7

제곱합과 자유도, 그리고 기대평균제곱

7.1 제곱합과 자유도의 분할

실험설계에서 가장 중요한 고려 사항은 물론 분산분석의 수행이라 할 수 있다. 분산분석을 하기 위해서는 실험설계에 대응되는 모형을 구성하는 각 효과들에 대한 제곱합과 그 자유도를 결정해야 한다. 그 다음 적절한 검증법을 구축하려면 기대평균제곱의 유도가 필수적이다. 특히 랜덤 모형이나 혼합 모형에서 이러한 제반 과정의 공식적인 절차 확립은 매우 중요하다.

따라서 이 절에서는 주어진 모형에서 제곱합과 자유도를 분할하는 공식적인 방법을 소개하고, 제 7.2 절에서는 기대평균제곱을 유도하는 방법을 설명한다. 이 방식들은 균형된 요인 실험(factorial experiment), 지분 실험(nested experiment), 또는 지분 요

인 실험(nested factorial experiment)에서 사용할 수 있다.

이 장에서 소개하는 EMS 규칙은 라틴 정방이나 불완전 블록 설계와 같이 부분적으로 균형된 실험에는 적용할 수 없다.

제곱합의 분할시 모형에 관계없이 공통적으로 준수해야 하는 몇 가지 규칙이 있다.

- 오차항에서 수준 조합 내 반복 실험을 나타내는 첨자는 다른 모든 첨자에 지분된다.

예를 들어 반복이 2 이상인 이요인 요인 설계 모형에서 지금까지 오차항은 ϵ_{ijk}로 표기했으나 여기서는 $\epsilon_{k(ij)}$로 표기하여, 첨자 k가 첨자 i, j에 지분됨을 명시한다. 그러나 반복이 1이라면 첨자 k는 불필요하고 따라서 이런 경우에는 그냥 ϵ_{ij} 라고 표시한다.

- 어떤 모형에서도 반드시 전체 평균 μ와 오차 ϵ은 항상 나타난다. 그리고 실험에 개재된 모든 요인들의 주효과와 상호작용들을 나열한다. 그러나 지분 요인과 그 상위요인 간에는 상호작용이 없다. 이때 상위요인에 해당하는 첨자(들)은 괄호 속에 표시한다.

따라서 반복이 2 이상인 이요인 요인 설계 모형은 다음과 같이 작성한다.

$$Y_{ijk} = \mu + \alpha_i + \beta_j + (\alpha\beta)_{ij} + \epsilon_{k(ij)}$$

또 이 경우 만일 요인 B가 요인 A에 지분되어 있다면 모형은 다음과 같다.

$$Y_{ijk} = \mu + \alpha_i + \beta_{j(i)} + \epsilon_{k(ij)}$$

이 모형은 가장 간단한 형태의 지분 모형인데 이런 경우에 상호작용 AB는 존재할 수 없다. 따라서 다음 모형은 잘못된 것이다.

$$Y_{ijk} = \mu + \alpha_i + \beta_{j(i)} + (\alpha\beta)_{ij} + \epsilon_{k(ij)}$$

간단히 말해서 괄호 속 첨자와 괄호 밖 첨자 간에는 상호작용이 존재하지 않는다.

- 첨자는 사용 방식에 따라 세 종류로 구분한다. 효과항에 존재하되 괄호 밖에 있으면 '살았다(live)'고 한다. 효과항에 존재하되 괄호 안에 있으면 '죽었다(dead)'고 한다. 또 효과항에 사용되지 않았으면 '빠졌다(absent)'고 한다.

현재 모형에서 사용되는 첨자가 i, j, k의 3 개라고 하자. 이때 효과항 $(\alpha\beta)_{ij}$에서 i, j는 살았고 k는 빠진 것이다. 또 오차항 $\epsilon_{k(ij)}$에서 i, j는 죽었고 k는 살았다. 그리고 효과항 $\gamma_{j(i)}$가 있다면 여기서는 i가 죽었고 j는 살았고 k는 빠진 것이다.

- 모형에 나타난 각 효과항의 자유도는 다음 공식으로 결정된다.

$$\text{자유도} = \{\prod_{\text{죽은 첨자}}(\text{수준수})\} \times \{\prod_{\text{살아있는 첨자}}(\text{수준수}-1)\}$$

예를 들어 $(\alpha\beta)_{ij}$에서 i, j는 살아있다. 그러므로 $(\alpha\beta)_{ij}$의 자유도는 $(a-1)(b-1)$이다. 또 $\epsilon_{k(ij)}$에서 i, j는 죽었고 k는 살았으므로 대응되는 자유도는 $ab(n-1)$이 된다. 마찬가지로 $\gamma_{j(i)}$의 자유도는 $a(b-1)$이다. 요컨대 괄호 속 첨자는 수준수를 그대로 쓰고, 괄호 밖 첨자는 (수준수 - 1)을 하여 모두 곱하면 바로 자유도가 된다.

