

医学統計セミナー2019
量的データの解析(1)
単群・2群比較のための諸種の検定

臨床研究センター
谷岡健資

2019年度 医学統計セミナー

開催日	講義内容
5月21日 (火)	記述統計・グラフ表示・統計ソフトウェア
6月18日 (火)	単群・2群比較のための諸種の検定
7月25日 (木)	分散分析と多重比較
8月22日 (木)	相関分析と回帰分析
9月26日 (木)	クロス集計表の解釈と諸種の検定
10月17日 (木)	ロジスティック回帰分析
11月14日 (木)	生存曲線の推定と比較
12月 3日 (火)	Cox比例ハザードモデル
1月23日 (木)	感度・特異度・ROC曲線とその比較
2月27日 (木)	繰り返し測定分散分析、混合効果モデル

目次

□ はじめに

- 推定・検定の概観
- 推定とは
- 検定とは
- 本セミナーの目的

□ 推定

- 推定について
- 母平均の区間推定
- 母平均の差の区間推定

□ 検定の導入

□ 単群の検定

- 平均値の検定

□ 2群の検定

- 母平均の差の検定 おわりに

□ おわりに

1.はじめに

□ 推定・検定の概観

推定の例（丹後，2015）

ある病院の腎機能障害の患者に対する血清クレアチニン濃度（mg/dl）の平均値を知りたい

➤ 結果

血清クレアチニン濃度の平均値：4.0（5名）

- 本結果からある病院の腎機能障害の患者に対する血清クレアチニン濃度の平均は4.0と考えるよいか。また，その理由について考えてみてください



1.はじめに

□ 推定・検定の概観

検定の例（丹後，2015）

ある病院の検査において**非飲酒群**と**飲酒群**で γ -GTPの値に差があるかを検討することを考える

➤ 結果

非飲酒群の平均値：31.79（34名）

飲酒群の平均値：54.28（64名）

➤ 本結果から飲酒習慣により γ -GTPの値は変動を受けるといえるか？
また，その理由について考えてみてください



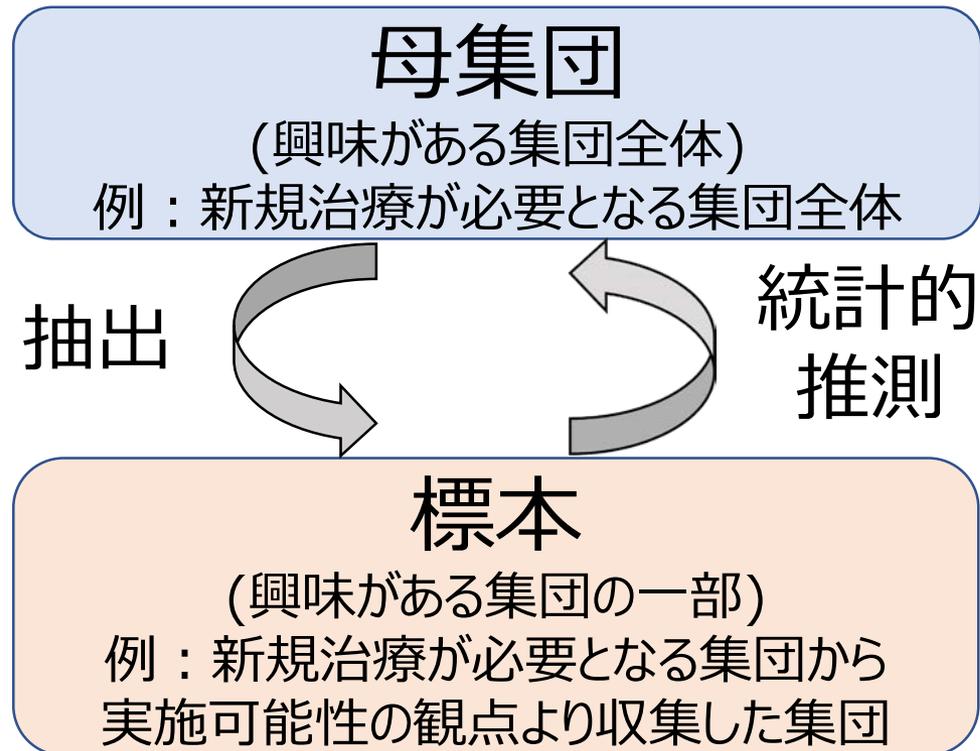
1.はじめに

□推定・検定の概観

医学系研究・看護研究等において，興味がある集団全員を調査することは時間・コスト面・実施可能性の観点から一般には不可能

調査した被験者の数値から興味ある集団の特徴を把握する必要性

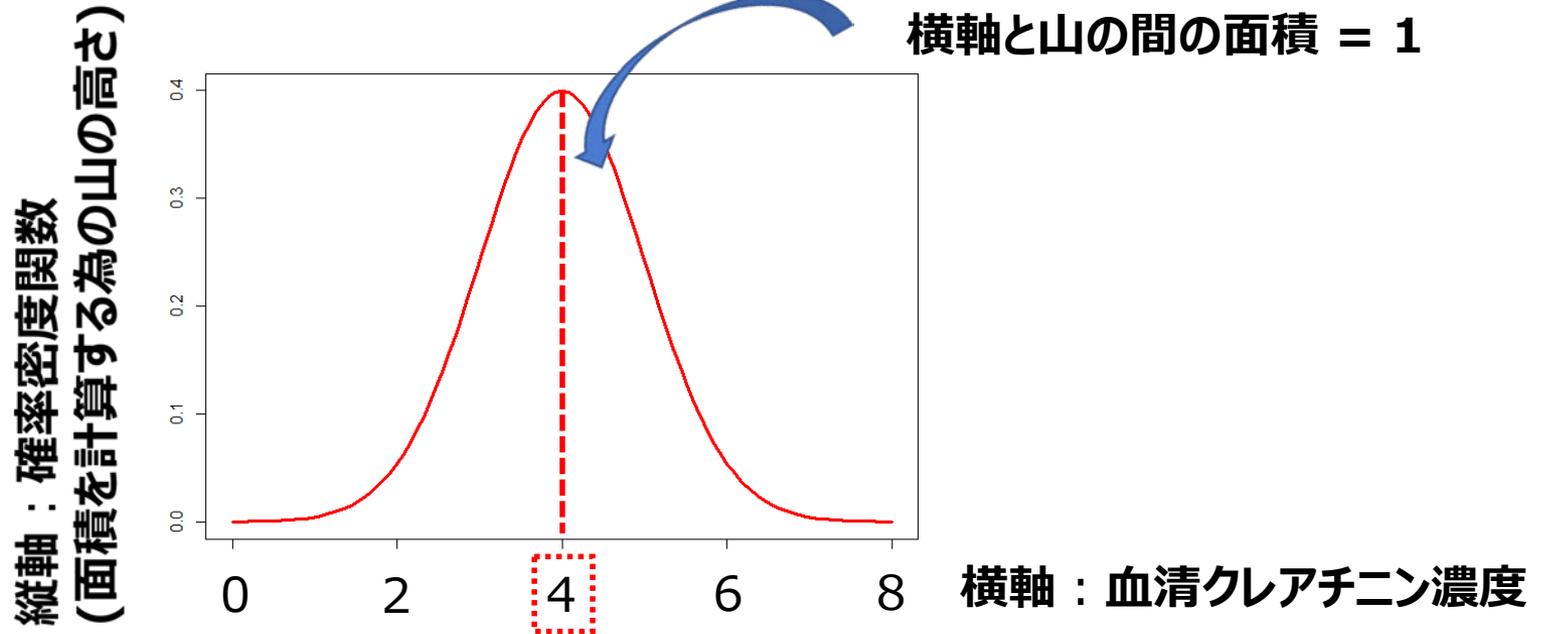
⇒ 推定・検定が必要となる



1.はじめに

□ 推定・検定の概観

母集団の記述について(例：血清クレアチニン濃度)



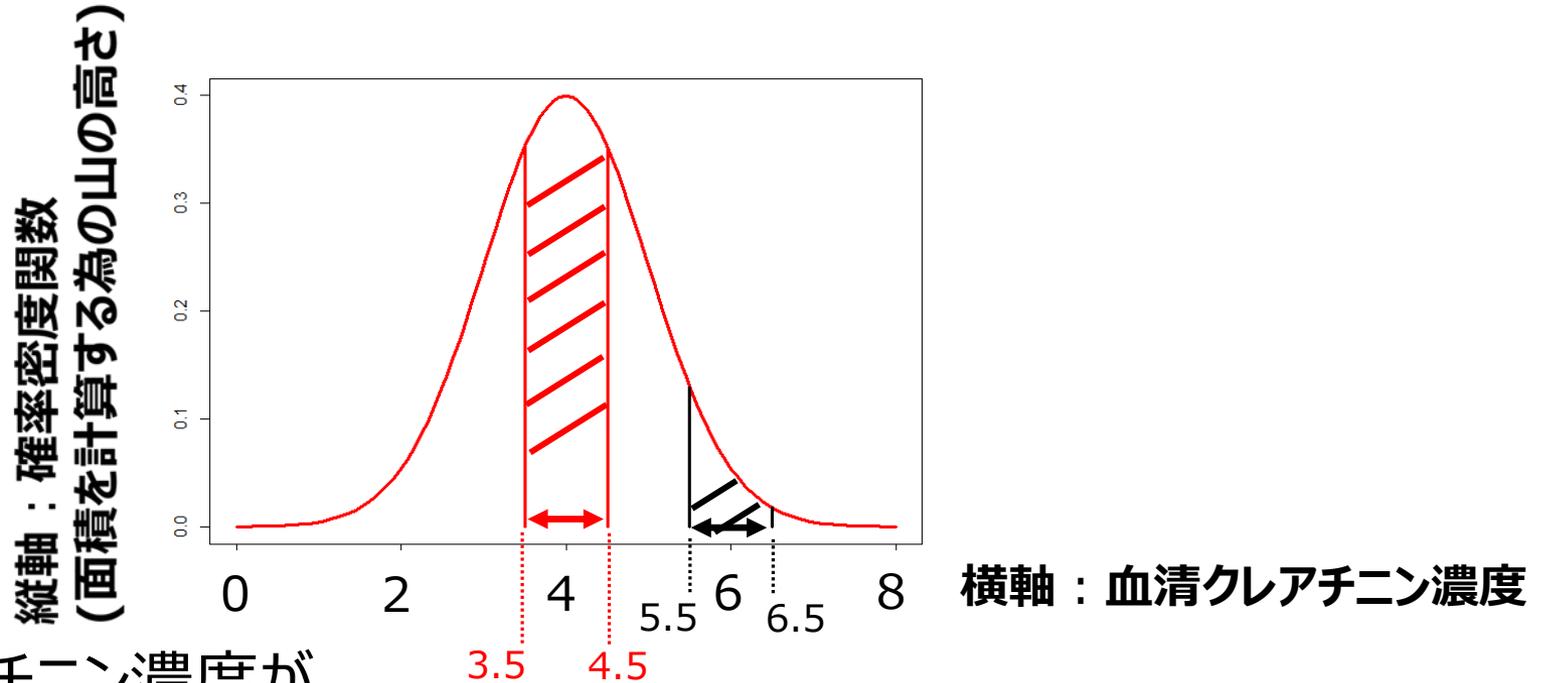
母集団から観測値（データ）が得られる経緯を確率分布によって表現することを仮定している

母集団の期待値（平均）は4であるとこの図から解釈できる

1.はじめに

□ 推定・検定の概観

母集団の記述について(例：血清クレアチニン濃度)



