

이단계 지분계획법의 오차제곱합 유도와 그 활용[†]

김대학¹

¹대구가톨릭대학교 수학과

접수 2013년 10월 23일, 수정 2013년 11월 15일, 게재확정 2013년 11월 20일

요약

확률화 블록계획법이나 교차된 이원분류표에 대한 분산분석과 오차제곱합의 성질은 널리 알려져 있다. 본 논문에서는 인자에 따른 수준에 또 다른 인자들이 내포되어 있는 지분계획 특히 이단계 지분계획에 대하여 구조적 특징을 설명하고 오차제곱합의 성질에 대하여 살펴보았다. 또한 이단계 지분계획의 활용으로 크로스오버 계획을 소개하고 생물학이나 약학 등의 분야에서 많이 사용되는 동등성 검정의 신뢰구간 구축에 대하여 설명하였고 실제자료와 SPSS 통계 패키지를 이용하여 분석함으로써 응용성을 부각시켰다.

주요용어: 랜덤효과모형, 상수효과모형, 신뢰구간추정, 오차제곱합, 지분계획법, 크로스오버 계획, 평균제곱.

1. 서론

역사적으로 실험의 계획과 분석 영역에서는 다양한 분야에 대한 응용과 실생활에서의 적용을 목적으로 이루어져 왔다. 특히 F 분포의 활용으로 여러 종류의 제곱합 (sum of squares)의 계산이 컴퓨터에 의해 쉽게 계산되고 가설검정을 통하여 결과가 쉽게 해석되어지고 있다. 이들 실험의 계획과 분석의 영역이 점점 실질적인 문제에 응용되어 지분계획법과 분할구 (split plot) 실험계획법등으로 확장되고 마찬가지로 F 통계량을 구성하는 오차제곱합의 유도가 계속 연구되고 있다. Searle (1971)은 행렬을 이용하여 오차제곱합의 구성을 다루었으며 그 이후로 많은 사람들에 의해 다양한 형태의 실험계획모형에 대한 오차제곱합의 유도가 연구되고 있다. 최근 Montgomery (2001)는 여러 가지 실험계획의 모형에 대하여 계획방법과 분석방법에 관하여 자세히 정리한 바 있다. 한편 Paik (1987)은 실험의 계획과 분석이란 제목으로 학부생들이 쉽게 이해할 수 있는 수준으로 상세하게 설명한 교재를 발간하기도 하였다. Kim과 Park (2004)은 통계전문 패키지인 SPSS를 이용하여 여러 가지 실험계획의 모형에 대하여 분산분석을 수행할 수 있도록 SPSS를 이용한 분산분석을 저술한 바 있으며 Park (2009)은 최근 현대실험계획법을 발간하였다. 실험계획 분야의 최근의 연구로는 Choi (2010), Choi (2011), Choi (2012) 등이 있다.

랜덤효과모형 (random effect model)과 상수효과모형 (fixed effect model) 그리고 혼합모형 (mixed effect model) 등을 이용한 이단계 지분계획법의 경우 오차제곱합에 대한 유도과정이 큰 의미를 지니고 있고 또 이를 이용한 신뢰구간의 형성이 여러 가지 정보를 제공하게 되므로 컴퓨터를 활용하는 유도과정의 이해와 실제 계산 그리고 이를 이용한 신뢰구간의 활용측면은 아주 중요하다. 일반적으로 교차된 이원분류자료나 다원분류자료에 대한 분산분석법과 그에 따른 오차제곱합 (error sum of squares)의 성

[†] 이 논문은 2012년도 대구가톨릭대학교 교내연구비 지원에 의한 것임.

¹ (712-702) 경상북도 경산시 하양읍 하양로 13-13, 대구가톨릭대학교 자연과학대학 수학과, 교수.
E-mail: dhkim@cu.ac.kr

질은 잘 알려져 있다. 그러나 대부분의 실질적 상황이나 실용적 활용차원을 고려하면 인자에 따른 수준에 또 다른 인자들이 내포되어 있는 지분계획 (nested design)이 필요한 경우가 많이 나타난다. 각종 조사에서 얻어지는 조사자료 등에서도 이런 경우를 자주 접하게 된다. 예로서 대도시 근교의 낙농가에서 생산된 우유의 품질이 균일한지를 조사하는 경우 여러 농가를 랜덤으로 또 이들 농가의 소들을 랜덤으로 추출하고 반복실험을 통하여 얻은 우유에 대한 지방함량을 측정하는 자료의 구조는 지분계획으로 접근가능하다. 이와 같이 첫 인자의 수준 내에 둘째 인자가 내포되어 있는 자료는 이단계 지분계획 (two stage nested design)으로 통칭되고 이와 관련된 분산분석은 이단계 지분계획법으로 잘 알려져 있다. Pereira와 Teixeira (2001)는 지분계획을 이용하여 팝콘의 유전자 성분을 추정하였으며 Biswas와 Raghavarao (1996)는 지분계획의 잔차 효과에 대하여 연구하였으며 Meyers-Levy와 Barnes (2001)는 지분계획법의 자료를 어떻게 분석할 수 있는지를 연구하였다. 한편 Kim 등 (2005)은 지분계획법에 대한 베이지안 연구를 수행한 바 있다. 지분계획에 관한 참고도서로는 Kempthorne (1952), Box 등 (2005)과 Box 등 (1978)이 있다.

