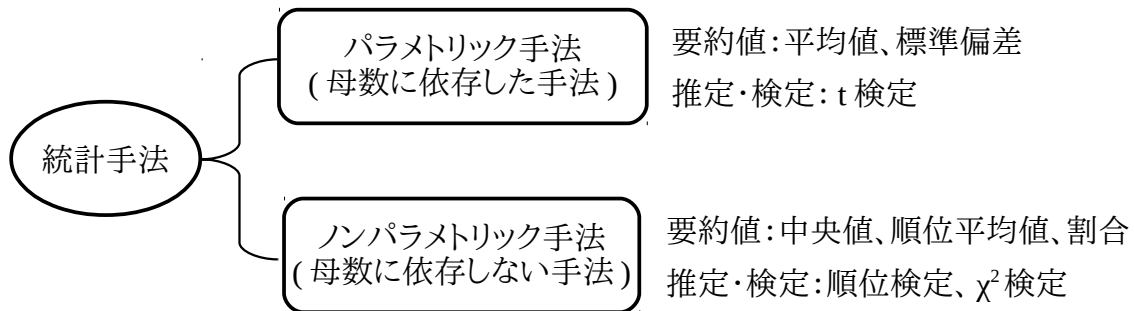


### 3. 統計学の落とし穴

#### 3.1 パラメトリック手法とノンパラメトリック手法

「データが正規分布しない時はノンパラ手法を適用せよ」を鵜呑みにしてはいけない



##### <パラメトリック手法の特徴>

- ・母集団のデータが特定の分布に従うことを前提にしている。  
例: t 検定…母集団のデータは近似的に正規分布
- ・要約値がデータの分布状態を反映する→分布状態によって要約値の値が変化する。  
例: 平均値…全データの重心を表すため、飛び離れたデータも反映
- ・結果の精度が高い。
- ・結果の普遍化が容易。  
標本集団のデータを特定の分布をする母集団から抽出されたものと仮定して計算するため、結果の外挿がある程度可能で普遍化しやすい。

##### <ノンパラメトリック手法の特徴>

- ・母集団のデータがどんな分布をしていても良い。  
例: 順位検定…母集団のデータに順位さえ付けられれば適用可能
- ・要約値がデータの分布状態を反映しにくい  
分布状態によって要約値の値が変化しにくい。

### 3. 統計学の落とし穴

例: 中央値…全データの中央の値を表すため、

分布の両端のデータは反映しない

- ・結果の精度が低い。
- ・結果の普遍化は困難。

母集団について特定の仮定をせず、標本集団のデータに基づいて

計算するため、結果の外挿が不可能で結果を普遍化しにくい。

統計学の解説書などに、

データが正規分布しない時はノンパラメトリック手法を適用しなければならない

または

データ数が少ない時はノンパラメトリック手法を適用しなければならない

ということがよく書かれているが、これを鵜呑みにしてはいけない!

<理由1>

- ・中央値の場合

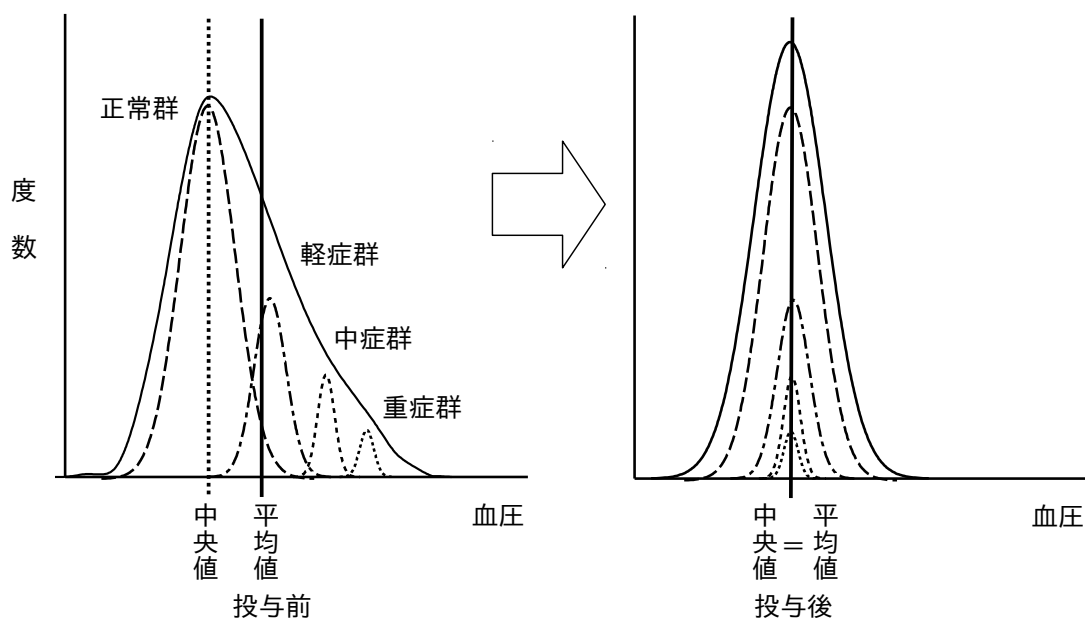
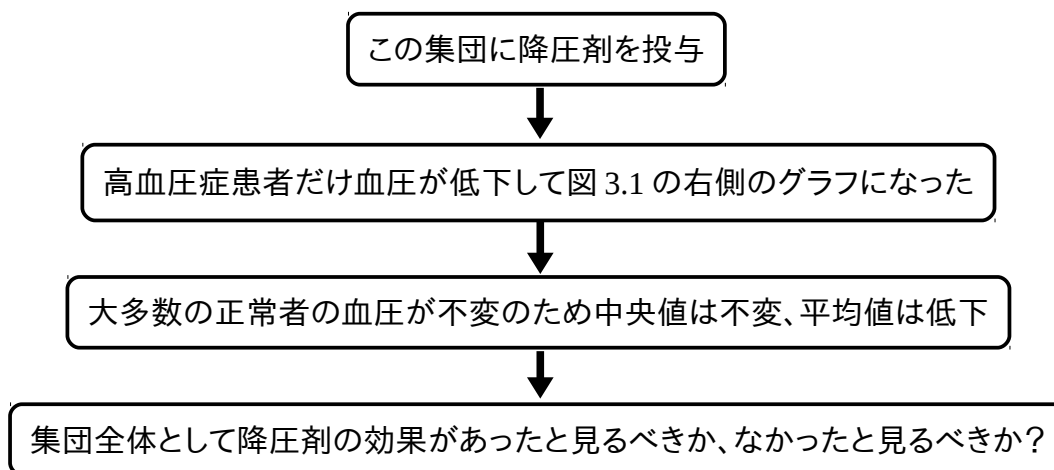


