An Analysis of Variance Procedure for the Split-Plot Design Using SPSS Syntax Window

Byoung-Chul Choi

Abstract

In conducting the analysis of variance for the split-plot design using the statistical package SPSS, users including statisticians are faced with difficulties because of no appropriate example in the SPSS applications guide book. In this paper, therefore, we present an analysis of variance procedure for the split-plot design using SPSS syntax window.

Keywords: split-plot design, analysis of variance, SPSS syntax window

1. 서론


분할법은 인자의 수준 변경이 쉽고 여러원 정을 고려하여 단계별로 인자를 배치하여 실험을 하기 때문에 실험 전체를 완전확률화하는 요인실험계획법과 구별되는 것이다. 이 때 인자를 두 단계로 나누어 하는 실험이면 단일분할법, 세 단계로 나누어 하는 실험이면 2단 분할법 등으로 불린다. 먼저 단일분할법의 분산분석이 어떻게 이루어지는지 다음 예1을 통해 살펴보자.

예1 어떤 실험이에서 <그림1>과 같은 순서로 혼합계급 \( A_1, A_2, A_3 \) 중 하나를 먼저 만든 후 세 가지 원도 \( B_1, B_2, B_3 \)으로 열처리를 하여 각도를 측정하는 방식으로 시험을 만들고, 9가지 모든 처리

---

1) Professor, division of Mathematics and Statistical Informatics, Chonbuk National University, Chonju, Chonbuk, 560-756, Korea.
E-mail: chbcn@chonbuk.ac.kr
조합에서 실험이 끝나면 전체 실험을 한 번 더 반복한다고 하자.

이와 같은 단일분할법에서는 인자 $A$를 1차단위 인자, 인자 $B$를 2차단위 인자라 하며, 전체 실험을 별도로 반복하는 반복인자 ($R$)는 1차단위 인자로 취급한다.

<table>
<thead>
<tr>
<th>반복1</th>
<th>반복2</th>
<th>실험순서</th>
</tr>
</thead>
<tbody>
<tr>
<td>1차단위</td>
<td>1차단위</td>
<td>B1 B2 B3</td>
</tr>
<tr>
<td></td>
<td></td>
<td>B1 B2 B3</td>
</tr>
<tr>
<td></td>
<td></td>
<td>B1 B2 B3</td>
</tr>
<tr>
<td>2차단위</td>
<td>2차단위</td>
<td>8 7 9</td>
</tr>
<tr>
<td></td>
<td></td>
<td>4 6 5</td>
</tr>
<tr>
<td></td>
<td></td>
<td>1 2 3</td>
</tr>
<tr>
<td></td>
<td></td>
<td>12 11 10</td>
</tr>
<tr>
<td></td>
<td></td>
<td>14 13 15</td>
</tr>
<tr>
<td></td>
<td></td>
<td>16 18 17</td>
</tr>
</tbody>
</table>

(그림1) 실험순서

이와 같은 단일분할법의 실험에는 인자가 $R$, $A$, $B$ 세이므로 반복이 없는 3원배치법을 연상할 수 있으나 (그림1)처럼 실험의 완전확률화가 1차단위 인자 $R$과 $A$에 의해 제한되어 있으므로 18회 전체 실험을 완전확률화하는 3원배치법과는 전혀 다르다. 따라서 자료의 모평식도 다음과 같이 3원배치법과는 다르며, 분산분석 방법도 다르게 된다.

3원배치법: $y_{ik} = \mu + r_i + a_i + (ar)_i + b_j + (ab)_j + e_{ijk}$

단일분할법: $y_{ik} = \mu + r_i + a_i + e_{1ijk} + b_j + (ab)_j + e_{2ijk}$

여기서

$r_i \sim N(0, \sigma_r^2)$, $e_{1ijk} \sim N(0, \sigma_{E_1}^2)$, $e_{2ijk} \sim N(0, \sigma_{E_2}^2)$

이와 서로 독립이다. 이 때 $e_{1ijk}$와 $e_{2ijk}$를 각각 1차단위 오차와 2차단위 오차라 부르며 예1의 분산분석표는 (표1)과 같다.

