

医学統計セミナー ベーシックコース  
量的データの解析

下川 敏雄

和歌山県立医科大学 臨床研究センター

2016年度 医学統計セミナー

■ ベーシック・コース

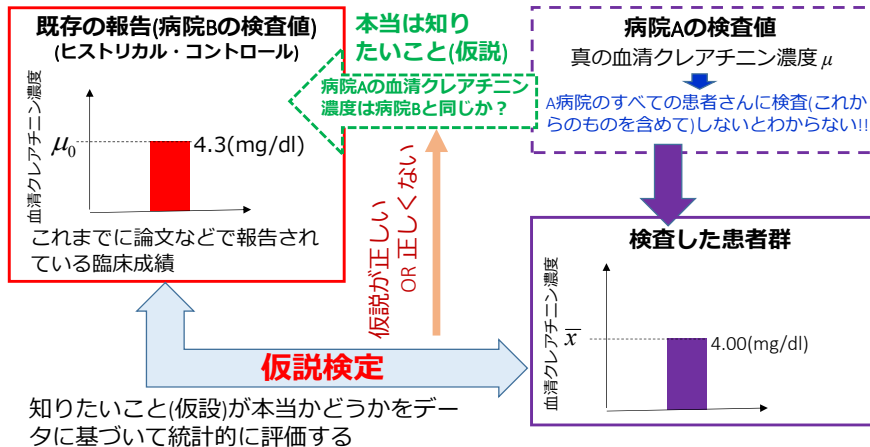
- ▶ 基礎統計学 (6月15日・住金棟5F 大研修室)
- ▶ 量的データの解析 (7月27日・住金棟5F 大研修室)
- ▶ 質的データの解析 (8月24日・住金棟5F 大研修室)
- ▶ 共変量調整を伴う解析 (11月2日・病院棟4F 臨床講堂1)
- ▶ 生存時間・臨床検査データの解析(11月16日・住金棟5F 大研修室)

■ アドバンス・コース

- ▶ 多群・経時データの解析と多重比較 (11月30日・病院棟4F 臨床講堂1)
- ▶ 臨床試験における症例数設定とガイドライン (12月28日・住金棟5F 大研修室)
- ▶ アンケート調査データの解析 (2月1日・病院棟4F 臨床講堂1)
- ▶ 統計的因果推論と傾向スコア (2月22日・住金棟5F 大研修室)
- ▶ メタアナリシス (3月22日・病院棟4F 臨床講堂1)

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。



帰無仮説と対立仮説

病院Aの血清クレアチニン濃度の平均値 $\mu$ 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 $\mu$ ) = 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 $\mu$ )  $\neq$  4.3 (病院Bの値)  
である。「異なる」という状況は、研究によって違ってくるかもしれない。  
\*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。

## 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。

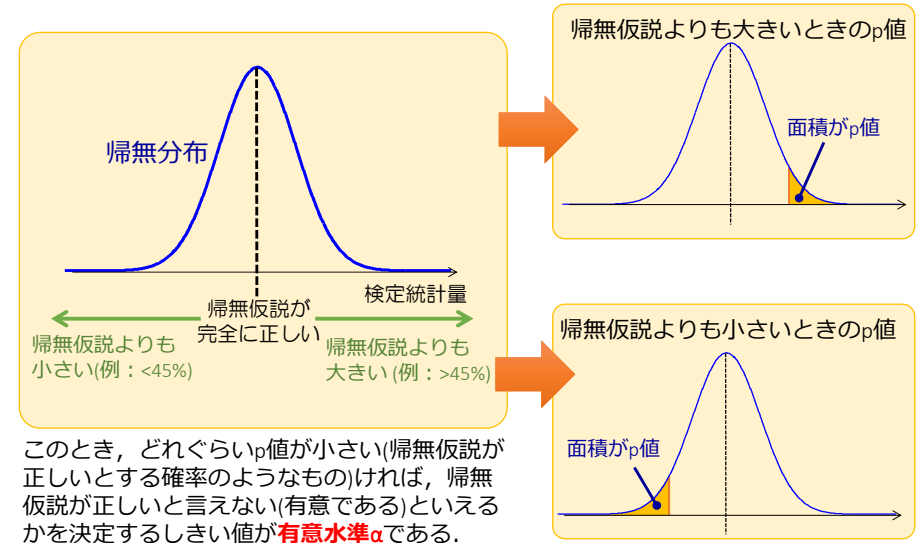
**p値が小さいとは、帰無仮説 $H_0$ が正しいとはいえないほどに差があることを意味する。つまり、対立仮説 $H_1$ が正しい状況にある。**