血清クレアチニン濃度が

3.5以上4.5未満である確率：赤斜線部分の面積

5.5以上6.5未満である確率：黒斜線部分の面積

この場合、「5.5以上6.5未満の値」よりも「3.5以上4.5未満の値」を調査によって得られる確率が高いと解釈できる

1.はじめに

□推定・検定の概観

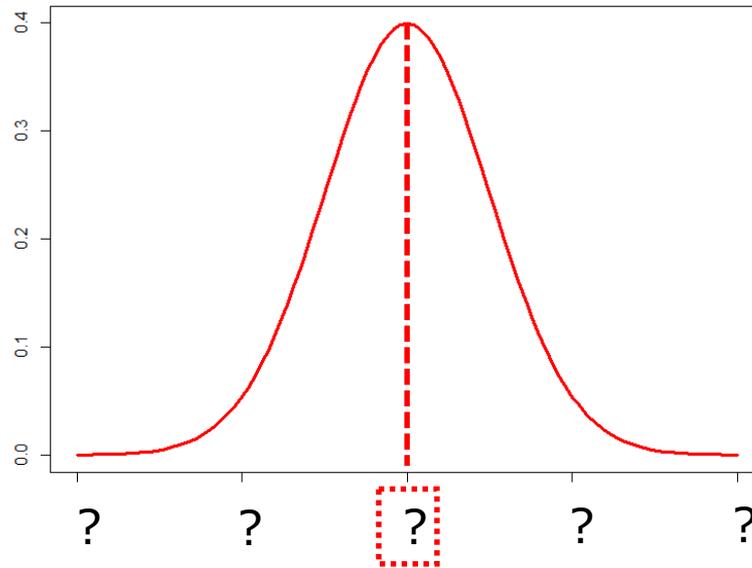
母集団の記述について(例：血清クレアチニン濃度)

一般に母集団の山の形状は**未知**

母集団の特徴を予測することは山の形状を固定すること！

= 形状を固定しないと確率を計算できない！！

縦軸：確率密度関数
(面積を計算する為の山の高さ)



山の形状：？(未知)

横軸：血清クレアチニン濃度

山の形状を予測(固定)するには・・・

よく使用される分布例) 正規分布の形状:2つの値により固定(平均・分散)

t 分布の形状を:1つの値により固定(自由度)

1.はじめに

□ 推定とは

例 再掲 (丹後, 2015)

ある病院の腎機能障害の患者に対する血清クレアチニン濃度 (mg/dl) の平均値を知りたいという例を通して説明

点推定とは

母集団の興味ある特徴を1つの値で推定する方法

例：標本からある病院の当該患者の血清クレアチニン濃度が 4.0 であるという一つの値で推測

区間推定とは

母集団の興味ある特徴が〇〇%の確率で当該区間に入っているという形式で推定する方法 (〇〇%は解析者が決定)

例：標本からある病院の当該患者の血清クレアチニン濃度が 95%の確率で区間 [3.78, 4.22] に入ると推測

1.はじめに

□検定とは

例（丹後，2015）

あるタイプの肝炎患者のアルカリフォスファターゼ（ALP）値が別のタイプの肝炎患者のALP平均値226(IU/l)と差があるか否か

検定

標本から母集団の特徴に関する仮説が成り立っているか否かを判断するための推測方法

例：

帰無仮説（否定したい仮説）：ALP値の母平均 = 226(IU/l)

対立仮説（興味ある仮説）：ALP値の母平均 \neq 226(IU/l)

主張したい仮説を対立仮説に設定する必要がある（後述）

1.はじめに

□検定とは

検定の論理（背理法）

「対立仮説が成り立つ」と証明したいときに、
「対立仮説が成り立たない = 帰無仮説が成り立つ」と仮定して
標本から矛盾を導くことより「対立仮説が成り立つ」と結論づける方法

検定の概要

Step 1: **帰無仮説**と**対立仮説**を設定する

Step 2: **帰無仮説**が成り立つと仮定する

Step 3: **帰無仮説**のもとで**標本（データ）の結果**が、矛盾するか否かを判断する

Step 4: 標本の結果と**帰無仮説**が矛盾する場合
⇒**帰無仮説**を棄却（**対立仮説**を主張）

標本の結果と**帰無仮説**が矛盾しない場合
⇒**帰無仮説**を否定する根拠がない

1.はじめに

□検定とは

対立仮説と帰無仮説について

例【再掲】：

帰無仮説（否定したい仮説）：ALP値の母平均 = 226(IU/l)

対立仮説（興味ある仮説）：ALP値の母平均 \neq 226(IU/l)

背理法をもちいて

「**等しい**」ということと「**違う**」ということとはどちらが示しやすいでしょうか？



1.はじめに

□検定とは

検定結果を信用できるのか？

質問

下記のような**帰無仮説**と**対立仮説**が与えられ、**帰無仮説**を棄却したとする。検定結果を100%信用してよいか？

= 「ALP値の母平均が226ではない」という結論を信用してよいか？

帰無仮説 (否定したい仮説) : ALP値の母平均 = 226(IU/l)

対立仮説 (興味ある仮説) : ALP値の母平均 \neq 226(IU/l)



1.はじめに

□検定とは

検定結果を信用できるのか？

⇒100%信用できないものの**第I種の過誤**と**第II種の過誤**を可能な限り制御したい（意思決定を行うのに重要）

第I種の過誤

本当は**帰無仮説**が正しい ⇒ 検定で**帰無仮説**を棄却してしまう誤り

対策：有意水準を定めることで制御

例)有意水準を0.05と定める ⇒ **第I種の過誤**を0.05に設定する

第II種の過誤

本当は**対立仮説**が正しい ⇒ 検定で**帰無仮説**を受容してしまう誤り

対策：サンプルサイズ（標本数）を増やすことで第II種の過誤が減少する。

1.はじめに

□検定について

検定結果を信用できるのか？

質問

有意水準を0.05として設定し、サンプルサイズを十分にとって第II種の過誤が0.20以下となるようなもとの検定を行った。

他に検討すべきことはないか？



1.はじめに

□本セミナーの目的

量的尺度に基づくデータを観測された際に、下記の方法について紹介する.

- 点推定と区間推定
- 一標本 t 検定
- 二標本対応無し t 検定
- 二標本対応あり t 検定
- ウィルコクソンの順位和検定
- ウィルコクソンの符号順位検定

2. 推定

□ 推定について

一般に研究発表や論文発表等では結論を主張するために、「**検定**」が重宝されることが多い。

質問

推定はどこで必要になってくるのか？

⇒ 主に区間推定について下記が使われる

1. 検定を目的にしてサンプルサイズを設計しても、必要症例数を収集できない場合、信頼区間がある程度の範囲に収まるよう、サンプルサイズを設計する場合
2. 群が多数あり、比較する場合
3. 群間の背景因子に差がないか示す場合

2.推定

□母平均の区間推定

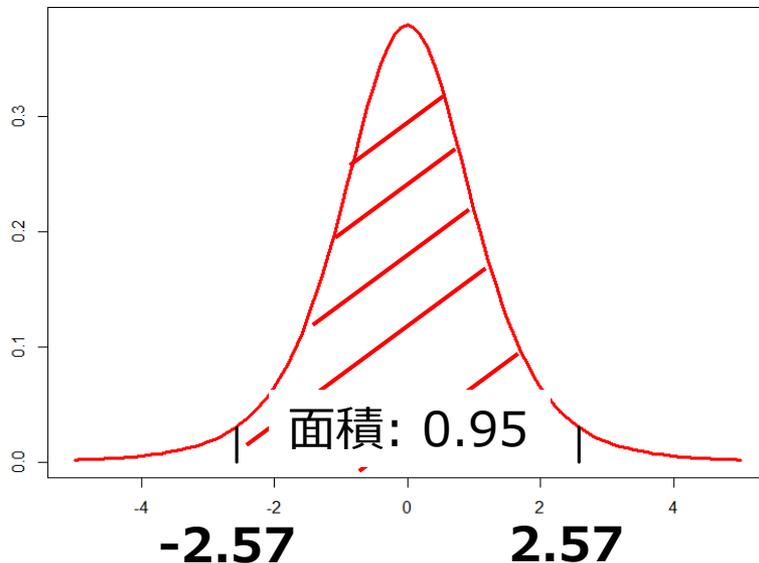
例 **再掲** (丹後, 2015)

ある病院の腎機能障害の患者に対する血清クレアチニン濃度 (mg/dl) の平均値を知りたいという例を通して説明

表：標本

	被験者1	被験者2	被験者3	被験者4	被験者5	被験者6
クレアチニン濃度	4.0	3.9	3.8	4.0	4.4	3.9

自由度5のt分布



標本平均：4 標準偏差：0.21

$$4 - 2.57 \times \frac{0.21}{\sqrt{6}} \leq \text{母平均} \leq 4 + 2.57 \times \frac{0.21}{\sqrt{6}}$$
$$= [3.78, 4.22]$$

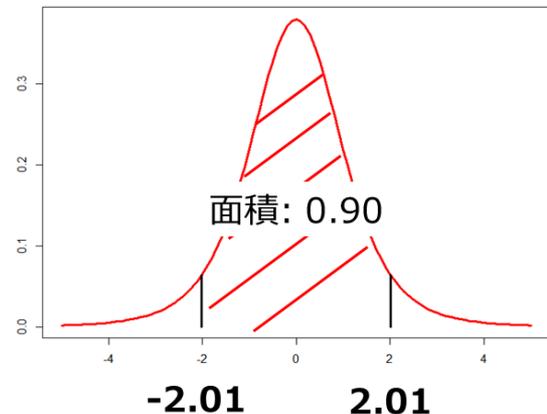
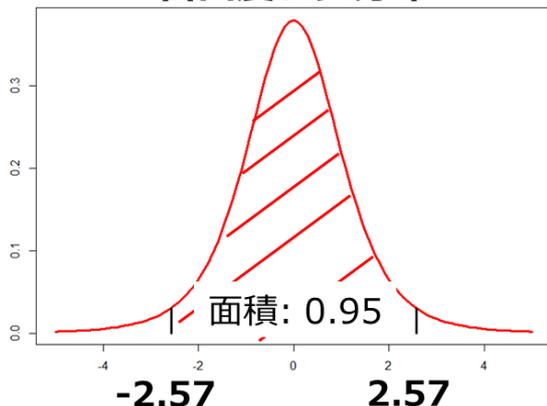
解釈：

血清クレアチニン濃度の母平均は95%の確率で3.78から4.22の間に入っている

2.推定

母平均の区間推定

自由度5のt分布



* 下記の他にサンプルサイズの増加によって区間幅は短くなる
* また、標準偏差が小さくなれば区間幅はやはり短くなる

95%信頼区間 [3.78, 4.22]

血清クレアチニン濃度の母平均は95%の確率で3.78から4.22の間に入っている

90%信頼区間 [3.83, 4.17]

血清クレアチニン濃度の母平均は90%の確率で3.83から4.17の間に入っている

区間幅は〇〇%を低くすると短くなる傾向にある(ただし、信頼度は ↓)

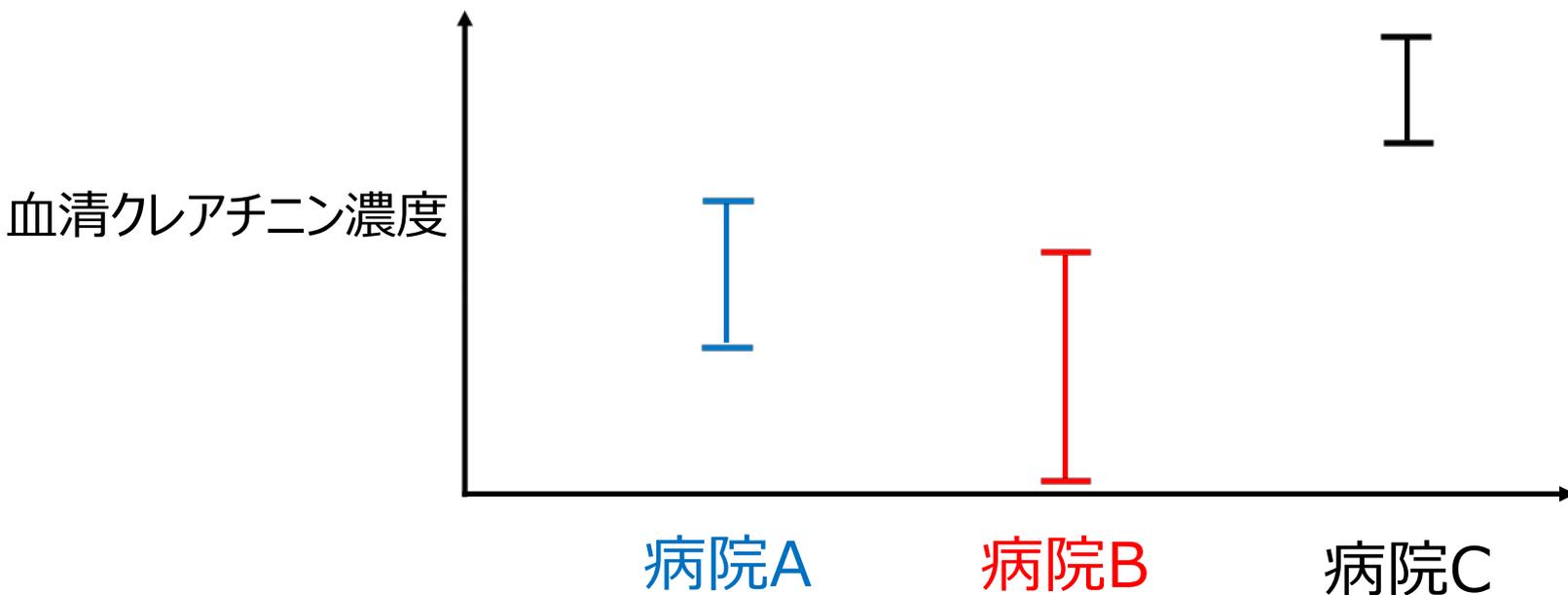
上記のように区間幅を用いて母平均の推定について評価することができる

信頼区間によって、サンプルサイズを設計することができる！

2.推定

□母平均の区間推定

95%信頼区間とそのplotについて（「2.の説明」）



図は病院ごとの「血清クレアチニン濃度」の95%信頼区間を表現している

解釈

病院A vs 病院B: 信頼区間が重複しているので差があるとは言えない

病院A vs 病院C: 信頼区間が重複していないので差がありそう

病院B vs 病院C: 信頼区間が重複していないので差がありそう

2.推定

□母平均の差の区間推定

例（丹後，2015）

透析患者の免疫グロブリンの一つIgG値(mg/100ml)が健常者に比べて高いか否かを調査

表：調査結果

透析患者	1326	1418	1820	1516	1635	1720	1580	1452	1600
健常者	1220	1080	980	1420	1170	1290	1116		