<table>
<thead>
<tr>
<th>단위</th>
<th>요인</th>
<th>측정값</th>
<th>자유도</th>
<th>평균제곱</th>
<th>분산제곱</th>
<th>F</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>$R$</td>
<td>SS$_R$</td>
<td>1</td>
<td>$MS_R$</td>
<td>$\sigma_R^2 + 3\sigma_{A-B}^2 + 9\sigma_{B}^2$</td>
<td>$MS_R/MS_{E_1}$</td>
</tr>
<tr>
<td></td>
<td>$A$</td>
<td>SS$_A$</td>
<td>2</td>
<td>$MS_A$</td>
<td>$\sigma_A^2 + 3\sigma_{A-B}^2 + 6Q(A)$</td>
<td>$MS_A/MS_{E_1}$</td>
</tr>
<tr>
<td></td>
<td>$E_1$</td>
<td>SS$_{E_1}$</td>
<td>2</td>
<td>$MS_{E_1}$</td>
<td>$\sigma_{E_1}^2 + 3\sigma_{A-B}^2$</td>
<td>$MS_{E_1}/MS_{E_2}$</td>
</tr>
<tr>
<td>2</td>
<td>$B$</td>
<td>SS$_B$</td>
<td>2</td>
<td>$MS_B$</td>
<td>$\sigma_B^2 + 6Q(B)$</td>
<td>$MS_B/MS_{E_1}$</td>
</tr>
<tr>
<td></td>
<td>$A \times B$</td>
<td>SS$_{A \times B}$</td>
<td>4</td>
<td>$MS_{A \times B}$</td>
<td>$\sigma_{A-B}^2 + 2Q(A \times B)$</td>
<td>$MS_{A \times B}/MS_{E_1}$</td>
</tr>
<tr>
<td></td>
<td>$E_2$</td>
<td>SS$_{E_2}$</td>
<td>6</td>
<td>$MS_{E_2}$</td>
<td>$\sigma_{E_2}^2$</td>
<td>$MS_{E_2}/MS_{E_1}$</td>
</tr>
<tr>
<td></td>
<td>계</td>
<td>SST</td>
<td>17</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
분산분석표에서 \( Q(A) \)는 인자 \( A \)의 모수효과(fixed effect)로

\[
Q(A) = \frac{\sum a_i^2}{l-1}
\]

이며, 일반적으로 수준수가 \( l, m \)인 두 모수인자 \( A \)와 \( B \)의 교호작용의 모수효과는

\[
Q(A \times B) = \frac{\sum \sum (ab)^2}{(l-1)(m-1)}
\]

과 같이 나타낸다.

<표1>의 분산분석표에서 제곱합이

\[
SST = SS_R + SS_A + SS_{E_1} + SS_B + SS_{A \times B} + SS_E
\]

와 같이 분해 되었는데 그 이론적 근거에 대해 알아보자. 예1을 인자가 \( R, A, B \)이고 실험순서가 완전확화된 반복이 없는 3원배치법으로 보면 제곱합은

\[
SST = SS_R + SS_A + SS_{A \times B} + SS_B + SS_{A \times B} + SS_E
\]

와 같이 분해 되는데 단일분할법에서는 인자 \( A \)와 \( B \)가 인자 \( R \)에 지분된 형태라 교호작용 \( A \times R \)과 \( B \times R \)이 고려되지 않아 \( A \times R \)을 1차단위 오차 \( E_1 \)으로 하고 \( B \times R \)을 2차단위 오차 \( E_2 \)에 포함시켜

\[
SS_{E_1} = SS_{B \times R} + SSE
\]

\( (E_2 \) 의 자유도) \( = (B \times R \) 의 자유도) \( + (E \) 의 자유도)\)

와 같이 된다. 자료의 모형에서 1차단위만 보면 인자가 \( A \)와 \( R \)인 반복이 없는 2원배치법과 같은 모형이라 교호작용 \( A \times R \)이 있다면 \( e_{(1)ab} \)와 교략되어 있어 별도로 검출할 수 없으므로 \( e_{(1)ab} \)으로 한 것이고, 교호작용 \( B \times R \)은 검출할 의미가 없어 \( e_{(2)ab} \)에 포함시킨 것이다.

