응용통계연구 제 5권 2호
1992년 pp 229-242

행-열 실험계획의 분석에 관한 연구

백 운 봉1)

< 요 약 >

Bradley & Stewart(1991)에서 인용하고 있는 예에 대해서 행간, 열간 정보를 회복하여 이것을 이용하고 블록내 분석 결과와 결합하는 방법을 생각해 볼 수 있다. 이것은 행효과와 열효과를 다같이 확률효과(random effects)로 간주하여 일반화 최소제곱법(Generalized least squares method)에 의해서 해를 구하는 것과 동일한 것이다. 이것이 Paik(1986)에서 논의되고 있다. 이 방법은 어떤 행-열 계획(Row-column design)에도 적용된다. 따라서 격자방격(格子方格, Lattice square)에도 그대로 적용된다. 그런데 이와 같은 방법은 보통 불완비 블록계획(incomplete block designs)에서의 방법을 확대 적용하여 얻을 수 있다. 이러한 블럭실험에 대한 SAS/IML을 이용한 분석법은 백운봉(1990a,1990b)에 의해서 제안된 바 있다. 그러나 이것이 개선될 필요가 있었고, 이 개선된 방법을 확대 적용한 것이 본 논문이다. 블록실험에 대한 개선된 방법은 본 논문 말미에 부록으로 제공 되어 있다.

1. 머 리 말

행-열 실험계획(Row-column design, 간단히 RCD) 중 우리와 가장 친숙한 것으로는 라틴방격(latin square), 유덴방격(Youden square)를 예로 들을 수 있다. 격자방격법(格子方格法, lattice square design)도 이에 속한다. 또 이들은 어느 것이나 소위 구조적으로 완비(structurally complete)된 것이라고 분류된다. 이에 대해서 다음에 우리가 다루게 될 실험계획을 구조적으로 불완비된(structurally incomplete) 계획이라고 한다. 라틴방격의 경우는 행과 열에 관한 결합행렬 N_{rc}의 모든 원소가 1로 된다. 이러한 경우 구조적으로 완비되어 있다고 한다. 이와 비교하여 다음 예의 경우는 그렇지가 않다. 그러나 이러한 것이 우리가 설명하는 분석법에 차이를 가져오지는 않는다.

다음 <표 1.1>은 미국 어느 도시내 10개 장소에서의 교통량의 관측결과 이다. 월요일 부터 금요일 까지, 8시 부터 9시 까지 사이의 6개 5분 구간 8:00-8:05, 8:10-8:15,...,8:50-8:55 에서의 통과 자동차수를 조사하고 이것의 저곱근을 구하여 분석대상 자료로 제시한 것이다. 이것은 Bradley & Stewart(1991)에서 다원블록계획(multidimensional block design, MBD)의 분석방법을 설명하는 과정에서 인용되고 있

1) (137-049) 서울특별시 서초구 반포본동 943, 반포아파트 101동 503호

는 예이다. <표 1.1>에서도 이 실험계획이 구조적으로 불완비한 것이라는 것이 잘 나타나고 있다.

〈표 1.1〉 장소, 요일, 시간대 별 교통량(제곱근변환한 것)

장소(Loca)	요일				
	Mon	Tues	Wedn	Thurs	Fri
1	8.49(1)	7.07(2)		9.43(3)	
2		7.00(1)	7.07(4)		7.42(2)
3	7.68(3)		7.87(1)		5.83(5)
4	7.28(4)		7.35(6)	7.21(1)	
5		6.16(5)		8.19(6)	7.55(1)
6	10.05(2)			6.78(5)	6.78(4)
7			7.00(3)	5.92(2)	6.93(6)
8	3.16(5)	3.46(6)	3.61(2)		
9		9.90(3)	8.54(5)	9.06(4)	
10	8.83(6)	9.59(4)			10.00(3)

* 괄호안의 수자는 시간(Time)대를 순서대로 수자번호로 나타낸 것이다.

