

**総合薬学講座
生物統計の基礎**

内田 吉昭 (神戸薬科大学薬学部)

2011年10月26日

1. 帰無仮説の概念を説明できる.
2. パラメトリック検定とノンパラメトリック検定の使い分けを説明できる.
3. 主な2群間の平均値の差の検定法 (t検定、**Mann-Whitney U検定**)について、適用できるデータの特徴を説明し、実施できる. (知識・技能)
4. χ^2 検定の適用できるデータの特徴を説明し、実施できる. (知識・技能)
5. 最小二乗法による直線回帰を説明でき、回帰係数の有意性を検定できる. (知識・技能)
6. 主な多重比較検定法 (分散分析、**Dunnnett検定**、**Tukey検定**など) の概要を説明できる.
7. 主な多変量解析の概要を説明できる.
8. 基本的な生存時間解析法 (Kaplan-Meier 曲線など) の特徴を説明できる.

赤字は教科書に載っていない.

今回が6年生の初めての国家試験なので、
問題に関する傾向と対策はよくわからない

青本のpp.662–665の問題と
教科書の付録Aを勉強していれば十分かもしれない

今回が6年生の初めての国家試験なので、
問題に関する傾向と対策はよくわからない

青本のpp.662–665の問題と
教科書の付録Aを勉強していれば十分かもしれない

えっ、教科書捨てちゃったって？ 知らんよ。そんなこと。

今回が6年生の初めての国家試験なので、
問題に関する傾向と対策はよくわからない

青本のpp.662–665の問題と
教科書の付録Aを勉強していれば十分かもしれない

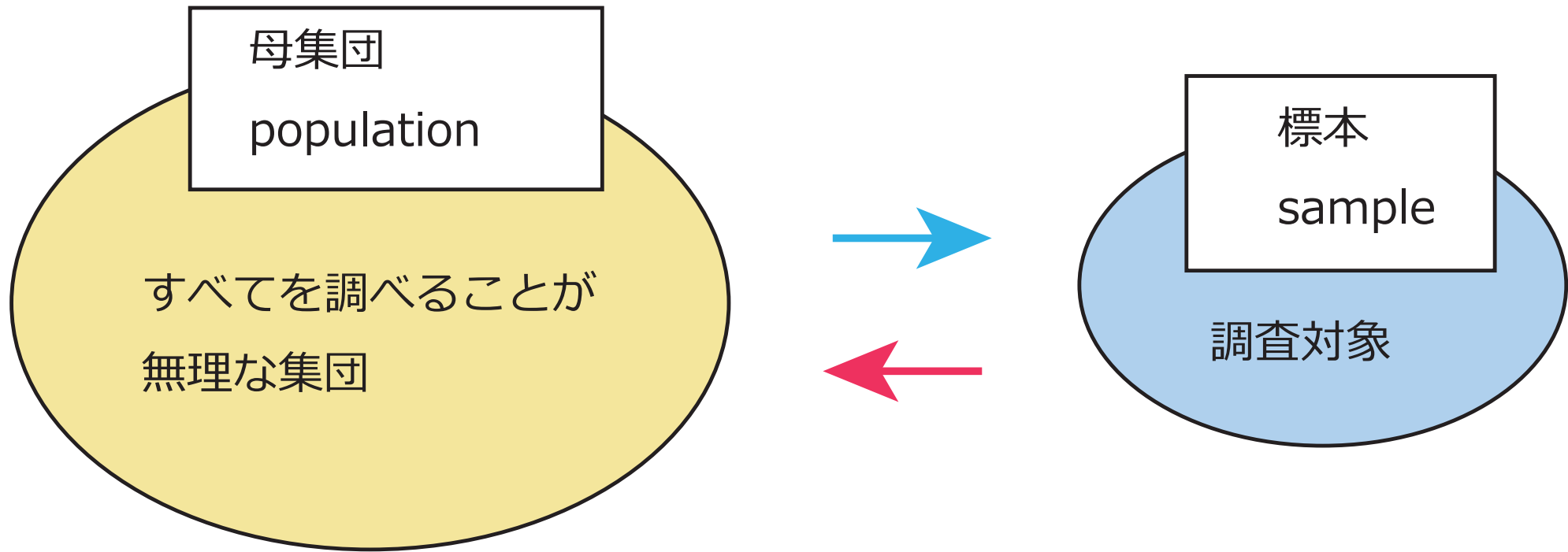
えっ、教科書捨てちゃったって？ 知らんよ。そんなこと。
青本の問題はやったほうがよいと思う。