2. SPSS를 이용한 단일분할법의 분석 절차

이제 SPSS를 이용한 단일분할법의 분석 절차에 대해 알아보자. 예1의 실험결과 다음 <표2>와 같은 자료를 얻었다고 하자
단일분할법에서는 인자 A와 B가 인자 R에 지분된 형태로 측정하여 SPSS(1999)에서 소개하고 있는 지분성시험의 기법을 이용하여 한다. 예1과 같은 모형의 자료를 SPSS를 이용해 분석하기 위해 <표2>의 자료를 입력하고 <그림2>의 절차에 따라 syntax 창의 /DESIGN 문장을

/DESIGN = r a a(r) b a*b

으로 수정하여 분산분석표를 작성하여 정리하면 <표3>과 같은 출력을 얻는다. <표3>의 결과를 보면 자유도, F검정통계량 등이 이론적으로 분석한 <표1>과 같이 제한된 분석방법이 타당함을 알 수 있다.

<그림2> 예1의 분석을 위한 SPSS 절차
<표3> 예1의 분산분석표

<table>
<thead>
<tr>
<th>단위</th>
<th>요인</th>
<th>계급합</th>
<th>자유도</th>
<th>평균계급</th>
<th>F</th>
<th>유의확률</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>R</td>
<td>가설</td>
<td>12.136</td>
<td>1</td>
<td>12.136</td>
<td>6.363</td>
</tr>
<tr>
<td></td>
<td></td>
<td>오차</td>
<td>3.815</td>
<td>2</td>
<td>1.907</td>
<td></td>
</tr>
<tr>
<td></td>
<td>A</td>
<td>가설</td>
<td>4.101</td>
<td>2</td>
<td>2.050</td>
<td>1.075</td>
</tr>
<tr>
<td></td>
<td></td>
<td>오차</td>
<td>3.815</td>
<td>2</td>
<td>1.907</td>
<td></td>
</tr>
<tr>
<td></td>
<td>A×R</td>
<td>가설</td>
<td>3.815</td>
<td>2</td>
<td>1.907</td>
<td>1.005</td>
</tr>
<tr>
<td></td>
<td></td>
<td>오차</td>
<td>11.390</td>
<td>6</td>
<td>1.898</td>
<td></td>
</tr>
<tr>
<td>2</td>
<td>B</td>
<td>가설</td>
<td>237.589</td>
<td>2</td>
<td>118.794</td>
<td>62.579</td>
</tr>
<tr>
<td></td>
<td></td>
<td>오차</td>
<td>11.390</td>
<td>6</td>
<td>1.898</td>
<td></td>
</tr>
<tr>
<td></td>
<td>A×B</td>
<td>가설</td>
<td>6.270</td>
<td>4</td>
<td>1.567</td>
<td>.826</td>
</tr>
<tr>
<td></td>
<td></td>
<td>오차</td>
<td>11.390</td>
<td>6</td>
<td>1.898</td>
<td></td>
</tr>
</tbody>
</table>

예2 1차단위 인자가 A와 B이고 2차단위 인자가 C이며 실험 전체의 반복인자가 R인 또 다른 형태의 단일분할법의 분석방법을 알아보자.