図 3.1 投与前後の血圧分布

図 3.1 の左のグラフのように、ある集団の中に血圧が正常範囲の正常者(大多数)と、血圧が少

### 3. 統計学の落とし穴

し高めの軽症高血圧患者(少数)、血圧がもう少し高めの中症高血圧患者(かなり少数)、血圧がかなり高めの重症高血圧患者(非常に少数)が含まれていたとする。



・数学者の見解…**中央値が変化していないから降圧剤の効果はなかつた。**

図 3.1 の左側の全体の分布を見ると、集団のデータは正規分布していない。このような場合は、分布状態に対して不変性——値が変わらない性質——がある中央値の方が平均値よりも数学的に意義があるから、要約値として中央値を採用。

・医学研究者の見解…**平均値が低下したから降圧剤の効果があつた。**

図 3.1 の左側の全体の分布を見ると、集団の中に血圧の高い人が少し含まれているようだ。このような場合は、分布状態の変化を反映する平均値の方が中央値よりも医学的に意義があるから、要約値として平均値を採用。

### 3. 統計学の落とし穴

<理由 2>

・順位検定の場合

データに順位を付けて、平均値の代わりに順位平均値を検定する

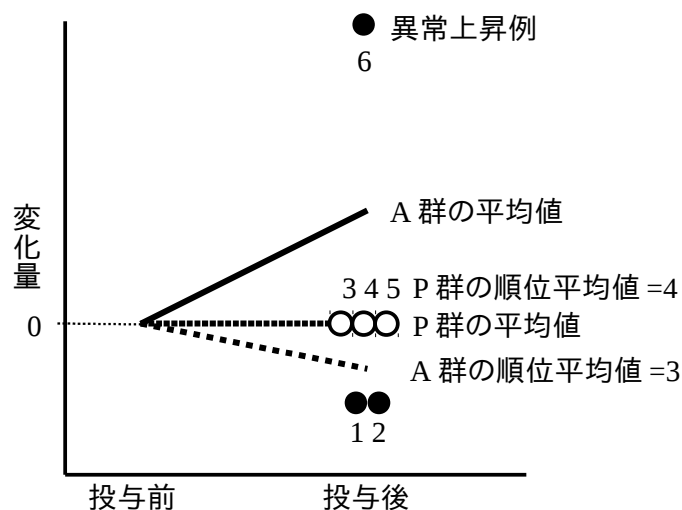


図 3.2 平均値と順位平均が矛盾する例

A 群に副作用による異常上昇例が発生した場合  
平均値を評価指標にした時: A 群は P 群よりも上昇したと評価  
順位平均値を評価指標にした時: A 群は P 群よりも低下したと評価



順位検定はデータの実測値には実質科学的な意味がなく  
データの順位や順位平均値だけが実質科学的な意味を持つ時に適用すべき手法

データの実測値ではなく順位だけが実質科学的な意味を持つデータは、最初から順位として定義されたデータ、つまり順序尺度のデータ以外には普通は有り得ない。

場合によっては、順序尺度のデータを計量尺度扱いした方が良いこともある。

### 3. 統計学の落とし穴

#### ○順序尺度のデータを計量尺度扱いした方が良い例

代表的な順位検定であるウィルコクソンの順位和検定は、外れ値を外れ値ではなくして解析するために開発された手法。そのため、次のような2種類の順序データに適用すると全く同じ結果になる。しかし、**医学的に見ると2種類の効果判定結果は明らかに違っている!**

＜薬剤の効果判定-1＞

群	著明改善	改善	不変	悪化	著明悪化	計
薬剤1投与群	0	40	40	0	0	80
薬剤2投与群	40	0	0	40	0	80
計	40	40	40	40	0	160

- ・ Wilcoxonの順位和検定(2標本検定、Mann-WhitneyのU検定)  
正規分布 $z=0$  有意確率 $p=1$

＜薬剤の効果判定-2＞

群	著明改善	改善	不変	悪化	著明悪化	計
薬剤1投与群	0	40	40	0	0	80
薬剤2投与群	40	0	0	0	40	80
計	40	40	40	0	40	160

- ・ Wilcoxonの順位和検定(2標本検定、Mann-WhitneyのU検定)  
正規分布 $z=0$  有意確率 $p=1$

著明改善を1、改善を2、不変を3、悪化を4、著明悪化を5と数量化し、対応のないt検定を適用すると次のような結果になる。**医学的に見ると、こちらの検定結果の方が合理的!**

#### ・薬剤の効果判定-1

薬剤1投与群: 平均値=2.5      標準偏差=0.5      標準誤差=0.06

薬剤2投与群: 平均値=2.5      標準偏差=1.5      標準誤差=0.17

対応のないt検定:  $t_0=0$        $p=1$

#### ・薬剤の効果判定-2

薬剤1投与群: 平均値=2.5      標準偏差=0.5      標準誤差=0.06

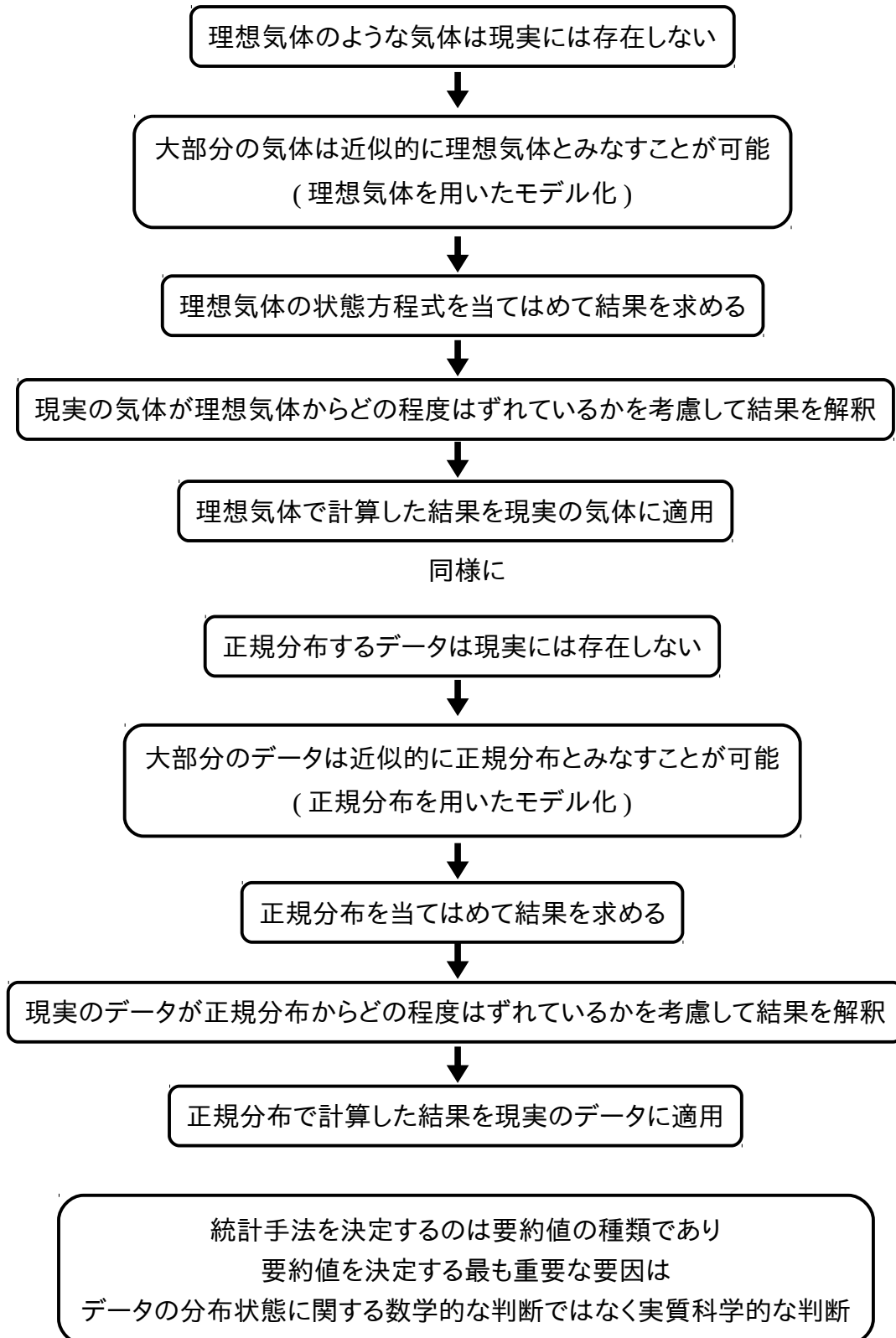
薬剤2投与群: 平均値=3.0      標準偏差=2.0      標準誤差=0.23

