

실험계획법의 전산화에 관한 연구 (III)

- Studies on the Computerization of Design of Experiments (III) -

정수일*

ABSTRACT

This paper studies computer programming for the analysis of data obtained by experiments using Orthogonal Arrays. The following items are considered in the computer programming :

- * significant digits in the computation of Sum of Squares, Mean Squares and Variance Ratios
- * containing the necessary F-distribution values in the program
- * matching the rules of KS A 0021 and 3251 in the digit treatments etc.

The running results of ANOVA Table and Pooled ANOVA Table of a fictitious example is added with the parts of a program. It should be mentioned that the main purpose of this paper is in the arousing of the discussion about significant digits concept in the computer programming for various kinds of Statistical Methods.

I. 서론

최근 자동화, Robot화 등이 급진전되면서 CIM, CAE, DSS 등도 빈번하게 회자되고 있다. 이러한 추세 속에서 Statistical Package, SPC 등의 전산 프로그램 또는 이와 관련되는 참고서적을 하나도 구비하고 있지 않은 기업은 이제 찾아보기가 힘들게 되었다.

공정이 점차 컴퓨터에 의해 자동 조절됨에 따라 이에서 얻어지는 데이터의 양상은 과거와 달라졌으며 이는 통계적방법의 활용습관을 대폭적으로 바꿀 수 밖에 없는 새로운 국면을 초래하게 되었다. 컴퓨터로 공정관리를 수행할 경우 전후공정은 보다 밀접하게 연결되고 또한 이에 수반하는 데이터는 상상을 초월하는 양이 되어 종전과 같은 손계산 또는 수작업 위주의 통계적방법의 사용은 그 활용가치를 잃게 된 데에서 기인[2]하는 필연적인 결과라고 하겠다.

국내 각 기업에서도 이와 같은 추세에 따라 전산부서의 보강과 함께, 각종 과학적관리기법의 전산화 및 전산 프로그램의 도입에 많은 노력을 경주하고 있다. 특히 컴퓨터 전문요원이나 SQC전문가만이 프로그램을 사용할 것이라는 전제 하에 중형 이상의 컴퓨터를 대상으로 했던 전산화습성에서 벗어나 기업 내의 전계층, 전부문에서 사용가능한 PC 위주의 프로그램 개발, 보급, 활용은 매우 바람직한 현상[6]이라고 할 수 있다. 그러나 이들 프로그램의 활용 계층 및 부문이 확대됨에 따라 기회 있을 때마다 거론해 온 계층기의 정확도 및 정밀도를 무시한 계층 데이터와 특히 이를 처리하는 과정에서의 유효숫자에 대한 무감각성은 그 심각함이 더욱 증대되고 있다고 하겠다.

여기에서는 "유효숫자자릿수 지정 및 소숫점아래 지정자릿수까지 반올림"과 "소숫점아래 지정자릿수까지 반올림"해 주는 subroutine을 이용하여, 지난번[7,8]에 발표한 반복있는 이원배치의 실험데이터에 대한 분산분석표작성과 이에 수반되는 추정외 프로그램에 이어 실험계획법에서 그 활용빈도가 가장 높은 직교배열표를 이용한 실험데이터의 분석을 위한 프로그래밍에 대해 다루었다. 유효숫자의 자릿수 처리에 대한 경각심을 불러일으켜 보고자 하는 점에 이 프로그래밍의 근본적인 의도가 있는 바, 이는 절대로 파소평가되어서는 안될 것이라고 생각한다.

* 인하대학교 산업공학과

접수 : 1993년 10월 22일

확정 : 1993년 10월 29일

II. 국내 프로그램의 문제점

국내에서 현재 개발 및 보급이 가장 많이 되고 있는 SQC 전산 프로그램은 다윈배치 실험데이터 처리를 위한 분산분석 및 구간추정용 프로그램이다. 이는 다윈배치 실험에 앞서, 실험인자의 선택을 올바르게 하기 위한 특성요인도의 작성 및 직교배열표를 이용한 실험의 실시를 무의식 중에 생략하게 하는 오류를 범하고 있음은 물론, 이들 전산 프로그램에는 지난번[7,8]에 거론한 바와 같이 다음과 같은 문제점들을 내포하고 있다.

