経時データの多重比較法

岸本　淳司
(ＳＡＳ／慶應義塾大学／東京大学)

Multiple Comparisons on Longitudinal Data

Junji Kishimoto
SAS Institute Japan / Keio Univ. SFC / Univ. of Tokyo
e-mail address: jpnjak@jpn.sas.com

反応を経時的に測定して得られたデータについて t 検定の反復あるいは Dunnett の多重比較を行うことには問題があるとされる．なぜそれが悪いのか，ではどうしたらよいのかについて SAS の具体的プログラムと共に解説する．

Multiple Comparisons, Repeated Measures, Quadrature

1　はじめに

1 群の各個体を時間を追って測定したデータ

表１ 1 群の経時的測定データ

時点	t_0	t_1	t_2	$\cdots$	t_k
個体$_1$	y_{10}	y_{11}	y_{12}	$\cdots$	y_{1k}
個体$_2$	y_{20}	y_{21}	y_{22}	$\cdots$	y_{2k}
$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$
個体$_n$	y_{n0}	y_{n1}	y_{n2}	$\cdots$	y_{nk}
平均	μ_0	μ_1	μ_2	$\cdots$	μ_k

について，各時点の平均を比較したいことがある．たとえば時点 0 と各時点との比較

$$H_{01}:\mu_0=\mu_1,\quad H_{02}:\mu_0=\mu_2,\ldots,H_{0p}:\mu_0=\mu_k \tag{1}$$

について，対応のある t 検定を反復して論文を提出すると「検定の多重性を考慮せよ」というコメントがつくことがある．しかし，通常の Dunnett の検定を実施してもよいものであろうか．

また，2 群の個体を経時的に測定したデータ

表 2 2 群の経時的測定データ

	時点	t_1	t_2	$\cdots$	t_k
群 1	個体$_1$	y_{11}	y_{12}	$\cdots$	y_{1k}
	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$
	個体$_m$	y_{m1}	y_{m2}	$\cdots$	y_{mk}
	平均	μ_{11}	μ_{12}	$\cdots$	μ_{1k}
群 2	個体$_{m+1}$	y_{m+11}	y_{m+12}	$\cdots$	y_{m+1k}
	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$
	個体$_n$	y_{n1}	y_{n2}	$\cdots$	y_{nk}
	平均	μ_{21}	μ_{22}	$\cdots$	μ_{2k}

について，各時点毎に対応のない t 検定

$$H_{01}:\mu_{11}=\mu_{21},\quad H_{02}:\mu_{12}=\mu_{22},\ldots,H_{0k}:\mu_{1k}=\mu_{2k} \tag{2}$$

を反復して実施すると，ここにも検定の多重性の問題が現れる．“臨床試験の統計解析に関するガイドライン” では「各時点ごとに 2 群間での対応のない検定を繰りかえしがちであるが，それでは第 1 種の過誤の確率が増大する．このようなデータの場合には時点毎の比較ではなく，トレンドやプロフィルを比較するための特別な手法が必要である.」と述べている．事前にトレンドモデルが想定できる場合には，群間の差を統計モデルのパラメタで表現することは解析法として本質的ではあるが，事前にはモデルを想定できないこともあるだろう．

本稿では，時点間の相関を考慮しつつ群平均の多重比較を行う方法について考察する．

2 反復測定時点間の多重比較

2.1 独立誤差の場合

表 1, 式 1 で示したような反復測定での時点間の平均を比較することを考えよう．次のデータを例題とする．

```
data repeated;
  input id @;
  do time=0 to 3;
     input y @;
     output;
  end;
  cards;
301  2.07 1.58 1.84 2.06
305  1.86 2.09 1.76 2.43
306  2.84 2.13 1.62 1.62
307  3.30 2.03 2.88 2.60
308  1.73 1.67 1.15 1.21
309  3.40 2.16 3.20 2.24
310  2.79 1.82 2.38 1.29
311  2.73 3.38 1.93 2.61
;
```

通常の一元配置モデルでの Dunnett の多重比較を PROC GLM で実行する.

```
proc glm data=repeated;
  class time;
  model y = time;
  lsmeans time / adjust=dunnett pdiff=control('0');
run;
```

出力 1.1 **一元配置の** Dunnett **比較**

```
General Linear Models Procedure
Least Squares Means
Adjustment for multiple comparisons: Dunnett

TIME            Y      Pr > |T| H0:
           LSMEAN   LSMEAN=CONTROL
0      2.59000000
1      2.10750000        0.2868
2      2.09500000        0.2683
3      2.00750000        0.1624
```

