The Korean Communications
in Statistics Vol. 12 No. 1, 2005
pp. 47-59

EMS Rules for Balanced Factorial Designs under No Restriction on Interaction

Byoung-Chul Choi[1)]

Abstract

Expected mean square(EMS) is an important part of conducting the analysis of variance in the experimental design problem, especially in mixed or random models. We present a set of EMS rules for balanced factorial designs under no restriction on interaction. Also we point out how to use the variance component of SPSS or SAS procedure to obtain EMS.

Keywords : EMS, factorial designs, mixed models. random models, variance component

1. 서론

어떤 화학제품의 생산 공정에서 인자 A와 B에 따른 제품의 수율(%)을 조사한 결과 <표1>과 같았다고 하자.

<표1> 인자 A와 B에 따른 제품의 수율

B	A		
	1	2	3
1	88 89	87 90	92 93
2	90 92	91 93	94 96
3	89 90	95 96	90 91

<표1>은 이원배치법의 실험자료인데 인자 A와 B가 모두 모수인자인 모수모형의 경우와, 인자 A는 모수인자이고 인자 B는 변량인자인 혼합모형, 그리고 인자 A와 B가 모두 변량인자인 변량모형의 경우 국내에서 널리 사용하고 있는 SPSS 패키지 한글판 버전10.1을 사용하여 분석하면 다음 <표2>와 같다.

1) Professor, division of Mathematics and Statistical Informatics, Chonbuk National University, Chonju, Chonbuk, 560-756, Korea.
E-mail : chbch@chonbuk.ac.kr

<표2> 이원배치법의 분산분석표

모형	요인		제곱합	자유도	평균제곱	F	유의확률
모수모형	A		25.444	2	12.722	8.808	.008
	B		29.778	2	14.889	10.308	.005
	A * B		50.222	4	12.556	8.692	.004
	오차		13.000	9	1.444		
	수정 합계		118.444	17			
혼합모형 또는 변량모형	A	가설	25.444	2	12.722	1.013	.441
		오차	50.222	4	12.556		
	B	가설	29.778	2	14.889	1.186	.394
		오차	50.222	4	12.556		
	A * B	가설	50.222	4	12.556	8.692	.004
		오차	13.000	9	1.444		

<표2>를 살펴보면 F검정통계량과 유의확률이 모수모형인 경우와 모수모형이 아닌 경우는 많은 차이가 있어 분석하는 과정에 주의해야 함을 알 수 있다. <표2>의 결과를 해석하기 위해 이원배치법 자료의 모형과 가설에 대해 알아보자. 모수인자 A의 i번째 수준과 변량인자 B의 j번째 수준에서의 k번째 반복 실험자료를 확률변수 y_{ijk}로 나타내면, 자료의 혼합모형은 다음과 같다.

$$y_{ijk} = \mu + a_i + b_j + (ab)_{ij} + e_{ijk} \quad \begin{cases} i = 1, 2, \cdots, l \\ j = 1, 2, \cdots, m \\ k = 1, 2, \cdots, r \end{cases}$$

여기서 μ는 총평균, a_i는 인자 A의 i번째 수준의 효과, b_j는 인자 B의 j번째 수준의 효과, $(ab)_{ij}$는 a_i와 b_j의 교호작용효과를 나타낸다. 또 A는 모수인자이므로 제한조건 $\sum_i a_i = 0$을 가정하지만 B는 변량인자라서 $\sum_i b_j = 0$, $\sum_i (ab)_{ij} = 0$, $\sum_j (ab)_{ij} = 0$ 등의 제한조건을 가정하지 않고, 대신 $E(b_j) = 0$, $E[(ab)_{ij}] = 0$, $b_j \sim N(0, \sigma_B^2)$, $(ab)_{ij} \sim N(0, \sigma_{A\times B}^2)$, $e_{ijk} \sim N(0, \sigma_E^2)$이고 서로 독립임을 가정한다. 이 가정으로부터 분산 σ_B^2과 $\sigma_{A\times B}^2$은

$$\sigma_B^2 = E(b_j^2), \qquad \sigma_{A\times B}^2 = E[(ab)_{ij}^2]$$

과 같이 정의된다.

