

統計・多変量解析と ソフトコンピューティング

第9章 多重比較法—検定統計量の見直し—

本稿掲載の Web ページ

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

古橋 武

目次

第9章 多重比較法—検定統計量の見直し—	1
9.1 母平均の差の検定（差の有無の検定）	1
9.1.1 分散分析による検定の例	1
9.1.2 分散分析のシミュレーション	3
9.1.3 分散分析の理論	4
9.2 母平均の差の検定（3群の場合の下位検定法）	7
9.2.1 フィッシャーのPLSD法による検定の例	8
9.2.2 フィッシャーのPLSD法のシミュレーション	8
9.2.3 フィッシャーのPLSD法の理論	9
9.3 母平均の差の検定（4群以上の場合の下位検定法）	9
9.3.1 ヘイター・フィッシャーの改良法による検定の例	10
9.3.2 ヘイター・フィッシャーの改良法のシミュレーション	13
9.3.3 ヘイター・フィッシャーの改良法の理論	15
9.4 母平均の差の検定（比較対象の組合せを自在にする検定法）	15
9.4.1 シェフェの方法による検定の例	16
9.4.2 シェフェの方法のシミュレーション	18
9.4.3 シェフェの方法の理論	18
9.5 シェフェの方法とヘイター・フィッシャーの方法の比較	21
9.6 まとめ	22
参考文献	24

第9章

多重比較法—検定統計量の見直し—

9.1 母平均の差の検定（差の有無の検定）

ボンフェローニ、シダックの方法は名義水準 α' を調整することで第1種の過誤の確率を公称の有意水準に一致させる手法であった。一方、テューキー、テューキー・クレーマー、ゲイムズ・ハウウェル、ダネットの方法は検定統計量の閾値 q_0 を調整して第1種の過誤の確率を公称の有意水準とする多重比較法であった。以上の他に、[検定統計量そのものを見直す](#) 考え方がある。本章で紹介する分散分析法、フィッシャーのPLSD法、ヘイター・フィッシャーの方法、シェフェの方法などがこの考え方に基づく多重比較法である。

9.1.1 分散分析による検定の例

ある日、ある工場のある製品のラインにおいて、ラインのスピードを3通りに変えながらあるパラメータの測定がなされたとする。スピード設定1, 2, 3からそれぞれ11, 9, 7個の標本を取り出すことができ、以下の値が得られたとする。

設定1: 3.3, 3.2, 3.4, 3.3, 3.3, 3.2, 3.5, 3.2, 3.2, 3.4, 3.5

設定2: 3.2, 3.1, 3.3, 3.2, 3.2, 3.1, 3.4, 3.0, 3.2

設定3: 3.1, 3.0, 3.3, 3.1, 3.2, 3.0, 3.2

ライン1, 2, 3の平均値 $\bar{x}_1, \bar{x}_2, \bar{x}_3$ はそれぞれ3.32, 3.19, 3.13であった。スピード設定を変えたことで平均値に差が出たと言えるか？ただし、いずれの母分散にも差がないとする。

これは8.2.1項の検定課題と全く同じである。検定の仮説は、データ群 i の母平均を μ_i

とすると

$$\begin{aligned}
 \text{帰無仮説} & : \mu_i = \mu_j \\
 \text{対立仮説} & : \mu_i \neq \mu_j \\
 & \text{ただし, } i, j = 1, 2, 3, i \neq j
 \end{aligned}
 \tag{9.1}$$

である。これに対して、分散分析の仮説は

$$\begin{aligned}
 \text{帰無仮説} & : \mu_1 = \mu_2 = \mu_3 \\
 \text{対立仮説} & : \mu_1 \neq \mu_2 \quad \text{or} \quad \mu_1 \neq \mu_3 \quad \text{or} \quad \mu_2 \neq \mu_3
 \end{aligned}
 \tag{9.2}$$

である。

母分散に差がない場合には、検定統計量

$$F = \frac{\frac{S_A}{a-1}}{\frac{S_E}{n_1+n_2+\dots+n_a-a}}
 \tag{9.3}$$

が自由度 $a-1, n_1+\dots+n_a-a$ の F 分布に従うことを利用する。ただし、 a はデータ群数、 n_i はデータ群 i のデータ数であり、

$$\begin{aligned}
 S_A & = \sum_{i=1}^a n_i (\bar{x}_i - \bar{x})^2 \\
 S_E & = \sum_{i=1}^a \sum_{j=1}^{n_i} (x_{ij} - \bar{x}_i)^2
 \end{aligned}
 \tag{9.4}$$

である。 S_A, S_E は水準間変動、水準内変動と呼ばれる。また、 x_{ij} は設定 i の第 j 番目の測定値、 \bar{x} は全体の平均値、 \bar{x}_i はデータ群 i の平均値であり、それぞれ

$$\begin{aligned}
 \bar{x} & = \frac{1}{\sum_{i=1}^a n_i} \sum_{i=1}^a \sum_{j=1}^{n_i} x_{ij} \\
 \bar{x}_i & = \frac{1}{n_i} \sum_{j=1}^{n_i} x_{ij}
 \end{aligned}
 \tag{9.5}$$

である。全変動 S_T を

$$S_T = \sum_{i=1}^a \sum_{j=1}^{n_i} (x_{ij} - \bar{x})^2
 \tag{9.6}$$

と定義すると、

$$S_T = S_A + S_E
 \tag{9.7}$$

の関係がある。

分散分析の実施例を図9.1に示す。セルB24の F 値は自由度 $a-1, \nu = n_1 + n_2 + n_3 - a$ の F 分布に従う。セルC24では

$$= \text{FDIST}(F, a - 1, \nu) \quad (9.8)$$

により、 F 値より大きな値の出る確率(p 値)を求めている。 p 値は0.0057である。セルB26ではこの p 値がセルB4に入力された公称の有意水準 $\alpha = 0.05$ を下まわった場合に1, そうでない場合に0を出力している。この例では帰無仮説： $\mu_1 = \mu_2 = \mu_3$ を棄却できる結果となっている。

	A	B	C	D
1	母平均の差の検定(分散分析)			
2				
3	入力	公称の有意水準 α		
4		0.05		
5		設定1	設定2	設定3
6		3.3	3.2	3.1
7		3.2	3.1	3.0
8		3.4	3.3	3.3
9		3.3	3.2	3.1
10		3.3	3.2	3.2
11		3.2	3.1	3.0
12		3.5	3.4	3.2
13		3.2	3.0	
14		3.2	3.2	
15		3.4		
16		3.5		
17	計算値	データ群数 a	自由度 ν	
18		3	24	
19		データ数 n1	n2	n3
20		11	9	7
21	(分散分析)	全変動 ST	水準間変動 SA	水準内変動 SE
22		0.4919	0.1723	0.3195
23	出力	F値	p値	=FDIST(F, a-1, ν)
24		6.4711	0.0057	
25		if p < α then 1 else 0	$\nu S_A / (a-1) S_E$	
26		1		

図 9.1: 分散分析 (分散分析.xlsx)

9.1.2 分散分析のシミュレーション

図9.2は分散分析のシミュレーションを行っているExcelの画面である。母平均 $\mu_i = 1$, 母標準偏差 $\sigma_i = 1$ の正規分布に従う乱数をデータ群1, 2, 3にそれぞれ11個, 9個, 7個生成させて p 値を求めて, 公称の有意水準 $\alpha = 0.05$ を下まわった場合に31行目のセルに1を出力している。この F 検定を1000組について行い, 第1種の過誤の起きた頻度をセルB33に出力している。この例では1000組中38組において第1種の過誤が起きていた。図9.3は, 以上の1000組のシミュレーションを100回繰り返したときの第1種の過誤の

