統計・多変量解析と
ソフトコンピューティング

第10章　多群の場合の当分散性の検定

本稿掲載のWebページ

http://mybook-pub-site.sakura.ne.jp/Statistics_Multivariate/index.html

古橋　武

第 10 章

多群の場合の等分散性の検定

第 8 章，第 9 章の多重比較法では，8.3 節のゲイムズ・ハウウェルの方法を除けば全て母分散が等しいことを前提としていた．したがって，これらの検定を適用する前に母分散が等しいとみなせるかどうかを検定しなければならない．

10.1　等分散性の検定（データ数が等しい場合）

10.1.1　ハートレーの方法による検定の例

8.1.1 項と同じ検定の課題に対して等分散性の検定を行う．検定の仮説は，データ群 i の母分散を σ_i とすると

$$\begin{aligned} \text{帰無仮説} &\quad : \quad \sigma_i = \sigma_j \\ \text{対立仮説} &\quad : \quad \sigma_i \neq \sigma_j \\ &\qquad\quad \text{ただし}, i, j = 1, 2, 3, i \neq j \end{aligned} \tag{10.1}$$

である．

データ群 i のデータ数を n_i とすると，この検定では $n_1 = n_2 = n_3 = n$ としている．データ群数が等しい場合の等分散性の検定法にハートレー (Hartley) の方法がある．データ群 i の不偏分散を v_{ei}^2 とすると，検定統計量

$$f_{ij} = \frac{v_{ei}^2}{v_{ej}^2} \quad (i, j = 1, 2, 3, i \neq j) \tag{10.2}$$

が自由度 $n-1, n-1$ の F 分布に従うことを利用する．

ここで f_{ij} の最大値 f_{max} を

$$f_{max} = \frac{\max\limits_i \{v_{ei}^2\}}{\min\limits_j \{v_{ej}^2\}} \tag{10.3}$$

とすると，ハートレーの方法では f_{max} に対する閾値 f_0 が数表により与えられている．公称の有意水準 $\alpha = 0.05$ のとき，表 A.12 のハートレーの方法の 5% 点を利用する．この f_{ij} の最大値という事象を本書では F_{max} と表す．

図 10.1 に実施例を示す．セル C4 には閾値 f_0 が入力されている．数表より設定数 $a = 3$，自由度 $\nu = n - 1 = 8$ のとき $f_0 = 6.00$ と読める．セル B24 では f_{max} の値を出力している．f_{max} の値は閾値 f_0 より小さく帰無仮説：$\sigma_1 = \sigma_2 = \sigma_3$ を棄却できない．

	A	B	C	D
1	等分散性の検定（ハートレーの方法）			
2				
3	入力	公称の有意水準α	閾値f0	
4		0.05	6	
5				
6		設定1	設定2	設定3
7		3.3	3.2	3.1
8		3.2	3.1	3.0
9		3.4	3.3	3.3
10		3.3	3.2	3.1
11		3.3	3.2	3.2
12		3.2	3.1	3.3
13		3.5	3.4	3.0
14		3.2	3.2	3.1
15		3.5	3.0	3.1
16				
17	計算値	設定数 a	データ数 n	自由度 ν
18		3	9	8
19		不偏分散ve1^2	ve2^2	ve3^2
20		0.0144	0.0136	0.0125
21		vmax	vmin	
22		0.0144	0.0125	
23	出力	fmax		
24		1.156		

図 10.1: 等分散性の検定の例（ハートレーの方法）（ハートレーによる検定.xlsx）

(略)

10.1.2　ハートレーの方法によるシミュレーション

(略)

10.1.3　ハートレーの検定の理論

互いに独立な事象 X_1, X_2, X_3 はそれぞれ平均 μ_1, μ_2, μ_3, 分散 $\sigma_1^2, \sigma_2^2, \sigma_3^2$ の正規分布に従うとする．すなわち，$X_1 \sim N(\mu_1, \sigma_1^2), X_2 \sim N(\mu_2, \sigma_2^2), X_3 \sim N(\mu_3, \sigma_3^2)$ である．また，各群のデータ数 n_1, n_2, n_3 は同じとする $(n_1 = n_2 = n_3 = n)$．事象 X_1, X_2, X_3 のデータ群の平均値をそれぞれ $\overline{x}_1, \overline{x}_2, \overline{x}_3$，不偏分散を $v_{e1}^2, v_{e2}^2, v_{e3}^2$ とすると，データ群 $i, j (i, j = 1, 2, 3, i \neq j)$

の不偏分散の比 f_{ij} は

$$f_{ij} = \frac{v_{ei}^2}{v_{ej}^2} \tag{10.4}$$

となる．f_{ij} は次の等分散の仮説の下で，式 (??) より自由度 $n-1, n-1$ の F 分布に従う．検定の仮説は，

$$\begin{aligned} \text{帰無仮説} \quad &: \quad \sigma_i = \sigma_j \\ \text{対立仮説} \quad &: \quad \sigma_i \neq \sigma_j \\ & \quad \text{ただし}, i, j = 1, 2, 3, i \neq j \end{aligned} \tag{10.5}$$

である．

(以下略)

10.2　等分散性の検定（データ数が異なる場合）

前節のハートレーの方法の理論はデータ数を等しいとして導出していた．したがって，各データ群でデータ数が異なる場合には検定の結果は保証されない．各群のデータ数が異なる場合の検定法にバートレット (Bartlett) の方法がある．