標本平均値の差

透析患者のIgG値の平均値 - 健常者のIgG値平均値 = 381

興味があるのは透析患者全体のIgG値の平均 - 健常者全体のIgG値の平均
⇒ 母平均の差が95%の確率で入っている区間（母平均の差の区間推定）

透析患者のIgG値の母平均 - 健常者のIgG値の母平均の95%信頼区間

$$218.9 \leq \text{透析患者のIgG値の母平均} - \text{健常者のIgG値の母平均} \leq 543.1 \\ = [218.9, 543.1]$$

2.推定

□母平均の差の区間推定

母平均の差の95%信頼区間の解釈の仕方

質問

95%信頼区間がどのような区間であるとき、「**差がある**」といえ、どのようなときに「**差があるとは言えない**」となるか？



2. 推定

□ 母平均の差の区間推定

母平均の差の95%信頼区間の解釈の仕方

質問

95%信頼区間がどのような区間であるとき、「**差がある**」といえ、どのようなときに「**差があるとは言えない**」となるか？

⇒

「差がある」場合

「母平均の差の95%信頼区間」に**0を含んでいない**場合

「差があるとは言えない」場合

「母平均の差の95%信頼区間」に**0を含んでいる**場合

例：前頁の例

$218.9 \leq \text{透析患者のIgG値の母平均} - \text{健常者のIgG値の母平均} \leq 543.1$

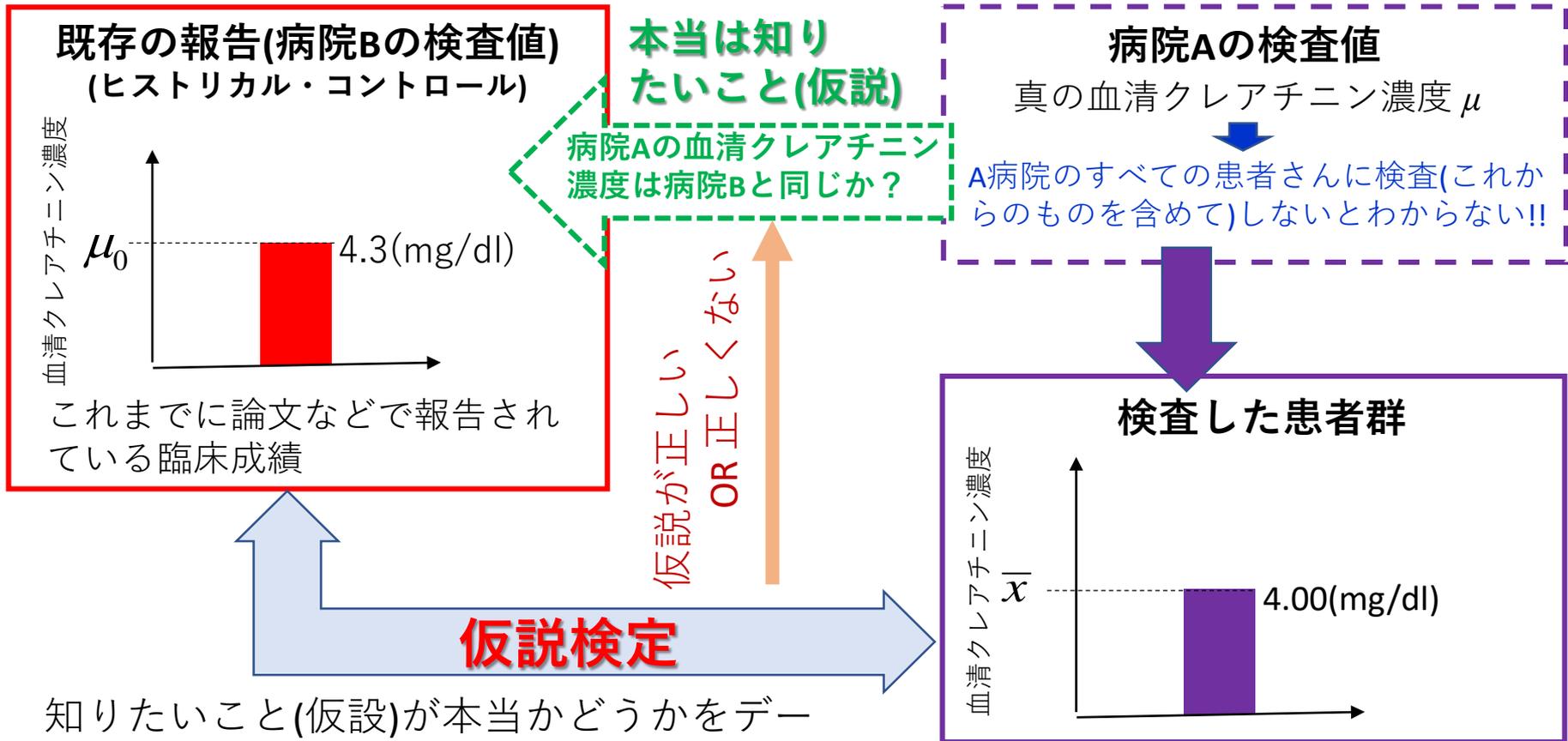
⇒ 母平均の差の95%信頼区間に0を含んでいないことから差がありそう・・・

3. 検定の導入

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。



知りたいこと(仮説)が本当かどうかをデータに基づいて統計的に評価する

3. 検定の導入

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。

3. 検定の導入

どのように「誤っている」と判断するのか？

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。

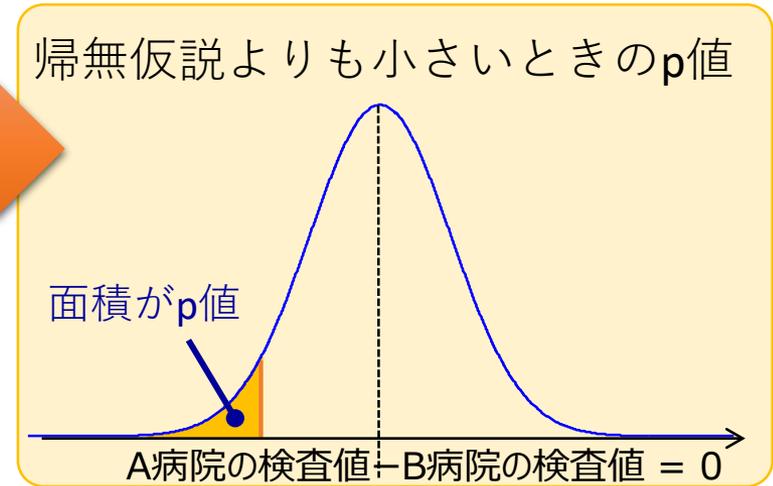
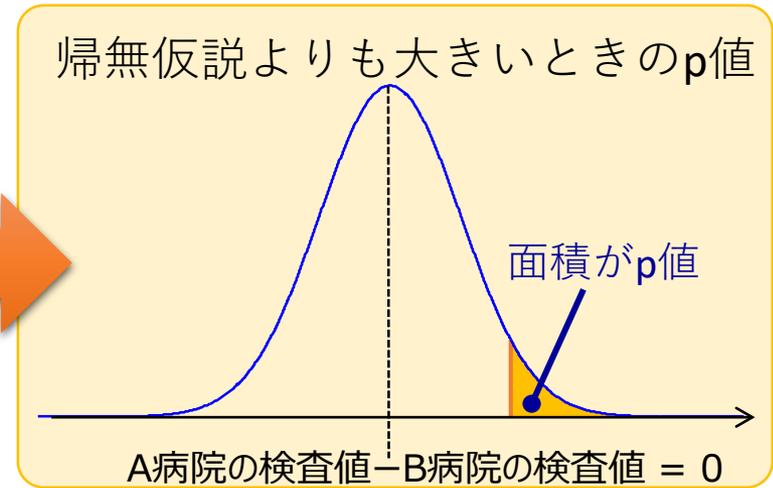
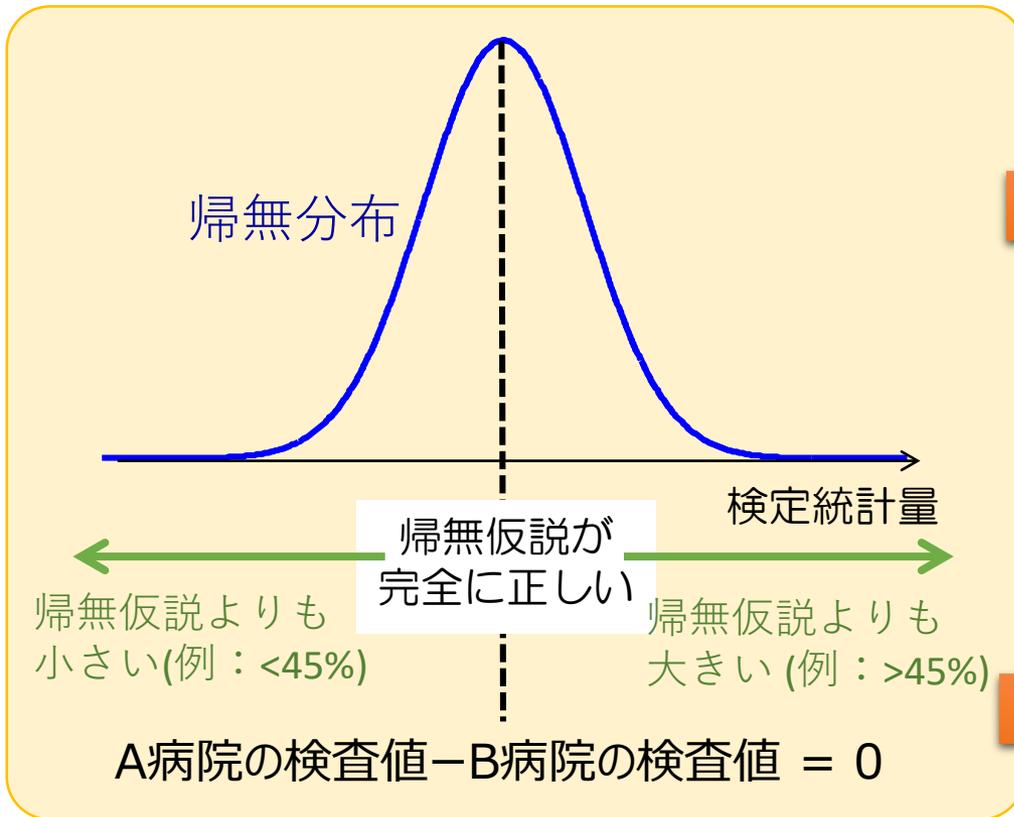
P値の計算には、**検定統計量**という数値を計算する。帰無仮説が正しいときに、検定統計量がどれぐらいの確率で得られるものなのかを計算する。こうして得られたものが、**p値**である。