確率の95%信頼区間を求めた結果である。第1種の過誤の真の確率 p^* は公称の有意水準 $\alpha = 0.05$ の辺りにある結果となった。図9.4は、1000組のシミュレーションにおける各組の F 値の頻度分布を示す。 F 値は小数点以下を四捨五入して、各整数値の出現割合を示してある。参考に自由度 $(a-1, \nu) = (2, 24)$ の F 分布 F_{24}^2 に基づく確率分布を併せて示してある。シミュレーションを再実行する度に頻度分布は理論値の周りで変化することが分かる。

	A	B	C	D	E	F	G
1	母平均の差の検定のシミュレーション(分散分析)						
2							
3	入力	$\mu 1$	$\mu 2$	$\mu 3$	$\sigma 1$	$\sigma 2$	$\sigma 3$
4		1	1	1	1	1	1
5		公称の有意水準 α					
6		0.05					
7							
8	計算値	1組目			2組目		
9		x1	x2	x3	x1	x2	x3
10		1.099	0.558	1.133	-0.307	-0.213	-0.031
11		-0.287	0.614	1.442	1.441	1.707	2.088
22		データ群数 a	自由度 ν	組数 G			
23		3	24	1000			
24		データ数 n1	n2	n3			
25		11	9	7			
26		全変動 ST	水準間変動 SA	水準内変動 SE	全変動 ST	水準間変動 SA	水準内変動 SE
27		20.07915	0.22509	19.85406	31.20424	2.26728	28.93696
28		F値	p値		F値	p値	
29		0.1360	0.8735		0.9402	0.4045	
30		if p < α then 1			if p < α then 1		
31		else 0			else 0		
32	出力	有意頻度 fr					
33		0.038					

図 9.2: 母平均の差の検定のシミュレーション (分散分析) (分散分析によるシミュレーション.xlsm)

53		有意頻度 frk	100回の平均		
54	fr1	0.052	0.0503		
55	fr2	0.047	95%信頼区間		
56	fr3	0.051	0.0488	$\leq p^* \leq$	0.0518

図 9.3: 母平均の差の検定, 分散分析における第1種の過誤の確率の95%信頼区間

9.1.3 分散分析の理論

互いに独立な事象 $X_{ij}(i = 1, \dots, a, j = 1, \dots, n_i) \sim N(\mu, \sigma^2)$ を考える。各事象の母平均 μ , 母分散 σ^2 は等しいとする。 a はデータ群数, n_i はデータ群 i のデータ数である。なお, 分散分析ではデータ群1, 2,..を水準1, 2,..とよぶ。式(9.4)の水準間変動, 水準内変

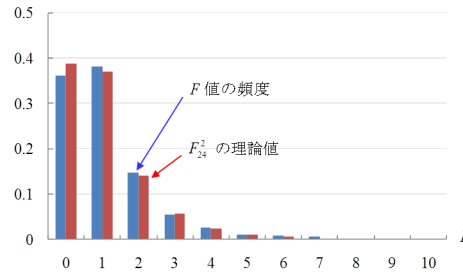


図 9.4: 母平均の差の検定, 分散分析における F 値の頻度分布 (F 値の分布 (分散分析) .xlsx)

動はデータ群間変動, データ群内変動とよんでもよいものである. 式 (9.7) の関係が成り立つことは以下による.

$$\begin{aligned}
 \sum_{i=1}^a \sum_{j=1}^{n_i} (x_{ij} - \bar{x})^2 &= \sum_{i=1}^a \sum_{j=1}^{n_i} (x_{ij} - \bar{x}_i + \bar{x}_i - \bar{x})^2 \\
 &= \sum_{i=1}^a \sum_{j=1}^{n_i} \{ (x_{ij} - \bar{x}_i)^2 + 2(x_{ij} - \bar{x}_i)(\bar{x}_i - \bar{x}) + (\bar{x}_i - \bar{x})^2 \} \\
 &= \sum_{i=1}^a \sum_{j=1}^{n_i} (x_{ij} - \bar{x}_i)^2 + \sum_{i=1}^a n_i (\bar{x}_i - \bar{x})^2
 \end{aligned} \tag{9.9}$$

式 (6.78) より

$$\begin{aligned}
 \frac{1}{\sigma^2} S_E &= \frac{1}{\sigma^2} \sum_{j=1}^{n_1} (x_{1j} - \bar{x}_1)^2 + \cdots + \frac{1}{\sigma^2} \sum_{j=1}^{n_a} (x_{aj} - \bar{x}_a)^2 \\
 &\sim \chi_{n_1 + \cdots + n_a - a}^2
 \end{aligned} \tag{9.10}$$

である. また,

$$\begin{aligned}
 \frac{1}{\sigma^2} S_A &= \frac{1}{\sigma^2} \sum_{i=1}^a n_i (\bar{x}_i - \bar{x})^2 \\
 &= \frac{1}{\sigma^2} \sum_{i=1}^a n_i \{ \bar{x}_i - \mu_0 - (\bar{x} - \mu_0) \}^2 \\
 &= \frac{1}{\sigma^2} \sum_{i=1}^a n_i (\bar{x}_i - \mu_0)^2 - \frac{1}{\sigma^2} \sum_{i=1}^a n_i (\bar{x} - \mu_0)^2 \\
 &= \sum_{i=1}^a \left(\frac{\bar{x}_i - \mu_0}{\sqrt{\frac{\sigma^2}{n_i}}} \right)^2 - \left(\frac{\bar{x} - \mu_0}{\sqrt{\frac{\sigma^2}{n_1 + \cdots + n_a}}} \right)^2
 \end{aligned} \tag{9.11}$$

であり、式(3.67)より、

$$\frac{\bar{X}_i - \mu_0}{\sqrt{\frac{\sigma^2}{n_i}}} \sim N(0, 1), \quad \frac{\bar{X} - \mu_0}{\sqrt{\frac{\sigma^2}{n_1 + \dots + n_a}}} \sim N(0, 1) \quad (9.12)$$

である。よって、式(4.122), (4.124)より、

$$\sum_{i=1}^a \left(\frac{\bar{X}_i - \mu_0}{\sqrt{\frac{\sigma^2}{n_i}}} \right)^2 \sim \chi_a^2, \quad \left(\frac{\bar{X} - \mu_0}{\sqrt{\frac{\sigma^2}{n_1 + \dots + n_a}}} \right)^2 \sim \chi_1^2 \quad (9.13)$$

と得られる。また、式(9.11)は

$$\sum_{i=1}^a \left(\frac{\bar{x}_i - \mu_0}{\sqrt{\frac{\sigma^2}{n_i}}} \right)^2 = \frac{1}{\sigma^2} S_A + \left(\frac{\bar{x} - \mu_0}{\sqrt{\frac{\sigma^2}{n_1 + \dots + n_a}}} \right)^2 \quad (9.14)$$

と変形できる。4.8節と同様にして、上式の右辺第1項と第2項の独立性を示すことができ、

$$\frac{1}{\sigma^2} S_A \sim \chi_{a-1}^2 \quad (9.15)$$

と求められる。したがって、式(9.10), (9.15), (5.22)より

$$\frac{\frac{S_A}{a-1}}{\frac{S_E}{n_1 + \dots + n_a - a}} \sim F_{n_1 + \dots + n_a - a}^{a-1} \quad (9.16)$$

である。水準間変動 S_A は各データ群の平均値のばらつきであり、これが大きいとき平均値間に差があることになる。したがって、 S_A を分子に持つ F 値が閾値 F_0 を超えた場合には、各データ群の平均値間に差がないとする帰無仮説を棄却する。

なお、分散分析の検定統計量 F 値とステューデントの t 検定の検定統計量 t 値の2乗値はデータ群数 $a = 2$ の時に一致する。 t 値は式(6.7)より

$$t = \frac{\bar{x}_1 - \bar{x}_2}{\sqrt{v_e^2 \left(\frac{1}{n_1} + \frac{1}{n_2} \right)}} \quad (9.17)$$

