医学データの統計解析の基本

多重比較の方法

朝倉こう子*・濱崎俊光*

Fundamentals of statistical analysis in biomedical research : Multiple comparisons: comparing more than two treatments

1 新しい化合物や治療法の特徴・性能を評価する場合、たいていの目的において、それらの化合物や治療法を無治療や標準治療と対比するといった2つの群の比較によれば、明瞭で直裁的に理解しやすい形で得たい情報を入手できる。しかし、新規化合物のヒトへの投与量や投与期間に関する情報を得ることを目的として実施される毒性試験、例えば反復投与毒性試験では、被験化合物を動物に繰り返し投与したときに毒性変化を惹起する用量(とその変化の内容)や毒性変化の認められない用量を評価し、化合物の毒性像を明らかにするために、毒性変化が認められる用量と毒性変化が認められない用量(無毒性量)、また用量反応関係が評価できるように、少なくとも3段階の投与群が設定される。また、必要に応じて対照群、無処理対照群や陽性対照群が加えられる(医薬品毒性試験法ガイドラインの改正,1999)。この場合、上述の毒性試験の目的に沿えば、対照群や無処理対照群と化合物の投与群の比較、化合物の投与群間での比較といった多群の比較が主たる関心事である。多群の比較では、2群の比較の手法として第2回の「2つの平均の比較」(朝倉・濱崎,2015a)や、第3回の「2つの割合の比較」(朝倉・濱崎,2015b)で紹介したt検定やカイ2乗検定を、関心のある比較対の組み合わせのすべてに対し繰り返して用いることができる一方で、「差がないにもかかわらず差がある」と結論づける確率(第1種の過誤確率)が、あらかじめ定めた検定の有意水準よりも大きくなることが知られている。第4回では、多群比較において、t検定やカイ2乗検定を繰り返して用いるときにどのようなリスクが生じるかを例示しながら、そのリスクを回避するための方法として多重比較の考え方を概説し、代表的な多重比較の手法を紹介する。また、事例を通して統計計算ソフトウェアでの多重比較の方法の実行の仕方を概説する。

2 いま、処理効果を測る何らかの評価指標に基づき、複数の処理群(例えば異なる化合物の種類や用量など)の処理効果を比較することにおいて、関心のある比較対の組み合わせのすべてに対し、繰り返して単純に検定を用いることの懸念を例示する。この例示のため、表1に示すOtt(1988)で紹介されている体重減少量データ(本例示のために若干、データを改変している)を用いる。このデータでは、A、B、Cの3つの処理をそれぞれ10例の被験者に対して施し、処理の前後で体重がどの程度

* Koko Asakura, Toshimitsu Hamasaki
Office of Biostatistics and Data Management
Department of Advanced Medical Technology Development
National Cerebral and Cardiovascular Center
国立循環器病研究センター先進医療・治験推進部 DM/ 統計室
〒565-8565 吹田市藤白台5-7-1 E-mail: toshi.hamasaki@ncvc.go.jp

減少したかを評価したものである。A、B、Cの3つの処理について、それらの体重減少量の平均をそれぞれ μ_A 、 μ_B 、 μ_C とし、処理間で体重減少量が異なるかどうかを、 $H_{ABC} : \mu_A = \mu_B = \mu_C$ を帰無仮説として有意水準0.05で検討する。考えられる対比較の組み合わせは、「処理A 対 処理B」、「処理A 対 処理C」、「処理B 対 処理C」の3つであり、体重減少量が連続変数であるからt検定を用いて、次の3つの帰無仮説の $H_{AB} : \mu_A = \mu_B$ 、 $H_{AC} : \mu_A = \mu_C$ 、 $H_{BC} : \mu_B = \mu_C$ の取捨選択を有意水準0.05で行うとする。各対比較について、体重減少量に差がないことが真実である場合に、正しく帰無仮説を棄却しない確率は $1 - 0.05 = 0.95$ であり、3つの対比較についてt検定をそれぞれ適用すれば、これらのすべてにおいて正しい結論を導く確率は、 $(1 - 0.05)^3 = 0.857$ である。これから、3つの処理の間で体重減少量に差がないことが真である場合に、 H_{AB} 、 H_{AC} 、 H_{BC} のいずれかを棄却する確率は、 $1 - 0.857 = 0.143$ と計算できる。これから、3つの対比較の実質有意水準は0.143であり、名目有意水準の0.05よりも大きくなることがわかる。図1に、検定の回数といずれかの検定で有意な結果が得られる確率の挙動を示す。図1から、検定の数が増えるにともない、実質有意水準が大きくなり、差がないにもかかわらず、いずれかの比較において差を見いだす可能性が高くなることがわかる。また、例えば、医薬品の推奨用量を決めるといった場面において、 m 個の複数の処理のなかから最も好ましい値を示した処理を選択するとすれば、実際はすべての処理効果がゼロであっても、図2に示すように、 m が大きくなるにともない効果サイズを正の方向へ偏って推定することがわかる(Bretz et al., 2010)。加えて、図2からは処理数 m は推定される効果サイズの分布の位置のみならず、変動に対しても影響を及ぼすことも確認できる。

