直交表を用いた場合の分散分析表 列選択可能な、２・３水準共用関数について

慶応義塾大学理工学部
竹内寿一郎

１．はじめに

前回、分散分析表を計算するＪ言語のプログラムとして Smillie によるものを紹介した [1]。そのとき彼のプログラムに手を加えて平均平方と F_0 の計算および罫線を加えるなど、実際に分散分析を行うときに便利な関数になるように工夫した [2]。ところで直交表による分散分析では計算がずっと楽だったので、これまでは各列の平方和を計算するに留めてきたのが現状であった [3]。Smillie のプログラムに手を入れているうちに、これは直交表でも同じように出来ると思い、これまでずっと懸案にしていた直交表を用いた場合の分散分析表を作成したものである。

２．２水準の直交表の列平方和について

WEB に載っている 2^n 系直交表の各列平方和の関数は以下の通りである。

●２水準の場合

```
NB.　表の定義
l4=: 1 2 2 1 2 1 2 1 2 2 1
l4=:2*1.5-|:3 4$l4
l8=:1 1 1 1 2 2 2 2 1 1 2 2 1 1 2 2 1 1 2 2 2 2 1 1 1 2 1 2 1 2 1 2 1
 2 1 2 2 1 2 1 1 2 2 1 1 2 2 1 1 2 2 1 2 1 1 2
l8=:2*1.5-|:7 8$l8
l=.1 1 1 1 1 1 1 1 2 2 2 2 2 2 2 2 1 1 1 1 2 2 2 2 1 1 1 1 2 2 2 2
l=.l,1 1 1 1 2 2 2 2 2 2 2 2 1 1 1 1 1 1 2 2 1 1 2 2 1 1 2 2 1 1 2 2
l=.l,1 1 2 2 1 1 2 2 2 2 1 1 2 2 1 1 1 1 2 2 2 2 1 1 1 1 2 2 2 2 1 1
l=.l,1 1 2 2 2 2 1 1 2 2 1 1 1 1 2 2 1 2 1 2 1 2 1 2 1 2 1 2 1 2 1 2
l=.l,1 2 1 2 1 2 1 2 2 1 2 1 2 1 2 1 1 2 1 2 2 1 2 1 1 2 1 2 2 1 2 1
l=.l,1 2 1 2 2 1 2 1 2 1 2 1 1 2 1 2 1 2 2 1 1 2 2 1 1 2 2 1 1 2 2 1
l=.l,1 2 2 1 1 2 2 1 2 1 1 2 2 1 1 2 1 2 2 1 2 1 1 2 1 2 2 1 2 1 1 2
l=.l,1 2 2 1 2 1 1 2 2 1 1 2 1 2 2 1
l16=:2*1.5-|:15 16$l
NB.　各列の平方和を纏めて計算する。ここがＪらしいところ。
t2=: 3 : 0
:
(*:+/x.*y.)%#y.
)
NB.　見やすくするために列番号を入れ、ボックススタイルでまとめる。
tt2=:,.@:(<"1)@":@(>:@i.@# ,: ])@t2
NB.　使い方は以下の通り。
```

```
NB.  l8 tt2 y
NB.  l16 tt2 y
```