이 예는 가장 단순한 MBD인 행-열 실험계획(RCD)에 관한 것이다. 이 예에서의 성분계획(component designs)는 행을 블록으로 하는 처리에 관한 블록계획, 열을 블록으로하는 처리에 관한 블록계획, 그리고 행을 처리로 하고 열을 블록으로 생각하는 블록계획 이다. 이들 각 블록계획의 결합행렬을 각각 N_r와 N_c 그리고 N_{rc}로 표시하면 다음과 같이 된다.

$$N_r=\begin{pmatrix} 1&1&1&1&1&0&0&0&0&0 \\ 1&1&0&0&0&1&1&1&0&0 \\ 1&0&1&0&0&0&1&0&1&1 \\ 0&1&0&1&0&1&0&0&1&1 \\ 0&0&1&0&1&1&0&1&1&0 \\ 0&0&0&1&1&0&1&1&0&1 \end{pmatrix}$$

$$N_c=\begin{pmatrix} 1&1&1&1&1 \\ 1&1&1&1&1 \\ 1&1&1&1&1 \\ 1&1&1&1&1 \\ 1&1&1&1&1 \\ 1&1&1&1&1 \end{pmatrix} \qquad N_{rc}=\begin{pmatrix} 1&1&0&1&0 \\ 0&1&1&0&1 \\ 1&0&1&0&1 \\ 1&0&1&1&0 \\ 0&1&0&1&1 \\ 1&0&0&1&1 \\ 0&0&1&1&1 \\ 1&1&1&0&0 \\ 0&1&1&1&0 \\ 1&1&0&0&1 \end{pmatrix}$$

여기에서 열을 블록으로한 계획은 행렬 N_c에서 알 수 있는 바와 같이 완비블록

(complete block)계획으로 되어 있으나 반드시 그래야 되는 것은 아니다. 또 위의 예가 유덴방격(Youden square)과 다른 점은 유덴방격의 경우는 행렬 N_{rc}의 모든 원소가 1로 되어 구조적으로 완비된 RCD이지만 여기에서는 그렇지가 않다는 것이다.

Bradley & Stewart(1991)에서는 일반적인 MBD의 블록내 분석(intrablock analysis)을 제시하고 아울러 실험계획의 효율성을 논의하고 있다. 이와같은 블록내 분석은 SAS/GLM으로 간다히 얻을 수가 있다. 그런데 이러한 방법은 장소(location)효과나 요일(days)효과가 다같이 상수효과(fixed effect)라는 가정아래 보통 최소제곱법(ordinary least squares method)에 의해서 풀이하는 방법이다. 이와는 달리 장소효과나 요일효과가 다같이 확률효과(random effects)라는 가정아래 일반화 최소제곱법(generalized least squares method, 간단히 GLS법이라 하자)에 의한 해(solution)를 구할 수가 있다. 본 논문에서는 블록내 분석을 위한 SAS/GLM절차에서 얻은 결과와 결합하여 SAS/IML에 의해서 일반화 최소제곱법을 실행하는 방법을 설명할 것이다. 그런데 여기에서 사용한 방법은 백운봉(1990a, 1990b)에서 제안한 것을 개선한 방법이다. 여기에서 개선하였다는 것은 전에는 SAS/IML만을 이용해서 프로그램을 작성하였으나 이것을 SAS/GLM과 SAS/IML를 연계해서 좀더 요령있게 그리고 간편하게 새로이 프로그램을 작성하였다는 말이다. 이 방법이 '부록'에서 설명되어 있다.

2. SAS/GLM에 의한 블록내 분석

위의 예에 대한 수학적 모형은 다음과 같다.

$$Y_{ijk} = \mu + \tau_i + \rho_j + \gamma_k + \varepsilon_{ijk} \tag{2.1}$$

$i=1,..,v;\ j=1,..,r;\ k=1,..,c.$

(위 예는 v=6, r=10, c=5의 경우 임)

μ=일반 평균

τ_i=i번 째 처리효과(time효과)

ρ_j=j번 째 행 효과(location효과)

γ_k=k번 째 열 효과(요일(day)효과)

ε_{ijk}=오차항으로서 평균이 0 분산이 σ_e^2인 정규분포에 따른다.

