

統計学の落とし穴

寺子屋・統計庵 其之二

杉本解析差有比数

杉本典夫

統計学の落とし穴

- 標準偏差と標準誤差
- 有意性検定と統計的仮説検定
- パラメトリック手法とノンパラメトリック手法
- ハンディキャップ方式の検定
- 相関分析と回帰分析
- 葬患係数と怪奇直線
- 選択例の相関

標準偏差はデータのバラツキ具合を表す要約値

偏差(バラツキの定義): $d_i = x_i - m$

$$\text{平方和: } SS = S_{xx} = \sum_{i=1}^n d_i^2 = \sum_{i=1}^n (x_i - m)^2$$

$$\text{分散: } V = \frac{SS}{n} = \frac{\sum d_i^2}{n} = \frac{\sum (x_i - m)^2}{n}$$

$$\text{標準偏差: } s = SD = \sqrt{V} = \sqrt{\frac{SS}{n}} = \sqrt{\frac{\sum d_i^2}{n}} = \sqrt{\frac{\sum (x_i - m)^2}{n}}$$

標準偏差の特徴

- データ1個あたりのバラツキ具合を表す
- 正規分布では平均値から分布の変曲点までの距離になる
- 正規分布では平均値±標準偏差の間に全データの約68%が含まれる
- 正規分布では平均値±2×標準偏差の間に全データの約95%が含まれる

標準誤差は推測統計学独特の要約値

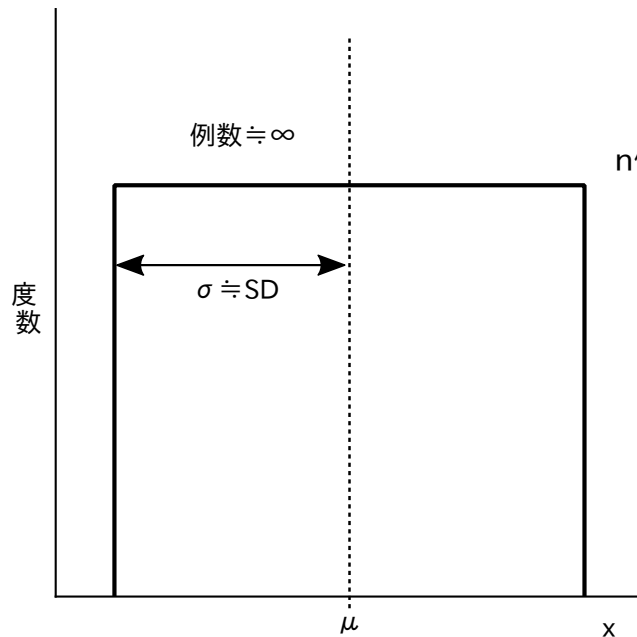


図1.3 母集団のデータ分布

n例を無作為抽出して
標本平均値mを
無限回求める

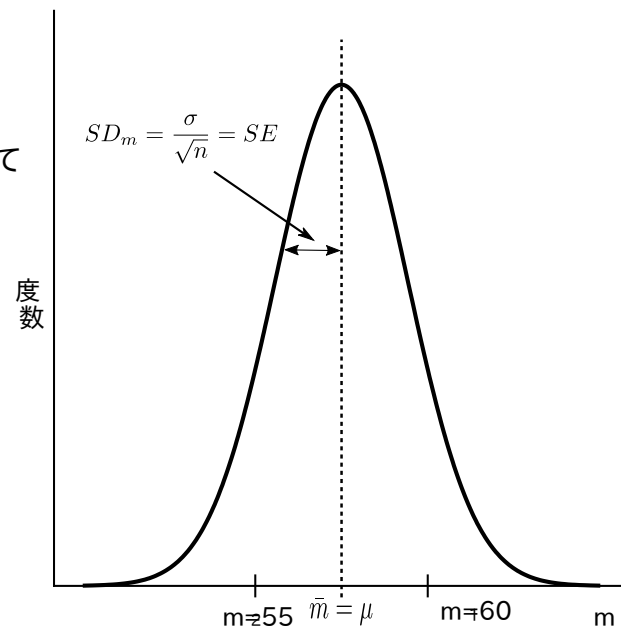
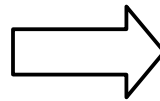


図1.4 標本平均値の分布

標準誤差の求め方

1. 母集団からn例の標本集団を無作為抽出する
2. 標本平均値を求め、それを m_1 として標本平均値の度数分布図にプロットする
3. n例の標本集団を母集団に戻す
4. 1番から3番を無限回繰り返す
5. 最終的に図1.4のような標本平均値の度数分布ができあがる

標準誤差は標本平均値の誤差を表す要約値

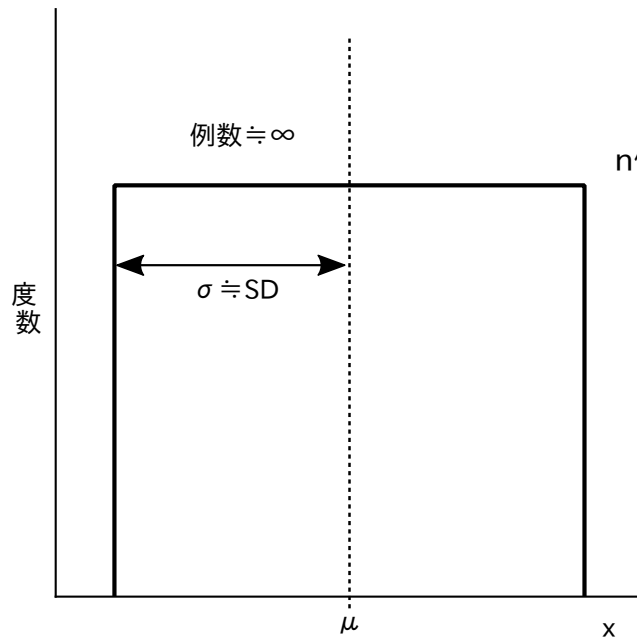


図1.3 母集団のデータ分布

n例を無作為抽出して
標本平均値mを
無限回求める

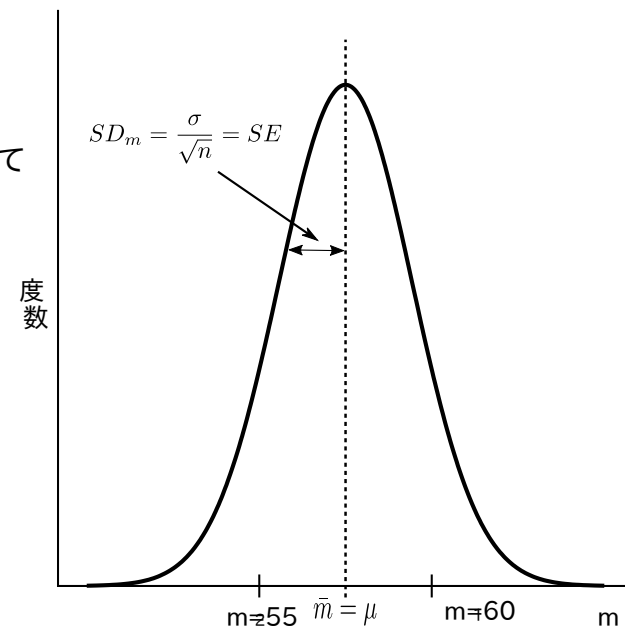
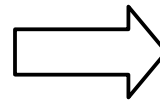


図1.4 標本平均値の分布

標本平均値の度数分布の特徴

- 母集団がどんな分布をしていても近似的に正規分布になる(nが多いほど近似が良い)
→ **中心極限定理(推測統計学の基本定理)**
- 標本平均値の平均値は母平均値と一致する
- 標本平均値の標準偏差は次のような値になる → **標準誤差**

$$SD_m = \frac{\sigma}{\sqrt{n}} \doteq \frac{SD}{\sqrt{n}} = SE$$

標準誤差と標準偏差の使い分け

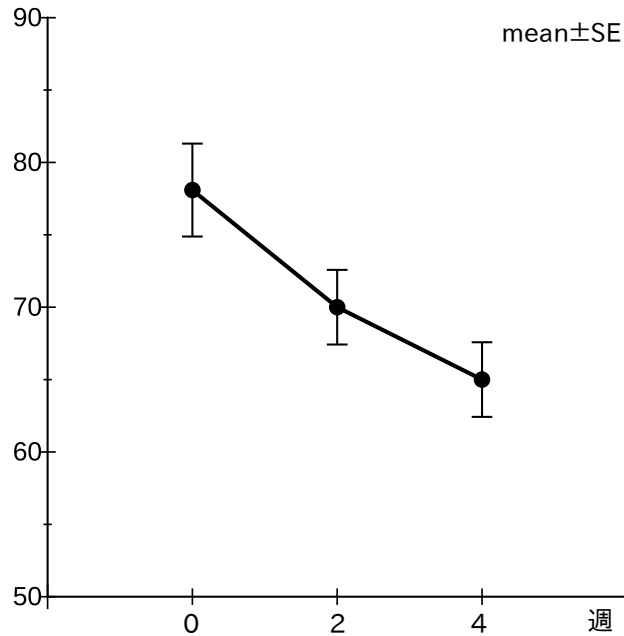


図1.5 体重の推移

母平均値の変化とその推測誤差を
表したい時は標準誤差

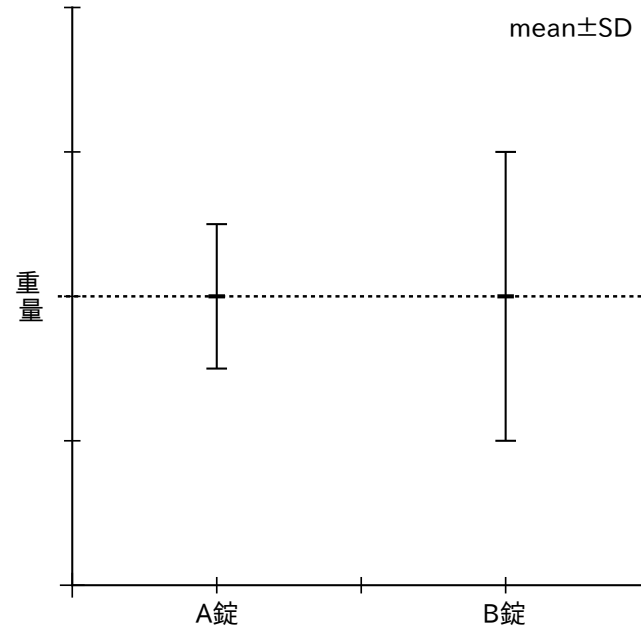


図1.6 錠剤の重量

データのバラツキ具合を
表したい時は標準偏差

$\mu = m \pm SE$: μ を m で推測すると SE 程度の推測誤差がある

※ $m \pm 2 \times SE$ (95% 信頼区間) を描く方が合理的 ← リスク比やオッズ比はこれが普通

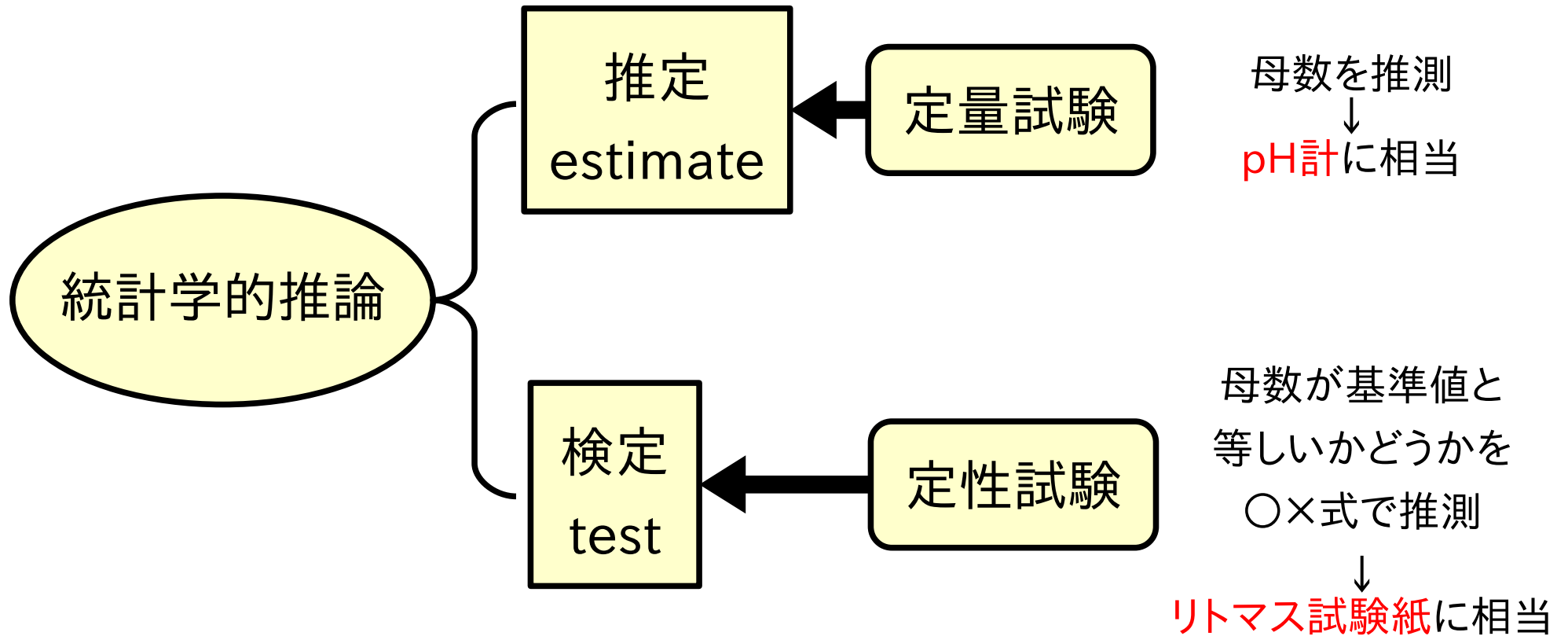


$m \pm SD$: データは m を中心にして SD 程度のバラツキがある

統計学の落とし穴

- 標準偏差と標準誤差
- 有意性検定と統計的仮説検定
- パラメトリック手法とノンパラメトリック手法
- ハンディキャップ方式の検定
- 相関分析と回帰分析
- 葬患係数と怪奇直線
- 選択例の相関

推定と検定



検定よりも推定の方が重要

しかし研究現場や厚労省では検定が偏重されている

∴ ○×式の方が採点が楽!