- 1) 실험 데이터의 유효숫자 자릿수, 더 명확하게는 데이터의 최대치와 최소치의 차이 즉, 범위의 자릿수가 지나치게 적은 경우에도 이를 간과하고 분산분석을 수행함
- 2) 분산분석에서 소숫점 이하 특성의 고정 자릿수로 모든 계산을 처리하고 있는 경우가 많은 바, 이는 승제(\times 및 \div)계산에서 매우 불합리한 처리방법임
- 3) PC의 보통 정도 유효숫자 자릿수인 7자리의 숫자를 컴퓨터에 내장시킨 채 프린트 아웃 만들 특성의 소숫점 아래 자릿수 또는 유효숫자 자릿수로 처리하는 경우가 있음
- 4) 변동의 유효숫자 자릿수, 특히 오차변동의 유효숫자 자릿수가 원래의 실험데이터 자릿수보다 너무 많거나 적음
- 5) 이에 따라 분산 및 분산비의 유효숫자 자릿수에도 일정한 규칙성이 없음
- 6) 변동보다 분산의 유효숫자 자릿수가, 그리고 분산보다 분산비의 유효숫자 자릿수가 많아지고 있는 바, 이는 계산이 진행되어 나가면서 유효숫자의 자릿수가 감소될 수는 있으나 증가되어서는 안된다는 일반원칙을 무시하는 수치처리방법임
- 7) F분포표의 값(일반적으로 유효숫자 3자리)과 비교해야 할 분산비의 유효숫자 자릿수가 너무 많거나 적음
- 8) 일반적으로 F분포표의 값이 분산분석표에 제시되지 않아 전산처리 후에 다시 F분포표를 참조해야만 하는 번거로움이 있음
- 9) KS A 3251 "측정치의 처리 방법"을 준수하고 있지 않음
- 10) KS A 0021 "수치의 댄음법"을 준수하고 있지 않음

이상과 같은 수치의 불합리한 처리는 다윈배치 실험데이터의 분산분석에서 뿐만 아니라 직교배열표를 이용한 실험데이터의 분산분석, 더 나아가서는 여타 모든 통계데이터의 처리에서도 공통적으로 범하고 있으며, 이는 흔히 PC의 특성 때문에 어떻게 해 볼 도리가 없다고 변명하는 경우가 많으나, 전[6]에 제시한 "유효숫자자릿수 지정 및 소숫점아래 지정자릿수까지 반올림"해 주는 subroutine과 "소숫점아래 지정자릿수까지 반올림"해 주는 subroutine을 이용하면 의외로 쉽게 해결될 수 있다.

III. 프로그램 작성시의 유의사항

상기의 문제점들 중 가장 중요한 것은 변동, 분산 및 분산비의 유효숫자 자릿수이다. 이는 오차분산의 값이 0.501~1.49인 경우 유효숫자 1자리로 잡는 경우 1로 되어 분산비의 값에 있어 거의 3배에 달하는 차이가 날 수 있다는 점을 생각해 보면 그 심각성이 쉽게 이해될 수 있으며, 그 무엇보다도 최우선적으로 유의하여 반드시 고려해야 할 사항이라고 생각된다.

이들의 유효숫자 자릿수에 대해서는 여러가지 관점에서의 논란이 있겠으나 분산분석은 분산비와 F분포표의 값을 비교함으로써 그 결론이 맺어진다는 점과 F분포표의 값이 대개의 경우 유효숫자 3자리로 주어지고 있는 관행을 고려한다면 이들의 유효숫자 자릿수도 역시 3자리로 잡는 것이 가장 타당할 것이다. 그러므로 가장 기본이 되는 변동의 값들은 모두 유효숫자 3자리 이상으로 잡아야 할 것이나, 분산비의 계산 및 분산분석 후의 각종 추정에 오차분산이 이용된다는 점을 감안하여 오차변동의 값이 유효숫자 3자리로 계산되도록 하고, 이를 이용하여 계산되는 수치도 모두 유효숫자 3자리로 통일하는 방법이 가장 바람직할 것이다. 그리고 다음과 같은 점들도 역시 고려하는 것이 좋을 것이다.