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。

3. 検定の導入

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。

3. 検定の導入

どのように「有意である」と判断するのか？

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*1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。

3. 検定の導入

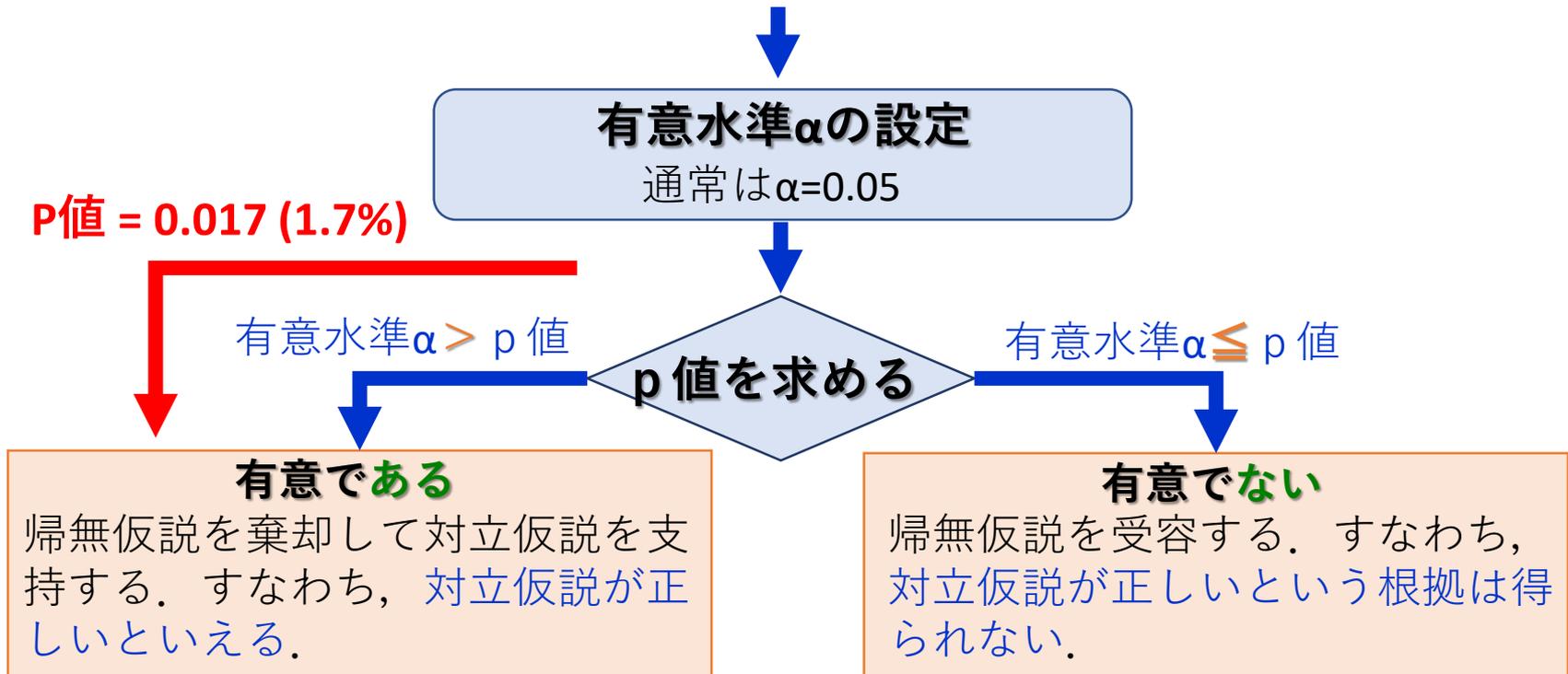
検定の流れの整理

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

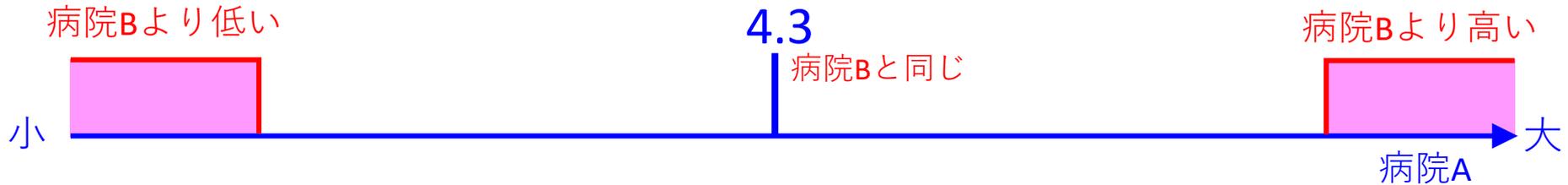
帰無仮説 H_0 : A病院の血清クレアチニン濃度の平均値 μ は, B病院の平均値 μ_0 と**同じ**.

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**.



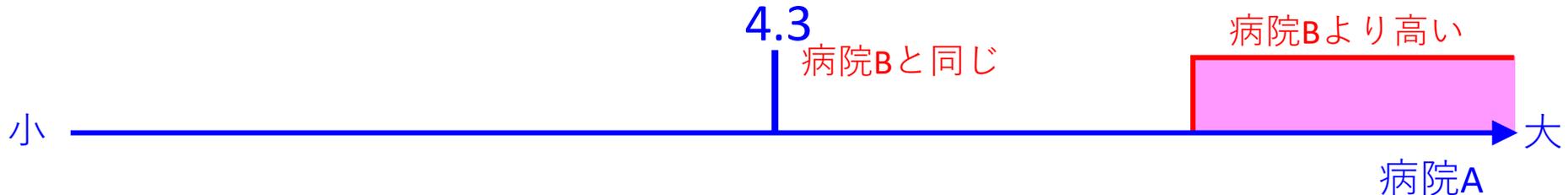
3.検定の導入

■ 両側対立仮説：病院Aと病院Bに違いがある。



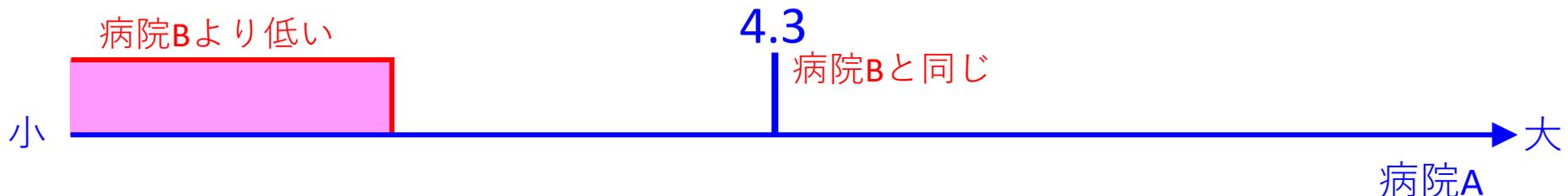
有意水準は「病院Bの4.3(mg/dl)よりも高い確率」「病院Bの4.3(mg/dl)よりも低い確率」の両方で設定されるので、片側では $\alpha/2$ になる。

■ 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。

3. 検定の導入

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

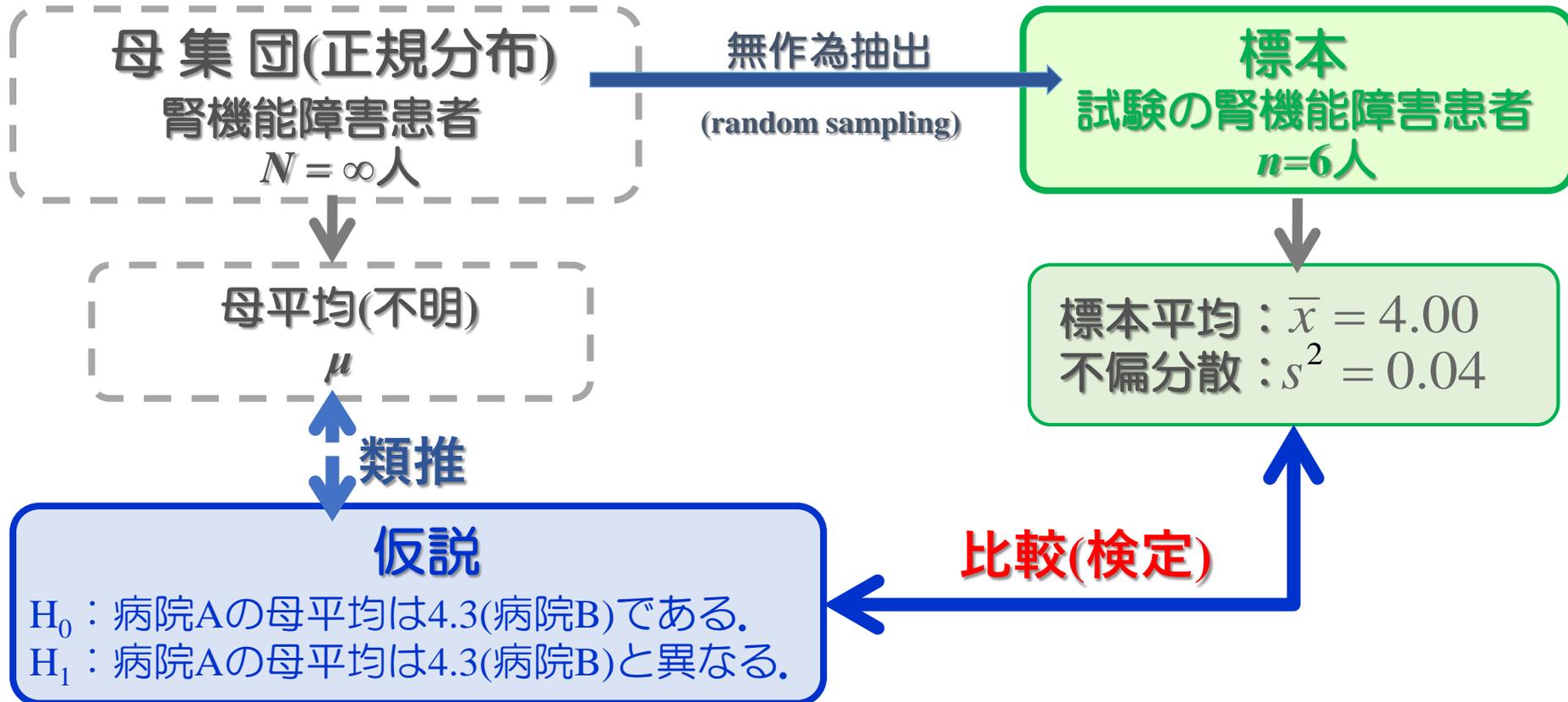
Welch検定

4. 単群の検定

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。



4. 単群の検定

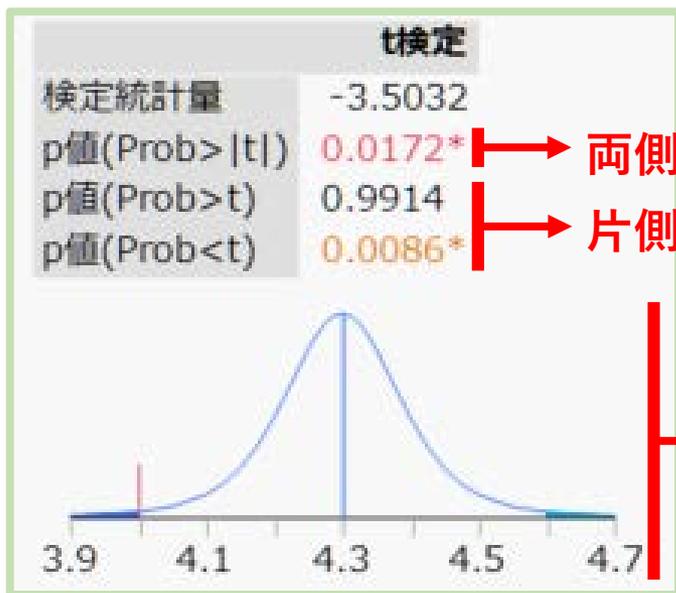
1標本 t 検定：JMPでの結果

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。

「分析」 → 「一変量の分布」 → 変数名横の「▼」をクリック → 「平均の検定」を選択



両側対立仮説の結果

片側対立仮説の結果

帰無分布の結果

両側対立仮説におけるp値が0.0172なので、有意水準 $\alpha=0.05$ よりも小さい。つまり、帰無仮説が棄却され、対立仮説が支持された(有意だった)。

