Journal of the Korean Data & Information Science Society
2009, **20**(2), 377–385

한국데이터정보과학회지

체중감량자료에 대한 적정 공분산형태모형 산출에 관한 실증연구†

조진남[1]

동덕여자대학교 정보대학 데이터정보전공

접수 2009년 1월 18일, 수정 2009년 3월 16일, 게재확정 2009년 3월 20일

요 약

서울시에 거주하는 25명의 여대생을 대상으로 식이요법에 대한 체중감량 효과를 비교하고자 식이요법과 운동을 병행하는 그룹과 식이요법만 실시하는 그룹으로 나누어서, 8주간에 걸쳐서 2주 간격으로 측정을 실시하여 각 그룹별로 4회 반복측정실험자료를 얻었다. 이 실험자료를 바탕으로 반복측정에 관한 혼합모형을 이용하여 분석한 결과 처리별 Toeplitz 공분산형태가 가장 적절한 모형으로 선택되었다. 처리별 Toeplitz 공분산형태를 가정하여 분석한 결과, 식이요법 이전의 체중값과 시간의 차이에 따른 효과는 대단히 유의하지만, 처리와 시간 간의 교호작용은 유의하지 않은 것으로 나타났으며, 식이요법과 운동을 병행한 그룹의 학생들이 식이요법만 섭취한 그룹의 학생들보다 좀더 효과적인 체중감량의 효과가 있었음이 판명되었다.

주요용어: 공분산형태모형, 반복측정 데이터, 우도비검정, 혼합모형.

1. 서론

의학, 약학 및 생물학 분야 등에서 사람이나 동물 등을 대상으로 실험을 실시하는 경우, 같은 개체를 대상으로 반복적으로 값을 측정하였을 때 얻어지는 자료를 반복측정자료라고 한다. 반복측정자료의 측정시점 간에는 시간별로 독립이 아닌 일정 관계의 연관성이 존재하며, 이러한 관계는 다양한 형태의 공분산 구조로 표현된다. 공분산구조의 다양한 변화양상에 대한 적절한 정보를 얻기 위해서는 혼합모형으로 가정하여 분석하는 것이 모수모형보다 각 추정치의 분산이 적은 안정된 계수들의 분석결과를 얻을 수 있으며, 각 시점 별로도 처리효과를 정확하게 도출해 낼 수 있다. 혼합모형을 이용하여 반복측정자료를 분석한 문헌들은 Brown (1999), Diggle (1989), Fitzmaurice (2004), Verbeke (2000) 등이 있다.

이 논문에서는 체중감량에 대한 반복측정 실험을 실시한 후, 혼합모형으로 분석하여 적절한 공분산 구조모형을 찾고자 한다. 식이요법과 운동이 체중감량에 미치는 효과에 관련된 논문은 장은재 (1997), 장은재 (1999) 등이 있으며, 구체적인 실험내용은 다음과 같다.

체중감량에 대한 식이요법과 운동을 병행하는 처리(이하 처리 A)와 식이요법만 실시하는 처리(이하 처리 B)에 대한 효과를 비교하고자 서울에 거주하는 25명의 여대생을 대상으로 무작위로 2개의 그룹으로 나누어서 8주간에 걸쳐서 4회 반복측정실험을 실시하였다. 이 실험에서 무작위로 배당한 결과 14명의 학생들이 처리 A 그룹에 배당되었으며, 나머지 11명의 학생들은 처리 B 그룹에 배당되었으며, 2주

† 이 논문은 2007년도 동덕여자대학교 학술연구비 지원에 의하여 수행된 것임.
[1] (136-714) 서울시 성북구 월곡동 23-1, 동덕여자대학교 정보대학 데이터정보전공, 교수.
E-mail: jinnam@dongduk.ac.kr

간격으로 4회 실험데이터를 얻었다. 반응변수는 2주, 4주, 6주, 8주 후의 학생들의 반복측정된 체중값이며, 설명변수로는 기준값(실험 실시 전 체중값), 처리, 시점, 그리고 처리와 시점간의 교호작용을 고려하였으며, 혼합모형으로 분석하기 위한 구체적인 실험데이터는 부록에 수록하였다.

따라서 이 논문에서는 다양한 공분산형태 중에서 이 실험에 가장 적합한 공분산형태를 찾은 후, 적정 공분산구조를 지닌 혼합모형을 바탕으로 기준값, 처리, 시간, 처리와 시간 간의 교호작용의 효과와 처리들 간의 체중감량 효과를 통계적으로 분석하고자 한다.

