

実践医学統計セミナー 2022

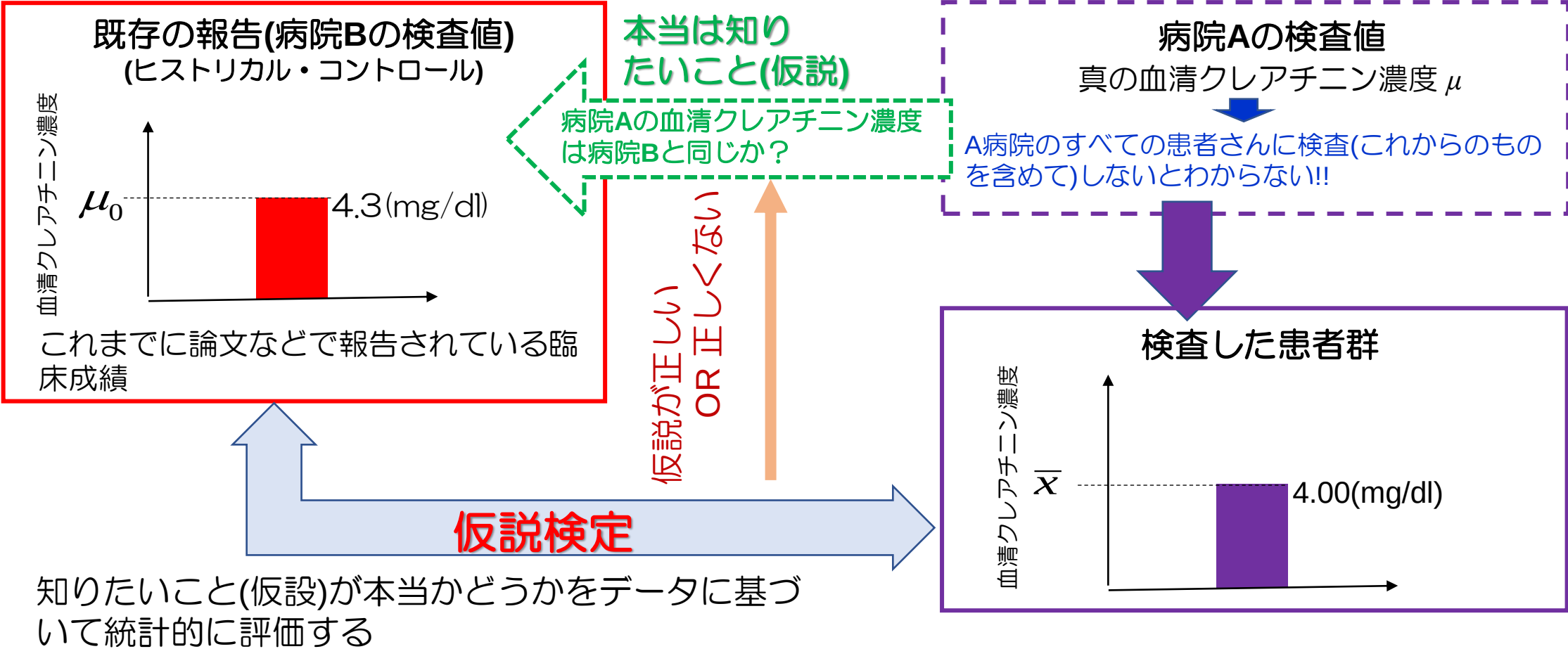
下川敏雄

和歌山県立医科大学附属病院 臨床研究センター

準備：統計学の復習

仮説検定とは何か？

病院A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 3.9
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Aの血清クレアチニン濃度の平均値 μ は、病院Bの値 $\mu_0=4.3$ と同じか？

帰無仮説

血清クレアチニン濃度が病院Bと同じであるとは、
(病院Aの平均値 μ) = 4.3 (病院Bの値)
である。「同じである」という状況は、どのような研究でも同じである。

唯一の数値で設定できるこちら側を評価の基準とする
(ただし研究の目的と逆)

対立仮説

血清クレアチニン濃度が病院Bと異なるとは、
(病院Aの平均値 μ) $\neq$ 4.3 (病院Bの値)
である。「異なる」という状況は、研究によって違うかもしれない。
*例えば、白血球の違いを見る場合と異なると解釈する大きさが異なる

上の仮説が誤っていることを示すことで、逆の仮説であるこちら
が正しいと判断する。

帰無仮説 H_0 : A病院の検査値はB病院と同じである

対立仮説 H_1 : A病院の検査値はB病院と異なる

帰無仮説 H_0 が正しいときに、データから得られた差が偶然に生じたものであるかどうかを表す確率のようなものが**P値**である。

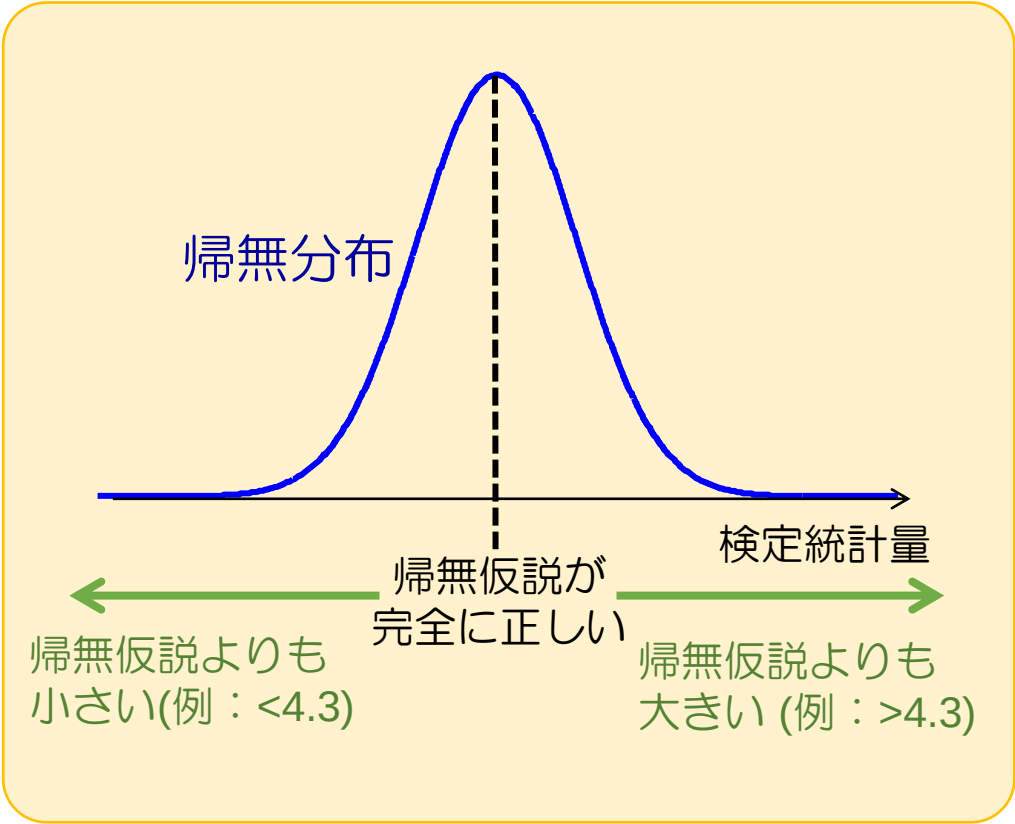
p値が小さいとは、帰無仮説 H_0 が正しいとはいえないほどに差があることを意味する。つまり、対立仮説 H_1 が正しい状況にある。

P値の計算には、**検定統計量**という数値を計算する。検定統計量もまた確率分布に従っており、これを**帰無分布**という。帰無仮説が正しいときに、検定統計量がどれぐらいの確率で得られるものなのかを計算する。これが、**p値**である。

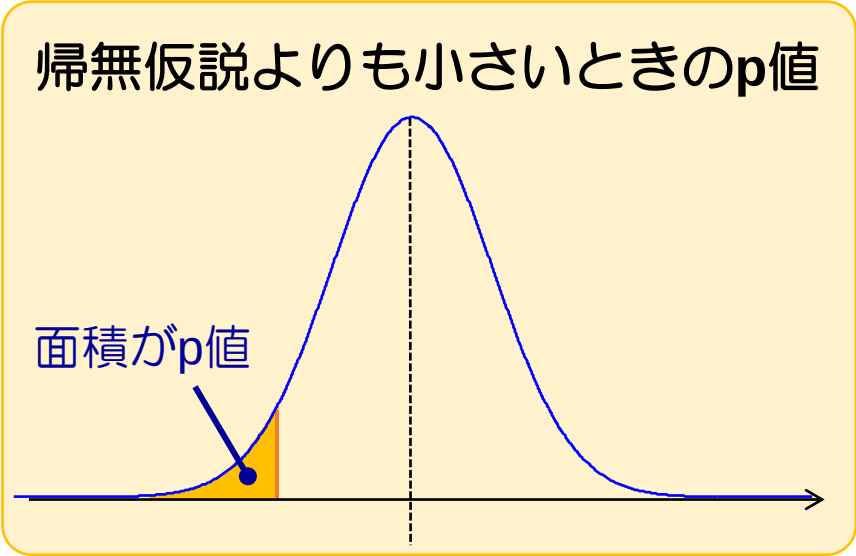
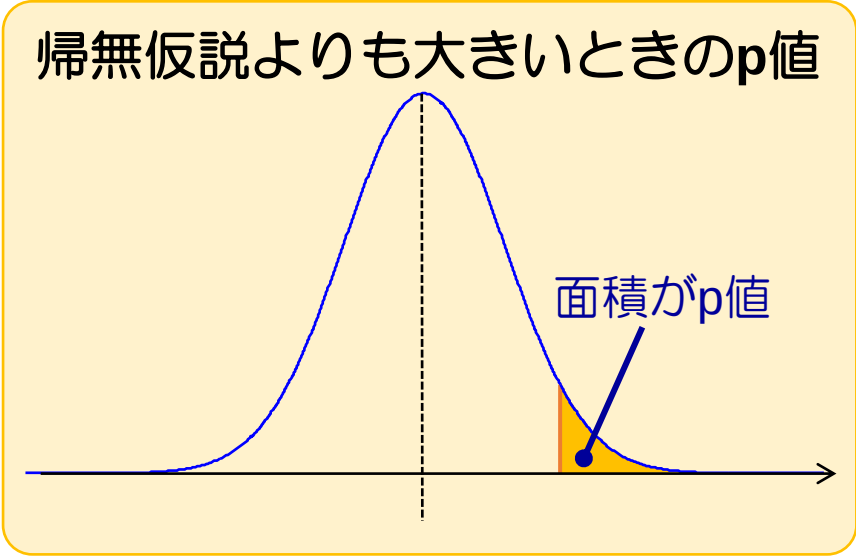
仮説検定では、帰無仮説の確からしさを表す確率(のようなもの)を計算し、その値が非常に小さい(帰無仮説が確からしくない)ことを示すことで、その反対の仮説である対立仮説が正しいことを提示する方法である。

対立仮説が支持されることを「**有意である(significance)**」という。

帰無分布とp値の関係



このとき、どれくらいp値が小さい(帰無仮説が正しいとする確率のようなもの)ければ、帰無仮説が正しいと言えない(有意である)といえるかを決定するしきい値が**有意水準 α** である。



どのように「有意である」と判定するのか？

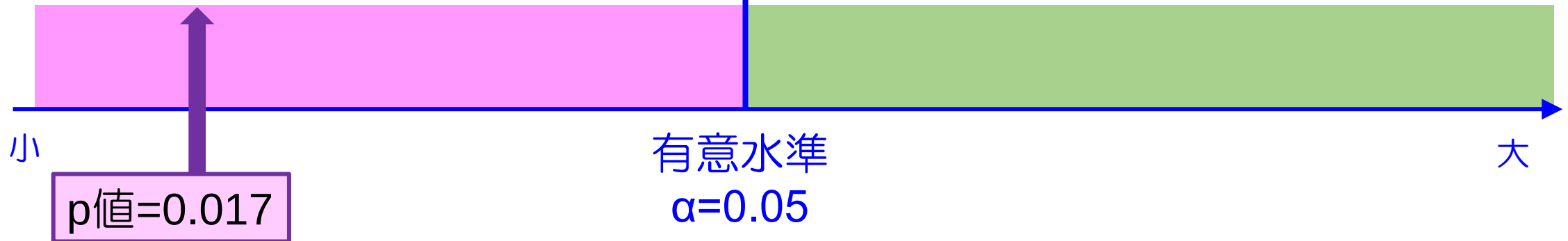
P値 = 0.017 (1.7%)

すなわち、病院Aの平均血清クレアチニン濃度がB病院と同じ4.3(mg/dl)であることが真実であるという可能性は、1.7%である。

有意水準 α とは、この結果が得られる状況が考えられ得るものなのか、そうでないのかを判断するカットオフ値のようなもの。

このような結果が得られるとは考えにくい

このような結果が得られるかもしれない



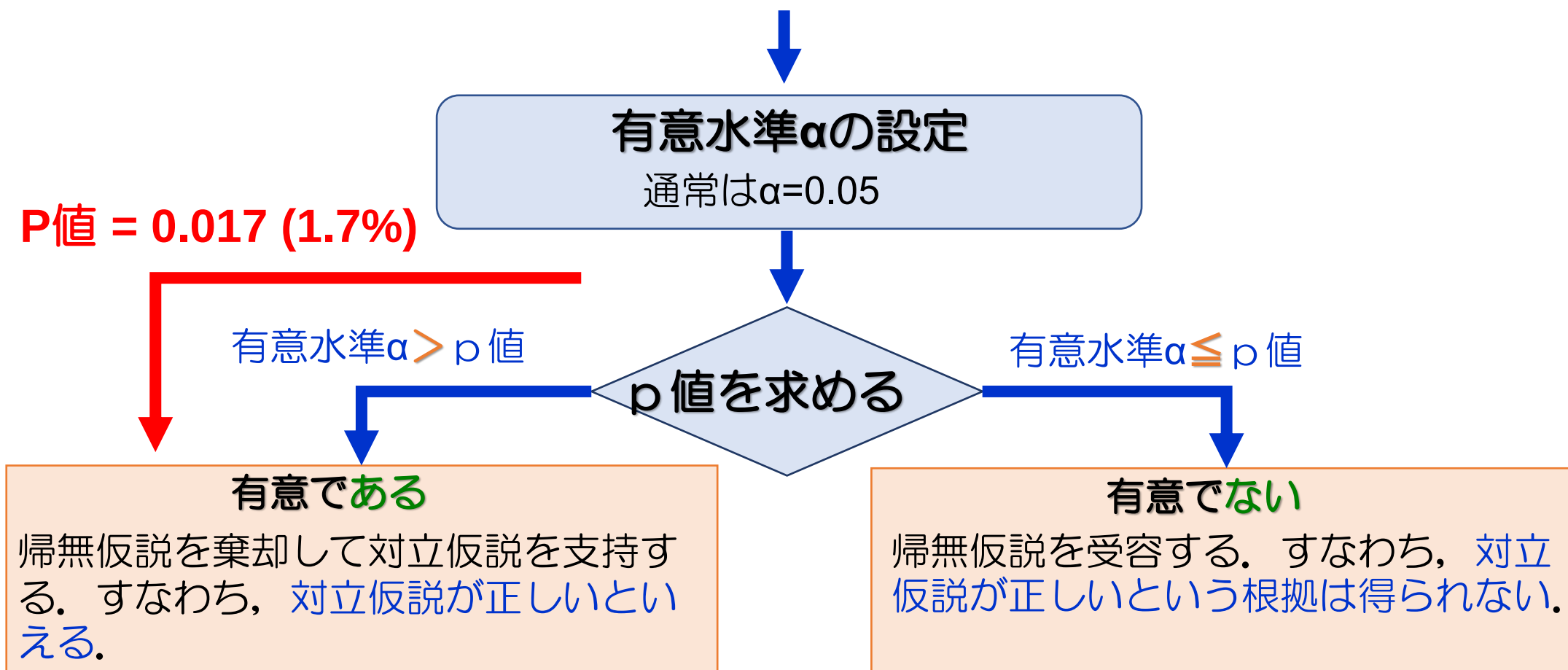
病院Aの平均血清クレアチニン濃度がB病院と同じ(帰無仮説)とすると、本試験の結果が得られる確率は1.7%しかない(有意水準5%未満である)。したがって、帰無仮説は誤っていると判断し(有意)、逆仮説である病院Aの血清クレアチニン濃度は病院Bの4.3(mg/dl)と異なる(対立仮説)と判断する。

p 値(帰無仮説が正しいとする確率のようなもの)がどれくらい小さければ対立仮説が正しいといえるのかを決定する閾値が必要である。これを有意水準 α という。

帰無仮説および対立仮説を設定

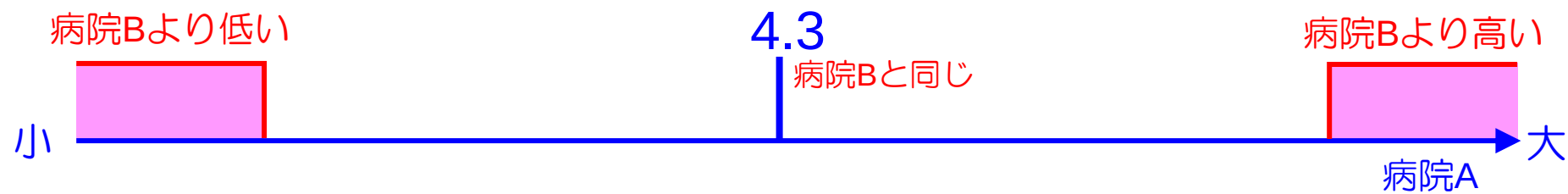
帰無仮説 H_0 : A病院の血清クレアチニン濃度の平均値 μ は, B病院の平均値 μ_0 と同じ。

対立仮説 H_1 : A病院の血清クレアチニン濃度の平均値 μ は, B病院の平均値 μ_0 と異なる。



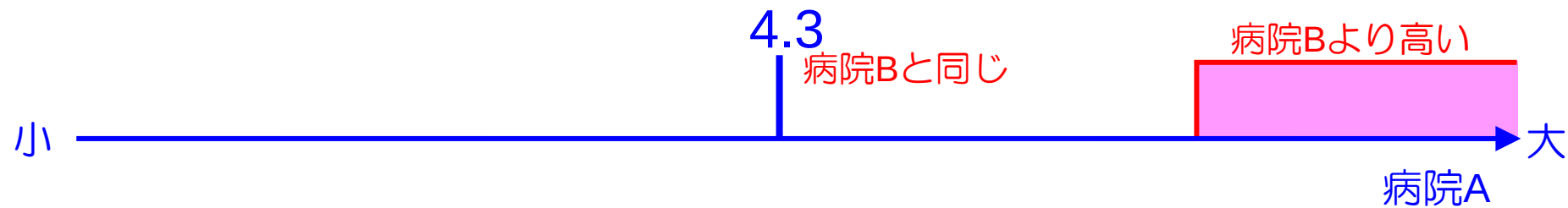
片側対立仮説と両側対立仮説

■ 両側対立仮説：病院Aと病院Bに違いがある。



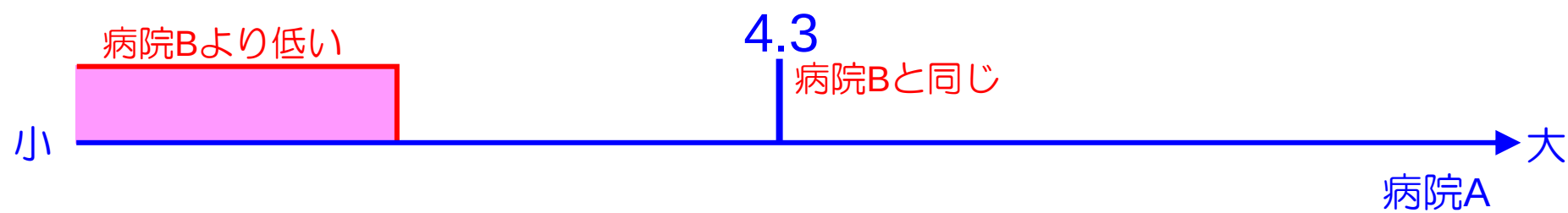
有意水準は「病院Bの4.3(mg/dl)よりも高い確率」「病院Bの4.3(mg/dl)よりも低い確率」の両方で設定されるので、片側では $\alpha/2$ になる。

■ 片側対立仮説(1)：病院Aのほうが病院Bよりも高い。



有意水準は「病院Bの4.3(mg/dl)よりも高い確率」で設定されるので、 α になる。

■ 片側対立仮説(2)：病院Aのほうが病院Bよりも低い。



有意水準は「病院Bの4.3(mg/dl)よりも低い確率」で設定されるので、 α になる。

このような場合はどう考える？

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p値が0.0466で**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？

ダメ!! 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。





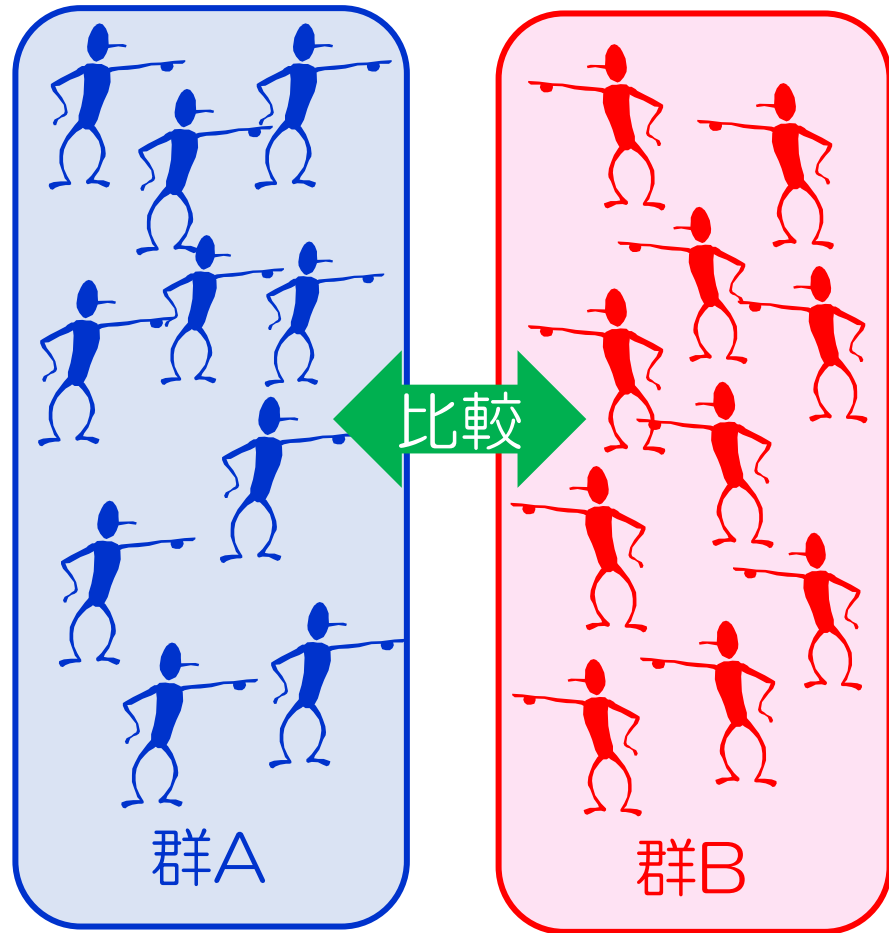
EZRによる演習(1)

テキスト「1.2 ヒストリカル・コントロールとの比較」

量的データにおける解析編(1章)

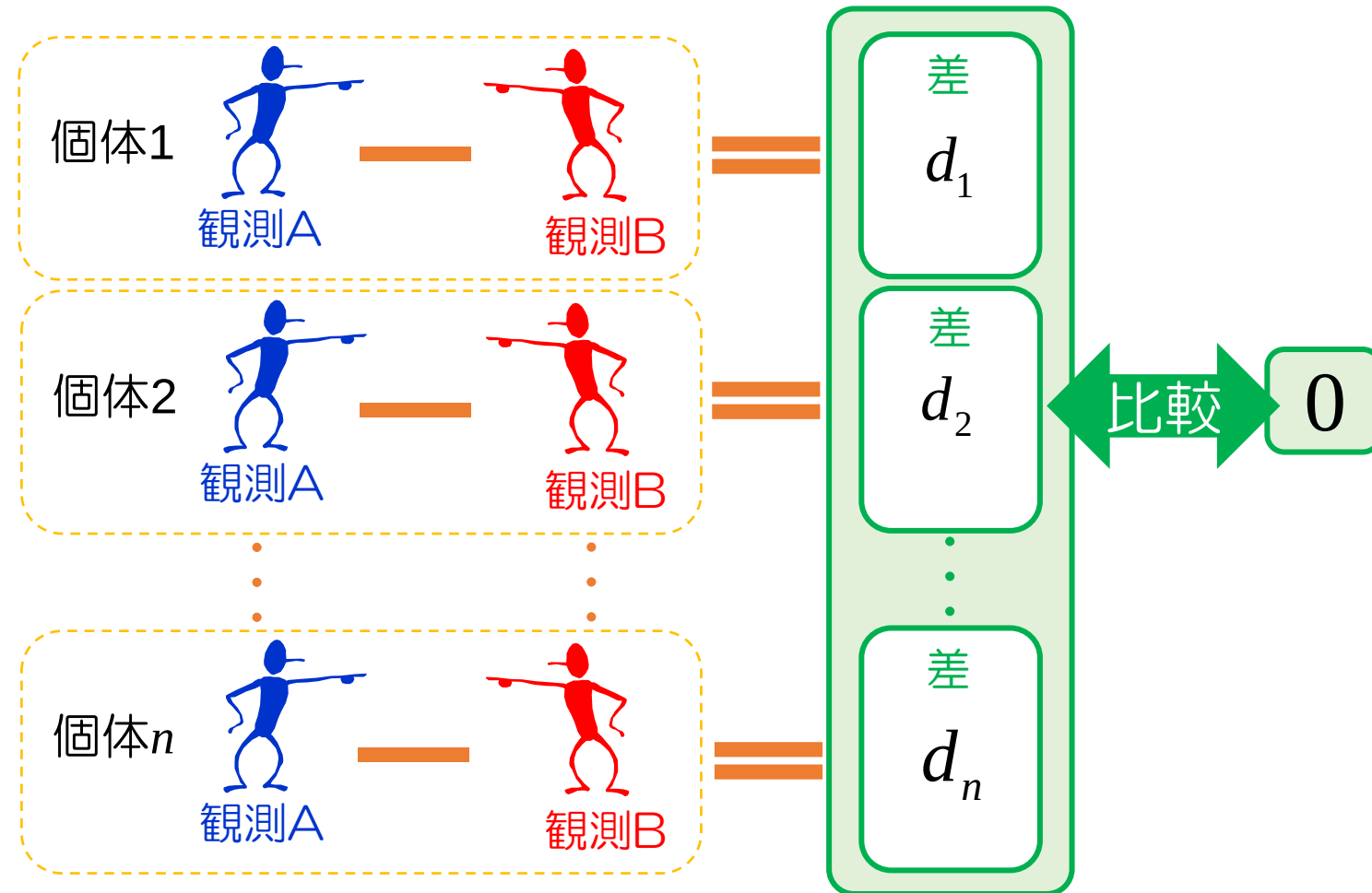
対応のあるデータと対応がない場合

独立2標本



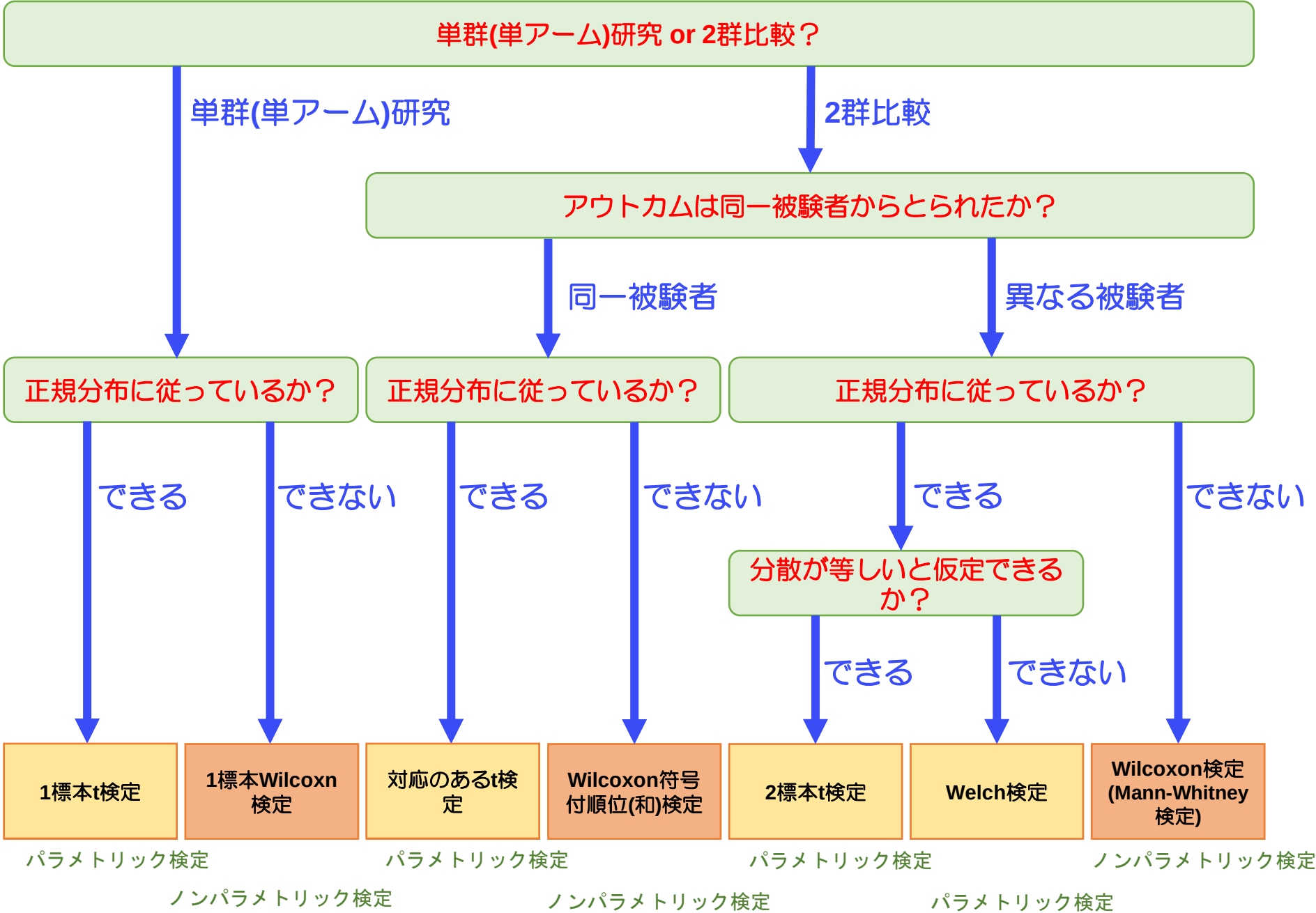
例えば、新薬と既存薬を2群に分けた被験者に投与する場合はこれにあたる。各個体が独立である。