表1 3つの処理における体重減少量(Ott, 1988)

群	体重減少量(単位: ポンド)									
A	12.4	10.7	11.9	11.0	12.4	12.3	13.0	12.5	11.2	13.1
B	12.7	13.2	11.8	11.9	12.2	11.2	13.7	11.8	12.2	11.7
C	9.1	11.5	11.3	9.7	13.2	10.7	10.6	11.3	11.1	11.7

図1 異なる検定の回数と有意水準のもとでの第1種の過誤確率の挙動

図2 m 個の処理から最大の効果サイズをもつ処理を選んだ場合の効果サイズの確率密度

上記の問題へ対処するための、全体の有意水準をあらかじめ規定した値に保つ方法として、多重比較の方法が用いられる。実地で最も簡単に適用できる代表的な方法として、Bonferroniの方法がある。Bonferroniの方法は、全体の有意水準を α 、対比較の個数を k とすれば、それぞれの検定の有意水準を、全体の有意水準を対比較の個数でわった α/k とする(有意水準を調整する)、あるいは同等に各検定で計算された p 値を k 倍する(p 値を調整する)ものである。Bonferroniの方法は、簡単に少なくとも1つの正しい帰無仮説を棄却する確率を全体の有意水準以下に制御できる。一方で、Bonferroniの方法は、検討したい対比較の数が多かったり、検定統計量間に強い正の相関が認められるときには(例えば、先の体重減少量のデータにおいては、「処理A 対 処理B」、「処理A 対 処理C」の対比較で、処理Aが重複して用いられるため、これらの対比較で用いる2つの t 統計量の間にはそれぞれの処理群のデータ数に依存した相関が存在する)、実質的な第1種の過誤確率が名目水準を下回ることがあり、また検出力の低い方法であることが知られている(Dmitrienko et al.,2009)。

そこで、単純であるといった理由で安易にBonferroniの方法を用いるよりも、比較の目的に応じて適切な方法を選択することが望ましい。パラメトリックな方法の代表的なものとしては、Tukeyの方法(Tukey,1953; Kramer,1956)、Dunnnettの方法(Dunnnett,1955)、Schefféの方法(Scheffé,1953)がある。Tukeyの方法はすべての処理について対比較を行う場合で、先の体重減少量データでは、「処理A 対 処理B」、「処理A 対 処理C」、「処理B 対 処理C」の3つの対比較を実施する。Dunnnettの方法は特定の処理群、例えば無処理・陽性対照や標準薬を基準として対比較を行う場合で、先の体重減少量データで処理Cを基準とすれば、「処理A 対 処理C」、「処理B 対 処理C」の2つの対比較を実施する。Schefféの方法は、Tukeyの方法やDunnnettの方法よりもさらに一般的な対比較で用いられ、例えば、先の体重減少量データでは、「処理Aと処理B 対 処理C」の比較(帰無仮説は $(\mu_A + \mu_B)/2 = \mu_C$ である)を扱うことができる。表2に体重減少量データへこれらの多重比較の方法を適用した結果を示す。表2から対照群との対比較に関心がある場合に、Dunnnettの方法でなく、Tukeyの方法