2. 이론적 배경

2.1. 혼합모형의 설정

반복측정 자료의 혼합모형은 다음과 같다.

$$y_i = \mu + \alpha_1 x_{i1} + \alpha_2 x_{i2} + \cdots + \alpha_p x_{ip} + \beta_1 z_{i1} + \beta_2 z_{i2} + \cdots + \beta_q z_{iq} + e_i,\ i = 1, 2, \cdots, n \tag{2.1}$$

n 개의 관찰치가 주어졌을 때 행렬형태로 표시하면 다음과 같다.

$$\underline{y} = X\underline{\alpha} + Z\underline{\beta} + \underline{e} \tag{2.2}$$

여기서 $\underline{y} = (y_1, y_2, \cdots, y_n)'$는 $n \times 1$의 관찰치 벡터, X는 $n \times p$크기의 모수인자들의 디자인 행렬(design matrix), $\underline{\alpha} = (\mu, \alpha_1, \alpha_2, \cdots, \alpha_p)'$는 모수인자 계수 벡터, Z는 $n \times q$크기의 변량인자 들의 디자인 행렬, $\underline{\beta} = (\beta_1, \beta_2, \cdots, \beta_q)'$는 평균 $\underline{0}$, 분산공분산행렬 G인 다변량 정규분포를 하는 $q \times q$변량인자 계수 벡터, 그리고 $\underline{e}$는 평균 $\underline{0}$, 분산공분산행렬 R인 다변량 정규분포를 하는 $n \times 1$ 크기의 오차벡터이다. 따라서 $\underline{y}$의 분산공분산행렬 V 는 다음식과 같다.

$$V = var\left(\underline{y}\right) = ZGZ' + R \tag{2.3}$$

이 모형에서 모수를 모수인자 계수 $\underline{\alpha}$ 와 변량인자 계수 $\underline{\beta}$ 를 추정하기 위한 우도함수(likelihood function)는 다음과 같다 Brown (1999).

$$L = (2\pi)^{-(1/2)n} \left|V\right|^{-(1/2)} \exp\left[-\frac{1}{2}(Y - X\alpha)' V^{-1} (Y - X\alpha)\right] \tag{2.4}$$

$$\log(L) = -\frac{n}{2}\log(2\pi) - \frac{1}{2}\left[\log\left|V\right| + (Y - X\alpha)' V^{-1} (Y - X\alpha)\right] \tag{2.5}$$

이 우도함수를 최대화시키는 제한된 최우추정법(REML : Restricted Maximum Likelihood) 에 의하여 모수인자 계수 $\underline{\alpha}$ 와 변량인자 계수 $\underline{\beta}$ 는 다음과 같이 추정된다.

$$\hat{\alpha} = \left(X'V^{-1}X\right)^{-1} X'V^{-1}\underline{y} \tag{2.6}$$

$$\hat{\beta} = GZ'V^{-1}\left(y - X\hat{\alpha}\right) \tag{2.7}$$

이 때 $\underline{\hat{\alpha}}$ $\underline{\hat{\beta}}$ 의 분산은 다음과 같다.

$$var\left(\hat{\alpha}\right) = \left(X'V^{-1}X\right)^{-1}$$

$$var\left(\hat{\beta}\right) = GZ'\left[V^{-1} - V^{-1}X\left(X'V^{-1}X\right)^{-1}X'V^{-1}\right]ZG$$

2.2. 반복측정자료의 공분산형태

반복측정자료에서 개체들 간에는 관련이 없으나, 각 개체의 시점들 간에는 연관성이 존재한다 따라서 $n \times n$크기의 오차의 공분산행렬 R은 각 개체내의 v개의 시점들 간의 공분산행렬 R_i 을 대각항으로 표시하며, 비대각항은 개체들 간의 연관성이 존재하지 않으므로 0행렬로 이루어진다. 이 때 개체의 공분산행렬 R_i 는 각 개체의 $v \times v$크기의 측정시점들 간의 공분산행렬이며, 따라서 전체 공분산행렬 R

$$R = \begin{bmatrix} R_1 & 0 & 0 & \cdots \\ 0 & R_2 & 0 & \cdots \\ 0 & 0 & R_3 & \cdots \\ \vdots & \vdots & \vdots & \ddots \end{bmatrix} = R_i \otimes I \tag{2.8}$$

로 표시할 수 있으며, 개체의 공분산행렬 R_i가 어떤 형태를 가지느냐에 따라서 결정된다. R_i 는 처리에 관계없는 개체의 공분산행렬과 처리별로 다른 개체의 공분산행렬로 구분된다.