- 실험 데이터의 유효숫자 자릿수(범위로 대변됨)가 너무 적으면 분산분석의 감도가 나빠지므로 이를 방지하기 위해 데이터의 범위는 유효숫자 2자리 이상이 되도록 한다.
- 사용 가능성이 있는 모든 자유도에 대해 F분포표의 값을 프로그램에 포함시킴으로써 분산분석의 F검정 및 분산분석후의 각종 추정이 별도의 F분포표나 t분포표를 참조하지 않고서도 프로그램에 의

```

1000 ' 2L8 : 'L8 - 2의 7승' ---- 직교배열표에 의한 실험결과데이터의 분석
1010 ' ===
1020 '
1030 '
1040 GOTO 1500 : '
1050 '
1060 ' *****
1070 ' Data Description
1080 ' *****
1090 ' ID$ ..... Problem ID ----- 영문 혹은 한글 --- 반각 40자 이내
1100 ' NK ..... Total # of Factors ( Main effects & Interactions ) ---
1110 ' ..... 6개 이내
1120 ' JK(K) ..... Column # at which a Factor is allocated ----- NK개
1130 ' JK$(K) ..... Factor ID's ----- NK개
1140 ' ..... Main effect --- A, B, D, F, G 등 ----- 반각 1자
1150 ' ..... Interaction --- AxB, AxD, DxF 등 ----- 반각 3자
1160 ' X(I) ..... Data ----- 8개
1170 ' ..... Measure Unit ----- 0.00000 0001 이상
1180 ' ..... ( Xmax - Xmin )의 값 ----- 유효숫자 2자리 이상
1190 '
1200 FOR IQAZ=1 TO JV:PRINT:NEXT IQAZ:RETURN : ' < 비디오 선택프 >
1210 FOR IQAZ=1 TO JP:LPRINT:NEXT IQAZ:RETURN : ' < 프린터 선택프 >
1220 FOR IUJ=1 TO 5555:NEXT IUJ: LOCATE 22,20:PRINT"다음 정보를 위해 <Space Bar>
1230 ' < 스페이스 바 >
1230 FOR IUJ=1 TO 5555:NEXT IUJ:YN$=INKEY$:IF YN$="Y" OR YN$="y" OR YN$="N" OR Y
1230 ' < y / n 확인 >
1230 ' < 윗첨자 시키기 >
1240 LPRINT CHR$(27);"SO";:RETURN : ' < 아랫첨자 시키기 >
1250 LPRINT CHR$(27);"SI";:RETURN : ' < 아랫첨자 해제 >
1260 LPRINT CHR$(27);"T";:RETURN : ' < 가로 확대 >
1270 LPRINT CHR$(27);CHR$(14);:RETURN : '
1280 IF YI01=0 AND YI02=0 THEN Y001=0:Y002=0:Y01$="#####":RETURN
1290 IF YI01<>0 THEN 1300 ELSE Y001=0:Y002=YI02:Y01$="":FOR Y003=1 TO 11-YI02:Y0
1300 ' $=Y01$+"#":NEXT Y003:Y01$=Y01$+"#":FOR Y003=1 TO YI02:Y01$=Y01$+"#":NEXT Y003:R
1300 ' ETURN
1300 IF YI01<0 THEN Y001=-1 ELSE Y001=1
1310 Y002=ABS(YI01):FOR IY03=20 TO 1 STEP -1:IF Y002>=10^(IY03-11) THEN 1320 ELS
1320 ' E NEXT IY03
-----< 중 략 >-----

2100 LCNT=0:FOR I=1 TO ICNT:IF F0(I)>=F15(IDF) OR SI(I)<0 THEN 2110 ELSE LCNT=LC
2110 ' NT+1:SI(7)=SI(7)+SI(I):IDF(7)=IDF(7)+1:SI(I)=-1:IDF(I)=0
2110 NEXT I:IF LCNT>0 THEN 2120 ELSE IF LPS<>1 THEN 2140 ELSE LPRINT CHR$(12):GO
2120 ' TO 2140
2120 FOR I=1 TO ICNT:VI(I)=SI(I):NEXT I:S1$=1SV$:YI01=SI(7)/IDF(7):YI02=3:GOSUB
1280:VI(7)=Y001:GOSUB 1220:CLS:LOCATE 1,21:PRINT"풀 립 된 본 산 본
2130 ' 섹 표":PRINT SPC(18);STRING$(44,"=")
2130 IF LPS<>1 THEN 1860 ELSE GOSUB 1210:LPRINT SPC(20):"풀 립 된 본
2140 ' 사 본 섹 표":LPRINT SPC(18);STRING$(44,"="):GOTO 1860
2140 LOCATE 22,31:PRINT"<<< Good Bye !!! >>>":END
2150 CLS:JV=3:GOSUB 1200:LIST 1070-1190
3000 ' ===== DATA =====
3010 DATA Running Example,5
3020 DATA 1,2,3,4,5
3030 DATA A,B,AxB,C,D
3040 DATA 25.3,24.8,21.3,22.3,25.9,21.9,20.4,16.9