모수인자 A의 수준별 평균이 같은가를 검정하기 위한 가설은

$$H_0: a_1 = a_2 = \cdots = a_l = 0 \quad vs \quad H_1: H_0\text{가 아니다}$$

또는

$$H_0: Q(A) = 0 \quad vs \quad H_1: Q(A) \neq 0$$

이다. 여기서 $Q(A)$는 인자 A의 모수효과(fixed effect)로

$$Q(A)=\frac{\sum_i a_i^2}{l-1}$$

이며, 일반적으로 수준수가 l, m인 두 모수인자 A와 B의 교호작용의 모수효과는

$$Q(A\times B)=\frac{\sum_i\sum_j(ab)_{ij}^2}{(l-1)(m-1)}$$

과 같이 나타낸다. 변량인자 B의 산포가 균질한가를 검정하기 위한 가설은

$$H_0:\ \sigma_B^2=0 \quad vs \quad H_1:\ \sigma_B^2\neq 0$$

이며, 교호작용효과가 있느냐를 검정하기 위한 가설은

$$H_0:\ \sigma_{A\times B}^2=0 \quad vs \quad H_1:\ \sigma_{A\times B}^2\neq 0$$

와 같다.

<표2>에서 모수모형이 아닌 경우 인자 A, B와 교호작용 $A\times B$의 F검정통계량은 차례로

$$F=\frac{MS_A}{MS_{A\times B}},\quad F=\frac{MS_B}{MS_{A\times B}},\quad F=\frac{MS_{A\times B}}{MSE}$$

와 같았다. 여기서 MS_A, MS_B, $MS_{A\times B}$는 차례로 A, B, $A\times B$의 평균제곱을, 그리고 MSE는 평균제곱오차를 나타낸다. F검정통계량들이 위와 같은 이유는 평균제곱의 기대값(expected mean square; EMS)들이 각각 다음과 같기 때문이다.

$$\begin{aligned}
E(MS_A)&=\sigma_E^2+r\sigma_{A\times B}^2+mrQ(A),\\
E(MS_B)&=\sigma_E^2+r\sigma_{A\times B}^2+lr\sigma_B^2,\\
E(MS_{A\times B})&=\sigma_E^2+r\sigma_{A\times B}^2,\\
E(MSE)&=\sigma_E^2.
\end{aligned}$$

또 다른 통계 패키지 SAS를 이용하여 분산분석을 할 경우, 모수모형의 분석과 해석은 SPSS와 마찬가지로 어려움이 없으나, 모수모형이 아닌 경우는 F검정통계량을 어떻게 구할지 사용자가 TEST 문장을 별도로 써서 지정을 해주어야 하기 때문에 전문가가 아니고는 분석하기가 쉽지 않은 점이 있다. SPSS 패키지에서는 모수모형이 아닌 경우라도 분석 과정에서 인자가 모수인자인지

변량인자인지를 구분해 주기만 하면 F검정통계량과 유의확률이 계산되기 때문에 전문가가 아닌 일반인들도 쉽게 사용할 수 있는 장점이 있다. 그래서 본 논문에서는 SPSS 패키지에서 F검정통계량을 구하기 위한 평균제곱의 기대값을 어떻게 계산하는지 또 오류는 없는지 검토하고 평균제곱의 기대값을 구하는 일반적인 법칙을 소개하고자 한다.

2 평균제곱의 기대값을 구하는 법칙

앞 절에서 평균제곱의 기대값이 분산분석에서 핵심적인 역할을 하는 것을 살펴보았는데, 인자의 수가 많은 경우에는 변량모형과 혼합모형의 평균제곱의 기대값을 이론적으로 유도하는 과정이 굉장히 복잡하다. 그래서 기계적으로 그 기대값을 구하는 방법을 반복이 있는 다음 3종류의 이원배치법 모형의 예를 들어 알아보자. 모형에서 인자 A와 B의 수준수는 각각 l, m이고, 반복수는 r이라 한다.