처리에 관계없는 개체의 공분산행렬은 모든 개체의 공분산행렬 R_i가 같은 경우이며, 단순형태 (simple), 복합대칭성(confound symmetry), AR(1), Toeplitz, 멱형태(power), 지수형태(exponential), 가우시안(Gaussian), 일반적 형태(unstructured) 등의 다양한 형태로 구성되며, 처리별로 다른 개체의 공분산행렬은, 처리별 단순형태, 처리별 복합대칭성, 처리별 AR(1), 처리별 Toeplitz 형태 등이 있다. 이 논문에서는 단순형태, 복합대칭성, AR(1), Toeplitz, Toeplitz(2)와 처리별 단순형태, 처리별 복합대칭성, 처리별 AR(1), 처리별Toeplitz, 처리별 Toeplitz(2)의 총 10가지 경우를 언급하고자 한다 Brown (1999).

처리에 관계없는 개체의 공분산행렬에서 단순형태는 각 시점간에 반복측정된 값들은 서로 독립인 경우이며, 복합대칭성은 측정시점 간의 간격에 관계없이 다른 시점간의 상관관계는 ρ 이며, AR(1)은 1차 자기회귀(1^{st} order autoregressive)로 시점간의 차이가 벌어질수록 관련정도가 낮아지는 형태이다. 또한 Toeplitz 형태 는 반복측정 시기의 간격에 따라 일정 정도의 연관성을 보이는 형태이며, Toeplitz(2) 형태는 측정간격이 하나차이일때만 관련성이 있으며 측정간격이 2 이상일 때는 관련이 없는 형태이다. 따라서 처리에 관계없는 개체의 공분산행렬과 처리별로 다른 공분산행렬을 표3.1에 정리하였다.

2.3. 모형의 적합성 검정과 유의성 검정

모형의 적합성 여부는 식 (2.4)의 우도값이 큰 모형이 작은 모형에 비해서 상대적으로 적합한 모형으로 판단되며, 기준으로는 $-2\log(L)$을 사용하며, 이 값이 작을수록 적합한 모형이라고 판단된다.

유의성 검정은 어떤 공분산 형태의 모수가 다른 공분산형태의 일부분으로 포함되는(Nested)경우에 다른 공분산형태의 포함되지 않는 모수의 우도비검정(LRT)을 이용하여 적합한 공분산구조모형을 찾을 수 있다.

3. 식이요법 실험자료의 실증분석

3.1. 공분산행렬의 추정

개체의 공분산 행렬 R_i는 처리에 관계없는 공분산행렬과 처리별로 다른 공분산행렬로 구분되며, 공분산행렬의 시점은 식이요법 실시후 2주, 4주, 6주, 8주의 4개 시점이다. 이 논문에서는 다양한 공분산행렬형태 중에서 처리에 관계없는 일정형태의 공분산행렬의 5가지 형태와 처리별로 공분산이 다른 경우의 5가지 형태를 선택하여 각각의 공분산 형태를 추정하여 표3.1에 정리하였다.

표 3.1 공분산행렬 형태와 추정 결과

(i)단순형태

$$R_i = \hat{\sigma}^2 I = 3.553 \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix}$$

(ii)복합대칭성

$$R_i = \hat{\sigma}^2 \begin{bmatrix} 1 & \hat{\rho} & \hat{\rho} & \hat{\rho} \\ \hat{\rho} & 1 & \hat{\rho} & \hat{\rho} \\ \hat{\rho} & \hat{\rho} & 1 & \hat{\rho} \\ \hat{\rho} & \hat{\rho} & \hat{\rho} & 1 \end{bmatrix} = 3.643 \begin{bmatrix} 1 & 0.751 & 0.751 & 0.751 \\ 0.751 & 1 & 0.751 & 0.751 \\ 0.751 & 0.751 & 1 & 0.751 \\ 0.751 & 0.751 & 0.751 & 1 \end{bmatrix}$$

(iii)AR(1)

$$R_i = \hat{\sigma}^2 \begin{bmatrix} 1 & \hat{\rho} & \hat{\rho}^2 & \hat{\rho}^3 \\ \hat{\rho} & 1 & \hat{\rho} & \hat{\rho}^2 \\ \hat{\rho}^2 & \hat{\rho} & 1 & \hat{\rho} \\ \hat{\rho}^3 & \hat{\rho}^2 & \hat{\rho} & 1 \end{bmatrix} = 3.680 \begin{bmatrix} 1 & 0.871 & 0.758 & 0.660 \\ 0.871 & 1 & 0.871 & 0.758 \\ 0.758 & 0.871 & 1 & 0.871 \\ 0.660 & 0.758 & 0.871 & 1 \end{bmatrix}$$