反復測定データを一元配置の方法で解析することは，時点間の相関をゼロと仮定することと同じである．PROC MIXED による反復測定解析で TYPE=VC の指定を行うと同様の解析が再現できる．

```
proc mixed data=repeated;
  class time;
  model y = time / ddfm=residual;
  repeated time / type=VC subject=id rcorr;
  lsmeans time / adjust=dunnett pdiff=control('0');
run;
```

出力 1.2 **独立誤差の相関構造**

```
R Correlation Matrix for Subject 1

Row          COL1          COL2          COL3          COL4
  1    1.00000000
  2                  1.00000000
  3                                1.00000000
  4                                              1.00000000
```

出力 1.3 **独立誤差を仮定した** Dunnett **比較**

```
Differences of Least Squares Means

Effect  TIME  _TIME  Difference   Std Error  DF      t  Pr>|t|  Adjustment  Adj P
TIME    1     0     -0.48250000  0.30616143  28  -1.58  0.1263  Dunnett    0.2868
TIME    2     0     -0.49500000  0.30616143  28  -1.62  0.1171  Dunnett    0.2683
TIME    3     0     -0.58250000  0.30616143  28  -1.90  0.0674  Dunnett    0.1624
```

出力 1.2 中，ブランクの相関係数はゼロであることを表している．すべての時点対間の相関係数をゼロとした反復測定の多重比較は，通常の一元配置多重比較と全く同じ結果をもたらすことがわかる．

2.2　球面誤差の場合

誤差分散が等しく時点間の共分散もすべて等しいとき，その共分散行列は複合対称であるという．このとき，通常の F 検定は妥当である．さらに一般的には，Huyhn and Feldt(1970) が示した球面性を満たしているときも F 検定は妥当である．誤差が球面であるとき，Tukey 法や Dunnett 法のような通常の多重比較法も妥当である．

PROC GLM でこのような多重比較を実行するには，個体差をモデルに入れたプログラムを指定する．

```
proc glm data=repeated;
  class id time;
  model y = id time;
  lsmeans time / adjust=dunnett pdiff=control('0');
run;
```

出力 2.1 個人差をモデルに入れた Dunnett 比較

```
                 General Linear Models Procedure
                        Least Squares Means
              Adjustment for multiple comparisons: Dunnett

                TIME              Y     Pr > |T| H0:
                             LSMEAN   LSMEAN=CONTROL
                0       2.59000000
                1       2.10750000        0.1189
                2       2.09500000        0.1073
                3       2.00750000        0.0502
```

PROC MIXED を使って，球面形の誤差構造を指定しても同じ結果が得られる．

```
proc mixed data=repeated;
  class time;
  model y = time;
  repeated time / type=HF subject=id rcorr;
  lsmeans time / adjust=dunnett pdiff=control('0');
run;
```

出力 2.2 球面性を仮定した相関構造

```
                  R Correlation Matrix for Subject 1

       Row         COL1          COL2          COL3          COL4
         1   1.00000000    0.44604378    0.49269359    0.51790125
         2   0.44604378    1.00000000    0.34194834    0.39242273
         3   0.49269359    0.34194834    1.00000000    0.45236522
         4   0.51790125    0.39242273    0.45236522    1.00000000
```

出力 2.3 球面性を仮定した Dunnett 比較

```
                      Differences of Least Squares Means

 Effect   TIME _TIME  Difference   Std Error  DF     t  Pr>|t|  Adjustment    Adj P
 TIME     1    0     -0.48250000  0.23037494  21 -2.09  0.0485  Dunnett-Hsu  0.1189
 TIME     2    0     -0.49500000  0.23037494  21 -2.15  0.0435  Dunnett-Hsu  0.1073
 TIME     3    0     -0.58250000  0.23037494  21 -2.53  0.0195  Dunnett-Hsu  0.0502
```

個人差をモデルに入れた検定と誤差に球面性の仮定をおいた検定とは全く同じ結果をもたらす．すなわち，誤差に球面性が仮定できるなら個人差をモデルに入れた多重比較は妥当である．一元配置型 Dunnett 検定で TIME=0 と TIME=3 の比較の p 値が 0.1624 であったのに対し，球面誤差の場合は 0.0502 になっており，差をより鋭敏に検出できることがわかる．反復測定型のデータに一元配置型の多重比較を施すのが不適切なのは，(相関がほとんどゼロの場合を除き) 検出力で損をするからである．

2.3 非球面誤差の場合

球面型の誤差構造が不適切な場合には，適当な誤差構造を仮定して最尤推定するか，あるいは標本からそのまま推定された相関構造を元に検定を構成することになる．ところが，特定の相関構造以外の積分計算は今日のコンピュータをもってしても難しい．そこで，いくつかの近似法がとられる．