모형식 : $y_{ijk} = \mu + a_i + b_j + (ab)_{ij} + e_{ijk}$

모수모형 : $a_i, b_j, (ab)_{ij}$: 미지의 상수, $e_{ijk} \sim N(0, \sigma_E^2)$이고 서로 독립

혼합모형 : a_i : 미지의 상수
$b_j \sim N(0, \sigma_B^2)$, $(ab)_{ij} \sim N(0, \sigma_{A\times B}^2)$, $e_{ijk} \sim N(0, \sigma_E^2)$이고 서로 독립

변량모형 : $a_i \sim N(0, \sigma_A^2)$, $b_j \sim N(0, \sigma_B^2)$, $(ab)_{ij} \sim N(0, \sigma_{A\times B}^2)$, $e_{ijk} \sim N(0, \sigma_E^2)$
이고 서로 독립

혼합모형의 경우 교호작용을 나타내는 항 $(ab)_{ij}$에 대해

$$\sum_i (ab)_{ij} = \overline{(ab)}_{.j} = 0$$

이라는 제한을 두는 경우도 있는데, 제한을 두는 경우에 평균제곱의 기대값을 구하는 법칙은 박성현(2003), Montgomery(2001), Hicks(1999) 등에 자세히 소개되어 있다. 제한을 두는 것이 좋은지 안 두는 것이 좋은지는 아직 연구되지 않은 열린 문제이다. Hartley와 Searle(1969)은 $(ab)_{ij}$에 제한을 두지 않는 것이 셀 반복수가 같지 않은 경우에도 평균제곱의 기대값에 일관성이 있다고 지적했다. 제한을 두지 않는 경우 Searle(1971)은 일반적인 실험계획에서 적용할 수 있는 평균제곱의 기대값을 구하는 법칙을 소개하고 있는데 난해하고 박성현(2003) 등의 경우처럼 자세하지 않아, 요인실험에서 평균제곱의 기대값을 구하는 법칙을 본 논문에서 이용하기 쉽게 자세히 소개하려고 한다. 참고로, 널리 사용되고 있는 통계 패키지 SPSS(1999)와 SAS(1988)에서도 제한을 두지 않는 경우의 평균제곱의 기대값을 제공하고 있다.

법칙 1. 교호작용이 변량인자를 하나라도 포함하고 있으면 변량인자로 간주한다. 인자 A가 모수인자이고 인자 B가 변량인자이면 인자 A의 모수효과를 $Q(A)$, 인자 B의 분산은 σ^2_B, 교호작용 $A\times B$의 분산은 $\sigma^2_{A\times B}$이라 놓는다.

법칙 2. 다음과 같은 표를 만든다. 첫 번째 칸에 모형에 있는 각 항의 효과를 기입한다. 첫 번째 줄에는 첫 번째 칸에 나온 오차항을 제외한 나머지 효과를 차례로 기입한 후 각각의 분산과 수준수를 기입하되, 교호작용의 경우에는 해당인자들의 수준수의 곱을 기입한다.

요인	A σ^2_A l	B σ^2_B m	$A\times B$ $\sigma^2_{A\times B}$ lm	EMS
절편				
A				
B				
$A\times B$				
E				

법칙 3. 절편 줄의 각 셀에 자료의 총수 lmr을 첫 번째 줄에 있는 수준수로 나누어 기입한다. 다음으로, 같은 인자가 하나이상 만나는 셀에 그 셀의 절편 줄에 있는 숫자를 기입한다.

요인	A σ^2_A l	B σ^2_B m	$A\times B$ $\sigma^2_{A\times B}$ lm	EMS
절편	mr	lr	r	
A	mr		r	
B		lr	r	
$A\times B$			r	
E				

법칙 4. 평균제곱의 기대값

(1) 변량모형의 경우

줄별로 셀의 숫자와 그 셀의 칸에 해당하는 효과의 분산을 곱하여 모두 합한 후 추가로 σ^2_E을 합해서 $E(MS)$ 칸에 기입한다.

