対応のある t 検定はなぜ使われにくいのか

鈴木 治郎 [*1]

[*1] 信州大学 全学教育機構

[*1]szkjiro@shinshu-u.ac.jp

キーワード データサイエンス教育，統計的検定，統計学の誤用

1 はじめに

2 群の平均値の差の検定に使われる t 検定は，検定手法の中ではおそらく，もっともよく使われるものの一つであろう．この t 検定を統計解析ソフトで処理しようとする場合，メニュー型であろうとなかろうと「対応のある／ない」の選択を指定する形になっているのがふつうである．

たとえば表計算ソフト Excel に用意されている分析ツールで t 検定を選択した場合，選択肢は次の通りである．

- 一対の標本による平均の検定
- 等分散を仮定した 2 標本による検定
- 分散が等しくないと仮定した 2 標本による検定

あるいは R の t 検定関数 `t.test` であれば Welch の t 検定が適用され，この関数のオプションとして

```
paired=TRUE/FALSE
```

の指定を行う．TRUE 指定が対応のある t 検定の指定であり，FALSE は対応のない場合である．省略時デフォルトは FALSE 指定に同じである．

この t 検定であるが，2 群の平均比較のレポート（学術論文でも）において，「運動前と後の被験者への効果」のような，各メンバーに関して明らかに対応のある場合でも，次節に示すような「対応のない t 検定」を適用している事例を少なからず目にする．以下ではこの問題をとりあげて議論したい．

2 サンプルデータと分析例

説明のためのサンプルデータを次により与える（表 1 参照）．$y = x + r$，ただし r は一様乱数 $0 \le r < 1$ であり，小数点以下第 2 位までで四捨五入したものを用意した．平均値は

$$\bar{x} = 12.0, \quad \bar{y} = 12.68, \qquad \overline{x-y} = \bar{x} - \bar{y} = -0.68$$

である．

サンプルデータのモデル $y = x + r$ は，同じ集団において，集団内の各個体に関する計測値に変化があり，その変化を分析する設定のつもりである．

Table 1 サンプルデータ

変数	1	2	3	4	5
x	10	11	12	13	14
y	10.7	11.2	12.6	13.9	15.0

2 群の平均値の比較では，次図のようなグラフに表すことが多い．

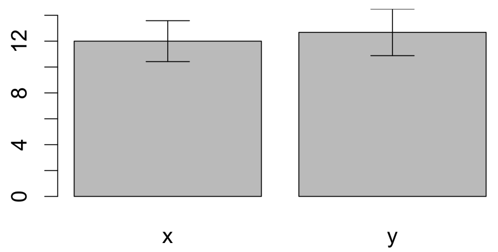

Fig.1 2 群の平均値の比較グラフ

このサンプルデータに対して R の t 検定関数 `t.test` を適用すると，出力（p 値）は次の通りとなる．

Table 2 サンプルデータに対する t 検定結果

paired 指定	p 値
TRUE	0.0081
FALSE	0.5439

ここで与えたサンプルデータは，x と y との変化，すなわち差 $x-y$ の平均値が 0 とみなせるかどうかが検討の対象になるべきである．だから分析対象とすべきは，次表にあげた $x-y$ とすべきである．

Table 3 サンプルデータの本質は差 $x-y$ である

変数	1	2	3	4	5	平均
x	10	11	12	13	14	12.00
y	10.7	11.2	12.6	13.9	15.0	12.68
$x-y$	-0.7	-0.2	-0.6	-0.9	-1.0	-0.68

3　統計学の教科書では

「対応のない t 検定」に関する教科書の記述では，数学的な表現を伴う場合

> 2 つの群が独立であるとき検定統計量は...

と条件に書いてあることがふつうである [1]．前節の例における 2 群 x と y は，明らかに独立でない．

なぜ 2 群の独立を仮定した解析方法の適用が行われるのであろうか．

3.1　実利的な記述例

統計解析にコンピュータ利用が当然のこととなった今日，実利的な側面から解説がなされることは珍しくない．たとえば嶋田 [2] では,「対応のない t 検定」による処理を説明した後に,

> ... 有意な結果にはならなかった．... 似た被験者を組みにして差を取る「対応のある t 検定」で分析をするとどうなるだろうか？ その結果，有意な差が得られた．