対応のある標本



治療開始前と治療開始後の臨床的指標の差を比較するような場合がこれにあたる。つまり、個体内で比較される。

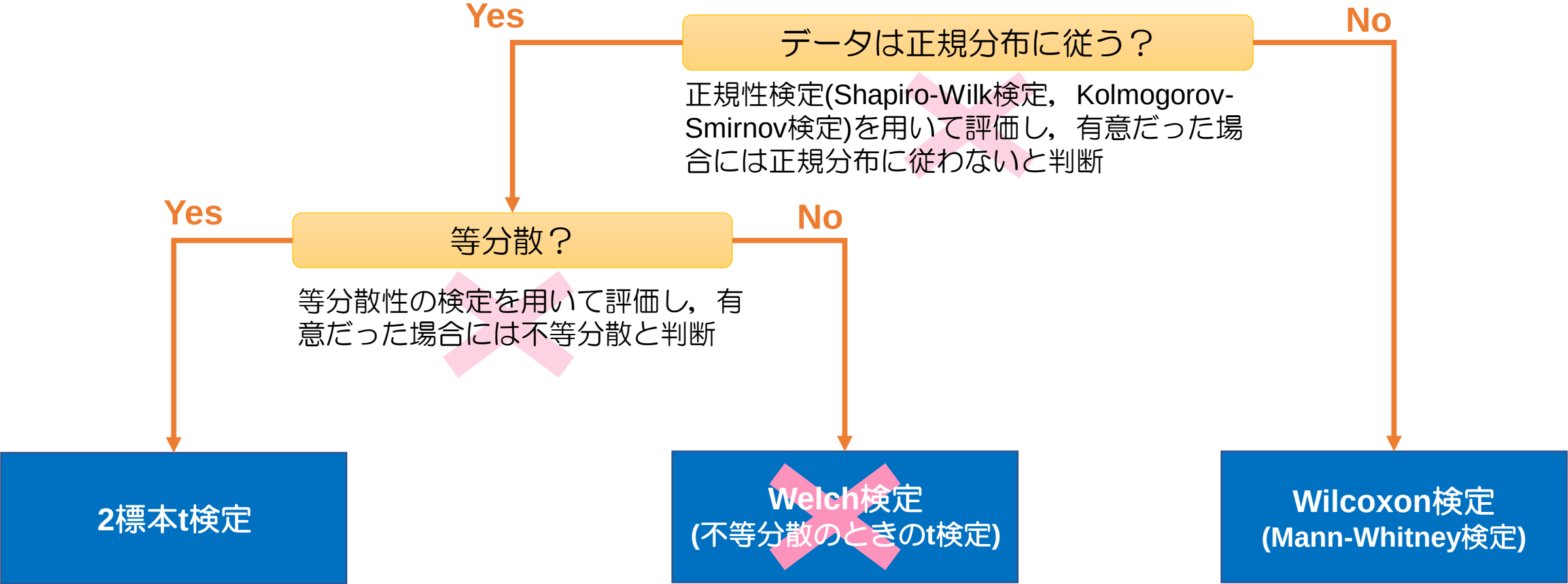
教科書的な統計的検定の取捨選択



教科書的な統計的検定に対するアンチテーゼ

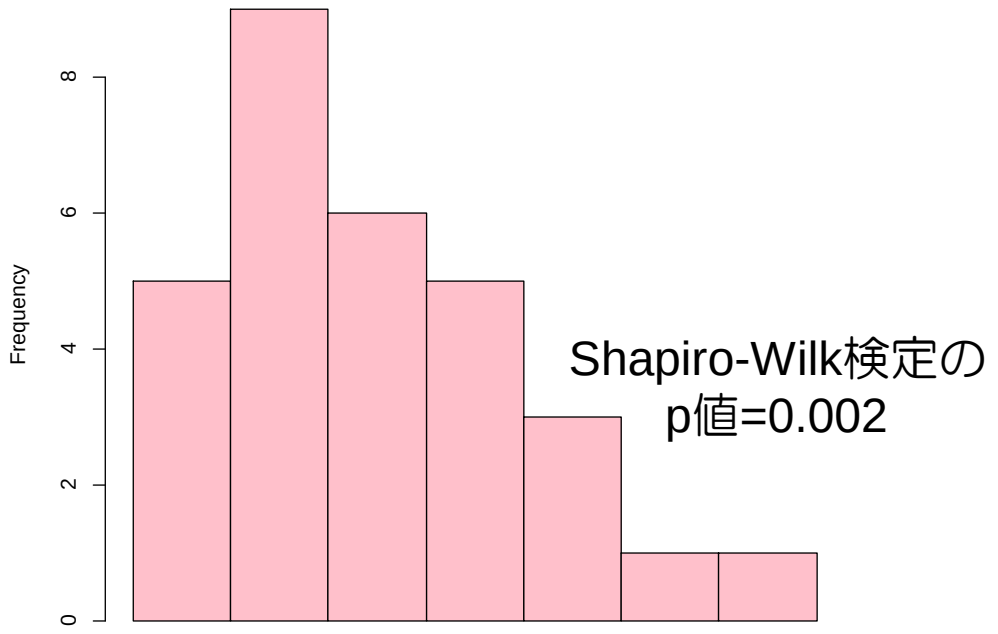


量的データの場合にはどの検定を使えばいいですか？
(2標本t検定, Welch検定(不等分散のときのt検定), Wilcoxon検定(Mann-Whitney検定))



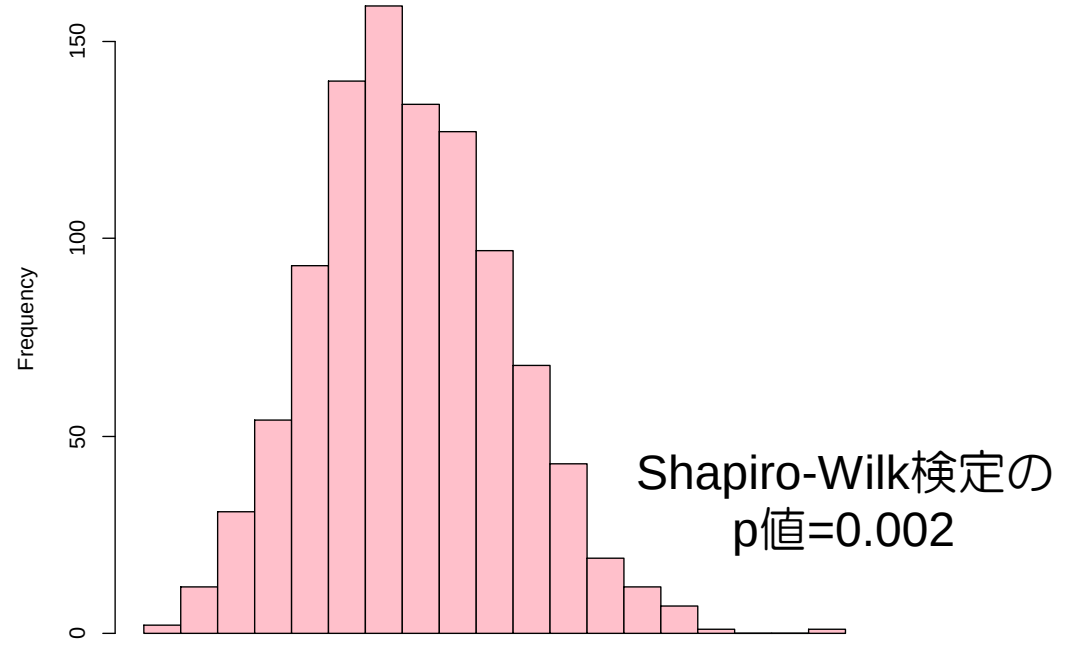
正規性の検定を用いて取捨選択した場合におきえること

N=30の調査研究



ヒストグラムを見る限りでは正規分布に近い形状だと**思いますか？**

N=1000の調査研究



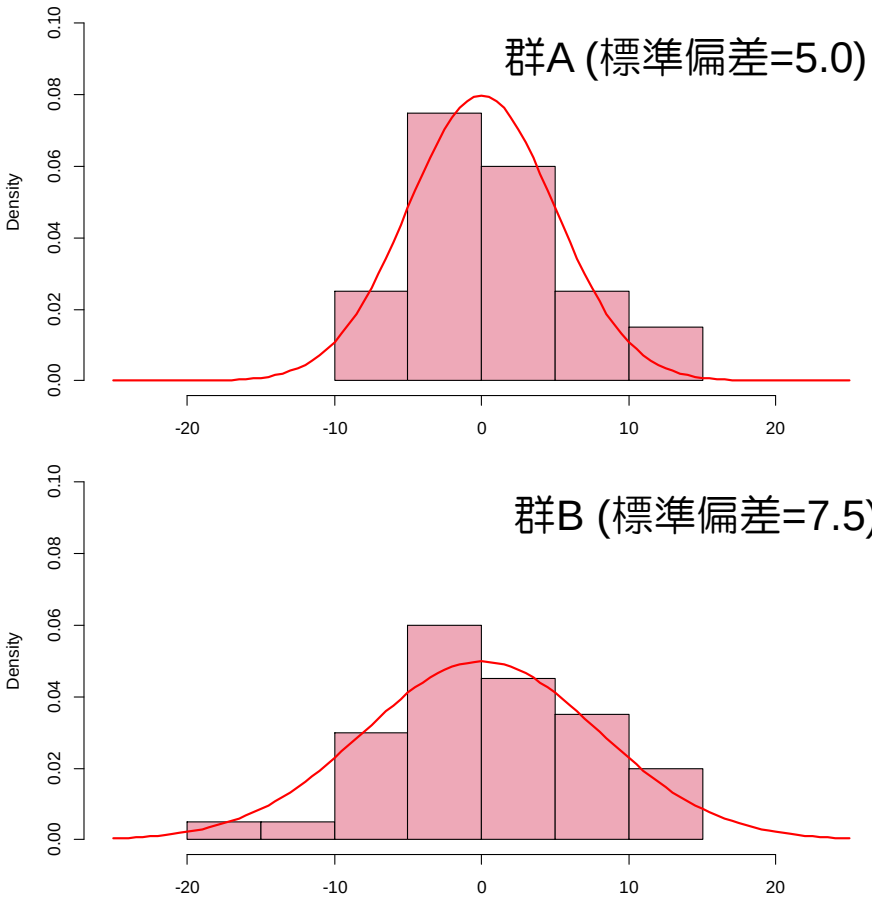
ヒストグラムを見る限りでは正規分布に近い形状だと**思いませんか？**

正規性検定を基準に検定法を選択してしまうと、「データの歪み」だけでなく、「サンプルサイズの大きさ」が影響を及ぼし、サンプルサイズが大きければ、僅かな歪みでも有意になる。そもそも、サンプルサイズで検定法を選択するというのがおかしい。

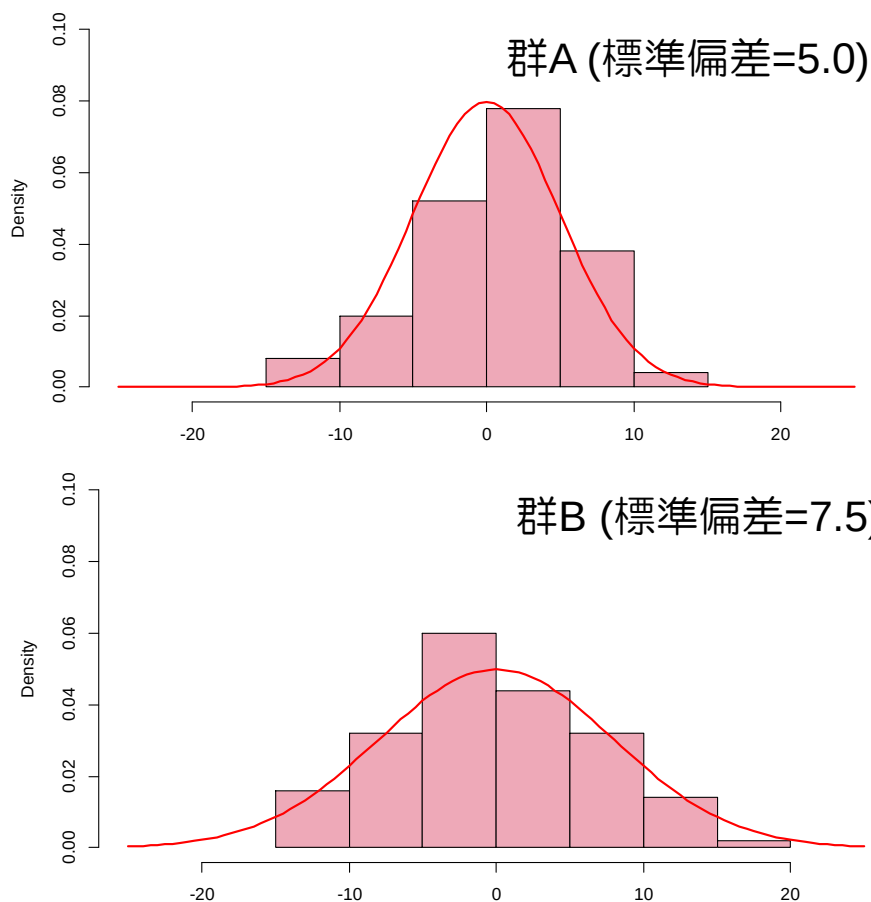
等分散性の検定を用いてt検定とWelch検定を選択することで起きえること

統計の教科書では、等分散性の検定を用いてt検定とWelch検定を取捨選択するような説明をしているものがある。ただし、等分散性の検定による取捨選択については、先ほどの正規性と同様に、サンプルサイズが増えれば有意になりやすくなる問題がある。

各群のサンプルサイズ = 40
等分散性の検定のp値=0.164



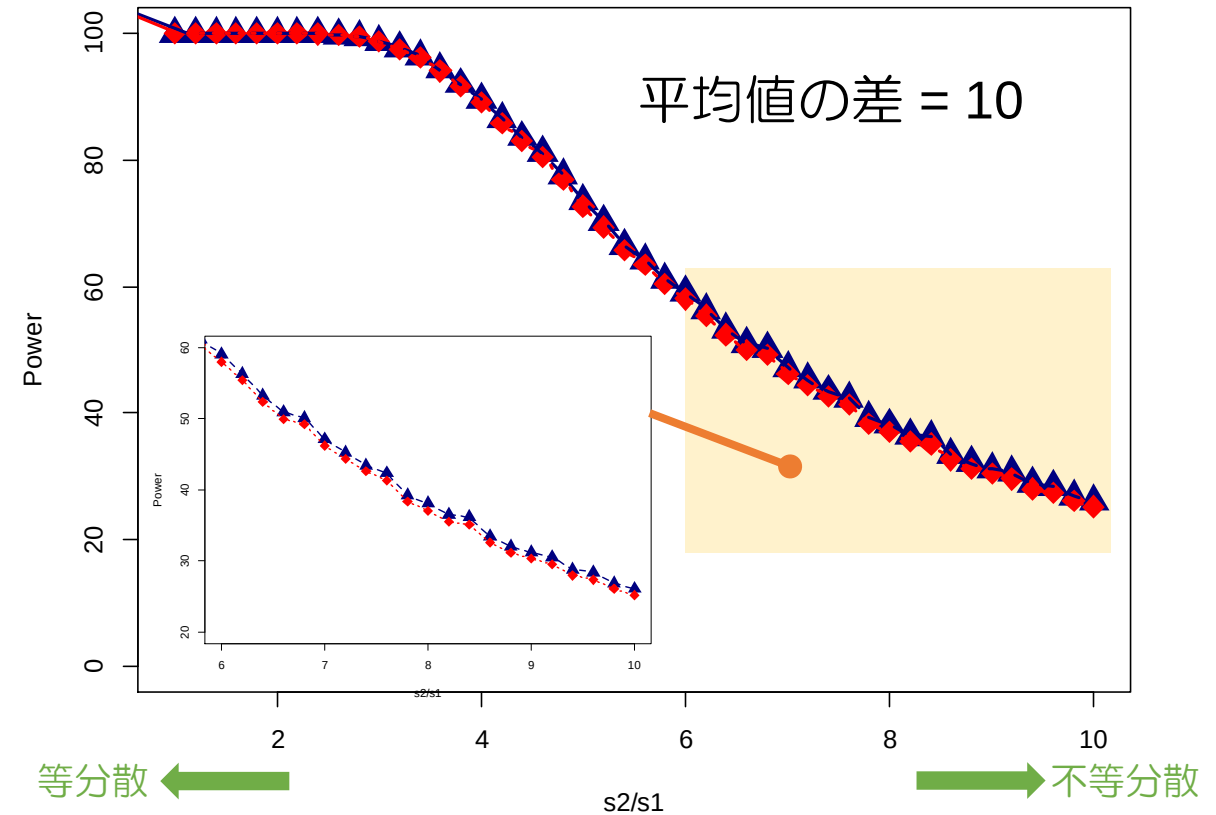
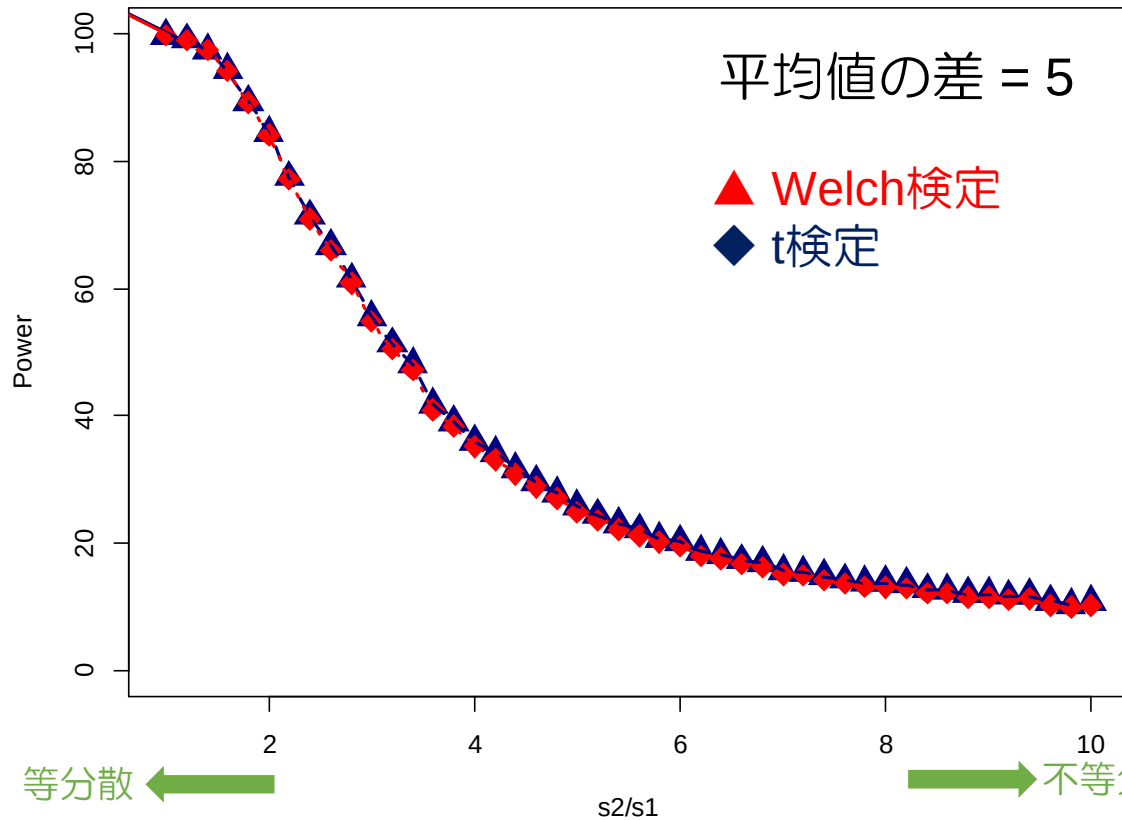
各群のサンプルサイズ = 60
等分散性の検定のp値=0.006



同じ母集団であっても、サンプルサイズの違いによって検定手法を変えるというのは、そもそもおかしいのでは？

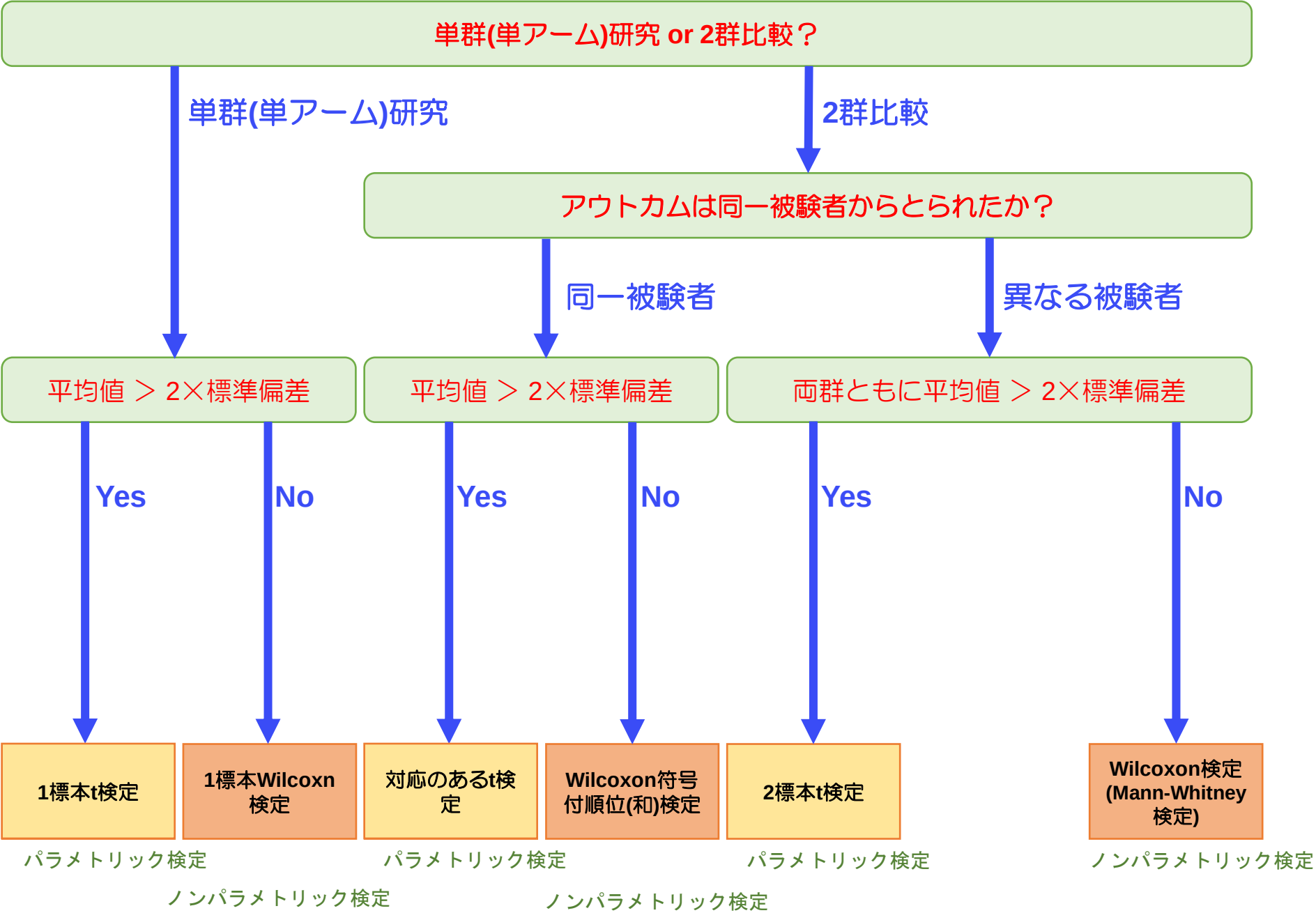
不等分散のときに t 検定とWelch検定の結果に違いが出るのか？

Welch検定は、未知の不等分散のもとでの母平均の比較(Behrens-Fisher問題)に対する近似解の一つであるが、そもそもWelch検定の利用には統計学者からの悲観的な意見が多い。



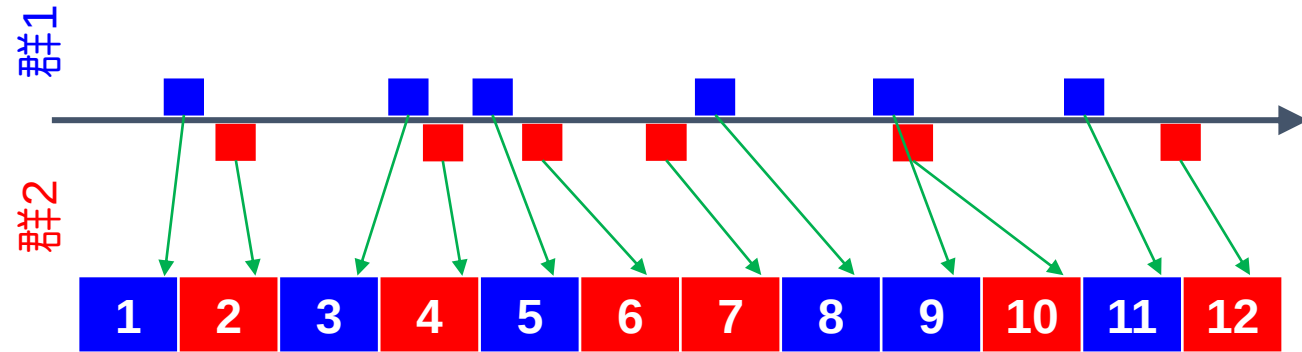
上の二つのグラフは、10,000回のシミュレーションにおいて、t検定とWelch検定が有意になった割合を表している。X軸が標準偏差の比率であり、左端が等分散を表す。右側に行くほど不等分散になる。また、Y軸は有意になった割合(パーセントで表示、実質検出力という)。実質検出力が高い手法のほうが良い方法であると判断される。不等分散であっても、**t検定とWelch検定の性能はほぼ変わらない(しいていえば、僅かにt検定の方が高い)**。