본 연구에서는 이단계 지분계획의 분산분석에 있어서 통계적 가설 검정에 사용되는 통계량에서 절대적인 역할을 수행하는 오차제곱합의 유도과 그 활용에 대하여 살펴보고자 한다. 2절에서는 이단계 지분계획법의 소개와 오차제곱합의 성질 등에 대하여 서술하였고 3절에서는 이단계 지분계획법의 오차제곱합을 유도하였다. 4절에서는 특히 약학과 의학 등의 분야에서 자주 활용되고 있는 이른바 동등성 한계와 생동성 검정에 대하여 살펴보고 지분계획을 통한 모수의 차이에 대한 신뢰구간 등에 대하여 언급하였고 결론은 5절에 나타내었다.

2. 이단계 지분계획법

처리나 요인의 효과 τ_i 를 상수로 가정하고 분산분석을 수행하는 상수효과모형

$$y_{ij} = \mu + \tau_i + \epsilon_{ij} \quad (2.1)$$

과는 달리 τ_i 를 상수로 고려하기 보다는 확률변수로 생각하는 것이 타당한 경우가 흔히 발생한다. 여러 가지의 가능한 수준을 가지는 처리에 대하여 a 개만을 랜덤하게 선택하여 실험하는 경우를 고려하여 보자. 예로서 순무 (turnip greens)에 함유되어 있는 칼슘함량을 알아보기 위하여 한 포기씩의 순무에서 랜덤으로 a 개의 잎을 채취하고 채취된 잎의 각각에 대하여 n 번씩 분석 측정된 경우를 고려할 수 있을 것이다. 이러한 모형을 랜덤효과모형이라고 부른다. i 번째 잎의 j 번째 측정치를 y_{ij} 로 나타내면 다음과 같은 통계적 모형

$$y_{ij} = \mu + A_i + \epsilon_{ij} \quad (2.2)$$

을 생각할 수 있게 된다. 여기서 $i = 1, 2, \dots, a, j = 1, 2, \dots, n$ 이다. 이 때 전체 데이터의 수는 $N = kn$ 이고 μ 는 전체에 대한 평균으로서 추정하고자 하는 모수이다. A_i 는 i 번째 잎의 칼슘함량의 μ 에 대한 편차로서 선택한 잎에 따라 그 값이 달라질 수 있는 확률변수로서 평균이 0, 분산이 σ_A^2 인 정규분포를 따른다고 생각한다. 물론 ϵ_{ij} 는 오차항으로서 $E(\epsilon_{ij}) = 0, V(\epsilon_{ij}) = \sigma^2$ 인 정규분포를 따르는 확률변수이다. 또한 확률변수 A_i 와 ϵ_{ij} 는 서로 독립이라고 가정한다. 즉 모든 i 와 j 에 대하여 $E(A_i \epsilon_{ij}) = 0$ 이라고 가정한다. 이러한 랜덤효과모형 (2.2)에 대하여 지분계획을 적용하여 이단계 지분계획 모형

$$y_{ijk} = \mu + A_i + B_{j(i)} + \epsilon_{ijk} \quad (2.3)$$

을 고려할 수 있다. 여기서 $i = 1, 2, \dots, a, j = 1, 2, \dots, b, k = 1, 2, \dots, n$ 이다. 이때 $A_i \sim N(0, \sigma_A^2), B_{j(i)} \sim N(0, \sigma_B^2)$ 그리고 $\epsilon_{ijk} \sim N(0, \sigma^2)$ 이다.