点推定はピンポイントの推測

母平均値： $\mu \approx m$ ：標本平均値

母標準偏差： $\sigma \approx SD = \sqrt{V} = \sqrt{\frac{SS}{n-1}}$ ：不偏標準偏差

点推定(point estimation)

母平均値を標本平均値で、母標準偏差を不偏標準偏差でそのまま推測

母標準偏差は点推定が普通

区間推定は幅を持たせた推測

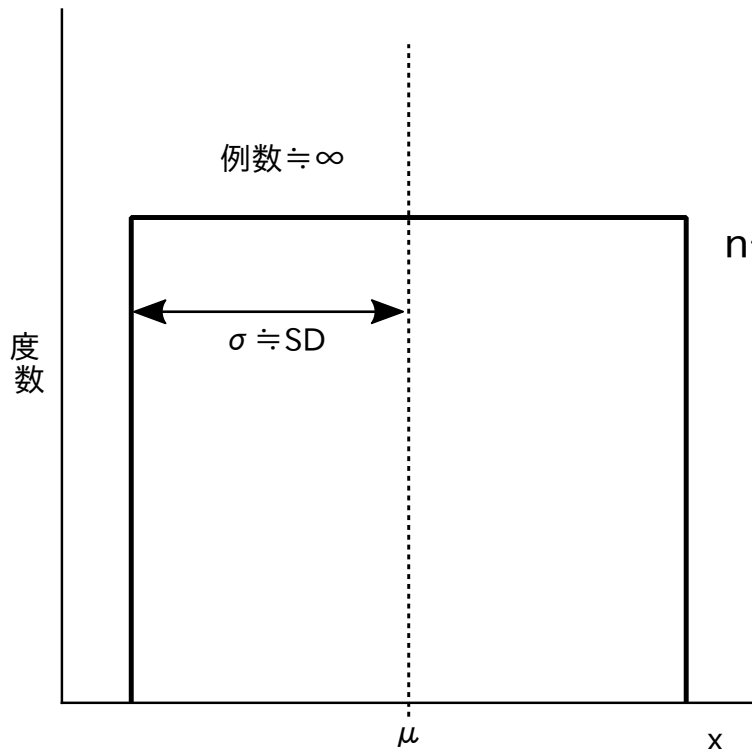


図1.3 母集団のデータ分布

n例を無作為抽出して
標本平均値mを
無限回求める
⇒

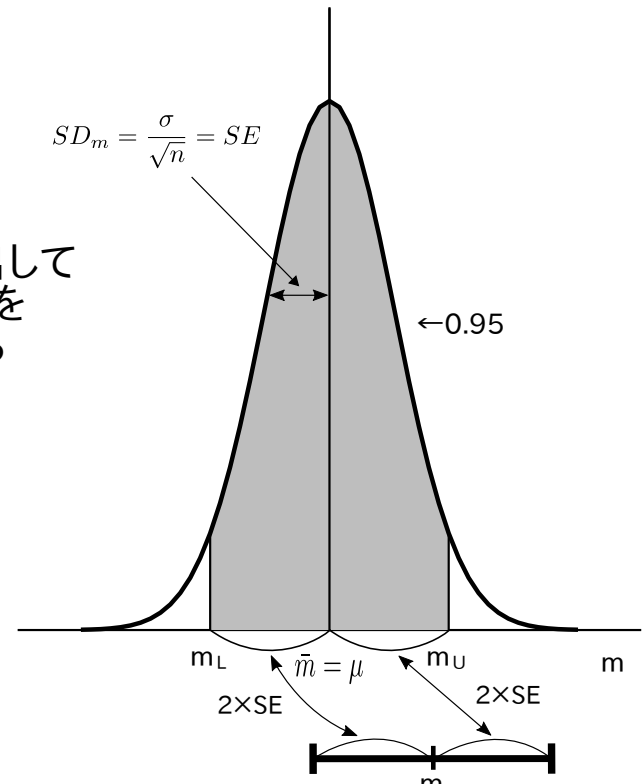


図1.8 標本平均値の分布と信頼区間

区間推定(interval estimation)
ある程度の幅を持たせて母数を推測する
母平均値は区間推定が普通

母平均値の区間推定法

標本平均値 m の分布は近似的に正規分布になる

m の平均値は母平均値 μ に、 m の標準偏差は標準誤差 SE になり
 $\mu \pm 2 \times SE$ の範囲に約95%の m が含まれる

ある標本平均値 m が $\mu \pm 2 \times SE$ の範囲に含まれる確率は約95%

逆に $m \pm 2 \times SE$ の範囲に μ が含まれる確率も約95%

95%信頼区間: $\mu \doteq m \pm 2 \times SE \rightarrow \mu_L = m - 2 \times SE \quad \mu_U = m + 2 \times SE$

95%信頼区間(95%CI): 母平均値が95%の確率で含まれる区間、95%信頼限界(95%CL)

μ_L : 信頼区間下限

μ_U : 信頼区間上限

95%: 信頼係数

点推定と区間推定

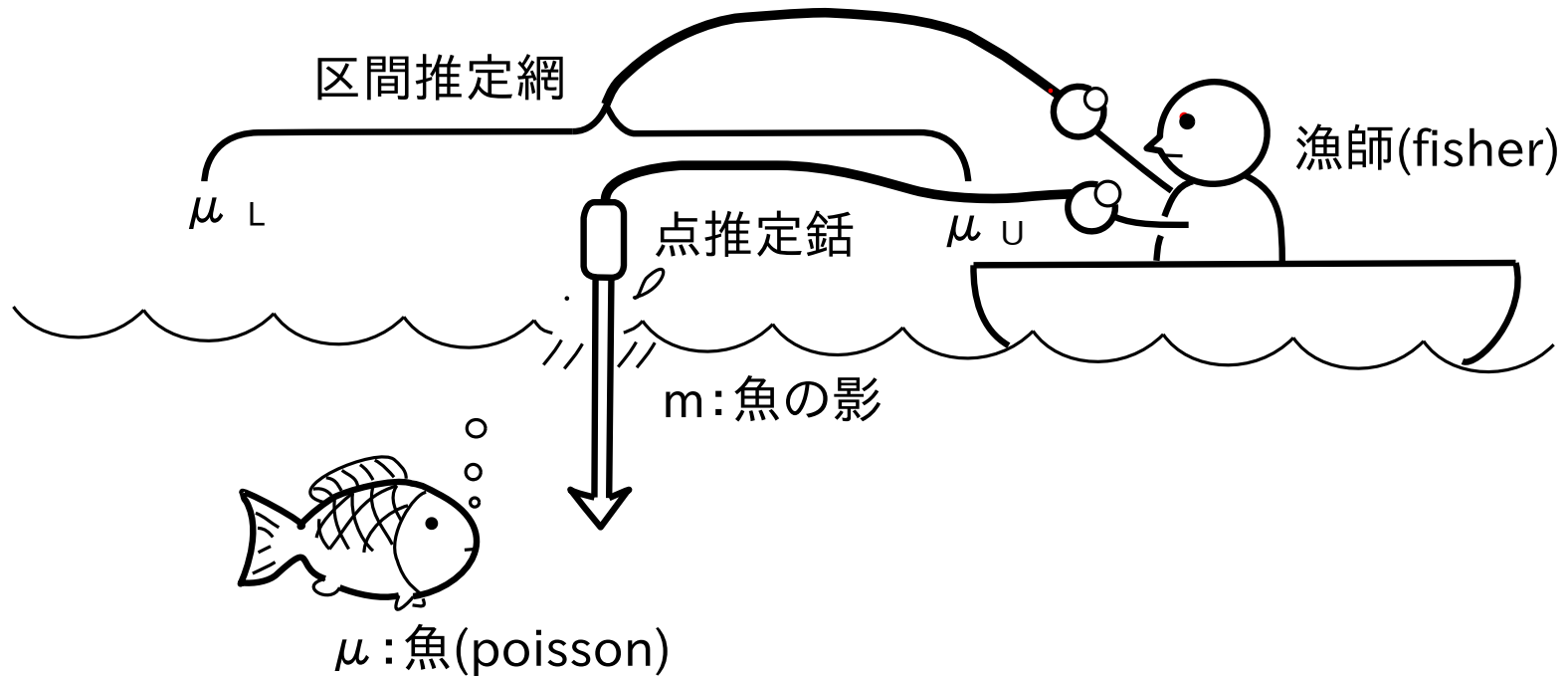


図1.9 点推定と区間推定

推定は漁師(Fisher)が
水面に映った魚(Poisson)の影 m を見て魚 μ を捕まえるようなもの
点推定は鉤で一突き、区間推定は投網を打つことに相当
普通は点推定を用い、重要な時だけ区間推定を行う

有意性検定は基準値を用いた○×式の定性試験

問題: 日本人の平均体重は50kgか?

帰無仮説 H_0 : 日本人の平均体重は50kgである ← 問題の答えは○



対立仮説 H_1 : 日本人の平均体重は50kgではない ← 問題の答えは×

有意性検定の手順

1. 問題を設定する → 医学的意義のある基準値 μ_0 を決める
2. 問題の答を○×式で設定する → 帰無仮説 H_0 と対立仮説 H_1 を設定する
3. データに基づいて仮説の妥当性を判定する

検定は定性試験だから定量試験である推定結果から判定可能

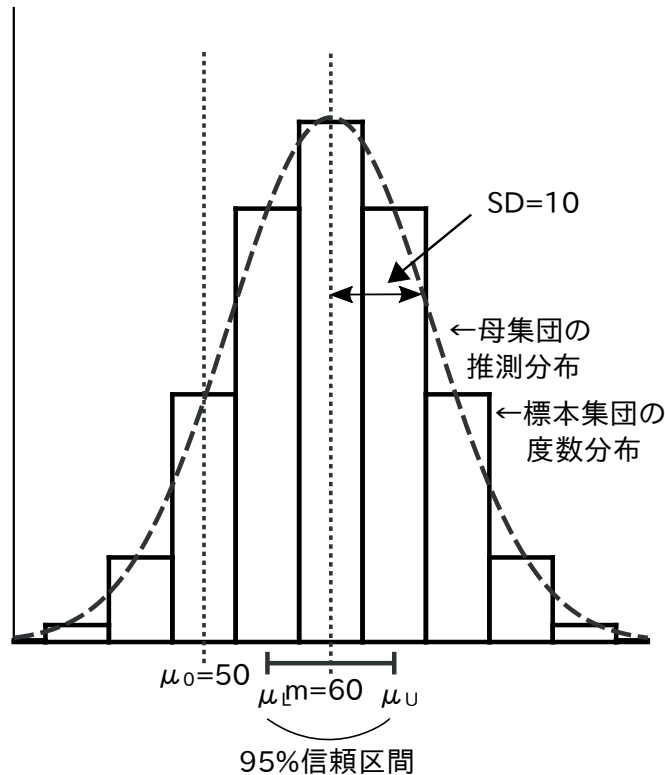


図1.10-a 信頼区間と有意性検定

$$95\% \text{信頼区間: } \mu = 60 \pm 2 \times \frac{10}{\sqrt{100}} = 60 \pm 2 \rightarrow \mu = 58 \sim 62$$

母平均値は95%の確率で58～62kgの間にある
→ 母平均値は95%以上の確率で50kgではない

母集団から見た有意性検定

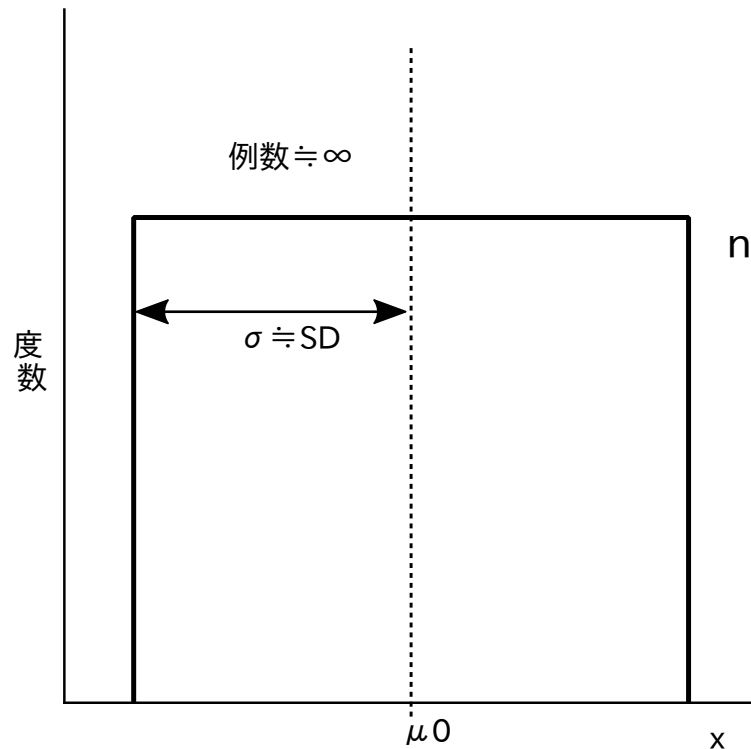


図1.3 母集団のデータ分布

n例を無作為抽出して
標本平均値mを
無限回求める

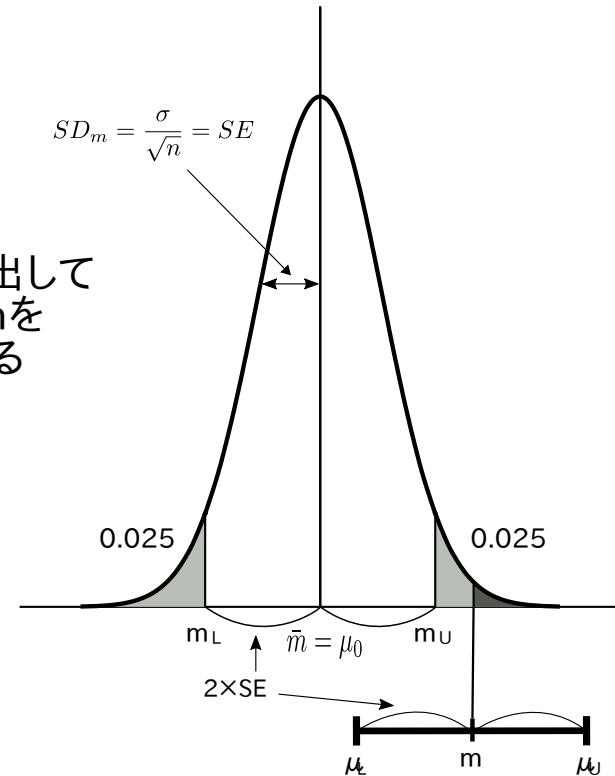
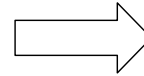