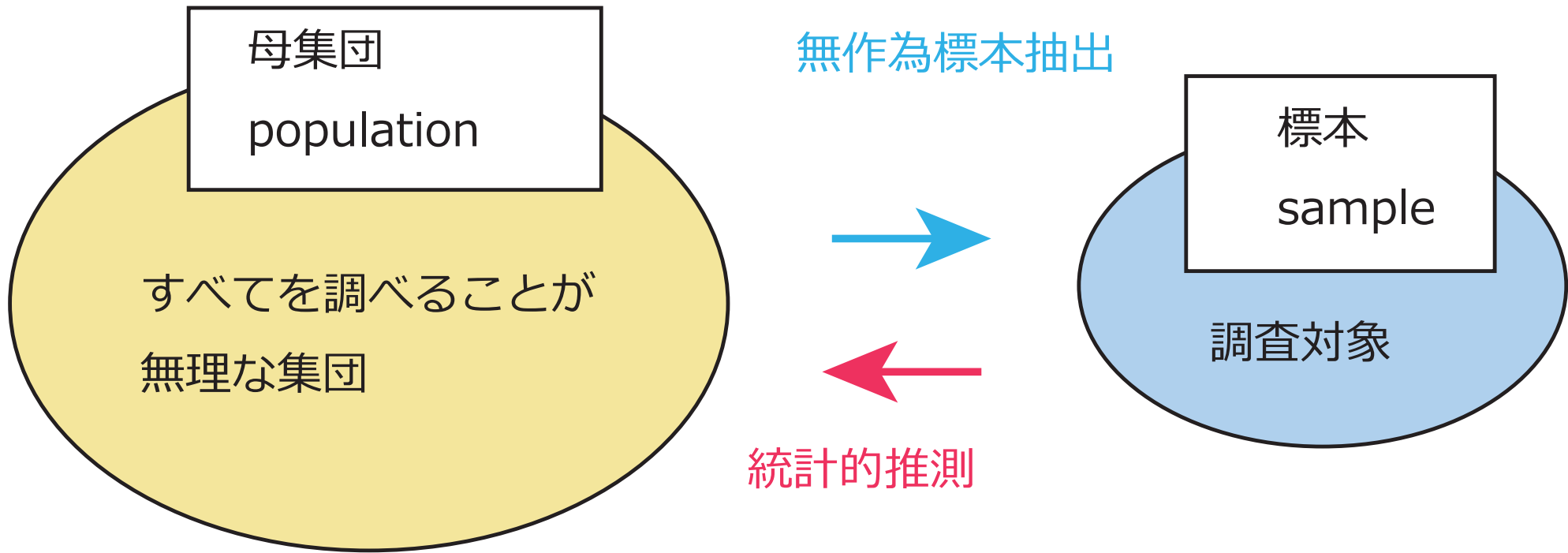
生物統計 統計学を生物学(医学・薬学を含む)に応用.
特に重要なのが「臨床試験」

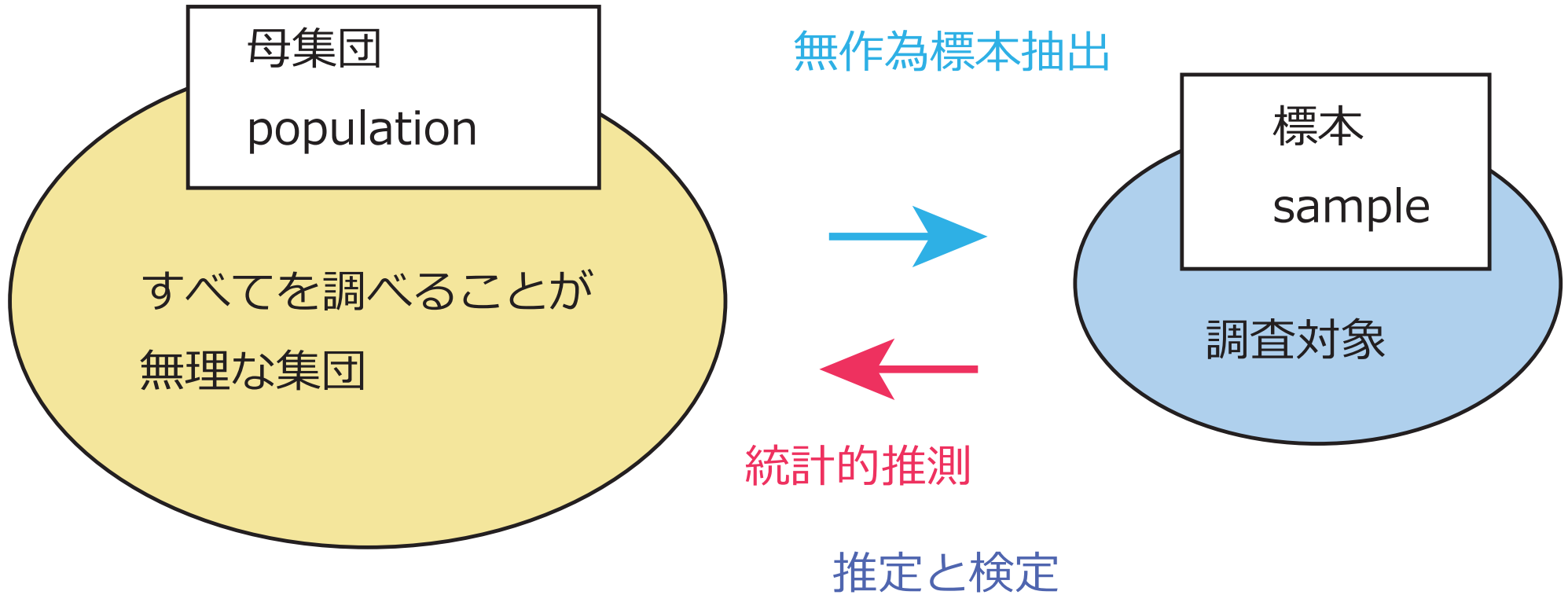
このpdfファイルをホームページに載せておきます.
大学→研究室→数学研究室→研究室のホームページ