である。ただし、 n_1, n_2 はデータ群1,2のデータ数である。また、

$$v_e^2 = \frac{(n_1 - 1)v_{e1}^2 + (n_2 - 1)v_{e2}^2}{n_1 + n_2 - 2} \quad (9.18)$$

である。 v_{e1}^2, v_{e2}^2 はデータ群1,2の不偏分散である。 t 値を2乗すると

$$t^2 = \frac{\frac{n_1 n_2}{n_1 + n_2} (\bar{x}_1 - \bar{x}_2)^2}{v_e^2} \quad (9.19)$$

となり、分母は検定統計量 F 値の分母と一致する。分子については

$$\frac{n_1 n_2}{n_1 + n_2} (\bar{x}_1 - \bar{x}_2)^2 = n_1 (\bar{x}_1 - \bar{x})^2 + n_2 (\bar{x}_2 - \bar{x})^2 \quad (9.20)$$

が成り立つことを示せばよい。データ群数 $a = 2$ のとき

$$\bar{x} = \frac{n_1 \bar{x}_1 + n_2 \bar{x}_2}{n_1 + n_2} \quad (9.21)$$

であるので、これを式 (9.20) の右辺に代入すると

$$\begin{aligned} n_1 (\bar{x}_1 - \bar{x})^2 + n_2 (\bar{x}_2 - \bar{x})^2 &= n_1 \left(\bar{x}_1 - \frac{n_1 \bar{x}_1 + n_2 \bar{x}_2}{n_1 + n_2} \right)^2 + n_2 \left(\bar{x}_2 - \frac{n_1 \bar{x}_1 + n_2 \bar{x}_2}{n_1 + n_2} \right)^2 \\ &= n_1 \left(\frac{n_2 \bar{x}_1 - n_2 \bar{x}_2}{n_1 + n_2} \right)^2 + n_2 \left(\frac{-n_1 \bar{x}_1 + n_1 \bar{x}_2}{n_1 + n_2} \right)^2 \\ &= \frac{n_1 n_2}{n_1 + n_2} (\bar{x}_1 - \bar{x}_2)^2 \end{aligned} \quad (9.22)$$

となる。よって、

$$t^2 = F_{n_1+n_2-1}^{a-1} \quad (9.23)$$

が示された。データ群数 $a \geq 3$ の場合には分散分析の F 値はステューデントの t 値の拡張とみなすことができる。

9.2 母平均の差の検定 (3群の場合の下位検定法)

前節の分散分析により式 (9.2) の帰無仮説が棄却されれば、いずれかのデータ群の母平均間に差があることがわかる。どの母平均間に差があるかは個々の母平均の差に対して検定を適用する必要がある。すなわち、式 (9.1) の仮説の下での検定を行うには、分散分析 (事前検定と呼ばれる) で有意差ありと判定された後で、さらなる検定 (事後検定、もしくは下位検定と呼ばれる) を行わなければならない。データ群が3群の場合に適用できる下位検定法にフィッシャー (Fisher) の PLSD (Protected Least Significant Difference) 法がある。これは分散分析で有意差ありと判定された後で、6.2節のステューデントの t 検定を適用する方法である。制約付き最小有意差法と訳される。制約付きとは事前に分散分析により有意差ありと判定された場合にのみ実施する検定であることを意味し、LSD とはステューデントの t 検定において有意差ありと判定できる最小の t 値を意味する。制約なし (unprotected) でステューデントの t 検定を3群間の平均値の差の検定に適用することは、図 8.2、図 8.3 に見たように第1種の過誤の確率を公称の有意水準よりも大きくしてしまう。また、4群以上の平均値の差の検定に対してこの PLSD 法を適用した場合にも第1種の過誤の確率を公称の有意水準より大きくしてしまう。フィッシャーの PLSD 法は3群の場合にのみ適用できる。

9.2.1 フィッシャーの PLSD 法による検定の例

図 9.5 は 9.1.1 項の検定課題に対してフィッシャーの PLSD 法を適用している Excel の画面である。3 行目から 26 行目までは図 9.1 の分散分析の実施例と全く同じであるので省略してある。ステューデントの t 検定の検定統計量 t_{kl} を

$$t_{kl} = \frac{\bar{x}_k - \bar{x}_l}{\sqrt{v_E^2 \left(\frac{1}{n_k} + \frac{1}{n_l} \right)}} \quad (9.24)$$

とする。 \bar{x}_k, \bar{x}_l はそれぞれデータ群 k, l の平均値であり、 n_k, n_l はデータ数である。ただし、

$$\begin{aligned} v_E^2 &= \frac{\sum_{i=1}^a \sum_{j=1}^{n_i} (x_{ij} - \bar{x}_i)^2}{\sum_{i=1}^a (n_i - 1)} \\ &= \frac{S_E}{\sum_{i=1}^a (n_i - 1)} \end{aligned} \quad (9.25)$$

である。セル C31 には公称の有意水準 $\alpha = 0.05$ に対応する t_0 値を $TINV()$ 関数により求め、これを最小有意差としてある。自由度 $\nu = \sum_{i=1}^a (n_i - 1)$ である。35 行目では分散分析で有意差ありとされ、かつ、検定統計量の絶対値 $|t_{kl}|$ が最小有意差 t_0 を上回った場合に 1 を出力している。図 8.18 との違いは、テューキー・クレーマーの方法では設定 1, 3 間に有意差ありとされていたのに対して、フィッシャーの PLSD 法では設定 1, 3 間に加えて、1, 2 間にも有意差ありとされている点である。テューキー・クレーマーの方法の閾値 $q_0 = 2.498$ であったのに対してフィッシャーの PLSD 法では閾値 $t_0 = 2.064$ と低くすることができる。第 1 種の過誤の確率を公称の有意水準以下に抑えながら、母平均の差に対する検定統計量の閾値を小さくできることは、母平均に差があるのに差がないと誤判定してしまう確率（第 2 種の過誤の確率）を小さくできることとなり望ましい。この PLSD 法では式 (9.2) の仮説の下での第 1 種の過誤の確率は事前の分散分析により公称の有意水準 $\alpha = 0.05$ に抑えられている。このことは次項のシミュレーションによっても確認できる。

9.2.2 フィッシャーの PLSD 法のシミュレーション

図 8.19 のテューキー・クレーマーの方法のシミュレーションと同様に、データ数がそれぞれ 11 個、9 個、7 個の正規乱数のデータ群 1, 2, 3 を生成させる。そして、各群の母平均の差の検定にフィッシャーの PLSD 法を適用することを 1000 組繰り返して第 1 種の過誤の頻度を求めるシミュレーションを行った。図 9.6 はこのシミュレーションを 100 回繰り返したときの第 1 種の過誤の確率の 95% 信頼区間の出力画面を示す。フィッシャーの PLSD 法により第 1 種の過誤の確率が公称の有意水準 $\alpha = 0.05$ 辺りにある結果が得られている。

	A	B	C	D
1	母平均の差の検定(フィッシャーのPLSD法)			
28	(t検定)	平均値x1	x2	x3
29		3.318	3.189	3.129
30		平均平方 νE	最小有意差 t_0	
31		0.0133	2.064	
32		検定統計量 t_{12}	t_{13}	t_{23}
33		2.49	3.40	1.04
34	出力	if $p < \alpha$ and $ t_{12} > t_0$ then 1 else 0	if $p < \alpha$ and $ t_{13} > t_0$	if $p < \alpha$ and $ t_{23} > t_0$
35		1	1	0

図 9.5: 母平均の差の検定の例（フィッシャーの PLSD 法）（フィッシャーの PLSD による検定.xlsx）

53		有意頻度 fr_k	100回の平均		
54	fr1	0.049	0.0498		
55	fr2	0.055	95%信頼区間		
56	fr3	0.05	0.0485	$\leq p^* \leq$	0.0512