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。

## 帰無分布とp値の関係

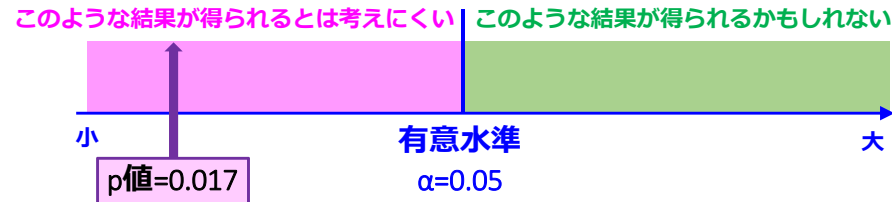


## どのように「有意である」と判定するのか？

**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、0.7%である。

有意水準 $\alpha$ 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、そうでないかを判断するカットオフ値のようなもの。



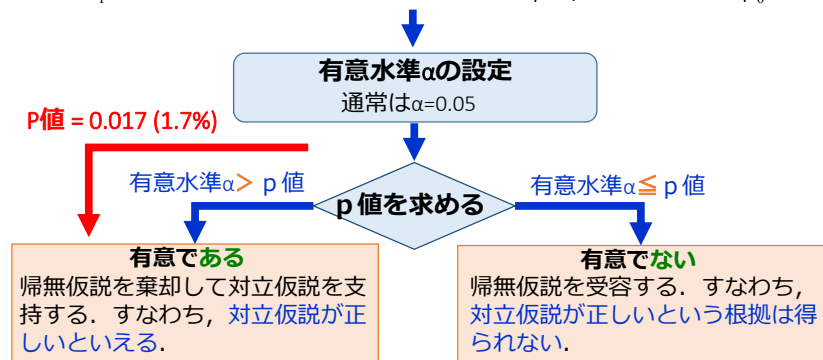
病院Aの平均血清クレアチニン濃度がB病院と同じ帰無仮説)とすると、本試験の結果が得られる確率\*1は1.7%しかない(有意水準5%未満である)。したがって、帰無仮説は誤っていると判断し(有意)、逆仮説である病院Aの血清クレアチニン濃度は病院Bの4.3(mg/dl)と異なる(対立仮説)と判断する。

## 仮説検定の流れを整理する

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 $\alpha$** という。

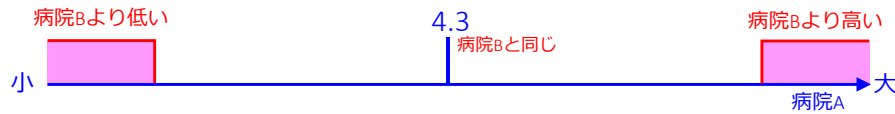
### 帰無仮説および対立仮説を設定

帰無仮説 $H_0$  : A病院の血清クレアチニン濃度の平均値 $\mu$ は、B病院の平均値 $\mu_0$ 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 $\mu$ は、B病院の平均値 $\mu_0$ と異なる。



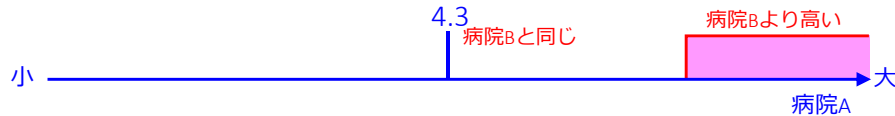
## 片側対立仮説と両側対立仮説

### ■ 両側対立仮説：新薬とプラセボの罹患率に違いがある。



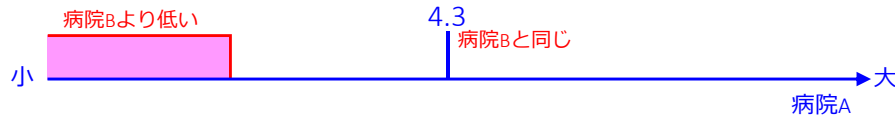
有意水準は「病院Bの4.3(mg/dl)よりも高い確率」「病院Bの4.3(mg/dl)よりも低い確率」の両方で設定されるので、片側では $\alpha/2$ になる。