요인	A σ_A^2 l	B σ_B^2 m	$A\times B$ $\sigma_{A\times B}^2$ lm	EMS
절편	mr	lr	r	
A	mr		r	$\sigma_E^2+r\sigma_{A\times B}^2+mr\sigma_A^2$
B		lr	r	$\sigma_E^2+r\sigma_{A\times B}^2+lr\sigma_B^2$
$A\times B$			r	$\sigma_E^2+r\sigma_{A\times B}^2$
E				σ_E^2

(2) 혼합모형(A 모수, B 변량)의 경우

모수인자나 모수인자들만의 교호작용은 그 분산성분을 모수효과 $Q(\cdot)$로 바꾼다. 나머지 효과의 $E(MS)$는 변량모형과 동일하다.

요인	A $Q(A)$ l	B σ_B^2 m	$A\times B$ $\sigma_{A\times B}^2$ lm	EMS
절편	mr	lr	r	
A	mr		r	$\sigma_E^2+r\sigma_{A\times B}^2+mrQ(A)$
B		lr	r	$\sigma_E^2+r\sigma_{A\times B}^2+lr\sigma_B^2$
$A\times B$			r	$\sigma_E^2+r\sigma_{A\times B}^2$
E				σ_E^2

(3) 모수모형의 경우

모든 효과의 분산성분을 모수효과 $Q(\cdot)$로 바꾸고 A와 B의 줄과 $A\times B$칸이 만나는 셀에서 r을 삭제한다.

요인	A $Q(A)$ l	B $Q(B)$ m	$A\times B$ $Q(A\times B)$ lm	EMS
절편	mr	lr	r	
A	mr			$\sigma_E^2+mrQ(A)$
B		lr		$\sigma_E^2+lrQ(B)$
$A\times B$			r	$\sigma_E^2+rQ(A\times B)$
E				σ_E^2

모수모형의 경우에는, 인자의 수에 관계없이 효과 X의 평균제곱의 기대값은 모두

$$\sigma_E^2+urQ(X)$$

의 형태가 된다. 여기서 r은 반복수이고, u는 효과 X에 들어 있는 인자를 제외한 나머지 인자들의 수준수의 곱이다. 예를 들어, 인자가 A, B, C, D, E, F인 6원배치법에서 각각의 수준수를 a, b, c, d, e, f라 하면 교호작용 $A\times C\times E\times F$의 평균제곱의 기대값은

$$\sigma_E^2 + bdrQ(A\times C\times E\times F)$$

와 같다.

법칙 5. 인자의 수가 세 개 이상인 혼합모형의 경우 X를 모수인자들만의 교호작용효과라 하고 Y를 X에 들어있는 인자들로 이루어진 효과나 인자라고 하면 $E(MS_X)$에는 $Q(X)$항이 포함되지만, $E(MS_Y)$에는 포함되지 않는다. 예를 들어, 인자가 A, B, C, D, E, F인 6원배치법에서 A, B, C가 모수인자이고 D, E, F가 변량인자인 경우 $E(MS_{A\times B\times C})$에는 $Q(A\times B\times C)$항이 포함되지만, $E(MS_A)$, $E(MS_B)$, $E(MS_C)$, $E(MS_{A\times B})$, $E(MS_{A\times C})$, $E(MS_{B\times C})$에는 포함되지 않는다. 같은 방법으로 $E(MS_{A\times B})$에는 $Q(A\times B)$항이 포함되지만, $E(MS_A)$, $E(MS_B)$에는 포함되지 않는다. 이 법칙의 이론적 근거는 부록을 참고하라.

법칙 6. 반복이 없는 경우($r=1$) 반복이 없는 경우에는 $r=1$로 놓고, 최고차 교호작용이 나타나는 줄과 칸을 없애고 EMS 칸에서 최고차 교호작용의 분산성분이나 모수효과 항을 제거한다. 예를 들어, A가 모수인자이고 B가 변량인자인 혼합모형의 경우 다음과 같다.