이와 같은 단일분할법의 모형식은 다음과 같다.

\[ y_{ijk} = \mu + r_i + a_j + (ab)_{ij} + e_{ijk} \]

모형에서 1차단위 오차 \( E_1 \)에 교호작용 \( A\times R \), \( B\times R \), \( A\times B\times R \)이 포함되어 있고 이들의 계급합과 자유도 합이 각각 1차단위, 오차의 계급합과 자유도가 된다. 또 2차단위 오차 \( E_2 \)에는 교호작용 \( C\times R \), \( A\times C\times R \), \( B\times C\times R \)이 포함되어 있고 이들의 계급합과 자유도 합이 각각 2차단위 오차의 계급합과 자유도가 된다.

이와 같은 모형의 SPSS 분석에서는 syntax 창의 /DESIGN 문장을

/DESIGN = r a b a*b a*b(r) c a*c b*c a*b*c .

으로 수정하면 된다. 여기서 \( a\times b(r) \)은 \( a(b\times r) \)이나 \( b(a\times r) \)과 같이 해도 되는데, 반드시 1차단위에 나타 있는 인자를 나타내는 문자가 하나나 있어야 한다.

위 표에서 $E_1$의 평균제곱의 기댓값이

$$E(MS_{E_1}) = \sigma^2_{E_1} + 3\sigma^2_{E_i}$$

과 같은 이유를 살펴보자. 이를 위해 먼저 $E_2$의 자료 모형을 인자가 $R, A, B, C$인 반복이 없는 4원배치법으로 보고 교호작용 $A \times R, B \times R, A \times B \times R$의 평균제곱의 기댓값을 구하면 $<표5>$와 같다.

$$<표5>\text{ 평균제곱의 기댓값(부분)}$$

<table>
<thead>
<tr>
<th>요인</th>
<th>자유도</th>
<th>$E(MS)$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$A \times R$</td>
<td>2</td>
<td>$\sigma^2_{E_1} + 3\sigma^2_{E_i} + 6\sigma^2_{A \times R}$</td>
</tr>
<tr>
<td>$B \times R$</td>
<td>2</td>
<td>$\sigma^2_{E_1} + 3\sigma^2_{E_i} + 9\sigma^2_{B \times R}$</td>
</tr>
<tr>
<td>$A \times B \times R = E_2$</td>
<td>2</td>
<td>$\sigma^2_{E_1} + 3\sigma^2_{E_i}$</td>
</tr>
</tbody>
</table>

그러면 교호작용 $A \times R, B \times R$이 의미가 없어 $\sigma^2_{A \times R} = \sigma^2_{B \times R} = 0$으로 간주하면, $<표5>$의 세 교호작용 $A \times R, B \times R, A \times B \times R$의 평균제곱의 기댓값이 모두 $\sigma^2_{E_1} + 3\sigma^2_{E_i}$으로 표현된다. 따라서 이들을 모두 $E_1$ 즉 1차단위 오차로 놓고 분석한다. 이때 $E_1$의 자유도는 5가 되는데, 이는 $<표5>$에서 세 교호작용의 자유도를 모두 합한 것이다. 또, $E_2$에 교차하여 있는 교호작용 $C \times R, A \times C \times R, B \times C \times R, A \times B \times C \times R$의 자유도를 계산할 때 각각 2, 4, 2, 4이므로 $E_2$의 자유도는 12이다. 다음과으로 제곱합에 대해서도

$$SS_{E_1} = SS_{A \times R} + SS_{B \times R} + SS_{A \times B \times R}$$
이 실험을 <표6>에서 확인할 수 있다.