이러한 선형모형(linear model)으로 표현되는 경우라면 어떤 형식의 불완비 블록계획이든 또는 어떤 조건을 갖춘 MBD이든 간에 PROC GLM에 의해서 블록내 분석이 가능하고 분산분석표가 작성 된다. 그러나 이러한 것이 처리 (위 예의경우는 Times) 의 LSMEANS가 구해지고 이들간의 차의 추정이 가능하다는 것을 의미하지는 않는다. 이와 같은 추정가능(estimability)문제는 실험계획과 모형설정에 달려있는 것이다. 우리는 이러한 것이 가능한 실험계획하에서 얻은 데이타에 대한 분석방법을 생각하고

있는 것이다.

우선 다음과 같이 입력하고 후에 이것을 활용하기 위하여 DAY의 수준을 수량화한 데이타셋 DATA NEW를 창출한다.

```
DATA MBD;
DO LOCA=1 TO 10;
 DO COL=1 TO 3;
 INPUT DAY $ TIMES Y @; OUTPUT;
END; END;
CARDS;
MON  1 8.49  TUES  2 7.07 THURS 3 9.43
TUES 1 7.00  WEDN  4 7.07 FRI   2 7.42
MON  3 7.68  WEDN  1 7.87 FRI   5 5.83
MON  4 7.28  WEDN  6 7.35 THURS 1 7.21
TUES 5 6.16  THURS 6 8.19 FRI   1 7.55
MON  2 10.05 THURS 5 6.78 FRI   4 6.78
WEDN 3 7.00  THURS 2 5.92 FRI   6 6.93
MON  5 3.16  TUES  6 3.46 WEDN  2 3.61
TUES 3 9.90  WEDN  5 8.54 THURS 4 9.06
MON  6 8.83  TUES  4 9.59 FRI   3 10.00
RUN;
DATA NEW;
SET MBD; DAYS=(DAY='MON')+2*(DAY='TUES')+3*(DAY='WEDN')
             +4*(DAY='THURS')+5*(DAY='FRI'); RUN;
```

프로그램 (1):

```
PROC GLM;
CLASS LOCA DAYS TIMES;
MODEL Y=TIMES DAYS LOCA;
LSMEANS TIMES;
OUTPUT OUT=NEW1 P=YHAT R=RESID1; RUN;
```

출력결과(블럭내 분석)의 일부:

Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	18	85.57525333	4.75418074	7.24	0.0009
Error	11	**7.22664333**	**0.65696758**		
Corrected Total	29	92.80189667			

R-Square	C.V.	Root MSE	Y Mean
0.922128	10.99230	0.810535	7.37366667

Source	DF	Type I SS	Mean Square	F Value	Pr > F
LOCA	9	75.56256333	8.39584037	12.78	0.0001
DAYS	4	1.66137333	0.41534333	0.63	0.6498
TIMES	5	8.35131667	1.67026333	2.54	0.0916

TIMES	Y LSMEAN
1	7.60366667
2	7.54366667
3	8.19783333
4	7.08033333
5	6.27283333
6	7.54366667

이들 값에서 평균 7.37366667를 빼면 다음과 같은 처리효과를 얻는다.

EFFECTS OF TIMES=(0.23, 0.17, 0.82, −0.29, −1.10, 0.17)

제 4절에 대한 준비로 다음 두 프로그램을 실행한다.

프로그램 (2):

```
PROC GLM DATA=NEW;
CLASS DAYS TIMES;
MODEL Y=TIMES DAYS/SS1;
OUTPUT OUT=NEW2 P=P R=RESID2; RUN;
```

출력결과의 일부:

Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	9	25.53937000	2.83770778	0.84	0.5863
Error	20	**67.26252667**	**3.36312633**		
Corrected Total	29	92.80189667			

프로그램 (3):

```
PROC GLM DATA=NEW;
CLASS LOCA TIMES;
MODEL Y=TIMES LOCA/SS1;
OUTPUT OUT=NEW3 P=P R=RESID3; RUN;
```

출력결과의 일부:

Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	14	83.91388000	5.99384857	10.12	0.0001
Error	15	**8.88801667**	**0.59253444**		
Corrected Total	29	92.80189667			