1. 各相関係数の平均を計算し，共通の相関係数とする．
2. Hsu(1992) の因子分析的方法を用いて近似的な積分を行う．
3. 標本相関係数を再現するような多変量 t 分布を乱数で再現しその最大値に対する調整 p 値を求める．

Release 6.10 以降の PROC GLM では，LSMEANS ステートメント中 ADJUST=DUNNETT オプションを指定すると，必要な時は自動的に Hsu の近似が行なわれる．たとえば，AR(1) の誤差構造を想定して Hsu 型の近似計算を行うためには次のようにする．

```
proc mixed data=repeated;
  class time;
  model y = time;
  repeated time / type=AR(1) subject=id rcorr;
  lsmeans time / adjust=dunnett pdiff=control('0');
run;
```

図 3.1 AR(1) を仮定した誤差構造

```
                      R Correlation Matrix for Subject 1

               Row          COL1          COL2          COL3          COL4
                 1    1.00000000    0.27679862    0.07661748    0.02120761
                 2    0.27679862    1.00000000    0.27679862    0.07661748
                 3    0.07661748    0.27679862    1.00000000    0.27679862
                 4    0.02120761    0.07661748    0.27679862    1.00000000
```

図 3.2 AR(1) を仮定し Hsu 型の近似を行った Dunnett 比較

```
                         Differences of Least Squares Means

 Effect  TIME _TIME  Difference   Std Error  DF     t  Pr>|t|  Adjustment   Adj P
 TIME    1    0     -0.48250000  0.26019292  21 -1.85  0.0778  Dunnett-Hsu  0.1783
 TIME    2    0     -0.49500000  0.29400632  21 -1.68  0.1071  Dunnett-Hsu  0.2385
 TIME    3    0     -0.58250000  0.30269911  21 -1.92  0.0680  Dunnett-Hsu  0.1575
```

特定の相関構造を仮定しないで，シミュレーションにより調整 p 値を求めるには，次のようにする．

```
proc mixed data=repeated;
  class time;
  model y = time;
  repeated time / type=UN subject=id rcorr;
  lsmeans time / adjust=SIMULATE pdiff=control('0');
run;
```

図 4.1 無構成の誤差構造

```
                          R Correlation Matrix for Subject 1

              Row         COL1          COL2          COL3          COL4
                1    1.00000000    0.31304700    0.84507324    0.31850774
                2    0.31304700    1.00000000    0.09341769    0.56166162
                3    0.84507324    0.09341769    1.00000000    0.44463342
                4    0.31850774    0.56166162    0.44463342    1.00000000
```

図 4.2 無構成誤差構造，シミュレーションによる Dunnett 比較

```
                               Differences of Least Squares Means

 Effect  TIME _TIME  Difference   Std Error  DF     t  Pr>|t|  Adjustment   Adj P
 TIME    1    0     -0.48250000  0.24828519   7 -1.94  0.0931  Simulate    0.2058
 TIME    2    0     -0.49500000  0.13038405   7 -3.80  0.0067  Simulate    0.0162
 TIME    3    0     -0.58250000  0.24890295   7 -2.34  0.0518  Simulate    0.1173
```

通常の臨床試験や前臨床試験でのデータでは，ランダムな誤差が後の誤差に与える影響よりも個体差の影響の方がずっと強いと思われる．また，少ない標本から推定された相関係数を真の値として検定を構成するのは不安定である．ゆえに，TYPE=AR(1) とか TYPE=UN のような相関構造が適切な例は少なく，球面性 (あるいは複合対称) を仮定した解析が良いことが多いだろうというのが著者の見解である．

3　いわゆる輪切りの検定

3.1　定式化と方法

表 2 と式 2 に表したように，2 つの処置群について k 時点測定した試験について考える．岸本 (1997) が示したように，時点間の相関構造に複合対称を仮定すれば，時点毎に行う対応のない t 検定について多重性を調整した p 値を得ることができる．その概略は次の通りである．

観測値 Y_{ihr} について，各セルの平均をパラメタとしたモデルを想定する．

$$Y_{\mathrm{ihr}} = \mu_{\mathrm{ih}} + \varepsilon_{\mathrm{ihr}}, \qquad i = 1, 2, \quad h = 1, ..., k, \quad r = 1, ..., n_{\mathrm{ih}}, \tag{3}$$