**<표6>** \( E_1 \)과 \( E_2 \)의 자유도와 계급함

<table>
<thead>
<tr>
<th>요인</th>
<th>계급함</th>
<th>자유도</th>
</tr>
</thead>
<tbody>
<tr>
<td>( A \times R )</td>
<td>10.056</td>
<td>2</td>
</tr>
<tr>
<td>( B \times R )</td>
<td>5.444</td>
<td>1</td>
</tr>
<tr>
<td>( A \times B \times R )</td>
<td>1.056</td>
<td>2</td>
</tr>
<tr>
<td>계(= ( E_1 ))</td>
<td>16.556</td>
<td>5</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>요인</th>
<th>계급함</th>
<th>자유도</th>
</tr>
</thead>
<tbody>
<tr>
<td>( C \times R )</td>
<td>7.222</td>
<td>2</td>
</tr>
<tr>
<td>( A \times C \times R )</td>
<td>6.778</td>
<td>4</td>
</tr>
<tr>
<td>( B \times C \times R )</td>
<td>7.222</td>
<td>2</td>
</tr>
<tr>
<td>( A \times B \times C \times R )</td>
<td>5.778</td>
<td>4</td>
</tr>
<tr>
<td>계(= ( E_2 ))</td>
<td>14.000</td>
<td>12</td>
</tr>
</tbody>
</table>

또한 분산분석표 <표4>에서 \( F \) 검정도 1차단위 인자나 교호작용은 1차단위 오차를 오차로 하여 분석하고 2차단위 인자나 교호작용은 2차단위 오차를 오차로 하여 분석해야하는 이론대로 출력되어 있어 제한한 분석 방법이 절결함을 보이고 있다.

**예3** 1차단위 인자가 \( A \) 하나이고 2차단위 인자가 없는 단일 분할법을 고려해 보자.

이와 같은 경우에는 완전확률화의 일원배치법과는 다르며 모형식은

\[ y_{ij} = \mu + r_i + a_{ij} + e_{(1)ij} + e_{(2)ij} \]

이고 1차단위 오차에 교호작용 \( A \times R \)이 포함되어 있다. 이는 경우 SPSS를 이용해서 자료를 분석하는 방법은 syntax 창을 이용하지 않고 반복이 있는 이원배치법에서 인자 \( A \)를 모수인자로 인자 \( R \)을 변량인자로 지정하여 분석하거나, syntax 창의 /DESIGN 부분을

```
/DESIGN = r a a(r)
```

으로 수정해서 분석할 수 있다.

**예4** 모수인자 \( A \)와 \( B \)의 수준수를 각각 3으로 하여 총 9회의 실험조건 \( A_iB_j (i, j = 1, 2, 3) \)을 반복 하게 배치하여 실험을 한 후 어떤 특성치를 측정하면 측정은 매 실험조건에서의 실험이 중나는 노로 \( r \)개의 표본을 취하여 하는 실험을 고려해 보자.

이와 같은 실험은 인자가 분할이 안 된 경우로 1차단위 인자만 2개 있는 경우인데 실험의 확률화 방법이 분할법을 따르고 있으므로 자료의 모형식은

\[ y_{ijk} = \mu + a_i + b_j + e_{(1)ij} + e_{(2)ijk} \]
와 같이 1차단위 오차에 교호작용 $A \times B$가 교차되어 있다. 이 모형의 분산분석을 위해서는 1차단위 오차에 교호작용 $A \times B$가 교차되어 있음에 착안하여 Choi(2005)의 법칙에 따라 [표7]과 같이 평균계급의 기댓값을 구한 후 $F$검정통계량을 결정해야 한다.

[표7]  예4의 평균계급의 기댓값과 $F$검정통계량

<table>
<thead>
<tr>
<th>요인</th>
<th>분산성분</th>
<th>$E(\text{MS})$</th>
<th>$F$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Q(A)</td>
<td>$\sigma^2_{E_1}$</td>
<td>$\sigma^2_{E_1} + 2\sigma^2_{E_1} + 6Q(A)$</td>
<td>$MS_A/MS_E$</td>
</tr>
<tr>
<td>Q(B)</td>
<td>$\sigma^2_{E_1}$</td>
<td>$\sigma^2_{E_1} + 2\sigma^2_{E_1} + 6Q(B)$</td>
<td>$MS_B/MS_E$</td>
</tr>
<tr>
<td>$E_1(=A \times B)$</td>
<td>$\sigma^2_{E_1}$</td>
<td>$\sigma^2_{E_1}$</td>
<td>$MS_E/MS_E$</td>
</tr>
<tr>
<td>$E_2$</td>
<td></td>
<td>$\sigma^2_{E_1}$</td>
<td></td>
</tr>
</tbody>
</table>