3. 일반화 최소제곱법에 의한 분석

불완비 블록계획에서 블록효과를 상수효과(fixed effect)로 다루어, 어떤 제약조건 하에서의 처리효과를 추정하는 방법을 블록내 분석(intrablock analysis)이라고 한다. 이것은 SAS/GLM을 이용해서 쉽게 실행할 수가 있다. 그러나 실제에 있어서는 블록효과를 확률적인 효과(random effect)라고 생각하는 것이 합당할 때가 많이 있다. 이렇게 생각하는 경우는 소위 블록간 정보를 이용한 처리효과의 추정도 가능하게된다(interblock analysis). 그러므로 실험자는 보다 나은 추정치를 얻기위하여 이들 두 추정치를 결합시킨다. 그런데 이와 같은 결합추정치(combined estimator)는 블록효과가 독립 확률변수라는 생각아래 GLS법으로 구해지는 것과 같다는 것을 밝인바 있다(Paik(1986)를 참조할 것).

행-열 계획에서도 이와 같은 행간-열간 정보를 이용한 추정치와 위에서의 블록내 분석에서 얻은 추정치와의 결합추정치가 다음에 설명하는 GLS법에 의해서 구해진다. 모형 (2.1)을 행렬표현으로 나타내면 다음과 같다.

$$Y=\mathbf{1}\mu + X_1\tau + X_2\rho + X_3\gamma + \varepsilon \tag{3.1}$$

여기에서 Y는 N×1 관측치 벡터이고 **1**는 원소가 전부 1인 N×1 벡터이며, X_1, X_2, X_3는 각 각 처리효과벡터 τ, 행효과벡터 ρ, 그리고 열효과벡터 γ에 대응하는 계획행렬들 이다. ε는 N×1 오차벡터를 나타낸다. 여기에서 ρ, γ, 그리고 ε은 행, 열, 그리고 오차에 관한 랜덤효과 벡터로서, 서로 독립이고 $E[\rho]=0$, $E[\gamma]=0$, $E[\varepsilon]=0$ 이며 $V[\rho]=I_r\sigma_r^2$, $V[\gamma]=I_c\sigma_c^2$, $V[\varepsilon]=I_N\sigma_e^2$이라고 가정한다. 그러면 Y의 분산은 다음과 같이 된다. 여기에서 I_a는 크기가 a인 단위행렬을 나타낸다.

$$V[Y]=X_2X_2'\sigma_r^2 + X_3X_3'\sigma_c^2 + I_N\sigma_e^2 \tag{3.2}$$

이것을 이용하여 일반화 최소제곱법을 실행하자면 (3.2)의 각 분산성분을 추정하지않으면 안된다. 이것은 위의 PROC GLM에서, OUTPUT-DATA OUT에서의 RESID1, RESID2, RESID3을 이용한 IML프로그램에서 구하여 이용한다.

처리 τ에 관한 축소된 정규방정식(reduced normal equation)은 $\Omega=V[Y]^{-1}$로 놓고 $R=X_1'\Omega X_1$, $K=\mathbf{1}'\Omega\mathbf{1}$, $N=X_1'\Omega\mathbf{1}$, $T=X_1'\Omega Y$, $B=\mathbf{1}'\Omega Y$와 같이 놓았을 때 다음과 같이 된다 (부록 2를 참조할 것).

$$[R - NN'/K]\tau = T - NB/K \tag{3,3}$$

V[Y]를 추정하고 이것의 역행렬을 구하고 이것을 일반화 최소제곱법을 적용하는데 이용한다는 것은 컴퓨터를 사용하지 않고는 실제로 실행하기 어려운 일이다. 그러나 이것은 SAS 패키지를 활용하여 컴퓨터를 자유롭게 사용할 수 있는 이 시대에 이러한 방법을 외면한다는 것도 어리석은 짓이라고 생각한다.