や Scheffé の方法を適用すれば、関心のない比較も考慮されるため、Tukey の方法や Scheffé の方法の p 値は、Dunnnett の方法に比べると大きいことがわかる。このことから、比較の目的に応じて適切な方法を選択することの重要性が確認できる。

表2 体重減少量データへの多重比較の方法の適用

方法	A 対 B	対比較 B 対 C	C 対 A
通常の t 検定	0.5988	0.0106	0.0313
Bonferroni の方法	1.0000	0.0182	0.0545
Tukey の方法	0.8886	0.0161	0.0463
Dunnnett の方法	—	0.0114	0.0336
Scheffé の方法	0.8984	0.0216	0.0584

実地での多重比較法の適用において、1元分類分散分析の F 検定で統計的に有意な結果が得られたが、その後 Tukey の方法を適用すれば、いずれの対比較においても統計的に有意な結果が見いだされないといった、1元分類分散分析の F 検定と Tukey の方法の結果が一致しないことがある。 $H_{ABC} : \mu_A = \mu_B = \mu_C$ のもとで F 検定の結果が有意となる確率は α であるが、さらにいずれかの対比較で有意になる確率は α 以下である。永田・吉田(1997)にあるように、1元分類分散分析と多重比較の方法は基本的には別の方法であり、F 検定の結果に依存し多重比較の方法を適用することは、検定を繰り返す多重性の観点からも好ましいとはいえない。Scheffé の方法は通常の1元分類分散分析の F 検定をその手順の中に含むため、1元分類分散分析の F 検定が有意になった場合に、Scheffé の方法を適用する意味がある。しかし、いずれの処理の間に差があるかに関心がある場合には、分散分析を実施せず、適切な多重比較の方法を適用することが望ましい。表2では対比較に用いるパラメトリックな方法について紹介したが、その他にも検定統計量の分布を仮定しないノンパラメトリックな方法も提案されている。例えば、すべての対比較を行う Tukey の方法に対応する Steel-Dwass の方法(Steel,1960 ; Dwass,1960)や Dunn の方法(Dunn,1964)、対照群との対比較を行う Dunnnett の方法に対応する Steel の方法(Steel,1959)があげられる。

③ 複数の処理群に関し対比較を行う場合について、先に紹介した方法のうち Bonferroni の方法、Tukey の方法および Dunnnett の方法を取りあげ、第3回までと同様に事例を通じて例示する。SAS[®] University Edition、R、エクセル統計、JMP[®] の4つの統計計算ソフトウェアを使用し、データ“vitcap2”を解析する*1(図3)。なおそれぞれのソフトウェアについて SAS[®] Studio 3.4、R version 3.2.0、エクセル統計2015および JMP[®] Pro 10.0.2にて動作確認を行っている。いま、「10年以上曝露されている群(group=1)」、「10年未満曝露されている群(group=2)」および「曝露されていない群(group=3)」の間で、肺活量の母平均に差があるか否かに関心があるとする。ただしここでは、肺活量を表す変数を「vital」としている(元データでは「vital.capacity」)。この場合、研究の目的により、㊦すべての群間(対)において差があるか否かを検討したい場合、㊧対照群(ここでは、曝露されていない群)とその他の群との間で差があるか否かを検討したい場合の2つが考えられる。いま、「10年以上曝露されている群」、

*1 Rの ISwRパッケージに組み込まれている、カドミウム産業における雇用者についての曝露状態と年齢および肺活量のデータである。

	group	age	vital
1	1	39	4.62
2	1	40	5.29
3	1	41	5.52
4	1	41	3.71
5	1	45	4.02
6	1	49	5.09
7	1	52	2.7
8	1	47	4.31
9	1	61	2.7
10	1	65	3.03
11	1	58	2.73
12	1	59	3.67
13	2	29	5.21
14	2	29	5.17
15	2	33	4.88
16	2	32	4.5
17	2	31	4.47
18	2	29	5.12
19	2	29	4.51
20	2	30	4.85
21	2	21	5.22
22	2	28	4.62
23	2	23	5.07
24	2	35	3.64
25	2	38	3.64
26	2	38	5.09
27	2	43	4.61
28	2	39	4.73
29	2	38	4.58
30	2	42	5.12
31	2	43	3.89
32	2	43	4.62
33	2	37	4.3
34	2	50	2.7
35	2	50	3.5
36	2	45	5.06