図1.5.3 標本平均値の分布と信頼区間

$\mu_0=50$ の時(帰無仮説が正しい時): 標本平均値の95%が含まれる区間= $50 \pm 2 = 48 \sim 52$

$m_L=48$: 下側棄却域の上限 $m_U=52$: 上側棄却域の下限 < 60: 標本平均値

標本平均値が棄却域に入っている=95%信頼区間に基準値 μ_0 が入っていない
 =標本平均値mから右側の面積(図1.5.3の濃い灰色の部分) $\times 2 = p$ 値(有意確率)が0.05以下
 → 母平均値は95%以上の確率で50kgではない

有意性検定は対立仮説だけを採用する

棄却域に標本平均値が入っている時=95%信頼区間に基準値が入っていない時
統計学的結論:日本人の平均体重は50kgではない ← 問題の答えは×

「有意水準5%で有意」または「危険率5%で有意」と表現する
→ 統計学的結論が間違っている確率が5%程度ある

棄却域に標本平均値が入っていない時=95%信頼区間に基準値が入っている時
統計学的結論:日本人の平均体重は50kgではないと断定できない ← 問題の答えは**保留**

「有意水準5%で有意ではない」または「危険率5%で有意ではない」と表現する
これは帰無仮説「日本人の平均体重は50kgである」の採用ではないことに注意!

※不確かなデータから得られた結果を解釈する時は確定的なことを断言する方が「非科学的」
※得られたデータから結論できる限界を明確にするのが「科学的」

「有意差あり」は「実質科学的に差がある」という意味ではない

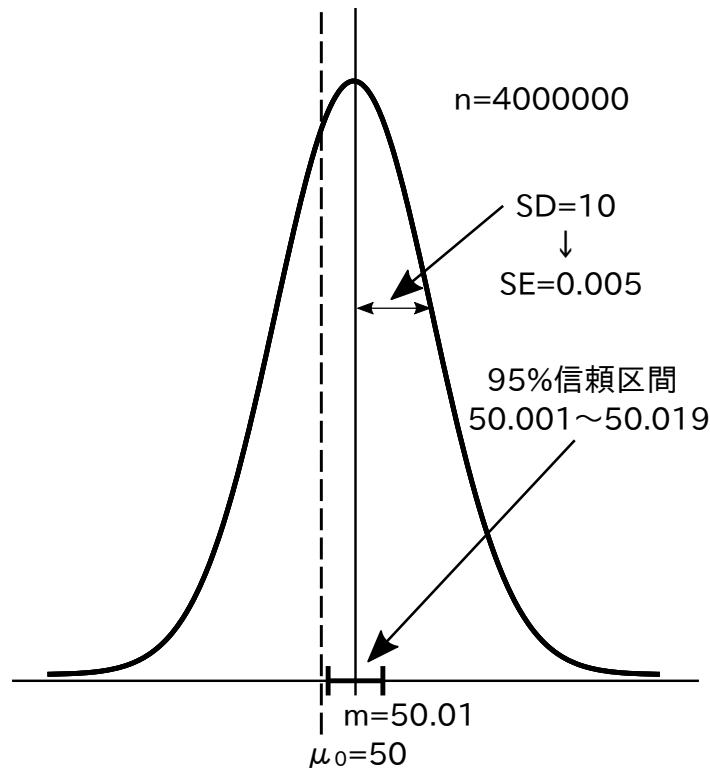


図1.11 有意でも実質科学的には無意味な差

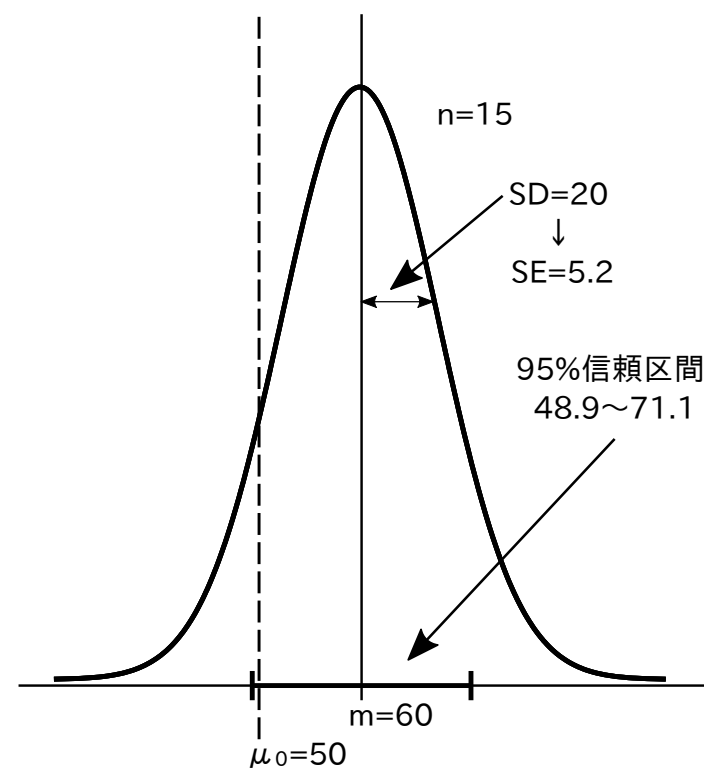


図1.12 実質科学的に意味があっても有意ではない差

有意: 数学的に意味が有る → 統計学的結論の信頼性が高い

有意ではない: 数学的に意味が無い

→ 統計学的結論の信頼性が低い → 統計学的結論を保留する

※母平均値が基準値とぴったり一致することは現実には有り得ない

→ わざわざ検定する必要はない! → 検定廃止論

有意症・有意症症候群は難治性疾患

有意症(significantosis)・有意症症候群(significant syndrome)

「有意差あり=実質科学的に有意義な差がある」とか

「有意差なし=実質科学的に有意義な差がない」と誤解する**難治性の疾患**

医学界や厚生労働省等大流行中!

有意症・有意症症候群の予防策

- 「有意差あり」や「有意差なし」という用語を使わず
「差は有意である」や「差は有意ではない」という用語を使う
- **推定結果**を重視する

統計的仮説検定は有意性検定の欠点を是正した手法

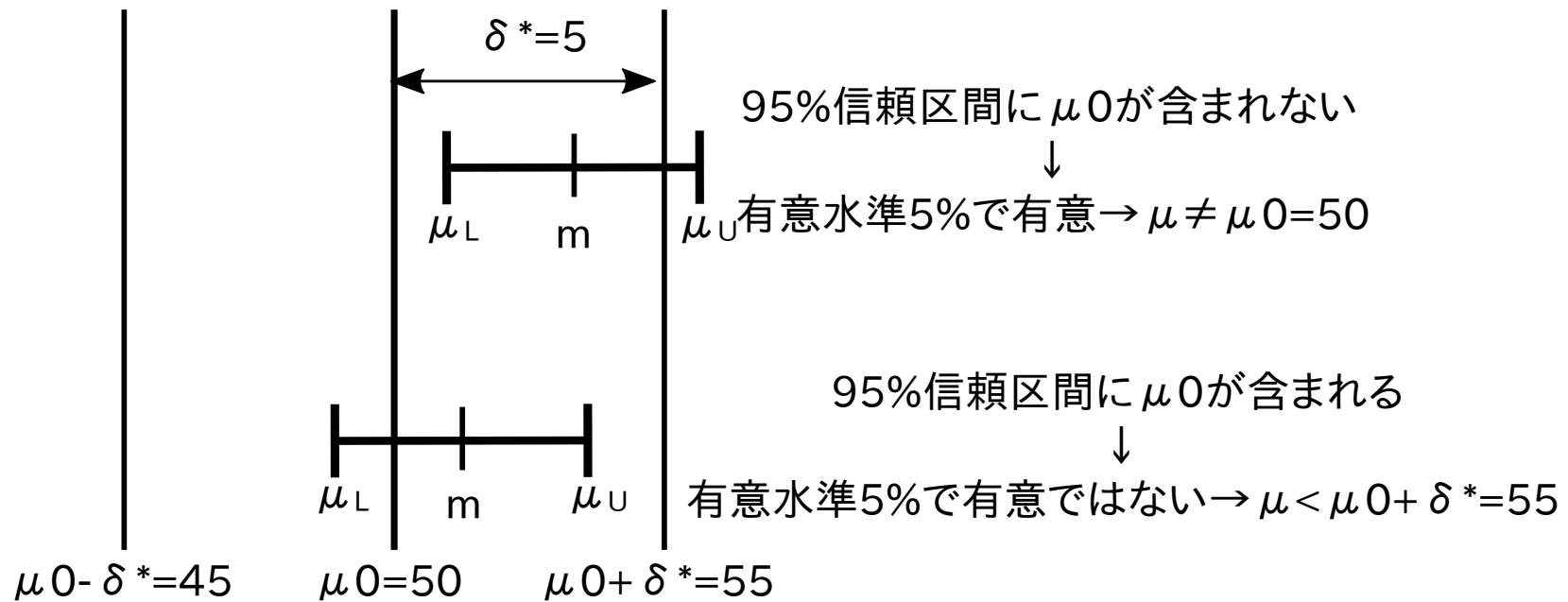


図1.13 信頼区間と統計的仮説検定

$\delta^* = \pm 5\text{kg}$: (最小)検出差 (scientific significant difference) \div 医学的な許容範囲

$$95\% \text{信頼区間: } \mu = 51 \pm 2 \times \frac{10}{\sqrt{100}} = 51 \pm 2 \rightarrow \mu = 49 \sim 53$$

母平均値は95%の確率で49~53kgの間にある

\rightarrow 母平均値は95%以上の確率で45kgよりも重く55kgよりも軽い

統計的仮説検定は基準値と許容範囲を用いた○×式の定性試験

問題: 日本人の平均体重は50kgか?

帰無仮説 H_0 : 日本人の平均体重は50kgである ← 問題の答えは○



対立仮説 H_1 : 日本人の平均体重は45kgまたは55kgである ← 問題の答えは×

統計的仮説検定の手順

1. 問題を設定する → 医学的意義のある基準値 μ_0 と許容範囲 δ^* を決める
2. 問題の答を○×式で設定する → 帰無仮説 H_0 と具体的な対立仮説 H_1 を設定する
3. データに基づいて仮説の妥当性を判定する

検定は定性試験だから定量試験である推定結果から判定可能

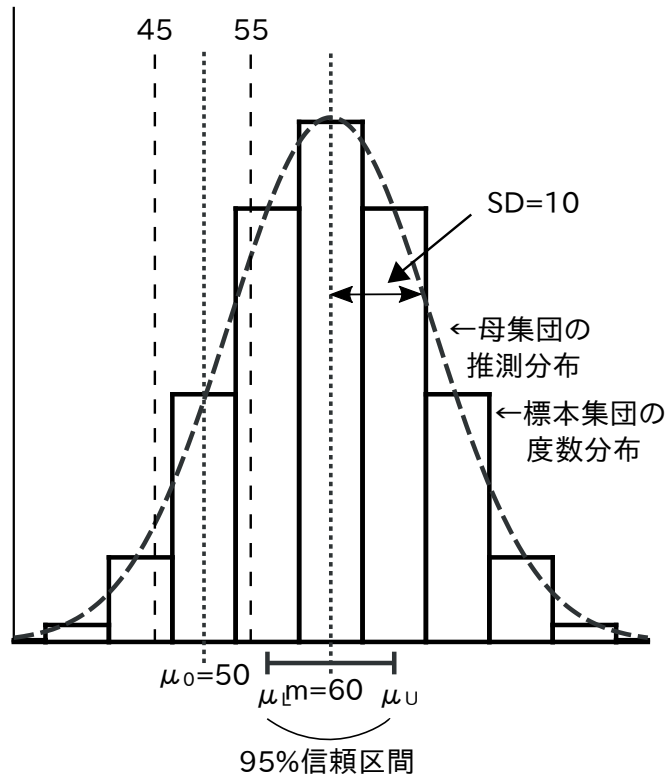


図1.10-b 信頼区間と統計的仮説検定

$$95\% \text{信頼区間: } \mu = 60 \pm 2 \times \frac{10}{\sqrt{100}} = 60 \pm 2 \rightarrow \mu = 58 \sim 62$$