4. SAS/IML을 이용한 분석

```
PROC IML WORKSIZE=150;
RESET NOLOG;
START RCD;
SEE=SUM(RESID1#RESID1); SEE2=SUM(RESID2#RESID2);
SEE3=SUM(RESID3#RESID3);
SER=SEE2-SEE; SEC=SEE3-SEE;
PRINT, SEE SEE2 SEE3; /* PROC GLM ①,②,③에서의 Error SS와 비교해 볼 것 */
v=NCOL(X1); rw=NCOL(X2); cl=NCOL(X3); k=cl;
Nr=X1`*X2; rk=SUM(Nr[,1]); r=rw*rk/v;
dfe=v*r-v-rw-cl+2; dfr=rw-1; dfc=cl-1;
MSe=SEE/dfe; MSr=SER/dfr; MSc=SEC/dfc;
sigma=MSe;
sigmaR=(MSr-MSe)*dfr/(v*r-v-cl+1);
sigmaC=(MSc-MSe)*dfc/(v*r-v-rw+1);
PRINT, SIGMA SIGMAR SIGMAC;
X22=X2*X2`; X33=X3*X3`; OBS=NROW(X22);
PRINT, "GENERALIZED LEAST SQUARE METHOD FOR R by C DESIGN ";
VY=X22*SIGMAR+X33*SIGMAC+I(OBS)*SIGMA;
VI=INV(VY);
J=J(OBS,1,1); Kn=J`*VI*J;
R=X1`*VI*X1; X12=J`*VI*X1; NN=X12`*X12;
C=R-NN/Kn;
Tt=X1`*VI*Y; BY=J`*VI*Y;  NB=X12`;
Q=Tt-NB*BY/Kn;
AT=INV(C+1)*Q; NT=NROW(AT);
ADJTIMES=AT-SUM(AT)/NT;
ADJMEAN=ADJTIMES+SUM(Y)/OBS;
ADJ=ADJTIMES || ADJMEAN;
FINISH;

USE NEW;
VNAME={TIMES LOCA DAYS Y};
READ ALL VAR VNAME INTO XYZ;
TIMES=XYZ[,1]; LOCA=XYZ[,2]; DAYS=XYZ[,3]; Y=XYZ[,4];
X1=DESIGN(TIMES); X2=DESIGN(LOCA); X3=DESIGN(DAYS);
USE NEW1;
```

```
READ ALL VAR {RESID1} INTO RESID1;
USE NEW2;
READ ALL VAR {RESID2} INTO RESID2;
USE NEW3;
READ ALL VAR {RESID3} INTO RESID3;
RUN RCD;
cl={'adj. time effect' 'adj.mean'};
PRINT, ADJ[COLNAME=cl FORMAT=10.4];
QUIT;
```

출력결과:

```
                    SEE       SEE2       SEE3
              7.2266433 67.262527 8.8880167

                  SIGMA    SIGMAR    SIGMAC*
              0.6569676 2.7061588 -0.064433
```

* SIGMAC 즉 σ_c^2이 음수로 되어 있다는 것에 주목할 필요가 있다. 그런데 여기에서는 분산분석표에서의 평균제곱(MS)과 그 기대치 즉 EMS를 이용해서 각 분산성분을 추정하는 방법을 택하고 있으므로(이 방법을 SAS/VARCOMP에서는 TYPE 1 METHOD라고한다), 성분값이 음수로 나타날 가능성은 항상 있는 것이다. 이러한 경우, 분산성분이 음수인 것은 0으로 간주하는 것이 보통이다. 그러므로, 위 예에서는 SIGMAC를 구하는 절차 다음에 SIGMAC=0을 삽입하고 다시 RUN시키면 될것이다. 그런데 위의 예의 경우는 이것에 의해서 달라지는 것이 거이 없다. 출력결과는 다음과 같다.

```
         GENERALIZED LEAST SQUARE METHOD FOR R by C DESIGN
      ADJ adj. time effect        adj.mean      LSMEANS(참고용)
                    0.2319          7.6055        7.6037
                    0.1030          7.4766        7.5437
                    0.8797          8.2533        8.1978
                   -0.2129          7.1608        7.0803
                   -1.1173          6.2564        6.2728
                    0.1156          7.4893        7.5437
```

부록 1. 블럭계획에서의 일반화 최소제곱법의 적용방법

저자는 BIB 또는 PBIB실험계획등 일바적인 불완비 블록계획에 대한 일반화 최소제곱법에 의한 분석법을 SAS/IML을 이용하여 제시한바 있다(1990a, 1990b). 다음은 이와 같은 본인이 제안한 방법을 개선한 것이다. 이 개선된 방법이 본문 제 4절에서 이용되고 있다.