Table 2.1 Two stage nested design structure and data

| site | 1 | | | | 2 | | | | 3 | | | |
|---------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|
| batch | 1 | 2 | 3 | 4 | 1 | 2 | 3 | 4 | 1 | 2 | 3 | 4 |
| | 13.70 | 19.15 | 20.77 | 16.20 | 10.70 | 15.53 | 9.42 | 18.53 | 16.44 | 14.15 | 9.81 | 6.24 |
| samples | 9.71 | 17.58 | 21.61 | 9.91 | 15.66 | 13.36 | 17.14 | 15.82 | 16.34 | 15.15 | 19.88 | 17.95 |
| | 12.17 | 11.76 | 14.04 | 7.42 | 10.87 | 15.05 | 14.04 | 10.75 | 13.97 | 17.07 | 16.32 | 13.39 |

보다 쉽게 지분계획을 이해하는 하나의 방법은 내포된 구조가 전반에 걸쳐서 모두 동일하다는 점을 기억하는 것이다. 예로서 서로 다른 세 장소에서 생산되는 어느 제약회사의 제품이 동일한가를 알아보는 실험을 생각하자. 각 장소는 네 개의 배치 (batch)가 있으며 이들 배치중에서 랜덤하게 배치를 선택하고 이로부터 제품을 랜덤하게 세 개 얻어서 성분을 조사하면 두 장소에서 생산되는 제품이 동일한지를 알 수 있을 것이다. 이를 구조화 하면 Table 2.1과 같이 나타낼 수 있다. 각 사이트 별로 모두 똑 같이 네 개의 배치가 존재하는 동일한 구조를 유지한다는 점이 지분계획의 특징인 셈이다. 다단계 지분계획 (multistage nested design)의 경우도 그 구조가 동일하여야 함을 짐작할 수 있다.

장소별로 제품을 선택하여 실험한후 Table 2.1과 같은 혈중농도를 얻었다. 모형 (2.3)에서 요인의 효과들을 모두 모수로 고려하는 경우와 상수효과모형과 랜덤효과모형의 혼합형으로 고려하는 경우 그리고 두 효과 모두를 랜덤효과로 고려하는 경우에 따라서 평균제품의 기댓값들은 각각 다르게 계산된다. A 를 상수효과로 B 를 랜덤효과로 고려하는 혼합모형의 경우 모형은

$$y_{ijk} = \mu + \tau_i + B_{j(i)} + \epsilon_{ijk} \tag{2.4}$$

로 되고 이때의 평균제품에 대한 기댓값은

$$\begin{aligned} E(MS_A) &= \sigma^2 + n\sigma_B^2 + \frac{bn \sum \tau_i^2}{a-1} \\ E(MS_{B(A)}) &= \sigma^2 + n\sigma_B^2 \\ E(MS_B) &= \sigma^2 \end{aligned}$$

이 된다. 두 효과 모두 랜덤효과로 고려하는 모형 (2.3)의 경우에는 평균제품의 기댓값은

$$\begin{aligned} E(MS_A) &= \sigma^2 + n\sigma_B^2 + bn\sigma_A^2 \\ E(MS_{B(A)}) &= \sigma^2 + n\sigma_B^2 \\ E(MS_B) &= \sigma^2 \end{aligned} \tag{2.5}$$

의 관계가 성립함이 알려져 있다.

3. 이단계 지분계획법의 오차제곱합 유도

랜덤효과모형이든 상수효과모형이든 관계없이 이단계 지분계획 모형 (2.4)에 대하여 수정된 전체제곱합 (adjusted total sum of squares)을 분해하면

$$\sum_i \sum_j \sum_k (y_{ijk} - \bar{y})^2 = bn \sum_i (\bar{y}_{i..} - \bar{y}_{...})^2 + n \sum_i \sum_j (\bar{y}_{ij.} - \bar{y}_{i..})^2 + \sum_i \sum_j \sum_k (y_{ijk} - \bar{y}_{ij.})^2 \tag{3.1}$$

로 항상 같은 결과를 얻게된다. 식 (3.1)은

$$SS_T = SS_A + SS_{B(A)} + SS_E \tag{3.2}$$

로 표현할 수 있다. 이들 제곱합들에 대한 구체적인 관계식은

$$\begin{aligned}
 SS_A &= \frac{1}{bn} \sum_i y_{i..}^2 - \frac{y_{...}^2}{abn} \\
 SS_{B(A)} &= \frac{1}{n} \sum_i \sum_j y_{ij.}^2 - \frac{1}{bn} \sum_i y_{i..}^2 \\
 SS_E &= \sum_i \sum_j \sum_k y_{ijk}^2 - \frac{1}{n} \sum_i \sum_j y_{ij.}^2 \\
 SS_T &= \sum_i \sum_j \sum_k y_{ijk}^2 - \frac{y_{...}^2}{abn}
 \end{aligned}$$

과 같이 나타난다. 이들 제곱합들을 이용하여 이단계 지분계획모형 (2.4)에 대한 분산분석표를 구하면 다음과 같은 결과를 얻게 된다.

