The Korean Communications
in Statistics Vol. 12 No. 1, 2005
pp. 61-69

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

Byoung-Chul Choi[1)]

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

통계학 전공자나 일반 사용자들이 통계패키지 SPSS나 SAS를 이용하여 분할법(split-plot design)의 분석을 하고자 할 경우 마땅한 지침서가 없어 고심하기가 일쑤이다. SAS를 이용하려면 SAS 지침서(1988)나 성내경(1989) 등이 소개한 방법에 따라 변량인자를 지정하여 먼저 평균제곱의 기대값(expected mean square; EMS)의 출력을 구하고 이를 이용하여 F검정통계량을 구하기 위한 TEST 문장을 작성해야 하므로 통계학 전문가가 아니고는 패키지 활용이 용이하지가 않다. 대화형 패키지 SPSS의 경우에는 SPSS지침서(1999)에도 분할법의 분석방법이 소개되어 있지 않고, 실험계획법의 참고서로 널리 쓰이는 박성현(2003), Hicks와 Turner(1999), Montgomery (2001) 등의 책에도 소개되어 있지 않아 SPSS를 사용한 분할법의 분석은 할 수 없는 것으로 알려져 있다. 그래서 이 논문에서는 SPSS를 이용하여 분할법을 분석할 수 있는 절차를 소개하고자 한다.

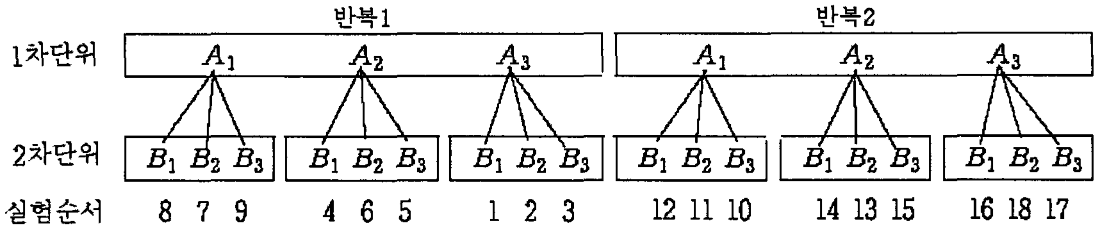
분할법은 인자의 수준 변경이 쉽고 어려운 점을 고려하여 단계별로 인자를 배치하여 실험을 하기 때문에 실험 전체를 완전확률화하는 요인실험계획법과 구별되는 것이다. 이 때 인자를 두 단계로 나누어 하는 실험이면 단일분할법, 세 단계로 나누어 하는 실험이면 2단 분할법 등으로 불린다. 먼저 단일분할법의 분산분석이 어떻게 이루어지는지 다음 예1을 통해 살펴보자.

예1 어떤 실험에서 <그림1>과 같은 순서로 혼합재료 A_1, A_2, A_3 중 하나를 먼저 만든 후 세 가지 온도 B_1, B_2, B_3로 열처리를 하여 강도를 측정하는 방식으로 시제품을 만들고, 9가지 모든 처리

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

조합에서 실험이 끝나면 전체 실험을 한 번 더 반복한다고 하자.

이와 같은 단일분할법에서는 인자 A를 1차단위 인자, 인자 B를 2차단위 인자라 하며, 전체 실험을 별도로 반복하는 반복인자(R)는 1차단위 인자로 취급한다.



<그림1> 실험순서

이와 같은 단일분할법의 실험에는 인자가 R, A, B 셋이므로 반복이 없는 3원배치법을 연상할 수 있으나 <그림1>처럼 실험의 완전확률화가 1차단위 인자 R과 A에 의해 제한되어 있으므로 18회 전체 실험을 완전확률화하는 3원배치법과는 전혀 다르다. 따라서 자료의 모형식도 다음과 같이 3원배치법과는 다르며, 분산분석 방법도 다르게 된다.

$$\text{3원배치법 :}\quad y_{ijk}=\mu+r_k+a_i+(ar)_{ik}+b_j+(ab)_{ij}+(br)_{jk}+e_{ijk},$$
$$\text{단일분할법:}\quad y_{ijk}=\mu+\underbrace{r_k+a_i+e_{(1)ik}}_{\text{1차단위}}+\underbrace{b_j+(ab)_{ij}+e_{(2)ijk}}_{\text{2차단위}}.$$