다음 예에서 반복(rep)이라고 한 것은 각 처리의 한 반복이 몇개 블록의 묶음으로 이루어지고(Design arranged in replication), 이러한 묶음의 수를 반복(rep)으로 나타낸 것이다.

<예 1>: BIB 계획(처리별 반복수는 2이나 계획의 반복(rep)은 1이라고 생각한다)

반복(rep)	블록	처리번호와 측정치			
1	1	(1)	15	(2)	17
	2	(1)	14	(3)	12
	3	(2)	20	(3)	23

<예 2>: Lattice 계획(처리별 반복수가 2, 계획반복(rep)수도 2 이다)

반복(rep)	블록	처리번호와 측정치					
1	1	(1)	11	(2)	12	(3)	14
	2	(4)	9	(5)	17	(6)	13
	3	(7)	15	(8)	14	(9)	12
2	1	(1)	21	(4)	23	(7)	22
	2	(2)	19	(5)	25	(8)	24
	3	(3)	20	(6)	22	(9)	18

데이타의 입력방법

<예 1>의 경우:

```
DATA D1;
INPUT REP @;
DO BLOCK=1 TO 3;
DO SIZE=1 TO 2;
INPUT TRT Y @; OUTPUT;
END; END;
CARDS;
1 1 15 2 17
  1 14 3 12
  2 20 3 23
RUN;
```

<예 2>의 경우:

```
DATA D1;
INPUT REP @;
DO BLOCK=1 TO 3;
DO SIZE=1 TO 3;
INPUT TRT Y @; OUTPUT;
END; END;
CARDS;
1 1 11 2 12 3 14
  4  9 5 17 6 13
  7 15 8 14 9 12
2 1 21 4 23 7 22
  2 19 5 25 8 24
  3 20 6 22 9 18
RUN;
```

<예 2>는 격자법(Lattice design)에서 처리수가 9, 2회 반복의 경우의 것이나 이와 같은 단순격자(2중 격자)을 p회 되풀이(repetition)한 경우는 반복내 블록 SS를 구하기 위하여 Component (a)와 Component (b)를 계산하는 것으로 되여 있다. 그러나 SAS/GLM을 사용하는 경우는 이것을 단순히 2p회 반복한 것이라고 생각하면 된다. 그러나 이때 관측치 수가 너무나 클 때(예로 100이상)는 V[Y]를 대각선에 따라 반복별로 분할하여 블록행렬의 역행렬을 구하여 이용하는 방법을 생각할 수 있다.

분석을 위한 PROC GLM과 PROC IML:

```
PROC GLM;
CLASS REP BLOCK TRT;
MODEL Y=REP BLOCK(REP) TRT/SS1 SS2;
LSMEANS TRT;
OUTPUT OUT=D11 P=YHAT R=RESID1;
RUN;
PROC GLM;
CLASS REP TRT;
MODEL Y=REP TRT/SS1;
OUTPUT OUT=D12 P=P R=RESID2;
RUN;

PROC IML WORKSIZE=150;
RESET NOLOG;
START PBIB;
SEE=SUM(RESID1#RESID1); SEE2=SUM(RESID2#RESID2);
SEB=SEE2-SEE;
PRINT, SEE SEE2 SEB;
N=X1'*X2;
v=NCOL(X1); b=NCOL(X2); rep=NCOL(X3);
k=SUM(N[,1]); r=b*k/v;
PRINT, v rep b k r;
dfe=v*r-v-b+1; dfb=b-rep;
MSe=SEE/dfe; MSb=SEB/dfb;
sigma=MSe;
sigmab=(MSb-MSe)*dfb/(v*r-v-k*rep+k);
PRINT, SIGMA SIGMAB;
X22=X2*X2'; OBS=NROW(X22);
PRINT, "GENERALIZED LEAST SQUARE METHOD FOR IB DESIGN ";
VY=X22*SIGMAB+I(OBS)*SIGMA;
VI=INV(VY);
J=J(OBS,1,1); Tn=J'*VI*J;
R=X1'*VI*X1; X12=J'*VI*X1; NN=X12'*X12;
C=R-NN/Tn;
Tt=X1'*VI*Y; BY=J'*VI*Y;  NB=X12';
Q=Tt-NB*BY/Tn;
Cov=INV(C+1);
AT=Cov*Q;
```

```
ADJTRT=AT-SUM(AT)/v;
ADJMEAN=ADJTRT+SUM(Y)/obs;
ADJ=ADJTRT || ADJMEAN; cl={'adj. trt. effect' 'adj. mean'};
PRINT, ADJ[COLNAME=cl FORMAT=8.4];
Chisq=AT'*Q;  DF=V-1;
PRINT, "Chisquare value and degrees of freedom", CHISQ DF;
FINISH;

USE D1;
VNAME={TRT BLOCK REP Y};
READ ALL VAR VNAME INTO XYZ;
TRT=XYZ[,1]; BLK=XYZ[,2]; REP=XYZ[,3]; Y=XYZ[,4];
X1=DESIGN(TRT); X2=DESIGN(BLK); X3=DESIGN(REP);
X2=HDIR(X3,X2);
USE D11;
READ ALL VAR {RESID1} INTO RESID1;
USE D12;
READ ALL VAR {RESID2} INTO RESID2;
RUN PBIB;
QUIT;
```