●３水準の場合

```
NB.  表の定義
l9=.1 1 1 2 2 2 3 3 3 1 2 3 1 2 3 1 2 3 1 2 3 2 3 1 3 1 2 1 2 3 3 1 2 2 3 1
l9=.|:4 9$l9
l=.1 1 1 1 1 1 1 1 1 2 2 2 2 2 2 2 2 2 3 3 3 3 3 3 3 3 3
l=.l,1 1 1 2 2 2 3 3 3 1 1 1 2 2 2 3 3 3 1 1 1 2 2 2 3 3 3
l=.l,1 1 1 2 2 2 3 3 3 2 2 2 3 3 3 1 1 1 3 3 3 1 1 1 2 2 2
l=.l,1 1 1 2 2 2 3 3 3 3 3 3 1 1 1 2 2 2 2 2 2 3 3 3 1 1 1
l=.l,1 2 3 1 2 3 1 2 3 1 2 3 1 2 3 1 2 3 1 2 3 1 2 3 1 2 3
l=.l,1 2 3 1 2 3 1 2 3 2 3 1 2 3 1 2 3 1 3 1 2 3 1 2 3 1 2
l=.l,1 2 3 1 2 3 1 2 3 3 1 2 3 1 2 3 1 2 2 3 1 2 3 1 2 3 1
l=.l,1 2 3 2 3 1 3 1 2 1 2 3 2 3 1 3 1 2 1 2 3 2 3 1 3 1 2
l=.l,1 2 3 2 3 1 3 1 2 2 3 1 3 1 2 1 2 3 3 1 2 1 2 3 2 3 1
l=.l,1 2 3 2 3 1 3 1 2 3 1 2 1 2 3 2 3 1 2 3 1 3 1 2 1 2 3
l=.l,1 2 3 3 1 2 2 3 1 1 2 3 3 1 2 2 3 1 1 2 3 3 1 2 2 3 1
l=.l,1 2 3 3 1 2 2 3 1 2 3 1 1 2 3 3 1 2 3 1 2 2 3 1 1 2 3
l=.l,1 2 3 3 1 2 2 3 1 3 1 2 2 3 1 1 2 3 2 3 1 1 2 3 3 1 2
l27=.|:13 27$l
NB.  Keyを用いて水準毎に２乗して合計する。これがＪらしいところ。
S3=.+/"1@*:@((|:@[)+//."1 ])
NB.  各列毎に平方和を計算する。
t3=. 3 : 0
:
((x. S3 y.)%((#y.)%3))-(*:+/y.)%#y.
)
NB.  見やすくするために列番号を入れ、ボックススタイルでまとめる。
tt3=.,.@:(<"1)@":@(>:@i.@# ,: ])@t3
NB.  使い方は以下の通り。
NB.    l9 tt3 y
NB.    l27 tt3 y
```

●例

```
   y8
6 5 9 2 4 9 0 7
   l8 tt2 y8
+--------------------+
|  1   2   3 4  5 6 7|
+--------------------+
|0.5 4.5 4.5 2 50 2 8|
```

```
+-------------------+
   y9
7 5 6 3 7 9 4 7 5
   l9 tt3 y9
+-------------------------------+
|      1       2       3       4|
+-------------------------------+
|1.55556 6.88889 16.8889 1.55556|
+-------------------------------+
```

例でも分かるようにこれから分散分析表をつくるのは結構計算も面倒である。

３．各モデルによる結果

【モデル１】$A_1 : x = 1$、$A_2 : x = 2$、$A_3 : x = 3$、$A_4 : x = 4$、繰り返し数 $n = 5$

```
   DATA1=.5 Fn1 point 1
   DATA1
 7.5652 5.38525 6.38359 7.15142 7.64221
6.73035 5.67498 3.32611 5.53201 4.03091
3.90604 6.39043 5.73018 4.26129 4.59514
8.04837 8.13776 6.28239 4.88261   7.902

   (1 2$'A ')AOV D1
-------------------------------------------------
Source   d.f.       S.S.     mean sq.       F0
-------------------------------------------------
A           3    18.58203     6.19401     4.19769
Error      16    23.60920     1.47558
-------------------------------------------------
Total      19    42.19124
-------------------------------------------------
```

図１．一元配置・完全無作為化法、4 水準、繰り返し数 5
$(a = 2.5,\ b = 5,\ n = 5,\ Interval = 1)$

x が１～４の幅であれば分散 1 の正規乱数を加えたとしても、水準による違いが大きく、ほぼ有意水準 5%で仮説が棄却される。

【モデル２】$A_1 : x = 1.45$、$A_2 : x = 2.15$、$A_3 : x = 2.85$、$A_4 : x = 3.55$、繰り返し数 $n = 5$

```
   DATA2=.5 Fn1 point 0.7
   DATA2
```

```
5.48515 6.75324 6.52646 8.53527 6.89115
6.38637 4.65813 4.62184  4.3171 6.63495
5.75182 4.58099 4.71279 5.74309 5.60617
6.79873 8.09027 6.90589 5.82505 4.83035


   (1 2$'A ')AOV DATA2
-------------------------------------------------
Source   d.f.       S.S.     mean sq.       F0
-------------------------------------------------
A           3     9.59464     3.19821     3.01097
Error      16    16.99500     1.06219
-------------------------------------------------
Total      19    26.58964
-------------------------------------------------
```

図 2．一元配置・完全無作為化法、4 水準、繰り返し数 5

$(a = 2.5,\ b = 5,\ n = 5,\ Interval = 0.7)$

最適値付近で x の幅が 1.45〜3.55 と区間幅が 0.7 に縮まることによって、特性値の変化が相対的に小さくなり、乱数の影響が大きくなって 5%検定では有意にならない場合が多く見られるようになる。

