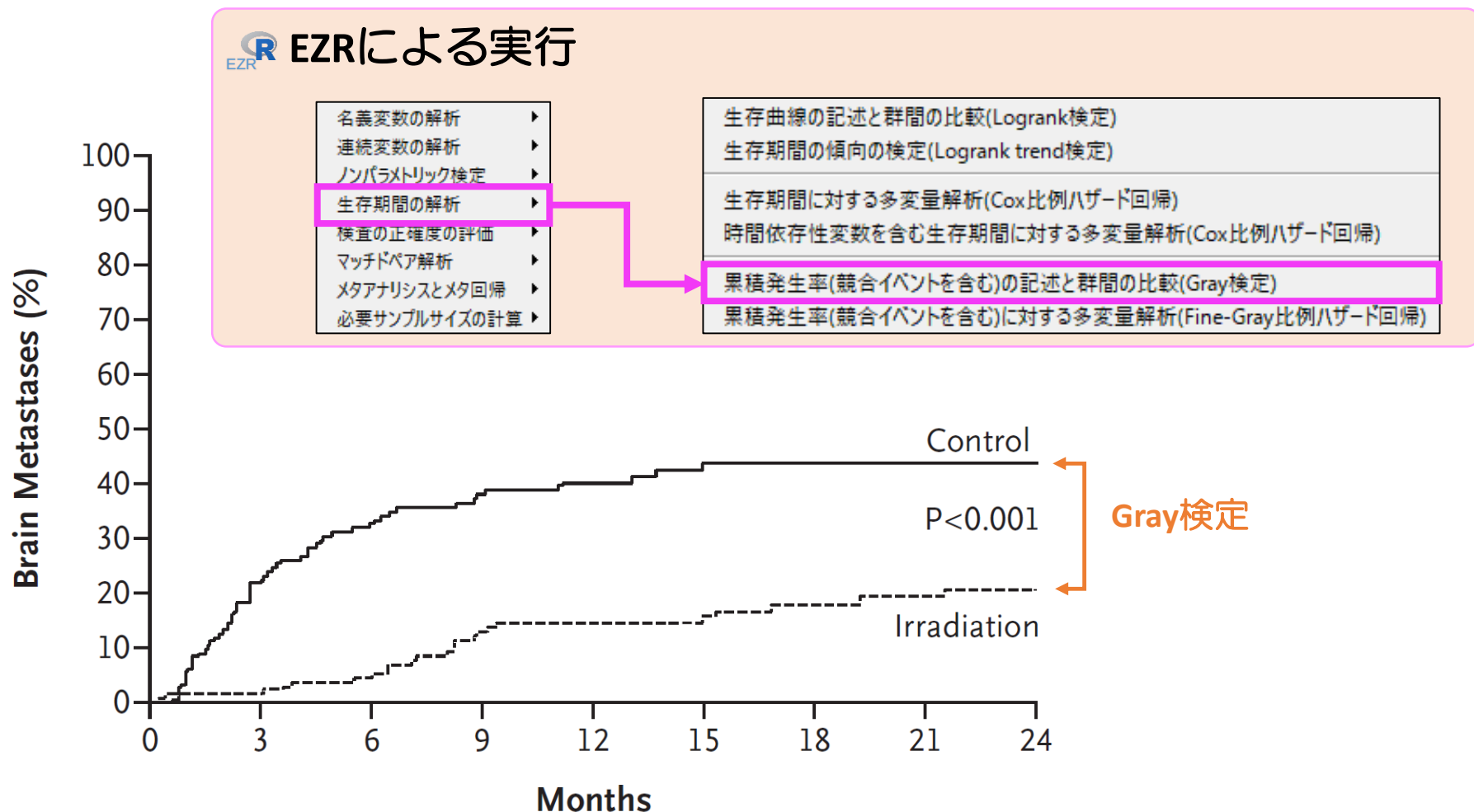
主要評価項目：脳転移をイベントとした生存期間



BKUP.5：打ち切り(censoring)の情報を活用する：競合リスク

累積発生関数(CIF; Cumulative Incident Function)