と解説する例題を紹介している．

コンピュータ利用の統計解析が手軽に行えるようになって以降，一部の利用者にみられる「いくつもの統計解析手法を実施した中で，有意な差を得たものを採用する」という嘆かわしい愚行を試すことが，この記述であれば促されると受け取る人もいるであろう．

分析方法の選択は，考察すべき統計モデルを利用者がどう判断するかにもとづくべきである．前節のサンプルデータで「対応のない t 検定」を採用した場合，モデルの x と y は独立ではないため，方法の誤用になる．

もちろん，2 群の集団の平均の差を t 検定で確かめたい状況の多くは，2 群が独立かどうかわからないから独立を仮定して検討するわけである．しかし，ここで問題にしている誤用とは，適当な処理の前後の平均値の比較問題であり，**2 群が独立ではありえない状況**を扱っていることに注意したい．

なぜ誤用が起こるのだろうか．

3.2　教科書のスタイル

一つには，ハウツー型の教科書が，各方法を採用するための数学的条件の説明を避けがちだという状況がある．

もう一つの理由は，正規分布の利用にもとづく平均値の点推定から進める標準的展開において，平均値の差の検定が紹介されるとき,「対応のある t 検定」は 1 標本の平均の検定の問題として先にとりあげることになる．もう一方の「対応のない t 検定」は 2 標本の平均値の差の検定でとりあげる [1]．このために，両者の検定を統合的な視点で眺める機会を持たないまま教科書を終える展開のよくあることも問題背景の一つであろう．

4　分散分析として考える

ここで論じてきた t 検定は，線形モデルであり，分散分析（ANOVA）の枠組みで捉えることができる．そして

- 対応のない t 検定 = 繰り返しのある 1 元配置分散分析
- 対応のある t 検定 = 繰り返しのない 2 元配置分散分析

に同等である．このことに言及している統計学入門の教科書もある [3]．

次表は 2 節で扱ったサンプルデータを「繰り返しのある 1 元分散分析」として分析して得た分散分析表である．先にあげた「対応のない t 検定」の p 値 0.5439 は，この値と小数点以下 3 桁まで一致している．

Table 4 繰り返しのある 1 元配置分散分析

項目	自由度	平方和	平均平方	F 値	p 値
集団間	1	1.156	1.1560	0.4023	0.5436
集団内	8	22.988	2.8735		

次表は 2 節で扱ったサンプルデータを「繰り返しのない 2 元分散分析」として分析して得た分散分析表である．先にあげた「対応のある t 検定」の p 値 0.0081 は，この値と小数点以下 4 桁まで一致している．

Table 5 繰り返しのない 2 元配置分散分析

項目	自由度	平方和	平均平方	F 値	p 値
集団間	1	1.156	1.1560	23.835	0.0081482
個体間	4	22.794	5.6985	117.495	0.0002125
誤差	4	0.194	0.0485		

コンピュータ利用が手軽となった現在，分散分析の枠組みの中で平均値の差をみるという考え方から，適切なモデルを考えること，そして群数が 2 に限って「平均の差をみる t 検定という手法」でも，同じ結果を得るが分析可能である，とするのも一つの見方ではないか．

5　おわりに

ここでは，明らかに独立でない 2 群の平均の差の検定に「対応のない t 検定」が使われることをよく見かける状

況をとりあげ，その背景を考察してきた．それは，データサイエンス教育の重要性が高まってきた今日において，統計学教育として何から学習すべきかという枠組みを設定するための，重要な問題を提起していると考える．今後，多くの分析では，統計解析のオペレーションを支援する人工知能の発達に伴って，ブラックボックス化が進行するだろうことは容易に想像できる．その際に人間が判断すべき根拠を明確に意識することは必要である．

ここでとりあげた問題は，人間が統計学の誤用をしないために欠かせない問題を提起していると考える．今後の研究の発展としては

- メタアナリシスあるいはインタビューの手法により，適当な領域の研究論文において，ここでとりあげた誤用の実態を明らかにすること
- 大学初年次学生などを対象とした統計学教育カリキュラムを設計すること
- 上で設計したカリキュラムを実践し，誤用の問題への気付きが高まるのか検証すること

などを考えたい．

参考文献

[1] 東京大学教養学部統計学教室,「統計学入門」，東京大学出版会，1991

[2] 嶋田正和他,「R で学ぶ統計学入門」，東京化学同人，2017

[3] ホーエル,「初等統計学第 4 版」，培風館，1981