Table 3.1 Analysis of variance table for two stage nested design

| source of variation | sum of squares | degrees of freedom | mean squares |
|---------------------|----------------|--------------------|--------------|
| A | SS_A | $a - 1$ | MS_A |
| B (within A) | $SS_{B(A)}$ | $a(b - 1)$ | $MS_{B(A)}$ |
| Error | SS_E | $ab(n - 1)$ | MS_E |
| Total | SS_T | $abn - 1$ | |

Table 3.1에서 평균제곱 (mean squares)들은 해당되는 제곱합을 각각의 자유도로 나누어 얻은 값으로

$$\begin{aligned}
 MS_A &= bn \sum_i (\bar{y}_{i..} - \bar{y}_{...})^2 / (a - 1) \\
 MS_{B(A)} &= n \sum_i \sum_j (\bar{y}_{ij.} - \bar{y}_{i..})^2 / a(b - 1) \\
 MS_E &= \sum_i \sum_j \sum_k (y_{ijk} - \bar{y}_{ij.})^2 / ab(n - 1)
 \end{aligned}$$

로 얻어진다. Table 2.1의 데이터를 IBM SPSS Statistics 19를 이용하여 분석한 결과는 다음과 같이 나타난다. Table 3.2로부터 오차제곱합은 63.333이고 오차분산의 추정치는 2.639임을 알 수 있다.

Table 3.2 Anova table for the data in Table 2.1

| source of variation | sum of squares | degrees of freedom | mean squares | F | P-value |
|---------------------|----------------|--------------------|--------------|-------|---------|
| site | 15.056 | 2 | 7.528 | 2.853 | 0.077 |
| site (batch) | 69.917 | 9 | 7.769 | 2.944 | 0.017 |
| error | 63.333 | 24 | 2.639 | | |
| Total | 148.306 | 35 | | | |

4. 지분계획법의 활용 : $\mu_T - \mu_R$ 의 신뢰구간추정

4.1. 지분계획법의 활용분야

제약회사에서 생산되는 바이오시밀러 의약품의 효능을 비교할 때 이단계 지분계획법은 빈번히 활용되고 있다. 특히 생동성 시험이나 동등성 한계 (equivalent margin)와 관련한 분야에서 지분계획법은 그

활용가치가 상당하다. 임상실험을 통한 실험계획의 방법에는 크게 환자가 단 하나의 처리만을 적용받게 계획하는 패러렐계획 (parallel design)과 이와는 상반되게 환자가 일정기간을 두고 여러 처리를 적용받게 되는 크로스오버 계획 (crossover design)이 있다. 두 표본 (two sample)의 경우를 고려하면 패러렐계획은 두 독립모집단의 모평균의 동일성 검정에 해당된다. 크로스오버 계획은 두 기간에 걸쳐 이루어질 수도 있고 여러 기간에 걸쳐 실험이 진행될 수도 있다. 두 기간에 걸쳐 실험이 진행되는 크로스오버계획을 두 기간 크로스오버 계획 (two period crossover design)이라 부르고 여러 기간에 걸쳐 실험이 진행되는 계획을 다중 기간 크로스오버 계획 (multi-period crossover design)이라 부른다. 지분계획의 활용을 연구하기 위하여 본 논문에서는 두 기간 크로스오버 계획만을 고려하기로 하자.

새로 만들어진 항생제의 생동성 시험을 고려하자. 최근 오리지널 생물학적 의약품들 중에서 특히 만료를 앞둔 제품들이 많아지면서 바이오시밀러 의약품에 대한 관심이 증가하고 있다. 바이오시밀러 의약품에서 동등성은 오리지널 의약품과 똑 같다는 의미가 아닌 오리지널 생물학적 의약품과 바이오시밀러 의약품 사이에 존재하는 화학적, 생물학적, 임상적 특성들의 차이가 그 이상을 넘어서면 유사 또는 비견할 만하다고 간주할 수 없는 어떤 한계보다 작은 경우를 가리킨다. 이러한 한계를 동등성한계 (equivalence margin)라고 부른다. 생동성 시험은 제네릭의약품의 허가를 위해 이미 허가를 받은 기존의약품과 제네릭의약품을 교차연구의 형태로 투여한 뒤 혈중 농도로부터 구한 약동학적 모수 (pharmacokinetic parameter)를 비교함으로써 동등성 유무를 판단하는 임상연구이다. 생동성 판단에 사용하는 약동학적 모수는 전신혈 흡수 (systemic absorption)의 속도와 정도를 각각 대변한다고 알려진 최고농도 (Cmax)와 농도시간곡선하면적 (area under the concentrationtime curve; AUC)이다. 최고농도와 농도시간곡선하면적의 제네릭의약품 대 기존의약품 기하평균비 (geometric mean ratio)의 90% 신뢰구간이 [0.8, 1.25] 범위 안에 들면 동등하다고 간주한다.

