Journal of the Korean Data & Information Science Society
2011, **22**(2), 287–296

한국데이터정보과학회지

변량계수모형을 이용한 체지방 실험자료에 관한 통계적 분석†

조진남[1]

[1]동덕여자대학교 정보대학 정보통계학과

접수 2011년 1월 20일, 수정 2011년 3월 15일, 게재확정 2011년 3월 19일

요 약

36명의 여대생을 대상으로 체지방 감소효과에 대한 실험을 실시하였다. 이 실험에서 처리는 매일 섭취하는 식사종류 및 양에 대한 식사일지 작성과 카메라폰으로 찍어 실험관리자에게 전송하여 매주 상담을 받는 것이다. 실험관리자는 체지방 및 관련된 자료를 일주일마다 측정하여 8주간의 반복측정 자료를 얻었다. 이 실험자료를 이용하여 혼합모형의 일종인 변량계수모형을 이용하여 추정 및 유의성 검정을 실시한 결과, 유의한 고정인자들은 처리전 체지방값, 비만지수, 확장기혈압, 총콜레스테롤 및 시간이다. 처리 후 시간에 따른 체지방감소는 2차함수의 관계가 성립된다. 변량인자인 개체효과와 개체와 시간과의 교호작용에서 1차함수의 관계가 존재한다. 처리 후 시간이 지남에 따라 체지방량은 점점 감소하였으며, 실험실시 8주 후에는 평균 2.1kg 감소한 효과가 있음을 보여주었다.

주요용어: 개체효과, 변량계수모형, 분산공분산행렬, 체지방.

1. 서론

여대생들의 체지방 감소를 위한 방법을 찾고자 다음과 같은 실험을 실시하였다. 실험에 참여한 36명의 여대생들은 매일 섭취하는 식사종류 및 식사량에 대하여 식사일지를 작성하며, 식사종류 및 식사량을 카메라폰으로 찍어서 실험관리자에게 전송하였다. 실험관리자는 식사일지와 전송된 사진에 찍혀있는 음식물을 바탕으로 실험참여자들이 하루에 섭취하는 총 칼로리를 계산하였으며, 매1주일마다8주간에 걸쳐서 실험 참여자들에게 체지방 감소에 관한 상담을 실시하였다. 또한 실험자는 8주동안 실험대상자들의 체지방에 관한 반복측정자료를 얻었으며, 허리엉덩이비율 (WHR), 비만지수 (BMI: Body Mass Index), 수축기혈압 (SBP: Systolic Blood Pressure), 확장기혈압 (DBP: Diastolic Blood Pressure), 중성지방 (TRG: Triglyceride mg/dL) 및 총콜레스테롤 (TC: Total Cholesterol mg/d) 값을 측정하였다. 이 실험에서 처리 (treatment)는 식사일지 작성, 카메라폰의 촬영 및 상담실시이며, 36명의 실험대상자 모두에게 똑같은 처리를 실시하였으므로 처리간의 비교는 할 수 없으며, 단지 처리 전과 처리 후 8주간의 변화가 어떤 형태로 이루어 지는지 알고자 한다.

일반적으로 반복측정자료를 분석할 시, 소수의 시점들이 일정하게 고정되어 있는 경우와 임의의 다수의 시점으로 나누어진 경우로 구분할 수 있다. 소수의 반복처리 시점이 고정된 경우 적정 공분산행렬형태모형 (covariance matrix pattern model)를 통하여 분석할 수 있으며, 임의의 시점들인 경우 변량계수모형 (random coefficient model)을 적용하여 실험자료를 분석한다. 해당 자료는 1주마다 측정된 반복측정자료이지만 8주간에 걸쳐서 얻은 자료이므로 소수의 시점들에 적용되는 공분산행렬형태모형보다

† 이 논문은 2010년도 동덕여자대학교 학술연구비 지원에 의하여 수행된 것임.
[1] (136-714) 서울시 성북구 월곡동 23-1, 동덕여자대학교 정보대학 정보통계학과, 교수.
E-mail: jinnam@dongduk.ac.kr

다수의 시점에 적용되는 변량계수모형을 이용하여 분석하고자 한다. 변량계수모형 적용 시 반응변수는 실험실시 후 매 주마다 측정된 체지방량이며, 독립변수는 실험실시 전 체지방값, 허리엉덩이비율, 비만지수, 수축기혈압, 확장기혈압, 중성지방, 총콜레스테롤 및 시간이 포함된다. 변량계수모형에서는 시간에 따른 반응값이 어떤 패턴을 보이는지 규명하는 것이 중요하다.