요인	A $Q(A)$ l	B σ_B^2 m	EMS
절편	m	l	
A	m		$\sigma_E^2 + mQ(A)$
B		l	$\sigma_E^2 + l\sigma_B^2$
E			σ_E^2

법칙 7. F검정통계량의 계산 모수모형의 경우는, 효과 X에 대한 검정통계량은 모두

$$F = MS_X / MSE$$

의 형태가 된다. 이는 인자의 수에 관계없이 효과 X의 평균제곱의 기대값 $E(MS_X)$가 법칙4(3)에 의해 $E(MS_X) = \sigma_E^2 + urQ(X)$이고 $E(MSE) = \sigma_E^2$이며, 그 비가

$$\frac{E(MS_X)}{E(MSE)} = \frac{\sigma_E^2 + urQ(X)}{\sigma_E^2} = 1 + ur\frac{Q(X)}{\sigma_E^2}$$

이기 때문이다. 모수모형이 아닌 경우에는 $E(MS)$의 적절한 비를 찾아서 검정통계량을 구해야 한다. 예를 들어 인자 A와 B가 모두 변량인자이고 각각의 수준수가 l, m, 반복수가 r인 2원배치법 변량모형의 경우 EMS와 검정통계량을 구하면 다음 <표3>과 같다.

<표3> 2인자 변량모형의 EMS와 검정통계량

요인	평균제곱	EMS	F
A	MS_A	$\sigma_E^2+r\sigma_{A\times B}^2+mr\sigma_A^2$	$MS_A/MS_{A\times B}$
B	MS_B	$\sigma_E^2+r\sigma_{A\times B}^2+lr\sigma_B^2$	$MS_B/MS_{A\times B}$
$A\times B$	$MS_{A\times B}$	$\sigma_E^2+r\sigma_{A\times B}^2$	$MS_{A\times B}/MSE$
E	MSE	σ_E^2	

반복이 없는 경우($r=1$)에는 <표3>에서 교호작용 $A\times B$가 오차항에 교락되어 구분이 안 되므로 교호작용 $A\times B$의 줄이 다음 표와 같이 오차항으로 합산된다.

요인	평균제곱	EMS	F
A	MS_A	$\sigma_E^{2\prime}+\sigma_{A\times B}^2+m\sigma_A^2$	$MS_A/(MSE+MS_{A\times B})$
B	MS_B	$\sigma_E^{2\prime}+\sigma_{A\times B}^2+l\sigma_B^2$	$MS_B/(MSE+MS_{A\times B})$
$E+(A\times B)$	$MSE+MS_{A\times B}$	$\sigma_E^{2\prime}+\sigma_{A\times B}^2$	

그런데 교호작용을 무시할 수 있다면 표에서 $\sigma_{A\times B}^2=0$으로 $MS_{A\times B}=0$이 되어, 인자 A와 B의 검출을 위한 검정통계량이 모두 모수모형과 같은 형태가 된다. 따라서 교호작용이 있으리라고 의심된다면 반드시 반복실험을 하여 <표3>과 같이 분석을 해야 한다.

3. SPSS 분산성분표를 이용한 EMS의 계산

인자 A와 B는 모수인자이고 인자 C는 변량인자며 각각의 수준수가 l, m, n(3, 2, 3)인 반복이 없는 3인자 혼합모형

$$y_{ijk}=\mu+a_i+b_j+c_k+(ab)_{ij}+(ac)_{ik}+(bc)_{jk}+e_{ijk}$$

의 자료가 다음 <표4>와 같이 주어진 경우를 예로 들어 SPSS 분산분석 절차에서 자동으로 출력되는 분산성분표에서 평균제곱의 기대값을 구하는 방법을 알아보자.

<표4> 3인자 혼합모형 자료

A		1		2		3	
B		1	2	1	2	1	2
C	1	34	25	27	18	19	9
	2	27	19	20	13	14	8
	3	23	18	17	12	12	6

<표4>와 같은 3인자 혼합모형의 자료를 SPSS 절차에 따라 실행하면 분산분석표에 이어 <표5>와 같은 분산성분표가 자동으로 출력된다.