■または内田吉昭で **google検索** で一発

1) 母集団と標本







2) 推定

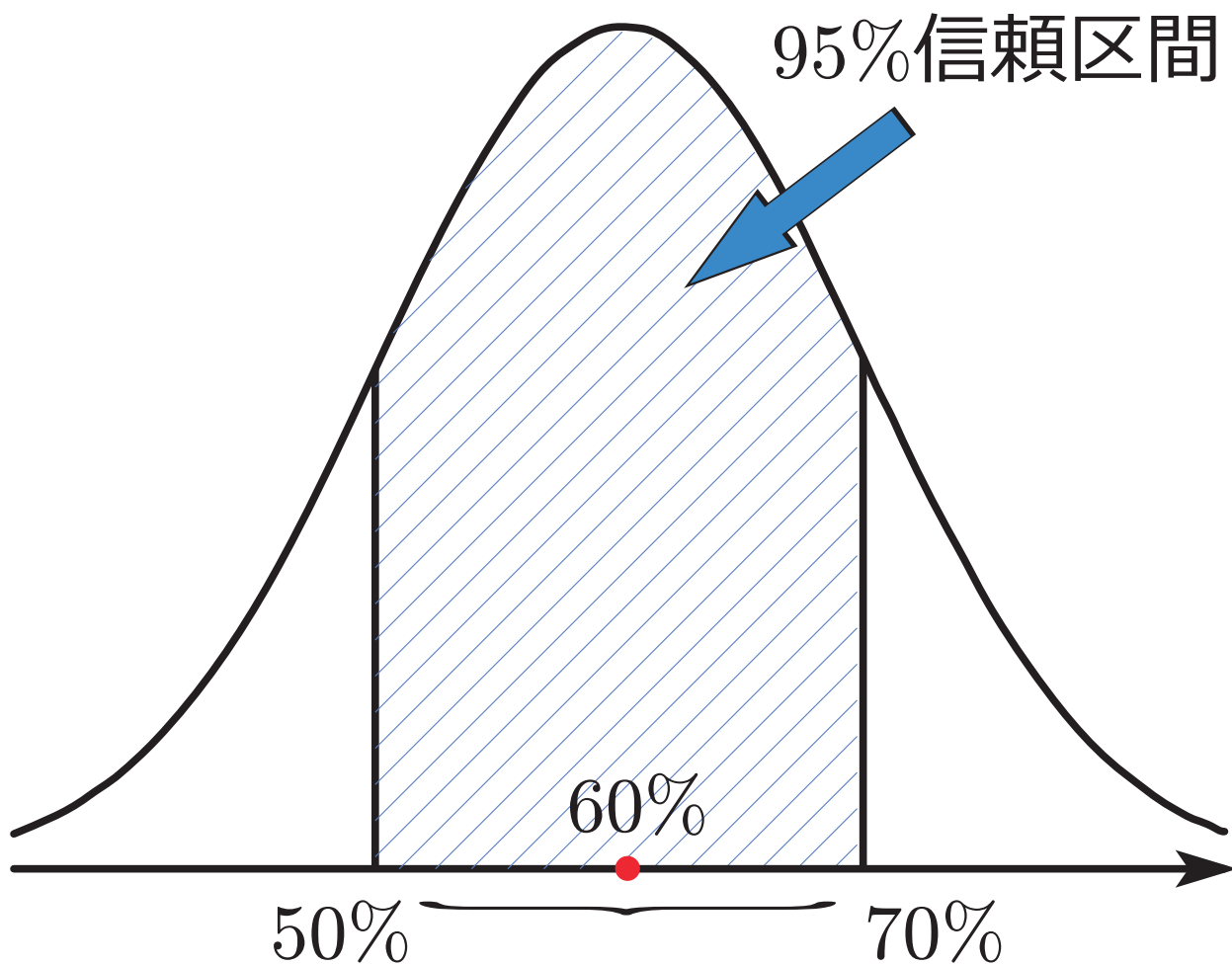
母集団 ➡ 100人の患者(標本)
薬剤Aを投与 60人に効果
有効率 $60\% = \frac{60}{100}$

100人の患者(標本)
薬剤Aを投与 60人に効果
母集団 ➡ 有効率 $60\% = \frac{60}{100}$

点推定 (1つの値だけで評価)

母集団の値と近いけれど, どれくらい近いかがわからない.
母集団の比率は62.5%かもしれない.