対応のないt検定:  $t_0=2$        $p=0.0338^*$

### 3. 統計学の落とし穴

#### <理由 3>

- ・正規分布は理想分布であり、厳密に正規分布するデータは現実には存在しない。



### 3. 統計学の落とし穴

#### 3.2 対数正規分布

データの分布が対数正規分布に似ているからといって無闇に対数変換をしてはいけない

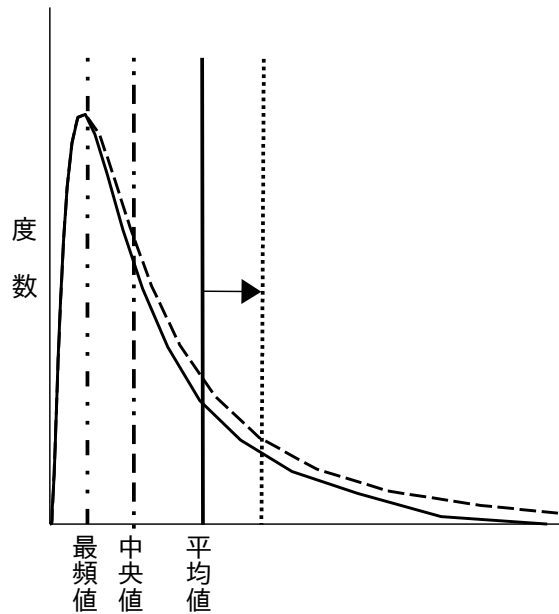
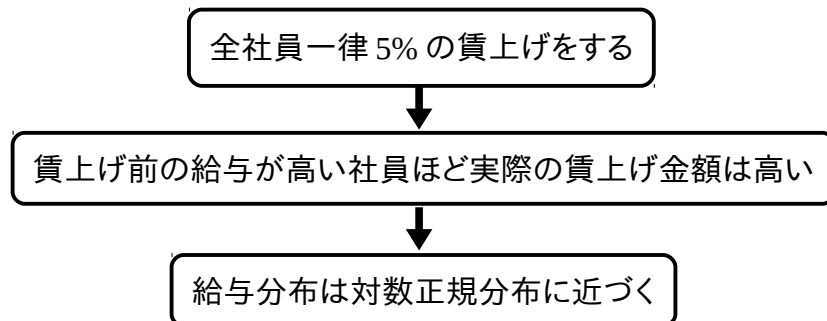


図 3.3 給与分布



・対数正規分布 (ジブラ分布) …対数変換したデータが正規分布をする分布

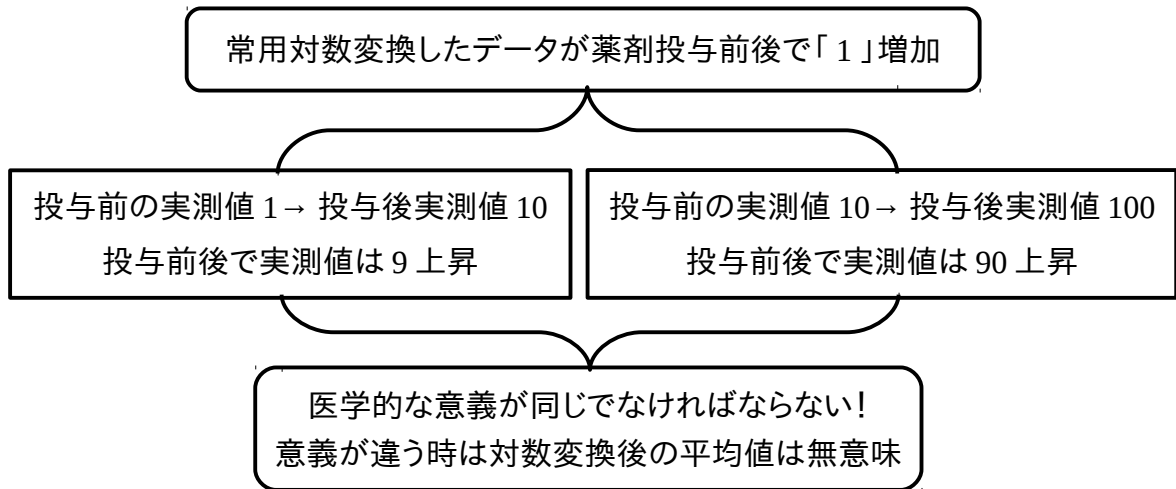
データ間に比例関係がある比例尺度のデータ。変動係数が一定。

対数変換した時の平均値を指数変換して元に戻すと、幾何平均値と一致する。

平均値、中央値、最頻値が一致しない。

通常、対数正規分布はデータを対数変換してから統計手法を適用する→医学・薬学分野で取り扱うデータは、分布の形が対数正規分布に似ているものが多い→その場合はデータを対数変換してから統計手法を適用する、という三段論法には注意が必要!