したがって、病院Aの血清クレアチニン濃度は病院Bの4.3(ml/dl)と異なっていた。

4. 単群の検定

このような場合はどのように考えたらよいか？

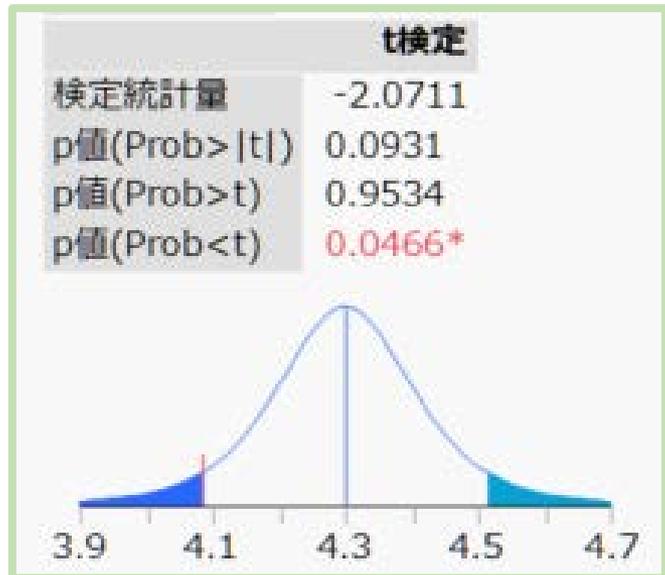
病院Cに通院する、腎機能障害の患者6名の血清クレアチニン濃度(mg/dl)を測定したところ

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)。

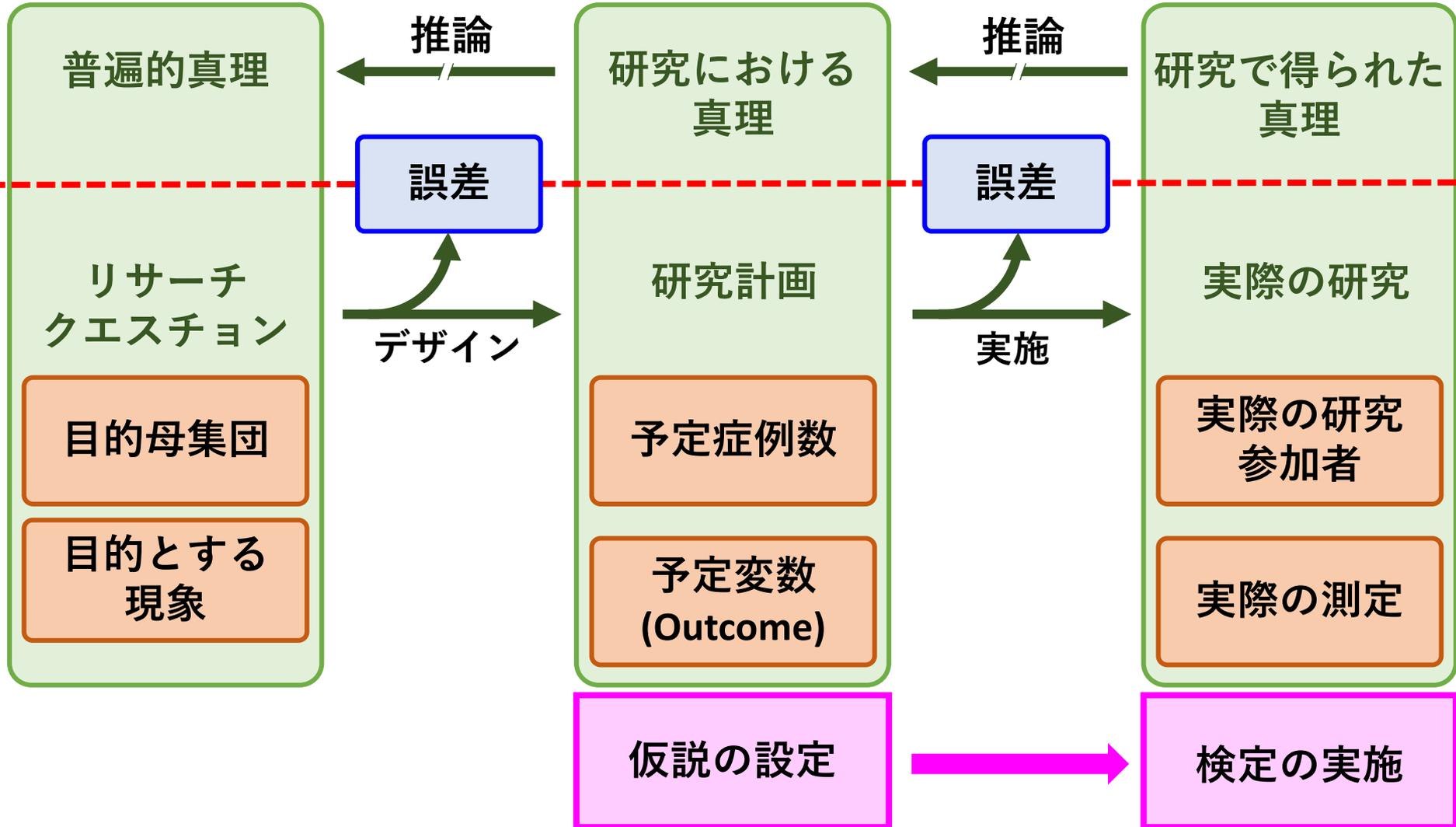
したがって、病院Cの血清クレアチニン濃度は病院Bの4.3(mg/dl)と異なるとはいえなかった。

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よいか？

4. 単群の検定

このような場合はどのように考えたらよいか？



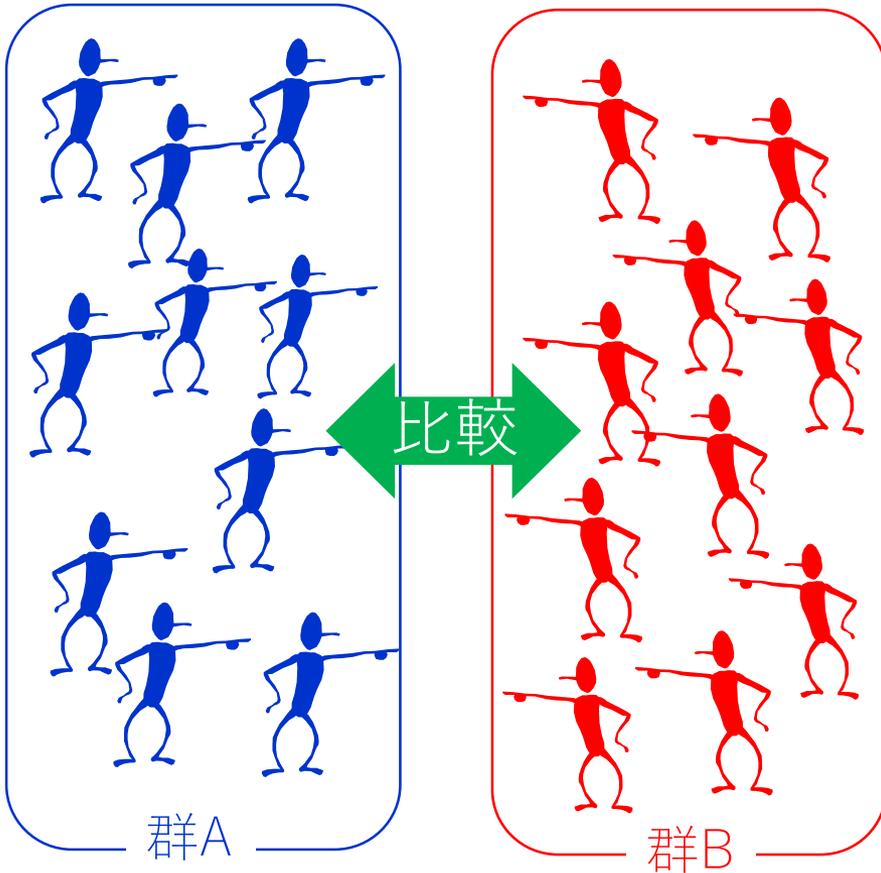
両側・片側の選択は、研究計画で行われるため、実際の研究において、それを変更することは許されない。

5.2 標本の検定

対応のないデータ(独立2標本)

■ 例：無作為化比較試験

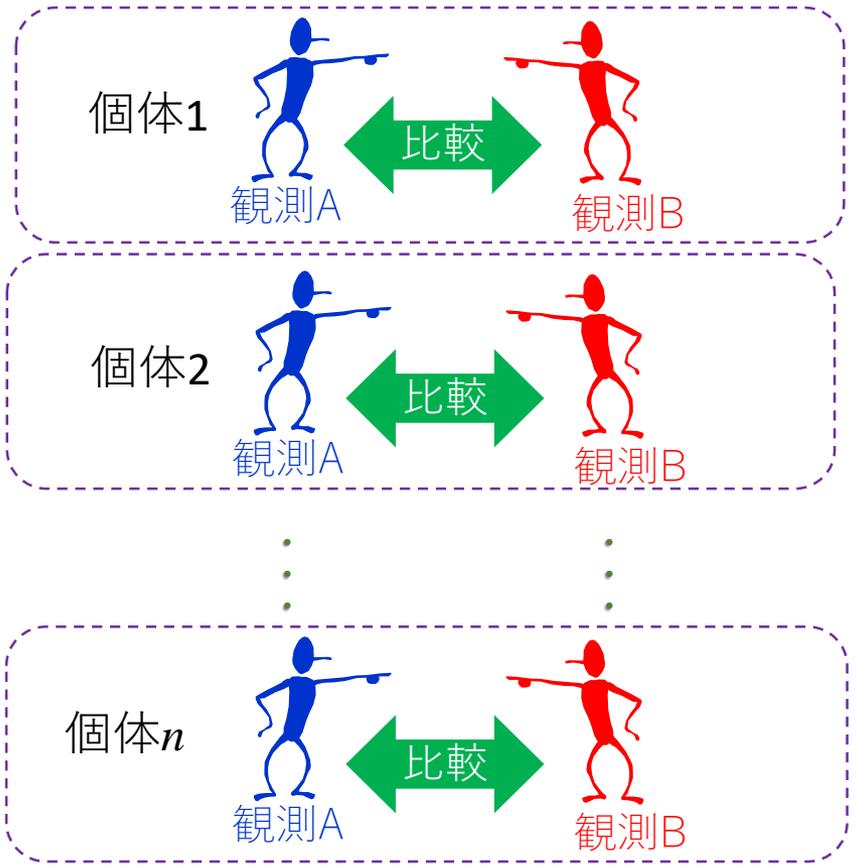
治療Aと治療Bを異なる被験者に実施して有効性を比較する。



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。



5.2標本の検定

対応のある t 検定

ある薬剤を投与し，その前後の検査値を10人の患者について調査した．この薬剤の効果(数値が増加したか)があったか否かを検討しなさい．

	A	B	C	D	E	F	G	H	I	J
治療前：	65	75	60	85	80	60	55	60	70	80
治療後：	70	75	70	80	90	85	60	70	70	85

	治療前	治療後	効果		治療前	補習後	効果
Aさん 	65	↔ 70	5	Fさん 	60	↔ 85	25
Bさん 	75	↔ 75	0	Gさん 	55	↔ 60	5
Cさん 	60	↔ 70	10	Hさん 	60	↔ 70	10
Dさん 	85	↔ 80	-5	Iさん 	70	↔ 70	0
Eさん 	80	↔ 90	10	Jさん 	80	↔ 85	5

平均効果
6.5

効果ある
といいて
良いか

5.2標本の検定

対応のある t 検定

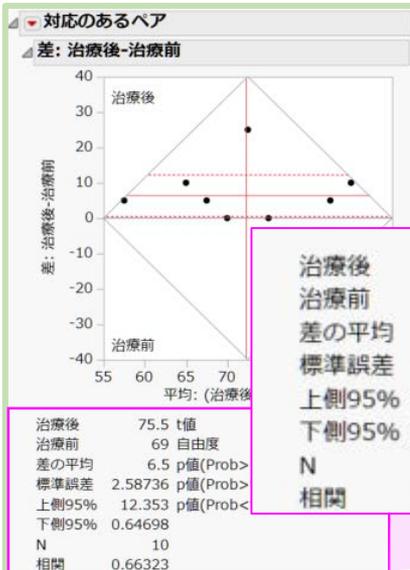
各個体(被験者)に対して、「治療後 - 治療前」を計算し、その平均変化量に対して、0であるか否かを検定する。変化量が正規分布に従うとするならば、帰無仮説での平均変化量 $\mu_0=0$ としたときの、1標本t検定と計算している内容は同じである。