여기서

$$r_p\sim N(0,\sigma_R^2),\quad e_{(1)ik}\sim N(0,\ {\sigma_{E_1}}^2),\quad e_{(2)ijk}\sim N(0,\ {\sigma_{E_2}}^2)$$

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

<표1> 예1의 분산분석표

단위	요인	제곱합	자유도	평균제곱	$E(MS)$	F
1	R	SS_R	1	MS_R	$\sigma_E^2+3\sigma_{A\times R}^2+9\sigma_R^2$	MS_R/MS_{E_1}
	A	SS_A	2	MS_A	$\sigma_E^2+3\sigma_{A\times R}^2+6Q(A)$	MS_A/MS_{E_1}
	E_1	SS_{E_1}	2	MS_{E_1}	$\sigma_E^2+3\sigma_{A\times R}^2$	MS_{E_1}/MS_{E_2}
2	B	SS_B	2	MS_B	$\sigma_E^2+6Q(B)$	MS_B/MS_{E_2}
	$A\times B$	$SS_{A\times B}$	4	$MS_{A\times B}$	$\sigma_E^2+2Q(A\times B)$	$MS_{A\times B}/MS_{E_2}$
	E_2	SS_{E_2}	6	MS_{E_2}	σ_E^2	
	계	SST	17			

분산분석표에서 $Q(A)$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.

<표1>의 분산분석표에서 제곱합이

$$SST=SS_R+SS_A+SS_{E_1}+SS_B+SS_{A\times B}+SS_{E_2}$$

와 같이 분해 되었는데 그 이론적 근거에 대해 알아보자. 예1을 인자가 R, A, B이고 실험순서가 완전확률화된 반복이 없는 3원배치법으로 보면 제곱합은

$$SST=SS_R+SS_A+SS_{A\times R}+SS_B+SS_{A\times B}+SS_{B\times R}+SSE$$

와 같이 분해 되는데 단일분할법에서는 인자 A와 B가 인자 R에 지분된 형태라 교호작용 $A\times R$과 $B\times R$이 고려되지 않아 $A\times R$을 1차단위 오차 E_1으로 하고 $B\times R$을 2차단위 오차 E_2에 포함시켜

$$SS_{E_2}=SS_{B\times R}+SSE$$

(E_2의 자유도) = ($B\times R$의 자유도) + (E의 자유도)

와 같이 된다. 자료의 모형에서 1차단위만 보면 인자가 A와 R인 반복이 없는 2원배치법과 같은 모형이라 교호작용 $A\times R$이 있다면 $e_{(1)ik}$와 교락되어 있어 별도로 검출할 수 없으므로 $e_{(1)ik}$으로 한 것이고, 교호작용 $B\times R$은 검출할 의미가 없어 $e_{(2)ijk}$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