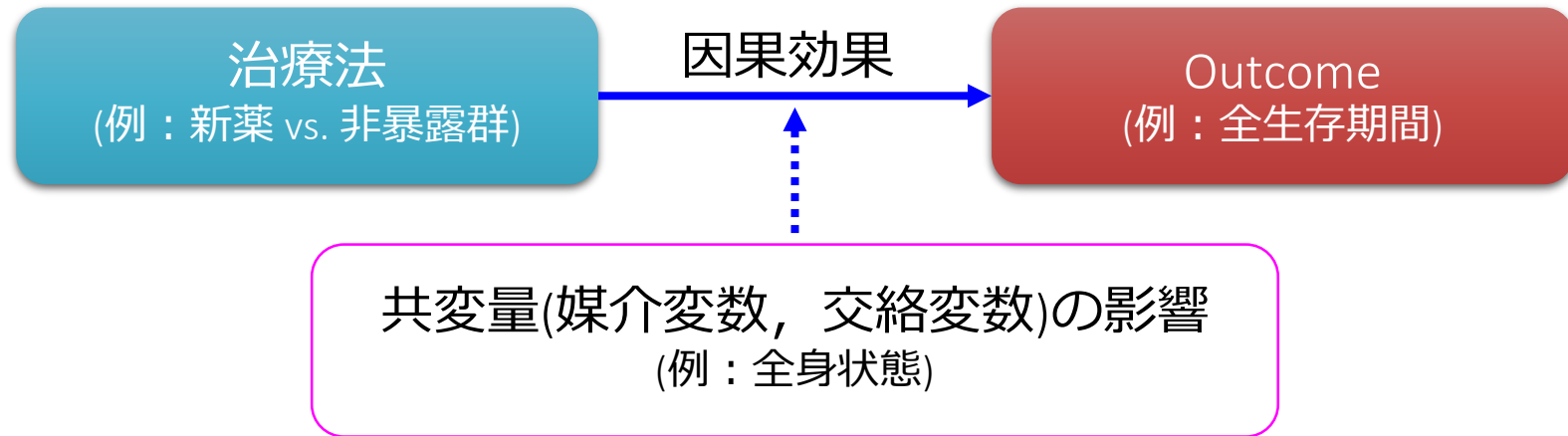
→ 時間tまでにイベントが発生する割合を表す曲線



CIFを比較する検定がGray検定である。Gray検定はログランク検定の競合リスク版として解釈できる。

BKUP.6：傾向スコア分析とはなにか？

□ 治療法の評価 = ある医学的介入による影響(因果効果)の検証



共変量が厄介な理由：コントロールできないものが多い

検証的な側面から共変量の影響を最小限にして因果効果を評価する

– 無作為化比較試験(RCT:Randomized Controlled Trial)

観察的な側面(観察研究)において(治療選択に対する)共変量の影響を最小限にして因果効果を評価する

- 統計学的な調整法 (共変量とOutcomeの線形性などの制約)
- 傾向スコア(Propensity score)の利用

BKUP.6：傾向スコア分析とはなにか？

傾向スコアとは、研究対象因子に暴露する確率で表される(つまり、傾向スコアは0～1の範囲をとる)。

暴露群

非暴露群

傾向スコア(propensity score)

暴露・非暴露が完全に分離しているため、比較可能性が担保できない(いいかえれば、患者の違い明確であり、アウトカムを比較する状況にないことを示している)

暴露群

非暴露群

傾向スコア(propensity score)

傾向スコアが重複している部分については、暴露・非暴露のいずれも候補になり得る。このような状況においては、比較可能性が担保されている。

BKUP.6：傾向スコア分析とはなにか？

傾向スコアの特徴

- 条件付き独立性：傾向スコアでの割付がアウトカムに影響しない
- バランシング：同じ傾向スコアの値に対応した共変量の分布は、それぞれの群で等しい

ただし、傾向スコアを計算するのに用いた共変量の範囲でのことであり、観測されていない(あるいは、傾向スコアの計算に用いていない)共変量があり、それが交絡因子である場合には、その限りではない。したがって、傾向スコアを計算するうえで、もちいる共変量は十分に吟味しなければならない。