부록 2. 전통적인 방법과의 관계

다음과 같은 블록(block)실험 모형을 생각한다.

$$Y=1\mu + X_1\tau + X_2\beta + \varepsilon \tag{부.1}$$

처리수=v, 반복수=r, 블록수=b, 블록의 크기=k, 따라서 vr=bk이고, Y는 관측치 벡터, μ는 일반평균, 1은 vr×1인 원소가 전부 1인 벡터, τ는 v×1 처리 벡터, β는 b×1 블록벡터 이고, ε는 오차 벡터 이다. 그리고 X_1, X_2는 각 모수 τ와 β에 대응하는 계획행렬 이다.

$Y=(Y_1',Y_2',\ldots,Y_b')'$이고 Y_j를 j번 블록에서의 관측치 벡터라고 하여도 일반성을 잃치 않을 것이다. 이때 $X_2X_2'=\mathrm{Diag}(J,J,\ldots,J)$와 같이 된다. 여기에서의 J는 모든 원소가 1인 k×k 행렬을 나타낸다.

여기에서 모든 j에 대해서 $E[\beta_j]=0$이고, $V[\beta_j]=\sigma_b^2$ 이며, 모든 i, j에 대해서 $E[\varepsilon_{ij}]=0$이고 $V[\varepsilon_{ij}]=\sigma_e^2$ 이라고 가정한다. 그러면 $E[Y]=1\mu + X_1\tau$ 이고

$$V[Y]=X_2X_2'\sigma_b^2 + I\sigma_e^2 \tag{부.2}$$

와 같이 된다(I는 단위행렬). $X_2X_2'=\mathrm{Diag}(J,J,\ldots,J)$이므로 $M=\sigma_e^2I+\sigma_b^2J$로 놓으면 $V[Y]=\mathrm{Diag}(M,M,\ldots,M)$와 같이 쓸 수 있다. 그리고 M^{-1}는 다음과 같이 된다.

$$M^{-1}=\omega I-(1/k)(\omega-\omega')J,\ \text{단}\ \omega=1/\sigma_e^2,\ \omega'=1/(\sigma_e^2 + k\sigma_b^2) \tag{부.3}$$

그런데 $V[Y]^{-1}=Diag(M^{-1},M^{-1},...,M^{-1})$와 같이 되므로 $\Omega=V[Y]^{-1}$ 로 놓으면

$$\Omega=\omega I-(1/k)(\omega-\omega')Diag(J,J,...,J) \quad (부.4)$$

$$=\omega I-(1/k)(\omega-\omega')X_2X_2' \quad (부.5)$$

일반화 최소제곱(GLS)법 이라는 것은 $(Y-E[Y])'\Omega(Y-E[Y])$를 최소로하는 μ,τ 값을 구하는 것을 말한다. 이것으로 다음과 같은 정규방정식을 얻는다.