母平均値は95%の確率で58~62kgの間にある
→ 母平均値は95%以上の確率で50kgではない

母集団から見た統計的仮説検定

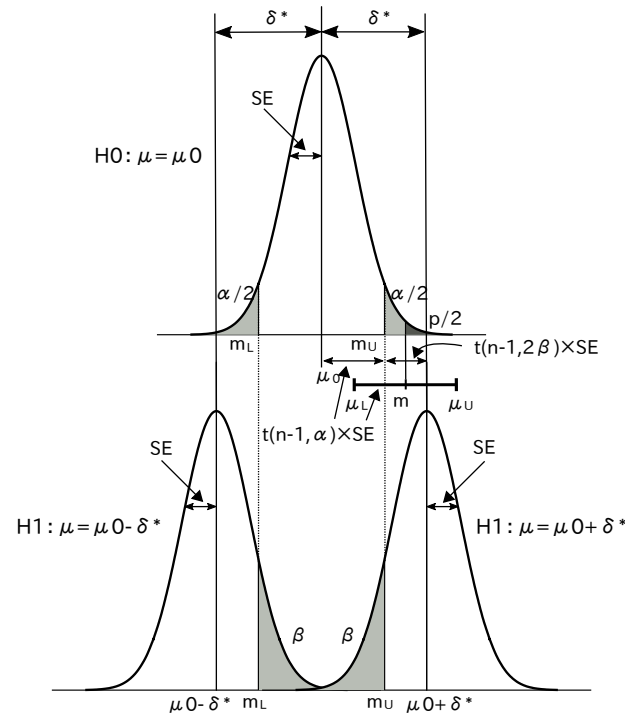


図1.14 統計的仮説検定の模式図

$\mu_0=50$ の時(帰無仮説が正しい時): 標本平均値の $(1-\alpha)$ が含まれる区間 $=50 \pm 2 \times SE = m_L \sim m_U$

$\mu_0 - \delta^* = 45$ の時(対立仮説が正しい時): 標本平均値の $(1-\beta)$ が含まれる区間 $= -\infty \sim m_L$

または $\mu_0 + \delta^* = 55$ の時(対立仮説が正しい時): 標本平均値の $(1-\beta)$ が含まれる区間 $= m_U \sim \infty$

m_L : 下側棄却域の上限 m_U : 上側棄却域の下限

α エラー=アワテの言い過ぎ: 帰無仮説が正しい時にアワテて $\mu \neq \mu_0 = 50$ と結論する確率

β エラー=ボンヤリの見逃し: 対立仮説が正しい時にボンヤリして $45 < \mu < 55$ と結論する確率

※ $(1-\beta)$: 検出力=対立仮説が正しい時に $\mu \neq \mu_0 = 50$ と結論する確率 ← 医学分野では80%

統計的仮説検定は有意ではない時も結論を採用する

棄却域に標本平均値が入っている時=95%信頼区間に基準値が入っていない時
統計学的結論:日本人の平均体重は50kgではない ← 問題の答えは×

「有意水準5%で有意」または「危険率5%で有意」と表現する
これは「日本人の平均体重は45kgまたは55kg」の採用ではないことに注意!
※統計学的結論が間違っている確率は5% ($\alpha = 0.05$)

棄却域に標本平均値が入っていない時=95%信頼区間に基準値が入っている時
かつ信頼区間が許容範囲内に収まっている時
統計学的結論:日本人の平均体重は45kgよりも重く55kgよりも軽い ← 問題の答えは△

「有意水準5%で有意ではない」または「危険率5%で有意ではない」と表現する
これは「日本人の平均体重は50kgである」の採用ではないが、実質的には同じ意味
※統計学的結論が間違っている確率は20% ($\beta = 0.2$) ← $\alpha = \beta$ にするのが理想

統計的仮説検定は事前に試験の必要例数を計算する

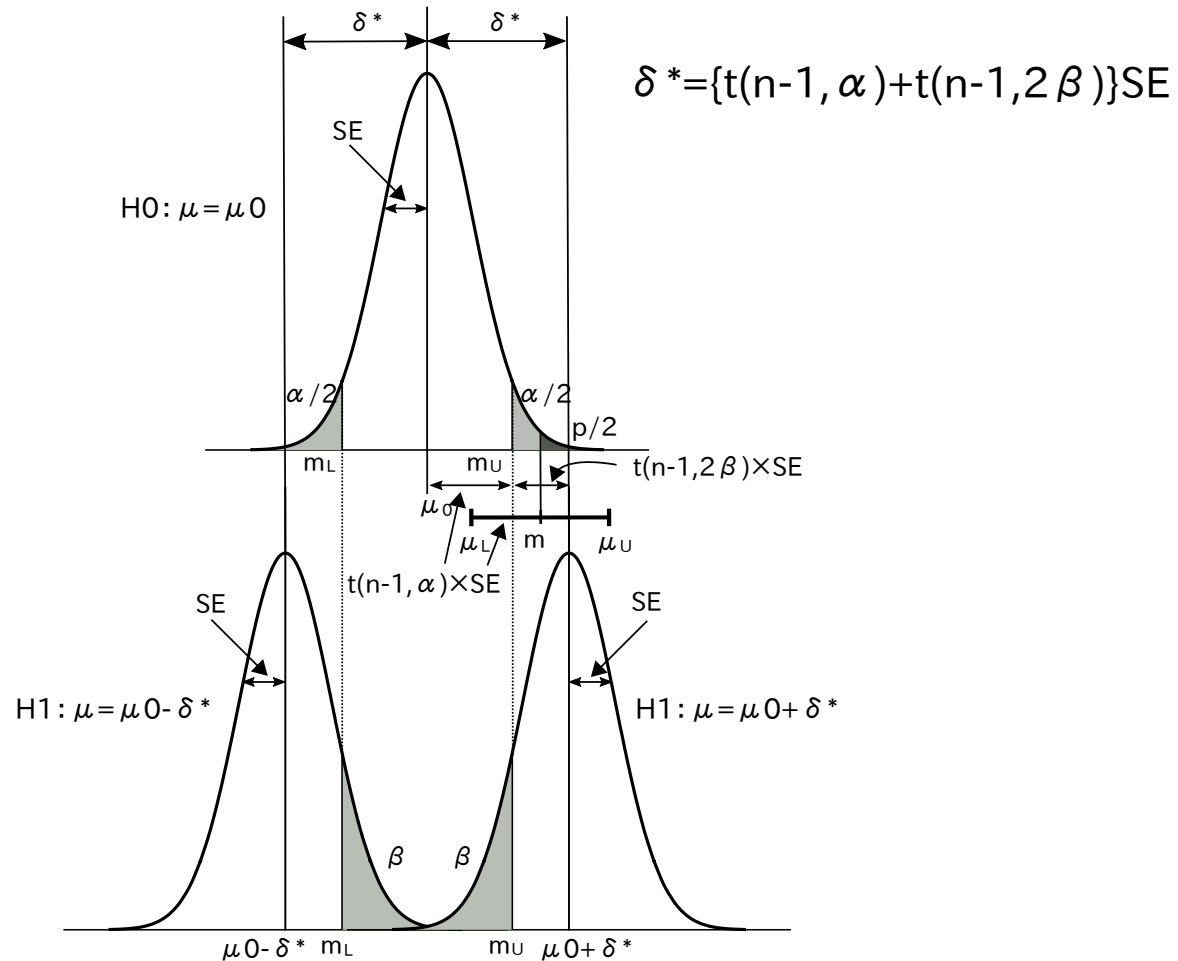


図1.14 統計的仮説検定の模式図

必要例数の計算式(お座敷):
$$n = \left[\{t(\infty, \alpha) + t(\infty, 2\beta)\} \frac{\sigma}{\delta^*} \right]^2$$

信頼区間を許容範囲よりも小さくする必要がある

統計的仮説検定は事後に試験の検出力分析を行う

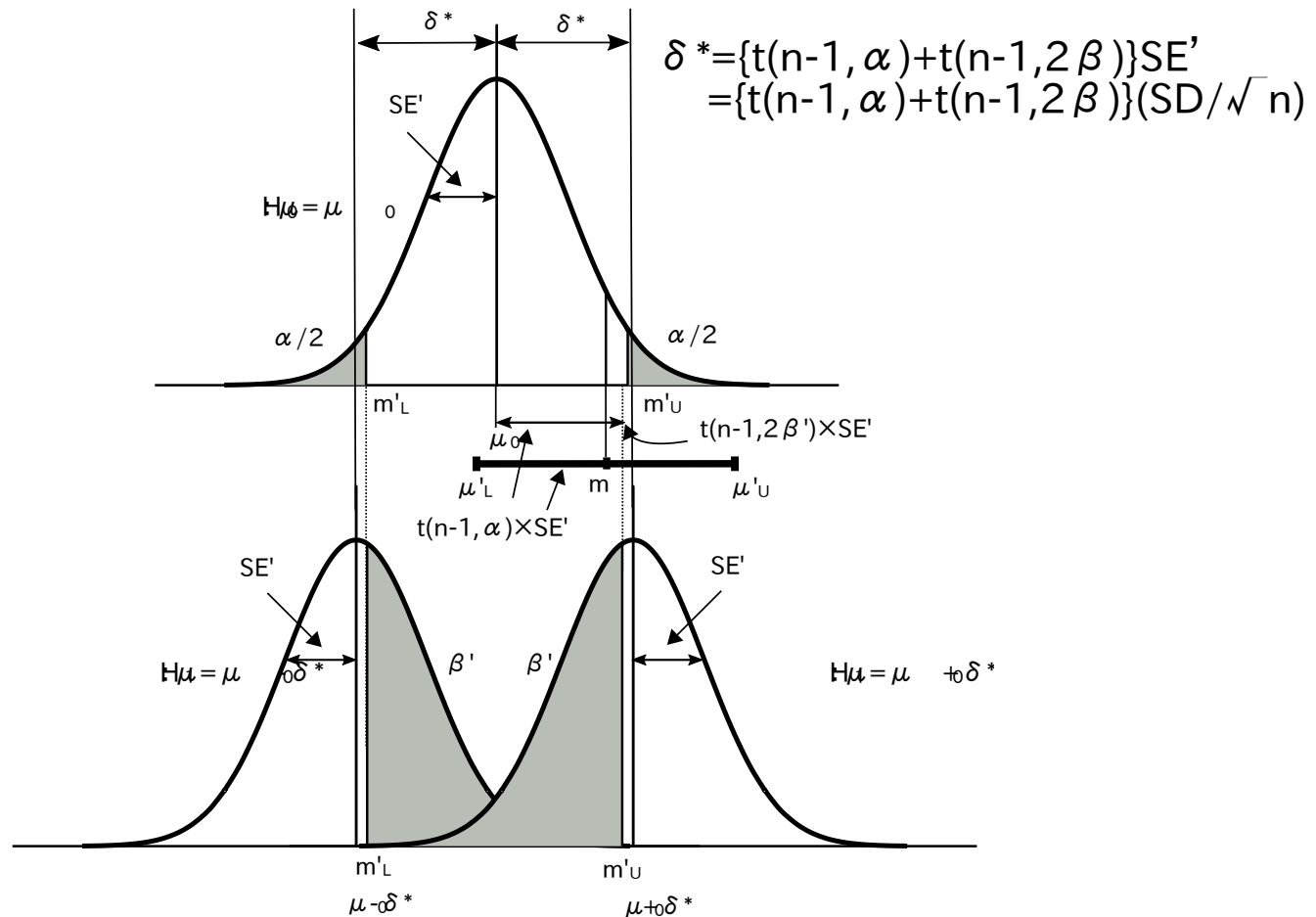


図1.6.8 統計的仮説検定の実際の模式図

実際のnとSDを用いた検出力: $\delta^* = \{t(n-1, \alpha) + t(n-1, 2\beta')\} (SD / \sqrt{n})$

実際の検出力(1-β')が事前に設定した(1-β)以上であることを確認する

検定結果だけから実質科学的な判断をするのは危険

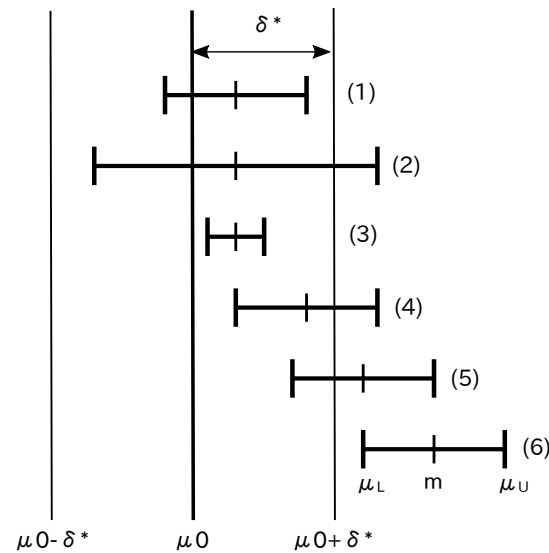


図1.15 検定結果と信頼区間

	検定結果	推定結果	実質科学的な判断
(1)	有意ではない	$\mu \doteq \mu_0$	母平均値は基準値とほぼ等しい
(2)	有意ではない	$\mu = \mu_0 \sim \mu_0 + \delta^*$	この結果だけでは判断できない 信頼区間をもっと狭くする必要がある(例数を増やす)
(3)	有意	$\mu_0 < \mu < \mu_0 + \delta^*$	母平均値は基準値と実質的に変わらない
(4)	有意	$\mu \doteq \mu_0 + \delta^*$	母平均値は基準値と実質的に変わらない可能性が高い
(5)	有意	$\mu \doteq \mu_0 + \delta^*$	母平均値は基準値よりも大きい可能性が高い
(6)	有意	$\mu_0 + \delta^* < \mu$	母平均値は基準値よりも大きい