<표5> 분산성분표

소스	분산성분				
	Var(C)	Var($A*C$)	Var($B*C$)	Var(오차)	2차항
A		2(m)		1	A, $A*B$
B			3(l)	1	B, $A*B$
C	6(lm)	2(m)	3(l)	1	
$A*B$				1	$A*B$
$A*C$		2(m)		1	
$B*C$			3(l)	1	
오차				1	

<표5>의 SPSS 분산성분표에서

"각 소스에 대한 평균제곱의 기대값은 셀의 계수와 분산성분을 곱한 값의 합에 2차항 셀 내의 효과와 관련된 2차항을 더한 값과 같습니다"

라는 안내문이 있는데 이에 따라 $E(MS_A)$와 $E(MS_B)$를 구하면

$$E(MS_A) = \sigma_E^2 + 2\sigma_{A\times C}^2 + 6Q(A) + 3Q(A\times B)$$
$$E(MS_B) = \sigma_E^2 + 3l\sigma_{B\times C}^2 + 9Q(B) + 3Q(A\times B)$$

가 된다. 여기서, <표5>에는 $A\times B$의 계수가 나와 있지 않지만 소개한 평균제곱의 기대값을 구하는 법칙에 의해 $n=3$이 된 것이다. 그런데 이 평균제곱의 기대값은 잘 못 계산된 것이며 올바로 계산하면 다음 <표6>과 같고, 그 이유는 부록에 별도로 기술하였다. 이런 SPSS 계산은 분산 σ_C^2, $\sigma_{A\times C}^2$, $\sigma_{B\times C}^2$, σ_E^2의 추정을 올바로 하지 못하게 하는 오류를 유발한다.

<표6> 평균제곱의 기대값, 3인자 혼합모형

요인	EMS	F
A	$\sigma_E^2+2\sigma_{A\times C}^2+6Q(A)$	$MS_A/MS_{A\times C}$
B	$\sigma_E^2+3\sigma_{B\times C}^2+9Q(B)$	$MS_B/MS_{B\times C}$
C	$\sigma_E^2+2\sigma_{A\times C}^2+3\sigma_{B\times C}^2+6\sigma_C^2$	$MS_C/(MS_{A\times C}+MS_{B\times C}-MSE)$
$A*B$	$\sigma_E^2+3Q(A\times B)$	$MS_{A\times B}/MSE$
$A*C$	$\sigma_E^2+2\sigma_{A\times C}^2$	$MS_{A\times C}/MSE$
$B*C$	$\sigma_E^2+3\sigma_{B\times C}^2$	$MS_{B\times C}/MSE$
오차	σ_E^2	

<표6>은 소개한 평균제곱의 기대값을 구하는 법칙에 따라 구한 것인데 A, B가 모수인자인 경우 SPSS의 출력과는 달리 $E(MS_{A\times B})$에는 $Q(A\times B)$가 포함되지만 $E(MS_A)$, $E(MS_B)$에는 포함되지 않았다. 이는 평균제곱의 기대값을 구하는 법칙5에 의한 것이다. SPSS 분산성분표에는 이런 언급이 없어 주의가 요한다. 그러나 F검정통계량은 모두 올바로 계산되기 때문에 분산분석표를 읽고 해석하는 데는 문제가 없다. 특히 인자 C의 효과에 대한 근사 F검정통계량을 <표6>과 같이 정확히 제공하고 있어 사용자가 이용하기에 편리한 점이 있다.

SAS 절차를 이용하여 분산성분을 계산해도 SPSS 절차에서와 비슷한 결과가 출력된다. 다음은 <표4>의 자료를 이용하여 평균제곱의 기대값을 구하기 위한 SAS 프로그램과 그 출력의 일부분이다.

[SAS 프로그램1]

```
PROC GLM;
  CLASS   a  b  c;
  MODEL  yield = a  b  c  a*b  a*c b*c;
  RANDOM  c  a*c  b*c;
RUN;
```

[프로그램1의 출력]

```
Source        Type III Expected Mean Square
  a           Var(Error) + 2 Var(a*c) + Q(a,a*b)
  b           Var(Error) + 3 Var(b*c) + Q(b,a*b)
  c           Var(Error) + 3 Var(b*c) + 2 Var(a*c) + 6 Var(c)
  a*b         Var(Error) + Q(a*b)
  a*c         Var(Error) + 2 Var(a*c)
  b*c         Var(Error) + 3 Var(b*c)
```

위와 같은 SAS 실행 결과에서는 특히 모수효과의 계산에 대한 언급이 없어 해석하기가 어렵다. 또한 프로그램의 RANDOM 문장에서 인자와 교호작용을 올바로 선택해야 하는 부담도 있다. SAS 절차에서 분산성분을 추정할 수 있는 또 다른 방법은 PROC VARCOMP 절차인데 그 프로

그램과 출력의 일부분은 다음과 같다.