### ■ 片側対立仮説(1)：新薬のほうがプラセボの罹患率よりも低い。



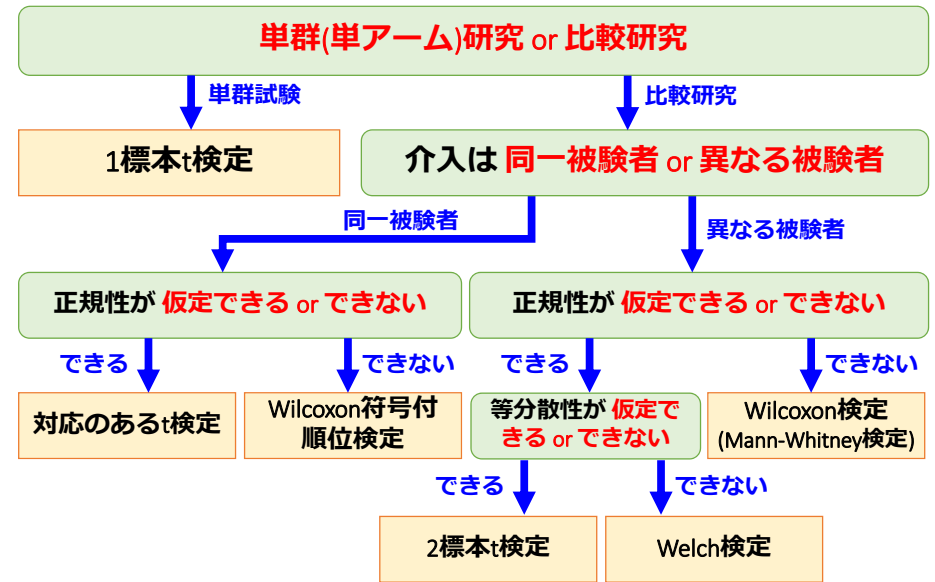
有意水準は「病院Bの4.3(mg/dl)よりも高い確率」で設定されるので、 $\alpha$ になる。