図 9.6: 母平均の差の検定，フィッシャーの PLSD 法における第 1 種の過誤の確率の 95% 信頼区間（フィッシャーの PLSD によるシミュレーション.xlsm）

9.2.3 フィッシャーの PLSD 法の理論

ここでは考え方を述べるにとどめる。厳密な証明は文献 [1] を参照されたい。検定の仮説は式 (9.1) により与えられているものとする。3 群の母平均値に差がない場合には分散分析により第 1 種の過誤の確率は公称の有意水準 α に抑えられる。いま、仮にデータ群 1 の母平均が大きく $\mu_1 \gg \mu_2, \mu_3$ であったとしよう。このとき分散分析の結果は μ_2, μ_3 の大小に関係なく有意差ありとなり、次のステューデントの t 検定により各群の平均値の差が比較される。この下位検定の有意水準は分散分析と同じ α である。このとき、 μ_1, μ_2 間、 μ_1, μ_3 間はステューデントの t 検定においても有意差ありと出る。これらの母平均の間にはもともと差があったのでこの結果は問題ではない。では、 μ_2, μ_3 間の差はどうであろうか？この両群における第 1 種の過誤の確率はステューデントの t 検定により有意水準 α に抑えられる。フィッシャーの PLSD 法は第 1 種の過誤の確率を公称の有意水準に抑えることができる手法である。

9.3 母平均の差の検定（4群以上の場合の下位検定法）

フィッシャーの PLSD 法が適用可能なのは 3 群間の母平均の差の検定のみである。これを 4 群以上の場合に適用すれば第 1 種の過誤の確率が公称の有意水準を上回ってしま

う。4群以上の場合に適用できるように改良された下位検定法がヘイター・フィッシャー (Hayter Fisher) の改良法 [1] である。

9.3.1 ヘイター・フィッシャーの改良法による検定の例

ある日、ある工場のある製品のラインにおいて、ラインのスピードを4通りに変えながらあるパラメータの測定がなされ、スピード設定1, 2, 3, 4からそれぞれ11, 9, 7, 5個の標本を取り出すことができ、以下の値が得られたとする。

設定1: 3.3, 3.2, 3.4, 3.3, 3.3, 3.2, 3.5, 3.2, 3.2, 3.4, 3.5

設定2: 3.2, 3.1, 3.3, 3.2, 3.2, 3.1, 3.4, 3.0, 3.2

設定3: 3.1, 3.0, 3.3, 3.1, 3.2, 3.0, 3.2

設定4: 3.2, 3.3, 3.1, 3.1, 3.2

ライン1, 2, 3, 4の平均値 $\bar{x}_1, \bar{x}_2, \bar{x}_3, \bar{x}_4$ はそれぞれ3.32, 3.19, 3.13, 3.18であった。スピード設定を変えたことで平均値に差が出たと言えるか？ただし、いずれの母分散にも差がないとみなしてよいとする。

検定の仮説は、データ群 i の母平均を μ_i とすると

$$\text{帰無仮説} : \mu_i = \mu_j$$

$$\text{対立仮説} : \mu_i \neq \mu_j$$

$$\text{ただし, } i, j = 1, 2, 3, 4, i \neq j \quad (9.26)$$

である。この仮説の下で、まず、分散分析を適用する。帰無仮説を棄却できた場合には、8.2節のテューキー・クレーマーの方法を適用する。ただし、検定統計量の閾値を数表から求める際にはデータ群数を $a - 1$ とする。設定 i と設定 j の平均値をそれぞれ \bar{x}_i, \bar{x}_j とすると、検定統計量 q_{ij} は次式により与えられる。

$$q_{ij} = \frac{|\bar{x}_i - \bar{x}_j|}{\sqrt{v_E^2 \left(\frac{1}{n_i} + \frac{1}{n_j} \right)}} \quad (9.27)$$

ただし、

$$v_E^2 = \frac{\sum_{k=1}^a \sum_{l=1}^{n_k} (x_{kl} - \bar{x}_k)^2}{\sum_{k=1}^a (n_k - 1)} \quad (9.28)$$

である. $n_k (k = 1, 2, 3, 4)$ は各設定のデータ数であり, a は設定の数である. この q_{ij} 値は自由度 ν が

$$\nu = \sum_{k=1}^a (n_k - 1) \quad (9.29)$$

である t 分布に従う.

検定統計量の閾値 q_0 はデータ群数 $a - 1$, 自由度 ν のステューデント化された範囲の 5 (もしくは 1) % 点とする. この改良法は, 母平均に差がある場合には, データ群数として a を用いる 8.2 節の方法と比較すると, 閾値 q_0 を小さく (緩く) できるため有意差を検出しやすい.

図 9.7, 図 9.8 はヘイター・フィッシャーの改良法による検定の実施例である. セル C4 に閾値 q_0 が入力されている. データ群数 $a = 4$, 自由度 $\nu = 28$ であるので, 表 A.1 において, $a - 1 = 3, \nu = 28$ のとき, ステューデント化された範囲の 5% 点は 3.499 と読める. これより, $q_0 = 3.499/\sqrt{2} = 2.474$ である. セル B28 では, 分散分析により平均値間に有意な差があるという結果が得られ, 1 を出力している. 30 行目以下においてヘイター・フィッシャーの改良法による検定を行っている. 35, 37 行目において式 (9.27) による検定統計量を求め, 39, 41 行目で q_0 値との比較を行い, 有意差ありの場合に 1 を出力している. 設定 1, 2 間, 1, 3 間に有意差ありという結果が得られている.

	A	B	C	D	E
1	母平均の差の検定 (ヘイター・フィッシャーの改良法)				
2					
3	入力	公称の有意水準 α	調整 q_0		
4		0.05	2.474	= 3.499/ $\sqrt{2}$	
5					
6		設定1	設定2	設定3	設定4
7		3.3	3.2	3.1	3.2
8		3.2	3.1	3.0	3.3
9		3.4	3.3	3.3	3.1
10		3.3	3.2	3.1	3.1
11		3.3	3.2	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	データ群数 a	自由度 ν	組数 G	
20		4	28	1	
21		データ数 n_1	n_2	n_3	n_4
22		11	9	7	5

図 9.7: 母平均の差の検定 (ヘイター・フィッシャーの改良法) (ヘイター・フィッシャーによる検定.xlsx)

図 9.9 はフィッシャーの PLSD 法をむりやり 4 群の平均値の差の検定に適用した例を示す. セル C4 に最小有意差 t_0 が入力されている. これは公称の有意水準 $\alpha = 0.05$, 自由度

23	(分散分析)	全変動 ST	水準間変動 SA	水準内変動 SE	
24		0.5287	0.1812	0.3475	
25		F 値	p 値		
26		4.8665	0.0076		
27		if $p < \alpha$			
28		1			
29					
30	(ヘイター・フィッシャー)	平均値 x1	x2	x3	x4
31		3.32	3.19	3.13	3.18
32		平均平方 νE			
33		0.012			
34		検定統計量 q12	q13	q14	
35		2.582	3.520	2.300	
36		q23	q24	q34	
37		1.074	0.143	-0.788	
38	出力	if $ q12 > q0$	if $ q13 > q0$	if $ q14 > q0$	
39		1	1	0	
40		if $ q23 > q0$	if $ q24 > q0$	if $ q34 > q0$	
41		0	0	0	

図 9.8: 母平均の差の検定 (ヘイター・フィッシャーの改良法 (つづき))

$\nu = 28$ のときの $TINV()$ 関数の値である。分散分析により平均値間に有意差ありという結果が得られ、かつ、ステューデントの t_{ij} 値が最小有意差 t_0 を上回った場合に設定 i, j 間に有意差ありとして 1 を出力している。設定 1, 2 間, 1, 3 間, 1, 4 間に有意差ありという結果が得られている。しかし、これは次項に見るように第 1 種の過誤の確率が公称の有意水準 α を上回ってしまった結果であり、間違っている。