傾向スコア推定の方法

傾向スコア推定の方法には、ロジスティック回帰を用いる方法から、機械学習(非線形分類手法)の方法を用いる方法まで様々な方法が提案されている。

■ ロジスティック回帰

解釈が分かりやすいものの、交互作用を捉えられない可能性がある。

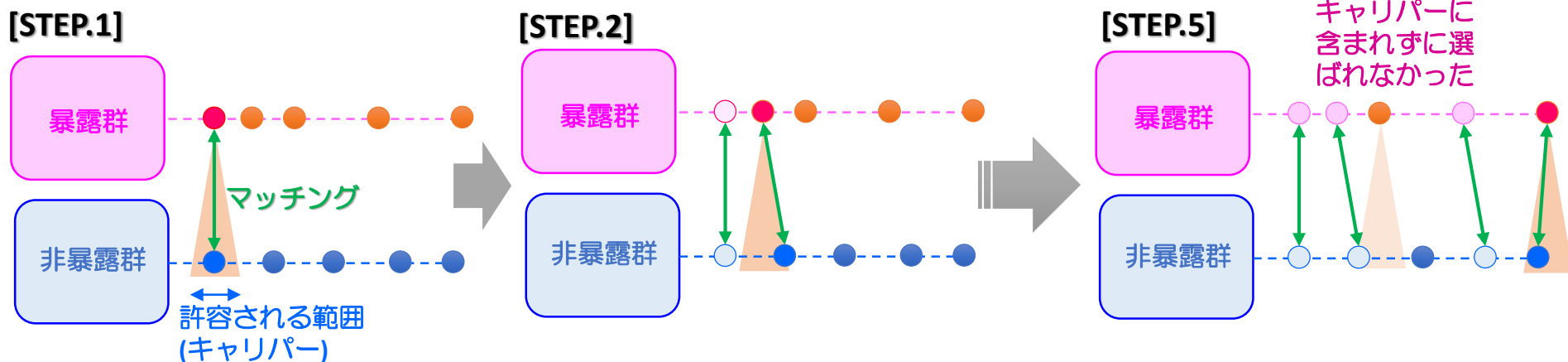
■ 機械学習の方法(例：Random Forests, Generalized Boosting Modeling)

交互作用を捉えられるものの、解釈ができない。

その他にも、Bayes流傾向スコア推定(An, 2010；Kaplan & Chen, 2012,2014; McCandless et al., 2009), 2段階(傾向スコア推定と実施)の同時推定(Imai & Ratkovic, 2014)などが提案されている(つまり、いっぱい提案されている)。

BKUP.6：傾向スコア分析とはなにか？

■ マッチングにおいて良く用いられる方法：強欲マッチング(Greedy matching)



■ 利点

- (1) わかりやすく、統計の知識がなくても使える。
- (2) 独立なサンプルよりも推定効率が良い(注意：1:1マッチングだからといって対応のある解析は行わない)。
- (3) データ解析が容易である。

■ 欠点

- (1) マッチングにコストがかかる。また、マッチングを行うと標本サイズが必ず少なくなる。
- (2) 標本サイズの減少に伴い、標準誤差(標準偏差を標本サイズの平方根で割ったもの)が増加するため、推定効率が下がる場合がある。
- (3) 他の方法(事後的な共分散分析や事後層別)に比べて必ずしも効率が良い方法ではない。

事例の結果

傾向スコア：ロジスティック回帰
マッチング：1対1非復元マッチング(キャリパーを伴う最近傍法：0.25SD)

マッチング前

		薬剤A (N=110)	薬剤B (N=71)	p値
性別	男性	69 (62.7%)	52 (73.2%)	0.1500
	女性	41 (37.3%)	19 (26.8%)	
年齢		53.4 (8.32)	54.1 (8.62)	0.6040
喫煙	有	30 (27.3%)	34 (47.9%)	0.0048**
	無	80 (72.7%)	37 (52.1%)	
BMI		24.1 (2.92)	25.0 (3.77)	0.0632
重症度スコア		7.3 (0.18)	8.8 (0.22)	<0.0001***

マッチング後

		薬剤A (N=38)	薬剤B (N=38)	p値
性別	男性	24 (63.2%)	24 (63.2%)	1.0000
	女性	14 (36.8%)	14 (36.8%)	
年齢		55.3 (1.36)	55.3 (1.36)	0.2947
喫煙	有	14 (36.8%)	13 (34.2%)	0.811
	無	24 (63.2%)	25 (65.8%)	
BMI		24.2 (0.48)	24.3 (0.48)	0.951
重症度スコア		8.2 (0.29)	8.2 (0.29)	1.0000