### ■ 片側対立仮説(2)：新薬のほうがプラセボの罹患率よりも高い。



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、 $\alpha$ になる。

## 量的データにおける基本的な仮説検定の種類

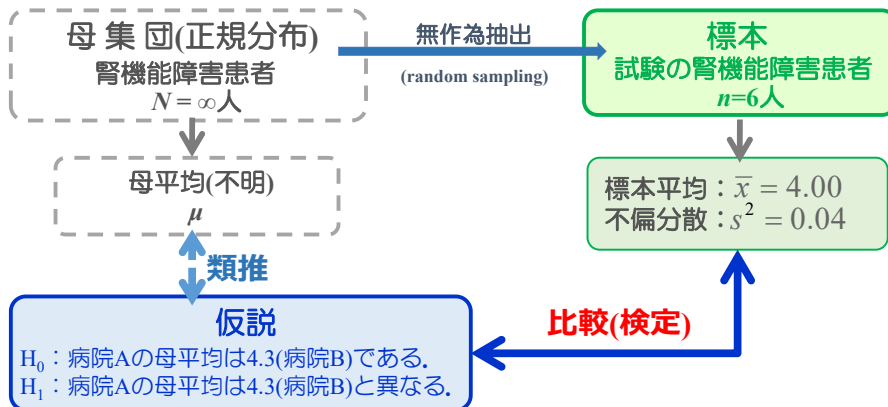


## 1標本t検定

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。



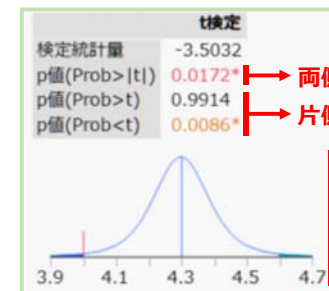
## 1標本t検定：JMPでの結果

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(mg/dl)と異なっていた。

## このような場合はどう考える？

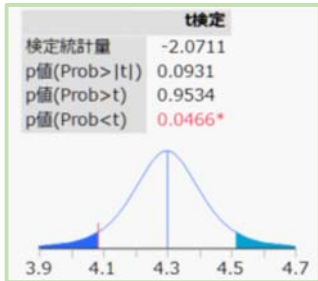
病院Cに入院する、腎機能障害の患者6名の血清クレアチニン濃度(mg/dl)を測定したところ

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： $\bar{x} = 4.05$

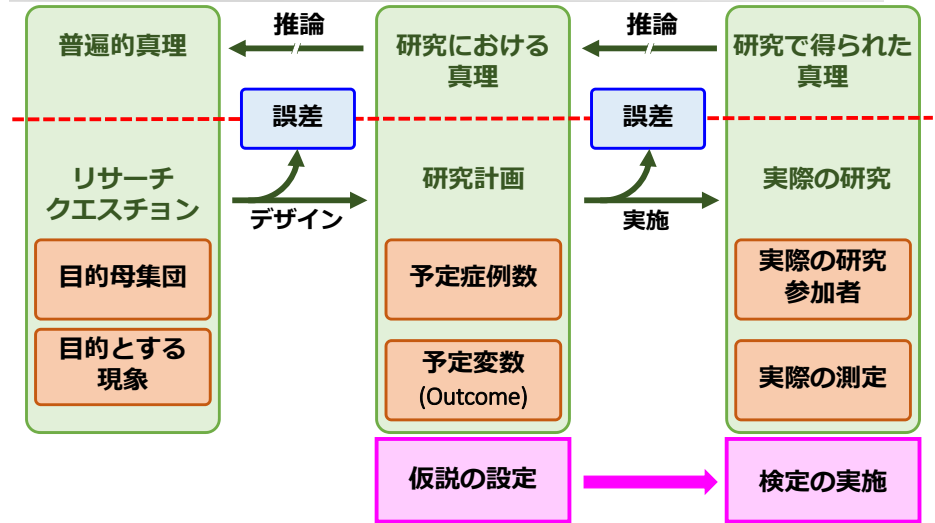
■病院Cの標準偏差： $s = 0.217$



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い。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？

## 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。

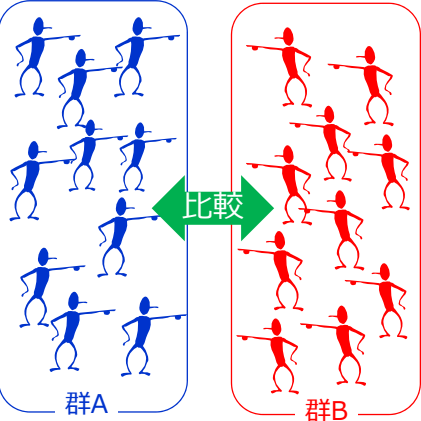
木原雅子・木原正博(訳)：医学的研究のデザイン 第4版，メディカル・サイエンス・インターナショナル，2014。

## 「対応がある／ない」とは何か？

### 対応のないデータ(独立2標本)

■例：無作為化比較試験

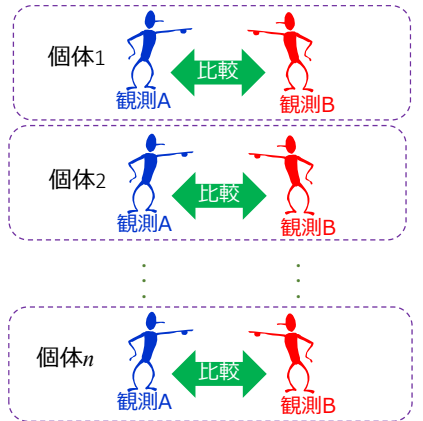
治療Aと治療Bを異なる被験者に実施して有効性を比較する。



### 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。



## 正規分布に従う場合の対応のあるデータの解析：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  
という  
良いか

## 対応のあるt検定の動機

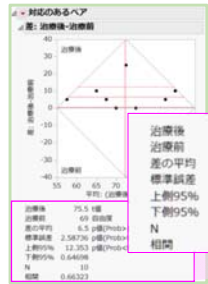
各個体(被験者)に対して、「治療後 - 治療前」を計算し、その平均変化量に対して、0であるか否かを検定する。変化量が正規分布に従うとするならば、帰無仮説での平均変化量 $\mu_0=0$ としたときの、1標本t検定と計算している内容は同じである。