기존의 대조약 (reference drug)과 새로 만들어진 시험하고자 하는 약 (test drug)의 두 모집단의 모평균을 각각 μ_R 과 μ_T 라 두자. 우리가 원하는 생동성 검정은

$$H_0 : \frac{\mu_T}{\mu_R} \leq 0.8 \quad \text{or} \quad H_0 : \frac{\mu_T}{\mu_R} \geq 1.25 \quad (4.1)$$

$$H_1 : 0.8 < \frac{\mu_T}{\mu_R} < 1.25$$

이다. 1.25는 $\frac{1}{0.8}$ 임을 쉽게 알 수 있고 위의 가설은 새로운 약이 기존의 약보다 평균적으로 80% 이상의 성능을 지니고 있는가를 나타내고 있음을 짐작할 수 있다. 동등성한계에 관한 기준으로는 Lee (2012)를 참고하면 된다. 일반적으로 바이오시밀러 의약품의 효능의 분포는 로그정규분포 (log-normal distribution)를 가정하고 있으므로 로그 변환을 고려하면 (4.1)은 $\eta_T = \log \mu_T$, $\eta_R = \log \mu_R$ 로 둘 때

$$H_0 : \eta_T - \eta_R \leq \log 0.8 \quad \text{or} \quad H_0 : \eta_T - \eta_R \geq \log 1.25 \quad (4.2)$$

$$H_1 : \log 0.8 < \eta_T - \eta_R < \log 1.25$$

와 같다. 우리는 $\eta_T - \eta_R$ 의 신뢰구간 추정에 관심을 두도록 하자.

4.2. 두 기간 크로스오버 실험계획

새로운 항생제 효능이 생동적인지를 검정하기 위하여 두 기간 크로스오버 계획을 적용하여 확인하기로 하자. 크로스오버 계획의 특징상 모든 환자는 두 기간 동안 대조약과 시험약을 다 복용하여야 한다. 이를 위하여 전체 대상자를 두 그룹으로 나누어 한 그룹은 대조약과 시험약의 순서로 또 다른 그룹은 시험약과 대조약의 순서로 복용하게 하여야 한다. 따라서 사람 (subject)들은 그룹에 내포되어 있고 지분계획의 요소가 되는 셈이다. 이를 표로 나타내면 Table 4.1과 같다.

Table 4.1 Crossover design and data structure

| group | 1 | 2 |
|----------|------------------------------------|------------------------------------|
| subject | 1, 2, ..., m | 1, 2, ..., n |
| period 1 | drug =R | drug =T |
| data | $X_{111}, X_{121}, \dots, X_{1m1}$ | $X_{211}, X_{221}, \dots, X_{2n1}$ |
| period 2 | drug =T | drug =R |
| data | $X_{112}, X_{1m1}, \dots, X_{1m2}$ | $X_{212}, X_{222}, \dots, X_{2n2}$ |

(R: reference, T: test)

이런 지분계획과 관련된 분산분석은 2절과 3절의 내용을 참고하여 이루어질 수 있다. 우리가 두 기간 크로스오버 실험계획에 대하여 고려하는 모형은 이단계 지분계획모형으로서

$$y_{ijk} = \mu + G_i + S_{j(i)} + p_k + q_l + \epsilon_{ijk} \tag{4.3}$$

이다. 여기서 $y_{ijk} = \log X_{ijk}$ 이고 G_i 는 그룹효과로서 랜덤효과로, $S_{j(i)}$ 는 사람효과로서 모수효과로, 또 p_k 는 기간효과로 모수효과로 마지막으로 q_l 은 약효과로 모수효과를 나타낸다. 물론 $i = 1, 2, j = 1, 2, \dots, m, k = 1, 2$ 이다. 이 때 l 은 i 와 k 에 의존하는 값으로서 q_l 은 약효과를 나타낸다. 실제로

$$q_l = \begin{cases} q_R & \text{if } i = 1, k = 1 \text{ or } i = 2, k = 1 \\ q_T & \text{if } i = 1, k = 2 \text{ or } i = 2, k = 2 \end{cases}$$

로서 각각 대조약과 시험약의 효과를 나타낸다. 이 경우에 평균에 영향을 미치는 요인과 각각에 해당되는 자유도는 Table 4.2와 같다.

이들 자료로부터 우리가 원하는 신뢰구간을 추정하여 보자. $\eta_T - \eta_R$ 의 신뢰구간은 $y_{ijk} = \log X_{ijk}$ 에 대하여 $\eta_T - \eta_R$ 의 추정치 $D = \bar{y}_{1.2} + \bar{y}_{2.1} - \bar{y}_{1.1} - \bar{y}_{2.2}$ 를 이용하여 계산한다. 여기서 $\bar{y}_{i.j}$ 는 i 번째 그룹의 j 번째 기간의 자료들의 평균이다.