同一群の観測値 Y_{ihr} は等分散で時点間で等相関と仮定し，さらに等サンプルサイズとする．パラメタ推定値 $\hat{\mu} = (\hat{\mu}_{11}, \ldots, \hat{\mu}_{1\mathrm{k}}, \hat{\mu}_{21}, \ldots, \hat{\mu}_{2\mathrm{k}})$ 間の分散共分散は，k=3 のとき次のように表される．

$$\mathrm{Var}(\hat{\mu}) = \sigma^2 \begin{pmatrix} 1 & \rho & \rho & 0 & 0 & 0 \\ \rho & 1 & \rho & 0 & 0 & 0 \\ \rho & \rho & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & \rho & \rho \\ 0 & 0 & 0 & \rho & 1 & \rho \\ 0 & 0 & 0 & \rho & \rho & 1 \end{pmatrix} \tag{4}$$

仮説のコントラスト行列を次のように設定する。

$$\mathrm{L} = \begin{pmatrix} 1 & 0 & 0 & -1 & 0 & 0 \\ 0 & 1 & 0 & 0 & -1 & 0 \\ 0 & 0 & 1 & 0 & 0 & -1 \end{pmatrix} \tag{5}$$

コントラスト間の分散共分散は次のようになる。

$$\mathrm{Var}(\mathrm{L}\hat{\mu}) = 2\sigma^2 \begin{pmatrix} 1 & \rho & \rho \\ \rho & 1 & \rho \\ \rho & \rho & 1 \end{pmatrix} \tag{6}$$

よく知られているように，このような等相関の多変量 t 分布の積分計算は実行可能であり，これにより各コントラストの最大値について確率評価することができる．SAS では，PROBMC 関数で Dunnett 型の指定をすることにより求められる．

表 3 5%臨界値の表

k=2	ρ					
ν	0	0.3	0.5	0.7	0.9	1
4	3.38	3.36	3.31	3.23	3.06	2.78
8	2.72	2.70	2.67	2.62	2.51	2.31
12	2.54	2.53	2.50	2.46	2.36	2.18
16	2.46	2.45	2.42	2.38	2.29	2.12
∞	2.24	2.23	2.21	2.18	2.11	1.96

k=3						
ν	0	0.3	0.5	0.7	0.9	1
6	3.19	3.16	3.10	2.99	2.80	2.45
12	2.75	2.73	2.68	2.61	2.46	2.18
18	2.62	2.60	2.56	2.50	2.36	2.10
24	2.56	2.54	2.51	2.44	2.31	2.06
∞	2.39	2.38	2.35	2.30	2.19	1.96
k=4						
8	3.13	3.09	3.02	2.91	2.69	2.31
16	2.78	2.76	2.71	2.62	2.44	2.12
24	2.68	2.66	2.61	2.53	2.37	2.06
32	2.63	2.61	2.57	2.49	2.34	2.04
∞	2.49	2.47	2.44	2.38	2.24	1.96

ρ = 0 のときは Studentized Maximum Modulus に、ρ = 1 のときは多重性を考慮しない場合に対応する。相関の大きさによる臨界値の変化は小さく，相関が相当大きくないと相関を考慮しない場合に比べて有利にならないことがわかる。実際の解析ではデータから推定した相関で ρ を置き換えるが，その際の推定誤差は問題にならない．また，多少の不等相関があっても影響ないことが示唆される．

3.2　輪切り検定の例題

多重性を考慮した輪切り検定の SAS による実施例を紹介する．データは遅効性の薬剤に関するもので，投薬後 5 時間～8 時間のあたりで実薬群と溶媒群とに効果の差があることが期待されたが，いつ効果があるかは事前には明らかになっていなかった．測定した 4 時点のうちいずれかで有意な差があれば効果ありと判定したいが，相関のある 4 時点で測定していることの多重性を考慮しなければならない．

```
data B;
  input group $ id $ @;
  do time=5 to 8;
     input y @;
     output;
  end;
  cards;
Vehicle 001  3.25 3.55 3.44 1.78
Vehicle 003  2.59 2.80 4.18 2.63
Vehicle 004  3.09 3.29 2.05 1.28
Vehicle 007  2.66 2.86 2.27 1.82
Vehicle 008  2.84 2.43 2.36 2.30
Vehicle 009  2.01 2.86 1.84 2.50
Vehicle 010  3.22 3.36 2.80 2.26
Vehicle 011  3.16 3.85 3.42 1.93
D+P(H)  301  2.07 1.58 1.84 2.06
D+P(H)  305  1.86 2.09 1.76 2.43
D+P(H)  306  2.84 2.13 1.62 1.62
D+P(H)  307  3.30 2.03 2.88 2.60
D+P(H)  308  1.73 1.67 1.15 1.21
```

```
D+P(H)  309  3.40 2.16 3.20 2.24
D+P(H)  310  2.79 1.82 2.38 1.29
D+P(H)  311  2.73 3.38 1.93 2.61
;
proc mixed data=B;
  class group time id;
  model y = group(time) / ddfm=residual;
  repeated time / type=cs subject=id RCORR;
  lsmeans group(time) / corr slice=time;
run;
```