[SAS 프로그램2]

```
PROC  VARCOMP  method=type1;
  CLASS  a  b  c;
  MODEL  yield = a  b  c  a*b  a*c b*c/fixed=2;
RUN;
```

[프로그램2의 출력]

Source	Mean Square	Expected Mean Square
a	253.555556	Var(Error) + 2 Var(a*c) + 3 Var(a*b) + Q(a)
b	227.555556	Var(Error) + 3 Var(b*c) + 3 Var(a*b) + Q(b)
c	89.388889	Var(Error) + 3 Var(b*c) + 2 Var(a*c) + 6 Var(c)
a*b	0.222222	Var(Error) + 3 Var(a*b)
a*c	2.972222	Var(Error) + 2 Var(a*c)
b*c	5.055556	Var(Error) + 3 Var(b*c)
Error	0.472222	Var(Error)

Variance component	Estimate
Var(c)	13.63889
Var(a*b)	-0.08333
Var(a*c)	1.25000
Var(b*c)	1.52778
Var(Error)	0.47222

위 출력에서도 인자 A, B에 모수효과 $Q(A\times B)$가 Var(a*b)로 나타나는 오류는 계속되며 Var(a*b)의 추정치가 -0.08333으로 음수인 점이 특이하다. 결국 SAS의 EMS 출력을 보고 F검정 통계량을 구하기 위한 TEST 문장을 작성해야 하는데 위 두 가지의 EMS 출력을 보고는 어떻게 TEST 문장을 작성해야 할지 난감해진다. TEST 문장 작성요령은 성내경(1989)에 일부 소개되어 있다.

분산 σ_C^2, $\sigma_{A\times C}^2$, $\sigma_{B\times C}^2$, σ_E^2의 추정값은 SPSS 절차에서는 제공하지 않고 있고, SAS 절차에서는 위 출력과 같이 제공하는데 Var(a*b)의 추정치가 -0.08333으로 음수인 점을 제외하고는 분산성분의 추정에 문제는 없다.

4. 결론

본 논문에서는 사용빈도가 높은 통계패키지인 SPSS와 SAS에서 F검정통계량을 구하기 위한 평균제곱의 기대값을 어떻게 계산하는지 또 오류는 없는지 검토하고 평균제곱의 기대값을 구하는

일반적인 법칙을 소개하였다. 여기서 소개한 법칙은 혼합모형의 교호작용을 나타내는 항에 대한 제한을 두지 않은 경우로 SPSS와 SAS에서도 제한을 두지 않는 경우의 평균제곱의 기대값을 제공하고 있다.

결론적으로, SPSS 나 SAS 절차를 이용하여 평균제곱의 기대값을 쉽게 구하려면 분산성분의 출력에 EMS 구하는 법칙5를 적용하여 포함되지 않을 교호작용의 모수효과를 제거하면 되겠다.

부록 법칙 5의 근거

인자 A와 B가 모수인자, 인자 C가 변량인자이고 각각의 수준수가 l, m, n인 반복이 없는 3원배치법 혼합모형의 예를 들어 $E(MS_A)$에는 $Q(A \times B)$가 포함되지 않음을 밝혀보자. 3원배치법 혼합모형

$$y_{ijk} = \mu + a_i + b_j + c_k + (ab)_{ij} + (ac)_{ik} + (bc)_{jk} + e_{ijk}$$