※生物学的同等性試験では推定結果を重視し、検定結果は参考程度 → 検定廃止論

例数が多いと検定結果は科学的な判断には使えない

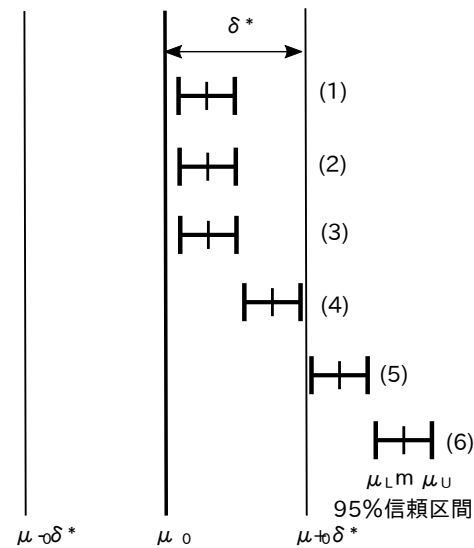


図1.7.6 例数が多い時の検定結果と信頼区間

	検定結果	推定結果	実質科学的な判断
(1)	有意	$\mu_0 < \mu < \mu_0 + \delta^*$	母平均値は基準値と実質的に変わらない
(2)	有意	$\mu_0 < \mu < \mu_0 + \delta^*$	母平均値は基準値と実質的に変わらない
(3)	有意	$\mu_0 < \mu < \mu_0 + \delta^*$	母平均値は基準値と実質的に変わらない
(4)	有意	$\mu_0 < \mu < \mu_0 + \delta^*$	母平均値は基準値と実質的に変わらない
(5)	有意	$\mu_0 + \delta^* < \mu$	母平均値は基準値よりも大きい
(6)	有意	$\mu_0 + \delta^* < \mu$	母平均値は基準値よりも大きい

※生物学的同等性試験では推定結果を重視し、検定結果は参考程度 → 検定廃止論

$p < 0.001$ になっても結果の信頼性は95%

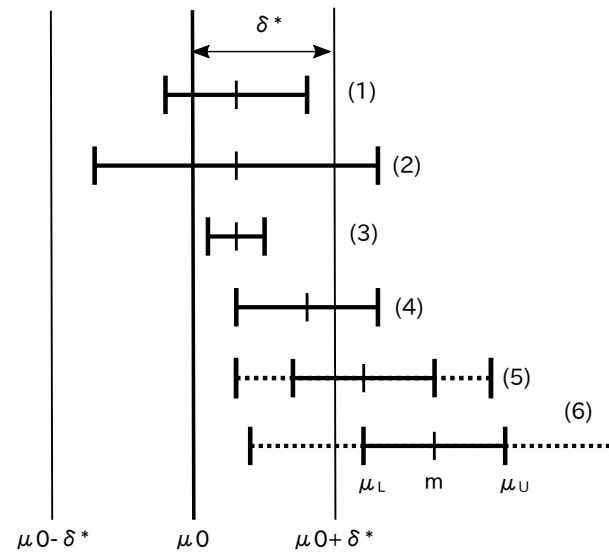


図1.20 検定結果と信頼係数を変えた信頼区間

(1)と(2): $p > 0.05$ (3)と(4): $p < 0.05$ (5): $p < 0.01$ (6): $p < 0.001$ になった時

(5)を「有意水準1%で有意」と表現すると99%信頼区間が対応

(6)を「有意水準0.1%で有意」と表現すると99.9%信頼区間が対応

→ 信頼区間の幅が広がって($\mu_0 + \delta^*$)が入ってしまい、結論が曖昧になる

必要例数を計算した時の有意水準(α エラー)によって結果の信頼性が決まる

統計学の落とし穴

- 標準偏差と標準誤差
- 有意性検定と統計的仮説検定
- パラメトリック手法とノンパラメトリック手法
- ハンディキャップ方式の検定
- 相関分析と回帰分析
- 葬患係数と怪奇直線
- 選択例の相関

パラメトリック手法とノンパラメトリック手法の特徴

統計手法

パラメトリック手法
母数に依存した手法
→ 数学的モデルを利用

要約値: 平均値、標準偏差

検定・推定: 平均値の検定(t検定)

- 母集団のデータが特定の分布に従うと仮定する
- 要約値がデータの分布状態を反映する
→ データの分布状態によって要約値の値が変化する
- 結果の精度が高い
- 結果の普遍化が容易

ノンパラメトリック手法
母数に依存しない手法
→ 数学的モデル利用せず

要約値: 中央値、順位平均値、割合

検定・推定: 順位和検定、出現率の検定(χ^2 検定)

- 母集団のデータがどんな分布をしていても良い
- 要約値がデータの分布状態を反映しにくい
→ データの分布状態によって要約値の値が変化しにくい
- 結果の精度が低い
- 結果の普遍化は困難

データの分布状態で適用する手法を選択してはいけない

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せよ
※データが正規分布しない時はパラメトリック手法が使えないのではなく
検定と推定の効率がノンパラメトリック手法よりも悪くなる時がある

要約値に関する科学的な考察を無視した乱暴な主張だから鵜呑みにしてはいけない

※腹痛で胃腸薬を求めたら「今は効果の弱い胃腸薬しかないので
代わりに効果の強い降圧剤を使った方が良い」と助言されるようなもの

統計手法は要約値の数学的な信頼性を評価するためのもの

統計手法と要約値は1対1で対応している

統計手法を決定する最も重要な要因は要約値の科学的な妥当性

標本平均値の分布

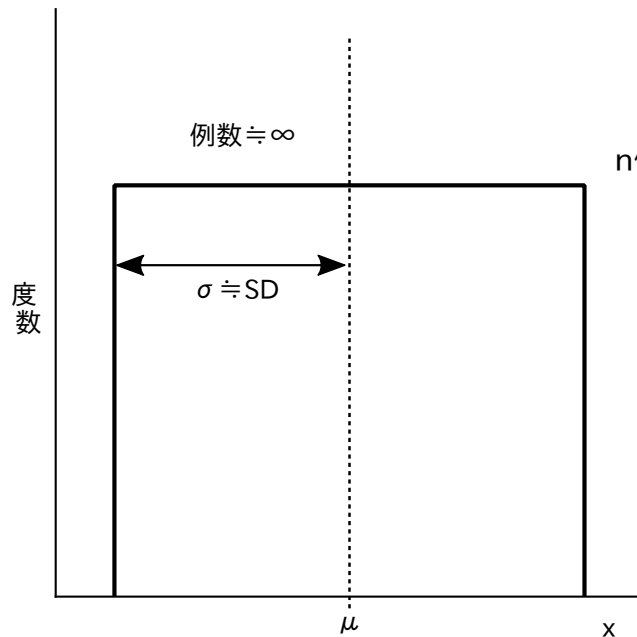


図1.3 母集団のデータ分布

n例を無作為抽出して
標本平均値mを
無限回求める

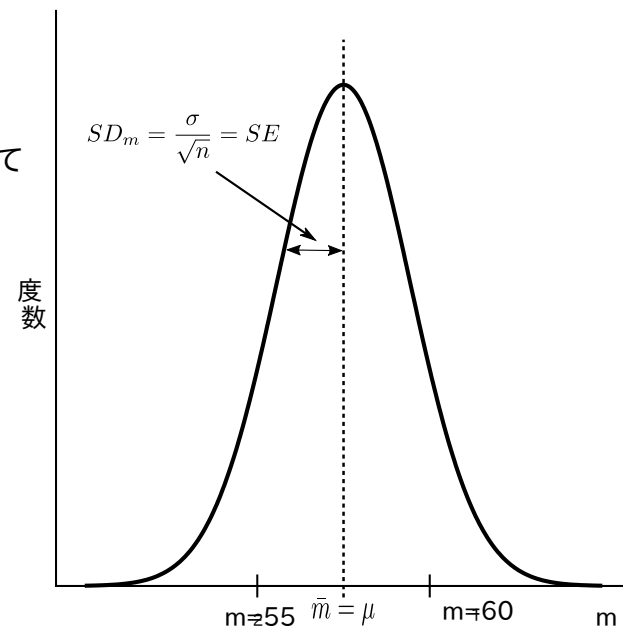
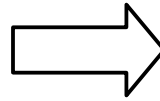


図1.4 標本平均値の分布

標本平均値の度数分布の特徴

- 母集団がどんな分布をしていても近似的に正規分布になる(nが多いほど近似が良い)
→ **中心極限定理(推測統計学の基本定理)**
- 標本平均値の平均値は母平均値と一致する
- 標本平均値の標準偏差は次のような値になる → **標準誤差**

$$SD_m = \frac{\sigma}{\sqrt{n}} \doteq \frac{SD}{\sqrt{n}} = SE$$

順位平均値の分布

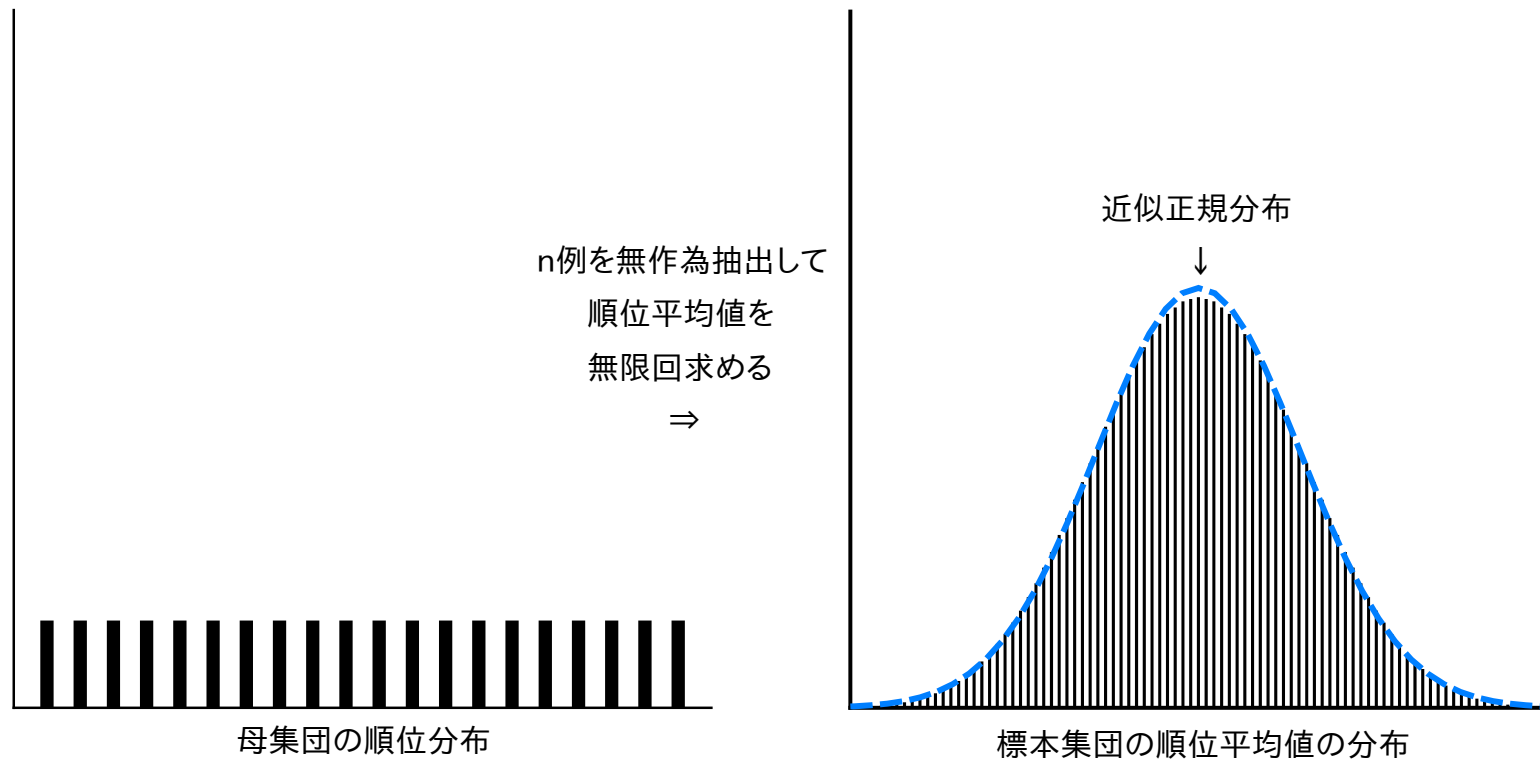


図4.10 順位分布と順位平均値の分布

- 図4.10の左のグラフ: 同位のデータがない時の母集団の順位分布
- 図4.10の右のグラフ: 順位平均値の理論分布
- 順位平均値の分布は中心極限定理によって近似的に正規分布になる

※パラメトリック手法もノンパラメトリック手法も要約値の近似正規分布を利用している!