Table 4.2 Factors and degrees of freedom for clarithromycin data

| factors | degrees of freedom |
|----------------|--------------------|
| group | 1 |
| subject(group) | m+n-2 |
| period | 1 |
| drug | 1 |

평균값의 차 $\eta_T - \eta_R$ 의 추정량의 분산은 지분계획법의 분산분석표에 나타나는 MS_E 를 이용하여 $V(D) = MS_E(1/m + 1/n)$ 와 같이 계산한다. 이때의 자유도는 $n + m - 2$ 이다. 이들 결과를 이용하여 $\eta_T - \eta_R$ 의 $100(1 - \alpha)\%$ 신뢰구간은

$$D - t_{1-\alpha, m+n-2} \sqrt{V(D)} < \eta_T - \eta_R < D + t_{1-\alpha, m+n-2} \sqrt{V(D)} \tag{4.4}$$

과 같이 추정된다.

4.3. 데이터 분석

위장병에 사용되는 항생제로서 헬리코박터 균의 살균에 사용되는 clarithromycin에 대하여 두 기간 크로스오버 실험계획을 통하여 얻어진 혈중농도 자료로부터 $\eta_T - \eta_R$ 의 신뢰구간을 추정하여 보자.

| group | subject | drug | period | auc | cmax | logauc | logcmax |
|-------|---------|------|--------|--------|------|--------|---------|
| 17 | 1 | 6 | 2 | 14.150 | 2.13 | 2.65 | .76 |
| 18 | 1 | 7 | 2 | 15.150 | 2.04 | 2.72 | .71 |
| 19 | 1 | 8 | 2 | 17.070 | 2.06 | 2.84 | .72 |
| 20 | 1 | 9 | 2 | 9.810 | 1.68 | 2.28 | .52 |
| 21 | 1 | 10 | 2 | 19.880 | 3.07 | 2.99 | 1.12 |
| 22 | 1 | 11 | 2 | 16.320 | 2.36 | 2.79 | .86 |
| 23 | 2 | 1 | 1 | 7.420 | .88 | 2.00 | -.13 |
| 24 | 2 | 2 | 1 | 10.700 | 1.56 | 2.37 | .44 |
| 25 | 2 | 3 | 1 | 15.660 | 2.27 | 2.75 | .82 |
| 26 | 2 | 4 | 1 | 10.870 | 1.64 | 2.39 | .49 |
| 27 | 2 | 5 | 1 | 15.530 | 2.29 | 2.74 | .83 |
| 28 | 2 | 6 | 1 | 13.360 | 1.59 | 2.59 | .46 |
| 29 | 2 | 7 | 1 | 15.650 | 2.01 | 2.75 | .70 |
| 30 | 2 | 8 | 1 | 9.420 | 1.38 | 2.24 | .32 |
| 31 | 2 | 9 | 1 | 17.140 | 2.55 | 2.84 | .94 |
| 32 | 2 | 10 | 1 | 14.040 | 2.14 | 2.64 | .76 |
| 33 | 2 | 11 | 1 | 18.530 | 2.85 | 2.92 | 1.05 |
| 34 | 2 | 1 | 2 | 6.240 | .83 | 1.83 | -.19 |
| 35 | 2 | 2 | 2 | 17.950 | 2.83 | 2.89 | 1.04 |
| 36 | 2 | 3 | 2 | 13.390 | 2.28 | 2.59 | .82 |
| 37 | 2 | 4 | 2 | 12.990 | 1.84 | 2.56 | .61 |
| 38 | 2 | 5 | 2 | 13.990 | 2.23 | 2.64 | .80 |
| 39 | 2 | 6 | 2 | 10.820 | 1.28 | 2.38 | .25 |
| 40 | 2 | 7 | 2 | 21.880 | 2.92 | 3.09 | 1.07 |
| 41 | 2 | 8 | 2 | 10.400 | 1.51 | 2.34 | .48 |
| 42 | 2 | 9 | 2 | 14.430 | 2.31 | 2.67 | .84 |
| 43 | 2 | 10 | 2 | 11.110 | 1.69 | 2.41 | .52 |
| 44 | 2 | 11 | 2 | 16.990 | 2.89 | 2.83 | 1.06 |

