総合薬学講座 生物統計の基礎

内田 吉昭

神戸薬科大学

2014 年 10 月 22 日

内容

青本 p.697

目標

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性を説明し、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性を説明し、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、Dunnett 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.

赤字は味村先生の教科書には載っていない. 香川のいるマン U と覚えれば良い.

問 **67** 正規分布が仮定できる数値データについて、2 群間の平均値の差の検定に用いる統計手法はどれか。**1** つ選べ。

1. 符号検定
2. カイ二乗検定
3. Student の t 検定
4. Fisher の直接確率法
5. Wilcoxon の順位和検定

解答　3 (できないと困る問題)

パラメトリック検定 (正規母集団)

3　Student の t 検定 (t は Student の最後の文字の t) 等分散を仮定

$\Rightarrow$ ウェルチの t 検定 等分散を仮定しない

$\Rightarrow$ t 検定は (母正規分布の) 平均値の検定

ノンパラメトリック検定 (母集団は正規母集団でなくて良い)

1　符号検定　母中央値 $= \xi_0$ (または 0) か
符号 $\pm$ の個数を調べる

2　カイ二乗検定　期待値と観測値のズレを調べる

4　Fisher の直接確率法　2×2 分割表の確率を直接計算する

5　Wilcoxon の順位和検定　順序も考えた符号検定

93 回 (平成 20 年)

問 230 データ解析に関する記述のうち正しいものの組合せはどれか.

- a t 検定はパラメトリック検定である
- b ノンパラメトリック検定ではデータが正規分布していなければならない
- c χ^2 検定は被験薬の投与群と非投与群との比較に用いられる
- d 最小二乗法により求められる相関係数は **−1.0** から **+1.0** の範囲の値として得られる
- e 有意水準とは, 対立仮説を棄却する確率のことである

b 誤り (ノンパラメトリック検定は正規分布でなくて良い)
e 誤り (有意水準とは帰無仮説を棄却するかどうかを判定する基準)

勉強方法

生物統計から,
国試に出題される量は多くない. (衛生薬学などの範囲が少しある)

⇒ 森脇先生担当の多重比較検定等がすこし出題.

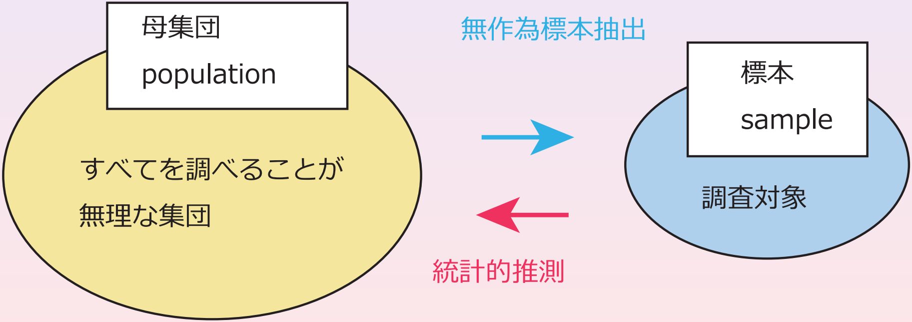
98 回国試
問 19 無作為化比較試験・横断研究・コホート研究　　問 67 交絡要因
問 68 EBM　　問 126 オッズ比　　問 292 単盲検試験

青本の pp.717–720 の問題と
国試過去問 (過去 10 年だと上の 2 題+α) を勉強していれば十分かもしれない

1) 母集団と標本

無作為 (ランダム) が重要

標本から全体像の母集団を推測する.



2 推定と信頼区間

青本 p.698

2) 推定

母集団 ➡

100 人の患者 (標本)
薬剤 A を投与 60 人に効果

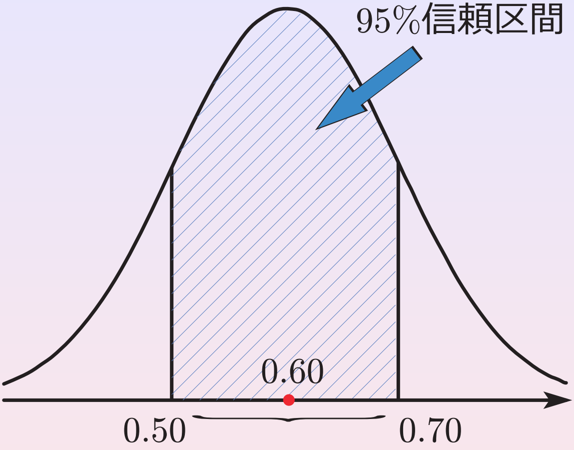
$$\text{有効率 } \mathbf{0.6} = \frac{\mathbf{60}}{\mathbf{100}}$$

⇧

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 **0.625** かもしれない ⇒ 確率付きで考える.