外れ値がある時のデータの分布と順位分布

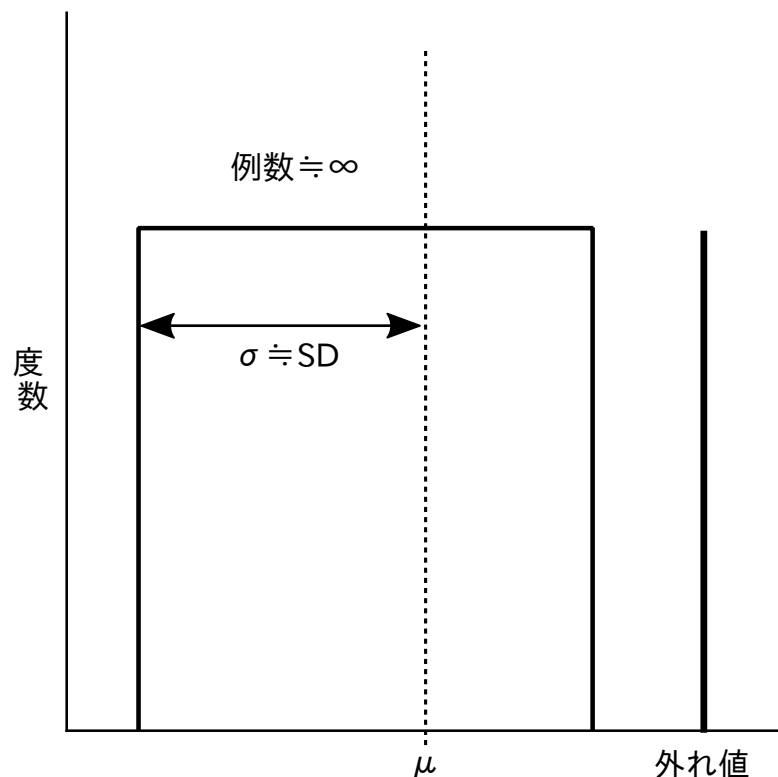


図1.18 外れ値がある母集団

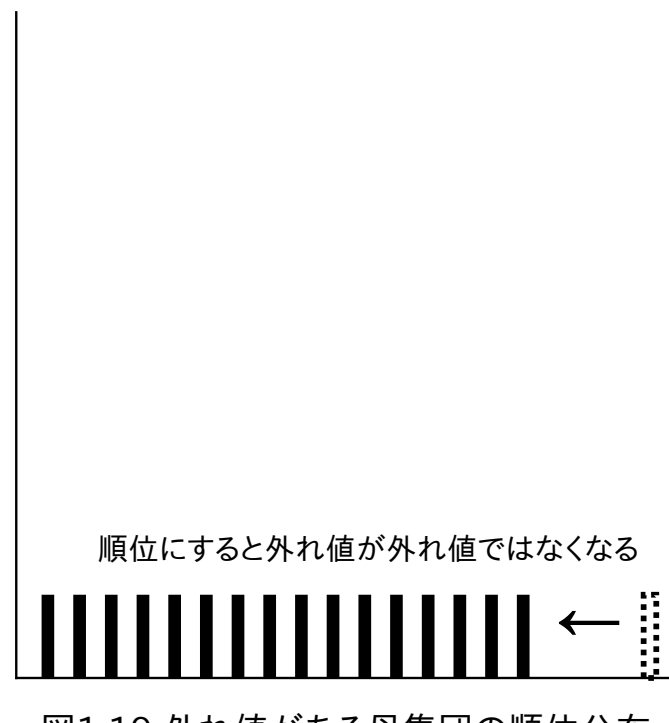
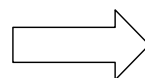


図1.19 外れ値がある母集団の順位分布

外れ値がある時

データの分布の標準偏差は大きくなるが順位分布の標準偏差は不変

→ 標本平均値の標準誤差は大きくなるが順位平均値の標準誤差は不変

∴ 順位平均値よりも平均値の方が推定と検定の効率が悪くなる時がある!

中央値の場合

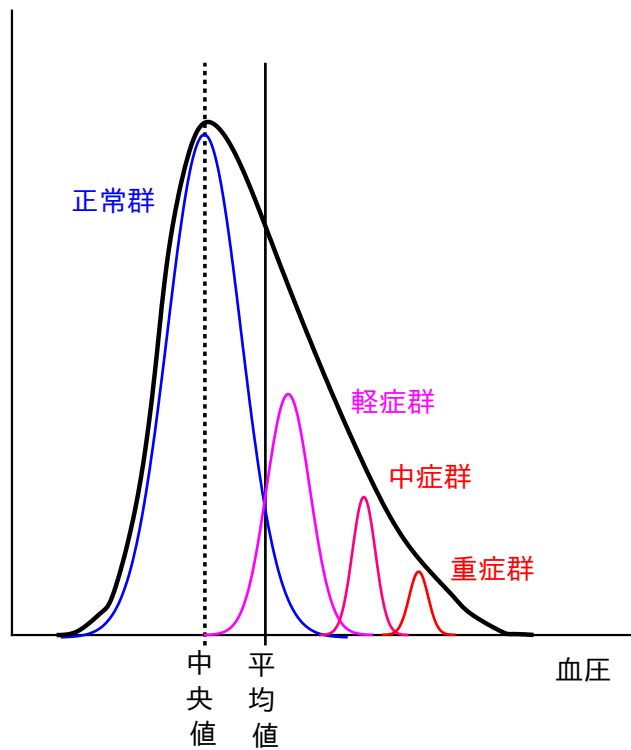


図3.1 投与前の血压分布

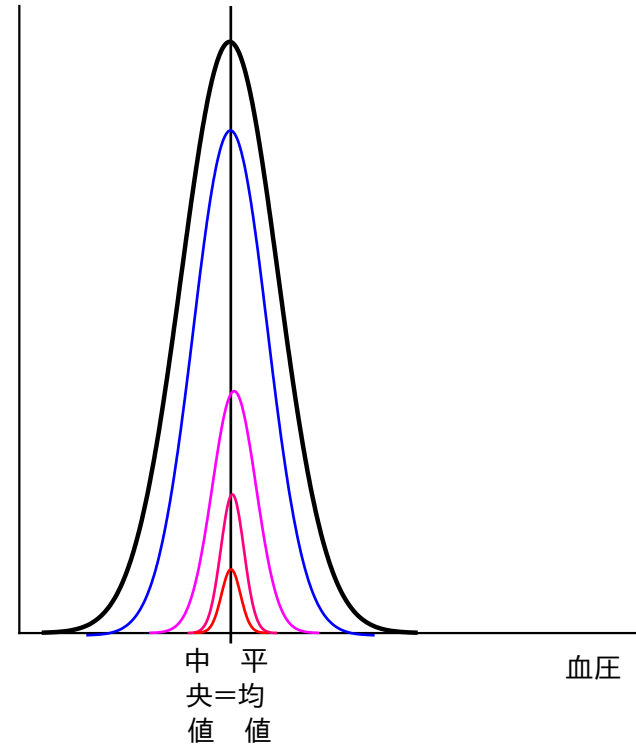
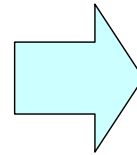


図3.2 投与後の血压分布

図3.1の左側の集団に降圧剤を投与したところ、高血圧患者だけ血圧が低下して図3.1の右側のようになった
→ 大多数の正常者は血圧が不変のため**平均値は低下、中央値は不変**

降圧剤の効果があったと評価すべきか、なかったと評価すべきか？

→ **医学的に妥当な評価指標は平均値か中央値か？**

順位和検定の場合

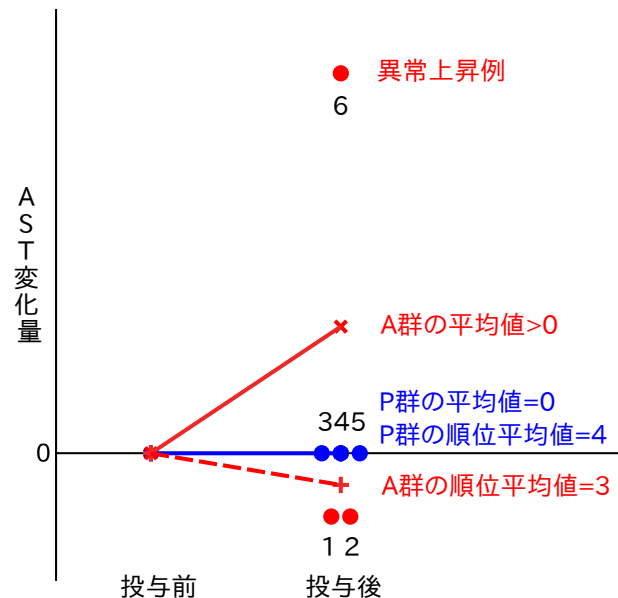


図3.3 平均値と順位平均値が矛盾する例

A群は副作用によるAST異常上昇例が発生し、それ以外はわずかに低下、P群は全例不変
変化量の平均値:A群は上昇、P群は不変 → A群はP群よりも上昇
変化量の順位平均値:A群は低下、P群は不変 → A群はP群よりも低下

A群に副作用があったと評価すべきか、なかったと評価すべきか？
→ 医学的に妥当な評価指標は平均値か順位平均値か？
※医学分野では疾患による異常値や副作用による異常値が必然的に発生する

※順位和検定はデータの実測値には科学的意義がなく、順位だけが科学的意義を持つ時に適用すべき手法

順序尺度のデータを計量尺度扱いした方が良い例

＜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

- ・Mann-WhitneyのU検定(Wilcoxonの2標本検定): 正規分布 $z_0=0$ 有意確率 $p=1$
- ・著明改善～著明悪化を1～5と数量化して計量尺度扱いした時(リッカート尺度)
 薬剤1投与群の平均値=2.5 薬剤2投与群の平均値=2.5 2標本t検定: $t_0=0$ $p=1$

＜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

- ・Mann-WhitneyのU検定(Wilcoxonの2標本検定): 正規分布 $z_0=0$ 有意確率 $p=1$
- ・著明改善～著明悪化を1～5と数量化して計量尺度扱いした時(リッカート尺度)
 薬剤1投与群の平均値=2.5 薬剤2投与群の平均値=3.0 2標本t検定: $t_0=2$ $p=0.0338$ *

医学的に見ると計量尺度扱いした方が妥当な結果

統計学の落とし穴

- 標準偏差と標準誤差
- 有意性検定と統計的仮説検定
- パラメトリック手法とノンパラメトリック手法
- ハンディキャップ方式の検定
- 相関分析と回帰分析
- 葬患係数と怪奇直線
- 選択例の相関

ハンディキャップ方式の検定の模式図

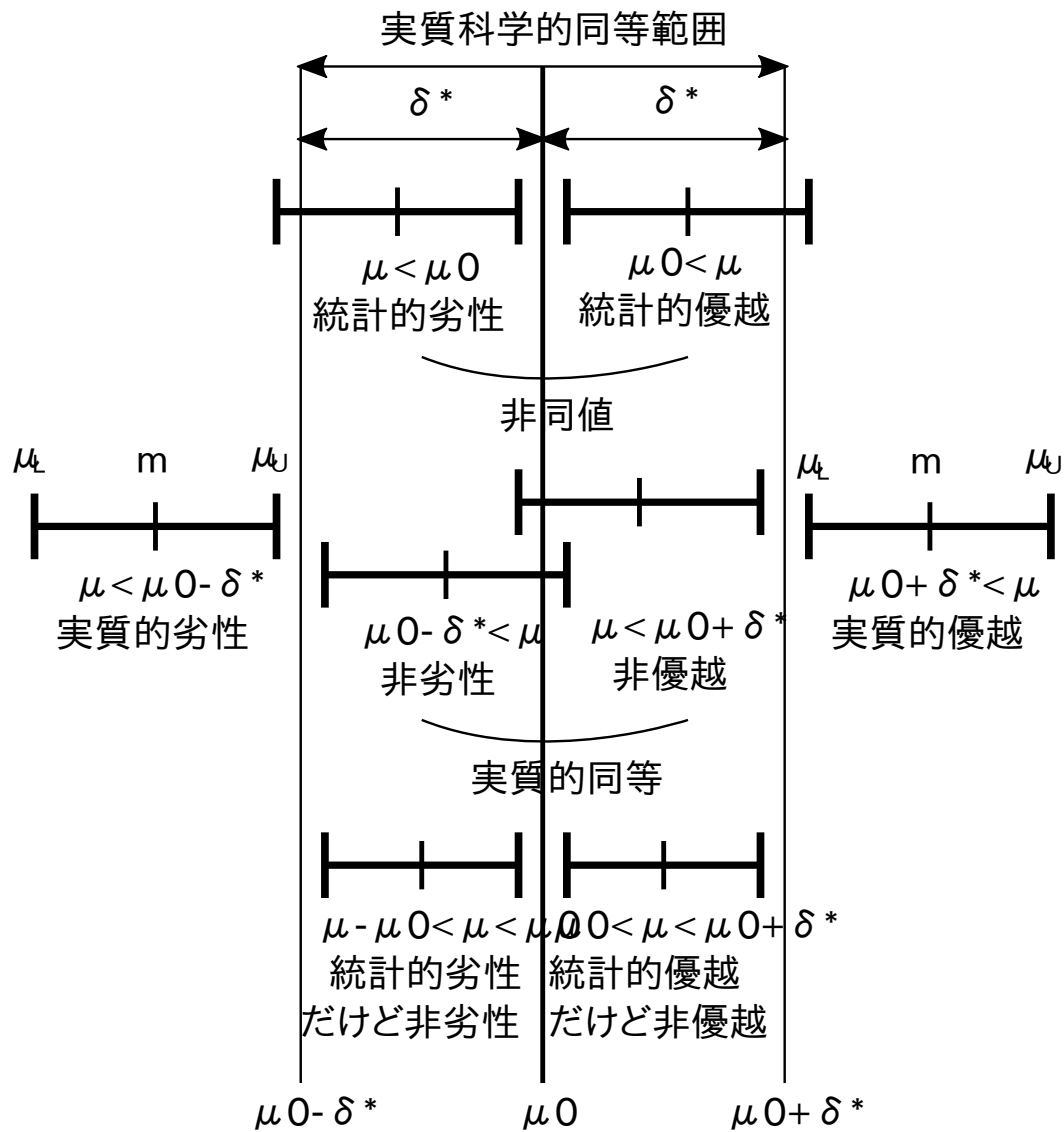


図3.13 ハンディキャップ方式の検定

ハンディキャップ方式の検定の名称

- **非同値検定**: μ_0 を基準値にした有意性検定
 $\mu_0 < \mu_L$ (95%CIの下限) → 有意: 統計的優越=非同値
 μ_U (95%CIの上限) $< \mu_0$ → 有意: 統計的劣性=非同値
- **同等性検定**: μ_0 を基準値にし95%CIの幅を δ^* 以下にした統計的仮説検定
 $\mu_0 < \mu_L$ (95%CIの下限) → 有意: 統計的優越=非同値
 μ_U (95%CIの上限) $< \mu_0$ → 有意: 統計的劣性=非同値
 $\mu_L < \mu_0 < \mu_U$ (95%CI内に μ_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
 $\mu_0 + \delta^* < \mu_L$ (95%CIの下限) → 有意: 実質的優越
 μ_U (95%CIの上限) $< \mu_0 + \delta^*$ → 有意: 実質的非優越
- **非劣性検定または劣性検定**: $\mu_0 - \delta^*$ を基準値にした有意性検定
 $\mu_0 - \delta^* < \mu_L$ (95%CIの下限) → 有意: 実質的非劣性
 μ_U (95%CIの上限) $< \mu_0 - \delta^*$ → 有意: 実質的劣性