区間推定



この区間が母比率を含む確率が95%

母集団と標本の評価が**確率**でできる. ⇒ 教科書 第4章

3) 検定

100人の標本(患者)に対して

薬剤Aを投与 ➡ 60人に効果

従来の薬剤Bの有効率は50%

薬剤Aの有効率は

薬剤Bより**本当**に高いか

もしかすると、薬剤Aの**本当の有効率**は45%なのだが、
偶然標本100人のうち60人に効いたのかもしれない。

検定を行なって調べる。

1.5.1 帰無仮説の概念

H_1 (対立仮説) 自分が主張したいこと

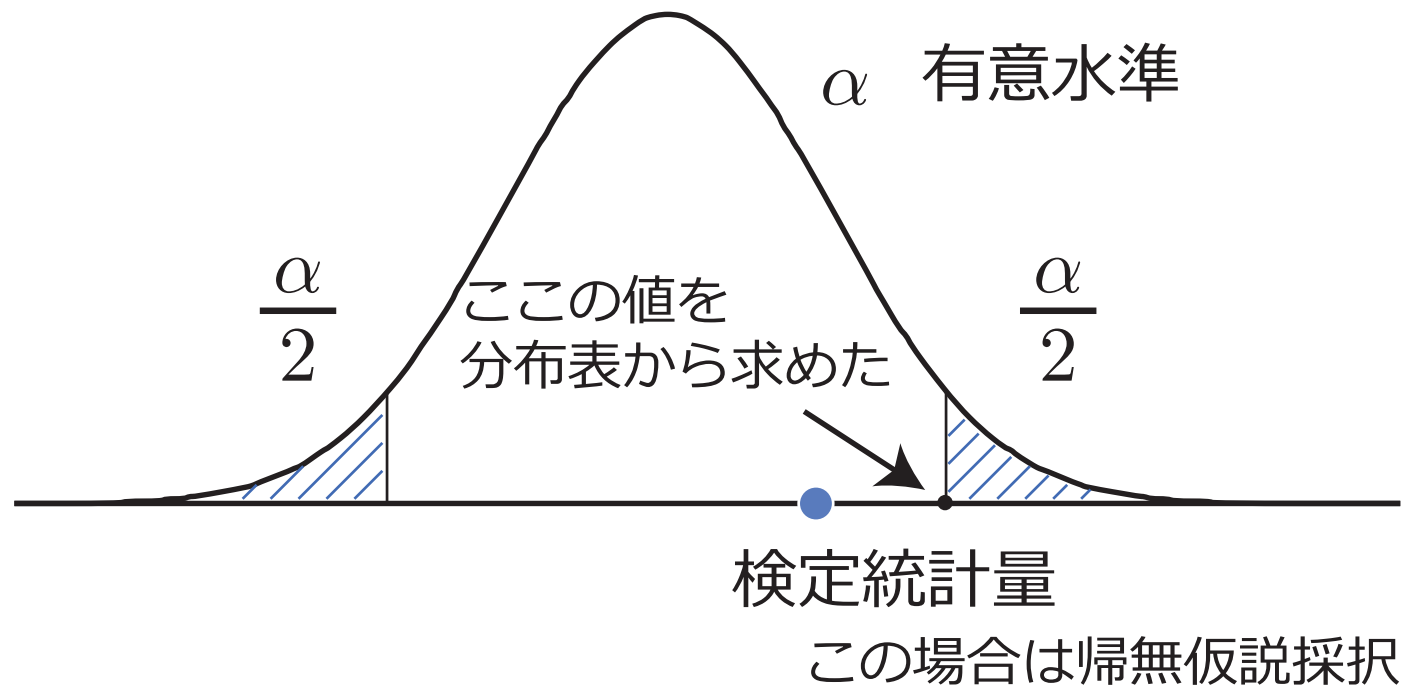
H_0 (帰無仮説) それとは反対のこと

有意水準(危険率) α

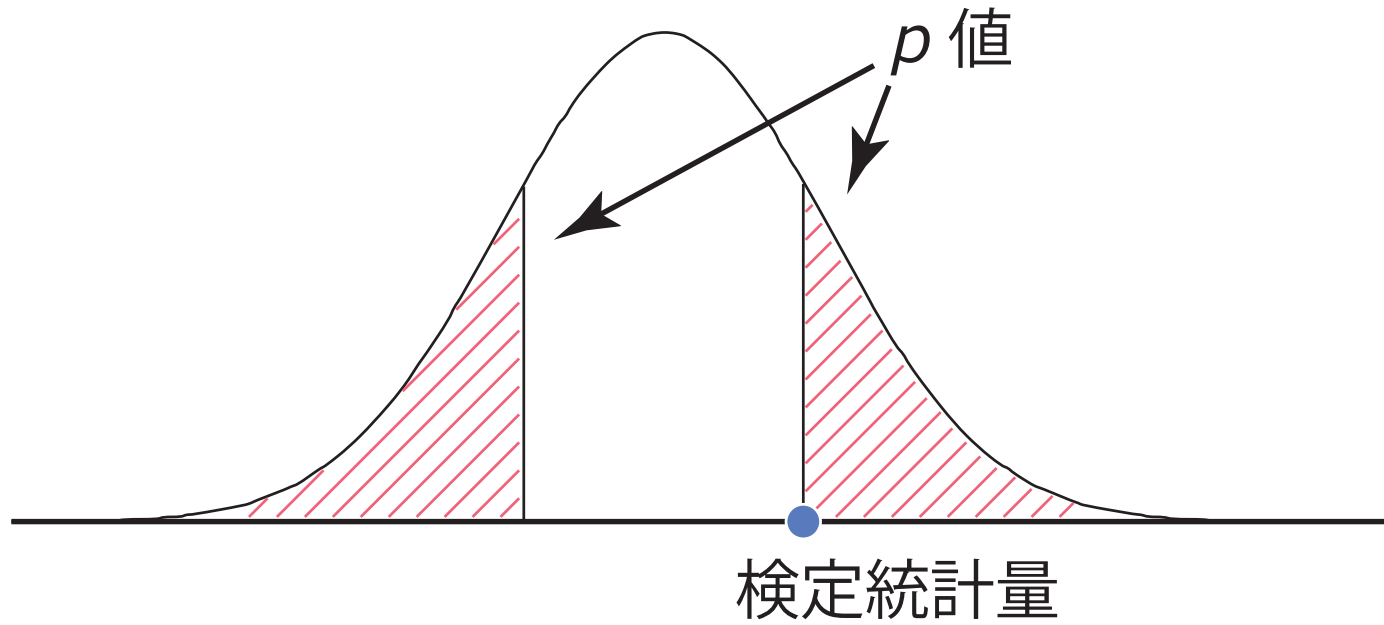
標本のデータから検定統計量を計算し p 値を求める.