経験則に基づく統計的検定の取捨選択



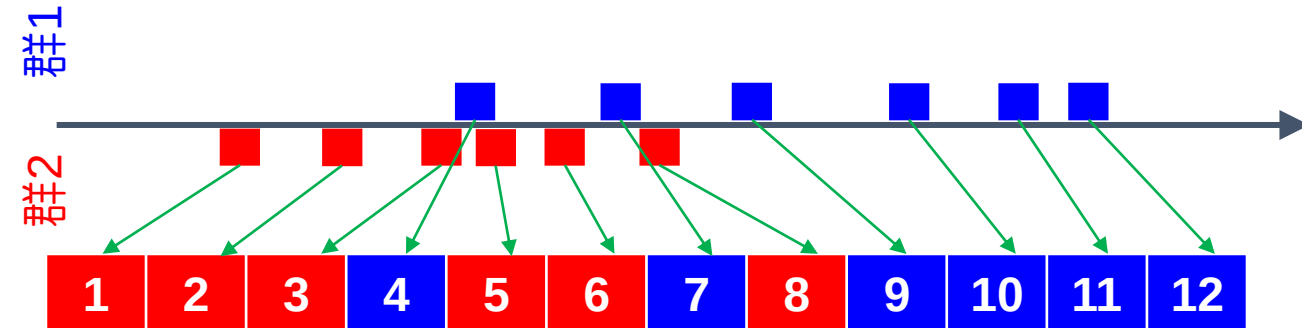
Wilcoxon(Mann-Whitney)検定の意味

群1と群2がほぼ同じである場合(有意でない場合)



群1と群2をプールして小さい順に並べ替えたとき、それぞれの群がおおよそ交互に並んでいる。

群1と群2が異なる場合(有意である場合)

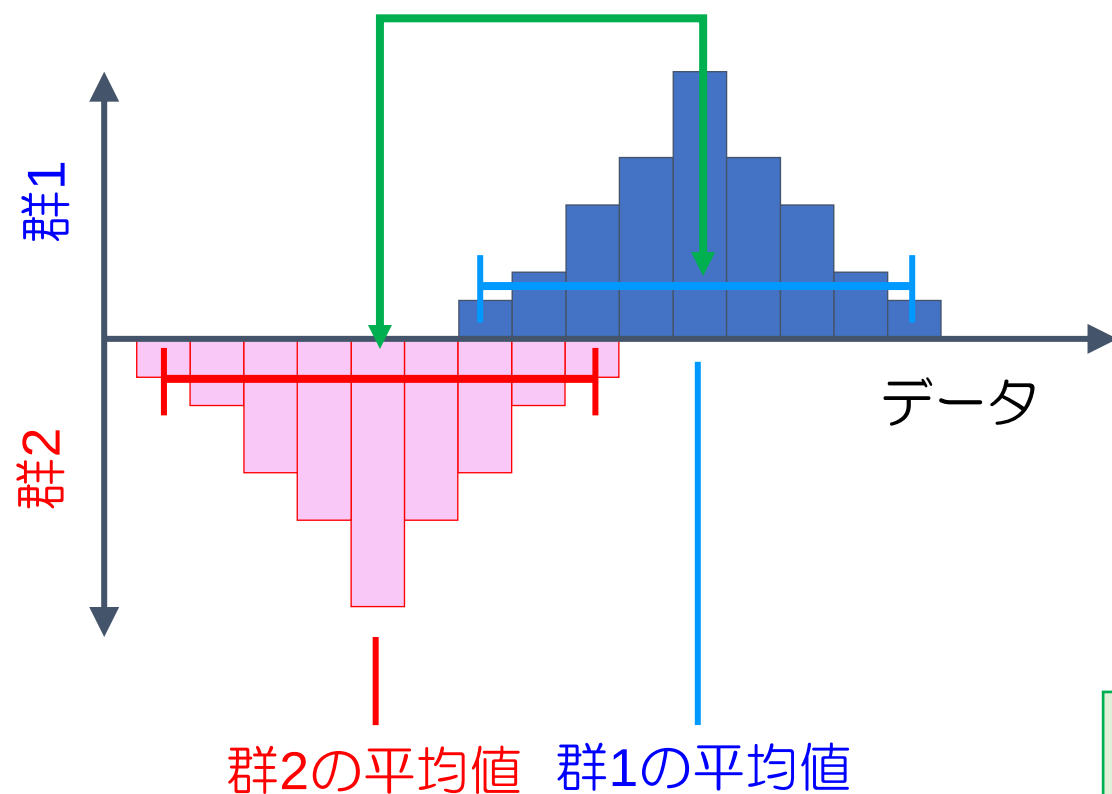


群1と群2をプールして小さい順に並べ替えたとき、群2が順位が小さい側、群1が大きい側に並んでいる。

Wilcoxon検定では、小さい順に並べ替えたときのいずれかの群の順位の和に基づいて検定を行う。Wilcoxon検定は中央値を比較しているわけではなく、分布の相対的な位置関係を評価している(結果の要約に中央値を使うのは便宜上)。

2標本t検定とWilcoxon検定(ノンパラメトリック検定)の違い

Wilcoxon検定では、相対的な位置関係を比較している。



2標本t検定では平均値(それぞれの群を代表する値)を比較している。

■ 2標本t検定で有意であるということ
「2群間の平均値が違う」と解釈できる。

■ Welch検定で有意であるということ
「2群間の相対的な位置関係が違う」と解釈できる。

中央値が違うとは言っていない。ノンパラメトリック検定を用いた場合に、中央値を用いるのは、その他に群の代表値として用いるものがないため。

上記の理由のほかに、分散分析や回帰分析(いわゆる多変量解析)においても、平均値で解釈される。ため、可能な限り2標本t検定を用いるほうが良い。

Wilcoxon (Mann-Whitney)検定のp値の計算方法には数種類存在する

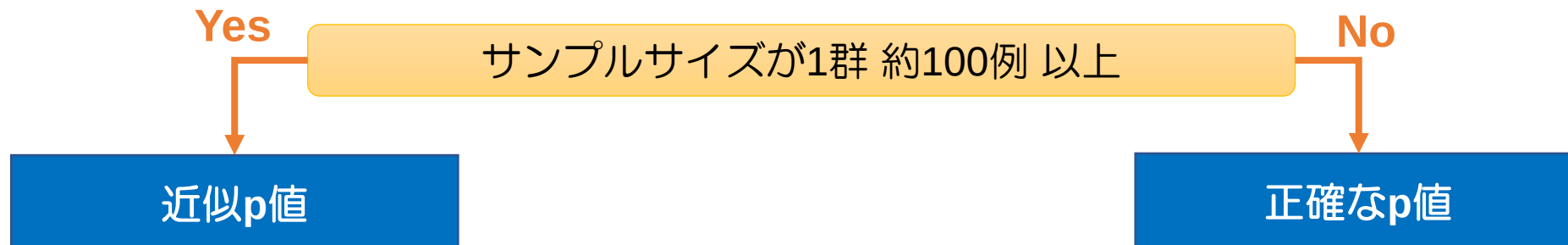
t検定では検定統計量がt分布に従うことが数学的に証明できるため、p値を正しく計算できる。一方で、Wilcoxon検定はp値を計算するための計算式を導出することはできない、

そのため、Wilcoxon検定では以下のいずれかによってp値を計算する。

- 近似式を用いる場合：
 - 正規分布を用いて近似計算を行う。
 - カイ2乗分布を用いて近似計算を行う。
➡ 近似式の計算の理屈は、ほぼ同じだが少しだけ値が違う
- コンピュータを用いて正確なp値(exact p-value)を計算する。

— サンプルサイズが大きければ、近似式でのp値と正確なp値はほぼ同じになるが、サンプルサイズが小さい場合には違いが生じる(近似式での近似精度が悪くなる)。

— 一方で、正確なp値はサンプルサイズが大きいと計算が膨大になり、PCがフリーズする(ときがある)。



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と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：対応がある場合

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
2群の平均値の比較(t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)
線形混合効果モデル

対応のあるt検定

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

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

「差は(第1の変数)-(第2の変数)」である

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
 - 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)
- 線形混合効果モデル

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 ☐ 差 < 0 ☐ 差 > 0 ☒ はい(t検定) ☐ いいえ(Welch検定)

↓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

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)
- 連続変数の傾向性検定(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 ☐ 差 < 0 ☐ 差 > 0 ☒ デフォルト ☐ 正確 ☐ 正規近似 ☐ 連続修正を用いた正規近似 ☐ Brunner-Munzel検定

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

EZRの更新に伴い、Brunner-Munzel検定が追加された
(バラツキが異なる場合の順位検定)

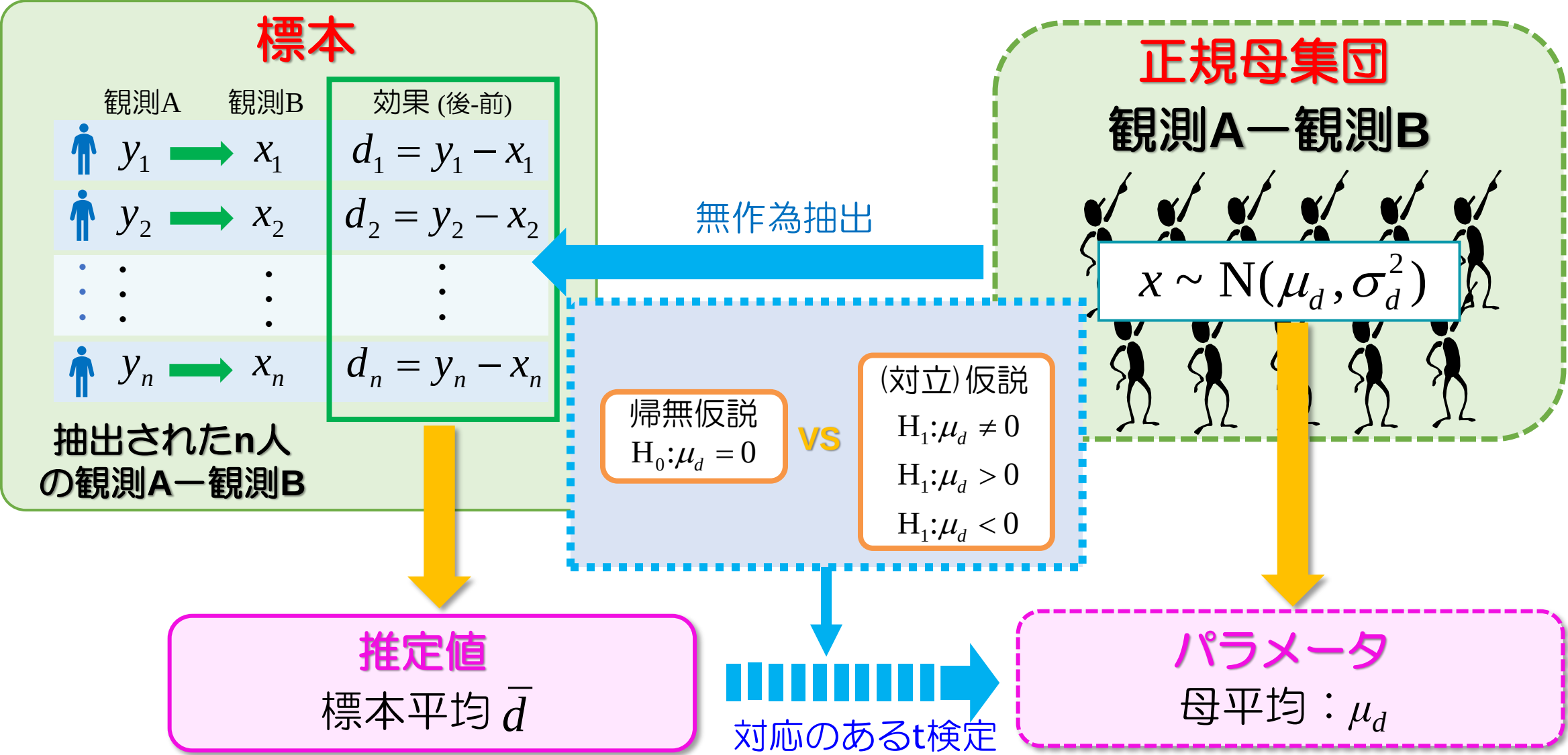


EZRによる演習(2)

「1.3.2 2標本における母平均の比較(2標本検定, Welch検定)」

「1.3.4 2標本におけるノンパラメトリック検定(Wilcoxon検定)」

対応のある場合の母平均に対する統計的推測：対応のあるt検定

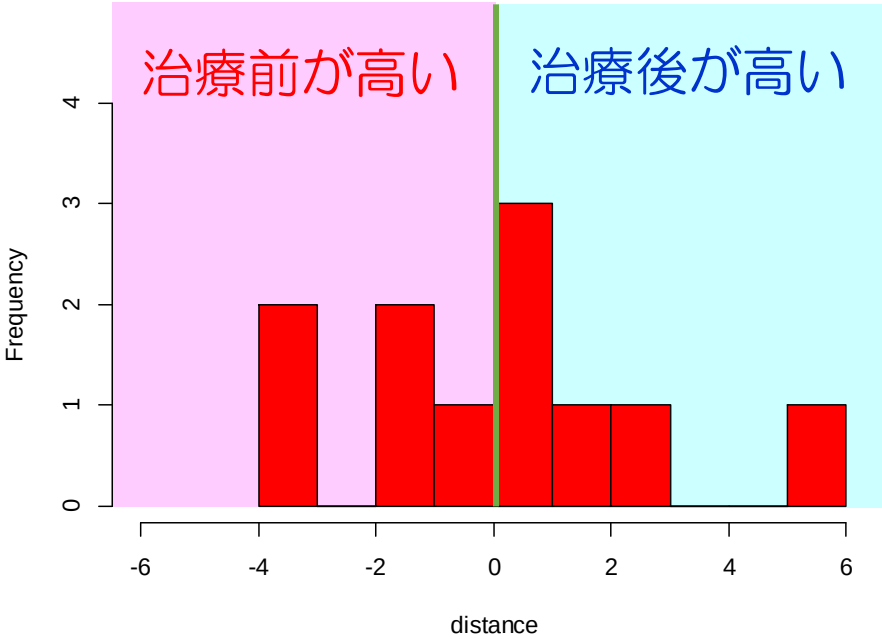


対応のある場合のノンパラメトリック検定：Wilcoxon符号付順位和検定

有意でない状況

No	前	後	差	No	前	後	差
1	5	4	-1	7	1	2	1
2	1	2	1	8	9	6	-3
3	7	8	1	9	4	3	-1
4	9	5	-4	10	2	2	0
5	3	9	6	11	3	6	3
6	2	4	2				

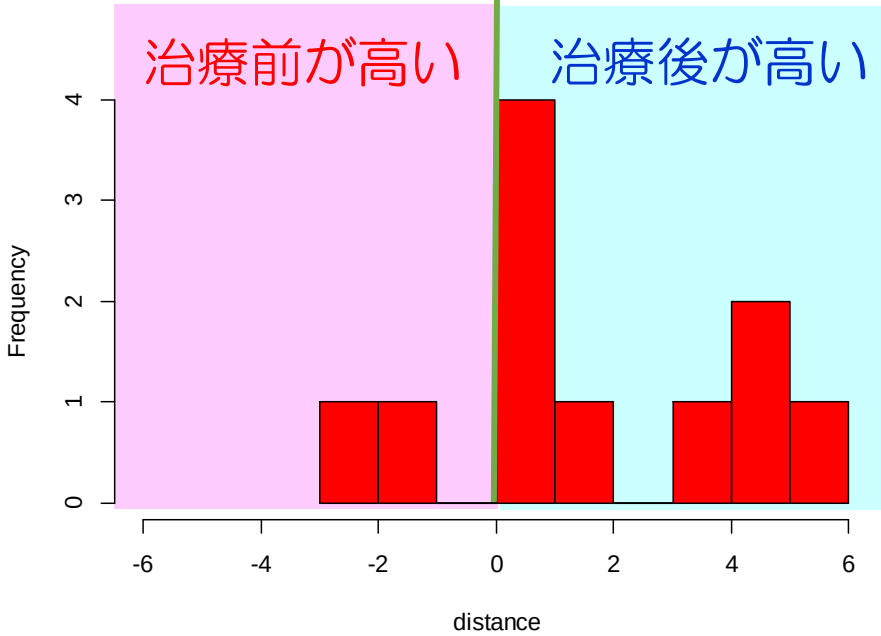
0



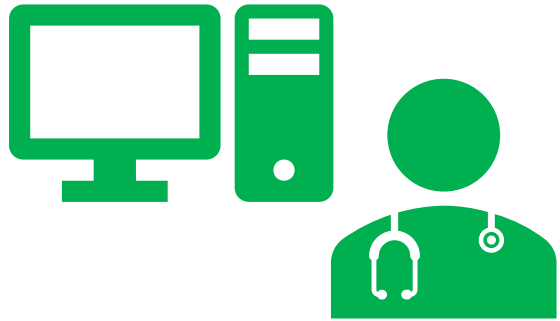
有意である状況

No	前	後	差	No	前	後	差
1	3	4	1	7	1	2	1
2	1	2	1	8	2	6	4
3	3	8	5	9	4	3	-1
4	4	5	1	10	5	2	-3
5	3	9	6	11	1	6	5
6	2	4	2				

0



Wilcoxon符号付順位和検定では、青色の領域の数と赤色の領域の数の偏りを検定している。



EZRによる演習(3)

「1.4.1 対応のあるt検定」

「1.4.2 Wilcoxon符号付順位検定」

EZRに実装されている多群比較(分散分析)

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) 線形混合効果モデル



今回のセミナーでは一元配置の分散分析のみ演習を行います

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

線形混合効果モデルの説明をテキストver.3において追記

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

分散分析の概念図

分散分析とは、3群以上の母平均に違いがあるか否かを評価する方法である。

分散分析における仮説

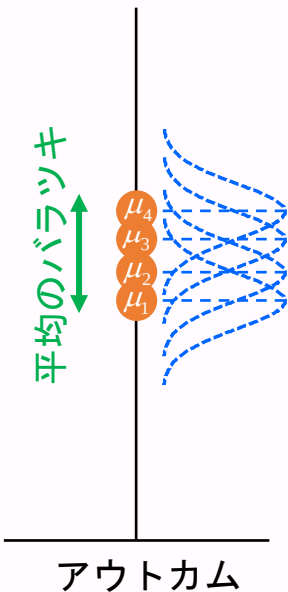
帰無仮説 H_0 ：すべての群の母平均が等しい

対立仮説 H_1 ：帰無仮説 H_0 ではない

なお、2群の場合の分散分析とt検定の結果は一致する。

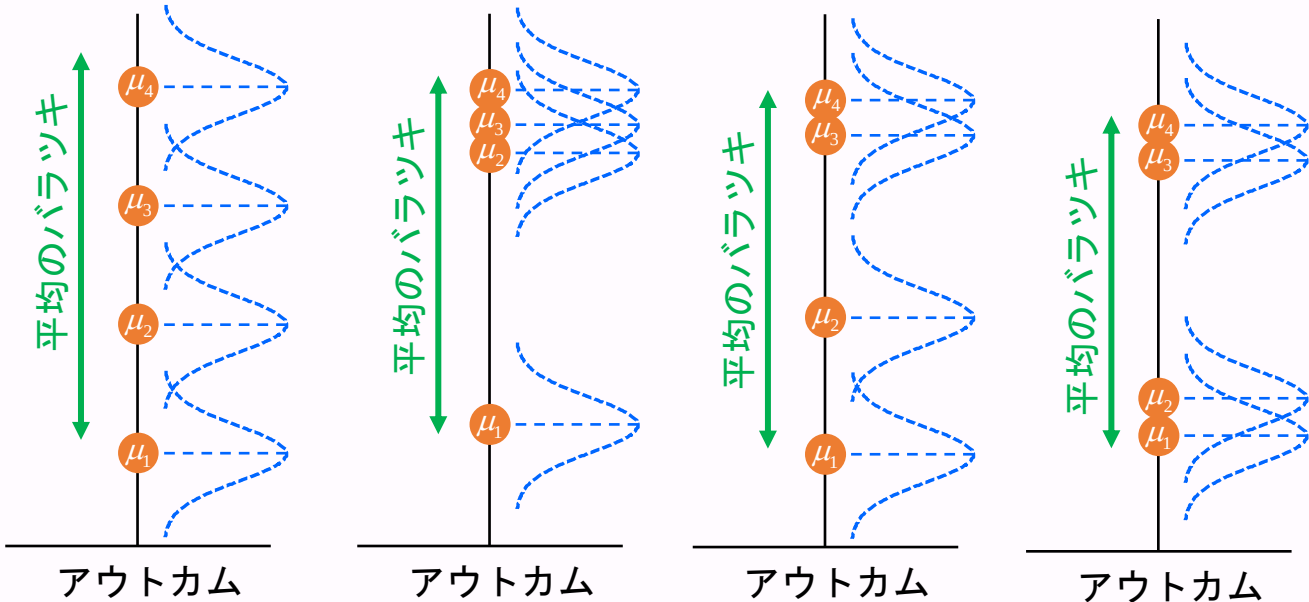
帰無仮説 H_0 が棄却できない(有意でない) 状況

平均値がほぼ同じ場合には、平均のバラツキ(=分散)が小さく、すべて同じであれば、0になる。

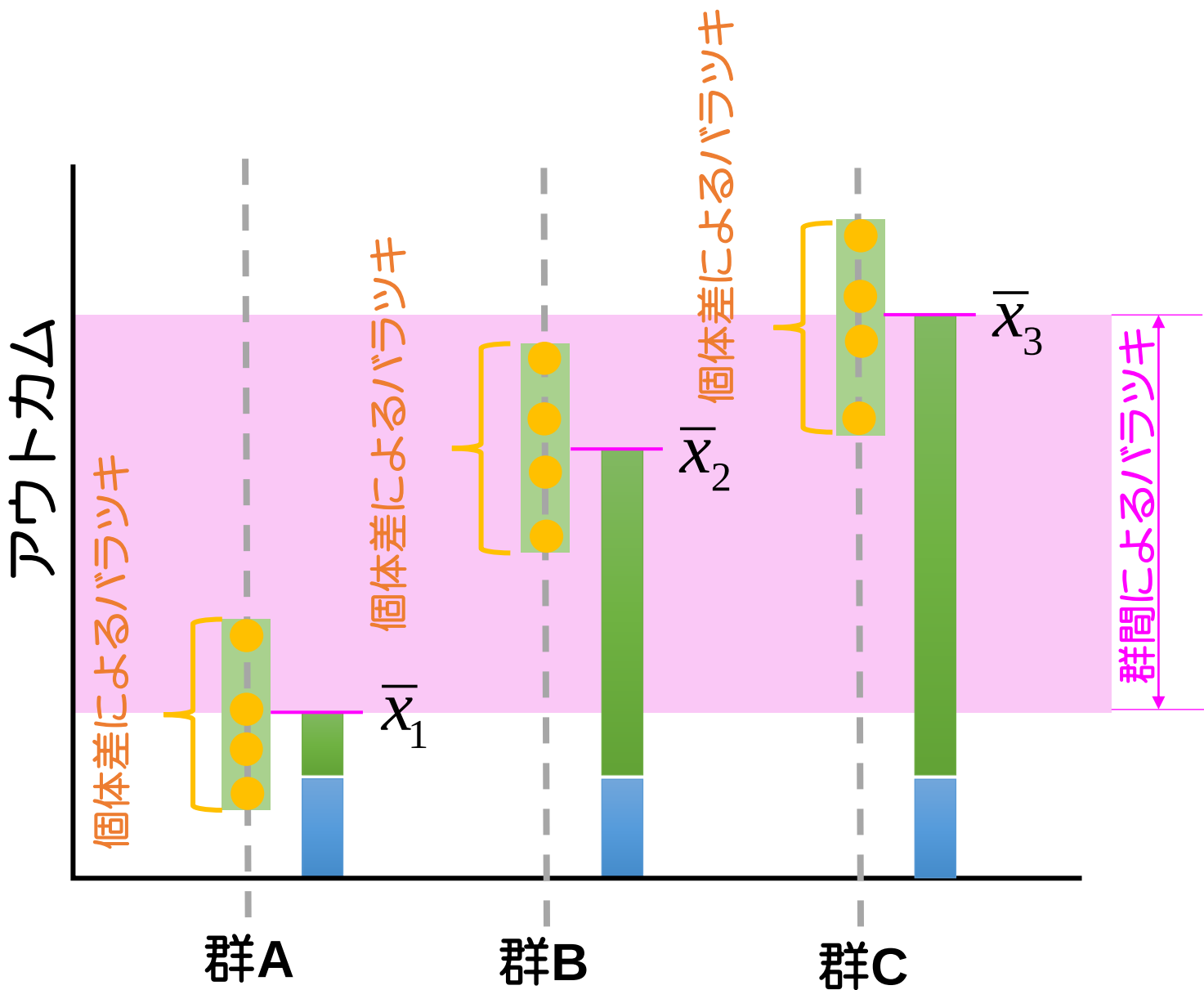


対立仮説 H_1 が正しい(有意)な状況

平均値が異なる場合には、平均のバラツキ(=分散)が大きい。



分散分析表の検定(F検定)の意味



分散分析では個体差によるバラツキ(分散)と群間のバラツキ(分散)が評価される。

□ 有意差が認められる場合

群間によるバラツキ(分散) $>$ 個体差によるバラツキ(分散)

□ 有意差が認められない場合

群間によるバラツキ(分散) $<$ 個体差によるバラツキ(分散)

分散分析ではこの関係性を応用して
$$F\text{値} = \frac{\text{群間によるバラツキ (分散)}}{\text{個体差によるバラツキ(分散)}}$$

が評価される。

分散分析の例

12人の被験者に対して、 $5\mu\text{g/ml}$ 、 $10\mu\text{g/ml}$ 、 $20\mu\text{g/ml}$ の濃度の薬剤を注入した。濃度によって疾病への効果に違いがあるのだろうか。

$5\mu\text{g/ml}$	$10\mu\text{g/ml}$	$20\mu\text{g/ml}$
64	82	55
72	78	64
68	77	66
77	85	49

分散分析表

要因	平方和	自由度	平均平方和	F値
A (因子)	969.5	2	484.8	13.52
E (誤差)	322.7	9	35.86	
全体	1292.2	11		

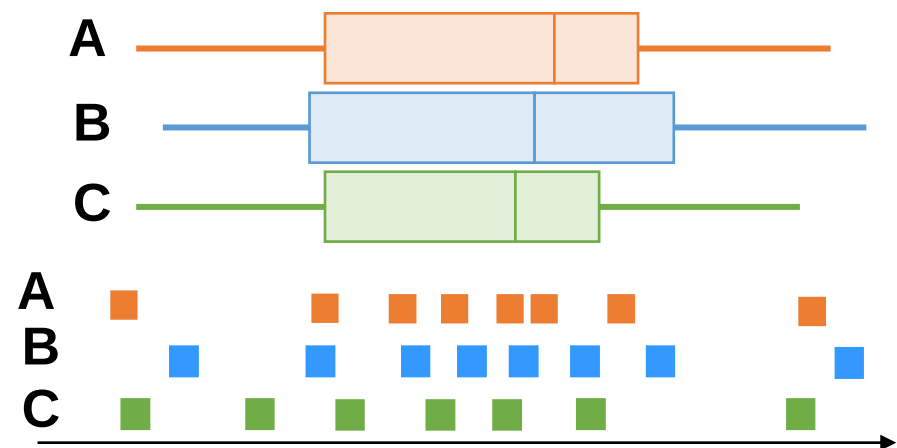
群間によるバラツキ(分散)

F 値に基づいて計算された p 値は 0.002 である。したがって、帰無仮説が棄却され、対立仮説が支持される。すなわち、濃度による効果の差異に違いが認められた。

個体差によるバラツキ(分散)