### 3. 統計学の落とし穴



医学分野で扱う集団は図 3.1 のように多数の正常者、少数の軽度の患者、かなり少数の中等度の患者、非常に少数の重症の患者が混ざっていることが多い→集団全体のデータは右裾が長くなり、見かけ上は対数正規分布のように見える。

このような集団に降圧剤を投与すると患者の血圧だけが低下し、投与後のデータは正規分布に近づくことが多い。

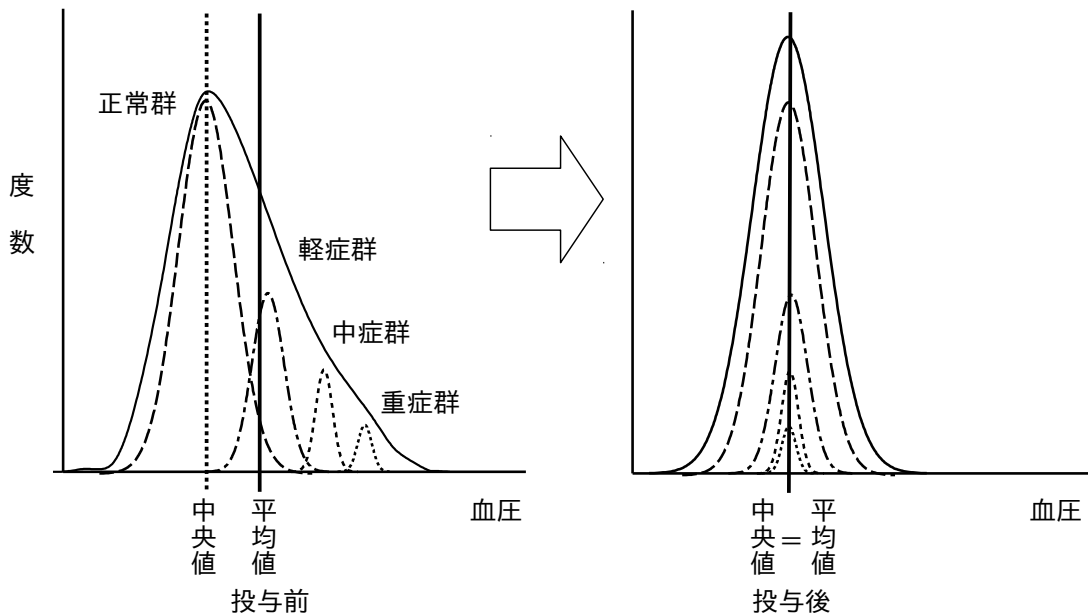


図 3.1 投与前後の血圧分布

データの分布が対数正規分布に似ているからといって  
実質科学的な意義を考慮せずに無闇に対数変換をしてはいけない



3.3 差と比の使い分け

データの差と比は実質科学的な意義を考慮して使い分ける

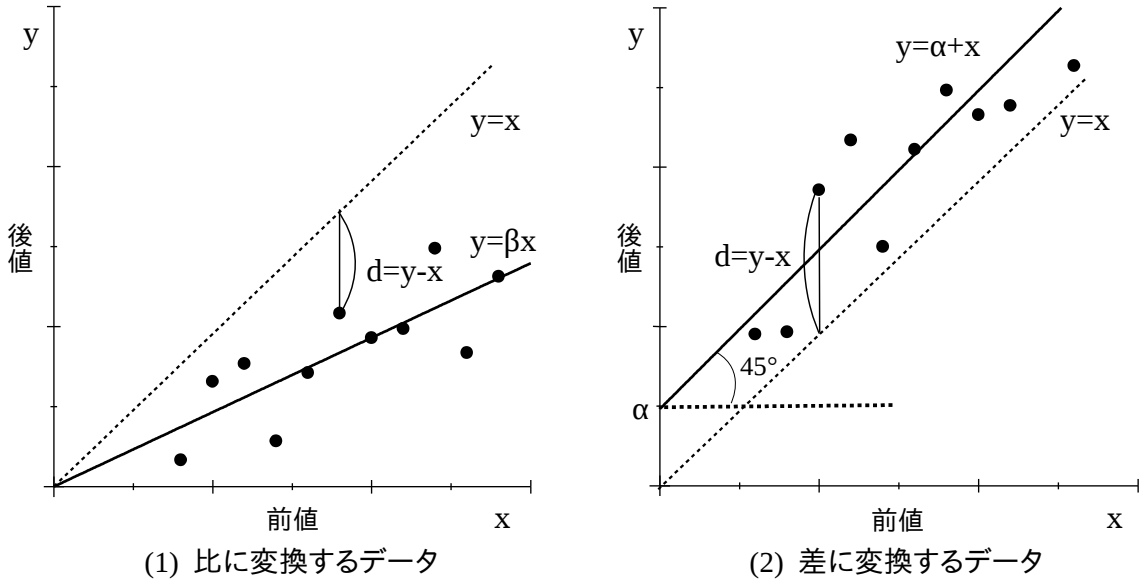


図 3.4 前値と後値の散布図

処置前後のデータは対応のある2標本データになり、普通は前後の差つまり変化量や、前後の比つまり前値に対する率または変化率に変換し、1標本にして統計処理する。

・**変化量(前後の差)**: 前値が後値に影響を与えないという暗黙の前提で計算

低下量 10: 前値が 10 の時も 100 の時も 10 低下するという意味

前値  $x$  と後値  $y$  の間に、

$$y = \alpha + \beta x$$

という直線関係がある時、変化量  $d$  は、

$$d = y - x = (\alpha + \beta)x - x = \alpha + (\beta - 1)x$$

より、 $\beta=1$ (直線の傾きが  $45^\circ$ )の時、 $d=\alpha$  となって、前値  $x$  とは無関係になる→図 3.4(2)の場合、間隔尺度的。

・**変化率(変化量の比)**: 前値と変化量が比例するという暗黙の前提で計算

低下率 10%: 前値が 10 の時は 1 低下し、100 の時は 10 低下するという意味

### 3. 統計学の落とし穴

値  $x$  と後値  $y$  の間に、

$$y = \alpha + \beta x$$

という直線関係がある時、変化率  $d(\%)$  は、

$$d(\%) = \frac{d}{x} \times 100 = \frac{\alpha + (\beta - 1)x}{x} \times 100 = \left\{ \frac{\alpha}{x} + (\beta - 1) \right\} \times 100$$

より、 $\alpha=0$ (直線が原点を通る)の時、 $d(\%)=(\beta-1)\times 100$  となって、前値  $x$  とは無関係になる→図

3.4(1)の場合、比例尺的。

前値と後値の比  $r$  は、

$$r = \frac{y}{x} = \frac{\alpha + \beta x}{x} = \frac{\alpha}{x} + \beta$$

$$d(\%) = \frac{y - x}{x} \times 100 = (r - 1) \times 100 = \left\{ \frac{\alpha}{x} + (\beta - 1) \right\} \times 100$$

となり、 $d(\%)$  は  $r$  に帰着する。

差・変化量または比・変化率に変換する時は  
前値と後値の関係を散布図で確認してから妥当な方を選択する  
どちらが妥当かわからない時は変化量の方が無難

※比と割合と率の定義

・比(ratio): お互いに相手を含まない別々の値を割ったもの

分子と分母の値の単位を組み合わせた次元を持ち、値に制限はない。

例: A/G 比(アルブミンをグロブリンで割った値)、BMI(体重を身長<sup>2</sup>で割った値)

・割合(proportion): 分子が分母に含まれる分数

次元を持たず、0~1 の間の値になる。

例: 有効率(有効例数を全例数で割った値)、有病率(疾病の患者数を全人口で割った値)