(iv)Toeplitz

$$R_i = \hat{\sigma}^2 \begin{bmatrix} 1 & \hat{\rho}_1 & \hat{\rho}_2 & \hat{\rho}_3 \\ \hat{\rho}_1 & 1 & \hat{\rho}_1 & \hat{\rho}_2 \\ \hat{\rho}_2 & \hat{\rho}_1 & 1 & \hat{\rho}_1 \\ \hat{\rho}_3 & \hat{\rho}_2 & \hat{\rho}_1 & 1 \end{bmatrix} = 3.688 \begin{bmatrix} 1 & 0.872 & 0.703 & 0.503 \\ 0.872 & 1 & 0.872 & 0.703 \\ 0.703 & 0.872 & 1 & 0.872 \\ 0.503 & 0.703 & 0.872 & 1 \end{bmatrix}$$

(v)Toeplitz(2)

$$R_i = \hat{\sigma}^2 \begin{bmatrix} 1 & \hat{\rho}_1 & 0 & 0 \\ \hat{\rho}_1 & 1 & \hat{\rho}_1 & 0 \\ 0 & \hat{\rho}_1 & 1 & \hat{\rho}_1 \\ 0 & 0 & \hat{\rho}_1 & 1 \end{bmatrix} = 2.912 \begin{bmatrix} 1 & 0.556 & 0 & 0 \\ 0.556 & 1 & 0.556 & 0 \\ 0 & 0.556 & 1 & 0.556 \\ 0 & 0 & 0.556 & 1 \end{bmatrix}$$

(vi)처리별 단순형태

$$R_{i(A)} = \hat{\sigma}_A^2 I = 4.558 \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix}, \quad R_{i(B)} = \hat{\sigma}_B^2 I = 2.246 \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix}$$

(vii)처리별 복합대칭성

$$R_{i(A)} = \hat{\sigma}_A^2 \begin{bmatrix} 1 & \hat{\rho}_A & \hat{\rho}_A & \hat{\rho}_A \\ \hat{\rho}_A & 1 & \hat{\rho}_A & \hat{\rho}_A \\ \hat{\rho}_A & \hat{\rho}_A & 1 & \hat{\rho}_A \\ \hat{\rho}_A & \hat{\rho}_A & \hat{\rho}_A & 1 \end{bmatrix} = 4.666 \begin{bmatrix} 1 & 0.745 & 0.745 & 0.745 \\ 0.745 & 1 & 0.745 & 0.745 \\ 0.745 & 0.745 & 1 & 0.745 \\ 0.745 & 0.745 & 0.745 & 1 \end{bmatrix}$$

$$R_{i(B)} = \hat{\sigma}_B^2 \begin{bmatrix} 1 & \hat{\rho}_B & \hat{\rho}_B & \hat{\rho}_B \\ \hat{\rho}_B & 1 & \hat{\rho}_B & \hat{\rho}_B \\ \hat{\rho}_B & \hat{\rho}_B & 1 & \hat{\rho}_B \\ \hat{\rho}_B & \hat{\rho}_B & \hat{\rho}_B & 1 \end{bmatrix} = 2.309 \begin{bmatrix} 1 & 0.764 & 0.764 & 0.764 \\ 0.764 & 1 & 0.764 & 0.764 \\ 0.764 & 0.764 & 1 & 0.764 \\ 0.764 & 0.764 & 0.764 & 1 \end{bmatrix}$$

(viii)처리별 AR(1)

$$R_{i(A)} = \hat{\sigma}_A^2 \begin{bmatrix} 1 & \hat{\rho}_A & \hat{\rho}_A^2 & \hat{\rho}_A^3 \\ \hat{\rho}_A & 1 & \hat{\rho}_A & \hat{\rho}_A^2 \\ \hat{\rho}_A^2 & \hat{\rho}_A & 1 & \hat{\rho}_A \\ \hat{\rho}_A^3 & \hat{\rho}_A^2 & \hat{\rho}_A & 1 \end{bmatrix} = 4.472 \begin{bmatrix} 1 & 0.891 & 0.793 & 0.706 \\ 0.891 & 1 & 0.891 & 0.793 \\ 0.793 & 0.891 & 1 & 0.891 \\ 0.706 & 0.793 & 0.891 & 1 \end{bmatrix}$$

$$R_{i(B)} = \hat{\sigma}_B^2 \begin{bmatrix} 1 & \hat{\rho}_B & \hat{\rho}_B^2 & \hat{\rho}_B^3 \\ \hat{\rho}_B & 1 & \hat{\rho}_B & \hat{\rho}_B^2 \\ \hat{\rho}_B^2 & \hat{\rho}_B & 1 & \hat{\rho}_B \\ \hat{\rho}_B^3 & \hat{\rho}_B^2 & \hat{\rho}_B & 1 \end{bmatrix} = 2.574 \begin{bmatrix} 1 & 0.820 & 0.673 & 0.552 \\ 0.820 & 1 & 0.820 & 0.673 \\ 0.673 & 0.820 & 1 & 0.820 \\ 0.552 & 0.673 & 0.820 & 1 \end{bmatrix}$$