ノンパラメトリックな多群比較：Kruskal-Wallis検定

□ 有意でない状況

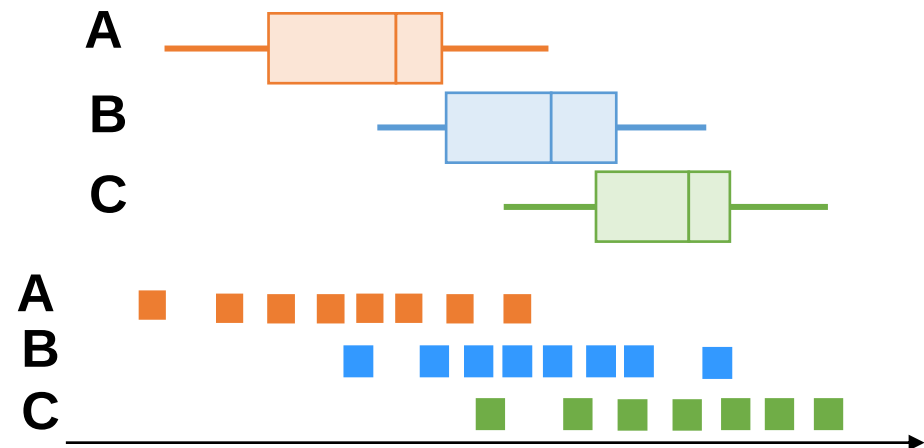


小さい順に並べると



データをプールして順位をつける。そしてその
順位の平均値をとる。差がなければ、**順位
の平均値が3群ともほぼ同じになるはず**である。

□ 有意である状況



小さい順に並べると



データをプールして順位をつける。そしてその
順位の平均値をとる。差があれば、**順位
の平均値がいずれの群で違いがでるはず**である。

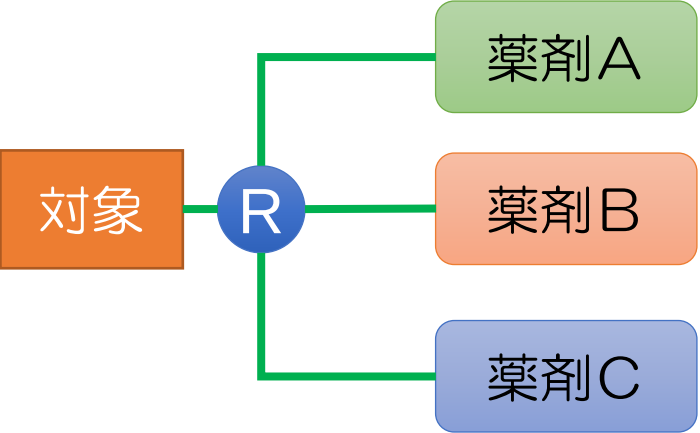
プールしたデータで順位をつけたときの群毎の平均値(順位和)を利用したのが
Kruskal-Wallis検定である。

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？

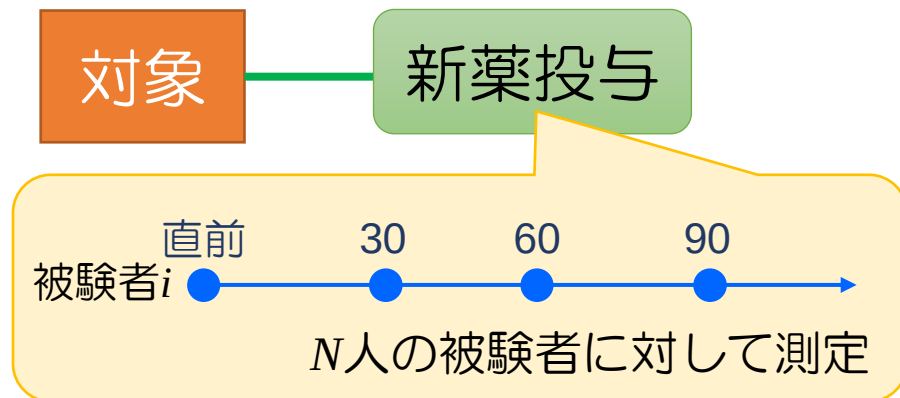
- パラメトリック検定：一元配置の分散分析
EZR：3群以上の間の平均値の比較(一元配置分散分析one-way ANOVA)
- ノンパラメトリック検定：Kruskal-Wallis検定
EZR：3群以上の間の比較(Kruskal-Wallis検定)

分散分析の種類(2/4)：繰り返し測定データの解析(1)

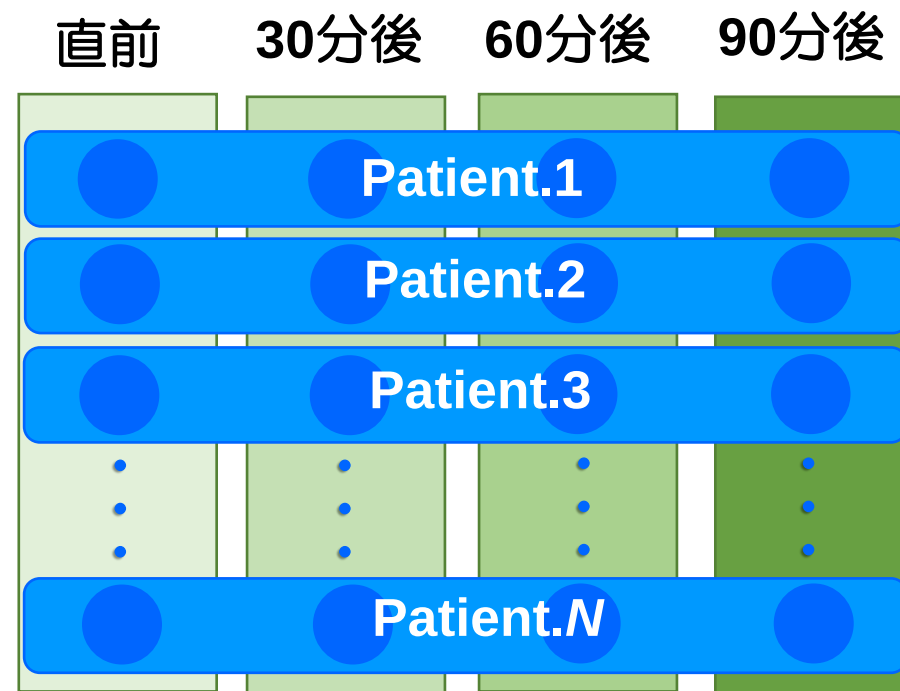


Clinical Question

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



■ パラメトリック検定：繰り返し測定分散分析

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)
線形混合モデル 

■ ノンパラメトリック検定：Friedman検定

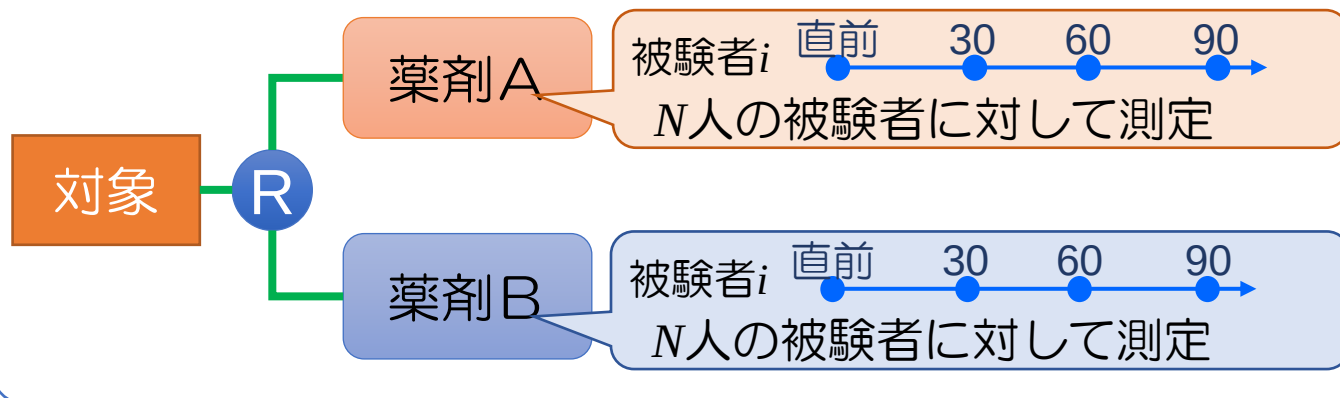
EZR：対応のある3群以上の間の比較(Friedman検定)

分散分析の種類(3/4)：繰り返し測定データの解析(2)

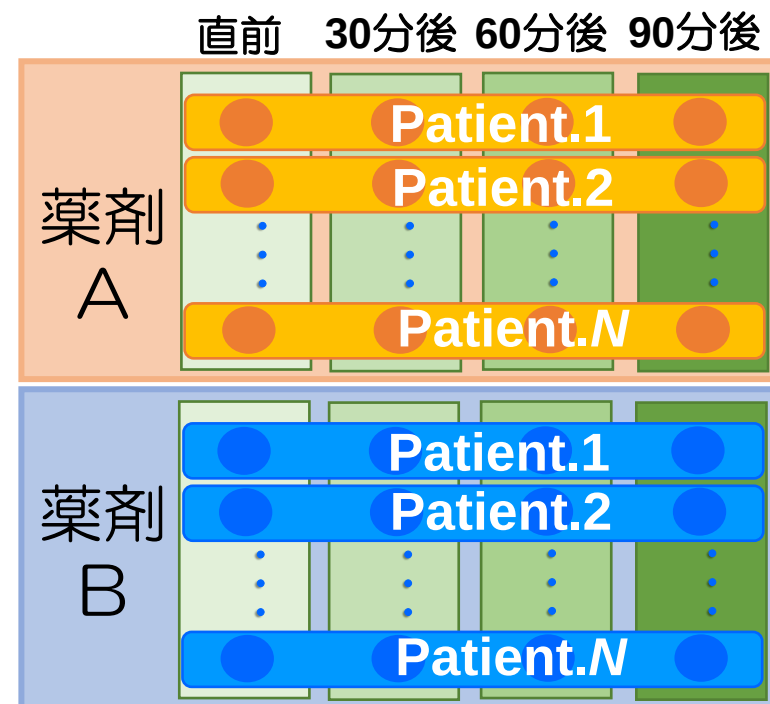


Clinical Question

神経障害性疼痛患者がランダムに割り付けられ、2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？
- (時間×薬剤の交互作用があるか？)
- VASに経時的変化があるか？

■ パラメトリック検定：繰り返し測定分散分析

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), 線形混合モデル

■ ノンパラメトリック検定：存在しない

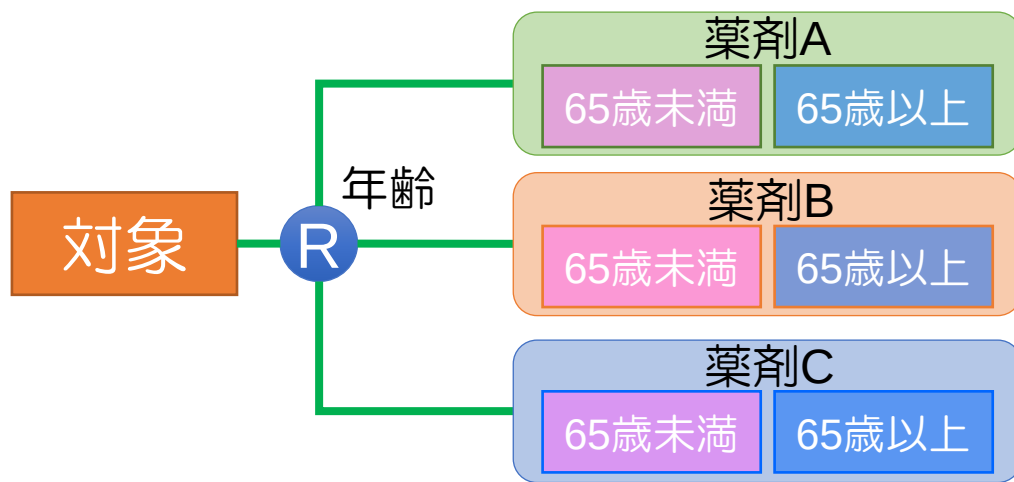
EZR：存在しない

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

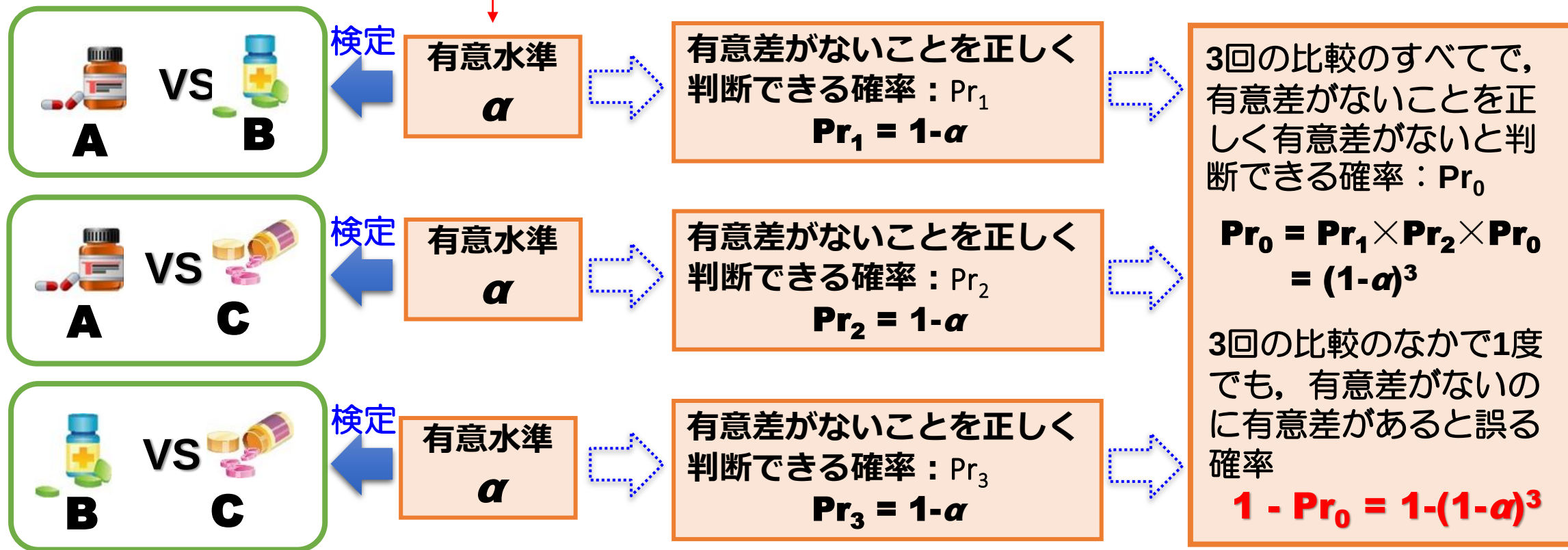
EZR：複数の因子での平均値の比較(多元配置分散分析multi-way ANOVA)

■ ノンパラメトリック検定：存在しない

EZR：存在しない

多重比較を数学的に考える

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 を次のように考える:

$\alpha = 0.05$ (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)

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、

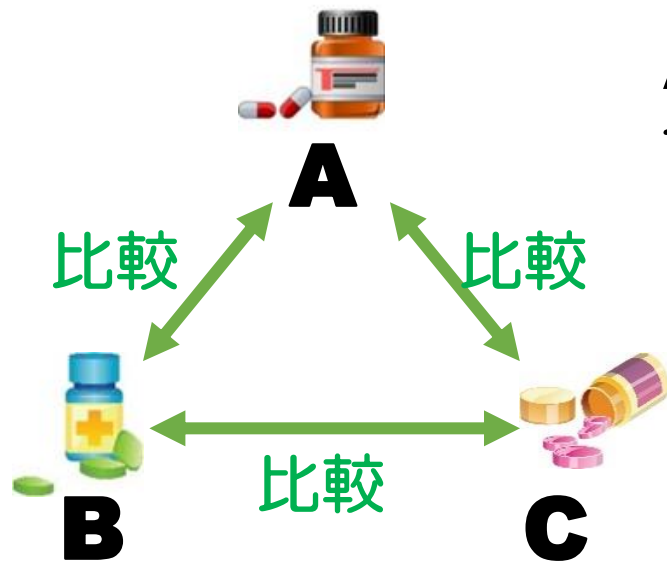
$$1 - (1 - \alpha)^3 = 1 - (1 - 0.05)^3 = 1 - (1 - 0.05)^3 = 0.142$$

となり、誤りの確率が増大してしまう。

多重比較の取捨選択について

多重比較の方法には、(1) p値(有意水準 α)に基づく方法、(2) 正規分布(or ノンパラメトリック)に基づく方法がある

P値に基づく方法(1)：Bonferroniの方法



A群, B群, C群の対比較の場合で全体での有意 p が α の場合には

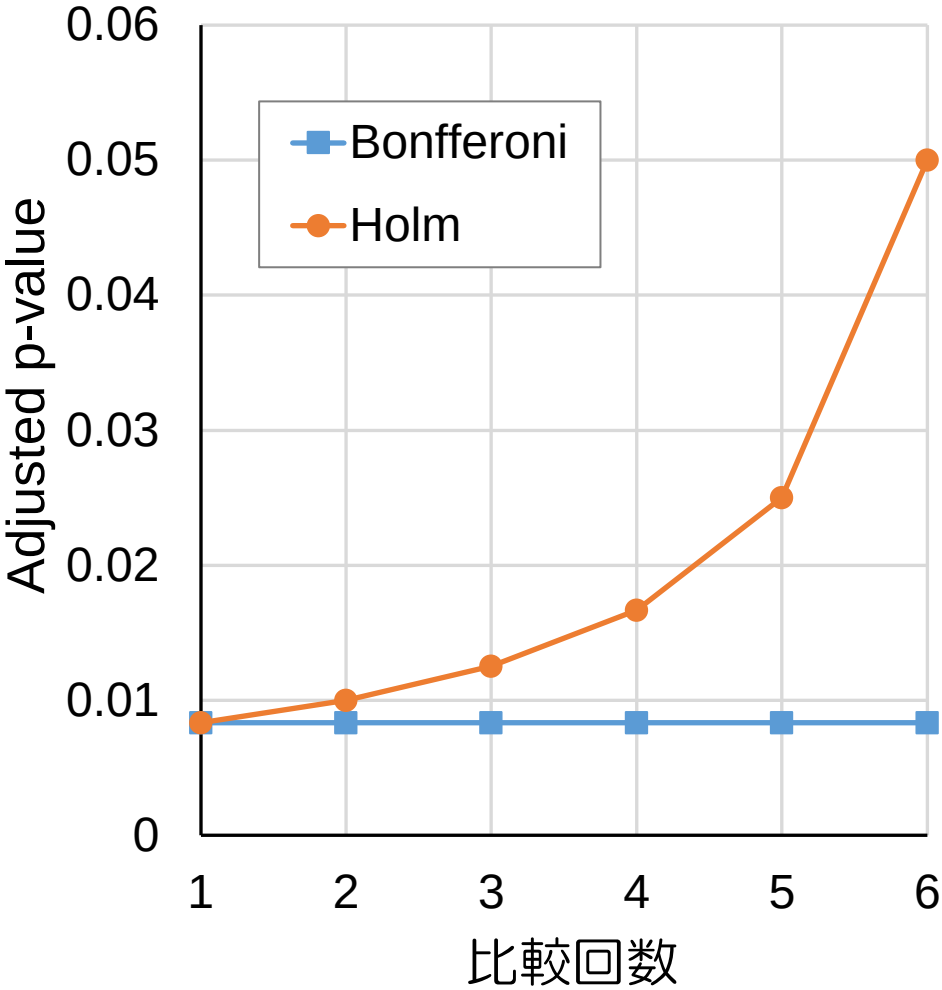
- A群 対 B群 → 有意水準 $\alpha/3$ と比較(or p値を3倍)
- A群 対 C群 → 有意水準 $\alpha/3$ と比較(or p値を3倍)
- B群 対 C群 → 有意水準 $\alpha/3$ と比較(or p値を3倍)

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Bonferroniの方法である。多重比較が簡単のため、最も用いられる方法の一つである。

P値に基づく方法(2)：Holmの方法

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。

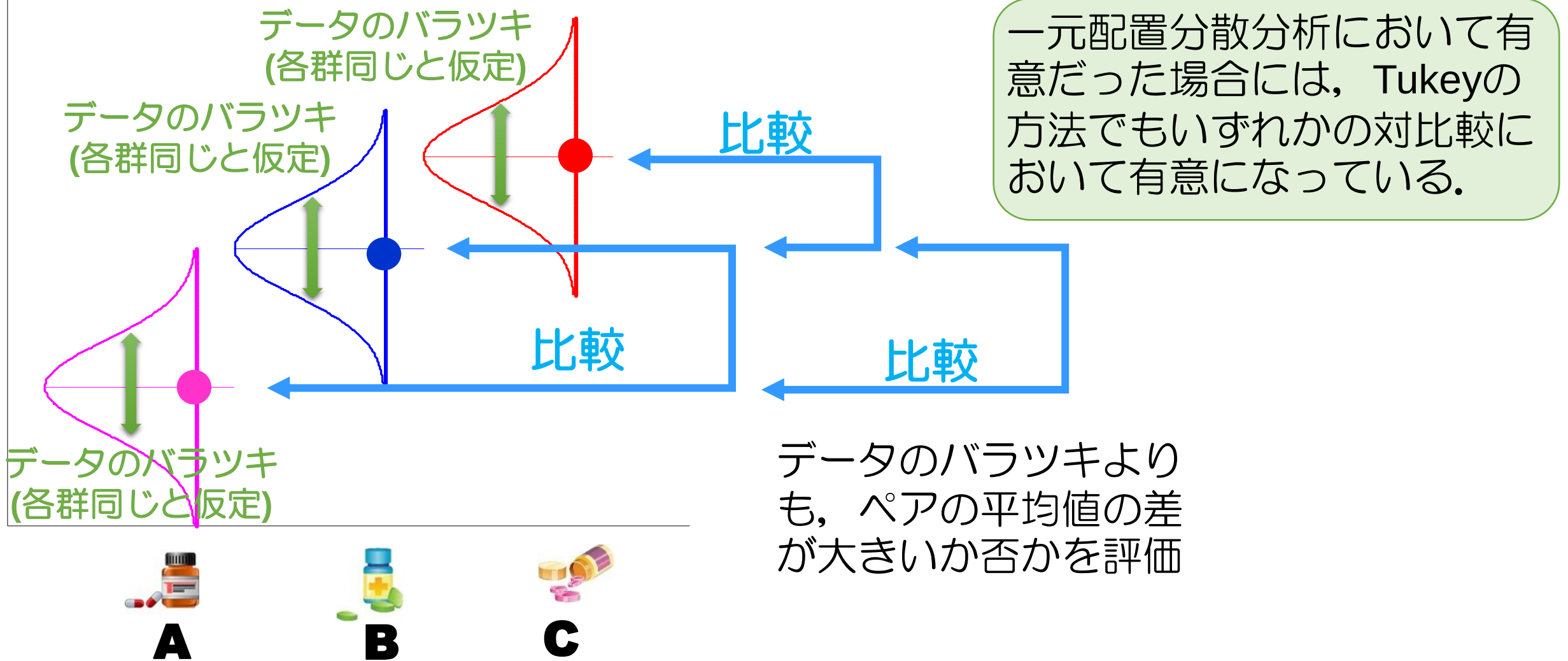
0.001 0.006 0.011 0.012 0.032 0.045



順位	p値	Bonfferoni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

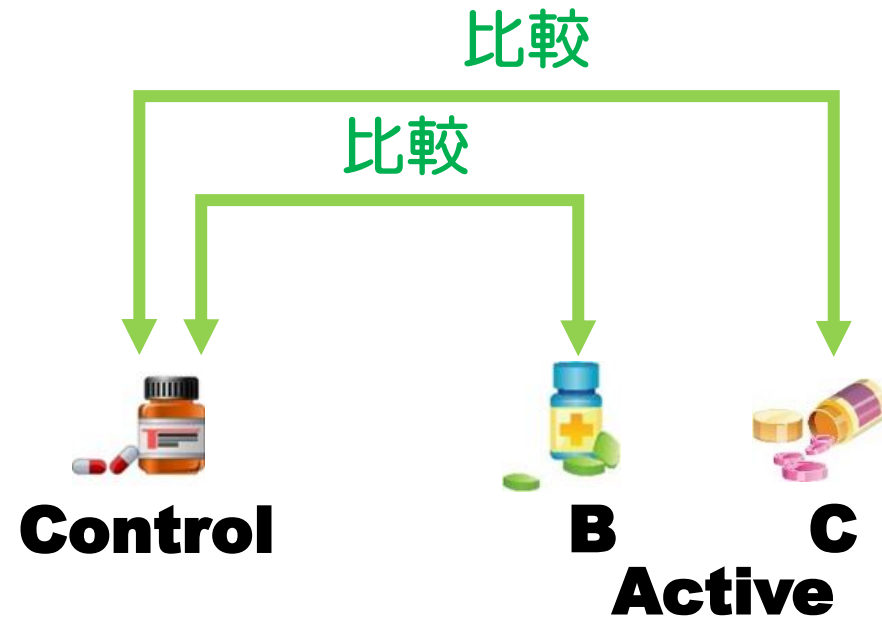
以降はすべて有意でない

正規分布に基づく方法(1)：Tukeyの方法（連続変数のみに利用できる）



Tukeyの方法のノンパラメトリック版には，Steel-Dwassの方法がある。

正規分布に基づく方法(2) : Dunnettの方法 (連続変数のみに利用できる)

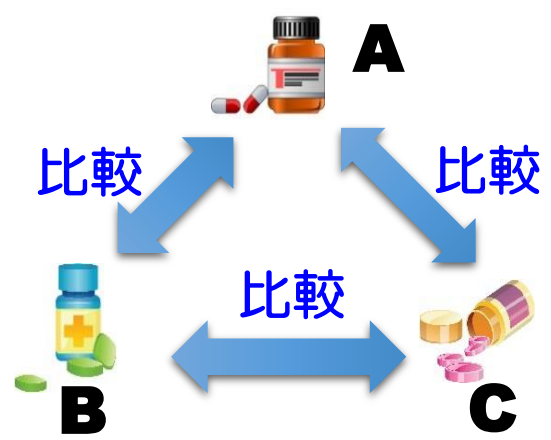


Controlとの比較のみを実施する多重比較がDunnettの方法である.

Tukeyの方法のノンパラメトリック版には, Steelの方法がある.

EZRで利用できる多重比較

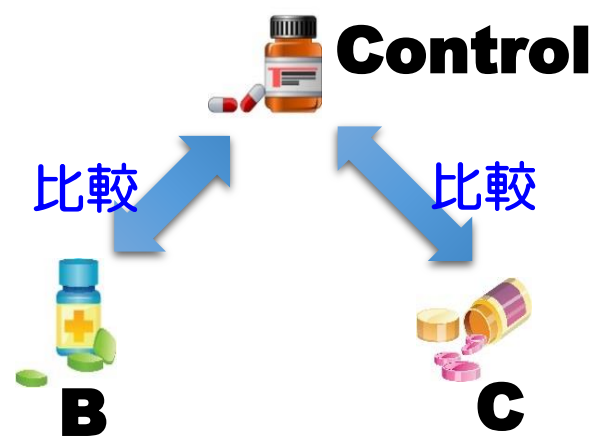
すべての群で比較
(ペアワイズ)



第1の過誤(α エラー)
を調整して多重比較
Bonferroniの多重比較
Holmの多重比較

- パラメトリック
- TukeyのHSD検定
- ノンパラメトリック
- Steel-Dwass検定

対照群と比較



- パラメトリック
- Dunnettの検定
- ノンパラメトリック
- Steel検定

Bonferroniの多重比較, Holmの多重比較は, 「すべての群で比較」, 「対照群と比較」のいずれにも適用できるが, EZRでは「すべての群で比較」のみに利用される.



