等分散の検定に対する同等な 2 つの説明

Two Equivalent Explanations for the Homogeneity Test for Variance

川口　雄一 †

Yuuichi KAWAGUCHI

Test of homogeneity of variance is one of statistical tests. For accepting / rejecting of null / alternative hypothesis, there are times when an explanation in a textbook is differnt from another one in other textbook. There is a chance for novice learners of statistics to be confused. This paper attempts to show two different explanations for dealing hypothesis, and show that they are equivalent.

統計的仮説検定の一つである等分散性の検定において、帰無仮説と対立仮説の採択および棄却に関し、統計学の教科書によって説明が異なることがある。これは、統計学の学習者に混乱をもたらす可能性がある。そこで本稿では、仮説の扱いに関して 2 つの異なる説明を示し、それらは同等であることを示す。

キーワード:

分散	variance
等分散	homogeneity of variance
仮説検定	statistical hypothesis testing
F-分布	F-distribution
統計学の学習者	learner of statistics

† 天使大学 看護栄養学部 教養教育科

(2014 年 6 月 30 日 受稿, 2014 年 12 月 4 日 審査終了受理)

1　はじめに

特に断わらない限り、F, α, f_1, f_2 などの文字は、本文中で初めて現れるときにその意味を説明し、それ以降は同じ意味で用いる。

一般的に、統計的仮説検定[*1]において、検定統計量が棄却域に入るとき、帰無仮説を棄却し対立仮説を採択する ([2] など)。なお、帰無仮説を棄却しないとき、帰無仮説を採択する[*2]と言う。

本稿では、仮説検定の一つである「等分散性の検定」[*3]について、両側検定における棄却域の扱いに関する 2 通りの異なる説明を示し、それら 2 つの説明は同等であることを示す。

これによって、統計初学者の混乱を避けることが本稿の目的である。

2　等分散の検定と F-分布

2.1　仮説検定の一般論

一般的に、仮説検定は次の手順に沿う。

(1) 標本から検定統計量を求める。
(2) 帰無仮説のもと検定統計量が従うべき分布において、あらかじめ定めた有意水準 α および対立仮説 ($<, \neq, >$) から棄却域を設定する。
(3) 検定統計量は棄却域に入るかどうかによって結果を判断する。

なお、この場合の結果とは、帰無仮説を棄却するかどうかのことである。帰無仮説を棄却する場合には対立仮説を採択する。

2.2　等分散の検定

二つの群 A, B は正規分布に従うと仮定する。

A, B それぞれの群の母分散を σ_A^2, σ_B^2 とする。このとき、等分散の検定における帰無仮説 H_0 と対立仮説 H_1 は次のとおりである。

$$\begin{cases} H_0: & \sigma_A^2 = \sigma_B^2 \\ H_1: & \sigma_A^2 \neq \sigma_B^2 \end{cases}$$

二つの群 A, B について、標本から点推定して求めた母分散を、それぞれ $\hat{\sigma}_A^2, \hat{\sigma}_B^2$ とする。このとき、その比 $F = \hat{\sigma}_A^2 / \hat{\sigma}_B^2$ は等分散の検定における検定統計量である。これは帰無仮説 H_0 のもと、F-分布に従う。

なお、F-分布は 2 つの自由度により特徴づけられる。群 x $(x = A, B)$ の大きさ[*4]を、それぞれ N_x とすると、自由度 d_x は $d_x = N_x - 1$ である。

F-分布の場合、検定統計量 F について、次で求められる値 f_1, f_2 [*5]が、対立仮説 H_1 に関する棄却域 $[0, f_1)$ および (f_2, ∞) を定める。

$$\begin{cases} f_1 & = \mathrm{qf}\left(1 - \dfrac{\alpha}{2}, d_A, d_B\right), \\ f_2 & = \mathrm{qf}\left(\dfrac{\alpha}{2}, d_A, d_B\right). \end{cases}$$

ここで、関数 qf [*6]は次のとおり定義される。

$$\mathrm{qf}(a, n_1, n_2) = F_{n_1, n_2}(a)$$

ここで、a, n_1, n_2 は出典 [2](p. 65) における説明に従い、$P(F \geq t) = a$ となる数 t を $F_{n_1, n_2}(a)$ と書く。

以上の説明を図 1 に示す。

[*1] 以降、本稿では「仮説検定」と呼ぶ。
[*2] 帰無仮説を棄却しないときに、帰無仮説をどう扱うかについては議論がある (例 [4][5])。本稿では立入らず、単純に真・偽の 2 値論理的な表現を用いる。
[*3] 以降、本稿では「等分散の検定」と呼ぶ。

[*4] 標本がもつデータの個数
[*5] 「パーセント点」または「分位点」と呼ぶ。
[*6] 関数名の出典は **R** である。ただし、引数として `lower.tail = F` を与える。

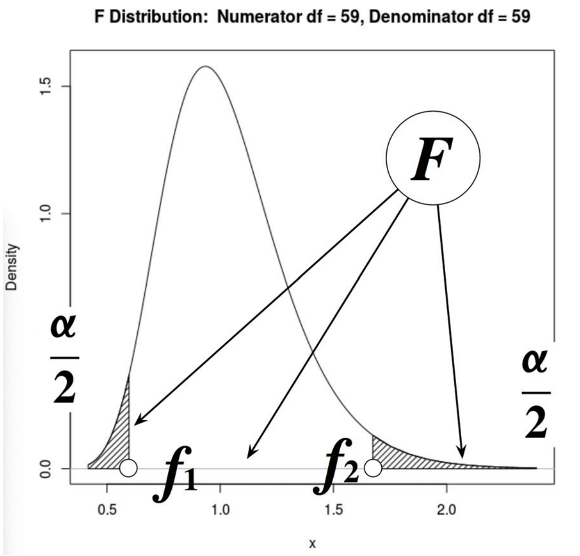

図 1　F-分布と棄却域 (自由度 59, 59 の例)