・率(rate): 単位あたりの変化量

単位の逆数の次元を持ち、値に制限はない。

例: 変化率(単位量あたりの変化量)、反応速度(単位時間あたりの反応量)

### 3. 統計学の落とし穴

#### 3.4 リスク差とリスク比の使い分け

リスク比は小さなリスクを顕微鏡で拡大して検討するような指標

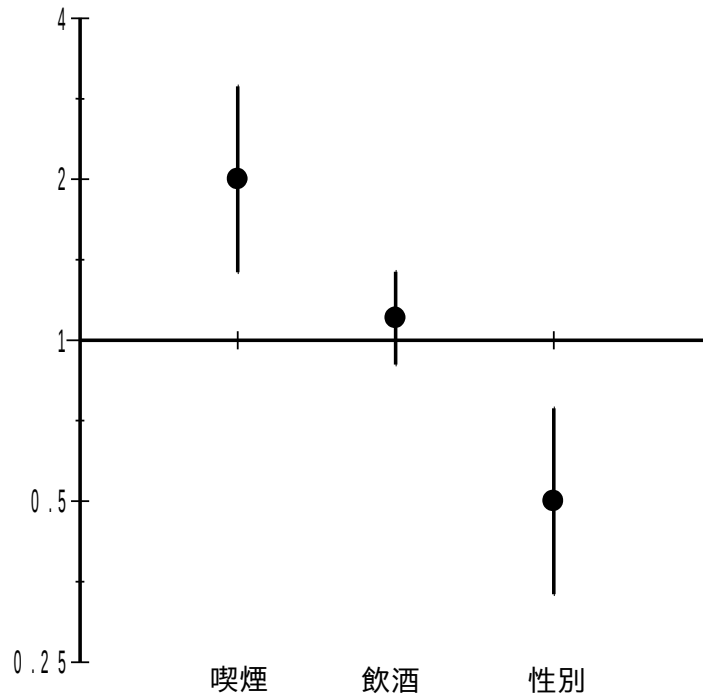


図 3.5 リスク比と 95% 信頼区間

・リスク差(絶対危険度) : 危険因子有群の疾患発症率  $p_+$  と危険因子無群の疾患発症率  $p_-$  の差

$$RD = p_+ - p_-$$

・リスク比(相対危険度) : 危険因子有群の疾患発症率  $p_+$  と危険因子無群の疾患発症率  $p_-$  の比

$$RR = \frac{p_+}{p_-}$$

・寄与リスク割合(寄与リスク) : 危険因子有群における危険因子による発症例の割合

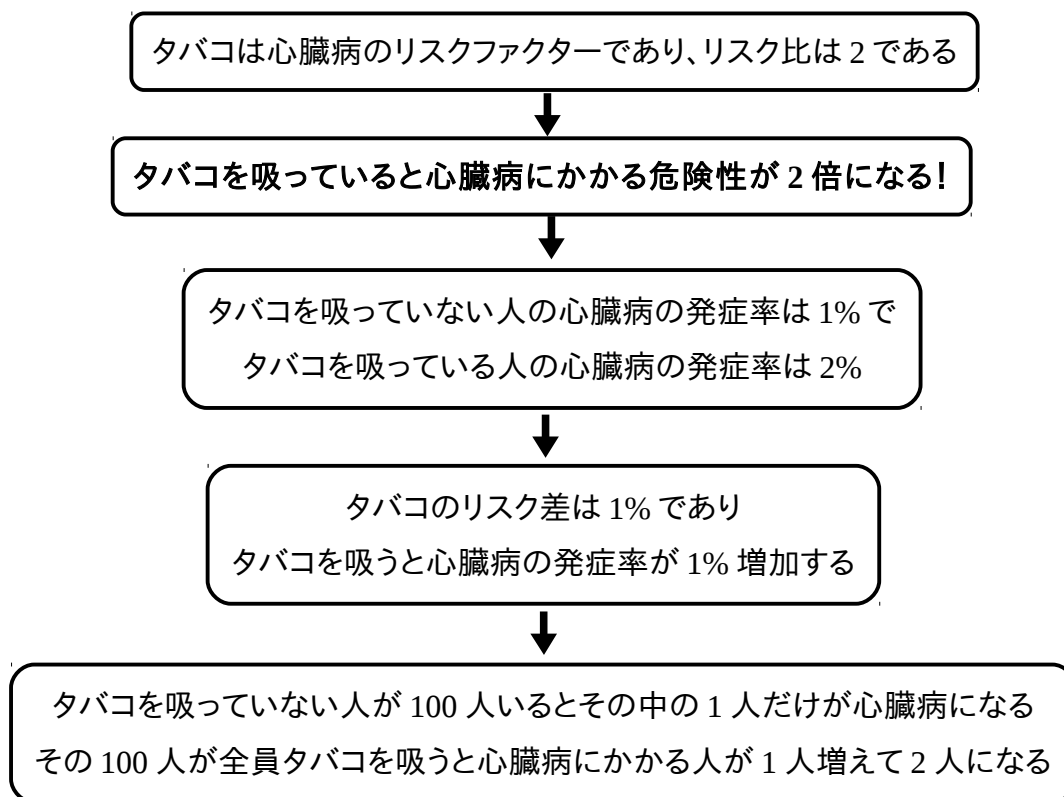
$$AR\% = \frac{p_+ - p_-}{p_+} \times 100 = \frac{RR - 1}{RR} \times 100$$

リスク比は発症率が高いと相対的に小さな値になり、発症率が小さいと相対的に大きな値になりやすい値→リスクの絶対値が小さい時に、リスクのわずかな差を顕微鏡で拡大して検討するような指標→リスクの絶対値を皮膚感覚で理解している医学研究者はリスク比を正しく評価でき

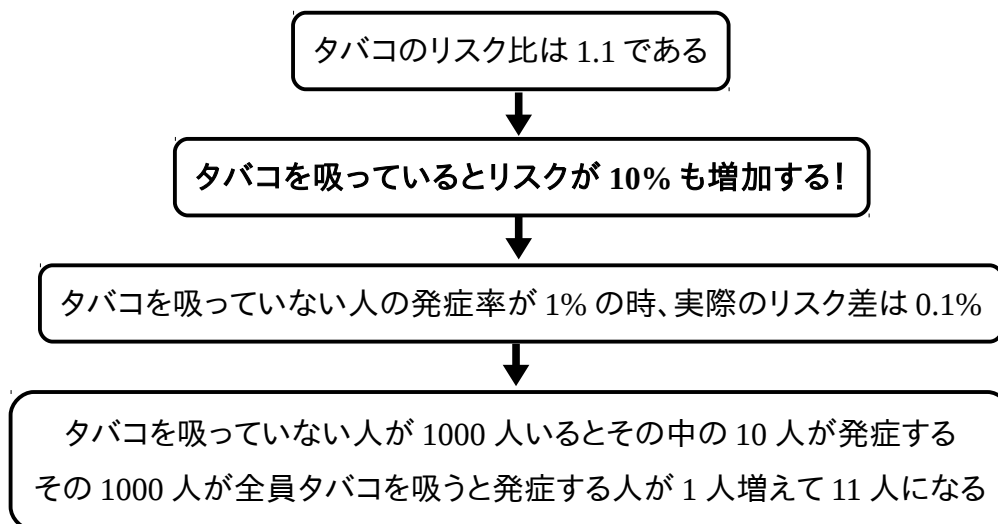
### 3. 統計学の落とし穴

るが、普通の人にはリスク比を正しく評価するのが難しいので注意が必要!