EZRによる演習(4)

「1.5.1 一元配置の分散分析」

「1.5.2 3群以上でのノンパラメトリック検定：Kraskal-Wallis検定」



EZRによる演習(4)における留意点

- 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) 線形混合効果モデル

パラメトリック

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)

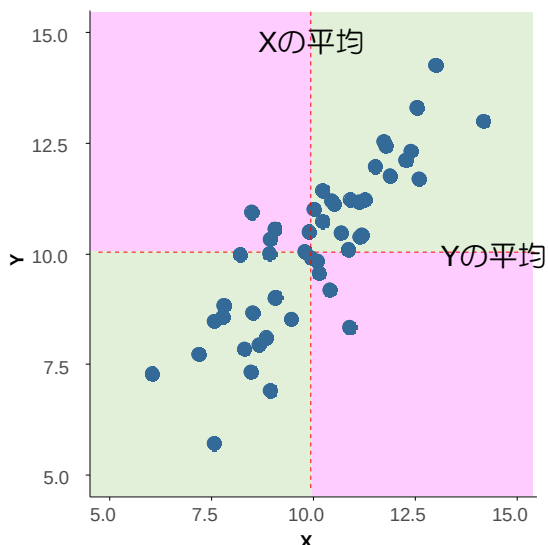
Kruskal-Wallis検定

Friedman検定

相関関係と共分散・相関係数

正の相関関係

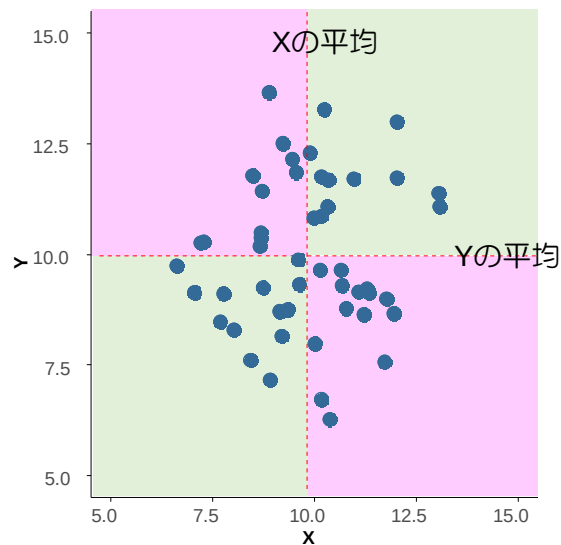
Xが**大きく**なるほど
Yが**大きく**なる



■ に多く布置する

無相関関係

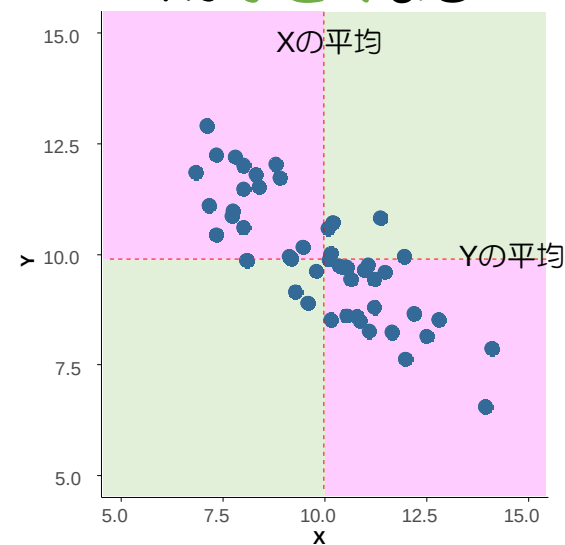
XとYには関連性
がない



同じぐらい布置する

負の相関関係

Xが**大きく**なるほど
Yが**小さく**なる



■ に多く布置する

■と■の布置の状況を表す指標 = 共分散

正值

1に近づく

0付近

0に近づく

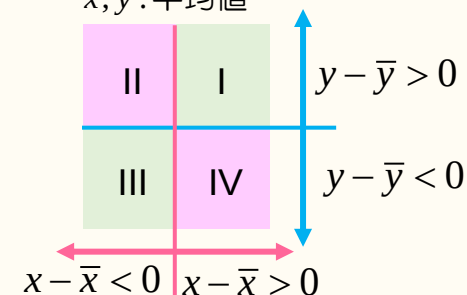
負値

-1に近づく

-1から1の範囲になるように調整 = 相関係数

各領域の説明

$\bar{x}, \bar{y}$: 平均値



Iの領域: $(x - \bar{x}) \times (y - \bar{y}) > 0$
正 正

IIの領域: $(x - \bar{x}) \times (y - \bar{y}) < 0$
負 正

IIIの領域: $(x - \bar{x}) \times (y - \bar{y}) > 0$
負 負

IVの領域: $(x - \bar{x}) \times (y - \bar{y}) < 0$
正 負

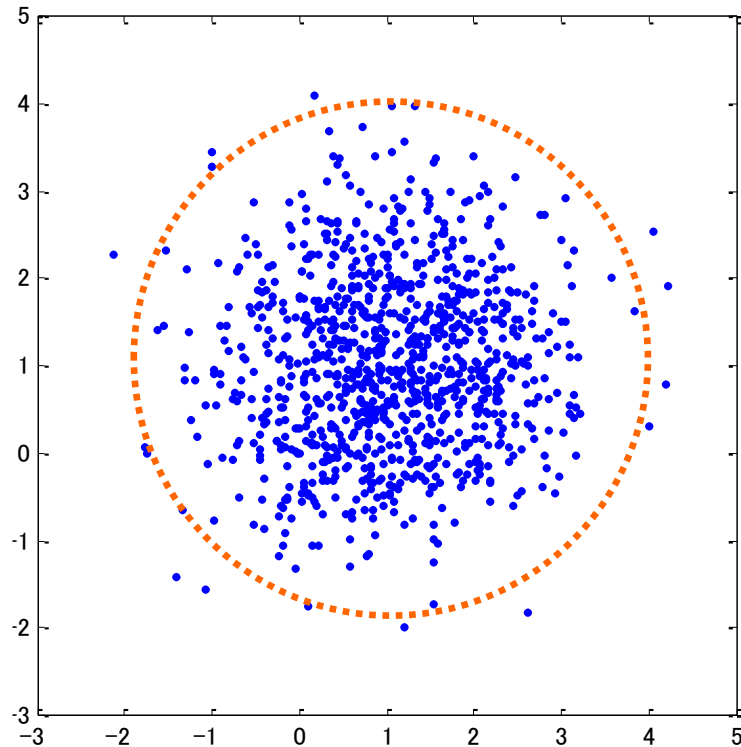
したがって、緑色の領域は正值、桃色の領域は負値になる。

無相関性の検定のp値だけを信用しない

標本の信憑性に自信がもてるときにのみ有意性検定や比較検定を行うほうがよい。とくに、無相関係数に関する検定は十分に吟味しなければならない。

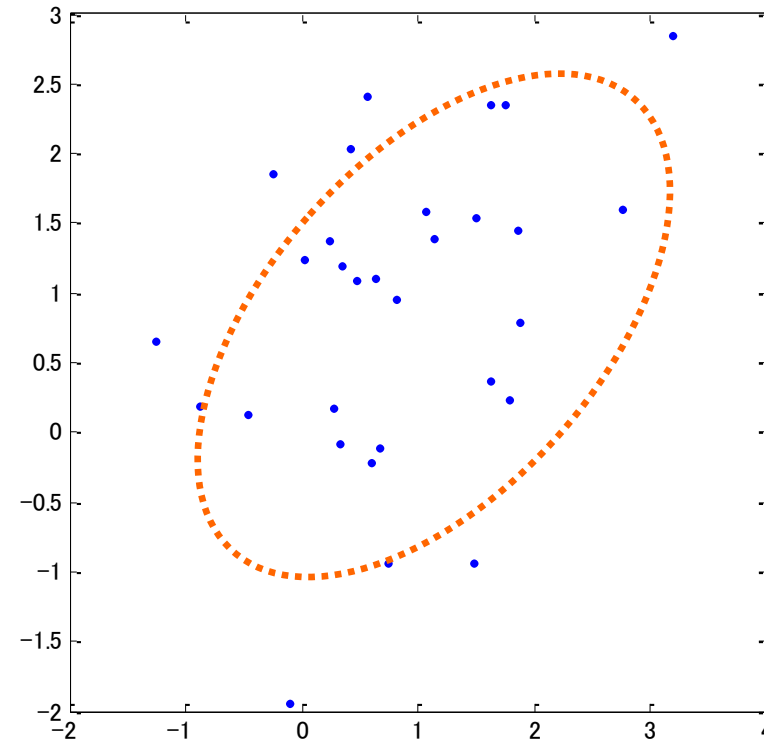
$$r = 0.05$$

有意水準0.001で有意



$$r = 0.5$$

有意水準0.05で有意

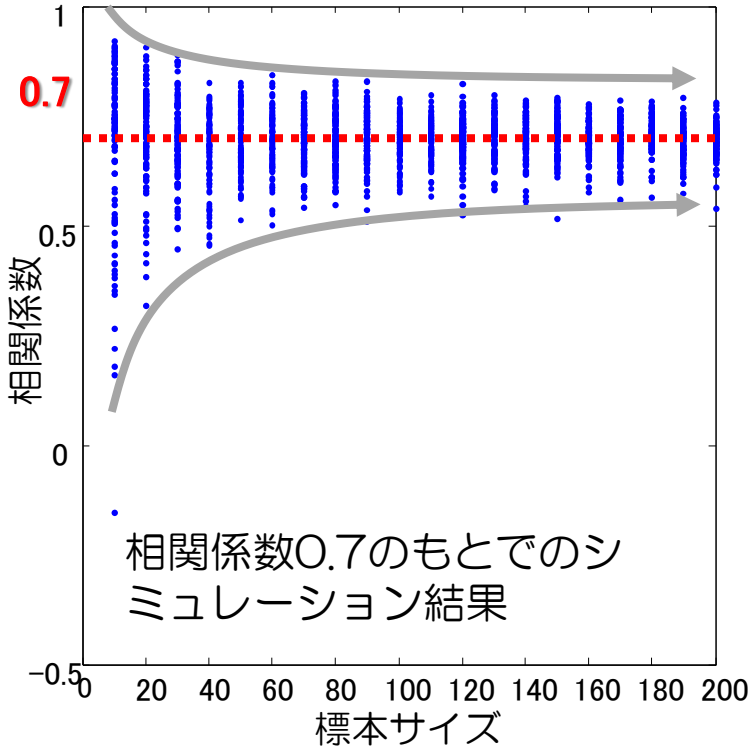


左側の散布図を見る限り、相関関係は認められない。それにもかかわらず無相関性の検定は有意である。つまり、無相関性の検定をもって、安易に相関関係があるというのは推奨されない。

相関係数に対する解釈と注意点

一般的な状況における相関係数の目安は以下の通りである。

$1.0 \geq \text{相関係数の絶対値} \geq 0.7$	かなり高い相関がある
$0.7 \geq \text{相関係数の絶対値} \geq 0.5$	高い相関がある
$0.5 \geq \text{相関係数の絶対値} \geq 0.4$	中程度の相関がある
$0.4 \geq \text{相関係数の絶対値} \geq 0.3$	ある程度の相関がある
$0.3 \geq \text{相関係数の絶対値} \geq 0.2$	弱い相関がある
$0.2 \geq \text{相関係数の絶対値} \geq 0.0$	ほとんど相関がない



一方で、小数例での相関係数にも注意が必要。
左の図は、真の相関係数が0.70のもとでシミュレーションを行った結果である。

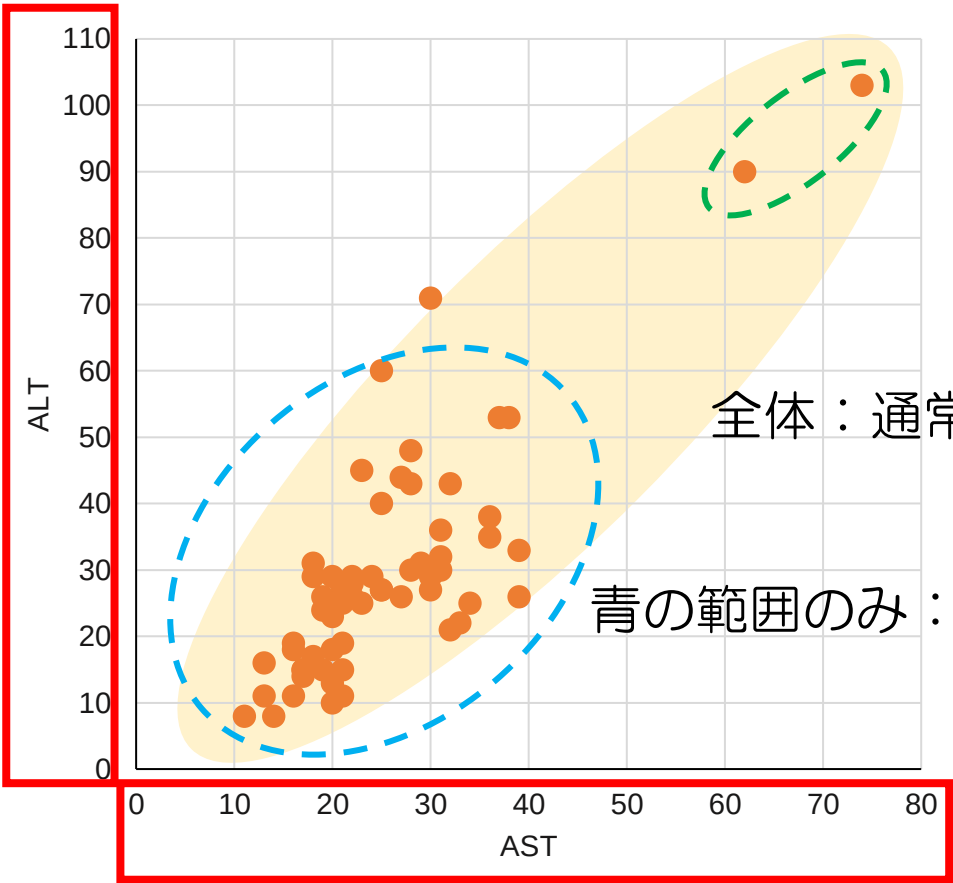
- 標本サイズが少ないときは、相関係数のバラツキが非常に大きい。
- 標本サイズが100程度でようやく安定している。

相関係数を使う場合の処方箋

- 可能な限り散布図で相関関係を確認する。
- 信頼区間を必ず併記する。
- p値で有意だからといって安易に「相関がある」とは考えない(無相関性の検定が有意であるということは母相関係数が0でないことを言っているに過ぎない)。

ノンパラメトリック相関とは

胃癌患者63名のALTとASTに関する散布図



全体：通常の相関係数=0.819

青の範囲のみ：通常の相関係数0.615

ALTが小さい観測値から順番に1,2,...と順位をつける

ASTが小さい観測値から順番に1,2,...と順位をつける

Spearmanの
順位相関係数

それぞれの変数の順位を用いて相関係数を計算したものがSpearmanの順位相関係数である(このデータにおけるSpearmanの順位相関係数 = 0.727).



EZRによる演習(5)

「1.6.1 Pearsonの相関係数」

「1.6.2 Spearmanの順位相関係数」



EZRによる演習(5)における留意点

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)

線形混合効果モデル

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)

回帰分析の動機

いま、たばこと肺がんの関係を明らかにするために、各地域のたばこの総販売個数と肺がんの罹患者数を調査した。



相関分析の結果、たばこの総販売個数と肺がんの罹患者数には**正の相関関係**があった。したがって、**たばこと肺がんには関連性がある!!**

相関分析では、たばこの総販売個数が多い地域では肺がんの罹患者が多い(肺がんが多い地域ではたばこの総販売個数が多い)という関係がわかる。

たばこの総販売個数



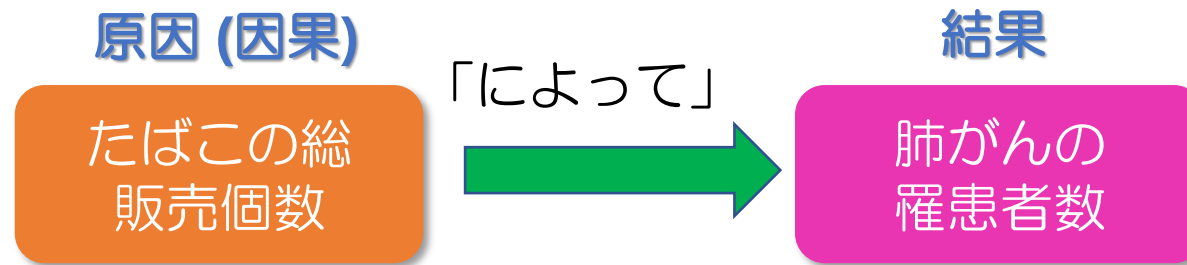
肺がんの罹患者数



このような調査では、「たばこが肺がんの原因である」ことを明らかにするために、**「たばこの総販売個数が高い(喫煙者が多い)ことで肺がんの罹患者数が高くなる」**ことを知りたいのでは？

回帰分析を用いる

仮説：たばこの総販売個数が高い(喫煙者が多い)ことで肺がんの罹患者数が多くなる



明確に原因と結果が明らかならば、「原因から結果を予測することができる」はずである。

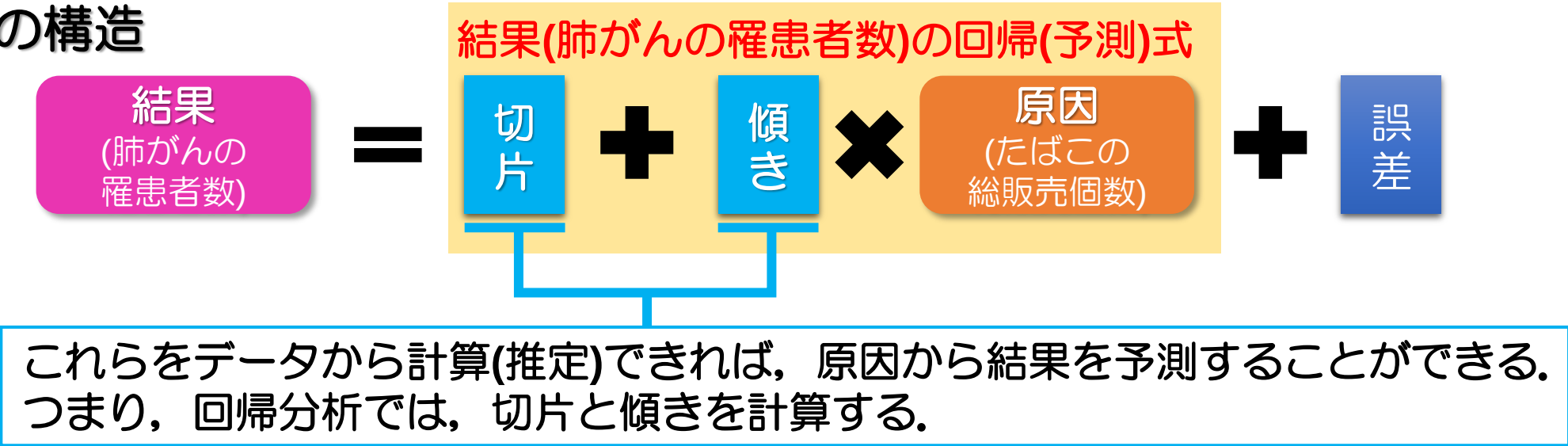
原因から結果を予測するための予測式を作る統計的なほうほうが回帰分析である。

回帰分析の目的は、

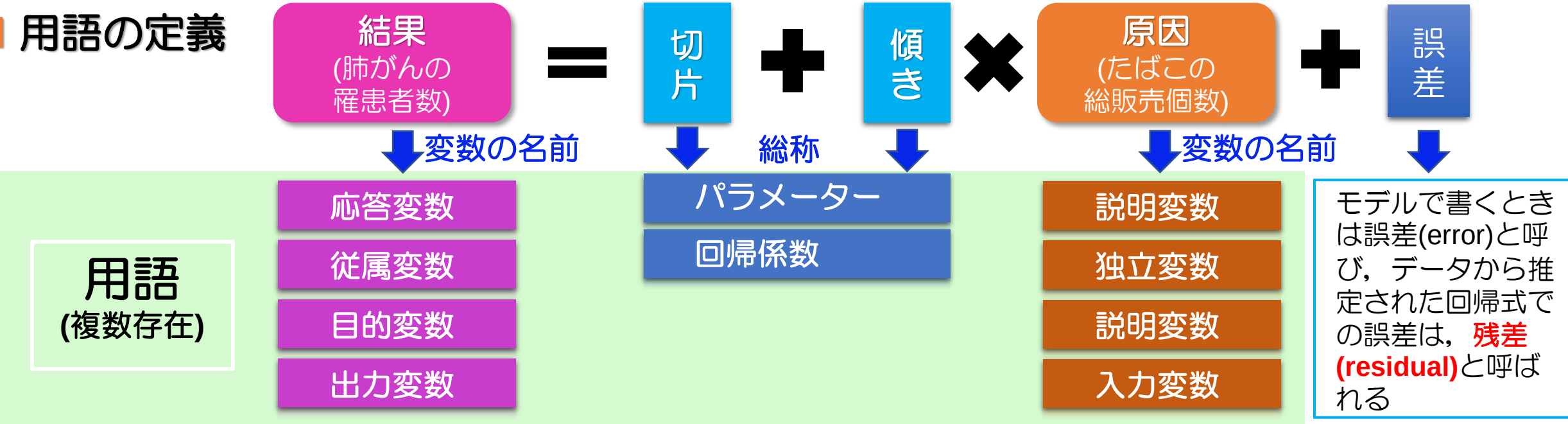
- (1) 原因から結果を予測するための予測式(回帰式／回帰モデル)を作成する [予測の観点].
- (2) 複数の原因がある場合には、複数の原因がどのように結果に影響するか(すなわち、要因構造)を明らかにする [要因構造探索の観点].

回帰式(回帰モデル)とは：単回帰分析を例に

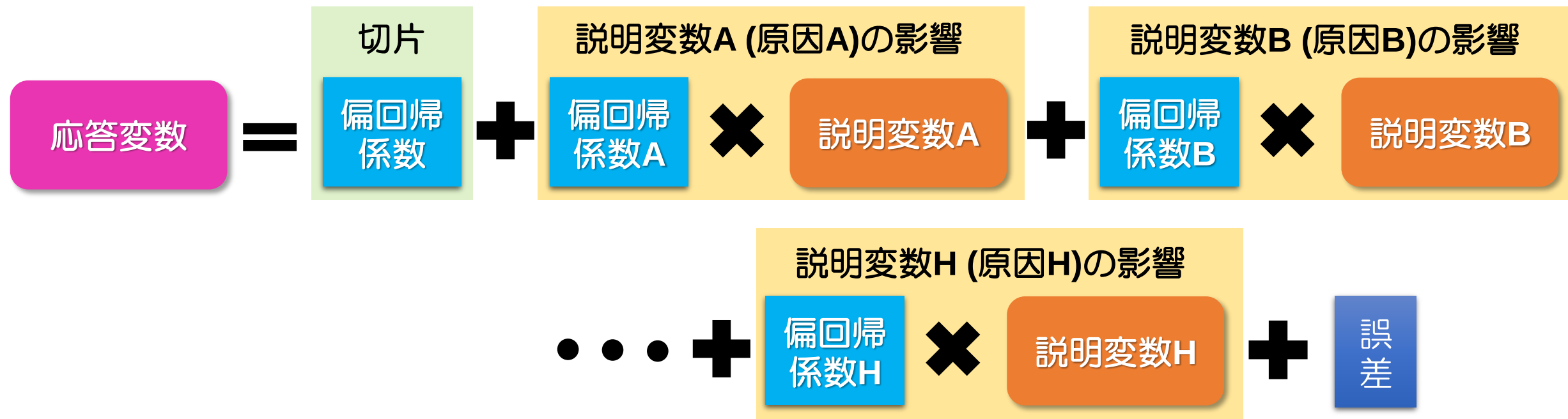
回帰式の構造



用語の定義

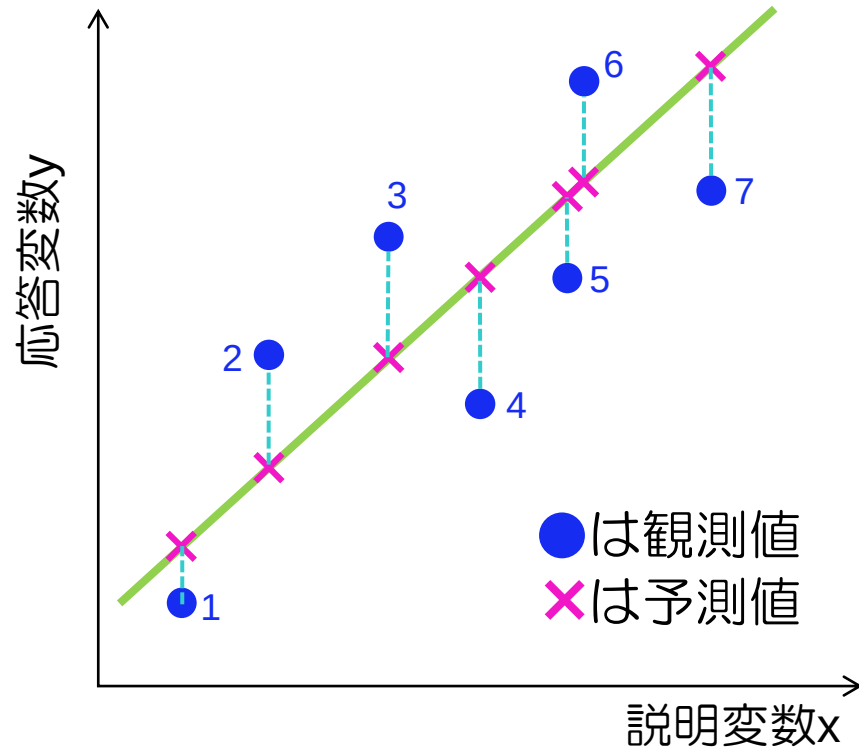


重回帰分析の場合の回帰式(回帰モデル)



- 重回帰分析では、回帰係数(回帰パラメータ)を偏回帰係数(偏回帰パラメータ)と呼ぶことがある).
- 重回帰分析の回帰式は、偏回帰係数×説明変数の総和で表される(線形結合).