この区間が母比率を含む確率が95%

95% を 信頼係数 CI とか 信頼水準

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 有効率 **60%**

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.

青本 p.699 訂正

危険率 (p 値) は危険率 (有意水準 α) に訂正

4 行目 (修正) 得られたデータに基づいて検定統計量を計算し p 値と呼ばれる数値~~(危険率)~~ を算出する. この~~危険率~~ p 値と予め設定した有意水準 (α) を比較し

■ 青本の危険率の書き方 に難あり.

多くの教科書で, 危険率=有意水準 α　危険率 $\neq$ p 値である

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.

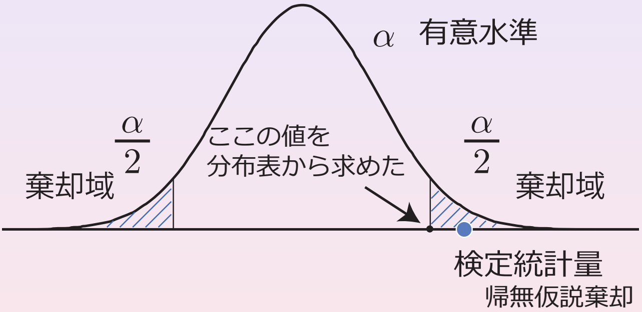
注意 2 回生の授業では検定統計量が棄却域に入るかどうかで判断した.

帰無仮説の考え方は
「宇宙怪人しまりす 医療統計を学ぶ 検定の巻」
第 1 話 がわかりやすい.

宇宙怪人しまりす

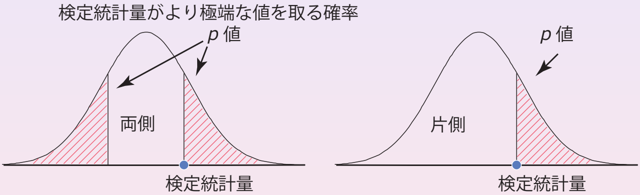
有意水準と棄却域

検定統計量が棄却域に入ると帰無仮説を棄却した.



p 値

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



青本 p.699 (訂正)

下から 8 行目　　確率を~~危険率~~ (p 値) とよぶ.

下から 5 行目　　~~危険率~~ (p 値) である.

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

有意水準 $\alpha = 0.05$ とは、
誤って帰無仮説を棄却する確率が 5 %を意味する

p.700
1 行目から 5 行目 マーカー

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.

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

$$\begin{cases} H_0\text{：棄却} & \Rightarrow \quad H_1\text{である. 積極的に支持} \\ H_0\text{：採択} & \Rightarrow \quad H_1\text{であるとはいえない. 積極的には支持していない} \end{cases}$$

結論は H_1 を使って述べれば良い.

教科書 p.55 表 5.2

	$\mu = \mu_0$ (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 が間違っているので捨てないといけないのに
ぼんやり (べーたー 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
母集団が正規分布を仮定する事が多い

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

平均値・最頻値・中央値

青本 p.704 参照

平均値 — データの和をデータ数で割ったもの
⇒ 正規分布のとき.

最頻値— モードともいう．データの中で一番多く現れる値
⇒ いびつな分布の時

中央値 — データを大小の順に並べた時ちょうど中央にくる値
⇒ いびつな分布の時

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

偏差値 ⇒ 点数のバラツキの位置を表す.

偏差値 70 なら上位　　偏差値 40 なら下位

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.

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

t 検定 (パラメトリック検定)

1) *t* 検定 (パラメトリック検定) 2 つの母集団の平均値は同じか ← 正規分布

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.63 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.65 5.4

帰無仮説・対立仮説

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　☜主張したいこと

2) Wilcoxon 順位和検定 (Mann-Whitney U 検定)

教科書 p.133 9.3 の Wilcoxon の符号付き順位和検定とは異なるが
どちらも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.