Figure 4.1 Used clarithromycin data

분산분석을 위하여 IBM SPSS Statistics 19 통계 패키지를 이용하여 다음과 같은 결과를 얻었다. 자료에 사용된 항생제는 시험약 및 대조약 각각 250mg*2캡셀이고 환자는 그룹 당 11명을 2 × 2 교차실험 하였고 고성능 액체크로마토그래피 (high performance liquid chromatography; HPLC)를 이용하여 분석한 혈중농도를 기록한 자료를 사용하였다. SPSS 통계 패키지를 사용하기 위하여 먼저 Figure 4.1과 같이 clarithromycin 자료를 입력하고 자료의 변환을 실행한다. 그런 다음 메뉴에서 [분석]-[일반선형모형]-[일반량]을 선택한다. 이후 Figure 4.2와 같이 나타나는 대화상자에서 변량요인으로 그룹을, 모수요인으로 환자 (subject)와 약 (drug) 그리고 기간 (period)을 선택하고 종속변수로서 로그변환된 변수 logcmax를 선택한다. logcmax는 혈중농도를 로그변환한 값이다. 이후 [모형]을 눌러 모형에 맞게 모든 주 효과들을 입력하고 확인을 누른다. [붙여넣기]를 눌러 지분계획모형입을 나타내기 위하여 Figure 4.3과 같이 맨 아래 줄에 subject(group)을 첨가한다.



Figure 4.2 Model setting and variable selection

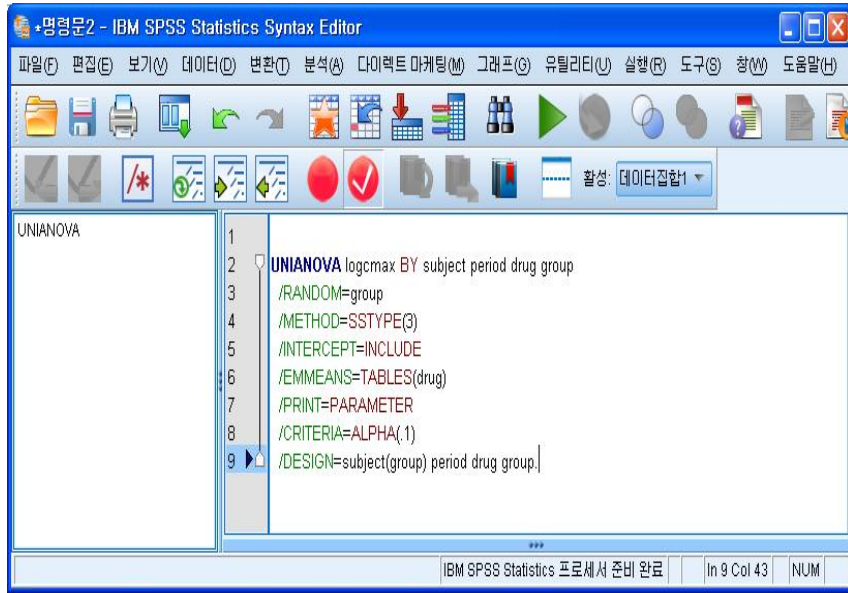


Figure 4.3 SPSS command application

Table 4.3은 SPSS의 출력결과로서 지분계획을 통한 두 기간 크로스오버 실험계획의 분산분석표를 나타낸다. 이로부터 오차분산의 추정값 MS_E 가 0.041임을 알 수 있고 간단한 기술통계량을 이용하여 $\eta_T = 0.723$, $\eta_R = 0.717$ 임을 확인할 수 있다 (Table 4.4).

이들 결과로부터 $\eta_T - \eta_R$ 의 신뢰구간을 계산하기 위하여 먼저 $D = 0.0064$, $t_{1-\alpha, m+n-2} = t_{0.05, 20} = 1.725$, $V(D) = 0.041(1/11 + 1/11)/2 = 0.0037$ 을 얻고 이로부터

$$-0.999 \leq \eta_T - \eta_R \leq 0.112$$

를 얻게 된다. 즉

$$0.9056 \leq \frac{\mu_T}{\mu_R} \leq 1.1185$$

를 얻게 되어 기준범위 (0.8 - 1.25)안에 있음을 알 수 있다.