(偏)回帰係数の推定



推定された回帰式の予測値と観測値の差(残差)は

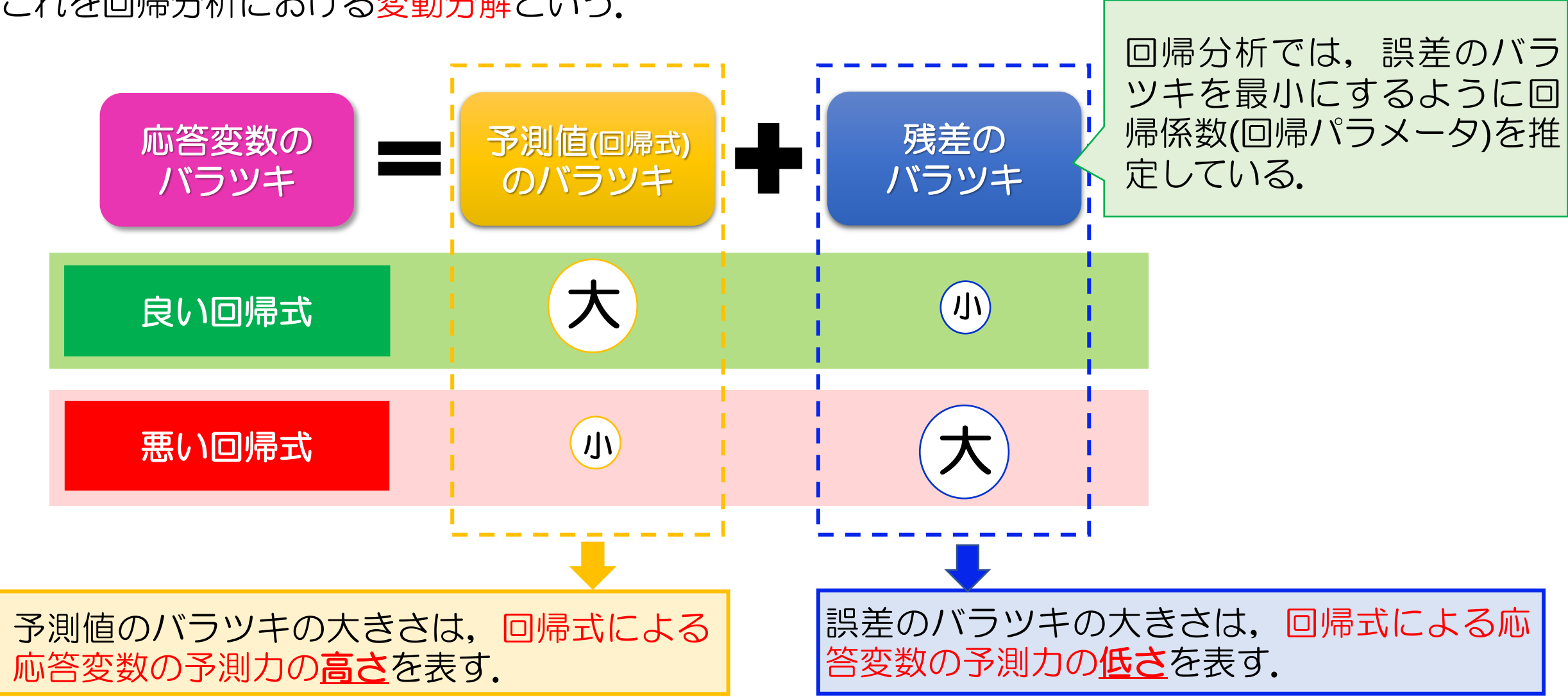
1番目の残差	=	1番目の応答変数の値	-	1番目の予測値
2番目の残差	=	2番目の応答変数の値	-	2番目の予測値
⋮		⋮		⋮
7番目の残差	=	7番目の応答変数の値	-	7番目の予測値

残差が全体的に最小にするような回帰係数を計算する？ しかし、残差の総和は0になる。

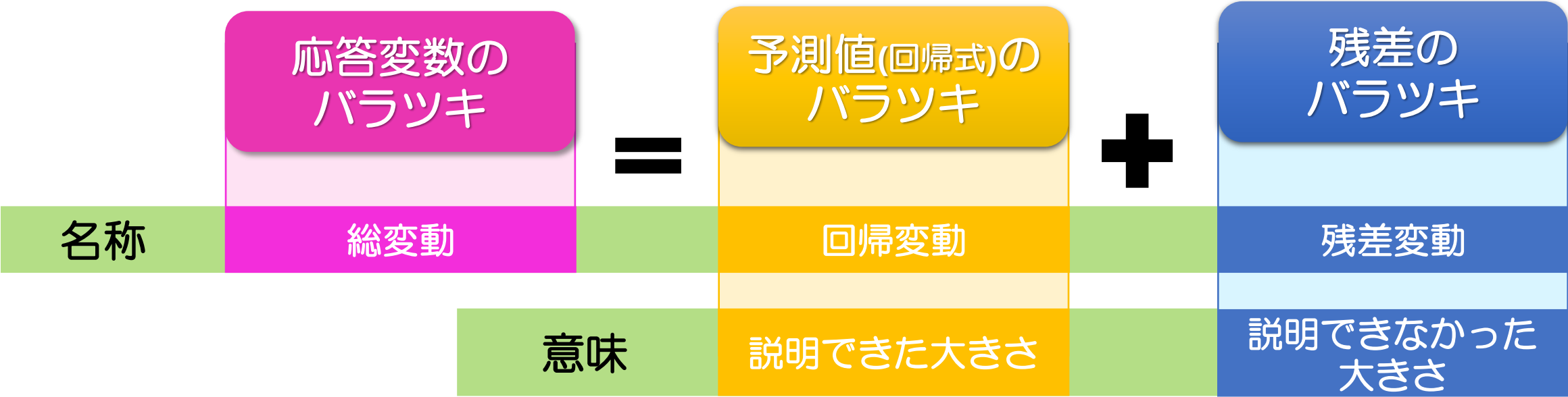
残差の2乗和 (残差のバラつき)を最小にするような回帰係数を推定する。これを**最小2乗法**という。

回帰モデルの変動分解

回帰分析では、応答変数のバラツキは、予測値のバラツキと残差のバラツキに分けることができる。
これを回帰分析における**変動分解**という。



寄与率



寄与率とは、 応答変数のバラツキの何割(何パーセント)を予測値のバラツキが示しているかを表したものである。 いかえれば、 **回帰式が応答変数をどの程度説明できたか**を表した値である。 したがって、 寄与率は0.0 (0%) ~ 1.0 (100%)の範囲をとる。

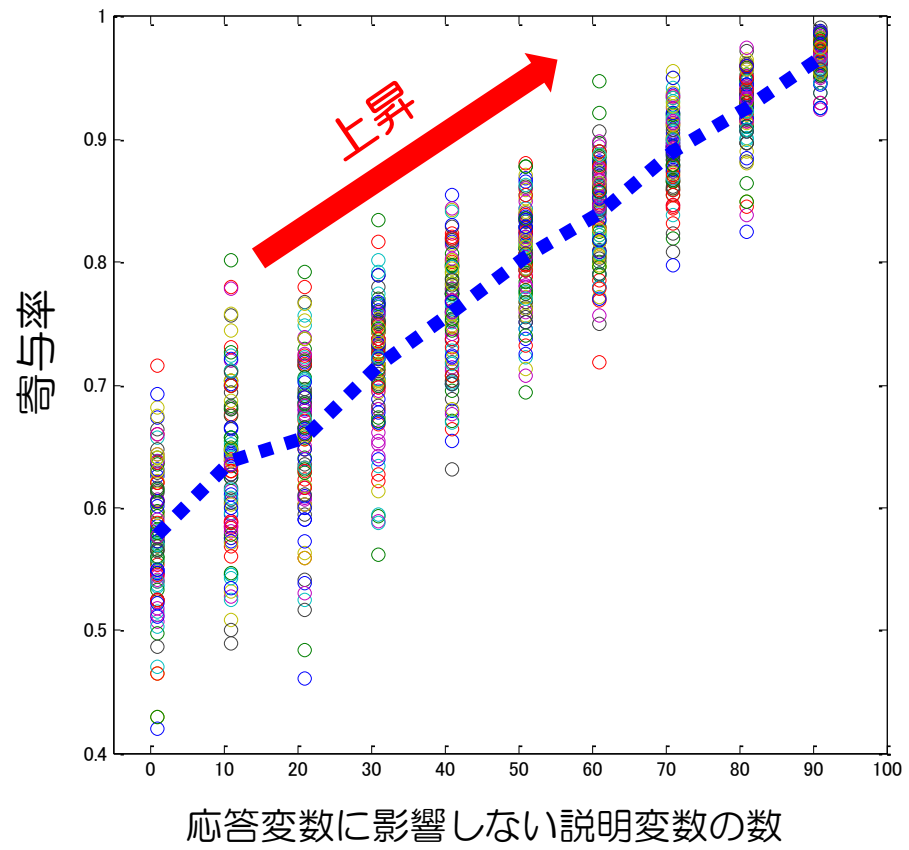
$$\text{寄与率} = \frac{\text{回帰変動}}{\text{総変動}} = 1 - \frac{\text{残差変動}}{\text{総変動}}$$

寄与率は、 **決定係数**とも呼ばれる。

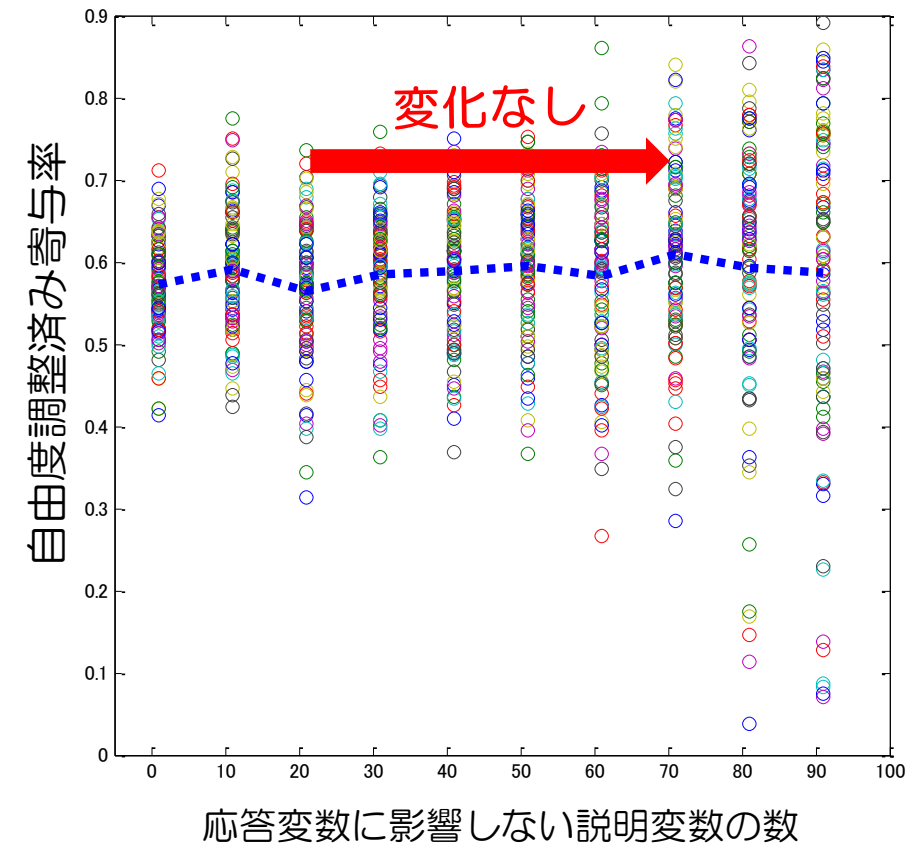
寄与率と自由度調整済み寄与率

説明変数の数が増えると、回帰変動が過大評価され、残差変動が過小評価される。この問題を調整した寄与率は、**自由度調整済み寄与率**と呼ばれる。

応答変数に影響しない説明変数の数を増加したときの寄与率及び自由度調整済み寄与率の推移(各説明変数の数に対して100回のシミュレートを実施している。青色の点線は平均値の推移を表している)



(a) 寄与率

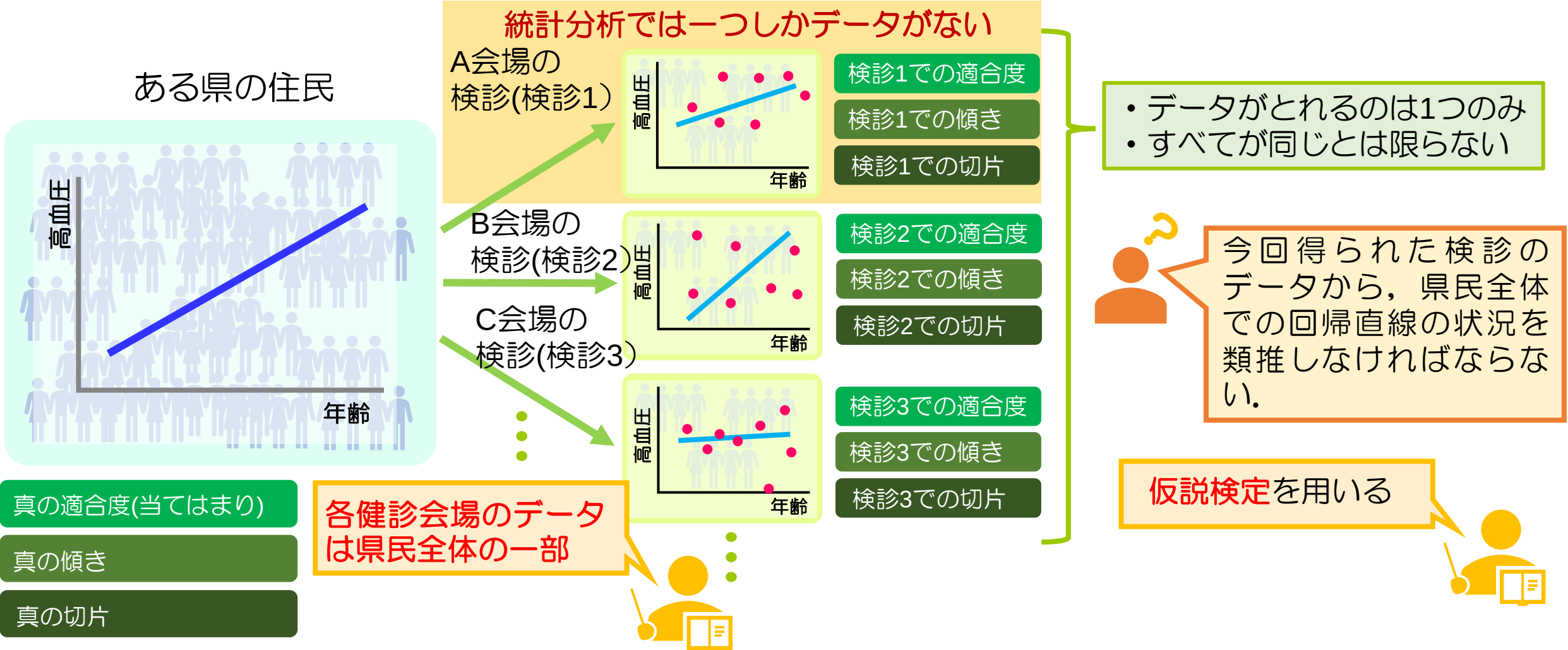


(b) 自由度調整済み寄与率

寄与率は、応答変数に影響しない説明変数を増やせば増やすほど増加するので、適合度評価にはならない。そのため、自由度調整済み寄与率を用いなければならない。

回帰分析における諸種の検定

仮想例：いま、ある県において実施された特定検診のデータを用いて、年齢と高血圧の関係が調査された。



仮説検定とは何か：傾きに対する有意性検定を例に考える

帰無仮説 H_0

真の回帰式における傾きは0である

二つのシナリオ

逆仮説

仮説が背反に
なっている

対立仮説 H_1

真の回帰式における傾きは0でない

シナリオのパターンは1種類

帰無仮説 H_0 のシナリオが正しいとしたもとで、今回のデータがどれぐらいの確率で得られるかを考える。

帰無仮説 H_0 のシナリオが正しいと仮定したときに、今回のデータから推定されるような傾き得られる確率

0.035 (3.5%)

これを**p値(有意確率)**という

シナリオのパターンは無限大

仮説検定とは、帰無仮説 H_0 が正しいと仮定したもとで、今回のデータが得られる確率を計算したもとで評価する方法である。

p値がどれぐらい小さいと帰無仮説 H_0 が誤っているといえるかを決定するカットオフ値が必要である。このカットオフ値を**有意水準 α** といい、一般的には**0.05 (5%)**とされている。

仮説検定による判定の方法

p値が有意水準 α (一般的には $\alpha=0.05$)未満

p値 (有意確率)

例示: p値=0.035

p値が有意水準 α (一般的には $\alpha=0.05$)以上

例示の場合

帰無仮説 H_0 が間違っている (棄却)

有意である (棄却される)

結論

(真の)傾きは0でない

有意であるということは、その説明変数は、応答変数に影響している。

帰無仮説 H_0 が間違っているとは言えない
(受容)

有意でない (棄却されない)

結論

(真の)傾きは0でないという根拠なし

仮説検定において0であるとは言えないので注意

F検定(回帰の分散分析)

帰無仮説 H_0 : (真の)回帰式に意味はない
対立仮説 H_1 : (真の)回帰式に意味はある

帰無仮説 H_0 が正しいと仮定したとき、**今回のデータにおける適合度はどれぐらいの確率(p値)で得られるのか**について計算する

p値

p値 $\geq$ 有意水準 α

比較

p値 $<$ 有意水準 α

帰無仮説 H_0 が受容

回帰式(モデル)に意味
あるとはいえない

帰無仮説 H_0 が棄却

回帰式(モデル)に意味
ある

傾きに対する有意性検定

帰無仮説 H_0 : (真の)傾きは0である
対立仮説 H_1 : (真の)傾きは0でない

帰無仮説 H_0 が正しいと仮定したとき、**今回のデータにおける傾きはどれぐらいの確率(p値)で得られるのか**について計算する

p値

p値 $\geq$ 有意水準 α

比較

p値 $<$ 有意水準 α

帰無仮説 H_0 が受容

傾きが0でないとはいえない

帰無仮説 H_0 が棄却

傾きが0でない

切片に対する有意性検定

帰無仮説 H_0 : (真の)切片は0である
対立仮説 H_1 : (真の)切片は0でない

帰無仮説 H_0 が正しいと仮定したとき、**今回のデータにおける切片はどれぐらいの確率(p値)で得られるのか**について計算する

p値

p値 $\geq$ 有意水準 α

比較

p値 $<$ 有意水準 α

帰無仮説 H_0 が受容

切片が0でないとはいえない

帰無仮説 H_0 が棄却

切片が0でない

「比較」の部分のロジックな説明

帰無仮説 H_0 が正しいと仮定したときに、今回の結果(適合度, 傾き, 切片)が得られる確率をp値(有意確率)という。このとき、どれぐらいの確率未満であれば、帰無仮説 H_0 が誤っている(逆仮説である対立仮説 H_1 が正しい)と判定するのかの基準が有意水準 α (通常は $\alpha=0.05$)である。

多重共線性

いま、奈良県における2016年と2017年の国別の外国人来訪者数から2018年の外国人来訪者数を予測できないかを検討している。

説明変数：2016年と2017年の国別外国人来訪者数， 応答変数：2018年の国別外国人来訪者数

Model.1：2016年と2017年の両方を加えた場合の回帰式

$$(\text{2018年の来訪者数}) = -995.870 - 1.082 \times (\text{2016年の来訪者数}) + 1.954 \times (\text{2017年の来訪者数})$$

2016年の回帰係数が負値なので、2016年の来訪者数が増えるほど、2018年の来訪者が減る。

Model.2：2016年のみの回帰式

$$(\text{2018年の来訪者数}) = -18643.979 + 2.401 \times (\text{2016年の来訪者数})$$

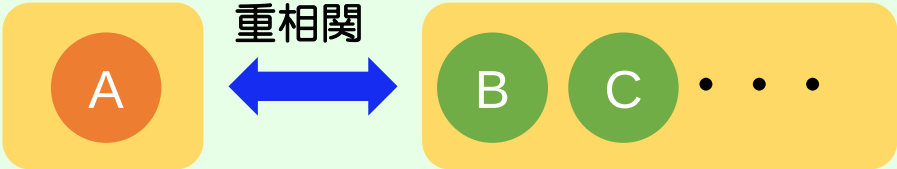
2016年の回帰係数が正值なので、2016年の来訪者数が増えるほど、2018年の来訪者が増える。

2016年＋2017年の回帰式における2016年の解釈と2016年のみの回帰式で解釈が逆になる。これは、2016年と2017年の来訪者の相関係数が0.994であり、非常に高いことに由来する。このように、同じような数値をとる説明変数を重回帰分析に用いた場合、相互に干渉してしまい、誤まった解釈をもたらすことがある。このことを**多重共線性**という。

多重共線性の診断：VIF (Variance Inflation Factor)

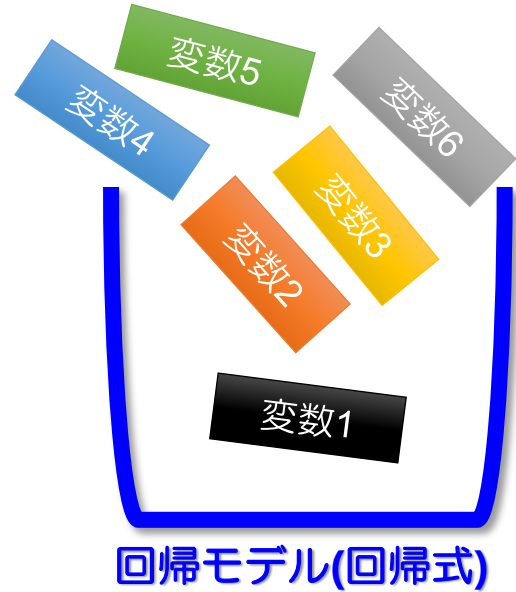
独立変数間に強い相関関係がある場合、変数間が干渉することで、実際の影響とは異なる(偏)回帰係数を推定したり、あてはまりが悪くなることを多重共線性という。それを評価する指標がVIFである。

VIFは、説明変数に用いられた「任意の変数A」と「変数A以外の変数」との重相関係数に基づいて計算できる。

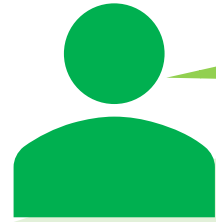
重相関係数 r =  を用いて $VIF = \frac{1}{1 - r^2}$ で計算できる

VIFは10を超えると、その変数とその他の変数のあいだに多重共線性が疑われると判断される。

重回帰分析における注意点：変数が多ければよいというものではない

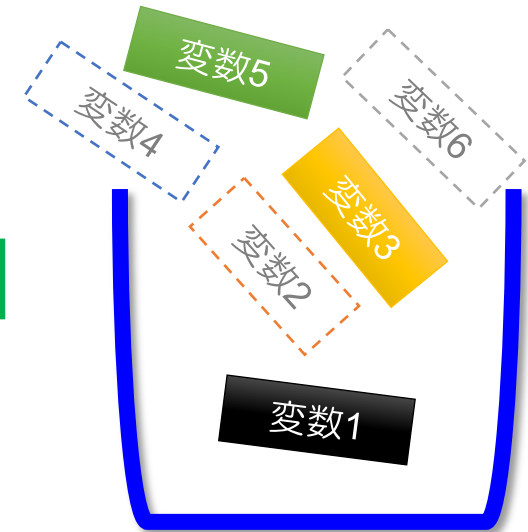


重回帰分析では、たくさんの説明変数を入れるほど情報がたくさんになるので、良い統計モデルになるということ？



必ずしもそうではない。説明変数が多いほど不必要な変数は単なるノイズでしかない。

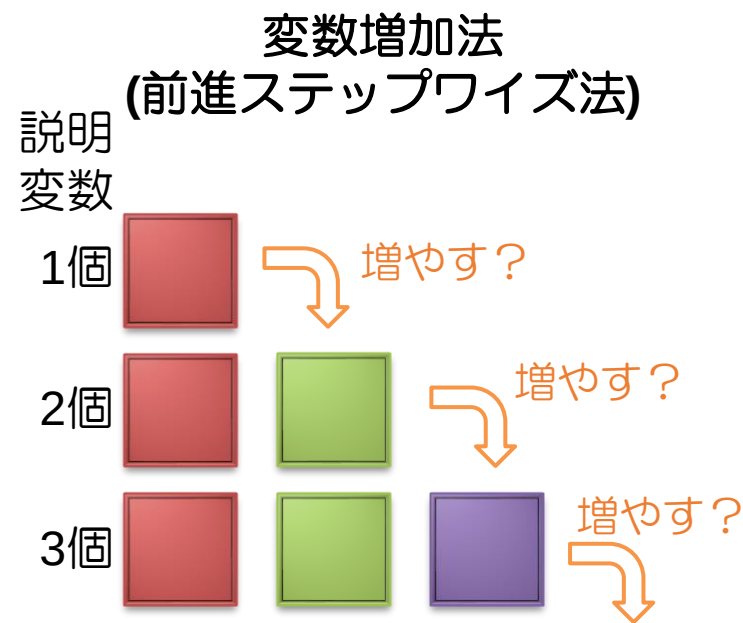
また、**多重共線性**の問題などがある。そのため、必要なものだけで回帰式をつくることが推奨されることが多い。そのための方法が**変数選択**である。



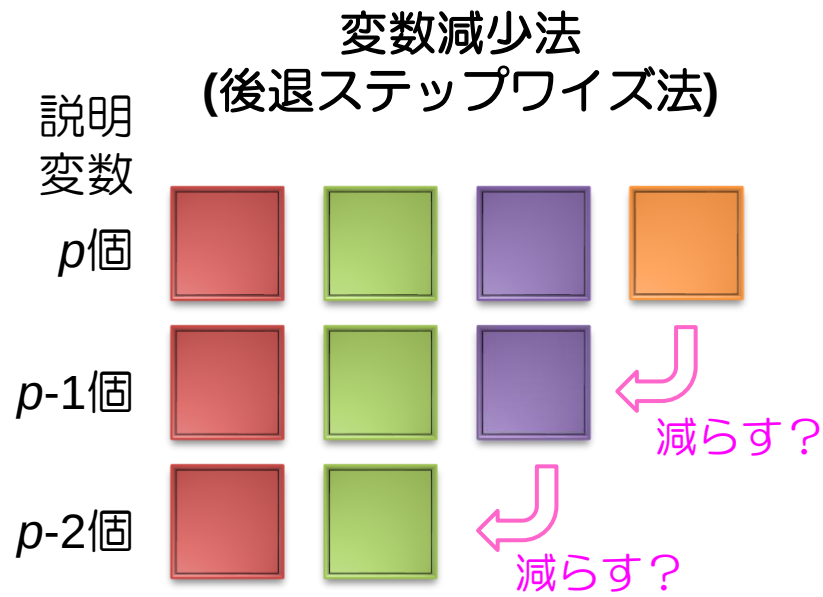
変数選択では、以下を決定したうえで実施しなければならない：

- ・どのようなアルゴリズムで変数を選択するのか (変数選択の方法)
- ・どのような基準で回帰モデル(回帰式)を評価するのか (評価基準)

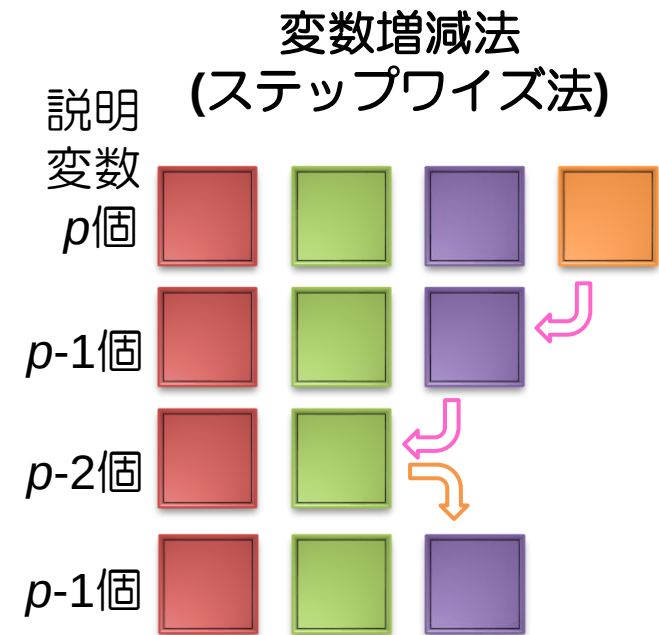
■ 変数選択のアルゴリズム



説明変数が1個の場合からスタートして、変数を追加したほうが良ければ増やし、そうでなければ変数の追加をしない。



全ての説明変数からスタートして、変数を減らしても影響がなければ減らし、そうでなければ変数の削除をしない。



変数減少法からスタートするが、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。

■ 変数選択の基準

- 検定を用いる方法 (最近では推奨されない)
- 情報量規準を用いる方法 (最近はこちらが一般的)
 - 赤池の情報量規準 (AIC; Akaike's Information Criteria)
 - Bayes流情報量規準 (BIC; Bayesian Information Criteria)

情報量規準は、**AICとBICのどちらをつかってかまわない。ただし、AICのほうがBICよりも多くの変数を選択する(BICのほうがドラスティックに削除する)。**

変数選択の留意点

(1) 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、本当は必要な説明変数を取り込まれる前に変数選択が終了する場合がある。

(2) 重回帰分析に用いることができる説明変数の数

重回帰分析解析に用いることができる説明変数の数は、標本サイズ(個体数)の1/10程度とされている(統計学的な意味はないが、慣例的に言われている)。

(3) 多数の説明変数がある場合の留意点

多数の調査項目(説明変数)が存在する場合には、全ての説明変数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、説明変数毎に単変量解析(1個の説明変数による回帰モデルを推定する)を実施し、その(偏)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値や回帰パラメータを用いる。このとき、有意水準 α は0.10あるいは0.20であっても許容される。なお、慣例的に標本サイズの1/10が重回帰分析で用いることができる説明変数の目安として考えられている。

(4) 欠測が多い説明変数(調査項目)には注意が必要である

重回帰分析では、説明変数のなかで1個でも欠測があれば、その個体を削除しなければならない。そのため、欠測が多い説明変数をモデルに含めると、多くの個体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、このような説明変数は、予め変数選択の候補から除外することが望ましい。

EZRによる変数選択

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
応答変数

説明変数
説明変数

☐ 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

変数減少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。



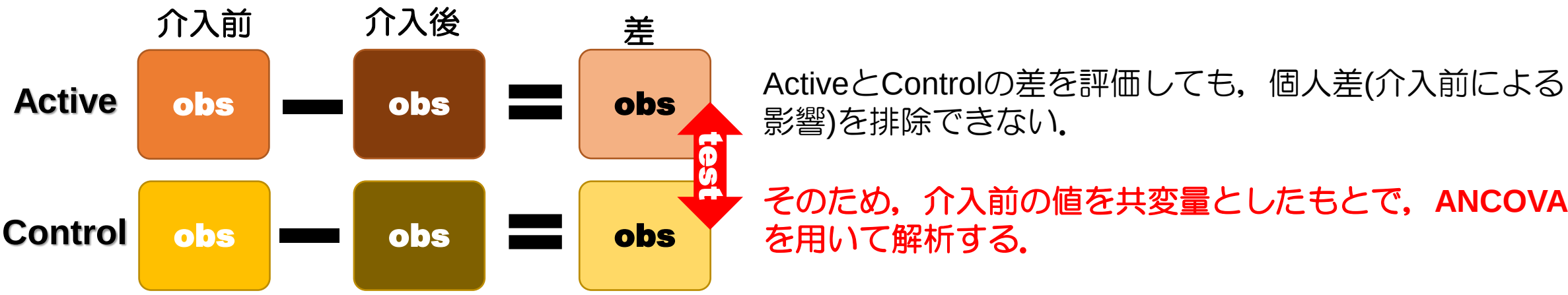
EZRによる演習(6)