	group	age	vital
37	2	48	4.06
38	2	51	4.51
39	2	46	4.66
40	2	58	2.88
41	3	27	5.29
42	3	25	3.67
43	3	24	5.82
44	3	32	4.77
45	3	23	5.71
46	3	25	4.47
47	3	32	4.55
48	3	18	4.61
49	3	19	5.86
50	3	26	5.2
51	3	33	4.44
52	3	27	5.52
53	3	33	4.97
54	3	25	4.99
55	3	42	4.89
56	3	35	4.09
57	3	35	4.24
58	3	41	3.88
59	3	38	4.85
60	3	41	4.79
61	3	36	4.36
62	3	36	4.02
63	3	41	3.77
64	3	41	4.22
65	3	37	4.94
66	3	42	4.04
67	3	39	4.51
68	3	41	4.06
69	3	43	4.02
70	3	41	4.99
71	3	48	3.86
72	3	47	4.68

	group	age	vital
73	3	53	4.74
74	3	49	3.76
75	3	54	3.98
76	3	48	5
77	3	49	3.31
78	3	47	3.11
79	3	52	4.76
80	3	58	3.95
81	3	62	4.6
82	3	65	4.83
83	3	62	3.18
84	3	59	3.03

図3 テストデータ "vitcap2"

「10年未満曝露されている群」、「曝露されていない群」の肺活量の母平均をそれぞれ μ_1 、 μ_2 および μ_3 とすれば、㉞の場面で検討したい帰無仮説は $H_{01} : \mu_1 - \mu_2 = 0$ 、 $H_{02} : \mu_2 - \mu_3 = 0$ 、 $H_{03} : \mu_3 - \mu_1 = 0$ と記述され、㉞の場面についてはそのうちの H_{02} と H_{03} である。なお対立仮説は $H_{1i} : \mu_i - \mu_j \neq 0, i, j = 1, 2, 3 ; i \neq j$ (両側対立仮説) とする。以降で

は、㉗の場面で Bonferroni の方法と Tukey の方法、㉘の場面で Bonferroni の方法と Dunnett の方法を適用する場合の、それぞれのソフトウェアにおける検定の実行と出力される結果について個々に述べる。

① SAS[®] University Edition

ここでは、一般線形モデルを適用する GLM プロシジャを用いて、Bonferroni の方法、Tukey の方法、Dunnett の方法を実行する手順を紹介する。まず㉗の場面で Bonferroni の方法を実行する場合には、群を示す変数(群変数)“group”を CLASS ステートメントで指定し、MODEL ステートメントで「vital=group」と指定する。さらに、LSMEANS ステートメントで群変数を指定し、オプションとして PDIFF を指定したうえで、ADJUST オプションを例えば「ADJUST=bonferroni」と指定すれば、すべての対についての差の推定値や Bonferroni の方法に基づく調整 p 値が算出される。なお、CL オプションを指定すれば 95% 信頼区間があわせて算出される。Tukey の方法を実行するためには、CLASS ステートメント、MODEL ステートメント、LSMEANS ステートメントを Bonferroni の方法の場合と同様に指定し、ADJUST オプションのみ「ADJUST=tukey」と変更する。次に、㉘の場面についても㉗と同様に CLASS ステートメント、MODEL ステートメント、LSMEANS ステートメントを指定するが、PDIFF オプションを「PDIFF=control(“3”)」と指定することにより、“group”が“3”である群を対照群とした比較に基づく p 値が出力される。Bonferroni の方法を実行する場合には「ADJUST=bonferroni」と指定し、Dunnett の方法を実行する場合には「ADJUST=dunnett」と指定する。SAS[®] では規定値で両側検定の結果が出力されるが、Dunnett の方法に関しては、PDIFF オプションを「PDIFF=controll(“3”)」または「PDIFF=controlu(“3”)」と指定することにより、片側検定の結果を表示できる。