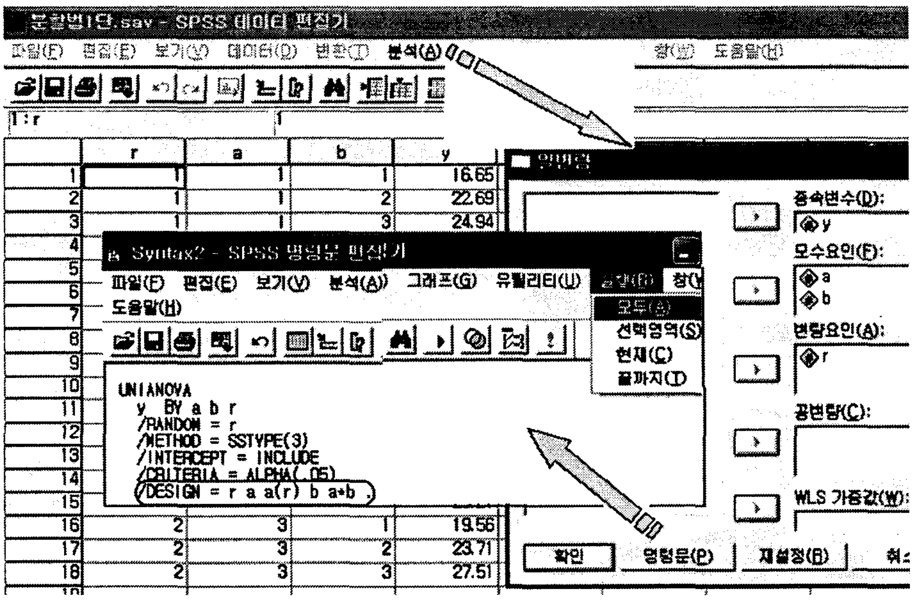
<표2> 예1의 자료

처리조합 R	A	B	강도	처리조합 R	A	B	강도
1	1	1	16.65	2	1	1	18.96
1	1	2	22.69	2	1	2	22.69
1	1	3	24.94	2	1	3	24.65
1	2	1	15.33	2	2	1	19.53
1	2	2	23.02	2	2	2	23.16
1	2	3	24.97	2	2	3	29.27
1	3	1	16.28	2	3	1	19.56
1	3	2	22.29	2	3	2	23.71
1	3	3	28.09	2	3	3	27.51

단일분할법에서는 인자 *A*와 *B*가 인자 *R*에 지분된 형태임에 착안하여 SPSS(1999)에서 소개하고 있는 지분실시법의 기법을 이용하려 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, *F*검정통계량 등이 이론적으로 분석한 <표1>과 같아 제안한 분석방법이 타당함을 알 수 있다.



<그림2> 예1의 분석을 위한 SPSS 절차

<표3> 예1의 분산분석표

단위	요인		제곱합	자유도	평균제곱	F	유의확률
1	R	가설	12.136	1	12.136	6.363	.128
		오차	3.815	2	1.907		
	A	가설	4.101	2	2.050	1.075	.482
		오차	3.815	2	1.907		
	$A \times R$ (E_1)	가설	3.815	2	1.907	1.005	.420
		오차	11.390	6	1.898		
2	B	가설	237.589	2	118.794	62.579	.000
		오차	11.390	6	1.898		
	$A \times B$	가설	6.270	4	1.567	.826	.554
		오차 (E_2)	11.390	6	1.898		

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

이와 같은 단일분할법의 모형식은 다음과 같다.

$$y_{ijkp} = \mu + r_p + a_i + b_j + (ab)_{ij} + e_{(1)ijp} + c_k + (ac)_{ik} + (bc)_{jk} + (abc)_{ijk} + e_{(2)ijkp}.$$

모형에서 1차단위 오차 E_1에 교호작용 $A \times R$, $B \times R$, $A \times B \times R$이 교락되어 있고 이들의 제곱합과 자유도 합이 각각 1차단위 오차의 제곱합과 자유도가 된다. 또 2차단위 오차 E_2에는 교호작용 $C \times R$, $A \times C \times R$, $B \times C \times R$, $A \times B \times C \times R$이 교락되어 있고 이들의 제곱합과 자유도 합이 각각 2차단위 오차의 제곱합과 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*b(r)을 a(b*r)이나 b(a*r)과 같이 해도 되는데, 반드시 1차단위에 나와 있는 인자를 나타내는 문자가 다 나오게 해야 된다.

이 분석방법의 타당성을 확인하기 위해 박성현(2003)의 적용사례(p203)를 위 SPSS 분석절차에 따라 분석하여 정리하면 <표4>와 같다. 참고로 이 적용사례에서 인자 R, A, B, C의 수준수는 각각 2, 3, 2, 3이다.