「1.7.2 重回帰分析」

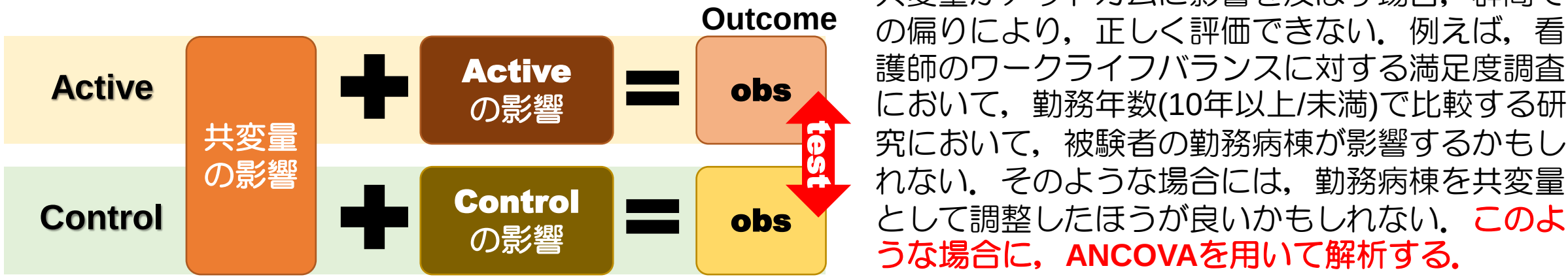
共分散分析

共分散分析 (ANCOVA; ANalysis of COVAriance)の動機

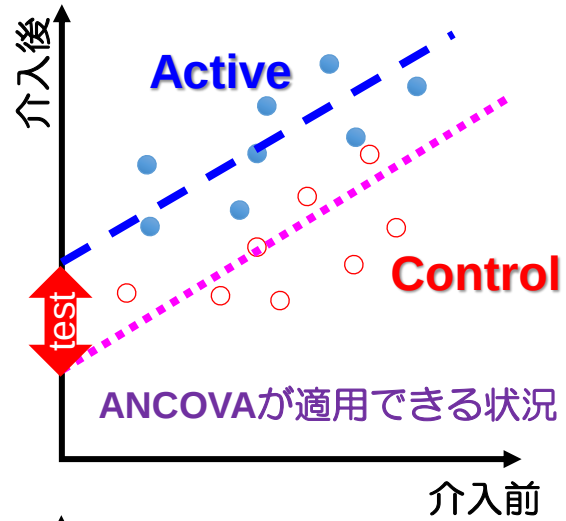
□ 介入前後での変化量 or 変化率を比較する



□ 介入前の共変量の影響を排除して解析する



- 共分散分析(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]

→ 有意な場合には治療前後での比較は不可能 (左下の図)

具体的には、

「応答変数」= 「群」+ 「共変量」+ 「群」× 「共変量」
の重回帰分析を行い、「群」× 「共変量」での交互作用のp値を評価する。

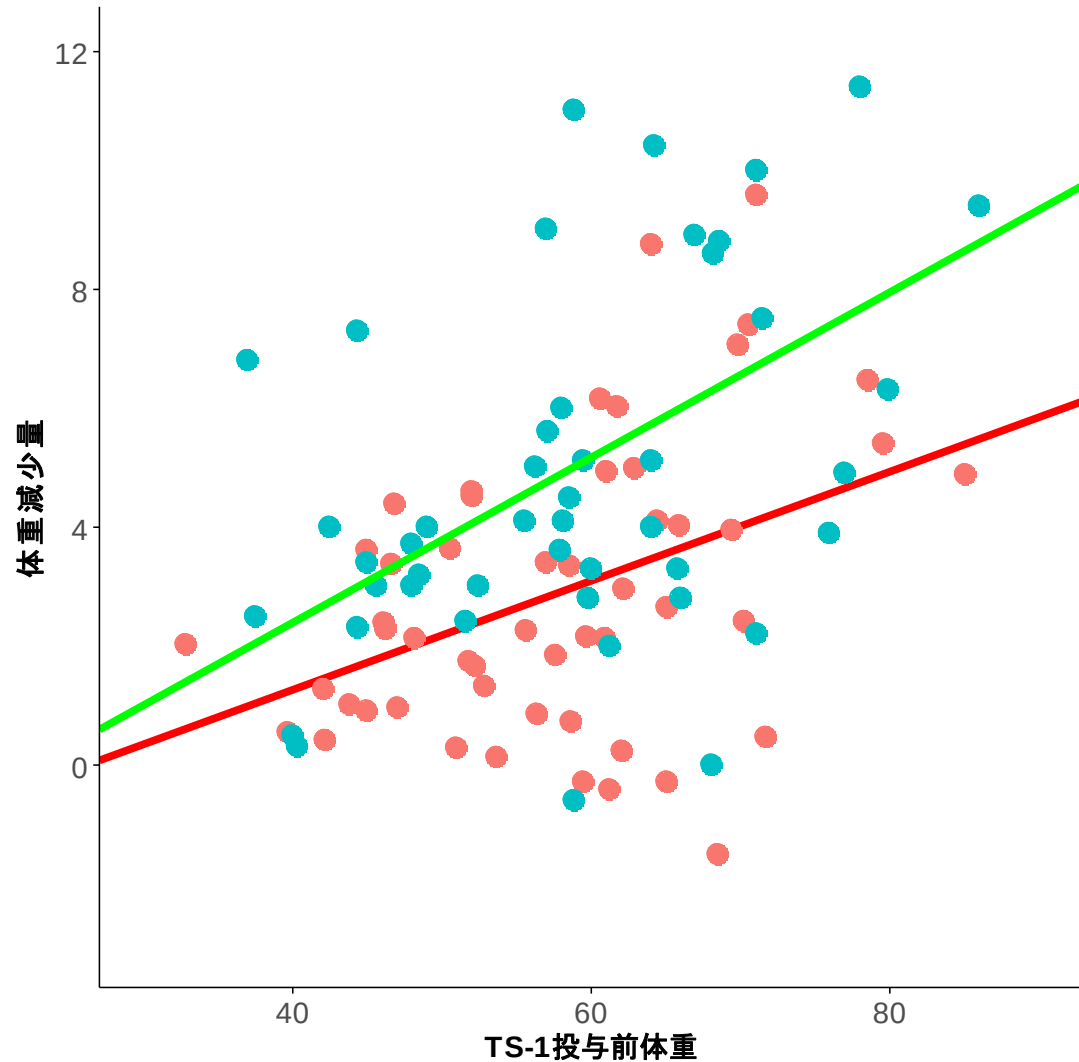
[STEP.2-1] 切片を比較する(介入前値の違いによる介入後値の変化が同じであるか検討する)[Test.2]

→ 有意な場合には変化に違いがあると解釈 (左上の図)

具体的には、

「応答変数」= 「群」+ 「共変量」
の重回帰分析を行い、「群」でのp値を評価する。

適用例



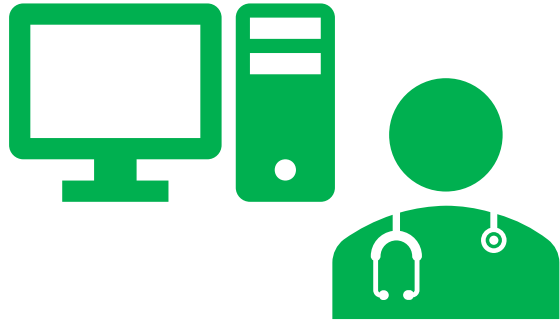
胃癌患者に対する術後のTS-1補助化学療法による体重減少の抑制を意図した成分栄養剤服用の有効性を検討している。

左の散布図において、●が成分栄養剤を服用してもらった患者であり、●が服用しなかった患者である。

共分散分析では、TS-1投与前の体重を共変量としたもとで、術後1年での体重減少を応答変数に用いている。

共変量×群の交互作用の評価ではp値が0.288であることから、交互作用があるとは言えない(もし、交互作用の評価で有意であれば、共分散分析が使えないとして評価を終了する)。

その後の共変量と群を用いた回帰分析における群に対するp値は <0.001 であった。すなわち、成分栄養剤を服用することで、TS-1による体重減少を抑制することができた。



EZRによる演習(7)

「1.8 共分散分析」

質的データにおける解析編(2章)

EZRによる質的データにおける解析一覧

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)
順序変数に対する多変量解析(順序ロジスティック回帰)
多項ロジスティック回帰

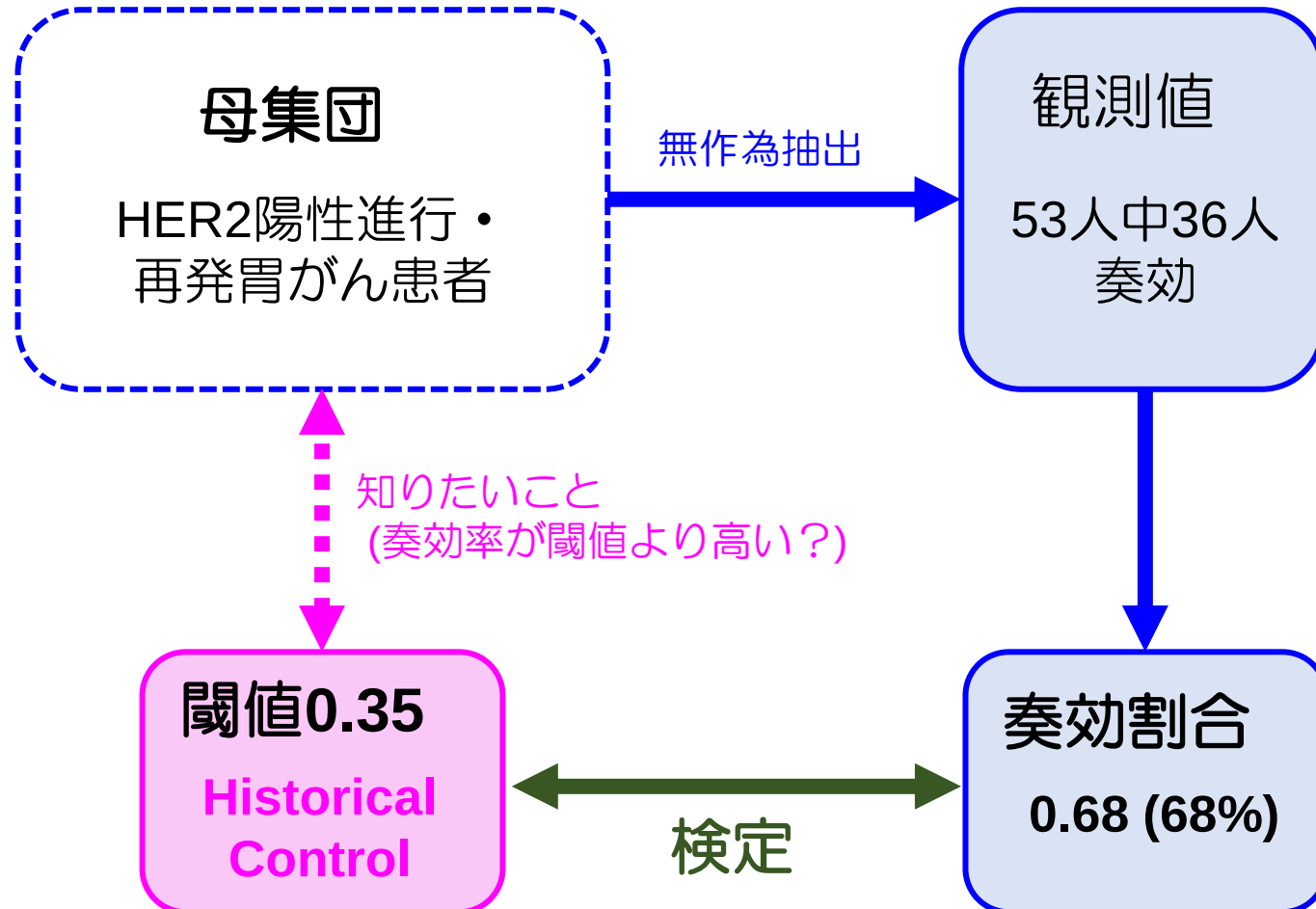
■ **Direct**: 標本サイズ(サンプル数), 成功回数 (イベント数)等を直接入力する.

■ **DataSet**: データ集合を読み込んで解析する.

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
順序変数に対する多変量解析(順序ロジスティック回帰)	DataSet
多項ロジスティック回帰	DataSet

母比率の検定

HER2陽性の進行・再発胃がん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)



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・再発胃がん患者に対して、非常に**有望**であることが証明された。

母比率の検定における仮説

帰無仮説

帰無仮説 H_0 「母比率は、ヒストリカル・コントロールでの比率と同じである
(新規レジメンでの奏効率は閾値奏効率35%と同じである)」

対立仮説

両側対立仮説 H_{1a} 「母比率は、ヒストリカル・コントロールでの比率と異なる
(新規レジメンでの奏効率は閾値奏効率35%と異なる)」

片側対立仮説 H_{1b} 「母比率は、ヒストリカル・コントロールでの比率を上回る
(新規レジメンでの奏効率は閾値奏効率35%よりも大きい)」

事例における対立仮説

片側対立仮説 H_{1c} 「母比率は、ヒストリカル・コントロールでの比率を下回る
(新規レジメンでの奏効率は閾値奏効率35%よりも小さい)」

母比率の検定は、臨床試験においては、単アームでの試験(第II相試験)の評価に用いられる。この場合、両側対立仮説にしなければならないことはなく、片側対立仮説のもとで評価されることも少なくない。



EZRによる演習(8)

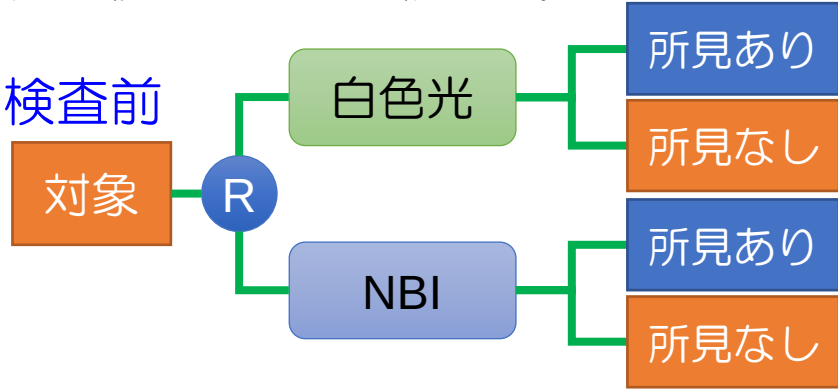
「**2.1** 2値変数に対する**1**標本データの解析：母比率に対する推測」

カイ2乗検定およびFisherの正確検定の動機



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

■ E Z R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。

カイ2乗検定とFisherの正確検定の使い分け

下表はある疾患に対する2種類の治療法の有効率を評価したときのクロス集計表である。治療法によって有効/無効に違いがあるか。

	有効	無効	計
治療A	204 (51.1%)	195 (48.9%)	399
治療B	158 (39.4%)	243 (60.6%)	401
計	362 (45.2%)	438 (54.2%)	800

帰無仮説：治療法によって有効/無効に違いが**ない**。
対立仮説：治療法によって有効/無効に違いが**ある**。



クロス集計表の解析において用いられる検定は、

- ・カイ2乗検定
- ・Fisherの正確検定 (Fisherの直接確率)

である。なお、今回の事例のように有効率を比較する場合には、母比率の差の検定というものがある。他方、母比率の差の検定は、正確にカイ2乗検定に一致する。

カイ2乗検定

仮説検定：帰無仮説の位置からどれだけ離れているか(どれだけ違っているか)を確率(p値)で表す

	有効	無効	計
治療A	204	195	399
治療B	158	243	401
計	362	438	800

帰無仮説でのクロス集計表を作らないといけない。
研究の結果

	有効	無効
治療A	204	195
治療B	158	243

期待度数(帰無仮説での位置)

有効	無効
$\frac{399 \times 362}{800} = 180.5$	$\frac{399 \times 438}{800} = 218.5$
$\frac{401 \times 362}{800} = 181.5$	$\frac{401 \times 438}{800} = 219.5$



各セルに対して、度数の差の2乗値を計算して、その和をとったものが帰無仮説の位置からの離れ度合いである(カイ2乗値という).

その結果、**p値は0.001**なので、有意でだった。つまり、治療Aと治療Bの有効率に違いが認められた。ちなみに、Fisherの正確検定でのp値は0.001であり結果に大差がない。

Fisherの正確検定

	有効	無効	計
治療A	12 (54.5%)	10 (45.5%)	22
治療B	3 (20.0%)	12 (80.0%)	15
計	15 (40.5%)	22 (59.5%)	37

	有効	無効	計
治療A	12	10	22
治療B	3	12	15
計	15	22	37

2×2クロス集計表では、周辺度数(青色の部分)を固定してしまうと、1個のセル(赤色)が決まれば、残りのセルは全て決まる。

こうなる確率 = 0.0000

$$a \times d - b \times c = -330$$

	有効	無効
治療A	0	22
治療B	15	0

こうなる確率 = 0.0000

$$a \times d - b \times c = -293$$

	有効	無効
治療A	1	21
治療B	14	1

こうなる確率 = 0.0000

$$a \times d - b \times c = -256$$

	有効	無効
治療A	2	20
治療B	13	2

こうなる確率 = 0.0001

$$a \times d - b \times c = -219$$

	有効	無効
治療A	3	19
治療B	12	3

こうなる確率 = 0.0011

$$a \times d - b \times c = -182$$

	有効	無効
治療A	4	18
治療B	11	4

こうなる確率 = 0.0084

$$a \times d - b \times c = -145$$

	有効	無効
治療A	5	17
治療B	10	5

こうなる確率 = 0.0399

$$a \times d - b \times c = -108$$

	有効	無効
治療A	6	16
治療B	9	6

こうなる確率 = 0.1172

$$a \times d - b \times c = -71$$

	有効	無効
治療A	7	15
治療B	8	7

こうなる確率 = 0.2197

$$a \times d - b \times c = -34$$

	有効	無効
治療A	8	14
治療B	7	8

こうなる確率 = 0.2659

$$a \times d - b \times c = 3$$

	有効	無効
治療A	9	13
治療B	6	9

こうなる確率 = 0.2074

$$a \times d - b \times c = 40$$

	有効	無効
治療A	10	12
治療B	5	10

こうなる確率 = 0.1028

$$a \times d - b \times c = 77$$

	有効	無効
治療A	11	11
治療B	4	11

こうなる確率 = 0.0314

$$a \times d - b \times c = 114$$

	有効	無効
治療A	12	10
治療B	3	12

こうなる確率 = 0.0056

$$a \times d - b \times c = 151$$

	有効	無効
治療A	13	9
治療B	2	13

こうなる確率 = 0.0005

$$a \times d - b \times c = 188$$

	有効	無効
治療A	14	8
治療B	1	14

こうなる確率 = 0.000

$$a \times d - b \times c = 225$$

	有効	無効
治療A	15	7
治療B	0	15

研究で得られた状況よりも極端な状況(「こうなる確率」が小さい状況)のクロス集計表になる確率の総和を計算する。その結果、p値は

$$0.0000 + 0.0000 + 0.0000 + 0.0001 + 0.0011 + 0.0084 + 0.0314 + 0.0056 + 0.0005 + 0.0000 = 0.0471$$

であるので、有意水準 $\alpha=0.05$ のもとで有意である。したがって、治療Aと治療Bの有効性に差異が認められた。

カイ2乗検定とFisherの正確検定の使い分け：まとめ

	治療A	治療B	計
治療A	12 (54.5%)	10 (45.5%)	22
治療B	3 (20.0%)	12 (80.0%)	15
計	15 (40.5%)	22 (59.5%)	37

カイ2乗検定

p値 = 0.0783

有意差が認められない

Fisherの正確検定

p値 = 0.0471

有意差が認められる

カイ2乗検定では、p値を近似で計算する。そのため、

- (1) 標本サイズが少ない場合 (50未満)
- (2) どこかのセルの度数が小さい場合 (5未満)

には、近似精度が悪くなる。そのため、Fisherの正確検定を用いるほうが良い。一方で、標本サイズが大きくなると、Fisherの正確検定は計算できなくなる(パソコンがフリーズする)ので、カイ2乗検定を推奨する。なお、標本サイズが増えると、カイ2乗検定のp値とFisherの正確検定のp値は一致する。



EZRによる演習(9)

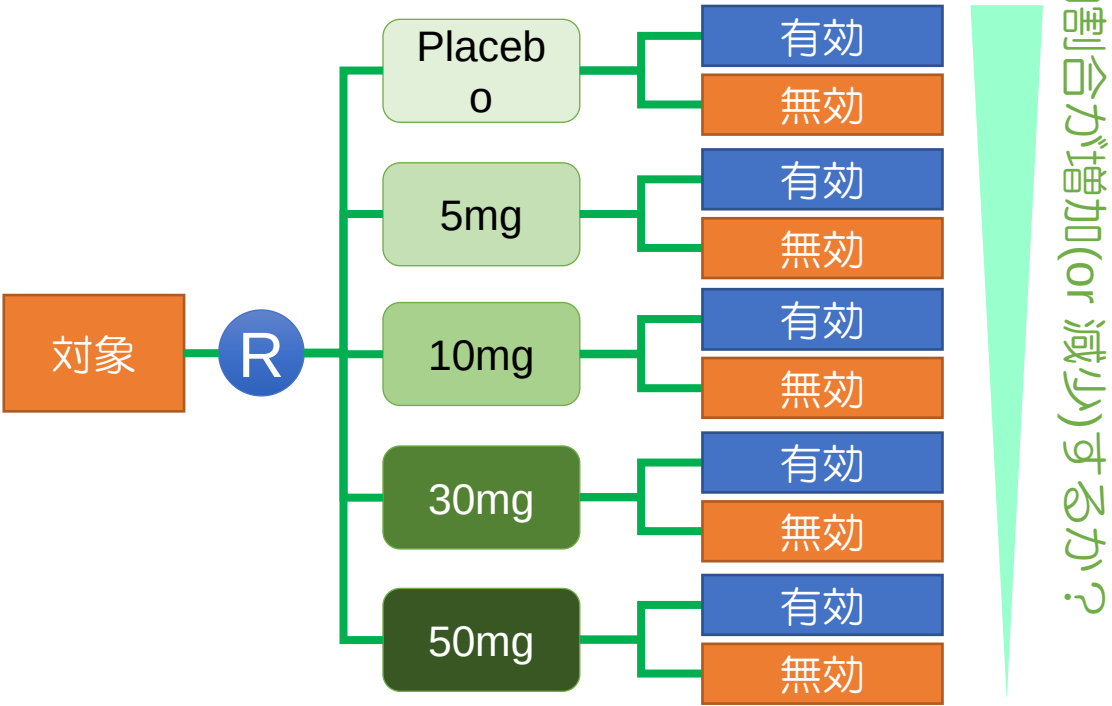
「2.2.6 EZRによるクロス集計表及び検定の実行」

Cochran-Armitage検定の動機



Clinical Question

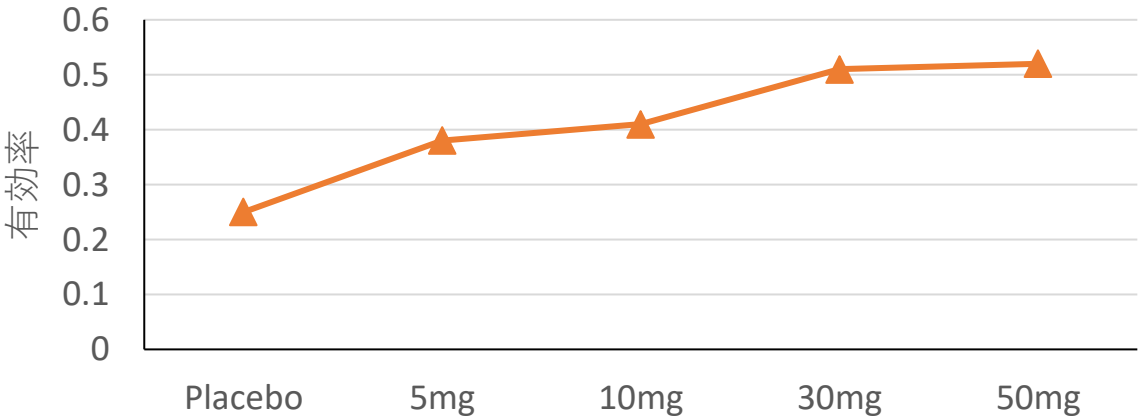
ある薬剤の用量反応関係を調査するために、Placebo, 5mg, 10mg, 30mg, 50mgの投与群とそのときの効果の有無を調べた。



統計的方法

検定：Cochran-Armitage検定
要約：有効割合

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



■ E Z R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)。

RS.10：質的データにおける様々な検定法について

傾向変化を評価するための検定：Cochran-Armitage検定

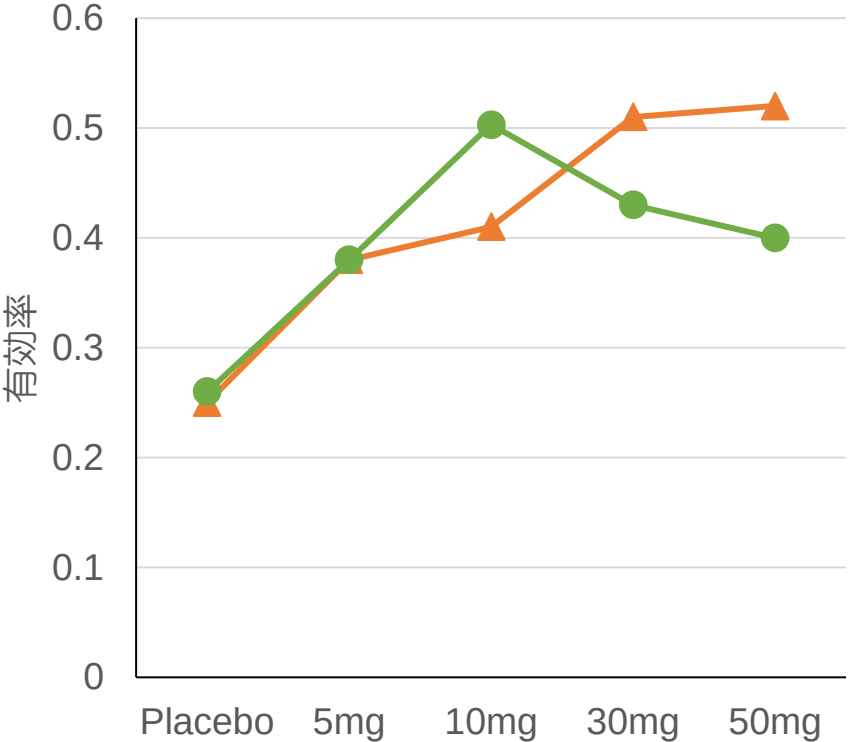
2種類の薬剤の投与量と有効性を評価したときのクロス集計表を下表に示す。

薬剤A

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

薬剤B

	Placebo	5mg	10mg	30mg	50mg
有効	82	81	220	268	233
無効	232	132	214	355	350



薬剤Aの場合には投与量が増加するにつれて有効性が増加しているものの、薬剤Bでは減少している。



このような研究を計画する理由としては、投与量による有効性の増加傾向 (あるいは減少傾向) を評価することにある。

Cochran-Armitage検定

薬剤A **p値<0.001** 投与量によって有効率の上昇傾向が認められた

薬剤B **p値=0.180** 投与量によって有効率の上昇傾向が認められなかった



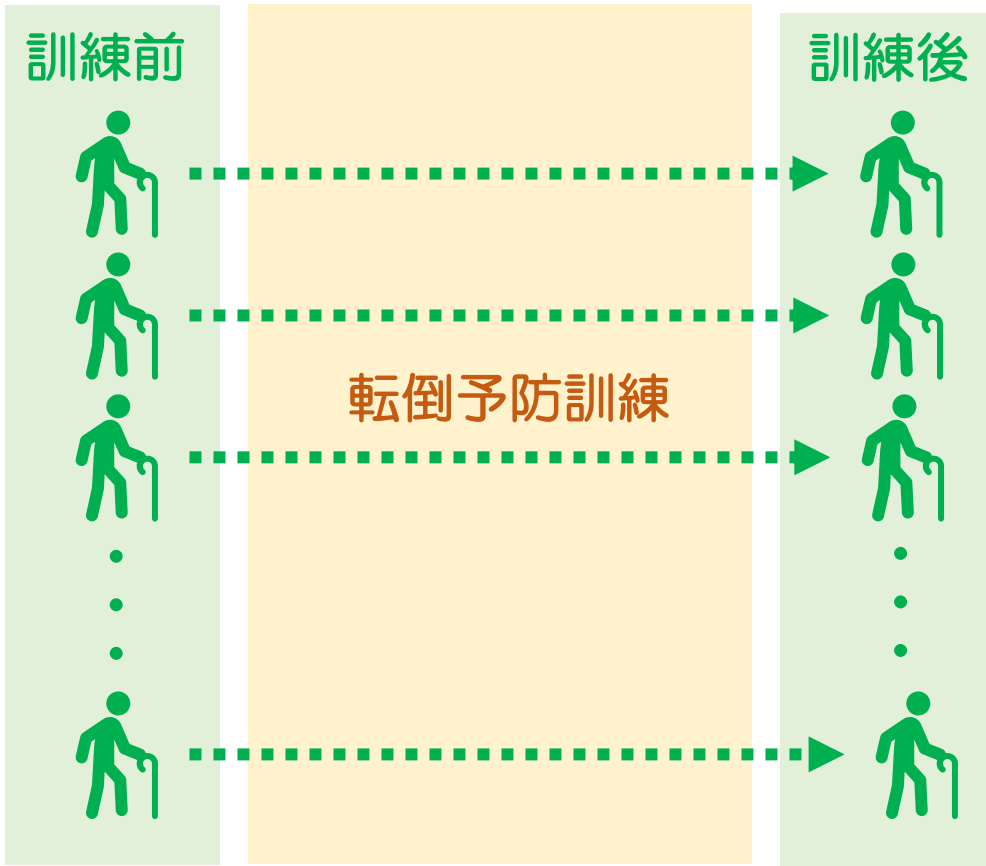
EZRによる演習(10)