[표7]와 같이 $F$검정통계량이 구해지게 하기 위한 SPSS 결차는 1차단위 인자가 하나 뿐인 예 3의 경우처럼 두 인자 중 하나를 인위적으로 변량인자로 지정하여 이원배치법으로 분석하거나, $B$를 변량인자로 지정한 경우 syntax 창의 /DESIGN 문장을

```
/DESIGN = a b a(b)
```

와 같이 수정해서 분석해도 된다.

3. SPSS를 이용한 2단 분할법의 분석 절차

예5 반복 $R$과 $A$를 1차단위 인자, $B$를 2차단위 인자, $C$, $D$, $F$를 3차단위 인자로 하는 2단 분할법을 고려해 보자.

인자의 수가 많으면 종 실험횟수가 많아지므로 최소의 실험으로 주효과와 주기 시험의 교호작용 $A \times B$와 $B \times F$ 만 검출하기 위해 부분요인을 하기로 하면 자료의 모형은

$$ y_{ijkmp} = \mu + r_i + a_i + e_{(1)ij} + b_j + (ab)_{ij} + e_{(2)ij} + c_{ik} + d_{ik} + f_{im} + (bf)_{im} + e_{(3)ijkmp} $$

1차단위 2차단위 3차단위

와 같다. 최소 요인의 실험조건을 구하기 위해서는 직교배열표법을 이용하여 배치하면 용이한데 여기서는 생략한다. 위 모형에 맞는 실험조건을 적절히 구하여 실험한 후 자료를 얻었다면 SPSS를 syntax 창의 /DESIGN 문장을

```
/DESIGN = r a a(r) b a*b b(a*r) c d f b*f
```
과 같이 수정하면 된다. /DESIGN 문장에서 \( b(a^r) \)을 \( a(b^r) \)과 같이 해도 분석결과는 같은데, 자료가 양어진 순서를 보면 인자 \( B \)가 인자 \( A \)에 지분된 형태이므로 \( b(a^r) \)을 사용하기를 권한다.

예6 예5의 2단 분할법에서 2차단위 인자가 \( B \) 와 \( C \)인 경우 syntax 창의 /DESIGN 문장을 수정해 보자.

예6의 경우 모형식은

\[
y_{ijk} = \mu + r_i + a_i + e_{ij} + b_j + c_k + (ab)_{ij} + e_{ijk} + d_{im} + (bf)_{jm} + e_{ijkm}
\]

와 같이지므로 syntax 창의 /DESIGN 문장을

```
/DESIGN = r a a(r) b c a*b b*c(a*r) d f b*f.
```

과 같이 수정해서 자료를 분석하면 된다. 여기서도 예5의 /DESIGN 문장에서처럼 \( b*c(a*r) \)을 \( a(b*c*r) \)이나 \( c(a*b*r) \), \( a*b(c*r) \), \( a*c(b*r) \)과 같이 해도 되는데, \( b*c(a*r) \)의 사용을 권하며 반드시 2차단위까지 나와 있는 인자를 나타내는 문장이 다 나오게 해야 한다.

3단 분할법 등 또한 다른 분할법의 분석에 대해서는 위 예제와 같은 방법으로 syntax 창의 /DESIGN 문장을 수정하여 어렵지 않게 할 수 있다. 그런데 /DESIGN 문장의 수정은 자료의 모형식을 보고 하는 것이므로, 결국, 자료의 운바른 분석을 위해서는 자료의 모형식에 대한 이해가 필수적이라고 본다.

참고문헌


[2004년 9월 점수, 2004년 12월 채택 ]