$H_0$ : 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 $\mu$ 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 $\mu$ 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。

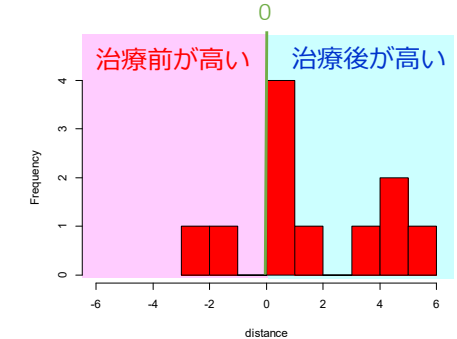
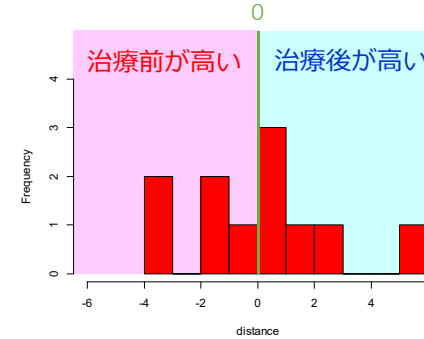
## 正規分布に従わない場合の対応のあるデータの解析: Wilcoxon符号付順位検定

有意でない状況

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

有意である状況

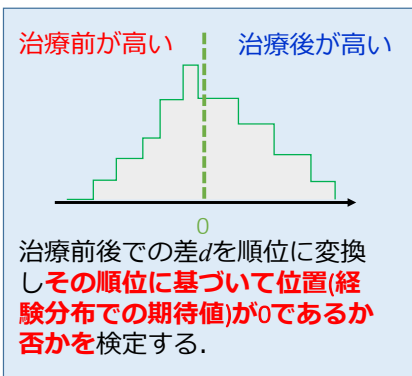
No	前	後	差	No	前	後	差
1	4	3	-1	7	2	1	-1
2	2	1	-1	8	6	2	-4
3	8	3	-5	9	3	4	1
4	5	4	-1	10	2	5	3
5	9	3	-6	11	6	1	-5
6	4	2	-2				



有意である状況では片側に多くのデータが存在する

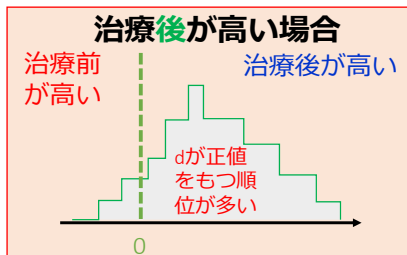
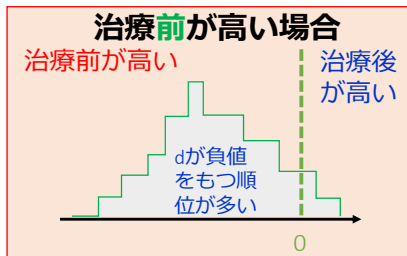
## Wilcoxon符号付順位検定における帰無仮説と対立仮説