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

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

標準薬に対する非劣性検定の非合理性

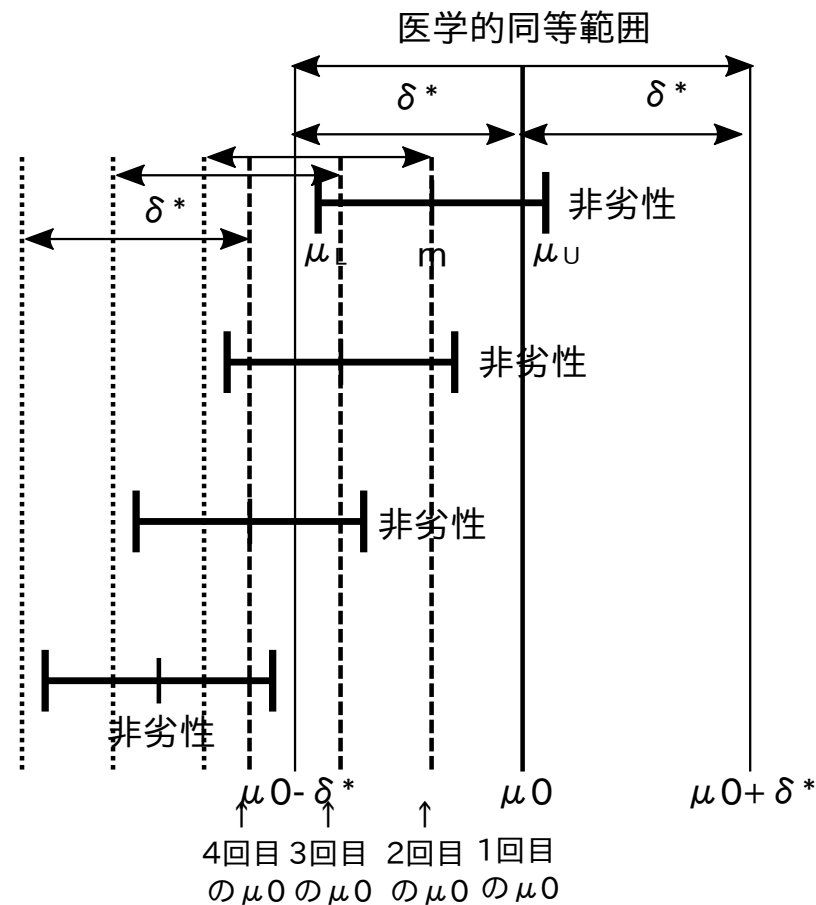


図3.14 標準薬に対する非劣性検定の非合理性

- 同等性検定の代わりに非劣性検定を行い、非同値検定を優越性検定の代わりに使用するのは欺瞞
- 新薬が標準薬と実質的に同等でも「新薬にはメリットがある」という結論が導けてしまう
 - 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

統計学の落とし穴

- 標準偏差と標準誤差
- 有意性検定と統計的仮説検定
- パラメトリック手法とノンパラメトリック手法
- ハンディキャップ方式の検定
- 相関分析と回帰分析
- 葬患係数と怪奇直線
- 選択例の相関

マウスの体長と尾長の相関分析

No.	体長	尾長
1	10	8.04
2	8	6.95
3	13	7.58
4	9	8.81
5	11	8.33
6	14	9.96
7	6	7.24
8	4	4.26
9	12	10.84
10	7	4.82
11	5	5.68
平均値	9	7.50

マウスの体長と尾長の関係を調べるために
11匹のマウスを無作為に選択して体長と尾長を測定した

横断的研究=体長にも尾長にも誤差がある

相関分析は横断的研究用

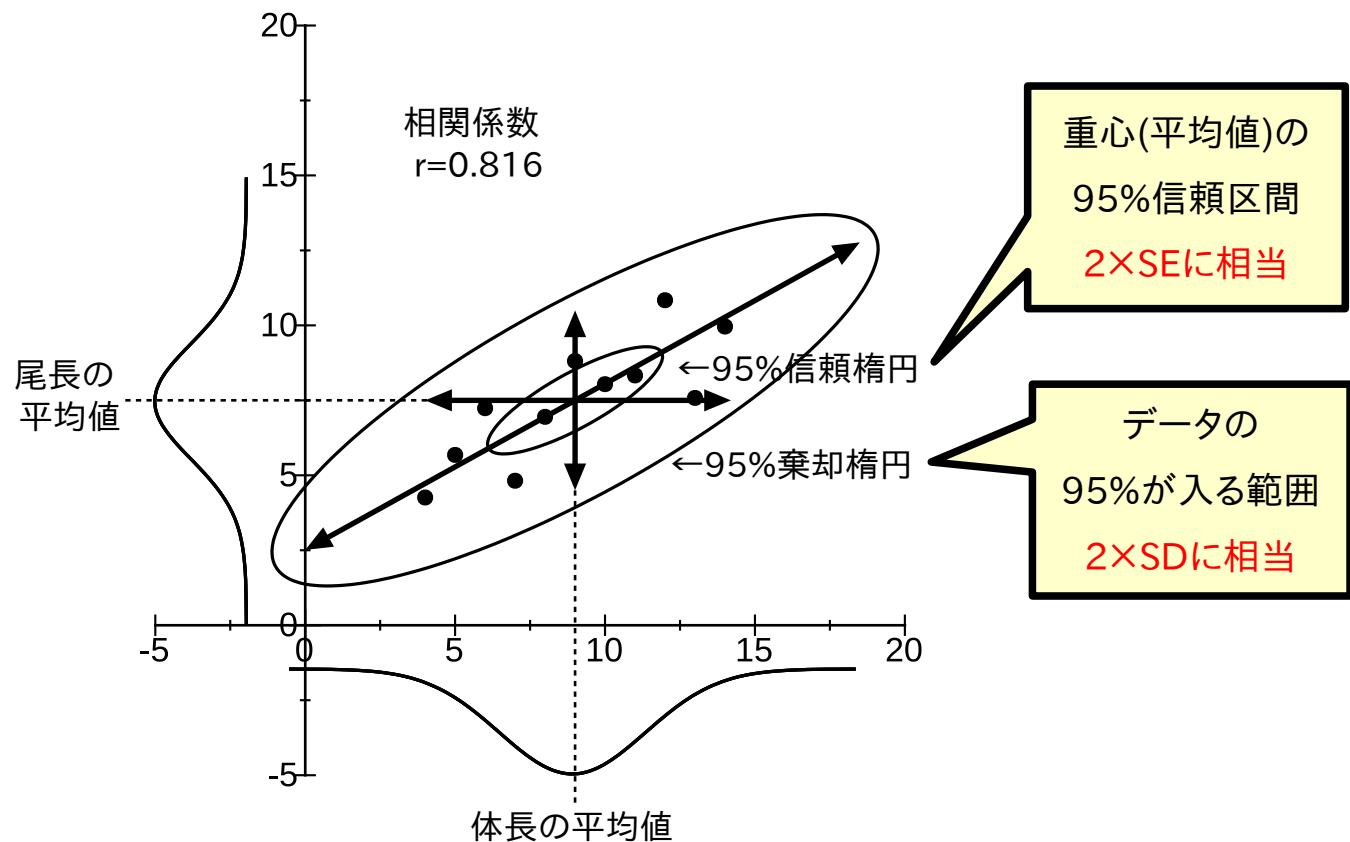


図3.1 体長と尾長の散布図

2つの項目のどちらにも誤差変動がある＝横断的研究から得られたデータ
→データは重心(2つの項目の平均値)を中心にして上下左右に変動する
→相関方向にはより変動しやすい
→2つの項目を入れ替えて計算しても結果は変わらない

相関関係と因果関係

異なる概念

相関関係

2つの項目がお互いに影響
を与え合っている状態
相関性=相互関連性

因果関係

一方の項目が原因になり
もう一方の項目に
結果として影響を与えている状態
因果関係=原因と結果の関係

相関係数: $-1 \leq r \leq +1$

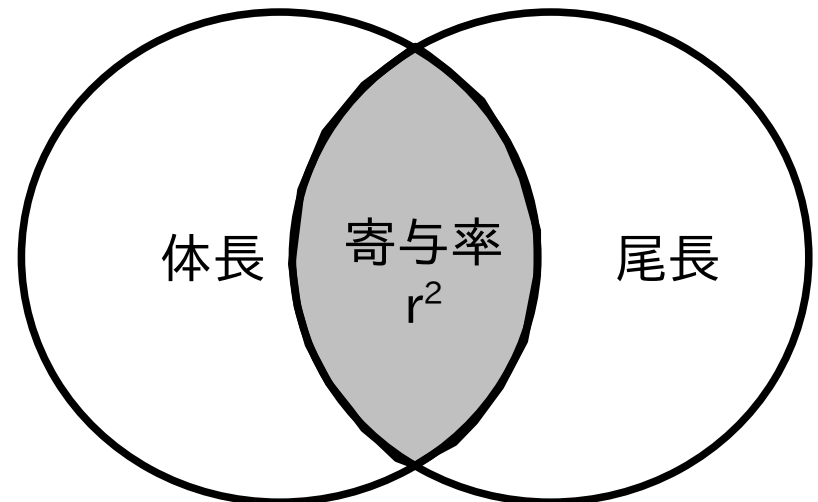
2つ項目の直線的な大小関係が
一致している程度を表す指標

関連あり: $r^2 \geq 0.5 (50\%) \rightarrow |r| \geq 0.7$

影響あり: $r^2 \geq 0.25 (25\%) \rightarrow |r| \geq 0.5$

r^2 : 寄与率 ← 共有情報量(共分散)の割合を表す

※統計学では平方した値が情報量を表す



マウスの生後日数と尾長の回帰分析

No.	x: 生後日数	y: 尾長
1	4	4.26
2	5	5.68
3	6	7.24
4	7	4.82
5	8	6.95
6	9	8.81
7	10	8.04
8	11	8.33
9	12	10.84
10	13	7.58
11	14	9.96
平均値	9	7.50