- 제곱합의 공식은 자유도 공식과 개념상 동일하다. 즉 어느 효과항에 대한 제곱합을 구할 때는 효과항에 붙어있는 모든 첨자에 대하여 살아있는 첨자에서는 1을 빼고, 죽은 첨자는 그대로 두고 모두 곱해서 전개한다. 이때 나타나는 첨자들과 그 부호는 제곱합 공식에서 평균에 그대로 반영된다. 빠진 첨자에 대해서는 그 첨자에 대하여 합산함을 나타내는 합산 부호 '•'를 붙인다.

예를 들어 반복이 2 이상인 이요인 요인 설계 모형에서 $(\alpha\beta)_{ij}$에 대응하는 제곱합의 공식을 구해보자. 두 첨자 모두 살았으므로 위에서 설명한 첨자곱의 규칙에 의거하면 $(i-1)(j-1) = ij - i - j + 1$을 얻는다. 따라서 대응되는 제곱합 공식은 다음과 같다.

$$\sum_i\sum_j\sum_k(\overline{Y}_{ij\cdot} - \overline{Y}_{i\cdot\cdot} - \overline{Y}_{\cdot j\cdot} + \overline{Y}_{\cdot\cdot\cdot})^2 \tag{7.1}$$

또 오차 $\epsilon_{k(ij)}$에 대한 제곱합은 $ij(k-1) = ijk - ij$이므로 다음과 같이 쓴다.

$$\sum_i\sum_j\sum_k(Y_{ijk} - \overline{Y}_{ij\cdot})^2$$

- 제곱합의 공식을 손계산으로 산출하려면 부득이 공식을 전개해야 한다. 손계산용 공식은 관측합을 기초로 주어지는데, 이 규칙은 제곱합 (7.1)을 변환하는 다음의 실례로 설명을 대신한다.

$$\sum_{i=1}^{a}\sum_{j=1}^{b}\frac{Y_{ij\cdot}^2}{n}-\sum_{i=1}^{a}\frac{Y_{i\cdot\cdot}^2}{bn}-\sum_{j=1}^{b}\frac{Y_{\cdot j\cdot}^2}{an}+\frac{Y_{\cdots}^2}{abn}$$

7.2 EMS 규칙

제 6.5 절에서 소개한 글자 표기 EMS 규칙은 요인 설계와 같은 교차 실험인 경우에 간편히 기대평균제곱을 산출하는 규칙이었다. 이 절에서는 지분 실험까지 포함하는 보다 일반적인 EMS 규칙을 소개한다.

다음 두 가지의 표기를 약속하자.

첫째로, Q를 효과라고 할 때 지시변수(indicator variable) δ_Q를 다음과 같이 정의한다.

$$\delta_Q=\begin{cases}0, & Q\text{는 고정효과}\\ 1, & Q\text{는 랜덤효과}\end{cases}$$

예를 들어 오차는 항상 랜덤이므로 언제나 $\delta_\epsilon=1$이다.

두 번째로, 편의상 모든 분산은 분산요소 형식으로 표현한다. 따라서 A가 고정일 때 σ_α^2란 표기는 실제로 $\sum\sum\alpha_i^2/(a-1)$을 의미한다. 마찬가지로 AB가 고정이라면 $\sigma_{\alpha\beta}^2=\sum\sum(\alpha\beta)_{ij}^2/(a-1)(b-1)$이다. 물론 랜덤이라면 그대로 사용하면 되겠다.

■ **1 단계**: 모형을 쓴다.

예를 들어 다음 모형을 기초로 EMS를 계산해보자.

$$Y_{ijkl}=\mu+\alpha_i+\beta_j+(\alpha\beta)_{ij}+\gamma_{k(j)}+(\alpha\gamma)_{ik(j)}+\epsilon_{ijkl}$$
$$(i=1,\cdots,a;\quad j=1,\cdots,b;\quad k=1,\cdots,c;\quad l=1,\cdots,n)$$

■ **2 단계**: 다음 형식으로 표를 만든다.

① 각 행의 제목에 모형의 효과항을 쓴다. μ는 쓰지 않는다.

② 각 행에 부제로 분산요소를 쓴다.

③ 열의 제목으로 모형에서 사용된 모든 첨자들을 나열한다.

④ 열의 부제로 첨자에 대응되는 지시변수를 쓴다.

