

基礎統計学

第3回 2群の比較

2016. 10. 21

正規性の問題

前回の復習

統計の手法は母集団が正規分布かどうかで大きく2つに分かれる。

- 1 正規性があるとき \implies パラメトリック検定
(t 検定, 分散分析など)
- 2 正規性がないとき \implies ノンパラメトリック検定
(順位和検定, U 検定など)

パラメトリック検定の方が理論上精密な検定 (検出力が高い) ができる。

正規性の確認法

- ヒストグラムや箱ひげ図をかいてみる.
- PPプロット（観測値を理論値のペアのプロット）
などをかいてみる.
- 正規性の検定を試してみる。（有意でなければ正規分布とみなす）

2群の平均の比較

パラメトリックな検定の代表例

2標本 t 検定 (対応のない場合)

前提

- 2つの母集団は正規分布に従う.
- 2つの母集団は独立 (互いに影響を及ぼさない) とする.

2つの正規母集団を Ω_1, Ω_2 とし,

母平均を μ_1, μ_2 , 母分散を σ_1^2, σ_2^2 で表す.

2群の平均の比較

問題

2つの母平均 μ_1, μ_2 が異なると言えるかどうか？

正規母集団は母平均と母分散の二つのパラメーターで決まる。(通常その値はどちらも未知である。) 従って、母平均の比較法は母分散の状況によって変わる。

2標本 t 検定

t 検定は2つの場合に分かれる.

- ① 母分散 σ_1^2, σ_2^2 が等しいとき (等分散)
- ② 母分散 σ_1^2, σ_2^2 が異なるとき \implies Welch の方法を用いる.

等分散の場合

2つの母集団からそれぞれ標本 X_{1i}, X_{2j} をとり、
標本のサイズを n_1, n_2 、標本平均を \bar{X}_1, \bar{X}_2 とする。
共通の分散の推定値として、合併分散

$$U^2 = \frac{S_1 + S_2}{n_1 + n_2 - 2}$$

を考える。 S_1, S_2 はそれぞれの標本の偏差平方和である。

等分散の場合

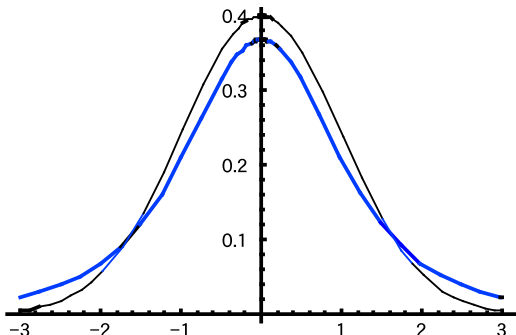
帰無仮説 $(H_0) : \mu_1 = \mu_2$.
対立仮説 $(H_1) : \mu_1 \neq \mu_2$.

(H_0) が正しいと仮定すると,

$$T = \frac{(\bar{X}_1 - \bar{X}_2)}{U \sqrt{\left(\frac{1}{n_1} + \frac{1}{n_2}\right)}} \quad (t \text{ 値})$$
 は自由度 $n_1 + n_2 - 2$ の t 分布に従う.

t 分布とは

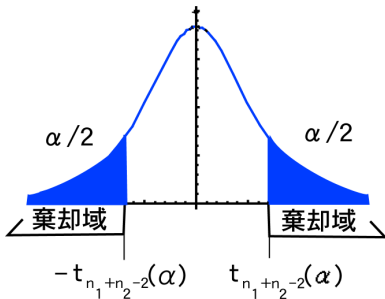
- t 分布は Gosset（ペンネーム Student）により発見された.
- 正規分布に似たグラフだが少し平坦
- 自由度が大きくなると正規分布に近づいていく



判定

- $|T| \geq t_{n_1+n_2-2}(\alpha) \implies (H_0)$ 棄却.
従って $(H_1) : \mu_1 \neq \mu_2$ といえる.
- $|T| < t_{n_1+n_2-2}(\alpha) \implies (H_0)$ は棄却できない.

ここで α は有意水準



判定

Rなどの統計ソフトでは p 値 ($|T| > t_0$ となる確率) が出力されるので

- $p < \alpha \implies (H_0)$ 棄却し (H_1) 採択
- $p \geq \alpha \implies (H_0)$ 棄却できない

等分散でないとき

母分散 σ_1^2, σ_2^2 が異なるときは **Welch の方法** を用いる.
実際, R の `t.test` コマンドは Welch の方法がデフォルト
になっている.

等分散検定

通常母分散は未知である。したがって母分散が等しいかどうかは直接確認できない。

そこで検定により等分散性を判定する。

R Commander には，等分散検定として

- 分散比を用いる F 検定 (R コマンド `var.test`)
- バートレット検定 (R コマンド `bartlett.test`)
- ルービン検定 (R コマンド `leveneTest`)

がある。

コメント

従来は、等分散検定をした後、通常の t 検定か Welch の t 検定を選択するという手順が推奨されたが、最近では、多重比較の問題や t 検定の頑健性の観点から、等分散検定はせず、始めから Welch の t 検定をすべきであるとの見解もある。

コメント

$n_1 \doteq n_2$ のときに、 t 分布は分散に関する頑健性 (robust 性) をもつ。また、第2種の過誤も小さくなることが知られてるので、2つのサンプルは同数の n 数にしておくのがよい。

例題： EZR (R commander) を用いて

ある地域の男女のヘモグロビン量の（仮想）データ
Hb.csv について男女でヘモグロビン量に差があるか？

- 1 Hb.csv を読み込む.
- 2 男女の等分散性の検定を行う.
- 3 男女の平均について t 検定を行う.