図4、図5より、㉗の場面における Bonferroni の方法に基づく調整 p 値は $p_{B1}=0.1355$ 、 $p_{B2}=1.0000$ 、 $p_{B3}=0.1124$ であり、Tukey の方法に基づく p 値は $p_{T1}=0.1105$ 、 $p_{T2}=0.9984$ 、 $p_{T3}=0.0930$ と算出される。また図6より、㉘の場面における Dunnett の方法に基づく p 値は $p_{D2}=0.9981$ 、 $p_{D3}=0.0718$ である。Bonferroni の方法に基づく結果はここでは示さないが、p 値は $p_{B2}=1.0000$ 、 $p_{B3}=0.0749$ と算出される。ここに、 p_{Bi} 、 p_{Ti} 、 p_{Di} は前記の帰無仮説 H_{0i} に対応する調整 p 値である。これより、Tukey の方法、Dunnett の方法のいずれについても、調整 p 値は Bonferroni の方法に基づく p 値に比較して小さいことが確認できる。また第2回で示したように、「10年以上曝露されている群(group=1)」と「曝露されていない群(group=3)」のみを比較すれば、有意水準5%のもとで帰無仮説が棄却され、「肺活量の母平均について処理群間で差がある」と結論づけるのに対し、すべての群間について、もしくは対照群とその他の群について比較した場合には、いずれの帰無仮説も棄却することができない。

なお、今回のように個々のデータから解析するのではなく、多重比較に関する調整を施していない p 値がすでに得られている場合には、MULTTEST プロシジャを用いて Bonferroni の方法に基づく調整 p 値を算出することができる。

GLM プロシジャ
最小 2 乗平均
多重比較の調整: Bonferroni

group	vital の最小 2 乗平均	最小 2 乗平均の番号
1	3.94916667	1
2	4.47178571	2
3	4.46204545	3

group	vital の最小 2 乗平均	95% 信頼限界	
1	3.949167	3.521561	4.376773
2	4.471786	4.191852	4.751720
3	4.462045	4.238736	4.685355

効果 group に対する最小 2 乗平均 H0: LSMean(i)=LSMean(j) の Pr > t			
従属変数 : vital			
i\j	1	2	3
1		0.1355	0.1124
2	0.1355		1.0000
3	0.1124	1.0000	

効果 group に対する最小 2 乗平均				
i	j	平均の差	LSMean(i)-LSMean(j) の 同時 95% 信頼限界	
1	2	-0.522619	-1.150588	0.105350
1	3	-0.512879	-1.105606	0.079849
2	3	0.009740	-0.430246	0.449726

図4 GLM プロシジャによる Bonferroni の方法の実行結果(SAS® University Edition)

GLM プロシジャ
最小 2 乗平均
多重比較の調整: Tukey-Kramer

group	vital の最小 2 乗平均	最小 2 乗平均の番号
1	3.94916667	1
2	4.47178571	2
3	4.46204545	3

group	vital の最小 2 乗平均	95% 信頼限界	
1	3.949167	3.521561	4.376773
2	4.471786	4.191852	4.751720
3	4.462045	4.238736	4.685355

効果 group に対する最小 2 乗平均 H0: LSMean(i)=LSMean(j) の Pr > t			
従属変数 : vital			
i\j	1	2	3
1		0.1105	0.0930
2	0.1105		0.9984
3	0.0930	0.9984	

効果 group に対する最小 2 乗平均				
i	j	平均の差	LSMean(i)-LSMean(j) の 同時 95% 信頼限界	
1	2	-0.522619	-1.135904	0.090666
1	3	-0.512879	-1.091746	0.065988
2	3	0.009740	-0.419957	0.439438

図5 GLM プロシジャによる Tukey の方法の実行結果(SAS® University Edition)

GLM プロシジャ 最小 2 乗平均 多重比較の調整: Dunnett		
group	vital の最小 2 乗平均	H0:LSMean=Control
		Pr > t
1	3.94916667	0.0718
2	4.47178571	0.9981
3	4.46204545	

図6 GLM プロシジャによる Dunnett の方法の実行結果 (SAS® University Edition)