	A	B	C	D	E
1	母平均の差の検定(フィッシャーのPLSD法を4群に適用)				
2					
3	入力	公称の有意水準 α	調整 t_0		
4		0.05	2.048		
5					
6		x1	x2	x3	x4
7		3.3	3.2	3.1	3.2
8		3.2	3.1	3.0	3.3
34	(フィッシャーのPLSD)	検定統計量 t12	t13	t14	
35		2.582	3.520	2.300	
36		t23	t24	t34	
37		1.074	0.143	-0.788	
38	出力	if $ q12 > t_0$	if $ q13 > t_0$	if $ q14 > t_0$	
39		1	1	1	
40		if $ q23 > t_0$	if $ q24 > t_0$	if $ q34 > t_0$	
41		0	0	0	

図 9.9: 母平均の差の検定 (フィッシャーの PLSD 法をむりやり 4 群に適用) (フィッシャーの PLSD による検定 (4 群) .xlsx)

図 9.10 はテューキー・クレーマーの方法を 4 群の平均値の差の検定に適用した例である。分散分析を行っていないので、セル C4 の検定統計量の閾値 q_0 にはデータ群数 $a = 4$, 自由度 $\nu = 28$ のステューデント化された範囲の 5% 点の値 (表 A.1 より 3.861) が入力されている。33 行目以下では式 (8.45) の検定統計量 q_{ij} が閾値 q_0 を超えた場合に有意差あ

りとして1を出力している。設定1, 3間にのみ有意差ありという結果が得られている。

これに対し図9.8のヘイター・フィッシャーの改良法の結果では設定1, 2間と1, 3間に有意差ありの結果が得られていた。これはヘイター・フィッシャーの改良法では、母平均に差があり、分散分析にて有意差ありと判定された場合には、その後の個々の母平均の差の検定において、テューキー・クレーマーの方法よりも閾値 q_0 を小さく（緩く）設定できることによる。第1種の過誤の確率を公称の有意水準 α 以下に抑えながら、閾値 q_0 を小さくできることは第2種の過誤（母平均に差があるのに差がないとする誤り）の確率を小さくできることを意味し、望ましい。

	A	B	C	D	E
1	母平均の差の検定(テューキー・クレーマーの方法を4群に適用)				
2					
3	入力	公称の有意水準 α	調整 q_0		
4		0.05	2.730	=3.861/ $\sqrt{2}$	
28		検定統計量 q_{12}	q_{13}	q_{14}	
29		2.582	3.520	2.300	
30		q_{23}	q_{24}	q_{34}	
31		1.074	0.143	-0.788	
32	出力	if $ q_{12} > q_0$	if $ q_{13} > q_0$	if $ q_{14} > q_0$	
33		0	1	0	
34		if $ q_{23} > q_0$	if $ q_{24} > q_0$	if $ q_{34} > q_0$	
35		0	0	0	

図 9.10: 母平均の差の検定（テューキー・クレーマーの方法を4群に適用）（テューキー・クレーマーによる検定（4群）.xlsx）

9.3.2 ヘイター・フィッシャーの改良法のシミュレーション

ヘイター・フィッシャーの改良法を用いたシミュレーションのExcelの画面を図9.11, 図9.12に示す。データ群1, 2, 3, 4に平均 $\mu_1, \mu_2, \mu_3, \mu_4$, 標準偏差 $\sigma_1, \sigma_2, \sigma_3, \sigma_4$ の正規乱数をそれぞれ11, 9, 7, 5個生成させ、公称の有意水準 $\alpha = 0.05$ で分散分析を行った後、35行目以下でヘイター・フィッシャーの改良法による検定を実施している。このときの閾値 q_0 はデータ群数 $a - 1 = 3$, 自由度 $\nu = 28$ の場合のステューデント化された範囲の5%点の値が入力されている。44行目と46行目では分散分析の結果にかかわらず、検定統計量 q_{ij} が閾値 q_0 を上回った場合に1を出力している。48行目では分散分析で有意差ありとされ、かつ、 q_{23}, q_{24}, q_{34} のいずれかが閾値を上回った場合に1を出力している。このシミュレーションを1000組について実施し、48行目の1の数を数えて、分散分析で有意差ありとされ、かつ、 q_{23}, q_{24}, q_{34} のいずれかが閾値を上回った頻度を求めてセルB50に出力している。

4群の平均値が等しい場合、第1種の過誤の確率は分散分析により公称の有意水準 $\alpha =$

0.05に抑えられる。そこで、このシミュレーションでは $\mu_1 = 5, \mu_2 = \mu_3 = \mu_4 = 1$ としてデータ群1の平均値を他群より大きくし、分散分析の結果がほぼ全ての組で有意差ありとなる設定としてある。この場合、検定統計量 q_{23}, q_{24}, q_{34} のいずれかが閾値を上回った頻度が第1種の過誤の頻度となる。図9.13は、以上の1000組のシミュレーションを100回繰り返して得られた第1種の過誤の確率の95%信頼区間を示す。ヘイター・フィッシャーの改良法により第1種の過誤の真の確率 p^* は5% 辺りにあることが分かる。

	A	B	C	D	E
1	母平均の差の検定のシミュレーション(ヘイター・フィッシャーの方法)				
2					
3	入力	μ_1	μ_2	μ_3	μ_4
4		5	1	1	1
5		σ_1	σ_2	σ_3	σ_4
6		1	1	1	1
7		公称の有意水準 α	調整 q_0		
8		0.05	2.474		
9					
10	計算値	1組目			
11		x1	x2	x3	x4
12		5.938	1.144	1.083	2.948
13		2.276	1.840	0.303	2.388

図 9.11: 母平均の差の検定のシミュレーション (ヘイター・フィッシャーの改良法) (ヘイター・フィッシャーによるシミュレーション.xlsx)

35	(ヘイター・フィッシャー)	平均値x1	x2	x3	x4
36		4.783	1.257	0.419	1.402
37		平均平方 ν^2			
38		1.057			
39		検定統計量 q_{12}	q_{13}	q_{14}	
40		7.631	8.781	6.098	
41		q_{23}	q_{24}	q_{34}	
42		1.619	-0.252	-1.633	
43		if $ q_{12} > q_0$ then 1 else 0	if $ q_{13} > q_0$ then 1 else 0	if $ q_{23} > q_0$ then 1 else 0	
44		1	1	1	
45		if $ q_{23} > q_0$ then 1 else 0	if $ q_{24} > q_0$ then 1 else 0	if $ q_{34} > q_0$ then 1 else 0	
46		0	0	0	
47		分散分析で有意差ありとされ、 かつ、 q_{23}, q_{24}, q_{34} のいずれ かで有意差あり			
48		0			
49	出力	有意頻度 fr			
50		0.051			

図 9.12: 母平均の差の検定のシミュレーション (ヘイター・フィッシャーの改良法 (つづき))

図9.14は、図9.11の設定においてヘイター・フィッシャーの改良法の部分をフィッシャーのPLSD法に置き換えて得られた第1種の過誤の確率の95%信頼区間を示す。データ群2, 3, 4間の平均値の差は分散分析では評価されないままに下位検定に渡され、これら3群に対してステューデントの t 検定を適用した場合には第1種の過誤の確率は公称の有意

55		有意頻度 frk		100回の平均		
56	fr1	0.058		0.0493		
57	fr2	0.058		95%信頼区間		
58	fr3	0.046		0.0478	< p* <	0.0507

図 9.13: 母平均の差の検定，ヘイター・フィッシャーの改良法における第1種の過誤の確率の95%信頼区間

水準 $\alpha = 0.05$ を超えてしまう。シミュレーション結果はこの考察に一致する結果となっている。

55		有意頻度 frk		100回の平均		
56	fr1	0.119		0.1151		
57	fr2	0.114		95%信頼区間		
58	fr3	0.095		0.1132	< p* <	0.1170