10.2.1　バートレットの方法による検定の例

8.2.1 項と同じ検定の課題に対して等分散性の検定を行う．検定の仮説は，データ群 i の母分散を σ_i^2 とすると

$$\begin{aligned} \text{帰無仮説} \quad &: \quad \sigma_i^2 = \sigma_0^2, \quad \forall i \\ \text{対立仮説} \quad &: \quad \sigma_i^2 \neq \sigma_0^2, \quad \exists i \end{aligned} \tag{10.23}$$

である．帰無仮説は全ての母分散が σ_0^2 に等しいとし，対立仮説は，母分散が σ_0^2 に等しくないデータ群があるとする．データ群数を a，データ群 i のデータ数を n_i．不偏分散を v_{ei}^2 とすると，検定統計量 B が次式で与えられる．

$$B = \frac{2\log L}{1 + \frac{1}{3(a-1)}\left(\sum_{i=1}^{a} \frac{1}{n_i - 1} - \frac{1}{\nu}\right)} \tag{10.24}$$

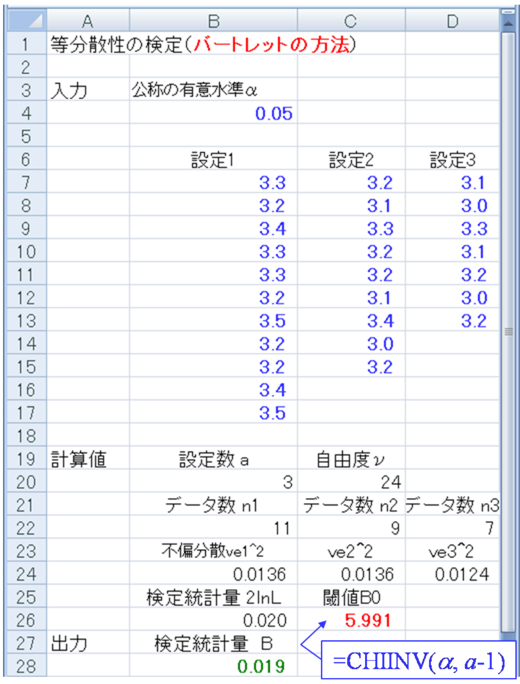
ただし，

$$2\log L = \nu \left\{ \log \frac{\sum_{i=1}^{a}(n_i - 1)v_{ei}^2}{\nu} - \frac{1}{\nu}\sum_{i=1}^{a}(n_i - 1)\log v_{ei}^2 \right\} \tag{10.25}$$

$$\nu = \sum_{i=1}^{a}(n_i - 1) \tag{10.26}$$

である．$\log x$ の底は e である．B が自由度 $a-1$ の χ^2 分布に従うことを利用する．

図 10.8 に実施例を示す．セル C26 には自由度 $a-1=2$ における閾値 B_0 が CHIINV() 関数を用いて求められている．セル B28 の検定統計量 B の値は閾値 B_0 より小さく帰無仮説：$\sigma_1^2=\sigma_2^2=\sigma_3^2=\sigma_0^2$ を棄却できない．



	A	B	C	D
1	等分散性の検定(バートレットの方法)			
2				
3	入力	公称の有意水準α		
4		0.05		
5				
6		設定1	設定2	設定3
7		3.3	3.2	3.1
8		3.2	3.1	3.0
9		3.4	3.3	3.3
10		3.3	3.2	3.1
11		3.3	3.2	3.2
12		3.2	3.1	3.0
13		3.5	3.4	3.2
14		3.2	3.0	
15		3.2	3.2	
16		3.4		
17		3.5		
18				
19	計算値	設定数 a	自由度 ν	
20		3	24	
21		データ数 n1	データ数 n2	データ数 n3
22		11	9	7
23		不偏分散ve1^2	ve2^2	ve3^2
24		0.0136	0.0136	0.0124
25		検定統計量 2lnL	閾値B0	
26		0.020	5.991	
27	出力	検定統計量 B		
28		0.019		

=CHIINV(α, a-1)

図 10.8: 等分散性の検定の例（バートレットの方法）（バートレットによる検定.xlsx）

10.2.2　バートレットの方法によるシミュレーション

(略)

図 10.11 は，1000 組のシミュレーションにおける各組の B 値の頻度分布を示す．B 値の小数点以下を切り捨てて，各整数値の出現割合を示してある．参考に自由度 2 の χ^2 分布に基づく確率分布を併せて示す．シミュレーションを再実行すると B 値の頻度が理論値の周りで変化する様子を見て取ることができる．

(略)

10.2.3　バートレットの検定の理論

(以下略)

著者
古橋　武
名古屋大学工学研究科計算理工学専攻

本稿の内容は，

古橋武・宮本定明著
「統計・多変量解析とソフトコンピューティング —超多自由度系解析を目指して—」
金田・笹井監修，計算科学講座 第 3 巻，共立出版, 2012
http://www.kyoritsu-pub.co.jp/bookdetail/9784320122680

から抜粋したものです．共立出版社の許可を得て Web ページに掲載しています．著作権法上で認められている例外を除き，出版社の許可なく複写することはできません．