H_0 : 治療後前後での検査値の平均変化量に違いはない。

➡ 変化量(治療後 - 治療前)の母平均 μ は0である。

H_1 : 治療後の検査値のほうが治療前よりも大きい。

➡ 変化量(治療後 - 治療前)の母平均 μ は0より大きい。



JMP : 「分析」 → 「対応のあるペア」 で自動的に計算される

片側対立仮説におけるp値が0.016なので、有意水準 $\alpha=0.05$ よりも小さい。つまり、帰無仮説が棄却され、対立仮説が支持された(有意だった)。

したがって、治療後の検査値のほうが治療前より大きかった。

5.2標本の検定

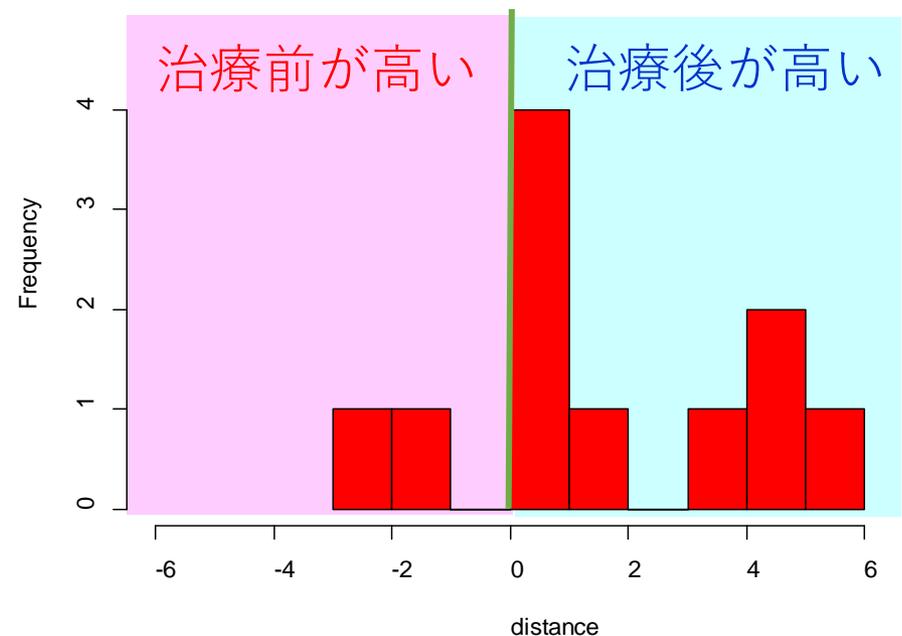
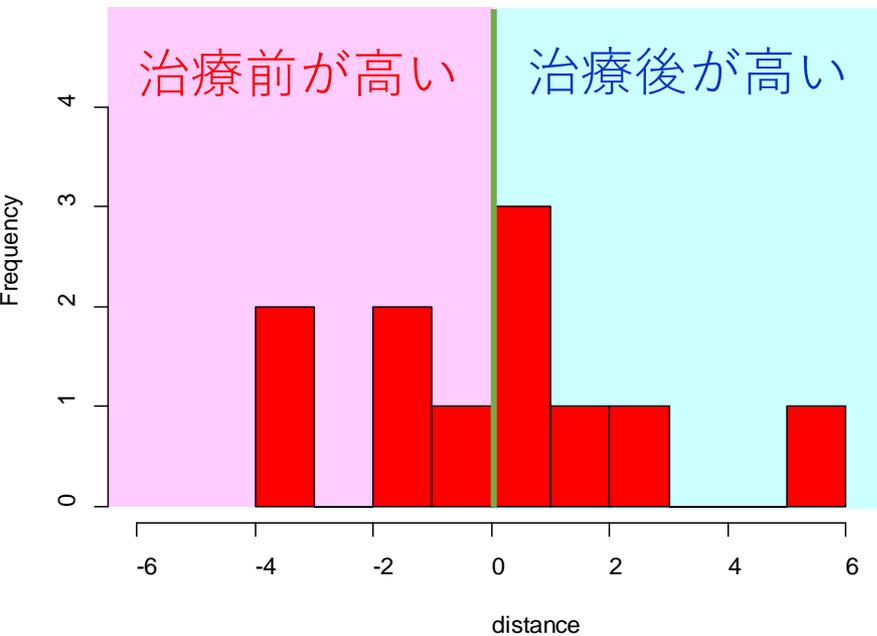
正規分布に従わない場合の対応のある検定(Wilcoxon符号付順位検定)

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				

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				

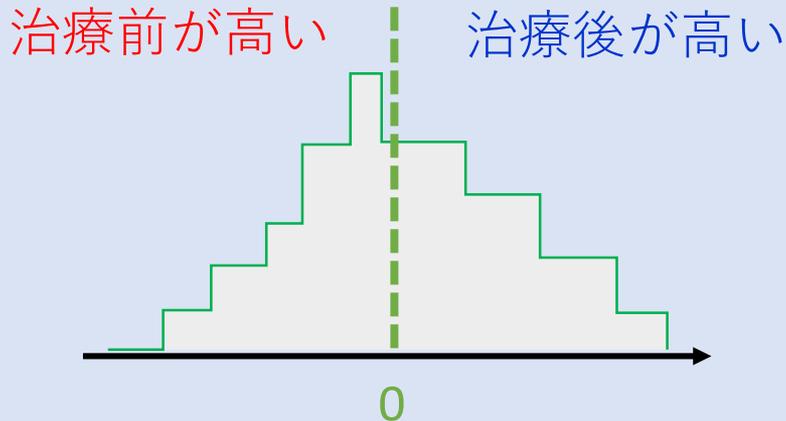


有意である状況では片側に多くのデータが存在する

5.2標本の検定

正規分布に従わない場合の対応のある検定(Wilcoxon符号付順位検定)

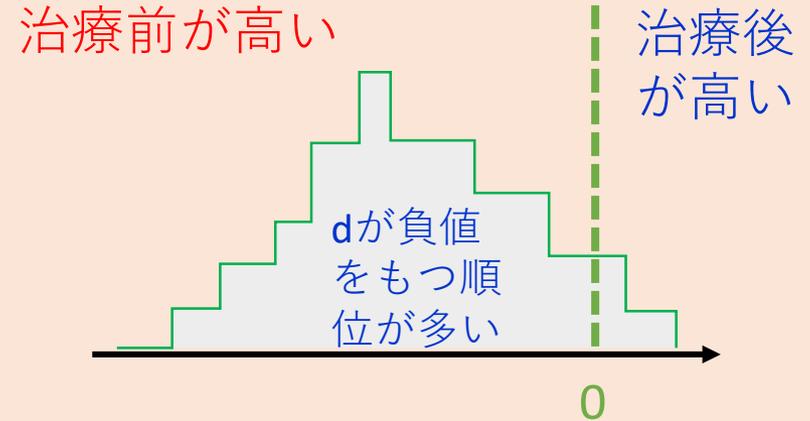
帰無仮説



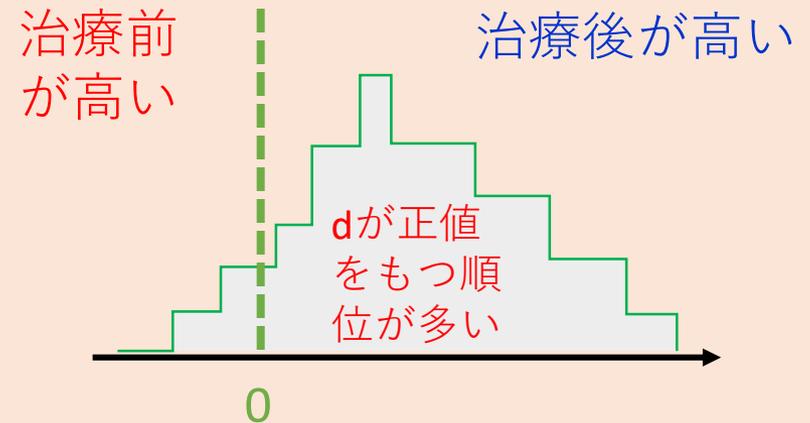
治療前後での差 d を順位に変換しその順位に基づいて位置(経験分布での期待値)が0であるか否かを検定する。

対立仮説

治療前が高い場合



治療後が高い場合



H_0 : 治療前後で検査値に違いはない。

H_1 : 治療後の検査値のほうが治療前よりも大きい。

5.2標本の検定

ある薬剤を投与し、その前後の検査値を10人の患者について調査した。この薬剤の効果(数値が増加したか)があったか否かを検討しなさい。

1. $d_i = x_{B_i} - x_{A_i}$ を計算する (0の場合には削除する)

	A	B	C	D	E	F	G	H	I	J
治療前	65	75	60	85	80	60	55	60	70	80
治療後	70	75	70	80	90	85	60	70	70	85
効果	5	0	10	-5	10	25	5	10	0	5



効果	5	10	-5	10	25	5	10	5
----	---	----	----	----	----	---	----	---

2. d_i の正負の符号により2群に分ける。

効果	5	10	-5	10	25	5	10	5
----	---	----	----	----	----	---	----	---

3. $|d_i|$ の正負の符号により2群に分ける。

効果	-5	5	5	5	10	10	10	25
順位	2.5	2.5	2.5	2.5	6	6	6	8

$$\leftarrow (1 + 2 + 3 + 4) / 4 = 2.5 \rightarrow \quad \leftarrow (5 + 6 + 7) / 3 = 6 \rightarrow$$

4. 検定統計量 $W = \min[\text{正值の群の順位和}, \text{負値の群の順位和}]$ を計算する。

$$W = \min[2.5, 33.5] = 2.5$$

得られた検定統計量Wに基づいてp値を計算する

5.2 標本の検定

ある薬剤を投与し、その前後の検査値を10人の患者について調査した。この薬剤の効果(数値が増加したか)があったか否かを検討しなさい。

JMP : 「分析」 → 「対応のあるペア」 → 対応のあるペア横「▼」をクリック
→ 「Wilcoxonの符号付検定」を選択

Wilcoxonの符号付順位検定	
治療後-治療前	
検定統計量S	21.500
p値(Prob> S)	0.0391*
p値(Prob>S)	0.0195*
p値(Prob<S)	0.9805

H_1 : 治療後と治療前は異なる :

H_1 : 治療後のほうが治療前よりも大きい : p値(Prob>S) 0.0195* → 片側対立仮説の結果

H_1 : 治療後のほうが治療前よりも小さい : p値(Prob<S) 0.9805 → 片側対立仮説の結果

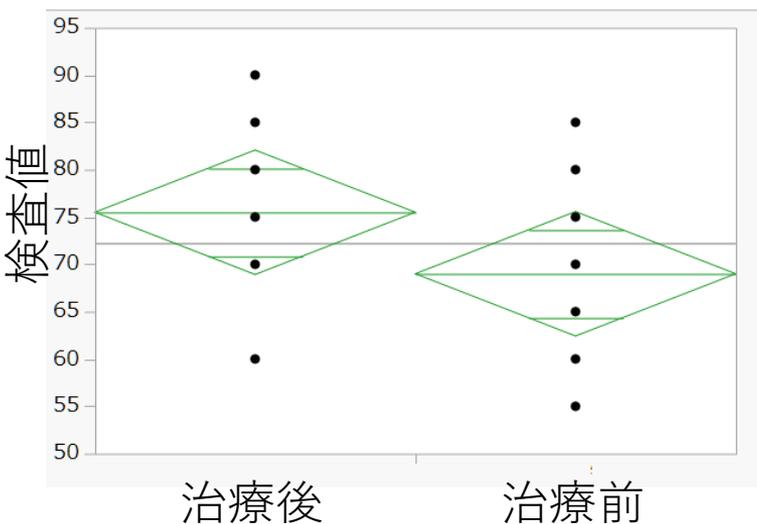
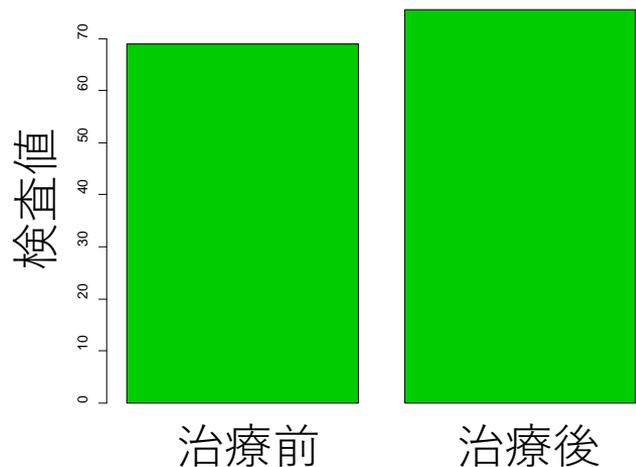
注意1 : JMPでは、検定統計量が異なるが、これは帰無分布の計算のために変換していることに由来