図 9.14: 母平均の差の検定，フィッシャーのPLSD法をむりやり4群に適用した場合の第1種の過誤の確率の95%信頼区間（フィッシャーのPLSDによるシミュレーション（4群）.xlsm）

9.3.3 ヘイター・フィッシャーの改良法の理論

ここでは考え方だけを述べる。証明の詳細は文献 [1] を参照されたい。 a 個のデータ群の平均値が等しい場合，第1種の過誤の確率は分散分析により公称の有意水準 $\alpha = 0.05$ に抑えられる。 a 個のデータ群の中の一つの群の平均値が他の群と比べて大きく異なっていて，分散分析ではほぼ100%の確率で有意差ありとされる場合を考える。このとき $a - 1$ 個のデータ群間の平均値の差は分散分析では評価されないままに下位検定が実施される。したがって，下位検定は $a - 1$ 個のデータ群間の平均値の差の検定において第1種の過誤の確率を公称の有意水準 α 以下に抑える多重比較法でなければならない。文献 [1] では，データ群数を $a - 1$ としてテューキー・クレーマーの方法を適用すれば，第1種の過誤の確率は公称の有意水準 α 以下となることが証明されている。なお，データ群数 $a = 3$ のときヘイター・フィッシャーの改良法はフィッシャーのPLSD法に一致する。

9.4 母平均の差の検定（比較対象の組合せを自在にする検定法）

以上の検定法はデータ群対に対する検定法である。データ群の対の検定だけでなく，ある特定のデータ群とそれ以外のデータ群間の母平均の差の検定，もしくは複数のデータ

群のグループ間の母平均の差の検定などデータ群の組み合わせを自在に変えた検定を可能とする検定法にシェフェ(Scheffe)の方法がある。

9.4.1 シェフェの方法による検定の例

9.3.1 項と同じ問題に対して次の検定の仮説を考える。

$$\begin{aligned} \text{帰無仮説} &: C_1\mu_1 + C_2\mu_2 + C_3\mu_3 + C_4\mu_4 = 0 \\ \text{対立仮説} &: C_1\mu_1 + C_2\mu_2 + C_3\mu_3 + C_4\mu_4 \neq 0 \end{aligned} \quad (9.30)$$

係数 $C_i, i = 1, \dots, 4$ には次の条件を設ける。

$$\begin{aligned} C_1 + C_2 + C_3 + C_4 &= 0 \\ C_1^2 + C_2^2 + C_3^2 + C_4^2 &= 1 \end{aligned} \quad (9.31)$$

上式の条件を満たす限り、 C_i は任意の実数をとることができる。このときの検定統計量は

$$S = \frac{(\sum_{i=1}^a C_i \bar{x}_i)^2}{(a-1) \left(\sum_{i=1}^a \frac{C_i^2}{n_i} \right) v_E^2} \quad (9.32)$$

である。ただし、 \bar{x}_i はデータ群 i の平均値、 a はデータ群数、 n_i はデータ群 i のデータ数である。また、平均平方 v_E^2 は

$$v_E^2 = \frac{\sum_{i=1}^a \sum_{j=1}^{n_i} (x_{ij} - \bar{x}_i)^2}{\sum_{i=1}^a (n_i - 1)} \quad (9.33)$$

である。 x_{ij} はデータ群 i 内の j 番目のデータである。

この S 値は係数 C_i の値によって変わるが、 C_i が式 (9.31) の条件を満たす限り、 S 値の最大値は自由度 $a-1, \nu$ の F 分布 F_ν^{a-1} に従う。ここで、自由度 $\nu = \sum_{i=1}^a (n_i - 1)$ である。したがって、 F_ν^{a-1} により公称の有意水準 α に対応する閾値 S_0 を定めれば、これを超える S 値による第1種の過誤の確率は公称の有意水準を上回ることはない。

図 9.15 は検定を実行している Excel の画面の抜粋である。6 行目で係数 C_i の値が入力されている。26 行目にて係数 C_i が式 (9.31) の条件を満たしていることを確認している。このとき検定の仮説は

$$\text{帰無仮説} : \frac{3}{2\sqrt{3}}\mu_1 - \frac{1}{2\sqrt{3}}(\mu_2 + \mu_3 + \mu_4) = 0 \quad (9.34)$$

である。画面では省略してあるが、参考に公称の有意水準 $\alpha = 0.05$ のもとで分散分析を実施している。分散分析により母平均の間に有意差ありという結果が得られている。セル

D22ではFINV()関数を用いて S 値の閾値 S_0 を求めている。これは、自由度3,28の F 分布の上側5%点の値である。シェフェの方法による検定統計量 S の値はセルB36にて求められている。 S 値は閾値 S_0 を超えている。したがって、帰無仮説は棄却でき、 μ_1 と μ_2, μ_3, μ_4 の間には差があるという結果が得られた。

	A	B	C	D	E
1	母平均の差の検定(シェフェの方法)				
2					
3	入力	公称の有意水準 α			
4		0.05			
5		C1	C2	C3	C4
6		0.8660	-0.2887	-0.2887	-0.2887
21	計算値	データ群数 a	自由度 ν	閾値 s_0	
22		4	28	2.947	
23		データ数 n1	n2	n3	n4
24		11	9	7	5
25		C1+C2+C3+C4	C1^2+O2^2+C3^2+O4^2		
26		0	1		
31	(シェフェ)	平均値X1	平均値X2	平均値X3	平均値X4
32		3.32	3.19	3.13	3.18
33		平均平方 $\nu^2 E$			
34		0.012			
35	出力	検定統計量 s			
36		4.411			

図 9.15: 母平均の差の検定（シェフェの方法）（シェフェによる検定.xlsx）

	A	B	C	D	E
1	母平均の差の検定(シェフェの方法)				
2					
3	入力	公称の有意水準 α			
4		0.05			
5		C1	C2	C3	C4
6		0.7071	0.0000	-0.7071	0.0000
21	計算値	データ群数 a	自由度 ν	閾値 s_0	
22		4	28	2.947	
23		データ数 n1	n2	n3	n4
24		11	9	7	5
25		C1+C2+C3+C4	C1^2+O2^2+C3^2+O4^2		
26		0	1		
31	(シェフェ)	平均値X1	平均値X2	平均値X3	平均値X4
32		3.32	3.19	3.13	3.18
33		平均平方 $\nu^2 E$			
34		0.012			
35	出力	検定統計量 s			
36		4.130			

図 9.16: 母平均の差の検定の例（シェフェの方法（その2））（シェフェによる検定（その2）.xlsx）

係数 C_i を変えることで平均値間の様々な検定を実行できる。図9.16は $C_1 = 1/\sqrt{2}$, $C_2 =$

$0, C_3 = -1/\sqrt{2}, C_4 = 0$ として、平均値 \bar{x}_1, \bar{x}_3 の差の検定を実行している Excel の画面である。このとき閾値 S_0 は C_i によらず一定である。 μ_1, μ_3 間に有意差があるという結果が得られている。

9.4.2 シェフェの方法のシミュレーション

図9.17, 図9.18は、係数 C_i の組を C_1, C_2, C_3 と C_x, C_y, C_z の2組を用意して、データ群間の組み合わせをデータ群1対データ群2, 3を1:2-3と略記することとして、1:2-3, 2:3-1, 3:1-2, 1:2, 1:3, 2:3と可能な組み合わせの全てについてシェフェの方法を適用しているシミュレーションの画面である。データ群1, 2, 3にそれぞれ正規乱数を11, 9, 7個生成させて、分散分析およびシェフェの方法による検定を1000組について実行している。セルE27では分散分析及びシェフェの方法の共通の閾値 S_0 を自由度2, 24として $\text{FINV}()$ 関数により求めている。