에서 A와 B가 모수인자이므로 제한조건 $\sum_i a_i = 0, \sum_i b_j = 0,\ \sum_i (ab)_{ij} = 0,\ \sum_j (ab)_{ij} = 0$이 만족되어

$$\begin{aligned}
\bar{y}_{ij.} &= \mu + a_i + b_j + \bar{c} + (ab)_{ij} + (\overline{ac})_{i.} + (\overline{bc})_{j.} + \bar{e}_{ij.} \\
\bar{y}_{i..} &= \mu + a_i \qquad + \bar{c} \qquad\quad + (\overline{ac})_{i.} + (\overline{bc}) + \bar{e}_{i..} \\
\bar{\bar{y}} &= \mu \qquad\qquad + \bar{c} \qquad\quad + (\overline{ac}) + (\overline{bc}) + \bar{\bar{e}}
\end{aligned}$$

가 성립한다. 따라서

$$\begin{aligned}
E\{SS_A\} &= E\Big\{\sum_i\sum_j\sum_k (\bar{y}_{i..} - \bar{\bar{y}})^2\Big\} \\
&= \sum_i\sum_j\sum_k E\{(a_i + [(\overline{ac})_{i.} - (\overline{ac})] + [\bar{e}_{i..} - \bar{\bar{e}}])^2\}
\end{aligned}$$

인데 교차곱의 항이 모두 0이 되므로

$$E\{SS_A\} = \sum_i\sum_j\sum_k a_i^2 + E\Big\{\sum_i\sum_j\sum_k [(\overline{ac})_{i.} - (\overline{ac})]^2\Big\} + E\Big\{\sum_i\sum_j\sum_k (\bar{e}_{i..} - \bar{\bar{e}})^2\Big\}$$

이다. 그런데 위 식의 각 항은

$$\sum_i\sum_j\sum_k a_i^2 = (l-1)\sum_j\sum_k\left\{\frac{\sum_i a_i^2}{l-1}\right\} = (l-1)\,mnQ(A)$$

$$E\left\{\sum_i\sum_j\sum_k[(\overline{ac})_{i.}-(\overline{ac})]^2\right\}=(l-1)\sum_j\sum_k E\left\{\frac{\sum_i[(\overline{ac})_{i.}-(\overline{ac})]^2}{l-1}\right\}$$

$$=(l-1)mn\frac{\sigma^2_{A\times C}}{n}=(l-1)m\,\sigma^2_{A\times C}$$

$$E\left\{\sum_i\sum_j\sum_k(\overline{e}_{i..}-\overline{\overline{e}})^2\right\}=(l-1)\sum_j\sum_k E\left\{\frac{\sum_i(\overline{e}_{i..}-\overline{\overline{e}})^2}{l-1}\right\}$$

$$=(l-1)mn\frac{\sigma^2_E}{mn}=(l-1)\,\sigma^2_E$$

이므로

$$E(MS_A)=\frac{E(SS_A)}{l-1}$$

$$=\frac{1}{l-1}\{(l-1)mnQ(A)+(l-1)m\,\sigma^2_{A\times C}+(l-1)\,\sigma^2_E\}$$

$$=mnQ(A)+m\sigma^2_{A\times C}+\sigma^2_E$$

이 되어 $E(MS_A)$에 $Q(A\times B)$가 포함되지 않음을 알 수 있다.

참고문헌

[1] 박성현 (2003). 「현대실험계획법」, 민영사, 서울.

[2] 성내경(1989). 「SAS/STAT-분산분석」, 자유아카데미, 서울.

[3] Hartley, H.O. and Searle, S.R. (1969). On Interaction Variance Components in Mixed Models. *Biometrics*, Vol. 25, 573-576.

[4] Hicks, C.R. and Turner, Jr, K.V. (1999). *Fundamental Concepts in the Design of Experiments,* 5^{th} *ed,* Oxford University Press, New York.

[5] Montgomery, D.C. (2001). *Design and Analysis of Experiments,* 5^{th} *ed,* John Willy & sons Inc, USA.

[6] SAS (1988). *SAS/STAT User's Guide, Release 6.03 Ed,* SAS Institute Inc, USA.

[7] Searle, S.R. (1971). *Linear Models*. John Willy & sons Inc, USA.

[8] SPSS (1999). *SPSS Base 10.0 Applications Guide,* SPSS Inc, USA.

[2004년 7월 접수, 2004년 12월 채택]