☞母集団が正規分布でないとき, 平均値 (中央値) は同じか

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.82 (6.5))

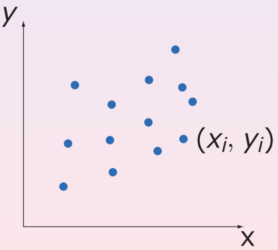
H_0（帰無仮説）: 母集団における薬剤 A 又は B による副作用発生率に差はない ☞差はない

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

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 $(\boldsymbol{x}, \boldsymbol{y})$ で与えられたデータ.

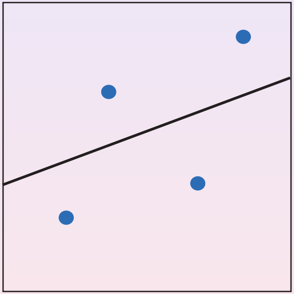


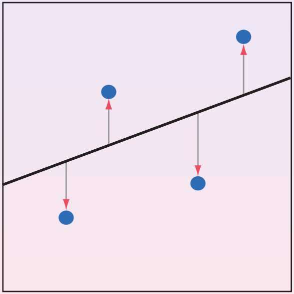
相関図

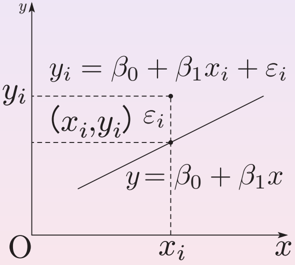
回帰直線 $y = \beta_0 + \beta_1 x$ で近似

『目的変数 (従属変数)』y = 『説明変数 (独立変数)』x

ビールの売上 y = 気温 x に関する式 $f(x)$







β_0 回帰直線の y 切片

β_1 傾き

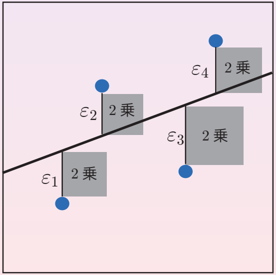
ε_i 誤差項

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

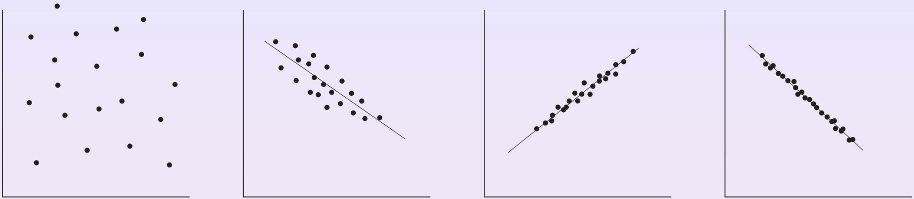


H_0（帰無仮説）: $\beta_1 = 0$ 等号で表示

H_1（対立仮説）: $\beta_1 \neq 0$

t 分布を使って検定.

相関係数



$r = -0.16$　　$r = -0.76$　　$r = 0.97$　　$r = -0.99$

1. 相関係数 $\boldsymbol{r}$ は $-1 \leqq \boldsymbol{r} \leqq 1$
2. 相関係数 $\boldsymbol{r}$ が 1 に近いほど，正の相関関係が強い．
   $\boldsymbol{r} = 1$ のときは右上がりの回帰直線上にデータ．
3. 相関係数 $\boldsymbol{r}$ が -1 に近いほど，負の相関関係が強い．
   $\boldsymbol{r} = -1$ のときは右下がりの回帰直線上にデータ．
4. データが直線的な分布から離れているとき $\boldsymbol{r}$ は 0 に近い値，
   逆もいえる．

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

パラメトリック検定

- (Student の) t 検定
  - 標本の大きさが 30 以上なら正規分布を使う時もある
- F 検定 (分散分析)

ノンパラメトリック検定

- χ^2 検定
- Fisher の直接確率法
- 符号検定 (Wilcoxon の順位和検定)

有意水準 (第 1 種・第 2 種の過誤)

連続データ

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

離散的データはデジタル表示 飛び飛びの値

(重) 回帰分析 ← 回帰直線 最小 2 乗法

相関係数