35行目では分散分析の結果 F 値が S_0 を上回った場合に1を出力している。41行目および43行目で各データ群の組み合わせにおける S 値を求め、45行目に S 値の中の最大値 S_{max} を出力している。47行目では、 S_{max} が S_0 を上回った場合に1を出力している。49行目の値は以上の検定を1000組について繰り返した結果、分散分析とシェフェのそれぞれの方法で有意差ありとされた頻度を求めている。これらはそれぞれ式(9.2), 式(9.30)の仮説の下での第1種の過誤の頻度である。

図9.19は以上の1000組のシミュレーションを100回繰り返して得られた第1種の過誤の確率 p^* の95%信頼区間を示す。毎回のシミュレーションにおいて、シェフェの方法により有意となる頻度は分散分析のそれよりも小さく、後述の式(9.41)と一致する結果である。シェフェの方法は p^* の値を公称の有意水準 α よりも少し小さな値としている。図9.20は、1000組のシミュレーションにおける各組の S_{max} 値の頻度分布を示す。 S_{max} 値は小数点以下を四捨五入して、各整数値の出現割合を示してある。参考に F 値の頻度分布および自由度2, 24の F 分布 F_{24}^2 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 S_{max} の頻度はその値の大きな ($S_{max} \geq 4$) 領域で F 値よりも小さな傾向を示すことを見て取ることもできる。

9.4.3 シェフェの方法の理論

互いに独立な事象 $X_{ij} (i = 1, \dots, a, j = 1, \dots, n_i) \sim N(\mu, \sigma^2)$ を考える。各事象の母平均 μ , 母分散 σ^2 は等しいとする。 a はデータ群数, n_i はデータ群 i のデータ数である。係

(差し替えです)

1	母平均の差の検定のシミュレーション(シェフェの方法)						
2							
3	入力	称の有意水準 α					
4		0.05					
5		C1	C2	C3	Cx	Cy	Cz
6		0.8165	-0.4082	-0.4082	0.7071	-0.7071	0.0000
7		$\mu 1$	$\mu 2$	$\mu 3$			
8		0	0	0			
9		$\sigma 1$	$\sigma 2$	$\sigma 3$			
10		1	1	1			
11							
12	計算値	1組目			2組目		
13		x1	x2	x3	x1	x2	x3
14		-0.692	0.809	1.232	0.109	0.817	0.273
15		1.357	-0.847	-0.551	0.335	0.098	0.263

26		データ群数 a	自由度 ν	組数 G	閾値 S_0		
27		3	24	1000	3.4028		
28		データ数 n1	n2	n3			
29		11	9	7			

図 9.17: 母平均の差の検定のシミュレーション（シェフェの方法）（シェフェによるシミュレーション.xlsxm）

(差し替えです)

30	(分散分析)	全変動 ST	水準間変動 SA	水準内変動 SE	全変動 ST	水準間変動 SA	水準内変動 SE
31		26.7398	6.0017	20.7380	23.6034	0.8631	22.7403
32		F値			F値		
33		3.4729			0.4555		
34		if(F > S0)			if(F > S0)		
35		1			0		
36	(シェフェ)	平均値x1	x2	x3	平均値x1	x2	x3
37		0.58	-0.52	0.10	0.61	0.50	0.17
38		平均平方 $\nu^2 E$			平均平方 $\nu^2 E$		
39		0.864			0.948		
40		1対2, 3	2対3, 1	3対1, 2	1対2, 3	2対3, 1	3対1, 2
41		2.34	2.53	0.02	0.27	0.04	0.41
42		1対2	1対3	2対3	1対2	1対3	2対3
43		3.47	0.57	0.88	0.03	0.44	0.23
44		Smax			Smax		
45		3.47			0.44		
46		if Smax > S0 then 1 else 0			if Smax > S0 then 1 else 0		
47		1			0		
48	出力	分散分析の 有意頻度 frF	シェフェの 有意頻度 frS				
49		0.054	0.051				

図 9.18: 母平均の差の検定のシミュレーション（シェフェの方法（つづき））

数 C_i は次の条件を満たすとする.

$$\sum_{i=1}^a C_i = 0$$

$$\sum_{i=1}^a C_i^2 = 1 \tag{9.35}$$

(差し替えです)

	分散分析の 有意頻度frFk	シェフェの 有意頻度frSk	分散分析 100回の平均		
54					
55	fr1	0.041	0.038	0.0497	
56	fr2	0.032	0.029	95%信頼区間	
57	fr3	0.049	0.045	0.0482	≤ p* ≤ 0.0512
58		0.054	0.05		
59		0.047	0.046	シェフェ 100回の平均	
60		0.053	0.051	0.0466	
61		0.035	0.034	95%信頼区間	
62		0.064	0.058	0.0452	≤ p* ≤ 0.0480

図 9.19: 母平均の差の検定, シェフェの方法における第1種の過誤の確率の95%信頼区間

(差し替えです)

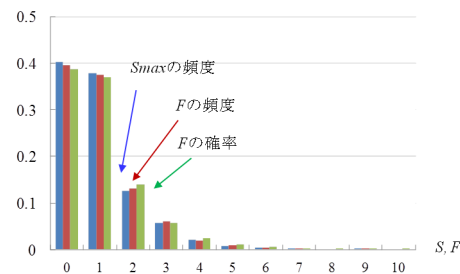


図 9.20: 母平均の差の検定, シェフェの方法におけるS値の頻度分布 (S値の分布 (シェフェ) .xlsx)

シェフェの方法の検定統計量 S は, 式 (9.32) より

$$S = \frac{(\sum_{i=1}^a C_i \bar{x}_i)^2}{(a-1) \left(\sum_{i=1}^a \frac{C_i^2}{n_i} \right) v_E^2} \quad (9.36)$$

である. \bar{x} を全事象の平均値 ($= \sum_{i=1}^a \sum_{j=1}^{n_i} x_{ij} / \sum_{i=1}^a n_i$) とすると, 式 (9.35) より

$$\sum_{i=1}^a C_i \bar{x} = 0 \quad (9.37)$$

であることに注意して,

$$\begin{aligned} S &= \frac{\left\{ \sum_{i=1}^a C_i (\bar{x}_i - \bar{x}) \right\}^2}{(a-1) \left(\sum_{i=1}^a \frac{C_i^2}{n_i} \right) v_E^2} \\ &= \frac{\left\{ \sum_{i=1}^a \frac{C_i}{\sqrt{n_i}} \sqrt{n_i} (\bar{x}_i - \bar{x}) \right\}^2}{(a-1) \left(\sum_{i=1}^a \frac{C_i^2}{n_i} \right) v_E^2} \end{aligned} \quad (9.38)$$

と変形できる. この式に, 次のシュバルツの不等式

$$x_1 y_1 + x_2 y_2 \leq \sqrt{x_1^2 + x_2^2} \sqrt{y_1^2 + y_2^2} \quad (9.39)$$

を適用すると

$$\begin{aligned} S &\leq \frac{\left(\sum_{i=1}^a \frac{C_i^2}{n_i}\right) \left(\sum_{i=1}^a n_i (\bar{x}_i - \bar{x})^2\right)}{(a-1) \left(\sum_{i=1}^a \frac{C_i^2}{n_i}\right) v_E^2} \\ &= \frac{\sum_{i=1}^a n_i (\bar{x}_i - \bar{x})^2}{(a-1) v_E^2} \end{aligned} \quad (9.40)$$

となる. 式 (9.33), (9.4) より

$$S \leq \frac{\frac{S_A}{a-1}}{\frac{S_E}{\sum_{i=1}^a (n_i - 1)}} \quad (9.41)$$

となり, 右辺は式 (9.3) の検定統計量 F に一致する. シェフェの方法の検定統計量 S の最大値 S_{max} は自由度 $a-1, \sum_{i=1}^a (n_i - 1)$ の F 分布の値を上限とする.