<例> 心臓病に対するタバコのリスク



リスク比が2未満の時、マスコミ報道等では寄与リスク割合の代わりにリスク比の増加率を全面に出す時があるので要注意!



リスク比が2.1の時は「リスクが110%も増加する!」ということになり、何となくウソ臭いので、普通はこの増加率を用いた表現はしない。

### 3. 統計学の落とし穴

#### 3.5 標準偏差と変動係数の使い分け

標準偏差と変動係数はデータの値とバラツキの関係を考慮して使い分ける

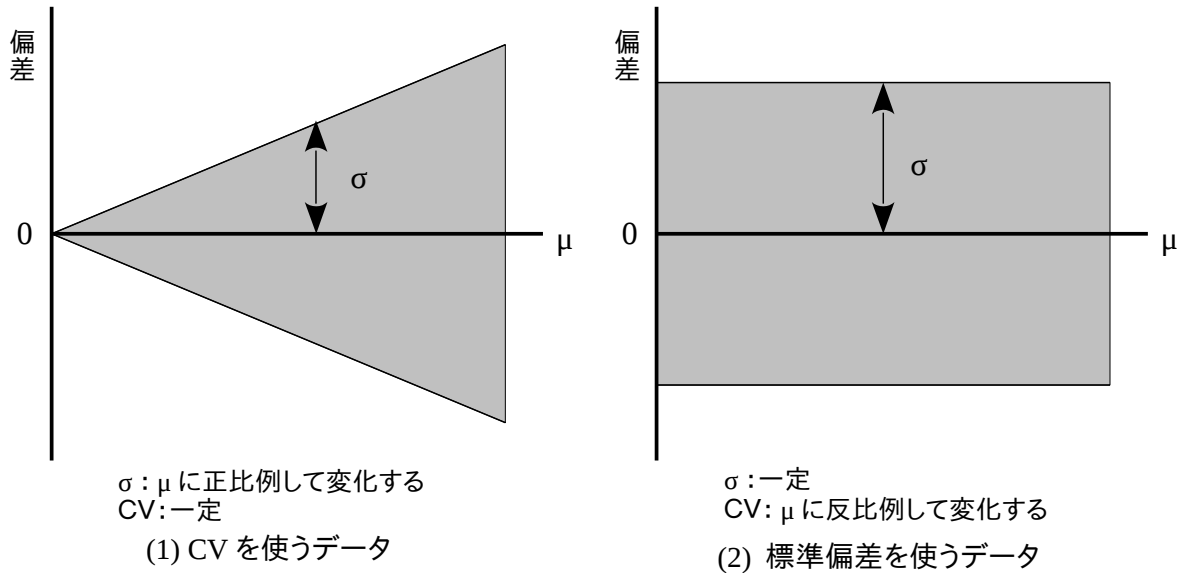


図 3.6 標準偏差と変動係数

・変動係数(CV:coefficient of variation)...平均値に対する標準偏差の率

$$CV = \frac{SD}{m}$$

男女 50 人について体重を測定し、平均値と標準偏差が次のようになった時、

男性:  $60 \pm 12\text{kg}$ (変動係数 20%)      女性:  $40 \pm 10\text{kg}$ (変動係数 25%)

データのバラツキがデータの値と比例する時  
女性の方がデータのバラツキが大きい



データのバラツキがデータの値と比例しない時  
男性の方がデータのバラツキが大きい

標準偏差と変動係数はデータの値とバラツキの関係を考慮して使い分ける  
どちらが妥当かわからない時は標準偏差の方が無難

### 3. 統計学の落とし穴

#### 3.6 尺度合わせ

尺度合わせは無闇にしない方が良い

全般改善度

薬剤	著明改善	中等度改善	軽度改善	不変	悪化	計	改善率
クダラン	8	12	45	35	0	100	65.0
プラセボ	0	25	25	35	15	100	50.0