Table 4.3 Anova table for clarithromycin data

| source of variation | sum of squares | degrees of freedom | mean squares | F | P-value |
|---------------------|----------------|--------------------|--------------|-------|---------|
| group | 0.309 | 1 | 0.309 | 2.075 | 0.165 |
| period | 0.028 | 1 | 0.028 | 0.667 | 0.424 |
| drug | 0.004 | 1 | 0.004 | 0.011 | 0.919 |
| subject(group) | 2.973 | 20 | 0.149 | 3.608 | 0.003 |
| error | 0.824 | 20 | 0.041 | | |

Table 4.4 Descriptive statistic

| drug | mean | standard erro | 95% confidence interval | |
|------|-------|---------------|-------------------------|-------------|
| | | | lower bound | upper bound |
| R | 0.717 | 0.043 | 0.626 | 0.807 |
| T | 0.723 | 0.043 | 0.633 | 0.813 |

5. 결론 및 토의

우리는 이단계 지분계획법에 대한 오차제곱합을 유도하고 분산분석표를 작성하였다. 또한 이단계 지분계획법의 활용으로 크로스오버 계획법도 살펴보았다. 그러나 실제로 대부분의 경우 그 계산량이 많기 때문에 오차제곱합은 데이터를 이용하여 컴퓨터 프로그램으로 구해진다. 랜덤효과모형과 상수효과모형 그리고 혼합모형 등을 이용한 이단계 지분계획법의 경우 오차제곱합에 대한 유도과정과 각 경우의 제곱합의 기댓값등이 큰 의미를 지니고 있고 또 이를 이용한 신뢰구간의 형성이 여러 가지 정보를 제공하게 되므로 통계 패키지를 활용하는 유도과정의 이해와 실제 계산 그리고 이를 이용한 신뢰구간의 활용측면은 아주 중요하다. SPSS의 경우도 지분계획을 적용하기 위해서는 모형설정과정에서 필요한 만큼 명령행에서 입력을 해야 하므로 구조적으로 모형을 이해하고 활용할 수 있도록 하여야 할 것이다.

References

- Biswas, N. and Raghavarao, D. (1996). Nested residual effects design. *The Indian Journal of Statistics*, **58**, 240-244.
- Box, G. E. P., Hunter, W. G. and Hunter, J. S. (1978). *Statistics for experimenters : An introduction to design, data analysis and model building*, John Wiley & Sons, New York.
- Box, G. E. P., Hunter, W. G. and Hunter, J. S. (2005). *Statistics for experimenters : Design, innovation, and discovery*, John Wiley & Sons, New York.
- Choi, J. (2010). A mixed model for repeated split-plot data. *Journal of the Korean Data & Information Science Society*, **21**, 1-9.
- Choi, J. (2011). Variance components in one factor random model by projection. *Journal of the Korean Data & Information Science Society*, **22**, 381-387.
- Choi, Y. H. (2012). Power study for 4×4 graeco-latin square design. *Journal of the Korean Data & Information Science Society*, **23**, 683-691.
- Kempthorne, O. (1952). *The design and experiments*, John Wiley & Sons, New York.
- Kim, D. H., Kang, S. G. and Lee, W. D. (2005). Noninformative priors for the nested design. *Journal of Statistical Planning and Inference*, **133**, 453-462.
- Kim, D. H. and Park, S. Y. (2004). *Analysis of variance using SPSS*, Kyowoosa, Seoul.
- Lee, H. (2012). Equivalence margin of the biosimilar product. *Journal of Korean Society for Clinical Pharmacology and Therapeutics*, **20**, 17-33.
- Meyers-Levy, J. and Barnes, J. (2001). How to analyze data from a nested design. *Journal of Consumer Psychology*, **10**, 21-24.
- Montgomery, D. C. (2001). *Design and analysis of experiments*, John Wiley & Sons, New York.
- Paik, W. B. (1989). *Experimental design and analysis*, Freedom Academy, Seoul.
- Park, S. H. (2009). *Modern experimental design*, Minyoungsa, Seoul.
- Pereira, M. G. and Teixeira, A. (2001). Estimation of genetic components in popcorn based on the nested design. *Crop Breeding and Applied Biotechnology*, **1**, 3-10.
- Searle, S. R. (1971). *Linear models*, John Wiley & Sons, New York.

Derivation of error sum of squares of two stage nested designs and its application[†]

Daehak Kim¹

¹Department of Mathematics, Catholic University of Daegu

Received 23 October 2013, revised 15 November 2013, accepted 20 November 2013

Abstract

The analysis of variance for randomized block design or two way classification data is well known. In this paper, particularly, we considered two stage nested design in which the levels of one factor is not identical for different levels of another factor. We investigate the structural properties of two stage nested design and the properties of error sum of squares for random effect model. For the application of two way nested design, we consider two-period crossover design which is used commonly for the equivalence test to bio-similar product. The confidence interval estimation of the difference of two population means in the crossover design is discussed based on statistical package SPSS.

Keywords: Confidence interval estimation, crossover design, error sum of squares, mean squared error, nested design, random effect model.

[†] This work was supported by research grants from the Catholic University of Daegu in 2102

¹ Professor, Department of Mathematics, Catholic University of Daegu, Kyungsan 712-702, Korea.
E-mail: dhkim@cu.ac.kr