<표4> 예2의 분산분석표

단위	요인	제곱합	자유도	평균제곱	$E(MS)$	F
1	R	.444	1	.444	$\sigma^2_{E_2}+3\sigma^2_{E_1}+18\sigma^2_R$	MS_R/MS_{E_1}
	A	206.056	2	103.028	$\sigma^2_{E_2}+3\sigma^2_{E_1}+12Q(A)$	MS_A/MS_{E_1}
	B	4.000	1	4.000	$\sigma^2_{E_2}+3\sigma^2_{E_1}+18Q(B)$	MS_B/MS_{E_1}
	$A\times B$	1.500	2	.750	$\sigma^2_{E_2}+3\sigma^2_{E_1}+6Q(A\times B$	$MS_{A\times B}/MS_{E_1}$
	E_1	16.556	5	3.311	$\sigma^2_{E_2}+3\sigma^2_{E_1}$	MS_{E_1}/MS_{E_2}
2	C	113.389	2	56.694	$\sigma^2_{E_2}+12Q(C)$	MS_C/MS_{E_2}
	$A\times C$	49.444	4	12.361	$\sigma^2_{E_2}+4Q(A\times C)$	$MS_{A\times C}/MS_{E_2}$
	$B\times C$	3.167	2	1.583	$\sigma^2_{E_2}+6Q(B\times C)$	$MS_{B\times C}/MS_{E_2}$
	$A\times B\times C$	9.333	4	2.333	$\sigma^2_{E_2}+2Q(A\times B\times C)$	$MS_{A\times B\times C}/MS_{E_2}$
	E_2	14.000	12	1.167	$\sigma^2_{E_2}$	

위 표에서 1차단위 오차 E_1의 평균제곱의 기댓값이

$$E(MS_{E_1})=\sigma^2_{E_2}+3\sigma^2_{E_1}$$

과 같은 이유를 살펴보자. 이를 위해 먼저 예2의 자료 모형을 인자가 R,A,B,C인 반복이 없는 4원배치법으로 보고 교호작용 $A\times R$, $B\times R$, $A\times B\times R$의 평균제곱의 기댓값을 구하면 <표5>와 같다.

<표5> 평균제곱의 기댓값(부분)

요인	자유도	$E(MS)$
$A\times R$	2	$\sigma^2_{E_2}+3\sigma^2_{E_1}+6\sigma^2_{A\times R}$
$B\times R$	1	$\sigma^2_{E_2}+3\sigma^2_{E_1}+9\sigma^2_{B\times R}$
$A\times B\times R\equiv E_1$	2	$\sigma^2_{E_2}+3\sigma^2_{E_1}$

그런데 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 $\sigma^2_{E_2}+3\sigma^2_{E_1}$으로 표현된다. 따라서 이들을 모두 E_1, 즉 1차단위 오차로 놓고 분석한다. 이 때 E_1의 자유도는 5가 되는데, 이는 <표5>에서 세 교호작용의 자유도를 모두 합한 것이다. 또, E_2에 교락되어 있는 교호작용 $C\times R$, $A\times C\times R$, $B\times C\times R$ $A\times B\times C\times R$의 자유도를 계산하면 각각 2, 4, 2, 4이므로 E_2의 자유도는 12이다. 다음으로 제곱합에 대해서도

$$SS_{E_1}=SS_{A\times R}+SS_{B\times R}+SS_{A\times B\times R}$$

$$SS_{E_2} = SS_{C\times R} + SS_{A\times C\times R} + SS_{B\times C\times R} + SS_{A\times B\times C\times R}$$

이 성립함을 <표6>에서 확인할 수 있다.

<표6> E_1과 E_2의 자유도와 제곱합

요인	제곱합	자유도
$A\times R$	10.056	2
$B\times R$	5.444	1
$A\times B\times R$	1.056	2
계$(=E_1)$	16.556	5

요인	제곱합	자유도
$C\times R$	.722	2
$A\times C\times R$	6.778	4
$B\times C\times R$	.722	2
$A\times B\times C\times R$	5.778	4
계$(=E_2)$	14.000	12

또한 분산분석표 <표4>에서 F검정도 1차단위 인자나 교호작용은 1차단위 오차를 오차로 하여 분석하고 2차단위 인자나 교호작용은 2차단위 오차를 오차로 하여 분석해야하는 이론대로 출력되어 있어 제안한 분석 방법이 절절함을 보이고 있다.

예3 1차단위 인자가 A 하나이고 2차단위 인자가 없는 단일 분할법을 고려해 보자.

이와 같은 경우는 완전확률화의 일원배치법과는 다르며 모형식은

$$y_{ijp} = \mu + r_p + a_i + e_{(1)ip} + e_{(2)ijp}$$

이고 1차단위 오차에 교호작용 $A*R$이 교락되어 있다. 이런 경우 SPSS를 이용해서 자료를 분석하는 방법은 syntax 창을 이용하지 않고 반복이 있는 이원배치법에서 인자 A를 모수인자로 인자 R을 변량인자로 지정하여 분석하거나, syntax 창의 /DESIGN 문장을

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

으로 수정해서 분석할 수 있다.