```

그림 1. "2L8" 프로그램

해 처리되도록 한다.

- KS A 3251에 따라 추정치의 자릿수는 원데이터 보다 1자리만 더 잡는다.
- 구간추정의 경우, 신뢰구간의 상한 및 하한과 점추정치와의 차이가 같아지도록 하기 위해 추정정밀도를 필요한 자릿수로 반올림한 후에 점추정치에 가감한다.
- KS A 0021에 따라 반올림치리를 한다.

IV. 프로그램 및 Running Example

위에서 거론한 문제점들을 보완하고, 유의사항들을 염두에 두어 작성한 "직교배열표를 이용한 실험데이터의 분석"을 위한 프로그램들 중의 하나가 그림 1의 "2L8" 프로그램(지면절약을 위해 부분만을 수록하고 Flow Chart는 생략 하였음)이며, 그림 2는 직교표 $L_8(2^7)$ 의 1, 2, 3, 4, 5열에 A, B, AxB, C, D를 배열하여 25.3, 24.8, 21.3, 22.3, 25.9, 21.9, 20.4, 16.9의 실험데이터를 얻은 가상의 예(역시 지면절약을 위해 실험의 크기를 작게 하였음)에 대한 분석의 Output이다.

Running Example

분 산 분 석 표						
요 인	변 동	자 유 도	불 편 분 산	분 산 비	F분포의 값	
A	9.245	1	9.24	29.6*	18.5	98.5
B	36.125	1	36.1	116**	18.5	98.5
C	6.125	1	6.12	19.6*	18.5	98.5
D	8.000	1	8.00	25.6*	18.5	98.5
AxB	2.000	1	2.00	6.41	18.5	98.5
Error	0.625	2	0.312		{5%}	{1%}
Total	62.120	7				

풀 링 된 분 산 분 석 표						
요 인	변 동	자 유 도	불 편 분 산	분 산 비	F분포의 값	
A	9.245	1	9.24	10.6*	10.1	34.1
B	36.125	1	36.1	41.3**	10.1	34.1
C	6.125	1	6.12	6.99	10.1	34.1
D	8.000	1	8.00	9.14	10.1	34.1
Error	2.625	3	0.875		{5%}	{1%}
Total	62.120	7				

풀 링 된 분 산 분 석 표						
요 인	변 동	자 유 도	불 편 분 산	분 산 비	F분포의 값	
A	9.245	1	9.24	2.76	6.61	16.3
B	36.125	1	36.1	10.8*	6.61	16.3
Error	16.750	5	3.35		{5%}	{1%}
Total	62.120	7				

풀 링 된 분 산 분 석 표						
요 인	변 동	자 유 도	불 편 분 산	분 산 비	F분포의 값	
B	36.125	1	36.1	8.34*	5.99	13.7
Error	25.995	6	4.33		{5%}	{1%}
Total	62.120	7				