(ix)처리별 Toeplitz

$$R_{i(A)} = \hat{\sigma}_A^2 \begin{bmatrix} 1 & \hat{\rho}_{1(A)} & \hat{\rho}_{2(A)} & \hat{\rho}_{3(A)} \\ \hat{\rho}_{1(A)} & 1 & \hat{\rho}_{1(A)} & \hat{\rho}_{2(A)} \\ \hat{\rho}_{2(A)} & \hat{\rho}_{1(A)} & 1 & \hat{\rho}_{1(A)} \\ \hat{\rho}_{3(A)} & \hat{\rho}_{2(A)} & \hat{\rho}_{1(A)} & 1 \end{bmatrix} = 4.295 \begin{bmatrix} 1 & 0.874 & 0.583 & 0.195 \\ 0.874 & 1 & 0.874 & 0.583 \\ 0.583 & 0.874 & 1 & 0.874 \\ 0.195 & 0.583 & 0.874 & 1 \end{bmatrix}$$

$$R_{i(B)} = \hat{\sigma}_B^2 \begin{bmatrix} 1 & \hat{\rho}_{1(B)} & \hat{\rho}_{2(B)} & \hat{\rho}_{3(B)} \\ \hat{\rho}_{1(B)} & 1 & \hat{\rho}_{1(B)} & \hat{\rho}_{2(B)} \\ \hat{\rho}_{2(B)} & \hat{\rho}_{1(B)} & 1 & \hat{\rho}_{1(B)} \\ \hat{\rho}_{3(B)} & \hat{\rho}_{2(B)} & \hat{\rho}_{1(B)} & 1 \end{bmatrix} = 2.308 \begin{bmatrix} 1 & 0.797 & 0.730 & 0.751 \\ 0.797 & 1 & 0.797 & 0.730 \\ 0.730 & 0.797 & 1 & 0.797 \\ 0.751 & 0.730 & 0.797 & 1 \end{bmatrix}$$

(x)처리별 Toeplitz(2)

$$R_{i(A)} = \hat{\sigma}_A^2 \begin{bmatrix} 1 & \hat{\rho}_{1(A)} & 0 & 0 \\ \hat{\rho}_{1(A)} & 1 & \hat{\rho}_{1(A)} & 0 \\ 0 & \hat{\rho}_{1(A)} & 1 & \hat{\rho}_{1(A)} \\ 0 & 0 & \hat{\rho}_{1(A)} & 1 \end{bmatrix} = 3.528 \begin{bmatrix} 1 & 0.574 & 0 & 0 \\ 0.574 & 1 & 0.574 & 0 \\ 0 & 0.574 & 1 & 0.574 \\ 0 & 0 & 0.574 & 1 \end{bmatrix}$$

$$R_{i(B)} = \hat{\sigma}_B^2 \begin{bmatrix} 1 & \hat{\rho}_{1(B)} & 0 & 0 \\ \hat{\rho}_{1(B)} & 1 & \hat{\rho}_{1(B)} & 0 \\ 0 & \hat{\rho}_{1(B)} & 1 & \hat{\rho}_{1(B)} \\ 0 & 0 & \hat{\rho}_{1(B)} & 1 \end{bmatrix} = 2.053 \begin{bmatrix} 1 & 0.511 & 0 & 0 \\ 0.511 & 1 & 0.511 & 0 \\ 0 & 0.511 & 1 & 0.511 \\ 0 & 0 & 0.511 & 1 \end{bmatrix}$$

3.2. 적합 공분산행렬의 산출과정

3.2.1. 공분산형태모형의 적합성 검정

공분산형태의 적합성여부를 판단하기 위하여 먼저 개별 공분산행렬에 대한 $-2\log(L)$ 값을 산출하였다. 표3.2에서 알 수 있듯이 Toeplitz 의 공분산구조행렬을 가지는 모형이 가장 적합한 것을 알 수 있다.