図 5.1 **各セルの最小二乗平均**

```
                              Least Squares Means

   Level                           LSMEAN      Std Error   DDF        T  Pr > |T|
   GROUP(TIME) D+P(H) 5       2.59000000    0.20761292    56    12.48    0.0001
   GROUP(TIME) Vehicle 5      2.85250000    0.20761292    56    13.74    0.0001
   GROUP(TIME) D+P(H) 6       2.10750000    0.20761292    56    10.15    0.0001
   GROUP(TIME) Vehicle 6      3.12500000    0.20761292    56    15.05    0.0001
   GROUP(TIME) D+P(H) 7       2.09500000    0.20761292    56    10.09    0.0001
   GROUP(TIME) Vehicle 7      2.79500000    0.20761292    56    13.46    0.0001
   GROUP(TIME) D+P(H) 8       2.00750000    0.20761292    56     9.67    0.0001
   GROUP(TIME) Vehicle 8      2.06250000    0.20761292    56     9.93    0.0001
```

図 5.2 **各セル間の相関係数**

```
             CORR1  CORR2  CORR3  CORR4  CORR5  CORR6  CORR7  CORR8
              1.00  -0.00   0.30   0.00   0.30   0.00   0.30   0.00
             -0.00   1.00  -0.00   0.30   0.00   0.30  -0.00   0.30
              0.30  -0.00   1.00   0.00   0.30   0.00   0.30   0.00
              0.00   0.30   0.00   1.00   0.00   0.30   0.00   0.30
              0.30   0.00   0.30   0.00   1.00   0.00   0.30   0.00
              0.00   0.30   0.00   0.30   0.00   1.00   0.00   0.30
              0.30  -0.00   0.30   0.00   0.30   0.00   1.00   0.00
              0.00   0.30   0.00   0.30   0.00   0.30   0.00   1.00
```

図 5.3 **各時点での多重性を調整しない検定**

```
                          Tests of Effect Slices

                Effect        Slice      NDF    DDF        F   Pr > F
                GROUP(TIME)   TIME 5       1     56     0.80   0.3751
                GROUP(TIME)   TIME 6       1     56    12.01   0.0010
                GROUP(TIME)   TIME 7       1     56     5.68   0.0205
                GROUP(TIME)   TIME 8       1     56     0.04   0.8521
```

TIME=6 のとき F=12.01, p=0.0010 で最も効果が大きい.

得られた相関と F 値 (t 値に変換) を基に，多重性を調整した p 値を求める．

```
data mult_t;
  k = 4;
  t = sqrt(12.01);
  rho = 0.30;
  nu = 56;
  array lambda{4};
  do i=1 to k;
     lambda{i} = sqrt(rho);
  end;
  p = 1 - probmc("DUNNETT2", t, ., nu, k, of lambda1-lambda4);
  put p= 6.4;
run;
```

P=0.0040

TIME=6 での生の p 値が 0.0010 であったのに対し，多重性を調整した p 値は 0.0040 になった．今の例では相関が 0.30 と小さかったことにより，Bonferroni 調整と同じ結果となった．Frison and Pocock(1992) が 10 の臨床試験を調べた報告によれば，visit 間の相関は 0.6–0.8 程度であることが多かった，そのような試験であれば Bonferroni 調整よりは精度よく検定できる．

4 おわりに

経時データの多重比較法は，従来そのための方法がわからないという理由で十分に行われてこなかったように思われる．原理は意外と簡単であり，SAS での実行も容易であることが示された．

参考文献

[1] Frison,L., and Pocock,S.J.(1992) Repeated Measures In Clinical Trials: Analysis Using Mean Summary Statistics And Its Implications For Design, Statistics In Medicine, 11, 1685-1704.

[2] Huynh,H. and Feldt,L.S.(1970) Conditions Under Which Mean Square Ratios In Repeated Measures Designs Have Exact F-distributions: JASA 11, 1582-1589.

[3] Hsu,J.C.(1992), The Factor Analytic Approach to Simultaneous Inference in the General Linear Model, Journal of Computatinal Statistics and Graphics, 1, 151-168.

[4] 岸本 淳司 (1997) 経時データにおける時点毎の比較の多重性調整法，日本統計学会第 65 回大会講演報告集.