②R

ここでは、1元分類分散分析を適用する aov関数を実行した結果に対し、multcomp パッケージに含まれる glht関数を用いて多重比較を適用する手順を紹介する。まず、“vitcap2”では群変数“group”が数字で与えられており、通常の手順でデータを読み込むだけでは“group”が連続変数として認識されるため、factor関数を用いてカテゴリー変数として認識させる必要がある。この際にオプションとして「levels=c(“3”,“2”,“1”)」と指定し、群変数のカテゴリーの順序を降順にしておく (Dunnett の方法を適用する際に“group”が“3”である群を対照群とするため)。ここでは、元の“group”と区別するため、カテゴリー変数として認識させた群変数を“group_c”とする。Bonferroni、Tukey、Dunnett のいずれの方法を適用する場合にも、まず aov関数を用い、応答として“vital”、群変数として“group_c”、およびデータ名として読み込んだ際に付けたデータ名を指定し、結果がオブジェクトとして保存されるよう名前を付け(ここでは、“vitcap.aov”とする)実行する。次に glht関数を用い、モデルとして“vitcap.aov”を指定し、㉞の場面であればオプションで比較の形式を「linfct=mcp(group_c=“Tukey”)」と指定する。これにより、“group_c”のすべての対について比較することを指定する。なお、ここでも結果がオブジェクトとして保存されるよう名前を付け(ここでは、“vitcap.mcl”とする)実行する。最後に summary関数を用いて「summary(vitcap.mcl)」と実行すれば、㉞の場面での(すべての対についての)差の推定値や Tukey の方法に基づく調整p値などが出力される。この時、オプションとして「test=adjusted(type=“bonferroni”)」と指定すれば、㉞の場面での Bonferroni の方法に基づく結果が出力される。㉞の場面については、glht関数を適用する際にオプションで「linfct=mcp(group_c=“Dunnett”)」と指定すれば、対照群との対比較のみが実行される。その他は㉞の場合と同様に、summary関数を用いれば規定値で差の推定値や Dunnett の方法に基づく調整p値が、オプションで“bonferroni”を指定すれば Bonferroni の方法に基づく結果が出力される。R では規定値で両側検定の結果が出力され、片側検定の結果を得たい場合には、glht関数を適用する際に例えば「alternative=“greater”)」と指定する。なお信頼区間を出力したい場合には、例えば“vitcap.mcl”に対し confint関数を適用すればよい。

図7、図8、図9より、㊦㊧の場面における Bonferroni の方法と Dunnett の方法に基づく調整p値は SAS® による実行結果に一致するが、Tukey の方法に基づく p値は SAS® による結果と若干異なる。これは glht関数内での計算方法により、出力されるp値の末尾が実行の度に異なることによるが、導かれる結論には大きく影響しない。なお Tukey の方法は他にも、“vitcap.aov”に対し TukeyHSD関数を用いても適用できる。

```

Simultaneous Tests for General Linear Hypotheses

Multiple Comparisons of Means: Tukey Contrasts

Fit: aov(formula = vital ~ group_c, data = vitcap2)

Linear Hypotheses:
      Estimate Std. Error t value Pr(>|t|)
2 - 3 == 0    0.00974    0.17997   0.054   0.9984
1 - 3 == 0   -0.51288    0.24245  -2.115   0.0909 .
1 - 2 == 0   -0.52262    0.25687  -2.035   0.1079
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
(Adjusted p values reported -- single-step method)

```

図7 glht関数による Tukey の方法の実行結果(R)

```

Simultaneous Tests for General Linear Hypotheses

Multiple Comparisons of Means: Tukey Contrasts

Fit: aov(formula = vital ~ group_c, data = vitcap2)

Linear Hypotheses:
      Estimate Std. Error t value Pr(>|t|)
2 - 3 == 0    0.00974    0.17997   0.054   1.000
1 - 3 == 0   -0.51288    0.24245  -2.115   0.112
1 - 2 == 0   -0.52262    0.25687  -2.035   0.135
(Adjusted p values reported -- bonferroni method)

```

図8 glht関数による Bonferroni の方法の実行結果(R)

```

Simultaneous Tests for General Linear Hypotheses

Multiple Comparisons of Means: Dunnett Contrasts

Fit: aov(formula = vital ~ group_c, data = vitcap2)

Linear Hypotheses:
      Estimate Std. Error t value Pr(>|t|)
2 - 3 == 0    0.00974    0.17997   0.054   0.9981
1 - 3 == 0   -0.51288    0.24245  -2.115   0.0718 .
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
(Adjusted p values reported -- single-step method)