표 3.2 형태별 공분산행렬의 우도에 관련된 값

공분산행렬 형태	모수의 수	$-2\log(L)$
(i)단순형태	1	402.3
(ii)복합대칭성	2	334.7
(iii)AR(1)	2	306.6
(iv)Toeplitz	4	302.5
(v)Toeplitz(2)	2	342.5
(vi)처리별 단순형태	2	396.9
(vii)처리별 복합대칭성	4	328.6
(viii)처리별 AR(1)	4	305.2
(ix)처리별 Toeplitz	8	280.4
(x)처리별 Toeplitz(2)	4	339.4

3.2.2. 공분산 행렬의 모수에 대한 유의성 검정

공분산행렬의 모수에 대한 유의성 검정을 실시함에 있어서 포함하는 모형의 우도값(L_1)과 포함되는 모형의 우도값(L_2)을 구한 후, $\chi^2 = 2(\log(L_1) - \log(L_2))$ 의 우도비 검정을 실시한다. 이 때 자유도는 두 모형의 비교에서 포함되지 않는 모수의 갯수이다.

표3.3은 포함하는 모형과 포함되는 모형을 산정하여 우도비 검정을 실시한 결과이다.

표3.3에서 알 수 있듯이 복합대칭성, AR(1), Toeplitz(2) 의 공분산행렬은 단순형태의 공분산행렬보다 더 적합하며, Toeplitz 는 복합대칭성, Toeplitz(2) 보다 더 적합한 모형으로 나타났다. 처리별 AR(1)과 처리별 Toeplitz(2)의 공분산행렬은 AR(1)과 Toeplitz(2) 의 공분산행렬보다 다소 유의하게 나타났으며, 처리별 Toeplitz 의 공분산행렬은 Toeplitz 의 공분산행렬보다 대단히 유의한 것으로 판명되었다. 따라서 종합적으로 AR(1)과 처리별 Toeplitz 를 선택할 수 있으나 보다 유의한 처리별 Toeplitz 의 공분산행렬을 최종모형으로 선택한다.

표 3.3 공분산행렬 모수의 우도비 검정

포함하는 모형	포함되는 모형	자유도	χ^2	유의확률
(ii)복합대칭성	(i)단순형태	1	67.6	< 0.0001
(iii)AR(1)	(i)단순형태	2	95.7	< 0.0001
(iv)Toeplitz	(ii)복합대칭성	2	32.2	< 0.0001
(iv)Toeplitz	(v)Toeplitz(2)	2	40.0	< 0.0001
(v)Toeplitz(2)	(i)단순형태	1	59.8	< 0.0001
(vi)처리별 단순형태	(i)단순형태	1	73.7	< 0.0001
(vii)처리별 복합대칭성	(ii)복합대칭성	2	6.1	0.0474
(viii)처리별 AR(1)	(iii)AR(1)	2	1.4	0.4966
(ix)처리별 Toeplitz	(iv)Toeplitz	4	22.1	0.0002
(x)처리별 Toeplitz(2)	(v)Toeplitz(2)	2	3.1	0.28

3.2.3. 적정 공분산행렬모형 산출

공분산행렬의 적합성 검정과 유의성 검정결과, 처리별 Toeplitz 공분산행렬 형태가 최적모형으로 선정된다. 처리별 Toeplitz 의 공분산행렬 R_i는 다음과 같다.

$$R_{i(A)} = \hat{\sigma}_A^2 \begin{bmatrix} 1 & \hat{\rho}_{1(A)} & \hat{\rho}_{2(A)} & \hat{\rho}_{3(A)} \\ \hat{\rho}_{1(A)} & 1 & \hat{\rho}_{1(A)} & \hat{\rho}_{2(A)} \\ \hat{\rho}_{2(A)} & \hat{\rho}_{1(A)} & 1 & \hat{\rho}_{1(A)} \\ \hat{\rho}_{3(A)} & \hat{\rho}_{2(A)} & \hat{\rho}_{1(A)} & 1 \end{bmatrix} = 4.2948 \begin{bmatrix} 1 & 0.8739 & 0.5833 & 0.1947 \\ 0.8739 & 1 & 0.8739 & 0.5833 \\ 0.5833 & 0.8739 & 1 & 0.8739 \\ 0.1947 & 0.5833 & 0.8739 & 1 \end{bmatrix}$$