全体: ウィルコクソンの順位和検定  $z_0=2.052$   $p=0.0402^*$

軽度改善以上を改善とした時の改善率:  $\chi^2$ 検定(2×2)  $\chi_0^2=4.010$   $p=0.0452^*$

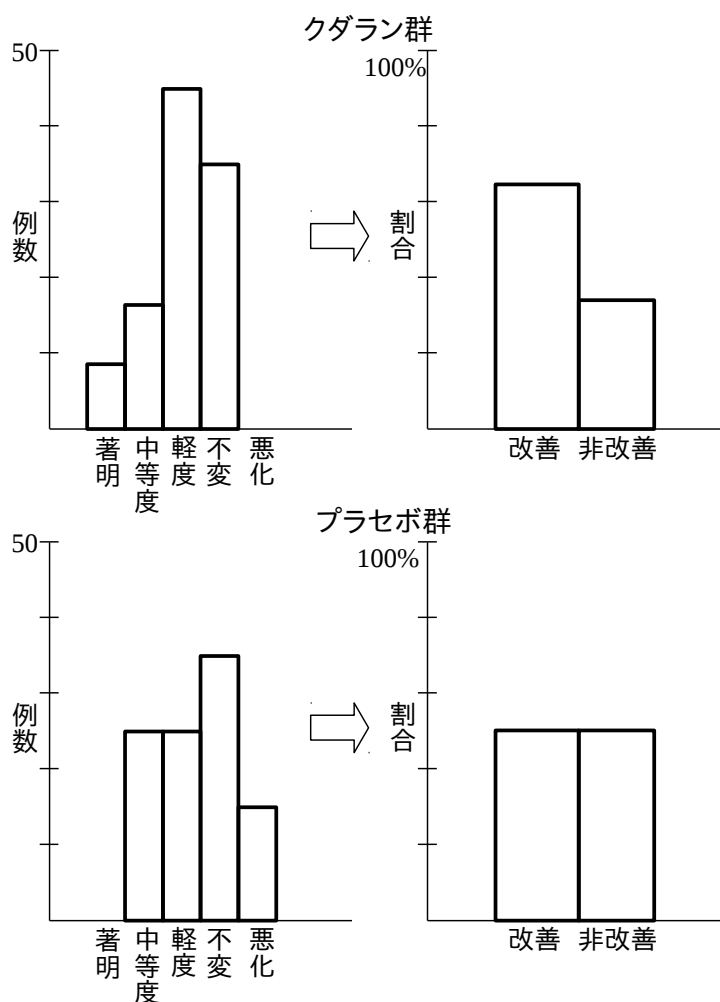
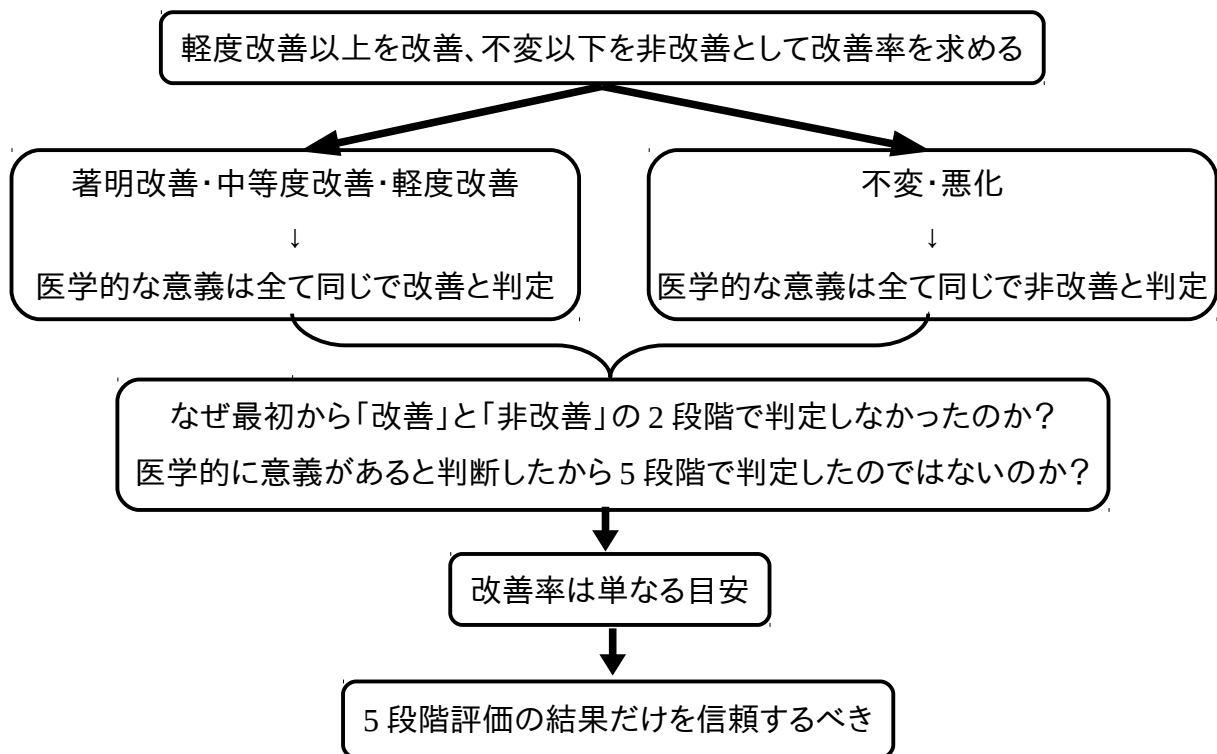


図 3.7 全般改善度の度数分布図

上表では「軽度改善以上を改善」とすることによって、順序尺度のデータを名義尺度のデータに変換している→尺度合わせの典型例

### 3. 統計学の落とし穴



改善率は単なる目安にすぎないのであまり信頼できない。その証拠に、中等度改善以上を改善とした時の改善率は結果が逆転する。

クダラン群の改善率: 20%      プラセボ群の改善率: 25%

レベルの高い尺度のデータほど多くの情報を持っている→その情報のある面だけを取り上げたのが尺度合わせ→別々の方法で尺度を合わせれば別々の結果になるのは当然。

・尺度合わせが実質科学的に意義を持つ場合

実験途中で計画段階には予測していなかった事態が発生し、どうしても評価基準を変えざるを得なくなった場合→その場合は元の評価基準による結果が実質科学的な意義を持たなくなる→尺度合わせをした結果だけを信頼すべき→このような場合、本来は新しい評価基準でもう一度実験をやり直すべきである。

原則として尺度合わせは行わず  
データが持っている情報を最大限有効に利用することが大切

### 3.7 外れ値

外れ値を簡単に除外してはいけない

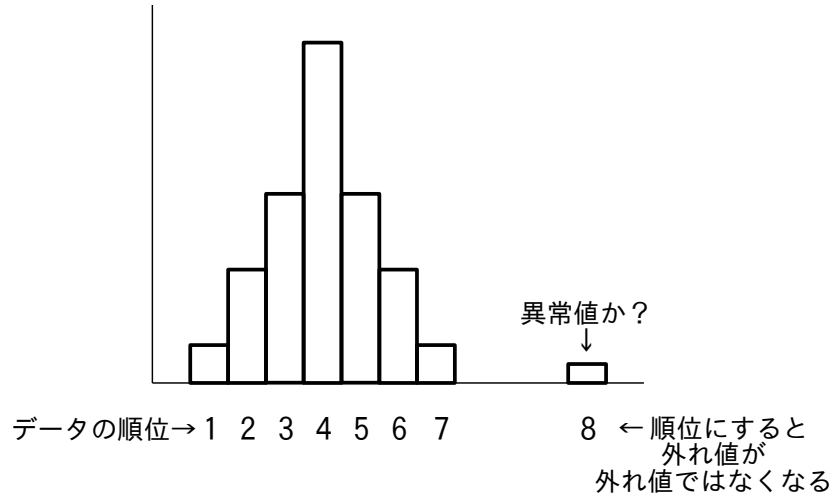


図 3.8 外れ値の例

#### ・外れ値(outlier)≡異常値の処理方法

- (1) 外れ値の原因を調べ、それに応じて除外するか除外しないかを検討する。
- (2) 外れ値を含めて統計解析を行った時の結果と、外れ値を除外して統計解析を行った結果を比較検討する。(感度分析)
- (3) そのまま統計解析を行う。
- (4) データに順位を付け、順位を用いて統計処理を行う。

※ウィルコクソンが順位和検定を開発したのは、外れ値の処理に困ったため。

- (5) 棄却検定によって外れ値を検定し、有意ならば除外して統計解析を行う。

薬剤の副作用は本質的に異常値→これを除外すると副作用のある薬剤はなくなってしまう。

過去の科学上の発見の多くは、外れ値について真剣に検討したからこそその結果→レントゲンによる X 線の発見、フレミングによるペニシリンの発見、パスツールによるワクチンの発見等々。