$$(1'\Omega 1)\mu + (1'\Omega X_1)\tau = 1'\Omega Y \quad (부.6)$$

$$(X_1'\Omega 1)\mu + (X_1'\Omega X_1)\tau = X_1'\Omega Y \quad (부.7)$$

(부.7) - $(X_1'\Omega 1)(1'\Omega 1)^{-1}\times$(부.6) 과 같은 계산으로 τ에 대한 축소된 정규방정식 (reduced normal equation)을 다음과 같이 얻는다.

$$[X_1'\Omega X_1-(X_1'\Omega 1)(1'\Omega 1)^{-1}(1'\Omega X_1)]\tau$$
$$=X_1'\Omega Y - (X_1'\Omega 1)(1'\Omega 1)^{-1}(1'\Omega Y) \quad (부.8)$$

이것이 Paik(1986)에서의 정규방정식의 표현(p.121)과 같다는 것이 다음과 같이 증명된다. (부.5)에서

$$X_1'\Omega X_1=\omega X_1'X_1 - (1/k)(\omega-\omega')X_1'X_2X_2'X_1$$
$$=\omega rI - (1/k)(\omega-\omega')NN',\ 단\ N=X_1'X_2$$

또, (부.4)에서 $\Omega J=\omega J - (\omega-\omega')J=\omega' J$, $\Omega J\Omega=\omega' J\Omega=\omega'\omega' J$ 이며

$1'\Omega 1=(\omega 1'-(1/k)(\omega-\omega')k1')1=bk\omega'$, 그리고 $X_1'JX_1=rJX_1=r^2J$, 따라서

$$X_1'\Omega 11'\Omega X_1=X_1'\omega'\omega' JX_1=\omega'\omega' X_1'JX_1=\omega'\omega' r^2J$$

이들 결과를 (부.8)의 좌변에 대입하고, (부.5)에서의 Ω를 (부.8)의 우변에 대입하여 계산하면 다음과 같은 전통적 표현방법으로 나타낸 정규방정식을 얻는다. 여기에서 $N=X_1'X_2$, $T=X_1'Y$, $B=X_2'Y$ 이고 g는 전체 관측치의 합계를 나타낸다.

$$[\omega rI-(1/k)(\omega-\omega')NN'-\omega' r^2J/(bk)]\tau$$
$$=\omega T-(1/k)(\omega-\omega')NB-\omega' r1g/(bk) \quad (부.9)$$

< 참고문헌 >

[1] 백운봉(1990a), *SAS/IML 행렬에 의한 통계분석*, 자유아카데미.

[2] 백운봉(1990b), "SAS/IML을 이용한 불완비블록계획의 분석", *응용통계 5, 17-34*, 고려대학교 통계연구소.

[3] Bradley, R. A. and Stewart, F. P.(1991), "Intrablock analysis of designs with multiple blocking criteria," *Journal of the American Statistical Association, 86*, 792-797.

[4] Clatworthy, W. H. (1973), *Tables of Two-Associate-Class Partially Balanced Designs*, NBS Applied Mathematics Series 63, U.S Dept. of Commerce.

[5] Cochran, W. G. and Cox, G. M.(1957), *Experimental Designs*, 2nd ed. John Wiley.

[6] John, J. A. (1987), *Cyclic Designs*, Chapman & Hall.

[7] Paik, U. B. (1986), "Another look at combined intrablock and interblock estimation in block designs," *Journal of the Korean Statistical Society 15*, 118-126.

On Analysis of Row-Column Designs

Uhn Boong Paik[2)]

< ABSTRACT >

Bradley and Stewart(1991) considered a large class of experimental designs as multidimensional block designs(MBD's). The simplest MBD could be considered to be a row-column design(RCD). They presented the intrablock analysis of variance for a general row-column design. In this article, a generalized least squares solution for Bradley & Stewart's example is considered. In this case, the assumption is that row and column effects are random. This is an application of revised Paik(1990a, 1990b)'s method. The Appendix is devoted to that revised method.

2) #101-503, Banpo Apt. Seoch-ku, Banpo-bondong, Seoul 137-049, KOREA.