$$R_{i(B)} = \hat{\sigma}_B^2 \begin{bmatrix} 1 & \hat{\rho}_{1(B)} & \hat{\rho}_{2(B)} & \hat{\rho}_{3(B)} \\ \hat{\rho}_{1(B)} & 1 & \hat{\rho}_{1(B)} & \hat{\rho}_{2(B)} \\ \hat{\rho}_{2(B)} & \hat{\rho}_{1(B)} & 1 & \hat{\rho}_{1(B)} \\ \hat{\rho}_{3(B)} & \hat{\rho}_{2(B)} & \hat{\rho}_{1(B)} & 1 \end{bmatrix} = 2.3077 \begin{bmatrix} 1 & 0.7972 & 0.7301 & 0.7509 \\ 0.7972 & 1 & 0.7972 & 0.7301 \\ 0.7301 & 0.7972 & 1 & 0.7972 \\ 0.7509 & 0.7301 & 0.7972 & 1 \end{bmatrix}$$

처리별 분산에서 A 처리일 때의 분산값 $\hat{\sigma}_A^2 = 4.2948$는 B 처리일 때의 분산값 $\hat{\sigma}_A^2 = 2.3077$ 보다 거의 2배나 되며, A 처리에서의 측정시점 간의 상관관계는 시점이 가까울수록 상관관계가 높고, 시점차이가 멀수록 상관관계가 낮은 반면, B 처리에서의 측정시점 간의 상관관계는 시점의 차이에 관계없이 073-0.79 로 대체로 비슷한 상관관계를 지닌다.

3.3. 최적모형의 처리효과에 대한 통계적 분석

모수인자의 효과에 대한 분산분석표는 표3.4와 같다. 식이요법을 복용하기 전의 체중값은 기준값(base line)으로 대단히 유의하며, 4개 시점들 간에도 상당히 유의함을 보여주지만, 처리와 시간 간의 교호작용은 거의 존재하지 않음을 알 수 있다.

표 3.4 모수인자의 효과에 대한 분산분석표

요인	분자의 자유도	본모의 자유도	F 값	유의확률
처리전 체중값	1	22	284.56	< 0.0001
처리	1	22	3.58	0.0719
시간	3	69	13.26	< 0.0001
처리 × 시간	3	69	1.16	0.3301

A 처리와 B 처리 효과의 차이에 대한 유의확률은 0.072 로 다소 차이가 있으며, 식이요법과 운동을 병행했을 때(처리 A)의 체중감소가 식이요법만 실시했을 때(처리 B)의 감소보다 1.115 Kg 의 효과가 더 있음을 알 수 있다.

표 3.5 모수인자의 효과에 대한 분산분석표

처리의 차이	추정치	표준오차	t 값	유의확률
A-B	-1.115	0.590	-1.89	0.072

4. 결론

체중감소를 위한 식이요법에 관련된 반복측정실험자료를 바탕으로 혼합모형의 분석기법을 적용하여 분석한 결과, 처리별 Toeplitz 가 가장 적합한 공분산행렬모형으로 채택되었으며, 이 경우 처리별 분산은 A 처리일 때의 분산값이 B 처리일 때의 분산값 보다 거의 2배 가까이 되며, A 처리에서의 측정시점 간의 상관관계는 시점차이가 가까울수록 상관관계가 높고, 멀수록 상관관계가 낮은 반면, B 처리에서의 측정시점 간의 상관관계는 시점의 차이에 관계없이 073-0.79 로 대체로 비슷한 상관관계를 지닌다.

처리별 Toeplitz 공분산행렬 모형을 채택하여 분석한 결과, 식이요법을 복용하기 전 체중값과 시간 효과는 대단히 유의한 결과를 보여주었으나, 처리와 시간 간의 교호작용은 거의 없는 것으로 나타났으며, 식이요법과 운동을 병행했을 때의 처리가 식이요법만 실시했을 때의 처리보다 체중감소 효과가 1.115 Kg 더 있음을 알 수 있다.

부록: 체중감량 실험데이터

개체	처리	키	사전값	y2	y4	y6	y8
1	A	165.5	64.9	62.8	63.1	63.6	63
2	A	154.8	59.1	57	55.3	55.4	54.9
3	A	158.1	62.7	60.9	59.1	58.7	57.5
4	A	161.6	61	58.8	58.1	59.3	58.2
5	A	158	57.5	56.9	55.1	54.5	53.6