⑤ 열의 두 번째 부제로 각 첨자에 대응하는 수준수를 쓴다. 이때 오차의 수준수는 처리 내 관측수다.

현재의 모형에 대하여 2 단계를 완성하면 표 7.1을 얻는다.

표 7.1 2 단계의 표 형식

EMS		i	j	k	l
		δ_A	δ_B	δ_C	1
		a	b	c	n
α_i	σ_α^2				
β_j	σ_β^2				
$(\alpha\beta)_{ij}$	$\sigma_{\alpha\beta}^2$				
$\gamma_{k(j)}$	σ_γ^2				
$(\alpha\gamma)_{ik(j)}$	$\sigma_{\alpha\gamma}^2$				
ϵ_{ijkl}	σ^2				

- **3 단계**: 오차항에 대한 행에서 각 열마다 1을 넣는다. (표 7.2)

- **4 단계**: 행 제목의 첨자 중 괄호 안의 첨자에 대응되는 열에 1을 넣는다. (표 7.3)

- **5 단계**: 행 제목의 첨자 중 괄호 밖의 첨자에 대응되는 열에 지시변수 δ 항을 넣는다. (표 7.4)

- **6 단계**: 나머지 모든 칸에는 열 별로 수준수를 쓴다. (표 7.5)

- **7 단계**: 임의의 항에 대한 기대평균제곱을 다음 순서로 구한다.
  ① 그 항에 붙은 모든 첨자들을 나열한다.
  ② 표의 각 행에서,
    (1) 만일 행 제목이 그 첨자들을 전부 포함하지 않으면 그 행을 무시한다.
    (2) 만일 행 제목이 그 첨자들을 전부 포함하면 그 첨자들의 열을 가리고 나머지 행의 원소들을 곱한다. 그리고 이 곱에 행의 부제인 분산요소를 곱한다.
  ③ 이 결과들을 더한다.

- **8 단계**: 7 단계를 모든 효과항에 대해서 반복한다.

표 7.2 3 단계의 표 형식

EMS		i	j	k	l
		δ_A	δ_B	δ_C	1
		a	b	c	n
α_i	σ_α^2				
β_j	σ_β^2				
$(\alpha\beta)_{ij}$	$\sigma_{\alpha\beta}^2$				
$\gamma_{k(j)}$	σ_γ^2				
$(\alpha\gamma)_{ik(j)}$	$\sigma_{\alpha\gamma}^2$				
ϵ_{ijkl}	σ^2	1	1	1	1

표 7.3 4 단계의 표 형식

EMS		i	j	k	l
		δ_A	δ_B	δ_C	1
		a	b	c	n
α_i	σ_α^2				
β_j	σ_β^2				
$(\alpha\beta)_{ij}$	$\sigma_{\alpha\beta}^2$				
$\gamma_{k(j)}$	σ_γ^2		1		
$(\alpha\gamma)_{ik(j)}$	$\sigma_{\alpha\gamma}^2$		1		
ϵ_{ijkl}	σ^2	1	1	1	1

표 7.4 5 단계의 표 형식

EMS		i	j	k	l
		δ_A	δ_B	δ_C	1
		a	b	c	n
α_i	σ_α^2	$\boldsymbol{\delta_A}$			
β_j	σ_β^2		$\boldsymbol{\delta_B}$		
$(\alpha\beta)_{ij}$	$\sigma_{\alpha\beta}^2$	$\boldsymbol{\delta_A}$	$\boldsymbol{\delta_B}$		
$\gamma_{k(j)}$	σ_γ^2		1	$\boldsymbol{\delta_C}$	
$(\alpha\gamma)_{ik(j)}$	$\sigma_{\alpha\gamma}^2$	$\boldsymbol{\delta_A}$	1	$\boldsymbol{\delta_C}$	
ϵ_{ijkl}	σ^2	1	1	1	1

표 7.5 6 단계의 표 형식

EMS		i	j	k	l
		δ_A	δ_B	δ_C	1
		a	b	c	n
α_i	σ^2_α	δ_A	b	c	n
β_j	σ^2_β	a	δ_B	c	n
$(\alpha\beta)_{ij}$	$\sigma^2_{\alpha\beta}$	δ_A	δ_B	c	n
$\gamma_{k(j)}$	σ^2_γ	a	1	δ_C	n
$(\alpha\gamma)_{ik(j)}$	$\sigma^2_{\alpha\gamma}$	δ_A	1	δ_C	n
ϵ_{ijkl}	σ^2	1	1	1	1

예를 들어 효과 $C(B)$에 대한 기대평균제곱은 다음과 같이 구한다. 효과 $C(B)$는 모형에서 $\gamma_{k(j)}$에 해당한다. 이 항의 첨자는 j와 k이다. 이제 표 7.5에서 각 행마다 순서대로 하나씩 연산을 한다.