9.5 シェフェの方法とヘイター・フィッシャーの方法の比較

分散分析の下位検定法としてフィッシャーの PLSD 法, ヘイター・フィッシャーの改良法を紹介した. ヘイター・フィッシャーの改良法はデータ群数 $a=3$ のときフィッシャーの PLSD 法と一致する. ヘイター・フィッシャーの改良法はデータ群の対比較のための下位検定法である. 一方, シェフェの方法はデータ群の任意の組み合わせに対する検定法である. シェフェの方法は, 前項, 前々項に見たように分散分析よりも有意差の出にくい検定法である. したがって, シェフェの方法を分散分析の下位検定に用いても間違いではない. では, シェフェの方法を対比較のためだけに適用した場合にはヘイター・フィッシャーの改良法と比べてどちらが有効な検定法と言えるだろうか?

この疑問に答えるために, まず式 (9.32) のシェフェの方法の検定統計量 S を見直す.

$$\begin{aligned} S &= \frac{\left(\sum_{i=1}^a C_i \bar{x}_i\right)^2}{(a-1) \left(\sum_{i=1}^a \frac{C_i^2}{n_i}\right) v_E^2} \\ &= \frac{\left\{\sum_{i=1}^a C_i (\bar{x}_i - \mu)\right\}^2}{\sum_{i=1}^a \frac{C_i^2 \sigma^2}{n_i}} \\ &= \frac{1}{(a-1) \frac{1}{\sigma^2} v_E^2} \end{aligned} \quad (9.42)$$

ただし, μ は母平均であり全てのデータ群で等しいとしている. $\sqrt{(a-1)S}$ を求めると

$$\sqrt{(a-1)S} = \frac{\frac{\sum_{i=1}^a C_i (\bar{x}_i - \mu)}{\sqrt{\sum_{i=1}^a \frac{C_i^2 \sigma^2}{n_i}}}}{\sqrt{\frac{1}{\sigma^2} v_E^2}} \quad (9.43)$$

となる。右辺の分子は式(3.63)～式(3.67)より、

$$\frac{\sum_{i=1}^a C_i(\bar{X}_i - \mu)}{\sqrt{\sum_{i=1}^a \frac{C_i^2 \sigma^2}{n_i}}} \sim N(0, 1) \quad (9.44)$$

であり、標準正規分布に従う。また、式(9.43)の分母において、式(6.78), (4.125)より

$$\frac{1}{\sigma^2} \left(\sum_{i=1}^a (n_i - 1) \right) v_E^2 \sim \chi_\nu^2 \quad (9.45)$$

である。ただし、 $\nu = \sum_{i=1}^a (n_i - 1)$ である。したがって、式(4.120)の t 分布の定義より

$$\sqrt{(a-1)S} \sim t_\nu \quad (9.46)$$

である。検定統計量 $\sqrt{(a-1)S}$ は自由度 ν の t 分布に従う。したがって、式(9.41)の右辺の値に $a-1$ を掛けてルートをとって得られる閾値 $\sqrt{(a-1)S_0}$ は、係数 C_i を様々に変えて比較を行った場合に(t_ν 分布に従う)検定統計量 $\sqrt{(a-1)S}$ がとり得る値の上限値である。

一方、ヘイター・フィッシャーの改良法において式(9.27)の検定統計量 q_{ij} 値も t_ν 分布に従う。この場合の閾値 q_0 はデータ群の全ての対を比較した場合に、(t_ν 分布に従う)検定統計量 q_{ij} がとり得る値の上限値(ステューデント化された範囲の%点)である。そこで、同じデータ群数 a 、自由度 ν の場合の閾値 $\sqrt{(a-1)S_0}$ 、 q_0 を比較してみる。図9.21はその比較を実行しているExcelの画面である。データ群数 $a = 3 \sim 9$ 、自由度 $\nu = 2 \sim 50$ の各組み合わせにおいて左側にシェフェの方法の閾値 $\sqrt{(a-1)S_0}$ を、右側にヘイター・フィッシャーの改良法の閾値 q_0 を並べて示してある。閾値 $\sqrt{(a-1)S_0}$ の計算では、公称の有意水準 $\alpha = 0.05$ 、自由度 $a-1, \nu$ に対応する F 値をFINV()関数により求めて、これを S_0 としている。閾値 q_0 はSheet2にステューデント化された範囲の5%点の表をコピーして用意し、これを読み出して $1/\sqrt{2}$ を掛けて表示している。いずれの組み合わせにおいてもシェフェの $\sqrt{(a-1)S_0}$ 値の方がヘイター・フィッシャーの q_0 値より大きな値となっている。したがって、対比較だけに限定してシェフェの方法を適用すると、ヘイター・フィッシャーの改良法よりも有意差が出にくいことが分かる。

9.6 まとめ

本章では検定統計量そのものを見直すアプローチである分散分析とシェフェの方法を解説した。分散分析の検定統計量は多重比較を考慮して導出された、その p 値は式(9.2)の仮説の下で第一種の過誤の確率に一致した。分散分析ではいずれかの群の母平均間に

	A	B	C	D	E	F	G
1	シェフェの方法とヘイター・フィッシャーの改良法の閾値の比較						
2							
3	公称の有意水準 α						
4	0.05						
5							
6	データ群数 a						
7		3		4		5	
8	自由度 ν	シェフェ	ヘイター・フィッシャー	シェフェ	ヘイター・フィッシャー	シェフェ	ヘイター・フィッシャー
9	2	6.164	4.303	7.582	5.891	8.774	6.928
10	3	4.371	3.183	5.275	4.179	6.039	4.826
11	4	3.727	2.776	4.447	3.564	5.055	4.071
12	5	3.402	2.570	4.028	3.254	4.557	3.690
13	6	3.207	2.447	3.778	3.068	4.258	3.462
14	7	3.078	2.365	3.611	2.945	4.060	3.310
15	8	2.986	2.306	3.493	2.857	3.918	3.202
16	9	2.918	2.262	3.404	2.792	3.812	3.122
17	10	2.865	2.228	3.335	2.741	3.730	3.060
18	11	2.822	2.201	3.281	2.701	3.664	3.009
19	12	2.788	2.179	3.236	2.668	3.611	2.969
20	13	2.759	2.160	3.199	2.640	3.566	2.935
21	14	2.735	2.145	3.167	2.617	3.528	2.907
22	15	2.714	2.131	3.140	2.597	3.496	2.882

図 9.21: シェフェの方法とヘイター・フィッシャーの改良法の閾値の比較 (シェフェとヘイター・フィッシャーの閾値の比較.xlsx)

差があることを見いだせるが、どの母平均間に差があるかはわからない。そこで、分散分析で有意差ありと判定された場合には、個々の母平均の差に対して下位検定を適用する必要がある。下位検定をまとめると

(i) 3群の場合で、母分散が等しいと見なせる場合.

フィッシャーの PLSD 法

(ii) 3群以上の場合で、母分散が等しいと見なせる場合.

ヘイター・フィッシャーの改良法

である。一方、シェフェの方法は3群以上の場合で、母分散が等しいと見なせる場合に、データ群の任意の組み合わせ間の母平均の差の検定に適用できる。

参考文献

- [1] A. J. Hayter, "The Maximum Familywise Error Rate of Fisher's Least Significant Difference Test," *Journal of the American Statistical Association*, Vol. 81, No. 396, pp. 1000-1004, 1986.

索引

FDIST() 関数, 3

下位検定, 7

シェフェの方法, 16

事後検定, 7

事前検定, 7

水準, 4

水準間変動, 2

水準内変動, 2

全変動, 2

フィッシャーの PLSD 方, 7

分散分析, 1

ヘイター・フィッシャーの改良法, 10

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。