6	A	168.3	75.2	73.2	70.3	70.2	69.9
7	A	163.2	72.1	69.9	65.5	63.6	61.2
8	A	153.5	57.5	55.8	55.3	56	55.2
9	A	154.1	69.3	67.9	67	67.5	66.8
10	A	152.2	72.7	72.7	70.6	71.2	69.2
11	A	161.1	70.1	68.1	66.6	66.3	66.4
12	A	148.6	51.3	51.2	50.9	51.8	52.1
13	A	157.7	80.2	80.7	81.3	82.1	80.5
14	A	154.9	59.4	58	56.1	56.1	56.2
15	B	159.5	63.9	61.9	60.6	60.3	58.5
16	B	166	65.4	63.4	63.2	63.4	62.9
17	B	156.8	55	56.7	54.7	57	57.9
18	B	155.5	56.3	54.5	55.3	54	53.9
19	B	160.6	73.7	72.4	70.6	70.9	70.7
20	B	157.5	60.7	61.4	60.3	59.8	60.5
21	B	154.1	66.4	66.4	65.6	66.9	64.5
22	B	171.4	72.7	73.1	71.5	72.1	71.9
23	B	157.6	58.3	57	56.4	56.3	56.3
24	B	161.7	62.5	61.1	60.8	59.5	59.5
25	B	157.5	69.1	68.6	67.8	67.6	67.7

참고문헌

장은재, 조진남, 황종현 (1997). 한국 여대생의 체지방 측정을 통한 측정기기들 간의 비교 연구. <한국식품영양과학회지>, **26**, 514-520.

장은재, 임경아, 한용봉 (1999). 영양교육이 체중조절프로그램에 미치는 효과에 관한연구. <한국식품영양학회지>, **12**, 177-183.

Brown, H. and Kempton, R. A. (1994). The application of REMLin clinical trials. *Statistics in Medicine*, **16**, 1601-1617.

Brown, H. and Prescott, R. (1999). *Applied mixed models in medicine*, John Wiley & Sons Inc, New York.

Diggle, P. J. (1989). Testing for random dropouts in repeated measurement data. *Biometrics*, **43**, 1255-1258.

Fitzmaurice, G. M., Laird, N. M. and Ware, J. H. (2004). *Applied longitudinal analysis*, John Wiley & Sons Inc, New York.

Hand, D. and Crowder, M. (1996). *Practical longitudinal data analysis,* Chapman & Hall, London.

Hartley, H. O. and Rao, J. N. K. (1967). Maximum likelihood estimation for the mixed analysis of variance model. *Biometrika*, **54**, 93-108.

Hinkelman, K. and Kempthorne, O. (1994). *Design and analysis of experimenrts*, Vol. 1, John Wiley & Sons Inc, New York.

Littell, R. C., Milliken, G. A., Stroup, W. W. and Wolfinger, R. D. (1996). *SAS system for mixed Models*, SAS Institute Inc., N.C., U.S.A.

Longford, N. T. (1993). *Random coefficient models*, Oxford University Press, Oxford.

Rao, C. R. (1971). Estimation of variance and covariance components-MINQUE theory. *Journal of Multivariate Analysis*, **1**, 257-275.

Rao, C. R. (1972). Estimation of variance and covariance components in linear models. *Journal of the American Statistical Association*, **67**, 112-117.

Satterthwaite, F. E. (1946). An approximate distribution of estimates of variance components. *Biometrics Bulletin*, **2**, 110-114.

Verbeke, G. and Molenberghs, G. (2000). *Linear mixed models for longitudinal data*, Springer Verlag, New York.

Journal of the Korean Data & Information Science Society
2009, **20**(2), 377–385

한국데이터정보과학회지

An empirical study on the selection of the optimal covariance pattern model for the weight loss data†

Jinnam Jo[1]

Department of Data Information Science, Dongduk Women's University

Received 18 January 2009, revised 16 March 2009, accepted 20 March 2009

Abstract

Twenty five female students in Seoul participated and were divided into two groups in the experiment of weight loss effect of two treatments. Fourteen students(Treatment A group), randomly chosen from the students, had fed on diet foods and exercised over 8 weeks, and the remaining students(Treatment B group) had fed on diet foods only for the same periods. Weights of 25 students had been measured repeatedly four times at an interval of two weeks during 8 weeks, It resulted from mixed model analysis of repeated measurements data that separate Toeplitz pattern for each treatment group was selected as the optimal covariance pattern. Based upon the optimal covariance pattern model, the baseline effect and time effect were found to be highly significant, but the treatment-time interaction effect was found to be insignificant. Finally, the students with diet foods and exercises were more effective in losing weight than the students with only diet foods were.

Keywords: Covariance pattern model, likelihood ratio test, mixed model, repeated measures data.

† This research was supported by the Dongduk Women's University Research Grants 2007.

[1] Professor, Department of Data Information Science, Dongduk Women's University, Seoul 136-714, Korea. E-mail: jinnam@dongduk.ac.kr