### 帰無仮説



$H_0$ : 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。

### 対立仮説



## 正規分布に従わない場合の対応のあるデータの解析: Wilcoxon符号付順位検定

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

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

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



## Wilcoxon符号付順位検定：JMPでの結果

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



$H_1$ ：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\*

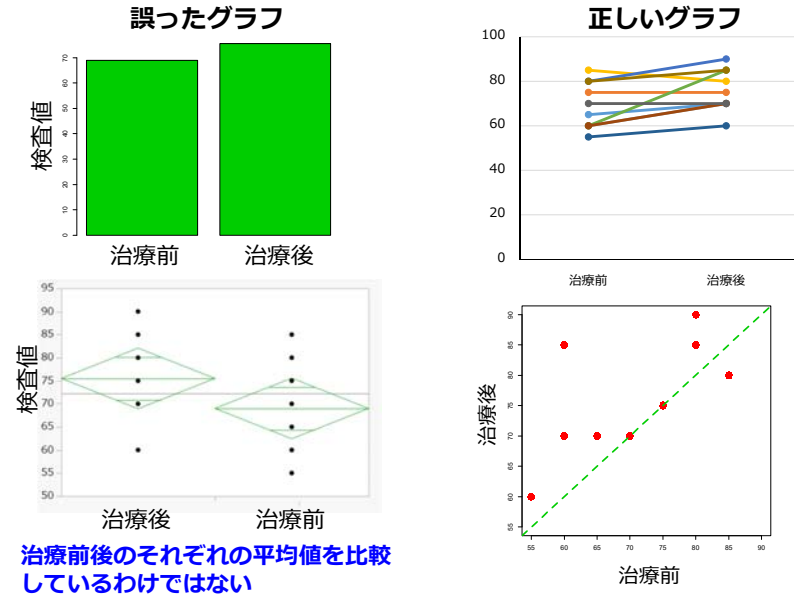
$H_1$ ：治療後のほうが治療前よりも小さい：p値(Prob<S) 0.9805

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。

## 対応のあるデータのグラフ表示の注意点



## 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・Welch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)

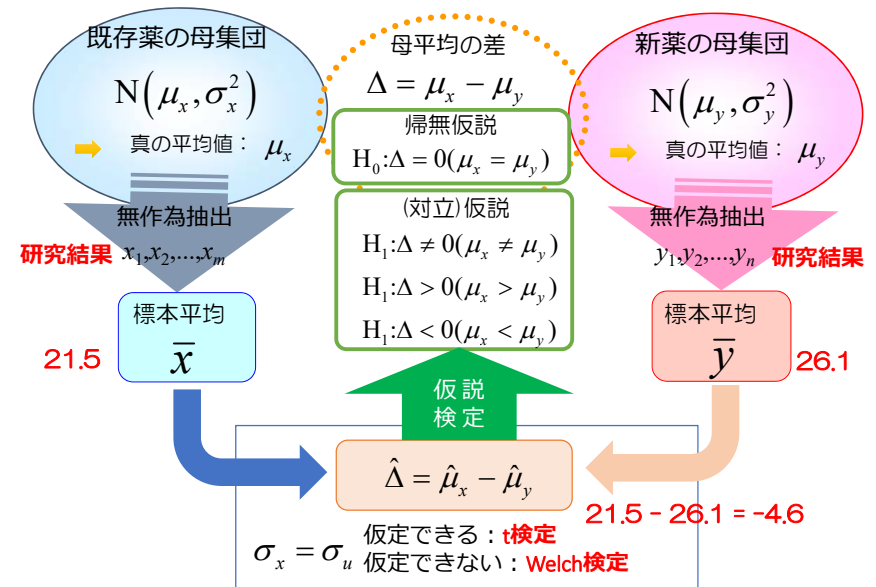
23 23 20 27 19 15 25 29 15 13 28 21

### 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。

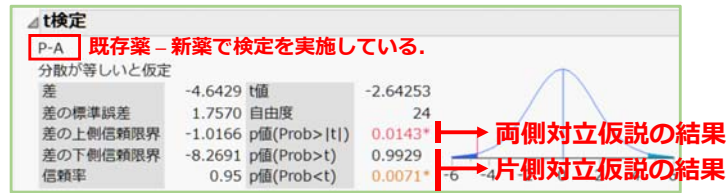
## 2標本t検定・Welch検定とは何か？



## 2標本t検定：JMPでの結果

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

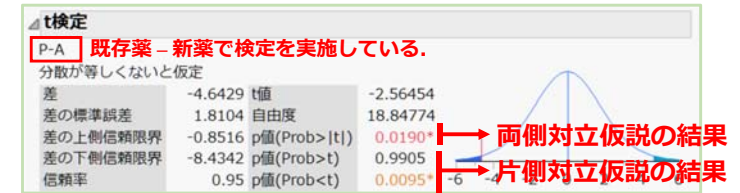


両側対立仮説におけるp値が0.0143なので、有意水準 $\alpha=0.05$ よりも小さい。つまり、帰無仮説が棄却され、対立仮説が支持された(有意だった)。したがって、新薬と既存薬には差があることが示された。