対応のある場合の t 検定

何か処置の前と後のように対応のある標本についての検定.

各データの差を新たな標本と思い処理するので, これは 1 標本 t 検定である.

R: `t.test(data1, data2, paired = TRUE)`

EZR: 統計解析 → 連続変数の解析 →

対応のある 2 群間の平均値の比較 (paired t 検定)

例題

- 1 Exp.csv を読み込む。(対応のあるデータ)
- 2 対応のある場合の t 検定を行う。

ノンパラメトリック検定

一般に正規分布でない場合（あるいは分布が不明の場合）の検定をノンパラメトリック検定という。

代表的なものに

- 1 Wilcoxon の順位和検定（or Mann-Whitney の U 検定）（対応のない場合）
- 2 Wilcoxon の符号付順位和検定（対応のある場合）

がある。

コメント

- Wilcoxon の順位和検定と Mann-Whitney の U 検定は本質的に同等な検定法である。(計算する統計量は異なるが、検定の結論は一致することが証明されている.)
- これらの検定はデータの値そのものではなく、順位の情報をもとにして、メディアンを比較する方法といえる.

R Commanderにて 対応のある場合

再び，Hb.csv を用いてノンパラメトリック検定を試みる。

EZRでは統計解析 → ノンパラメトリック検定 → 2群間の比較 (Mann-Whitney U検定 を選ぶ。

(Rから直接ならば

```
Hb <- read.csv("Hb.csv")  
wilcox.test(Hb$Hb ~ Hb$Sex)
```

と打ち込む。)

EZRにて

EZR:

統計解析 → ノンパラメトリック検定 →

対応のある2群間の比較 (Wilcoxon符号付順位和検定)

R:

```
data <- read.csv("Exp.csv")
```

```
wilcox.test(data$before, data$after, paired = TRUE)
```

検定における 2 種の誤り

統計的判断は常に誤りの可能性がある。

第 1 種の誤り (α エラー)

帰無仮説 H_0 が正しいのに、誤って棄却する誤り
その確率が有意水準 α である。

たとえば、検査における偽陽性の確率が α といえる。
 α は 0.05 や 0.01 など事前に設定でき、コントロール
可能

検定における 2 種の誤り

第 2 種の誤り (β エラー)

対立仮説 H_1 が正しいのに、 H_0 を棄却しない誤り
その確率を β で表し、 $1 - \beta$ をその検定の**検出力** (検定力, power) という。

たとえば、検査における偽陰性の確率が β といえる。
目安としては 0.2 (20%) 以下が望ましいとされる。
つまり、検出力は **0.8 (80%)** 以上あるのが望ましい。
(場合によっては 0.7 (70%) でも可.)

検定における 2 種の誤り

- ところが、 β は事前に定めることができない。したがって検出力も事前にはわからない。
- また、 n 数が一定ならば、 α と β を同時に小さくすることはできない。
- n 数を大きくすれば、 β は小さくなり、検出力は高くなる。有意でなかったとき、帰無仮説 H_0 を採択するためには十分大きな n 数が必要である。

検定における 2 種の誤り

- 一方, n 数が大きすぎると, 検出力が高くなりすぎ, 意味のない差を検出してしまう.
- また, 経済性や倫理性の問題もある.

サンプルサイズの評価

適切な n 数を事前に決めておくことが大事.

問

どれくらいの n 数を用意すべきか？

参考文献

- 豊田「検定力分析入門」東京書籍
- 永田「サンプルサイズの決め方」朝倉書店

サンプルサイズの評価

EZR

統計解析→必要サンプルサイズの計算 から必要なものを選択

サンプルサイズの評価

Rのコマンド

`power.t.test(n, delta, sd, power)`

- n 数
- delta 平均の差
- sd 標準偏差
- power 検出力 $1 - \beta$

これらのうち3つを与えると残りの値が返される。

2標本 t 検定の場合

$es = \frac{\text{delta}}{sd}$ を効果量という。

前提：

- 等分散性を仮定する。
(sd として合併分散のルートを考えればよい)
- $n = n_1 = n_2$ とする。

2標本 t 検定の場合

例として, Ht_MYpre.csv (各群 10 名) の身長について
2標本 t 検定を行うと有意にならない. 有意になるため
にはどれくらいの n 数が必要か?

サンプルサイズの評価

EZRでは

「2群の平均値の比較のためのサンプルサイズの計算」
を選ぶ

Ht_MYpre.csv の2つの群の身長データ（各群10件）

身長の差 $\text{delta} = 1.79$. $\text{sd} = 4.28$ となる. Rでは

`power.t.test(delta=1.79, sd=4.28, power=0.8)`

と打ち込むと . . .

サンプルサイズの評価

サンプルサイズの計算には Ht_MYpre.csv のような予備的なデータあるいは先行研究により得られているデータが必要である。