그림 2. Running Example

그림 2의 "분산분석표"에서 Error변동의 유효숫자 자릿수가 3자리이나 "풀링된 분산분석표"에서는 Error변동의 유효숫자 자릿수가 4 또는 5자리로 되어 있다. 이는 풀링의 개념상 별문제가 없을 것이다. 그림 2에서 특히 유의해야 할 사항은 풀링은 가능한 한 끝까지 시도하여 경우에 따라서는 모든 배치인자가 모두 다 풀링될 수도 있다는 것으로서, 이는 직교배열표를 이용하는 실험의 근본취지를 감안한 결과이다. 이 점에 대해서는 논란의 여지가 있겠으나 직교배열표를 이용한 실험은 어디까지나 예비적 실험으로서, 그 결과에 따라 고유기술 및 전문지식을 활용하여 더욱 상세한 수준의 본격적인 실험을 시행해야 한다는 점에서 쉽게 수궁이 될 것이며 아니더라도 풀링의 결과는 역시 참고자료가 될 것으로 생각한다.

V. 결 론

본고에서는 $L_6(2^7)$ 의 직교배열표를 이용한 실험데이터의 분석을 위한 프로그램만을 소개하였으나 이를 조금만 수정보완하면 여타의 2수준계 및 3수준계의 직교배열표를 이용한 실험데이터의 분석을 위한 프로그램 작성에 응용할 수 있을 것이다. 그리고 유효숫자 및 소숫점 아래 자릿수의 처리에 대해 앞에서 거론한 문제점들을 내포하고 있는 기존의 프로그램들에서도 본고에서 활용한 "유효숫자자릿수 지정 및 소숫점아래 지정자릿수까지 반올림"과 "소숫점아래 지정자릿수까지 반올림"해 주는 subroutine을 이용하면 역시 쉽게 그 문제점들이 제거될 수 있을 것이다.

국제시장에서의 가격경쟁력을 상실한 우리 기업이 해결해야 할 최우선의 과제는 품질경쟁력의 조속한 확보이며, 또한 이를 위해서는 불량율의 감소라는 소극적인 활동이 아니라 제품의 성능이나 기능 또는 디자인의 향상이라는 보다 적극적인 활동이 되어야 한다는 점에 대해서는 결코 이론의 여지가 없을 것이다. 최근 많이 거론되고 있는 ISO 9000 Series는 이 점에 있어서 큰 도움이 될 것이나, 표준류의 주목적이 "수준의 확보 및 유지"에 있다는 것을 감안하여 "수준의 탈피 및 향상"을 도모하기 위한 활동을 전개할 때 본고에서 거론된 내용은 적지 않은 도움이 될 것이며, 더 나아가서는 우리의 품질경쟁력향상을 위한 하나의 보조수단으로서 사용될 수 있다는 점에서도 그 의의는 재음미되어야만 할 것이다.

본고에서 제시한 전산프로그램은 결코 완벽한 것이라고는 할 수 없으며 단지 앞에서 거론한 문제점들을 개선하기 위한 하나의 시안으로서 제시한다는 점과, 이에 관심있는 이들의 토론의 장을 마련하는 계기가 되었으면 하는데에 그 근본취지가 있음을 재삼 강조하셔서 본 연구는 인하대학교부설 산업과학기술연구소의 지원하에 이루어졌음을 밝히고 아울러 관심있는 이들의 기탄없는 비판을 부탁드린다.

참 고 문 헌

- [1] 박성현, 현대실험계획법, 대영사, 1983
- [2] 서문원, Application of Statistical Methods and Current Practices of Statistical Quality Control among U.S. Industries —— New Theories and Trends for Renewed Objectives, '83 국내외 한국과학기술자학술회의 논문집, 하계심포지움편, 한국과학기술단체총연합회, pp.291-9, 1983
- [3] 송서일, 실험계획법, 영지문화사, 1980
- [4] 조남호, 김복만, 현대실험계획법, 창지사, 1982
- [5] 이상용, 공장실험계획법, 형설출판사, 1987
- [6] 정수일, 품질관리기법의 전산화에 관한 연구(I), 인하대학교 산업과학기술연구소 논문집, 제12집, pp. 52-61, 1984
- [7] 정수일, 실험계획법의 전산화에 관한 연구(I), 한국품질관리학회지, Vol.16, No.1, pp.23-31, 1988
- [8] 정수일, 실험계획법의 전산화에 관한 연구(II), 한국품질관리학회지, Vol.19, No.1, pp.163-9, 1991
- [9] 황의철, 최신품질관리, 박영사, 1977