## Welch検定：JMPでの結果

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。

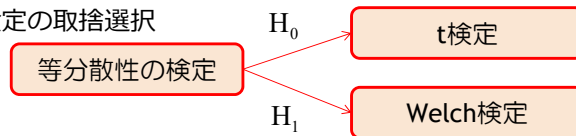
JMP：「分析」→「二変量の関係」→○○の一元配置の横の「▼」をクリック  
→「個々の分散を用いたt検定」を選択



両側対立仮説におけるp値が0.0190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。

## Welch検定に対する批判

次のような検定の取捨選択



は必ずしも適切であるとは限らない。等分散性を仮定するか否かは、以下のように決まる：

- (1) とるべきデータの性質に依るところが多いこと  
データをとり前のデザインの段階で、等分散性の仮定がおけるか否かは決定されるべきである(標本サイズもこれにより決まる)。
- (2) 当該の研究が目標とする「差」の意味、すなわち
  - (G1) 平均値の差が重要であり、バラツキは問題としない。  
Gans(1981)は、等分散性の検定は、平均値の差の検定の有意水準 $\alpha$ に対して、 $4\alpha$ でやるべきとの意見を述べている。これは、等分散の有無に関して、**明らかな差異が認められるとの結果がなければ、t検定でよい**ことを意味する。
  - (G2) 平均値の差が重要だが、バラツキも問題とする。  
この場合には、等分散性の結果、不等分散性が示唆されるならば、2群に差が認められると結論づけるべきである。

Behrens-Fisher問題(資料参照)があるため、Welch検定は避けるべき!!

## 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	p値(Prob>t)	0.0166*
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

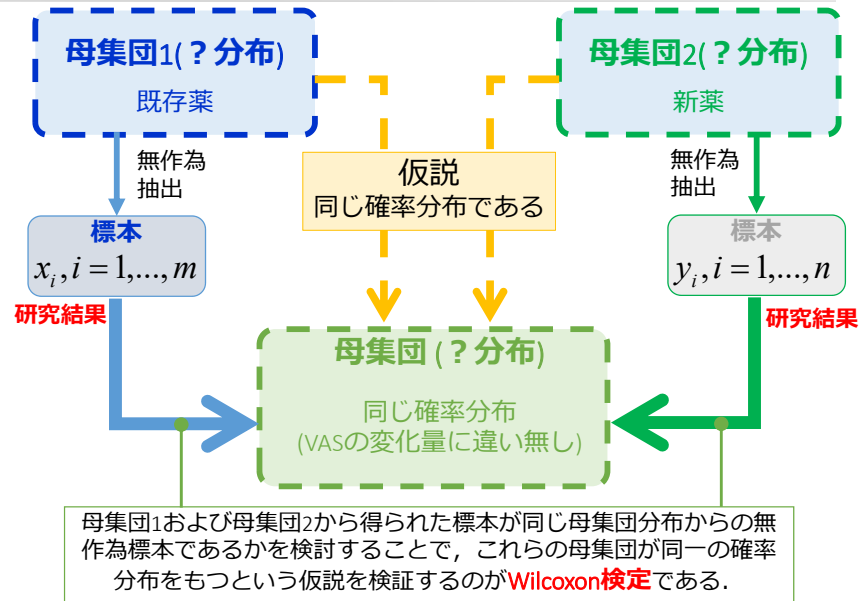
### 2標本t検定 (誤った解析)