예4 모수인자 A와 B의 수준수를 각각 3으로 하여 총 9회의 실험조건 $A_iB_j(i,j=1,2,3)$을 랜덤하게 택하여 실험을 한 후 어떤 특성치를 측정하되 측정은 매 실험조건에서의 실험이 끝나는 대로 r개의 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 같으며 1차단위 오차에 교호작용 $A\times B$가 교락되어 있다. 이 모형의 분산분석을 위해서는 1차단위 오차에 교호작용 $A\times B$가 교락되어 있음에 착안하여 Choi(2005)의 법칙에 따라 <표7>과 같이 평균제곱의 기댓값을 구한 후 F검정통계량을 결정해야 한다.

<표7> 예4의 평균제곱의 기댓값과 F검정통계량

요인	분산성분			$E(MS)$	F
	$Q(A)$	$Q(B)$	$\sigma^2_{E_1}$		
A	6		2	$\sigma^2_{E_2}+2\sigma^2_{E_1}+6Q(A)$	MS_A/MS_{E_1}
B		6	2	$\sigma^2_{E_2}+2\sigma^2_{E_1}+6Q(B)$	MS_B/MS_{E_1}
$E_1(\equiv A\times B)$			2	$\sigma^2_{E_2}+2\sigma^2_{E_1}$	MS_{E_1}/MS_{E_2}
E_2				$\sigma^2_{E_2}$	

<표7>과 같이 F검정통계량이 구해지게 하기 위한 SPSS 절차는 1차단위 인자가 하나 뿐인 예3의 경우처럼 두 인자 중 하나를 인위적으로 변량인자로 지정하여 이원배치법으로 분석하거나, B를 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 R과 A를 1차단위 인자, B를 2차단위 인자, C, D, F를 3차단위 인자로 하는 2단 분할법을 고려해 보자.

인자의 수가 많으면 총 실험횟수가 많아지므로 최소의 실험으로 주효과와 의심되는 교호작용 $A\times B$와 $B\times F$ 만 검출하기 위해 부분실험을 하기로 하면 자료의 모형은

$$y_{ijklmp}=\underbrace{\mu+r_p+a_i+e_{(1)ip}}_{1\text{차단위}}+\underbrace{b_j+(ab)_{ij}+e_{(2)ijp}}_{2\text{차단위}}+\underbrace{c_k+d_l+f_m+(bf)_{jm}+e_{(3)ijklmp}}_{3\text{차단위}}$$

와 같다. 최소 실험의 실험조건을 구하기 위해서는 직교배열표를 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 얻어진 순서를 보면 인자 B가 인자 A에 지분된 형태이므로 b(a*r)을 사용하기를 권한다.

예6 예5의 2단 분할법에서 2차단위 인자가 B와 C인 경우 syntax 창의 /DESIGN 문장을 수정해 보자.

예6의 경우 모형식은

$$y_{ijkp} = \mu + r_p + a_i + e_{(1)ip} + b_j + c_k + (ab)_{ij} + e_{(2)ijkp} + d_l + f_m + (bf)_{jm} + e_{(3)ijklmp}$$

와 같아지므로 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 되겠다. 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 인자들을 나타내는 문자가 다 나오게 해야 한다.

3단 분할법 등 또 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 올바른 분석을 위해서는 자료의 모형식에 대한 이해가 필수적이라고 본다.

참고문헌

[1] 박성현 (2003).「현대실험계획법」, 민영사, 서울.
[2] 성내경(1989).「SAS/STAT-분산분석」, 자유아카데미, 서울.
[3] Choi, B. C. (2005). EMS Rules for Balanced Factorial Designs under No Restriction on Interaction. *The Korean Communications in Statistics,* Vol. 12, 47-59.
[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] SPSS (1999). *SPSS Base 10.0 Applications Guide,* SPSS Inc, USA.

[2004년 9월 접수, 2004년 12월 채택]