注意 教科書では α の検定統計量を求めて比較した.

p 値とは,
帰無仮説のもとで, 標本から得られた検定統計量より
標本平均値 \bar{x} がより極端な値をとる確率



検定統計量より極端な値を取る確率



p.650 の注意

帰無仮説は等号で表される ($\mu = \mu_0$)

対立仮説は \neq または不等号で表される ($\mu \neq \mu_0$)

p 値が0.05未満 (5%未満) であれば小さい (統計的に有意) と判定される. \Rightarrow 有意水準 $\alpha = 0.05$ と同じ

このことは、誤って帰無仮説を棄却することが5%存在していることを意味する

検定では、対立仮説を支持するという結論は得られるが、
帰無仮説を積極的に支持するという結論は原則として得られ
ない. p.651 l.1-4 マーカー

そのために採択されたときには、対立仮説を棄却して
「 H_1 であるとはいえない」という表現をする

$$\begin{cases} H_0 : \text{棄却} & \Rightarrow & H_1 \text{である.} \\ H_0 : \text{採択} & \Rightarrow & H_1 \text{であるとはいえない.} \end{cases}$$

結論は H_1 を使って述べれば良い.

教科書 p.42 表5.1

| | $\mu = \mu_0$ (H_0 正しい) | $\mu \neq \mu_0$ (H_0 間違い) |
|-------------------------|----------------------------|-------------------------------|
| 有意差なし (H_0 棄却できず) | 正しい | 第2種の過誤 β |
| 有効差あり (H_0 を棄却) | 第1種の過誤 α | 正しい |

α (有意水準) : H_0 が正しいのに あわてて捨ててしまった
あわて者の誤り

β : H_0 が間違っているのに捨てないといけないのに
ぼんやり (ベーたーとして)として捨てなかった誤り

1.5.2 パラメトリック検定とノンパラメトリック検定の使い分け 教科書 第9章

パラメトリック手法 母集団に対して例えば正規分布のようなある特定の分布を仮定した手法
母集団に**パラメーター**が入っている

ノンパラメトリック手法 母集団に対して特定の分布を仮定しない手法

パラメトリック検定は母集団が**正規分布**であることを仮定する事が多い

1) 正規分布については教科書 p.21-23 を参照せよ

1.5.3 t 検定, Mann-Whitney U 検定

データの種類 p.653 データの種類と特徴 マーカー

| | |
|-------|---|
| 連続データ | 測定可能な連続量で表されるデータを連続データという. |
| 順序データ | グループ間の順序が意味を持つような幾つかのグループに分けられたデータを順序データという |
| 分類データ | 順番のない幾つかのカテゴリー(分類)に分けられたデータを分類データという |

〈平均値, 分散, 標準偏差, 標準誤差〉

データ $\{x_1, x_2, x_3, \dots, x_n\}$

$$\text{平均 } \bar{x} = \frac{x_1 + x_2 + x_3 + \dots + x_n}{n}$$

$$\text{分散 } s^2 = \frac{(x_1 - \bar{x})^2 + (x_2 - \bar{x})^2 + \dots + (x_n - \bar{x})^2}{n - 1}$$

$$\text{標準偏差 } s = \sqrt{\frac{(x_1 - \bar{x})^2 + (x_2 - \bar{x})^2 + \dots + (x_n - \bar{x})^2}{n - 1}}$$

標準偏差 (SD:Standard Deviation)は標本のばらつきを表す.

$$\text{標準偏差 } s = \sqrt{\frac{(x_1 - \bar{x})^2 + (x_2 - \bar{x})^2 + \cdots + (x_n - \bar{x})^2}{n - 1}}$$

標準誤差 (SE : Standard Error)は標本平均 \bar{x} (または統計量)のばらつきを表す.