注意2 : 対立仮説において平均値が出てこないことに注意(先ほどの例示参照)

片側対立仮説におけるp値が0.0195なので、有意水準 $\alpha=0.05$ よりも小さい。つまり、帰無仮説が棄却され、対立仮説が支持された(有意だった)。したがって、治療後の検査値のほうが治療前より大きかった。

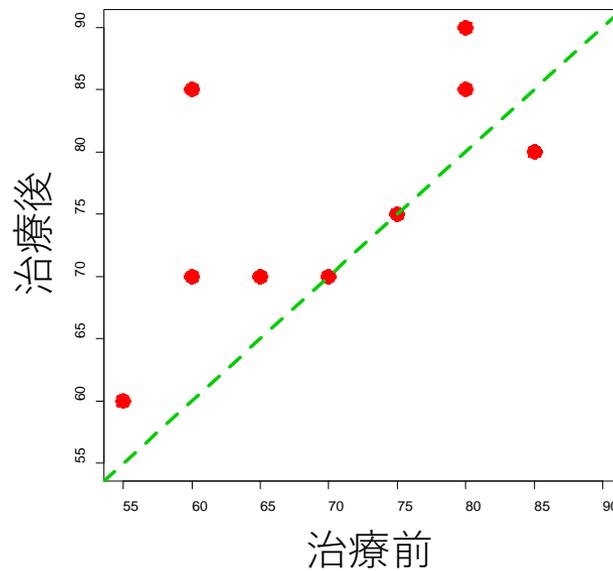
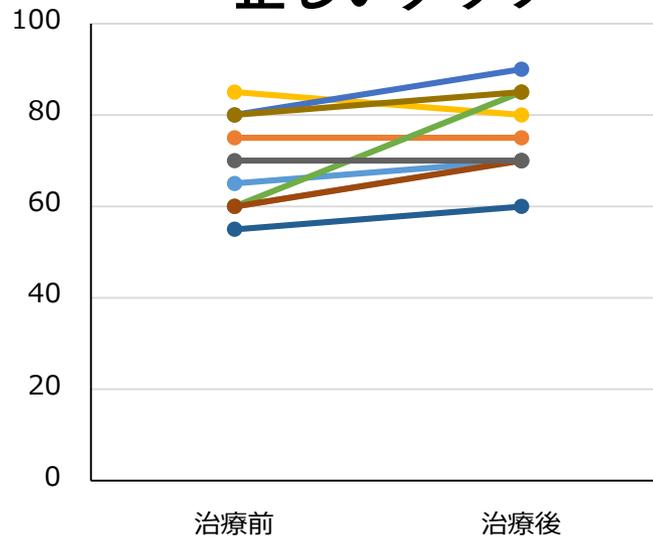
5.2 標本の検定

対応のあるデータのグラフ表示の注意点
誤ったグラフ



治療前後のそれぞれの平均値を比較
しているわけではない

正しいグラフ



5.2 標本の検定

正規分布に従う場合の2標本データの解析： 2標本t検定

神経障害性疼痛患者を対象に、2種類の除痛薬(新薬, 既存薬)の治療前後でのVAS (mm)の減少量を試験した。

新薬 (n=14)

31 25 28 29 23 25 30 25 29 27 30 20 20 24

既存薬 (n=12)

23 23 20 27 19 15 25 29 15 13 28 21

仮説の設定

H_0 : 新薬と既存薬でVASの減少量(の母平均)に差がない。

H_1 : 新薬と既存薬でVASの減少量(の母平均)に差がある。

5.2標本の検定

正規分布に従う場合の2標本データの解析： 2標本t検定

神経障害性疼痛患者を対象に、2種類の除痛薬(新薬, 既存薬)の治療前後でのVAS (mm)の減少量を試験した。

JMP：「分析」→「二変量の関係」→○○の一元配置の横の「▼」をクリック
→「平均/ANOVA/プーリングしたt検定」を選択



両側対立仮説におけるp値が0.0143なので、有意水準 $\alpha=0.05$ よりも小さい。つまり、帰無仮説が棄却され、対立仮説が支持された(有意だった)。したがって、新薬と既存薬には差があることが示された。

5.2標本の検定

注意点：対応のあるデータを2標本t検定で解析してはならない

ある薬剤を投与し、その前後の検査値を10人の患者について調査した。この薬剤の効果(数値が増加したか)があったか否かを検討しなさい。

対応のあるt検定 (正しい解析)

治療後	75.5	t値	2.512211
治療前	69	自由度	9
差の平均	6.5	p値(Prob> t)	0.0332*
標準誤差	2.58736	<u>p値(Prob>t)</u>	<u>0.0166*</u>
上側95%	12.353	p値(Prob<t)	0.9834
下側95%	0.64698		
N	10		
相関	0.66323		

P値が0.0166で有意である

2標本t検定 (誤った解析)

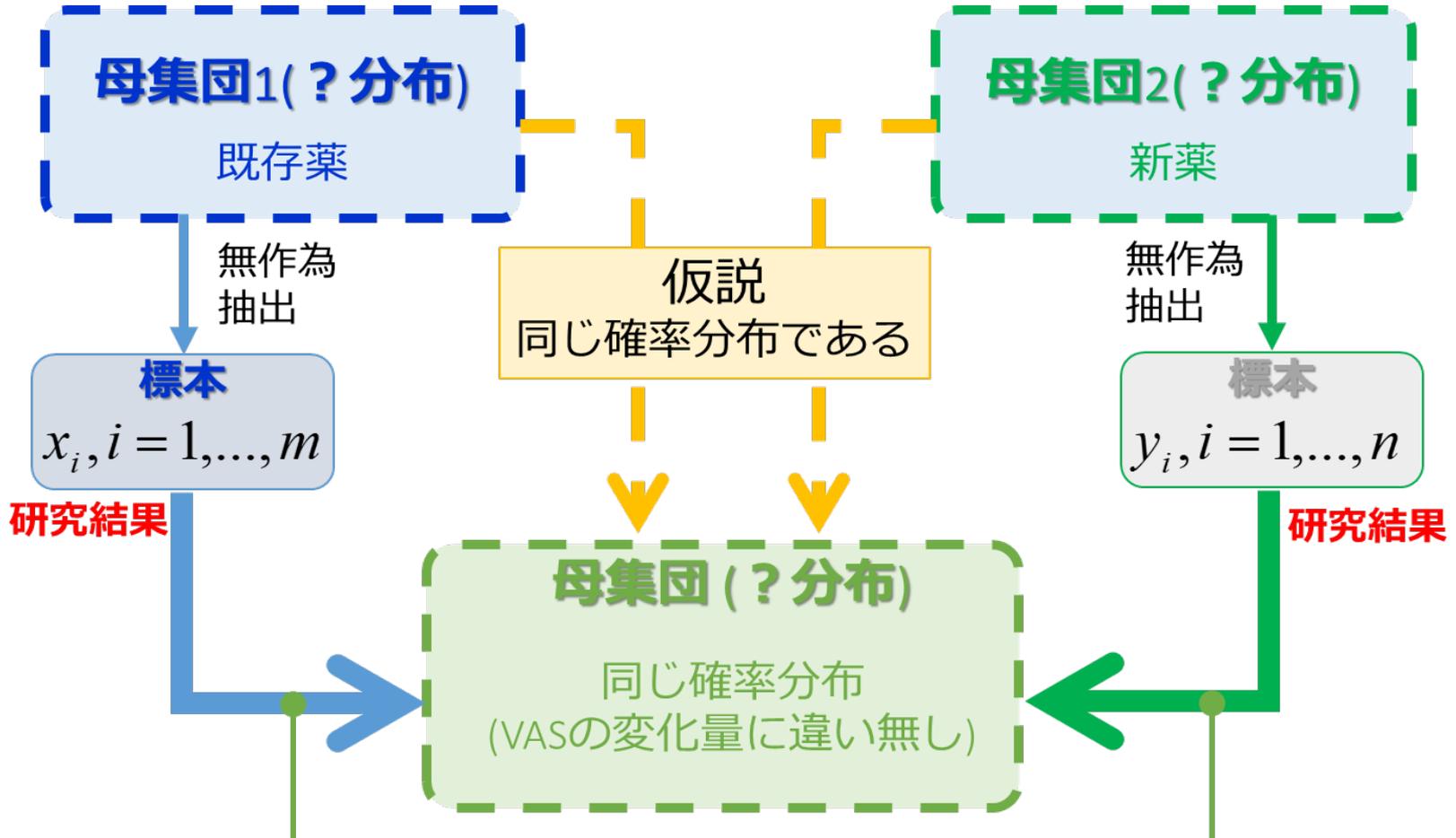
Pre-Past			
分散が等しいと仮定			
差	-6.500	t値	-1.46882
差の標準誤差	4.425	自由度	18
差の上側信頼限界	2.797	p値(Prob> t)	0.1591
差の下側信頼限界	-15.797	p値(Prob>t)	0.9204
信頼率	0.95	<u>p値(Prob<t)</u>	<u>0.0796</u>

P値が0.0796で有意でない

対応のあるデータを2標本t検定で検定する場合、分散(バラツキ)の影響によって、有意(positive)な結果を有意でない(negative)な結果にしてしまう恐れがある。

5.2 標本の検定

正規分布に従わない場合の2標本データの解析: Wilcoxon検定

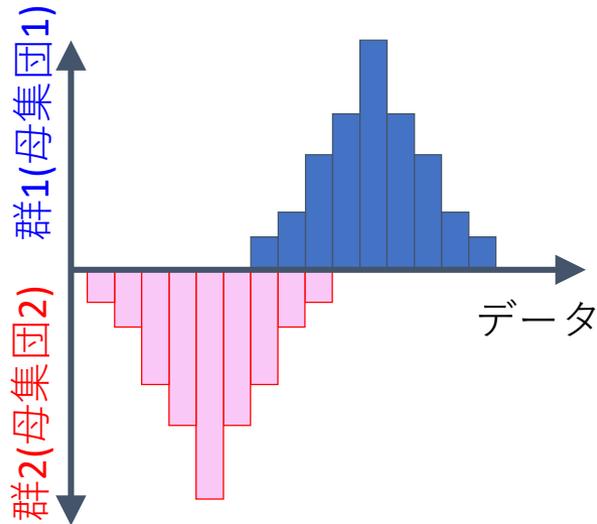


母集団1および母集団2から得られた標本が同じ母集団分布からの無作為標本であるかを検討することで、これらの母集団が同一の確率分布をもつという仮説を検証するのが**Wilcoxon検定**である。