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	p値(Prob<t)	0.0796

P値が0.0796で有意でない

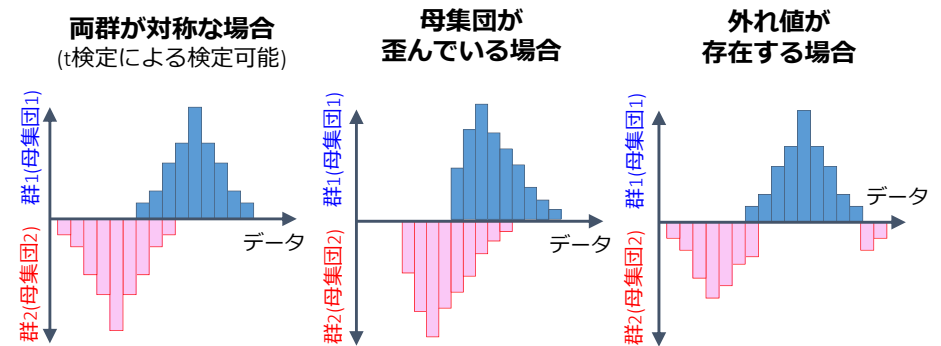
対応のあるデータを2標本t検定で検定する場合、分散(バラツキ)の影響によって、有意(positive)な結果を有意でない(negative)な結果にしてしまう恐れがある。

正規分布に従わない場合の2標本データの解析： Wilcoxon検定 (Mann-Whitney検定)



Wilcoxon検定では相対的な位置関係を評価する

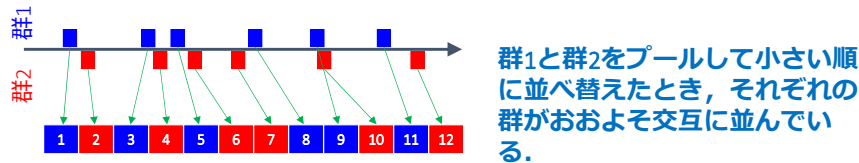
相対的位置関係には様々な状況が考えられる。ここでは母集団1が母集団2よりも大きな状況を考えて、ヒストグラムで表現する。



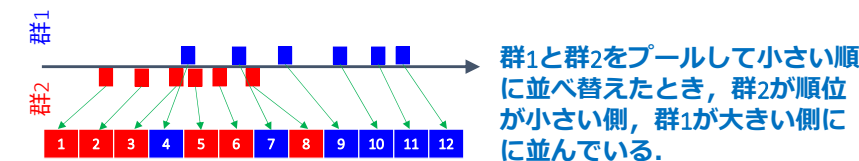
Wilcoxon検定では様々な母集団分布の状況で適用可能である。一方で「何を評価しているのかよくわからない」という状況になるため、t検定が使えない状況での代替と考えるべきである。

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が異なる場合(有意である場合)



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と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。

Wilcoxon検定：JMPでの結果(近似を用いてp値を計算した場合)

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*	→ p値の計算に正規近似を用いた場合

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*	→ p値の計算にカイ2乗近似を用いた場合

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。

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。



## Wilcoxon検定：JMPでの結果(正確法を用いてp値を計算した場合)

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

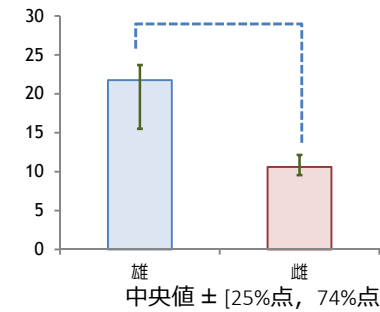
片側対立仮説 両側対立仮説

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。

## 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

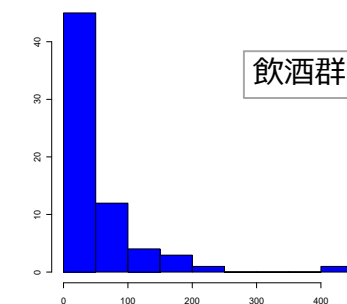
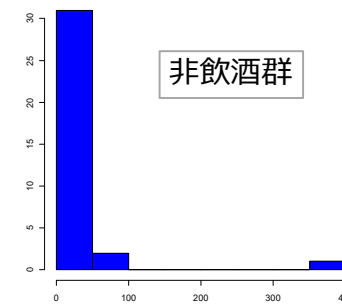
## 非正規分布での検定

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 $\gamma$ -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	

出典) 丹後敏郎: 医学への統計学 (第3版), 朝倉書店, 2013

## 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。