$$SE = \frac{s}{\sqrt{n}}$$

教科書 p.8 参照

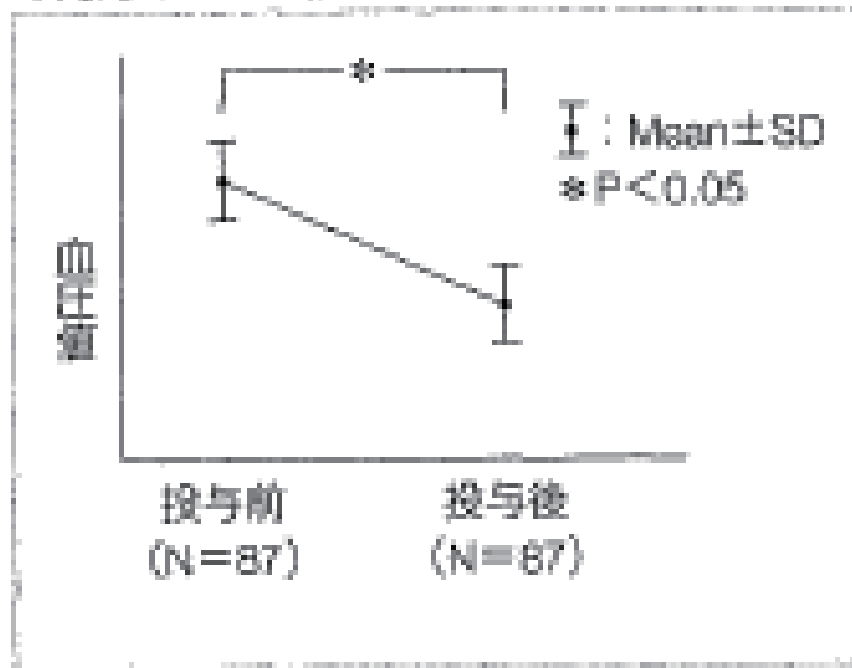
1) t 検定 (パラメトリック検定)

対応のない t 検定-薬剤Aと薬剤Bは別々の患者に投与され、各患者からは1つのデータだけが得られる。

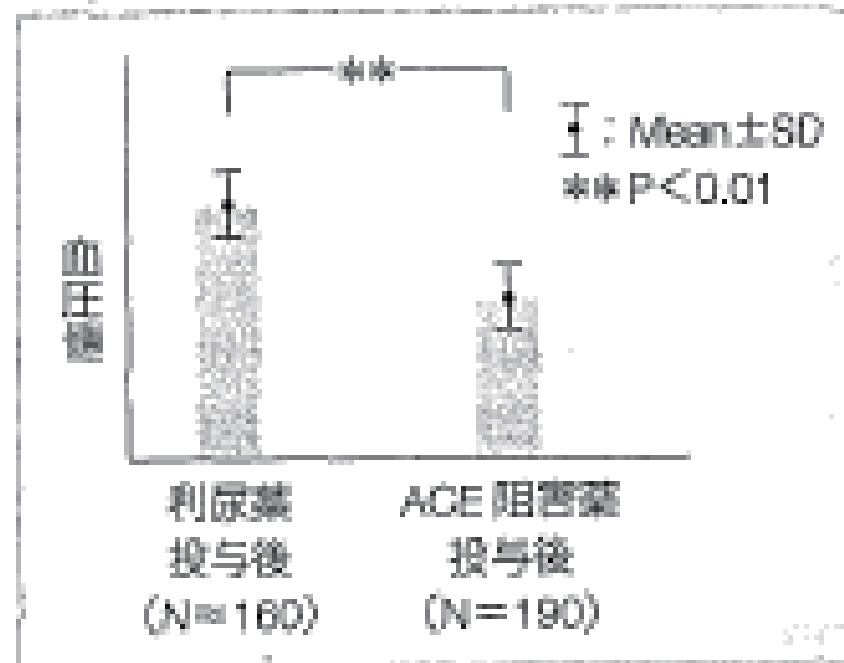
教科書 p.47 5.3.2

対応のある t 検定-標本の各々の患者に時期を違えて薬剤Aと薬剤Bを投与する方法。教科書 p.49 5.4

●対応のあるt検定



●対応のないt検定



対応のないt検定の計算式は難しい(p.47 (5.9)-(5.10))が、計算するソフトがあるので意味を理解すれば良い。

H_0 (帰無仮説) : 母集団における薬剤A又はB投与後の血圧値の平均値には**差がない** 帰無仮説は差がないに注意

H_1 (対立仮説) : 母集団における薬剤A又はB投与後の血圧値の平均値には**差がある**

2) Wilcoxon 順位和検定 (Mann-Whitney U 検定)

教科書 p.103 の Wilcoxon の符号付き順位和検定とは異なる
特定の分布を仮定しないノンパラメトリック検定である。

定義等を知りたい学生は後で研究室に来るように

H_0 (帰無仮説) : 母集団における薬剤 A 又は B 投与後の血圧
値の平均値には差がない 差がないに注意

H_1 (対立仮説) : 母集団における薬剤 A 又は B 投与後の血圧
値の平均値には差がある

1.5.4 χ^2 検定 教科書 第6章

χ^2 検定は観測値と期待値の差を見ている。

ノンパラメトリック検定である

2 × 2分割表 (cf. イエーツの補正 教科書 p.64 (6.5))