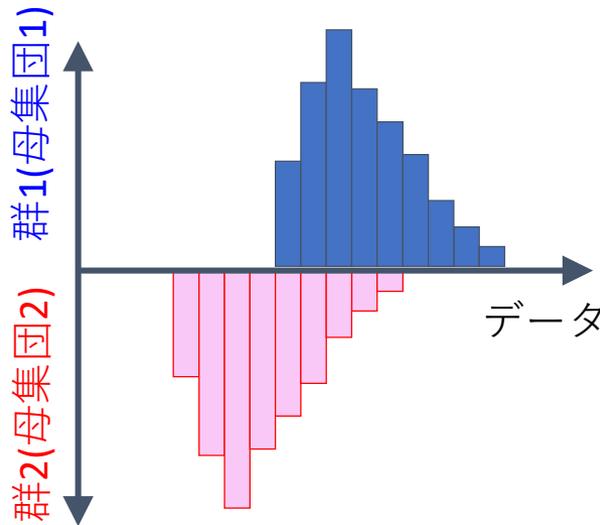
5.2 標本の検定

相対的位置関係には様々な状況が考えられる。ここでは母集団1が母集団2よりも大きな状況を考え、ヒストグラムで表現する。

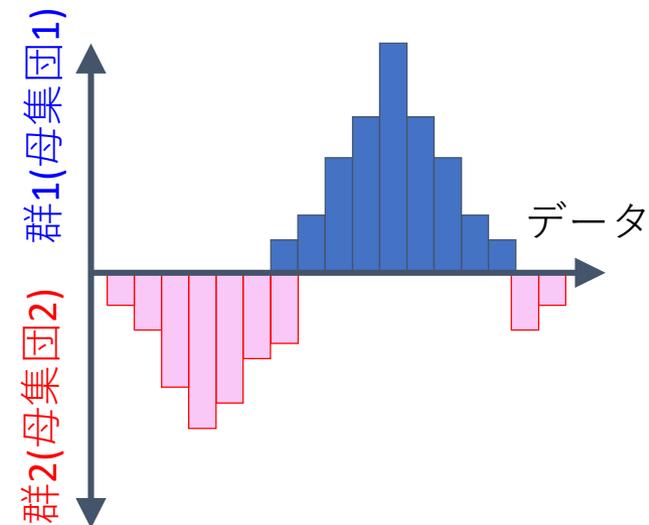
両群が対称な場合
(t検定による検定可能)



母集団が歪んでいる場合



外れ値が存在する場合

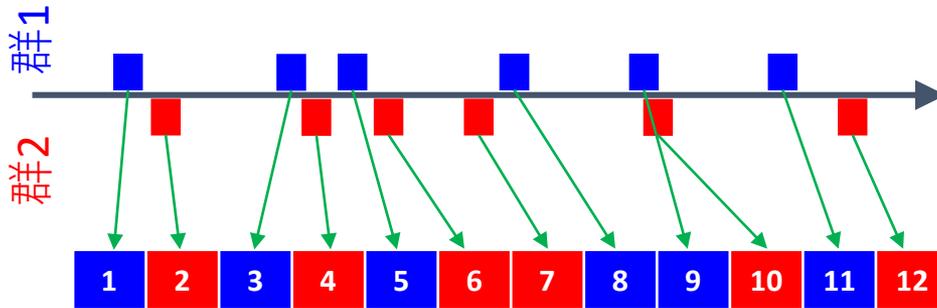


Wilcoxon検定では様々な母集団分布の状況で適用可能である。一方で「何を評価しているのかよくわからない」という状況になるため、t検定が使えない状況での代替と考えるべきである。

5.2 標本の検定

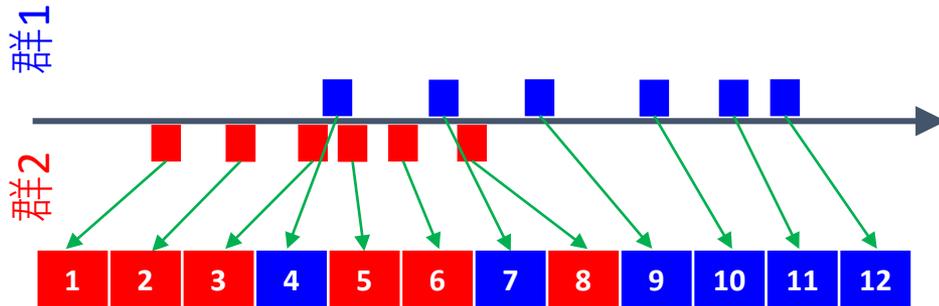
Wilcoxon検定で評価していること

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之和に基づいて検定を行う。そのため、外れ値や分布の歪みの影響を受けることはない。

5.2標本の検定

神経障害性疼痛患者を対象に、2種類の除痛薬(新薬, 既存薬)の治療前後でのVAS (mm)の減少量を試験した。

H_0 : 新薬と既存薬でVASの減少量(の分布の相対的位置関係)に差がない。

H_1 : 新薬と既存薬でVASの減少量(の分布の相対的位置関係)に差がある。

JMP : 「分析」 → 「二変量の関係」 → ○○の一元配置の横の「▼」をクリック
→ 「ノンパラメトリック」 → 「Wilcoxon検定」を選択

Wilcoxon/Kruskal-Wallisの検定(順位和)						
水準	度数	スコア和	スコアの期待値	スコア平均	(平均-平均0)/標準偏差0	
A	14	233.500	189.000	16.6786	2.273	
P	12	117.500	162.000	9.7917	-2.273	

2標本検定(正規近似)			
S	Z	p値(Prob> Z)	
117.5	-2.27324	0.0230*	

一元配置検定(カイ2乗近似)			
カイ2乗	自由度	p値(Prob>ChiSq)	
5.2857	1	0.0215*	

Wilcoxon検定では、p値の計算に近似を用いる場合と正確法を用いる場合の2種類がある。

→ p値の計算に正規近似を用いた場合

→ p値の計算にカイ2乗近似を用いた場合

JMPでは近似の方法に正規近似とカイ2乗近似の両方が存在する。解釈は、いずれの場合でも大差がない。

5.2標本の検定

神経障害性疼痛患者を対象に、2種類の除痛薬(新薬, 既存薬)の治療前後でのVAS (mm)の減少量を試験した。

JMP: 「分析」 → 「二変量の関係」 → ○○の一元配置の横の「▼」をクリック
→ 「ノンパラメトリック」 → 「正確検定」 → 「正確Wilcoxon検定」を選択

近似での結果

2標本検定(正規近似)		
S	Z	p値(Prob> Z)
117.5	-2.27324	0.0230*

一元配置検定(カイ2乗近似)		
カイ2乗	自由度	p値(Prob>ChiSq)
5.2857	1	0.0215*

正規近似でのp値=0.0230

カイ2乗近似でのp値=0.0215

正確法での結果

二標本: 正確検定		
S	p値(Prob≤S)	p値(Prob≥ S-Mean)
117.5	0.0102*	0.0203*

正確法でのp値=0.0203

片側対立仮説

両側対立仮説

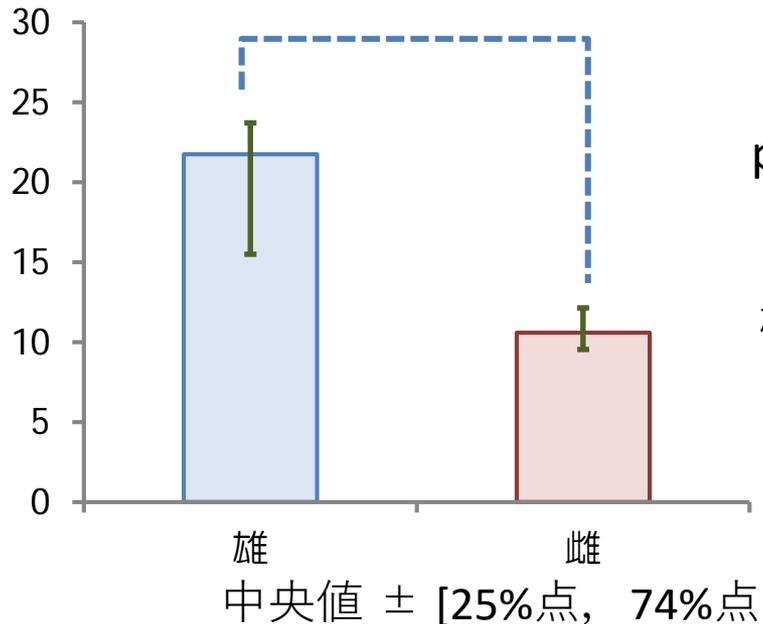
両側対立仮説におけるp値が0.0203なので、有意水準 $\alpha=0.05$ よりも小さい。つまり、帰無仮説が棄却され、対立仮説が支持された(有意だった)。したがって、新薬と既存薬には差があることが示された。

5.2標本の検定

少数例では正確なp値を利用する

雄	雌
32.3	13.9
27.2	13.5
23.8	11.7
23.4	10.7
22.6	10.5
20.9	9.7
18.2	9.1
14.6	6.2
13.4	
12.9	

Raisman & Field(1971)は、雄雌のラットの視索前野と腹内側核からの標本について、樹上突起上のシナプスの相対分布を計量的に求める研究を行った。ここでは、彼らが得た視索前野の神経網におけるシナプスからのデータを用いて、視索前野の神経網に形態的雄雌差が認められるかどうかを調べた。以下のデータを用いて、性別による違いを検定せよ。



漸近正規分布 : p値=0.0016

正確な p 値 : p値=0.0005

p値に若干の差異が認められる



標本サイズが少ないため正確なp値を用いた方が良い

5.2標本の検定

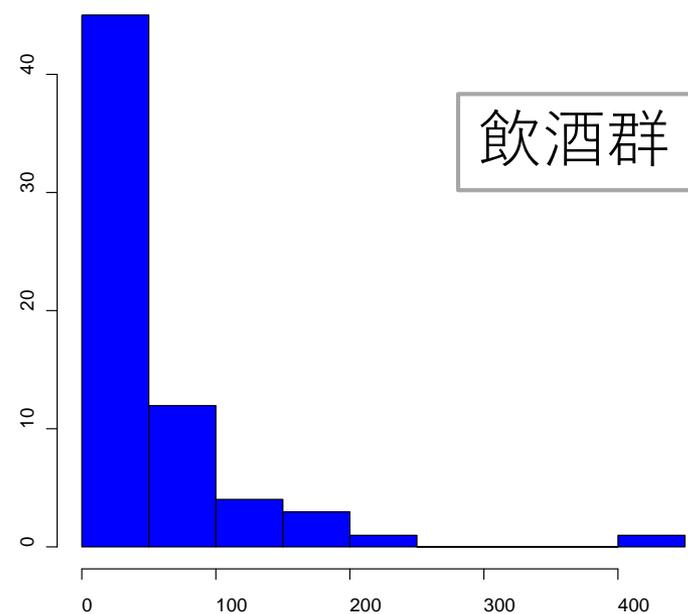
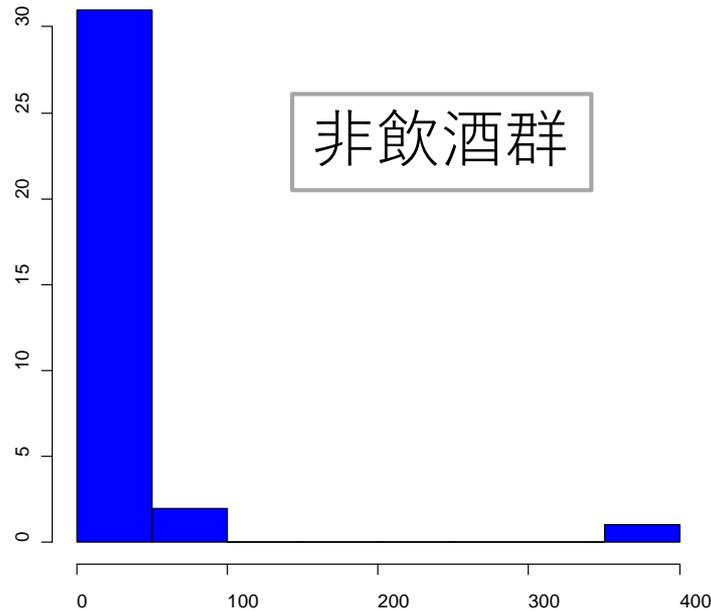
非正規分布での検定

病院Aの正常値を検討するために、病院職員を対象として生化学検査を測定した。男性98名について、飲酒群、非飲酒群に分けたところ下表の通りであった(丹後,2013)。飲酒の有無によって γ -GTPに差が認められるだろうか

非飲酒群			飲酒群					
22	11	13	23	15	116	87	18	95
13	21	19	52	17	87	22	67	12
0	3	0	14	33	238	33	18	16
16	16	18	32	24	103	20	25	21
22	27	22	88	55	0	29	106	34
23	3	18	98	48	55	0	14	22
44	45	17	190	32	66	35	45	
27	31	25	19	413	16	0	33	
19	30	16	37	13	17	102	33	
52	376	19	25	36	12	0	57	
27	68		174	40	83	21	18	
18	0		8	47	31	0	155	

5.2 標本の検定

非正規分布での検定の結果



	p 値
t 検定	0.1185
Welch検定	0.1185
Wilcoxon検定	0.0063

データが正規分布に従わない状況でt検定やWelch検定を用いるのは誤りである。この事例においても有意な結果が得られない。他方、Wilcoxon検定では、有意な結果が得られている。

6.おわりに

ご清聴ありがとうございました