外れ値はまだ知られていない新しい現象かもしれないので  
無闇に除外せず原因を検討することが大切



### 3. 統計学の落とし穴

#### 3.8 ハンディキャップ方式の有意性検定

非劣性検定を同等性検定の代わりに使ってはいけない

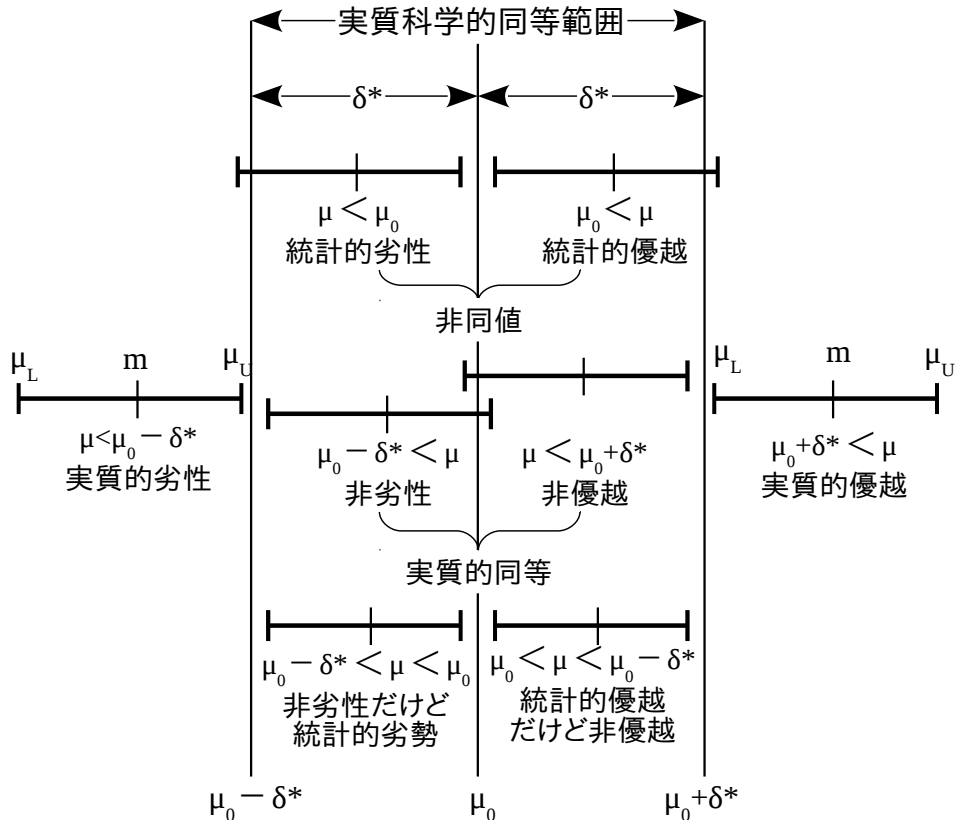


図 3.9 ハンディキャップ方式の有意性検定

・非同値検定… $\mu_0$ を基準値にした有意性検定

95%信頼区間の下限  $\mu_L$  が  $\mu_0$  よりも大きい(有意): 統計的優越=非同値

95%信頼区間の上限  $\mu_U$  が  $\mu_0$  よりも小さい(有意): 統計的劣性=非同値

・同等性検定… $\mu_0$ を基準値にし、95%信頼区間幅を  $\delta^*$ 以下にした統計的仮説検定

95%信頼区間の下限  $\mu_L$  が  $\mu_0$  よりも大きい(有意): 統計的優越=非同値

95%信頼区間の上限  $\mu_U$  が  $\mu_0$  よりも小さい(有意): 統計的劣性=非同値

95%信頼区間の中に基準値  $\mu_0$  が含まれる(有意ではない): 実質的同等

・優越性検定または非優越性検定… $\mu_0 + \delta^*$ を基準値にした有意性検定

95%信頼区間の下限  $\mu_L$  が  $\mu_0 + \delta^*$  よりも大きい(有意): 実質的優越

### 3. 統計学の落とし穴

95%信頼区間の上限  $\mu_U$  が  $\mu_0 + \delta^*$  よりも小さい(有意): 実質的非優越

・劣性検定または非劣性検定…  $\mu_0 - \delta^*$  を基準値にした有意性検定

95%信頼区間の上限  $\mu_U$  が  $\mu_0 - \delta^*$  よりも小さい(有意): 実質的劣性

95%信頼区間の下限  $\mu_L$  が  $\mu_0 - \delta^*$  よりも大きい(有意): 実質的非劣性

ハンディキャップ方式の有意性検定は有意性検定に統計的仮説検定の考え方を一部導入したもので、基本的に必要例数の計算(例数設計)をする必要はない。

薬業界には非同値検定のことを「優越性検定」と呼び、同等性検定の代わりに非劣性検定を行うという悪しき慣習がある。そして新薬と標準薬を比較して、新薬の有効性が非劣性検定で非劣性になり、安全性などが非同値検定で統計的優越つまり非同値になれば「新薬にはメリットがある」と結論する欺瞞が横行している。

図 3.9 を見れば、新薬が標準薬と実質的に同等でもこの結論が導けてしまうことがわかる。

そしてこのようにして許可された新薬を標準薬として、次の新薬の試験を行うという手順を繰り返すと、数回後には最初の標準薬と比較すると実質的に劣性な新薬が許可されてしまう!

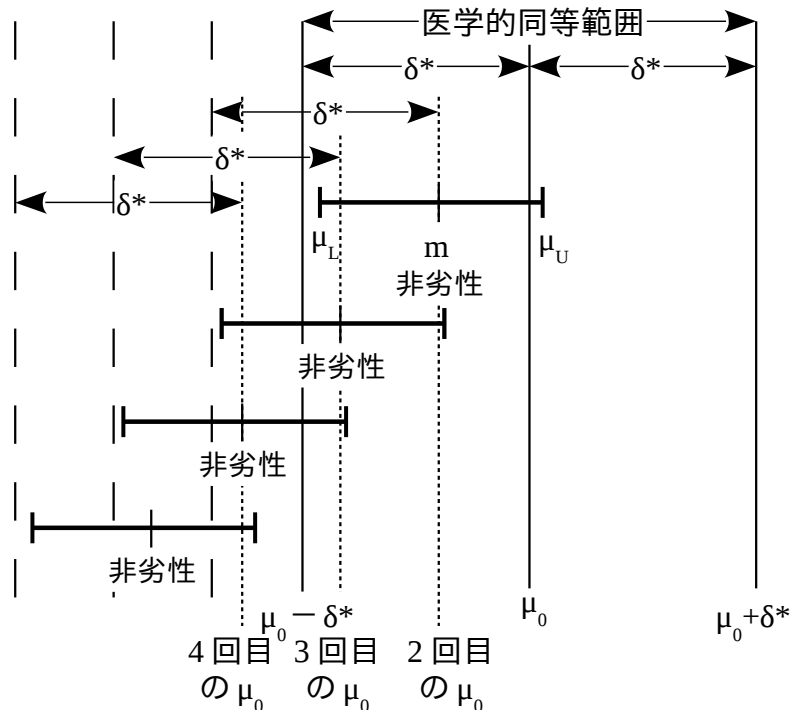


図 3.10 標準薬に対する非劣性検定の非合理性

### 3. 統計学の落とし穴

非劣性検定を同等性検定の代わりに使用してはいけないし  
非同値検定を優越性検定の代わりに使用してもいけない



非劣性検定は優越性検定とペアで使用しなければならない