棄却域が両側にあるのは、対立仮説で H_1 を $\neq$ により設定しているためである。このとき、この仮説検定を特に「両側検定」と呼ぶ。

両側検定の他に、対立仮説の設定 (‘<’ または ‘>’) により、右片側検定、および、左片側検定をとり得る。本稿では両側検定の場合のみを考える。

3　異なる 2 つの説明

等分散の検定について、両側検定における棄却域の扱いについて、異なる 2 つの説明を示す。

3.1　説明 1

例えば文献 [1](p. 143) では、帰無仮説の採択・棄却に関して、次のとおり定めている。

$$\begin{cases} f_1 \leqq F \leqq f_2 \iff & H_0, \\ F < f_1 \text{または } f_2 < F \iff & H_1. \end{cases}$$

なお、 $\iff$ の右辺で H_i $(i = 0, 1)$ とあるのは、「仮説 H_i を採択する」ことを意味する。

3.2　説明 2

一方、例えば文献 [3](p. 313) では、次のとおり定めている。

$$\begin{cases} 1 \leqq F \text{ then} & f_2 < F & \iff H_1, \\ F < 1 \text{ then} & f_1' < \dfrac{1}{F} & \iff H_1, \\ \quad \text{otherwise} & & \iff H_0. \end{cases}$$

なお、2 行目にあるパーセント点 f_1' は、関数 qf により次のとおり求められる。

$$f_1' = \mathrm{qf}\left(\frac{\alpha}{2}, d_B, d_A\right).$$

$F = \hat{\sigma}_A^2/\hat{\sigma}_B^2$ において、f_1' を求める際に、f_1, f_2 の場合と自由度の順番が逆である。なお、一般的に次の関係 (†1) [*7]が成立つ。

$$\begin{aligned} f_1' &= \mathrm{qf}\left(\frac{\alpha}{2}, d_B, d_A\right) \\ &= \frac{1}{\mathrm{qf}\left(1 - \dfrac{\alpha}{2}, d_A, d_B\right)} = \frac{1}{f_1} \qquad \cdots (\dagger 1) \end{aligned}$$

4　考察

これまで示した 2 つの異なる説明が同等であることを示す。

4.1　H_0 を採択する場合

◇ 説明 1 $\Longrightarrow$ 説明 2

$f_1 \leqq F \leqq f_2$ のときを考える。$1 \leqq F$ とする。$F \leqq f_2$ であるから、説明 2 において otherwise の場合であり、H_0 を採択する。$F < 1$ とすれば、$f_1 \leqq F$ であること、および、関

[*7] 例えば、**http://www.econ.hokudai.ac.jp/~takagi/2005_July_2nd.pdf** に証明がある。

係 (†1) により、$1/F \leqq 1/f_1 = f_1'$ が成立つ。これは otherwise の場合であり、H_0 を採択する。

◇ 説明 2 $\Longrightarrow$ 説明 1

$1 \leqq F$ かつ $F \leqq f_2$ のとき、$F \leqq f_1 < f_2$ となる f_1 は存在しない。このことは、**R**[*8]により、数値計算的に確認した (†2)。実際には、常に $f_1 < 1.00$ となった。$1 \leqq F$ なので、$F < f_1$ とはならない。すなわち、常に $f_1 < F < f_2$ が満足され、説明 1 において H_0 を採択する。

$F < 1$ かつ $1/F \leqq f_1'$ のとき、関係 (†1) により、$f_1' = 1/f_1$ であり、$f_1 \leqq F$ が満足される。このとき、$f_1 < f_2 < F$ となるような f_2 はない。このことも、**R** により確認した (†3)。やはり、常に $f_1 < F < f_2$ が満足され、説明 1 において H_0 を採択する。

4.2 $\boldsymbol{H_1}$ を採択する場合

$\boldsymbol{H_0}$ を採択する場合で見たとおり、次の関係が成立つ。

$$\text{説明 1 で } H_0 \iff \text{説明 2 で } H_0 \qquad \cdots (\dagger\dagger)$$

説明 i で H_0 (ただし、$i = 1, 2$) の否定は定義 i で H_1 であり、したがって、関係 (††) の対偶をとることにより、次が成立つ。

$$\text{説明 1 で } H_1 \iff \text{説明 2 で } H_1$$

5 おわりに

本稿では、関係 (†1), (†2), (†3) に基づき、等分散の検定に関する、両側検定における棄却域の扱いについて、(1) 異なる 2 通りの説明を示し、(2) それらは同等であることを示した。

[*8] `http://www.R-project.org`

参考文献

[1] 松原 望, 入門 統計解析 [医学・自然科学編], 東京図書, 2007 年.

[2] 松本, 宮原, 数理統計学入門, 学術図書出版社, 1990 年.

[3] 森田 優三, 久次 智雄, 新統計概論 改訂版、日本評論社, 1993 年.

[4] ディヴィッド・サルツブルグ, 統計学を拓いた異才たち, 日本ビジネス人文庫, 2010 年.

[5] 豊田 秀樹 編著, 検定力分析入門, 東京図書, 2009 年.

謝辞

本稿に関し、金沢大学 小栗栖 修 准教授から有益な示唆をいただいた。ここに感謝します。