```

図9 glht関数による Dunnett の方法の実行結果(R)

③エクセル統計

「分散分析・多重比較」の中の「一元配置分散分析」または「多元配置分散分析」を用いる。いずれを用いるかはデータの形式によるが、ここでは他のソフトウェアで解析する場合と同じ形式(図3の形式)でデータが入力されているとし、「多元配置分散分析」を用いる。出てくるウィンドウの「変数」タブで「目的変数」に“vital”を、「固定因子」に“group”を指定し、「多重比較」タブで「多重比較を行う因子」に“group”を指定する。また比較の形式として「Bonferroni」、「Tukey-Kramer」、「Dunnett」を指定し、対照群としてここでは「最後の水準」を指定する。すると分散分析表や多重比較などの結果が出力される(図10)。エクセル統計では Bonferroni の方法と Tukey の方法については両側検定のみ、Dunnett の方法については両側検定と片側検定の両方について差の推定値や調整p値が出力される。なお、図10は、出力のうち多重比較の結果のみを抜粋したものである。㊦の場面での Bonferroni の方法に基づく結果、Tukey の方法に基づく結果と、㊦の場合の Dunnett の方法に基づく結果は、SAS[®] による実行結果に一致するが、㊦の場面での Bonferroni の方法に基づく調整p値は出力できない。

	A	B	C	D	E	F	G	H	I	J	K	L	M	N
48	多重比較検定													
49	因子	目的変数	手法	水準1	水準2	平均1	平均2	差	標準誤差	統計量	P 値	*: P<0.05	** : P<0.01	
50	group	vital	Bonferroni	1	2	3.9492	4.4718	0.5226	0.2569	2.0346	0.1355			
51				1	3	3.9492	4.4620	0.5129	0.2425	2.1154	0.1124			
52				2	3	4.4718	4.4620	0.0097	0.1800	0.0541	1.0000			
53			Tukey-Kramer	1	2	3.9492	4.4718	0.5226	0.2569	2.0346	0.1105			
54				1	3	3.9492	4.4620	0.5129	0.2425	2.1154	0.0930			
55				2	3	4.4718	4.4620	0.0097	0.1800	0.0541	0.9984			
56														
57	因子	目的変数	手法	対立仮説	水準1	水準2	平均1	平均2	差	標準誤差	統計量	P 値	*: P<0.05 ** : P<0.01	
58	group	vital	Dunnett	対照群≠	3	1	4.4620	3.9492	0.5129	0.2425	2.1154	0.0718		
59					3	2	4.4620	4.4718	0.0097	0.1800	0.0541	0.9981		
60				対照群<	3	1	4.4620	3.9492	0.5129	0.2425	2.1154	0.9985		
61					3	2	4.4620	4.4718	-0.0097	0.1800	-0.0541	0.6815		
62				対照群>	3	1	4.4620	3.9492	0.5129	0.2425	2.1154	0.0359 *		
63					3	2	4.4620	4.4718	-0.0097	0.1800	-0.0541	0.7246		
64														

図10 「分散分析・多重比較」による Bonferroni の方法、Tukey の方法および Dunnett の方法の実行結果(エクセル統計)

④JMP[®]

まず事前にホームウィンドウの「ファイル」から「環境設定」を開き、「プラットフォーム」中で「一元配置」の「Student の t 検定を使ったペアごとの比較」、「すべてのペア、Tukey の HSD 検定」、「コントロール群との比較(Dunnett)」にチェックを入れ指定しておく。「分析」の「二変量の関係」を用い、X と Y に群変数(名義変数として指定しておく)と応答変数を指定する。次に出てくる「コントロール群を選択」のウィンドウで“3”を選択すれば、図11のように結果が出力される。ここで、コントロール群を選択するウィンドウが表示されない場合は、事前にホームウィンドウの「テーブル」から「並べ替え」を選択し、“group”が降順となるようデータを並び替える(“group”をダブルクリックで指定し、「▽」にて降順を指定する)。結果のウィンドウにはすべての対についての差の推定値や調整が施されていない p 値(「Student の t 検定を使ったペアごとの比較」、Dunnett の方法に基づく調整p値、および Tukey の方法に基づく調整p値と95%信頼区間が表示される(図11)。JMP[®]では、規定値で両側検定の結果が出力される。なお、図11は、出力のうち例示に必要な部分のみを抜粋したものである。また Bonferroni の方

