

# 独立 2 標本における等分散性の検定

青木繁伸

2020 年 3 月 17 日

## 1 目的

独立 2 標本における等分散性の検定を行う。

## 2 使用法

原データを用いる場合

```
import sys
sys.path.append("statlib")
from xtest import var_test
var_test(x, y, alternative="two_sided", conflevel=0.95, verbose=True)
```

二次データを用いる場合

```
import sys
sys.path.append("statlib")
from xtest import var_test2
var_test2(nx, ux, ny, uy, alternative="two_sided", conflevel=0.95, verbose=True)
```

### 2.1 引数

<code>x</code>	データベクトル
<code>y</code>	データベクトル
<code>alternative</code>	対立仮説の種類 デフォルトは "two_sided"。他に "less", "greater" を選択できる。
<code>conflevel</code>	信頼率 (デフォルトは 0.95)
<code>verbose</code>	必要最小限のプリント出力をする
<code>nx</code>	サンプルサイズ
<code>ux</code>	不偏分散
<code>ny</code>	サンプルサイズ
<code>uy</code>	不偏分散

## 2.2 戻り値の名前

"F"	検定統計量 ( $F$ 分布にしたがう)
"df1"	第 1 自由度
"df2"	第 2 自由度
"pvalue"	$p$ 値
"lcl"	標本分散の下側信頼限界値
"ucl"	標本分散の上側信頼限界値
"alternartive"	対立仮説の種類
"method"	検定手法名

## 3 使用例

### 3.1 原データを用いる場合

```
x = [4, 3, 2, 5, 4, 6, 5, 4, 3, 2, 7, 6, 5, 4]
y = [2, 1, 4, 3, 2, 5, 8, 6, 8, 4, 5, 6, 6, 4, 2, 5, 3, 2, 3]

import sys
sys.path.append("statlib")
from xtest import var_test

a = var_test(x, y)
```

```
F test to compare two variances
F = 0.53613, df1 = 13, df2 = 18, p value = 0.25614
alternative hypothesis: true ratio of variances is not equal to 1
95 percent confidence interval:
0.19637 1.5994
sample estimates: ratio of variances = 0.53613
```

### 3.2 二次データを用いる場合

以下の例では、元データのサンプルサイズと不偏分散を用いて同じ結果が出ることを示す。

この関数は本来は、元データがなくサンプルサイズと不偏分散しか分からないときに用いるものである。当然ながら使用する不偏分散の数値の精度によって検定結果 ( $p$  値) の精度にも影響が出る。

```
import numpy as np
from xtest import var_test2

np.random.seed(123) # 乱数の種 (本来は不要)
b = var_test2(len(x), np.var(x, ddof=1), len(y), np.var(y, ddof=1))
```

```
F test to compare two variances
```

```
F = 0.53613, df1 = 13, df2 = 18, p value = 0.25614
alternative hypothesis: true ratio of variances is not equal to 1
95 percent confidence interval:
0.19637 1.5994
sample estimates: ratio of variances = 0.53613
```

```
import numpy as np
from xtest import var_test2

print(len(x), np.var(x, ddof=1), len(y), np.var(y, ddof=1))
```

```
14 2.21978021978022 19 4.140350877192983
```

```
np.random.seed(123) # 乱数の種 (本来は不要)
c = var_test2(14, 2.22, 19, 4.14)
```

```
F test to compare two variances
F = 0.53623, df1 = 13, df2 = 18, p value = 0.25628
alternative hypothesis: true ratio of variances is not equal to 1
95 percent confidence interval:
0.19641 1.5997
sample estimates: ratio of variances = 0.53623
```

平均値と不偏分散の4つの数値を転記せずにそのまま使うと、元のデータを用いたのと丸めの誤差の範囲内で同じ結果になる。

```
np.random.seed(123) # 乱数の種 (本来は不要)
d = var_test2(len(x), np.var(x, ddof=1), len(y), np.var(y, ddof=1))
```

```
F test to compare two variances
F = 0.53613, df1 = 13, df2 = 18, p value = 0.25614
alternative hypothesis: true ratio of variances is not equal to 1
95 percent confidence interval:
0.19637 1.5994
sample estimates: ratio of variances = 0.53613
```

検定結果の戻り値に納められた  $p$  値 `b["pvalue"]`, `c["pvalue"]`, `d["pvalue"]` はそれぞれ  
0.25614284142458243  
0.2562821224748987  
0.25614284142458243  
である。