「2.3 傾向変化の検定：Cochran-Armitage検定」

質的データにも「対応がある場合」が存在します



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。ここでの目標は、転倒予防訓練後に運動機能の低下が改善していることを確認することにある(柳川他, 2011)。



このデータでは、被験者のそれぞれにおいて、訓練前の運動機能の低下の有無、訓練後の運動機能の低下の有無がとられている。

このように、個々の被験者において2個のデータがとられている場合を対応のあるデータという。また、今回の場合には、以下の4パターンが存在する。

- 訓練前に運動機能低下あり，訓練後に運動機能低下あり
- 訓練前に運動機能低下あり，訓練後に運動機能低下なし
- 訓練前に運動機能低下なし，訓練後に運動機能低下あり
- 訓練前に運動機能低下なし，訓練後に運動機能低下なし

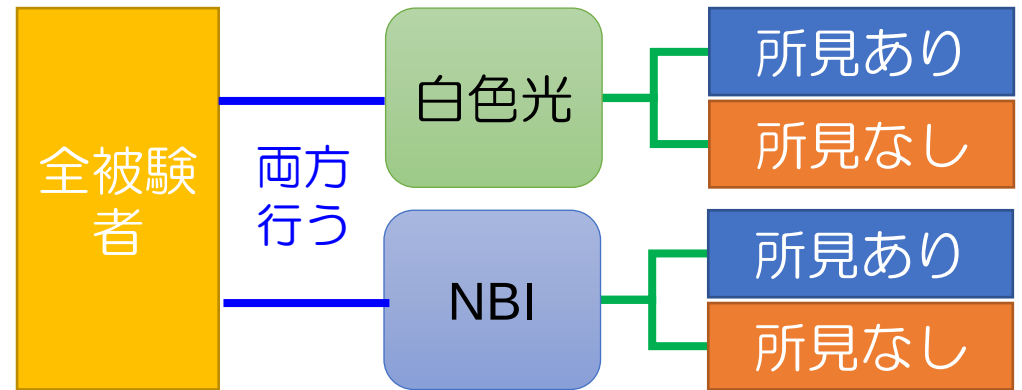
上記のパターンをクロス集計表で表したものを対応のあるクロス集計表という。

対応のあるクロス集計表の動機



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)

1. 「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。
2. 「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。

対応のあるクロス集計表

		訓練後		合計
		低下あり	低下なし	
訓練前	低下あり	40	60	100
	低下なし	30	70	100
合計		70	13	200

• 訓練前に運動機能低下あり，訓練後に運動機能低下あり
 $40/200 = 0.2$ (20%)

■ 訓練前に運動機能低下あり，訓練後に運動機能低下なし
 $60/200 = 0.3$ (30%)

■ 訓練前に運動機能低下なし，訓練後に運動機能低下あり
 $30/200 = 0.15$ (15%)

• 訓練前に運動機能低下なし，訓練後に運動機能低下なし
 $70/200 = 0.35$ (35%)



評価の方法 = McNemar検定

McNemar検定では，■の割合(30%)と■の割合(15%)が比較される。

その結果，p値は0.002であった。したがって，歩行訓練前後で運動機能に変化が認められた。また，歩行機能低下の患者が低下なしになった割合のほうが高いことから，歩行訓練の効果が認められた。

ちなみに，誤ってカイ2乗検定を用いた場合のp値が0.182なので，検定法を誤ることで，本来は有効であった歩行訓練を有効でないと誤認してしまうので注意。

「違いを見る」と「一貫性を見ることの違い」

「一貫性」を見るということ

検査Aと検査Bがあったとする。検査Aは精度の高い検査方法であるが費用がかかり、検査Bは新しい検査方法で簡便に行えるとする。この2つの検査を同じ被験者に行った場合、次のような対応のあるクロス集計表が構成される。

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

この例の場合には、(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

評価の方法 = カッパ係数

「違い」を見るということ

手術による不安感に関する調査を実施したとする。この調査では、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った。この場合には、次のような対応のあるクロス集計表が作成される。

		説明後	
		不安あり	不安なし
説明前	不安あり	(a)	(b)
	不安なし	(c)	(d)

この例の場合には、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。そのため、(b)のほうが(c)よりも高くなることが評価される。

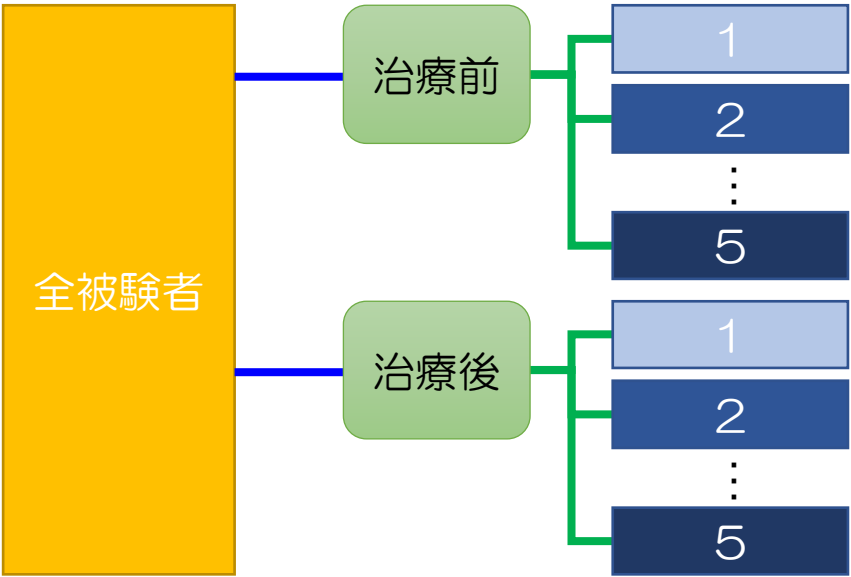
評価の方法 = McNemar検定

補足：McNemar検定が用いられるもう一つの状況



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)

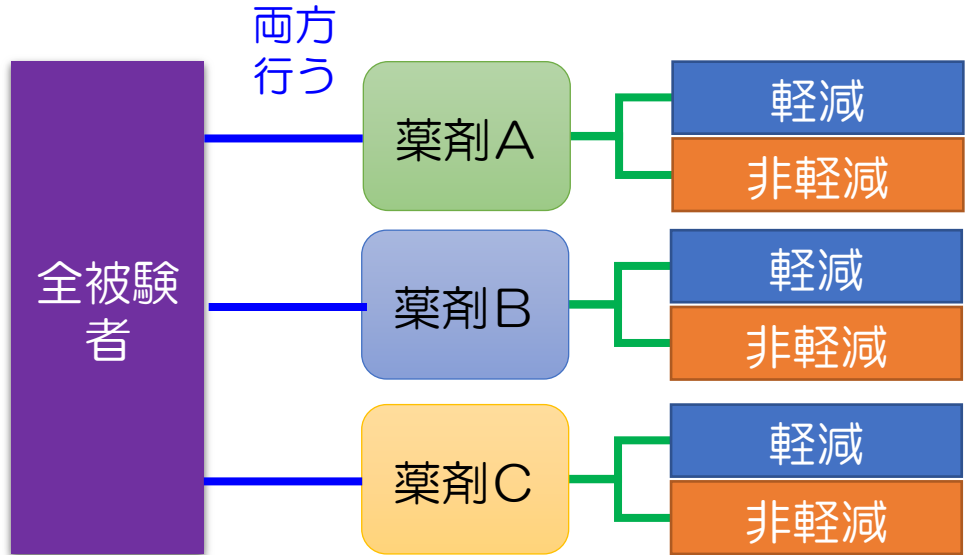
1. 「対応のある比率の比較(二分割表の対称性の検定、McNemar検定)」を選択して解析すると、[対応のあるクロス集計表](#)及び[McNemar検定](#)の結果が表示される。
2. 「分割表の直接入力と解析」を選択し「パーセントの計算」で「総計のパーセント」を選択する([全体パーセント](#))。

3群以上の比較の場合にはどうなるか？：CochranのQ検定



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)

- 1. 「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。
- 2. 「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。

テキストに適用手順あり，ここでは割愛



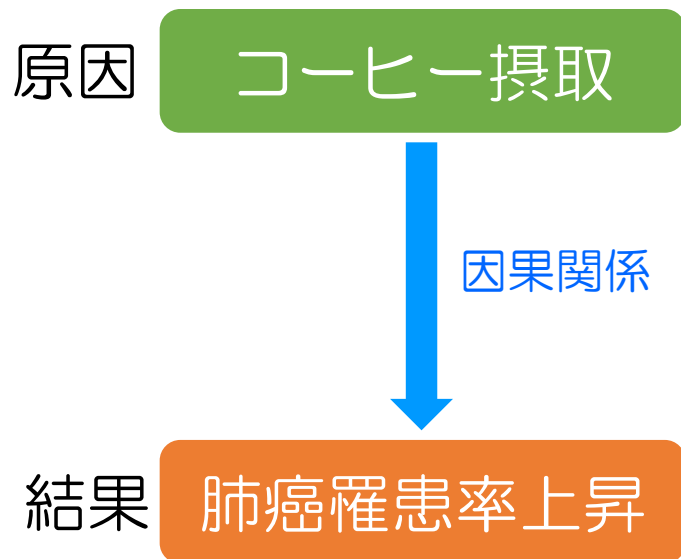
EZRによる演習(11)

「2.4.1.2 EZRによるMcNemar検定の実行」

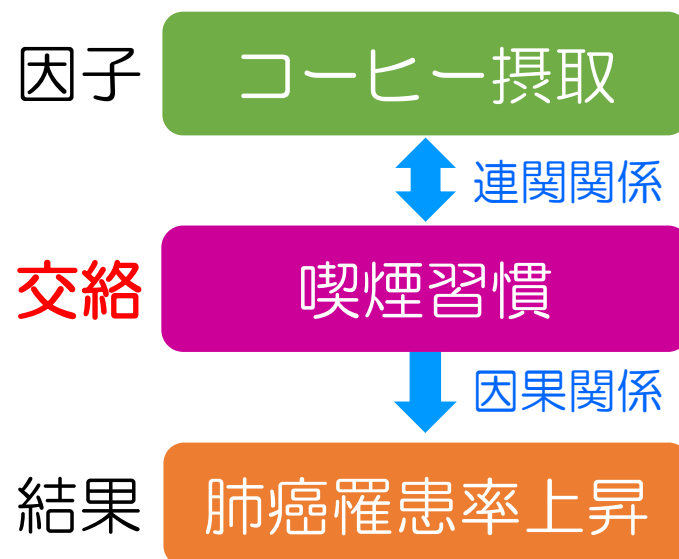
交絡とは

コーヒー摂取が肺癌の罹患率を上昇させるという研究結果が得られたとする。

研究において想定されたこと



実際に考えること



以下の関係がある場合に要因Xは交絡因子と呼ばれる。

1. 要因Xは疾患の既知のリスク要因である (例：喫煙習慣は肺癌のリスク因子である)。
2. 要因Xは因子と関連があるが、因子の結果ではない (例：喫煙はコーヒー摂取と連関関係はあるが、コーヒー摂取の結果ではない)。

共変量による調整を行った場合の検定：Cochran-Mantel-Haenszel検定

消化器癌(上部消化器, 下部消化器)に対する外科治療に対する縫合術(真皮縫合術, スキンステープラー)の比較の結果を以下に示す. 評価項目は, 創合併症の発現である(Tsujinaka et al., Lancet, 2013).

部位	縫合の方法	創合併症		計
		あり	なし	
上部	真皮縫合術	29 (0.076)	353 (0.924)	382
	スキンステープラー	39 (0.094)	374 (0.906)	413
	計	68 (0.086)	727 (0.914)	795
下部	真皮縫合術	18 (0.102)	158 (0.898)	176
	スキンステープラー	20 (0.198)	81 (0.802)	101
	計	38 (0.137)	239 (0.863)	277

■ 上部消化器がんにおけるスキンステープラーに対する真皮縫合術のオッズ比

$$OR = \frac{29 \times 374}{353 \times 39} = 0.788$$

■ 下部消化器がんにおけるスキンステープラーに対する真皮縫合術のオッズ比

$$OR = \frac{18 \times 81}{158 \times 20} = 0.461$$

部位別で縫合術の比較をした結果, いずれの部位でも真皮縫合術のほうが創合併症の発現は低く, 下部消化器がんでは半分未満になる.

共変量

部位	縫合の方法	創合併症		計
		あり	なし	
上部	真皮縫合術	29 O_{111}	353 O_{112}	382
	スキンステープラー	39 O_{121}	374 O_{122}	413
	計	68	727	795 N_1
下部	真皮縫合術	18 O_{211}	158 O_{212}	176
	スキンステープラー	20 O_{221}	81 O_{222}	101
	計	38	239	277 N_2

Mantel-Haenszel推定値

$$OR_{\text{adj}} = \frac{\frac{O_{111} \times O_{112}}{N_1} + \frac{O_{211} \times O_{212}}{N_2}}{\frac{O_{112} \times O_{121}}{N_1} + \frac{O_{212} \times O_{222}}{N_2}} = \frac{\frac{29 \times 374}{795} + \frac{18 \times 81}{277}}{\frac{353 \times 39}{795} + \frac{158 \times 20}{277}} = 0.658$$

共通オッズ比とは、共変量の層別を行ったうえで、それぞれの層に共通なオッズ比を表す。調整なしのオッズ比 $OR=0.709$ なので、手術部位で調整することで、オッズ比が小さくなっている。

Cochran-Mantel-Haenszel検定とは、共通オッズ比(Mantel-Haenszel推定量)が1であるという帰無仮説に対して、1でないという対立仮説を検定する。

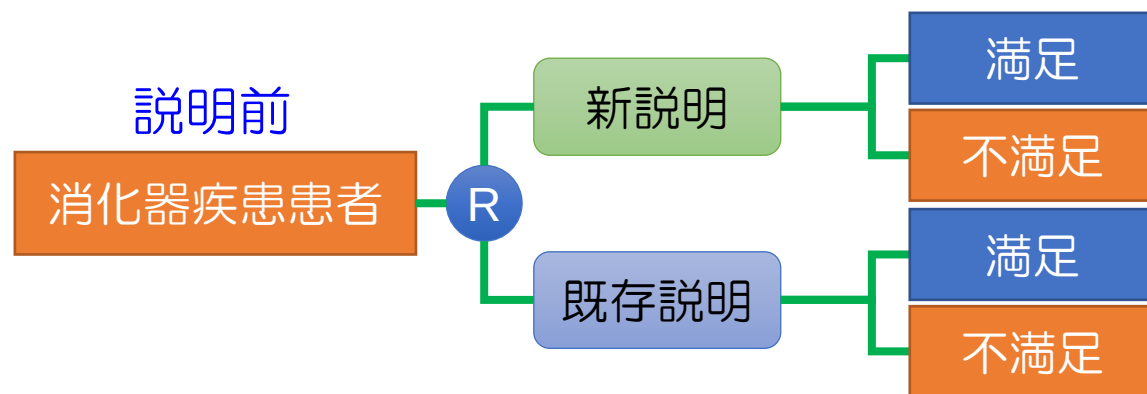


EZRによる演習(12)

「2.6 共変量調整を伴うクロス集計表の解析：Mantel-Haentzel検定」

ロジスティック回帰分析の説明の前に：オッズ比

いま、消化器疾患により、手術を行う患者に対して、これまでの事前説明を行った群と、新たに作成された説明プログラムに基づいて説明(新説明)を行った群にランダムに分けて行った。そして、説明後の満足度について調査した。



	満足	不満足	満足割合
新説明	43 (a)	7 (b)	$p_1=0.86$
既存説明	34 (c)	16 (d)	$p_2=0.68$

オッズとは、ある結果が生じる比率とある結果が生じない比率との比である。

$$Odds_{\text{新説明}} = \frac{p_1}{1-p_1} = \frac{0.86}{1-0.86} = 6.14, \quad Odds_{\text{既存説明}} = \frac{p_2}{1-p_2} = \frac{0.68}{1-0.68} = 2.13$$

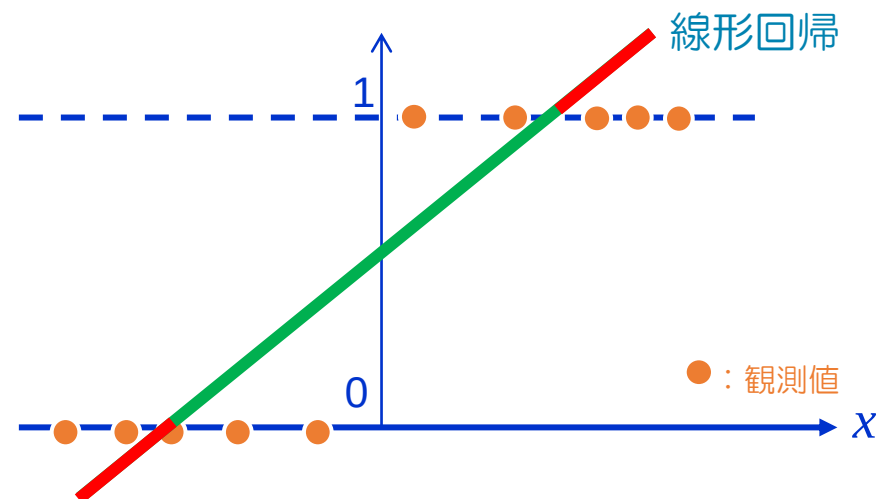
オッズは、関心がある事象の比率が0.5であれば1.0で、小数点以下であれば、その事象が起こる可能性が0.5未満 (半分未満でしか起こらない)、小数点以上であれば0.50超(半分以上に起こる)ことを意味する。

オッズ比はオッズの比率で表される指標である (たすき掛けでも計算できる)。

$$OR = \frac{Odds_{\text{新説明}}}{Odds_{\text{既存説明}}} = \frac{(a) \times (d)}{(b) \times (c)} = \frac{43 \times 16}{7 \times 34} = 2.891$$

つまり、新説明は既存説明に比べて、**2.891倍満足**していることがわかった。

ロジスティック回帰分析の動機

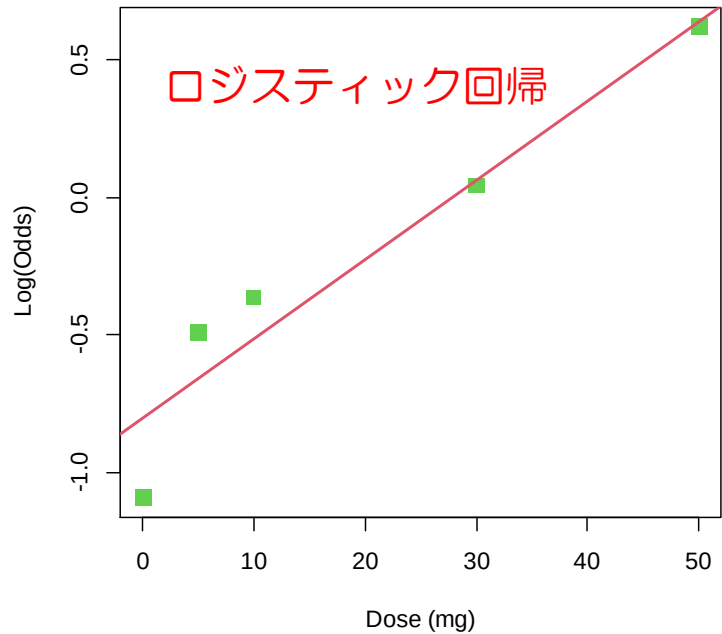


2値で与えられた応答変数に線形回帰分析を行うと. . .
➡ 応答変数の確率が1を超えたり, 0を下回ってしまう(赤色の部分)

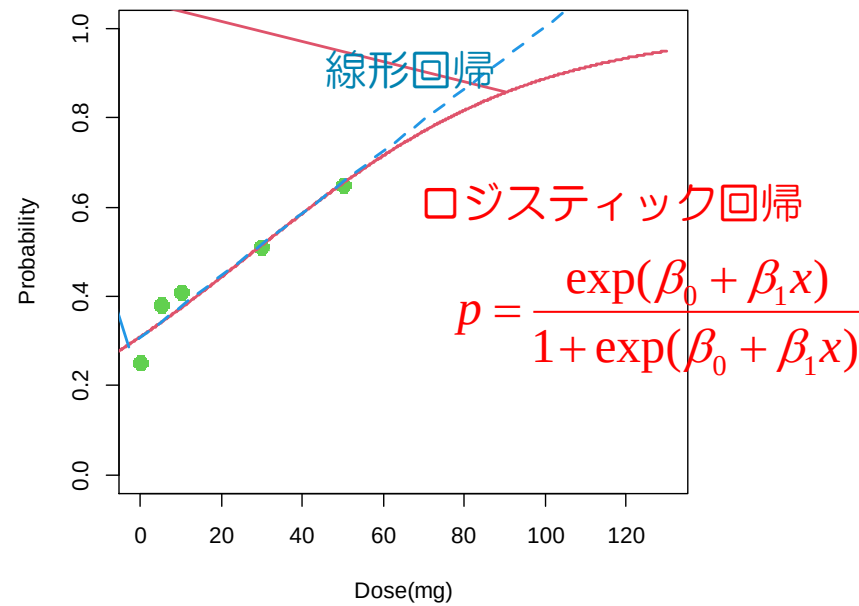
応答変数に対する対数オッズに対する線形モデルを考える

$$\log \frac{\Pr(y = 1)}{1 - \Pr(y = 1)} = \log \frac{p}{1 - p} = \beta_0 + \beta_1 x$$

薬剂量 (mg)	有効	無効	有効率
0	79	235	0.252
5	81	132	0.380
10	169	243	0.410
30	318	305	0.510
50	379	204	0.650



確率pに変換
➡



ロジスティック回帰では, 対数オッズに対して直線を当てはめる(最小2乗法ではない点に注意).
確率pに変換した後に描かれる曲線のことを, ロジット曲線(ロジスティック曲線)という.

ロジスティック回帰分析における調整オッズ比 (1/2)

いま，消化器疾患により，手術を行う患者に対して，これまでの事前説明を行った群と，新たに作成された説明プログラムに基づいて説明(新説明)を行った群にランダムに分けて行った．そして，説明後の満足度について調査した．このとき，説明を行った説明者のキャリアについても考慮した．

キャリア	説明の種類	満足／不満足		
		満足(1)	不満足(0)	合計
10年以上(1)	新説明(1)	77	7	84
	既存説明(0)	65	18	83
	合計	142	25	167
10年未満(0)	新説明(1)	70	17	87
	既存説明(0)	60	27	87
	合計	130	44	174

(多重)ロジスティック回帰分析のモデルは

説明の満足度に対する対数オッズ

$\log \frac{p}{1-p}$

=

切片

β_0

+

回帰係数1

β_1

×

説明種類

x_1

既存説明：0
新説明：1

+

回帰係数2

β_2

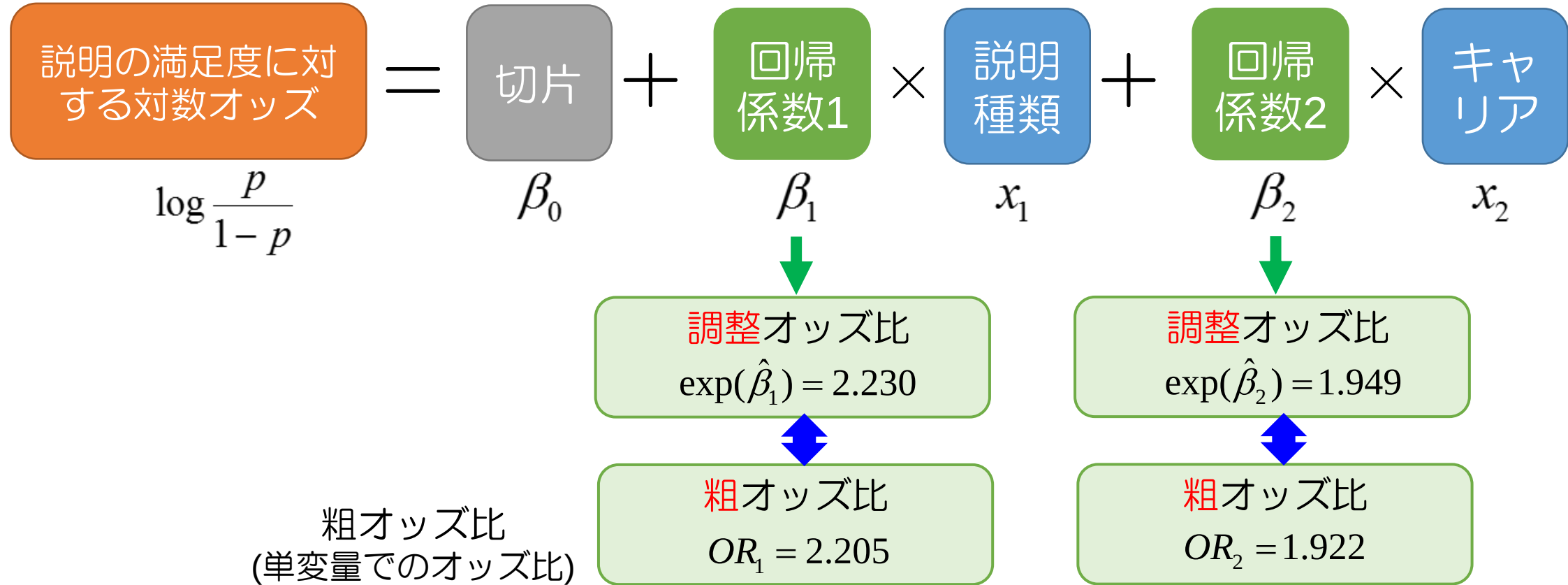
×

キャリア

x_2

10年未満：0
10年以上：1

ロジスティック回帰分析における調整オッズ比 (2/2)



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

- ・説明の種類のオッズ比で説明者のキャリアによる違いを考慮 (2.230倍 新説明に満足)
 - ・キャリアのオッズ比では説明の種類による違いを考慮 (1.922倍 10年以上の説明者の説明に満足)
- してオッズ比を計算している。

EZRによる計算における補足：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。



EZRによる演習(13)

テキストP87 「2.5.2 EZRによるロジスティック回帰の実行」



ご清聴ありがとうございます
toshibow2000@gmail.com