図11 「2変量の関係」による Tukey の方法および Dunnett の方法の実行結果 (JMP®)

法に基づく調整p値は出力できないため、「Studentのt検定を使ったペアごとの比較」に出力されたp値に検定の数(ここでは、㊦の場面は3、㊩の場面は2)を乗じ、調整p値を算出する。

文献

- 1) Bretz, F., Hothorn, T. & Westfall, P. (2010). Multiple Comparisons Using R. London: Chapman & Hall.
- 2) Dmitrienko, A., Tamhane, A.C. & Bretz, F. (2009). Multiple Testing Problems in Pharmaceutical Statistics. Boca Raton: Chapman & Hall.
- 3) Dunn, O.J. (1964). Multiple comparisons using rank sums. *Technometrics* 6, 241-252.
- 4) Dunnett, C.W. (1955). A multiple comparison procedure for comparing several treatments with a control. *Journal of the American Statistical Association* 50, 1096-1121.
- 5) Dwass, M. (1960). Some k-sample rank-order tests. *Contributions to Probability and Statistics* (Eds. I. Olkin et al.). Stanford: Stanford University Press, 198-202.
- 6) Kramer, C.Y. (1956). Extension of multiple range tests to group means with unequal numbers of replications. *Biometrics* 12, 307-310.
- 7) Ott, R.L., Longnecker, M. (1988). *An Introduction to Statistical Methods and Data Analysis*. 6th Edition. Boston: PWS-Kent.
- 8) Sheffé, H. (1953). A method for judging all contrasts in the analysis of variance. *Biometrika* 40, 87-104.
- 9) Steel, R.G.D. (1959). A multiple comparison rank sum test: Treatments versus control. *Biometrics* 15, 560-572.
- 10) Steel, R.G.D. (1960). A rank sum test for comparing all pairs of treatments. *Technometrics* 2, 197-207.
- 11) Tukey, J.W. (1953). The problem of multiple comparisons. Unpublished manuscript. The Collected Works of John W. Tukey, Volume 8, 1994. H. I. Braun (Ed.). New York: Chapman & Hall.
- 12) 朝倉こう子・濱崎俊光(2015a). 2つの平均の比較. *Drug Delivery System* 30(2), 149-158.
- 13) 朝倉こう子・濱崎俊光(2015b). 2つの割合の比較. *Drug Delivery System* 30(3), 235-245.
- 14) 永田靖・吉田道弘(1997). 統計的多重比較法の基礎. サイエンス社.
- 15) 厚生省 (1999). 反復投与毒性試験に係るガイドラインの一部改正について (平成11年4月5日医薬審第655号).

“DDS”用語解説 No.296

エバンスブルー-アルブミン複合体 (Evans Blue-Albumin Complex)

エバンスブルー-アルブミン複合体とは、ジアゾ色素であるエバンスブルー(分子量960.8)が血液内に存在するアルブミンとの相互作用により形成される複合体のことである。投与されたエバンスブルーは速やかにアルブミンと結合して分子量約70kDのアルブミンとして振舞う。1943年に Rawson らにより、エバンスブルーがアルブミンに選択的に結合することが明らかにされ、色素の吸光度測定($\lambda_{\max}=610\text{nm}$ 付近)によりアルブミンの動態が測定可能となった。これにより血液内のアルブミンの消失・生成計測、毛細血管の血管透過性評価に使用されてきている。0.3%濃度に希釈されたエバンスブルー水溶液(0.1mL)は、マウス血液内に投与されると血中アルブミンとほぼ同モル量となる。エバンスブルー-アルブミン複合体を DDS分野に用いる意義は、高分子(アルブミン)を漏出しない血管と、血管透過性が亢進され、高分子を漏出する血管がエバンスブルーにより明確に区別できることである。

白石貢一 (Koichi Shiraishi)

東京慈恵会医科大学 総合医科学研究センター医用エンジニアリング研究部