처음 세 행은 제목이 α_i, β_j, $(\alpha\beta)_{ij}$로서 첨자 j와 k를 포함하지 않는다. 따라서 이들 행은 무시한다. 표 7.6에서 무시된 행을 지웠다.

나머지 세 행의 제목은 모두 첨자 j, k를 포함한다. 따라서 이 세 행에 대해서는 첨자 j, k의 열을 지운다. 표 7.6을 보라.

$\gamma_{k(j)}$ 행에서 남아있는 칸 내 원소들의 곱에 분산요소를 곱하면 $an\sigma^2_\gamma$이다.

$(\alpha\gamma)_{ik(j)}$ 행에서 남아있는 칸 내 원소들의 곱에 분산요소를 곱하면 $n\delta_A\sigma^2_{\alpha\gamma}$이다.

ϵ_{ijkl} 행에서 남아있는 칸 내 원소들의 곱은 $(1)(1)\sigma^2 = \sigma^2$이다.

표 7.6 $C(B)$에 대한 기대평균제곱의 계산

EMS		i	j	k	l
		δ_A	δ_B	δ_C	1
		a	b	c	n
α_i	σ^2_α	δ_A	b	c	n
β_j	σ^2_β	a	δ_B	c	n
$(\alpha\beta)_{ij}$	$\sigma^2_{\alpha\beta}$	δ_A	δ_B	c	n
$\gamma_{k(j)}$	σ^2_γ	a	1	δ_C	n
$(\alpha\gamma)_{ik(j)}$	$\sigma^2_{\alpha\gamma}$	δ_A	1	δ_C	n
ϵ_{ijkl}	σ^2	1	1	1	1

이 결과들을 모두 더하면 $E[MS_{C(B)}]$를 얻게 된다. 즉

$$E[MS_{C(B)}] = \sigma^2 + n\delta_A\sigma^2_{\alpha\gamma} + an\sigma^2_\gamma$$

이제 남은 작업은 효과항들이 고정이냐 랜덤이냐에 따라 δ 값으로 0 또는 1을 대입하고 또 고정일 경우에는 분산요소 공식을 바꾸는 것이다.

이 같은 연산을 모든 항에 대하여 수행하면 표 7.7과 같은 결과를 얻는다.

표 7.7 완성된 EMS 표

source	E(MS)
A	$\sigma^2 + n\delta_C\sigma^2_{\alpha\gamma} + cn\delta_B\sigma^2_{\alpha\beta} + bcn\sigma^2_\alpha$
B	$\sigma^2 + n\delta_A\delta_C\sigma^2_{\alpha\gamma} + an\delta_C\sigma^2_\gamma + cn\delta_A\sigma^2_{\alpha\beta} + acn\sigma^2_\beta$
AB	$\sigma^2 + n\delta_C\sigma^2_{\alpha\gamma} + cn\sigma^2_{\alpha\beta}$
$C(B)$	$\sigma^2 + n\delta_A\sigma^2_{\alpha\gamma} + an\sigma^2_\gamma$
$AC(B)$	$\sigma^2 + n\sigma^2_{\alpha\gamma}$
E	σ^2

< 연 습 문 제 >

1. 다음 모형에 대해서 EMS 표를 만들라.

$$Y_{ijk} = \mu + \alpha_i + \beta_{j(i)} + \epsilon_{ijk} \quad (i=1,\cdots,a\,;\ j=1,\cdots,b\,;\ k=1,\cdots,n)$$

① 고정 모형일 경우 ② 랜덤 모형일 경우
③ 혼합 모형일 경우: A는 고정, $B(A)$는 랜덤

2. 다음 모형에 대해서 EMS 표를 만들라. 효과항의 자유도를 명시하라.

$$Y_{ijkl} = \mu + \alpha_i + \beta_{j(i)} + \gamma_{k(ji)} + \epsilon_{ijkl} \quad (i=1,\cdots,a\,;\ j=1,\cdots,b\,;\ k=1,\cdots,c\,;\ l=1,\cdots,n)$$

3. 다음 모형에 대해서 EMS 표를 만들라.

$$Y_{ijkl} = \mu + \alpha_i + \beta_j + (\alpha\beta)_{ij} + \gamma_{k(ij)} + \epsilon_{ijkl}$$
$$(i=1,\cdots,a\,;\ j=1,\cdots,b\,;\ k=1,\cdots,c\,;\ l=1,\cdots,n)$$