マウスの生後日数による尾長の変化を調べるために
生後4～14日のマウスを各1匹ずつ選択して尾長を測定した
前向き研究＝尾長にだけ誤差がある

回帰分析は前向き研究用

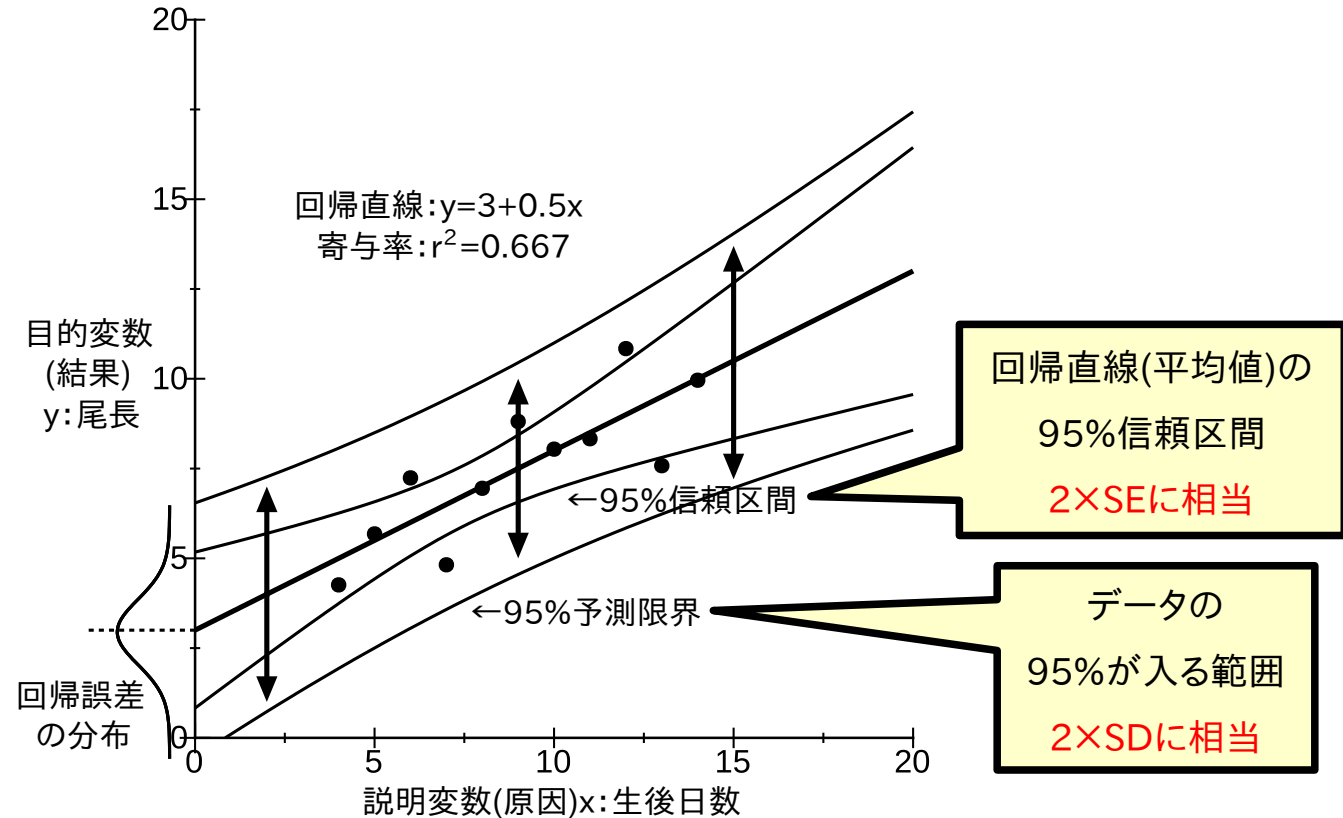


図3.2 生後日数と尾長の回帰直線

2つの項目に因果関係があり、目的変数(結果)にだけ誤差変動がある=前向き研究から得られたデータ

→説明変数(原因)の値を研究者が指定するため誤差変動はない

※xとyを入れ替えて計算すると別の回帰直線になる

→データは回帰直線を中心にして上下に変動する

※回帰直線の傾きにも誤差があるため平均値から離れるほど誤差が大きくなる

両方に誤差がある時の回帰直線

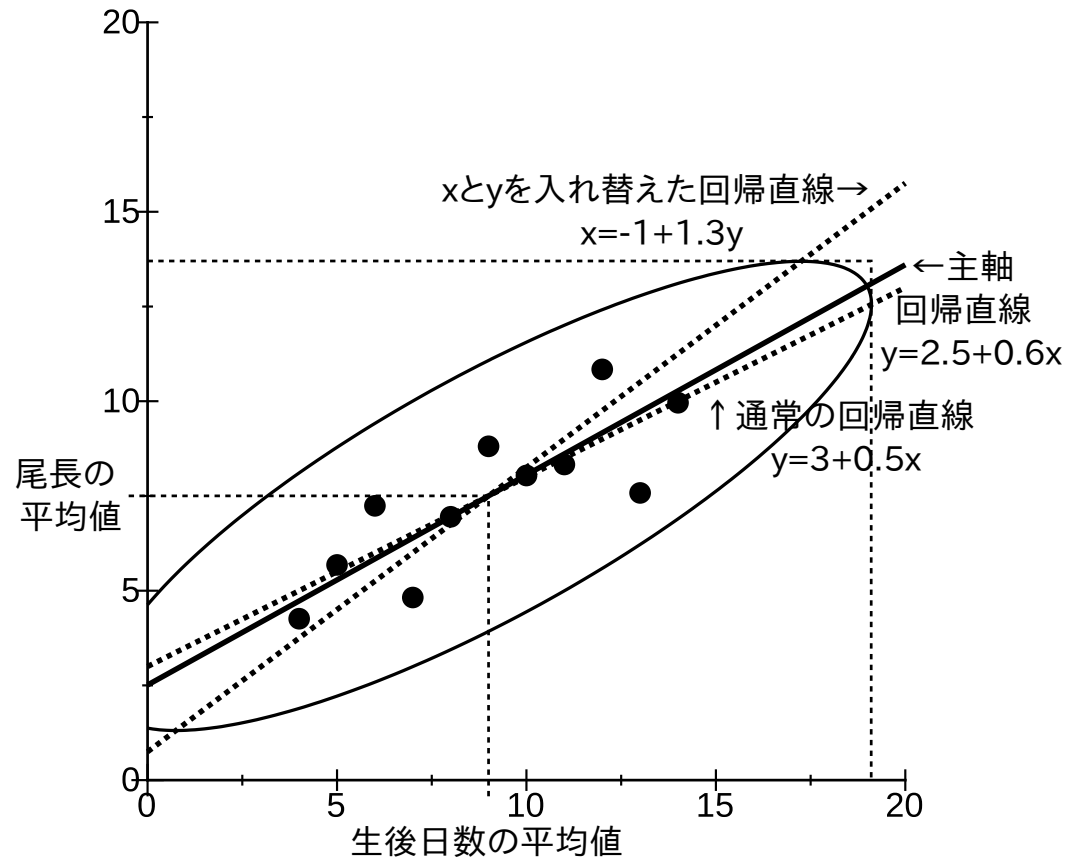


図3.3 両方に誤差がある時の回帰直線

2つの項目のどちらにも誤差変動がある＝**横断的研究から得られたデータ**

主軸回帰直線(主成分回帰直線):棄却楕円の長軸に相当(**xとyを入れ替えても変わらない**)

xからyを推測する回帰直線(yにだけ誤差がある時):棄却楕円とy軸に平行な直線の接点を通る

yからxを推測する回帰直線(xにだけ誤差がある時):棄却楕円とx軸に平行な直線の接点を通る

相関分析と回帰分析

目的が異なり
適用するデータも
異なる手法

相関分析

2つの項目の間に
相関性がある時
または相関性を想定している時
横断的研究から得られたデータ

回帰分析

2つの項目の間に
因果関係がある時
または因果関係を想定している時
前向き研究から得られたデータ

回帰直線

2つの項目の間の因果関係を直線 $y=a+bx$ で近似したもの

回帰係数: 説明変数が「1」増加した時に目的変数が平均的にいくつ変化するかを表す値

寄与率: 目的変数の全変動のうち説明変数によって説明できる変動の割合

統計学の落とし穴

- 標準偏差と標準誤差
- 有意性検定と統計的仮説検定
- パラメトリック手法とノンパラメトリック手法
- ハンディキャップ方式の検定
- 相関分析と回帰分析
- 葬患係数と怪奇直線
- 選択例の相関

葬患係数と怪奇直線

No.	x	y1	y2	y3	y4
1	10	8.04	9.14	7.46	9.42
2	8	6.95	8.14	6.77	5.18
3	13	7.58	8.74	12.74	9.08
4	9	8.81	8.77	7.11	9.53
5	11	8.33	9.26	7.81	9.30
6	14	9.96	8.10	8.84	8.97
7	6	7.24	6.13	6.08	5.40
8	4	4.26	3.10	5.39	5.62
9	12	10.84	9.13	8.15	9.20
10	7	4.82	7.26	6.42	5.29
11	5	5.68	4.74	5.73	5.51
平均値	9	7.50	7.50	7.50	7.50

xと4種類のyに関する相関係数と回帰直線

相関係数: $r=0.816$ 寄与率: $r^2=0.667$

回帰直線: $y=3+0.5x$

4種類の散布図

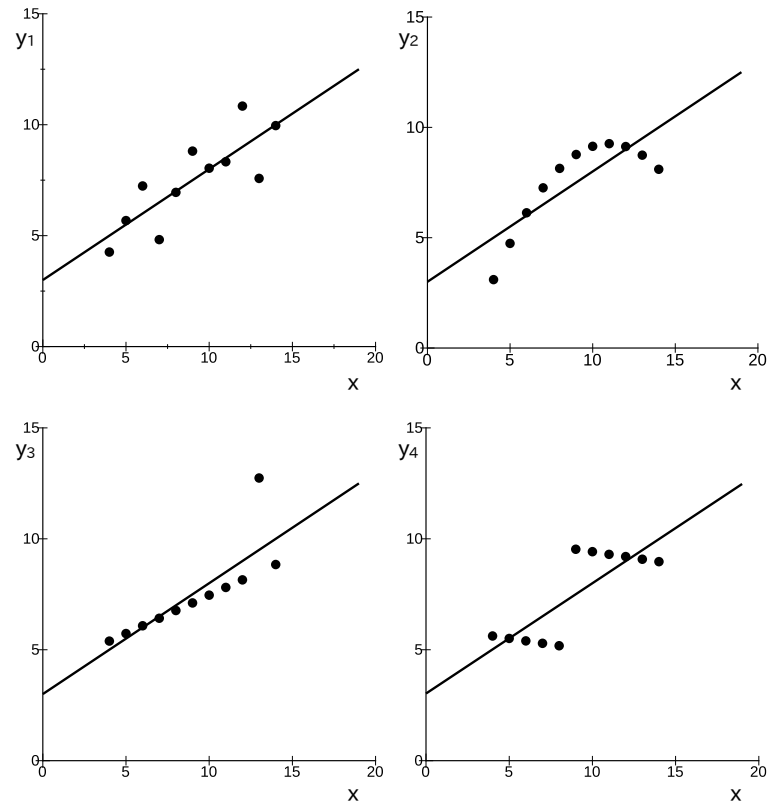


図3.4 4種類の散布図

y2: 放物線で回帰すると $y_2 = -6 + 2.78x - 0.13x^2$ $R^2 = 1$

y3: 3番目のデータ12.74を除外すると $y_3 = 4 + 0.35x$ $r^2 = 1$

※3番目のデータはx以外の要因(例えば副作用)で変動している可能性が高い(無闇に除外しない)

y4: 2,7,8,10,11番のデータはオス、残りはメスとして別々に計算すると

オス: $y_4 = 6.06 - 0.11x$ $r^2 = 1$ メス: $y_4 = 10.54 - 0.11x$ $r^2 = 1$

※これを単純化したものがシンプソンのパラドックス(層別解析の結果と全体解析の結果が矛盾する)

統計学の落とし穴

- 標準偏差と標準誤差
- 有意性検定と統計的仮説検定
- パラメトリック手法とノンパラメトリック手法
- ハンディキャップ方式の検定
- 相関分析と回帰分析
- 葬患係数と怪奇直線
- 選択例の相関

薬剤投与前後の収縮期血圧と拡張期血圧

No.	収縮期血圧		拡張期血圧	
	投与前	投与後	投与前	投与後
1	160	130	87	81
2	150	121	92	75
3	142	119	96	79
4	141	119	89	78
5	140	120	90	80
6	139	116	91	77
7	139	140	88	90
8	136	109	82	72
9	129	104	85	86
10	124	122	100	82
平均値	140	120	90	80

降圧剤の効果を調べるために次のような症例だけ選択

収縮期血圧が120mmHg以上または拡張期血圧が80mmHg以上

選択例の相関

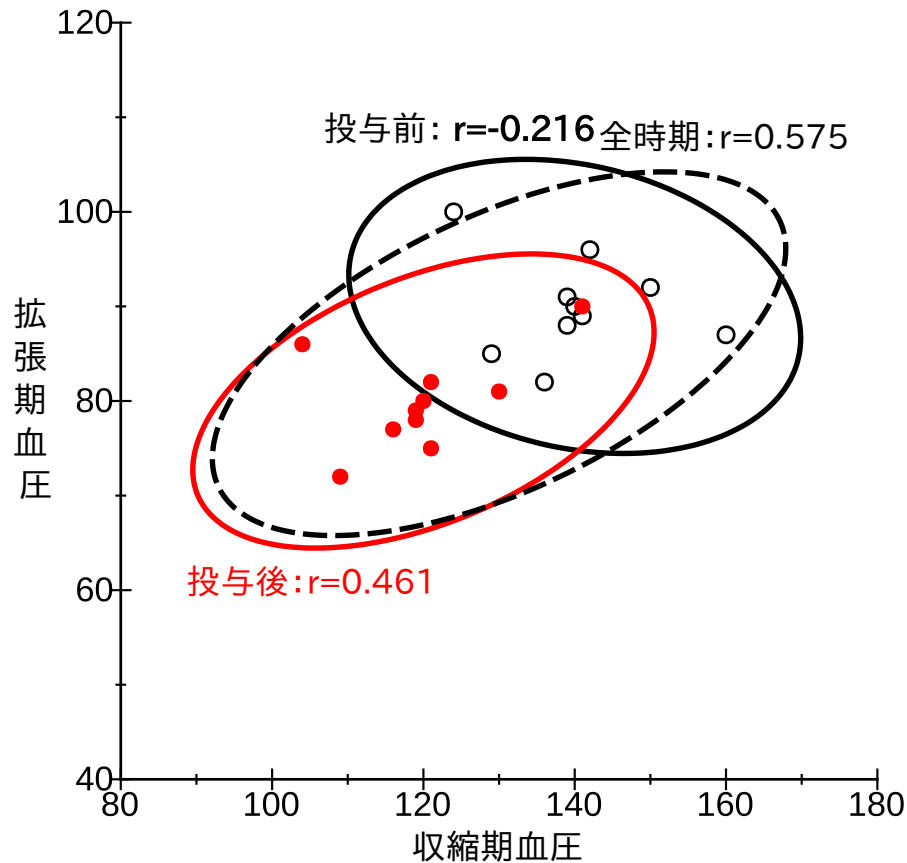


図3.5 血圧の全時期分布

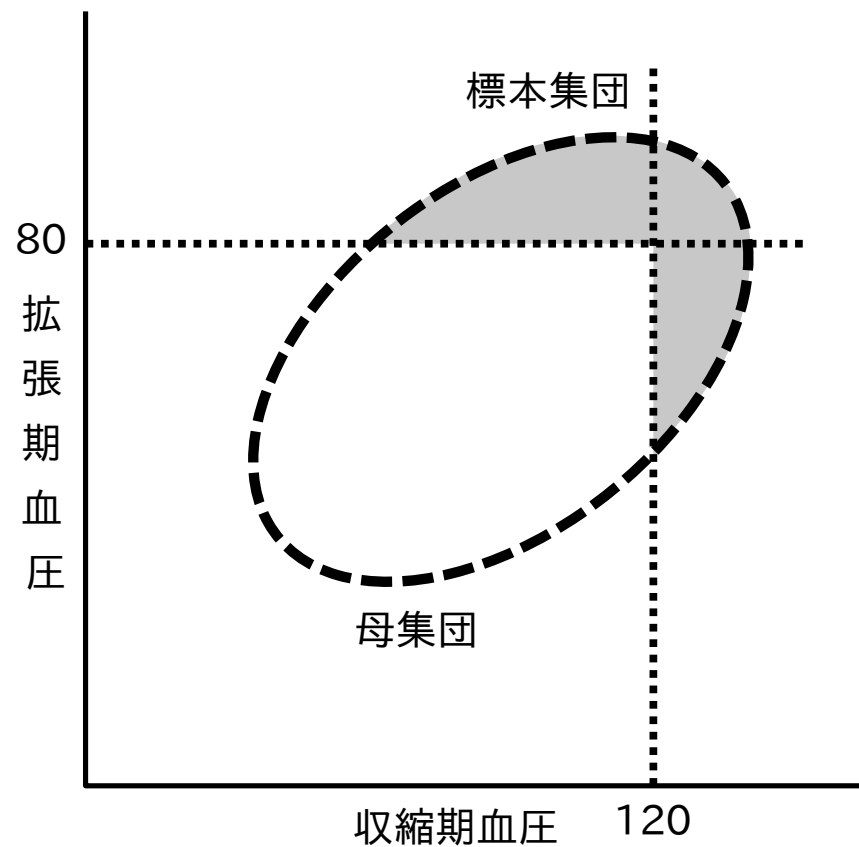


図3.6 血圧の初期分布

どちらかの項目をコントロールしたデータは
自然状態の相関関係が崩れている可能性がある
→ 正確な相関係数を求めることはできない!

本日の結語

ノンパラメトリック手法と

非劣性検定はできるだけ使わず

推定結果を重視しましょう!

ご清聴ありがとうございました