H_0 (帰無仮説) : 母集団における薬剤A又はBによる副作用発生率に**差はない** 差はない

H_1 (対立仮説) : 母集団における薬剤A又はBによる副作用発生率には**差がある**

χ^2 検定に使う分布は χ^2 分布である

訂正

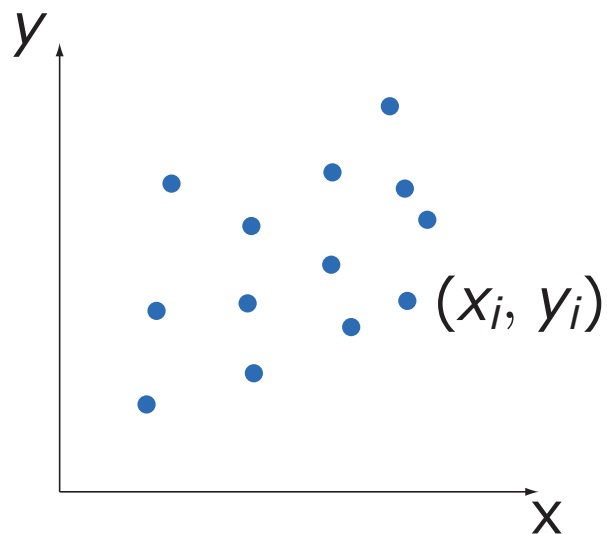
p.657 l.12 計算された検定統計量が正規分布に

⇒ 計算された検定統計量が χ^2 分布に

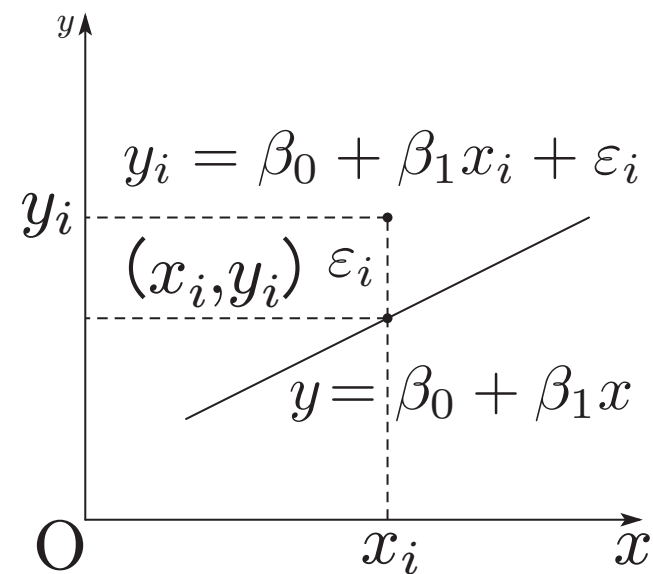
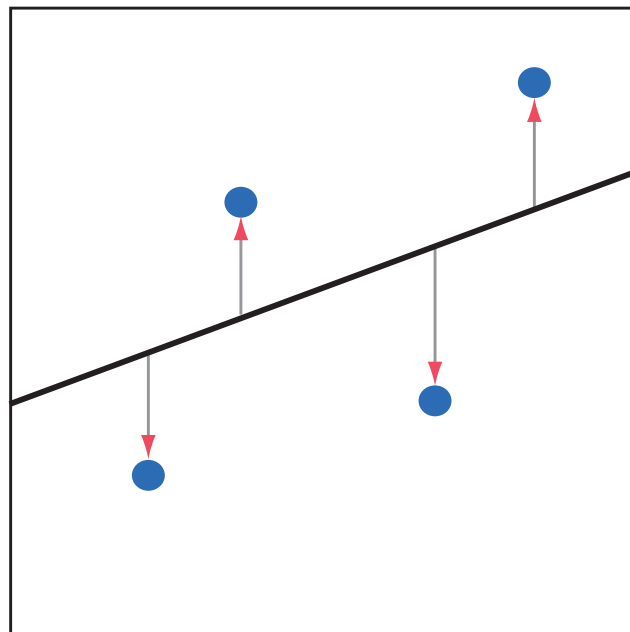
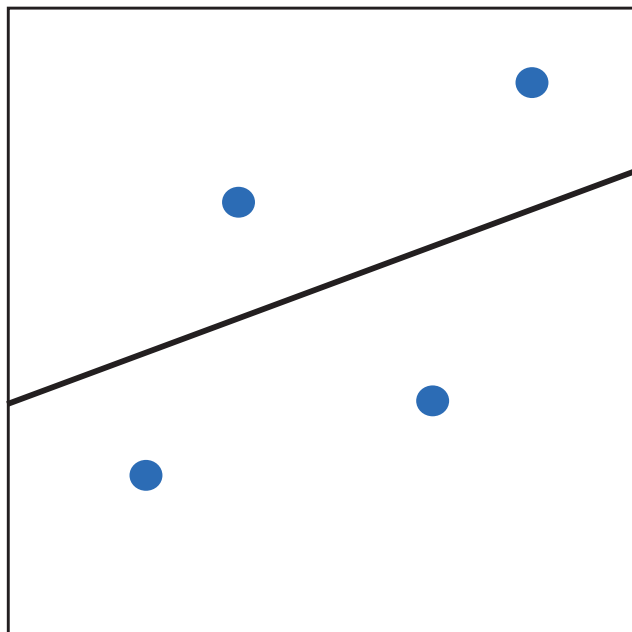
1.5.5 最小2乗法による直線回帰

回帰直線は進化論で有名なダーウィンの従兄弟である
ゴールトンによって発見された。

体重と身長など2つの変量 (x, y) で与えられたデータ。
相関図



回帰直線 $y = \beta_0 + \beta_1 x$ で近似



β_0 回帰直線の y 切片

β_1 傾き

ε_i 誤差項

最小2乗法

点と直線の y 軸方向の差は $\varepsilon_i = y_i - (\beta_0 + \beta_1 x_i)$ より, 差の平方和は

$$\delta = \varepsilon_1^2 + \varepsilon_2^2 + \cdots + \varepsilon_n^2$$

となる. この値が最小になるように β_0 , β_1 を求める

H_0 (帰無仮説) : $\beta_1 = 0$ 等号で表示

H_1 (対立仮説) : $\beta_1 \neq 0$

t 分布を使って検定.