【モデル 3】$A_1 : x = 1.45$、$A_2 : x = 2.15$、$A_3 : x = 2.85$、$A_4 : x = 3.55$、繰り返し数 $n = 10$

```
   DATA3=.10 Fn1 point 0.7
   |:DATA3
5.48515 6.38637 5.75182 6.79873
6.75324 4.65813 4.58099 8.09027
6.52646 4.62184 4.71279 6.90589
8.53527  4.3171 5.74309 5.82505
6.89115 6.63495 5.60617 4.83035
6.58019 5.83067 5.11892 7.15034
6.22389 4.27484 5.19373 3.32011
6.09753 4.90435 6.10052 9.54786
 8.1165 5.05414 3.09461 5.15649
8.05117 3.93814 4.71554 5.77138
   (1 2$'A ')AOV D1
-------------------------------------------------
Source   d.f.       S.S.     mean sq.       F0
-------------------------------------------------
A           3    26.39713     8.79904     6.20075
Error      36    51.08505     1.41903
-------------------------------------------------
```

```
Total        39    77.48218
-------------------------------------------------
```

図 3．一元配置・完全無作為化法、4 水準、繰り返し数 10

$(a = 2.5,\ b = 5,\ n = 10,\ Interval = 0.7)$

繰り返し数を増やすことにより検出力が増加し、最適値付近で誤差に比べて変化の割合が相対的に小さくても、仮説は高い確率で棄却される。

4．分散分析表を計算するプログラム

```
NB. Appendix Analysis-of-variance calculations by Smillie
NB. Revised by J.Takeuchi 2004/10/22
T=: 3 : 0
:
+/^:(+/-.x.)(/:x.)|:y.
)
tt=: ,@{@(]@#&(<0 1))
S=: (+/@((*:@,)@T)) % (*/@(-.@[ # $@]))
allS=: (>@tt@#@$) S"1 _ ]
stt=: <"1@|."1@(/: +/"1)@(>@tt)
alphabet=: 'ABCDEF'
tag=: ]#({.&alphabet)@#@]
numtag=: (({.&alphabet)@(#@$@])) e. [
alltags=: (tag &. >)@(1&}.)@stt
expand=: /:@\:@[{#@[{.]
ps=: <"1@expand"1 >@stt@(+/)
ss=: | @ (-/) @ (>@ps@([numtag]) (+/"1@[ +//. S"1 _) ])
df=: ([:*/<:)@(numtag # $@])
term=: ss (] , [ , %) df
allterms=: (>@[) term"1 _ ]
   AOV=: 3 : 0
s0=: (alltags #$y.) AOV y.
:
Tyt=:,:'Source   d.f.       S.S.     mean sq.       F0   '
AOVtable=.x. allterms y.
Labels=.8{."1 >x.
dfTotal=.<:*/$y.
ssTotal=.(+/*:,y.) - '''' ss y.
dfError=.dfTotal - +/0{"1 AOVtable
if. dfError>0 do.
ssError=.ssTotal - +/1{"1 AOVtable
AOVtable=.AOVtable,ssError(],[,%)dfError
Labels=.Labels,'Error'
```

```
end.
AOVtable=.AOVtable,.({:"1 AOVtable)%ssError%dfError
r=.Labels,"1 (5 12.5 12.5 12.5)":AOVtable
r=.(12#' ')(<_1;12{.i:12)}r
AOVtable=.AOVtable,TotalRow=.dfTotal,ssTotal
r=.(r,'-'),'Total    ',5 12.5":TotalRow
'-',Tyt,'-',r,'-'
)
```

5．参考文献

[1] Smillie,K.(1993)　Some Notes on Introducing J with Statistical Examples Technical Note of Dep. of Computing Science Univ. of Alberta Edomonton,Alberta T6G 2H1

[2] 竹内寿一郎 (2004)： 2次応答関数を用いた分散分析表 (一元配置・完全無作為化法の場合)　JAPLA 研究会 10 月例会、統計数理研究所

[3] 竹内寿一郎 (2004)：WEB サイト、
http://www.ae.keio.ac.jp/lab/soc/takeuchi/japla/contents/anova.zip
現在は修正して、ここで紹介した関数が載っている。