1.5.6 主な多重比較検定法（分散分析, Dunnett 検定, Turkey 検定など）

多重比較 例えば有意水準 α で群 (A, B, C) の検定で A-B, A-C, B-C と3回すると第1種の過誤の確率がほぼ 3α と上がってしまう。

差がない群間にも有意差が生じやすくなる。

⇒ 多重性 (multiplicity) の問題

第1種の過誤の確率を制御する方法 ⇒ 多重比較法

1) 分散分析法 教科書 p.75

H_0 (帰無仮説) : 4群の母集団における平均値はすべて等しい

H_1 (対立仮説) : 4群の母集団における平均値はすべて等しいとはいえない(等しくない組み合わせがある)

H_0 :棄却 \Rightarrow 母集団における平均値に異なるものがある

どの対の平均値が異なるのかはわからない

2) Bonferroni法

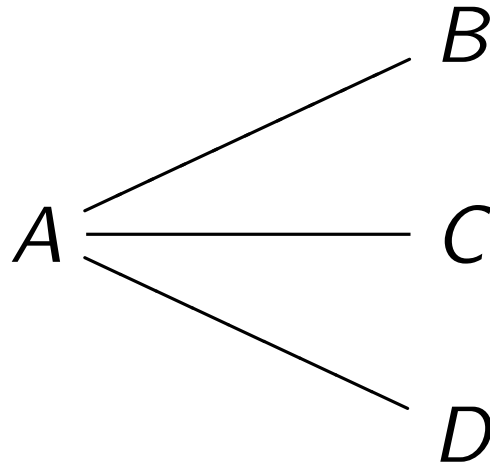
3群の比較では合計3回の検定が必要

全体の有意水準を0.05としたい場合、各比較の有意水準を
 $0.05 \div 3 = 0.01$ として検定を行う

1. 単純でわかりやすい
2. かなり保守的
3. 検出力が低下する欠点

3) Dunnett 法

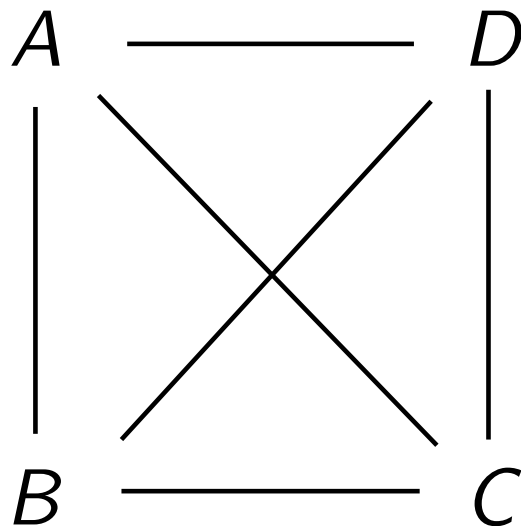
A群（コントロール群）とそれ以外の群（B, C, D）



対ごとに平均値の差を検定するために，検定統計量及び有意水準の棄却限界値を定める．

全体としての第一種の過誤を制御しながら検定を行う

4) Tukey 法



対ごとに平均値の差を検定するために、検定統計量及び有意水準の棄却限界値を定める。

全体としての第一種の過誤を制御しながら検定を行う

1.5.7 主な多変量解析の概要

医学・薬学のデータ

多くの変数(年齢・性別・体重・身長,...e.t.c.)

説明変数(原因となる事柄に関する変数)

⇒ 応答変数(結果となる事柄に関する変数)

1) 主成分分析

多くの説明変数 ⇒ 少数の無相関な合成変数に縮約して分析

2) 因子分析

多くの説明変数から少数の変数(因子)によって説明

3) 重回帰分析

(単)回帰分析を2つ以上の説明変数に拡張したもの

1.5.8 生存時間解析法

1) Kaplan-Meier法 教科書 p.109参照

Kaplan-Meier曲線

2) ログランク (Log-rank) 検定

(コクラン)マンテル・ヘンツェルと同じ

教科書 p.113参照

生存-死亡の複数の 2×2 分割表を併合して検定を行う.

ノンパラメトリック検定

3) Cox比例ハザードモデル

生存時間に影響を及ぼすと考えられる共変量(試験参加者の個人的な特徴(年齢、性別など)や疾患歴・治療歴など)の影響を含めて生存時間をモデル化し、2群間の生存状況を比較したり、共変量の影響を検討するための手法

比例ハザードの説明はわからなくても(今は)よい。
(説明にかなり時間が掛かる)

確認問題練習問題過去の国家試験問題を見て

標準偏差 ⇒ データのばらつきを表す

標準誤差 ⇒ 検定統計量のばらつきを表す

ある検定がパラメトリック検定かノンパラメトリック検定を
答えさす

連続データはどれか(値がアナログで表示される-アナログ
メモリの体重計・身長計など)

試験対策としては, pp.662-665 の問題を行えばよいだろう