따라서 이 논문에서는 체지방에 관련된 여러가지 가능한 변량계수모형들을 설정하고, 개별 모형에 대한 유의성 검정을 실시하며, 적합성 판단 기준에 따른 최적의 변량계수모형을 도출하고자 한다. 최적 변량계수모형을 바탕으로 체지방에 유의한 독립변수들과의 관계를 살펴보고 시간과의 관계를 규명해보고자 한다. 이에 따르는 실제 데이터분석은 혼합모형에 관련된 통계프로그램인 SAS 의 PROC MIXED 를 이용하였다 (Littell 등, 1996).

혼합모형에 관련된 문헌들은 Brown과 Kempton (1994), Brown과 Prescott (1999), Choi (2008), Diggle (1989), Frees (2006), Hand와 Crowder (1996) 등이 있다. 반복측정자료를 이용하여 공분산행렬형태를 분석한 문헌으로는 Fitzmaurice 등 (2004), Verbeke와 Molenberghs (2000), 조진남 (2009), 조진남과 백재욱 (2010) 등이 있으며, 변량계수모형을 이용하여 반복측정자료를 분석한 문헌들은 Frees (2006), Hay 등 (1998), Longford (1993), Smyth 등 (1997), 조진남과 백재욱 (2009) 등이 있다.

2. 실험자료의 분석과 변량계수 모형 설정

2.1. 체지방량의 변화 추이와 상관관계

먼저 시간에 따른 체지방량의 변화 추이를 표 2.1과 그림 2.1에서 보듯이, 처리 실시 1주일 후에 체지방량은 처리실시 전보다 0.5kg감소하였으며, 1주부터 7주까지는 시간이 지남에 따라 체지방량이 점점 감소하였으나, 8주에는 평균 0.29kg이 오히려 증가함을 보여주었다. 그러나 전반적으로 실험실시 전보다 실험실시 8주 후에는 평균 2.1kg 감소하였음을 보여주었다

표 2.1 체지방량의 변화 추이 (단위: kg)

시간	평균	표준편차
사전값(0주)	24.31	5.95
1 주	23.81	6.03
2 주	23.61	5.90
3 주	22.98	5.60
4 주	22.58	5.77
5 주	22.40	5.79
6 주	22.21	5.43
7 주	21.91	5.46
8 주	22.20	6.11

다음 행렬은 처리 후 1주에서 8주까지의 체지방량 간의 상관계수행렬이다.

$$r = \begin{bmatrix} 1.00 & 0.980 & 0.976 & 0.973 & 0.972 & 0.953 & 0.936 & 0.952 \\ 0.980 & 1.00 & 0.976 & 0.979 & 0.974 & 0.955 & 0.944 & 0.961 \\ 0.976 & 0.976 & 1.00 & 0.979 & 0.977 & 0.956 & 0.947 & 0.959 \\ 0.973 & 0.979 & 0.979 & 1.00 & 0.989 & 0.973 & 0.965 & 0.980 \\ 0.972 & 0.974 & 0.977 & 0.989 & 1.00 & 0.975 & 0.967 & 0.986 \\ 0.953 & 0.955 & 0.956 & 0.973 & 0.975 & 1.00 & 0.991 & 0.978 \\ 0.936 & 0.944 & 0.947 & 0.965 & 0.967 & 0.991 & 1.00 & 0.974 \\ 0.952 & 0.961 & 0.959 & 0.980 & 0.986 & 0.978 & 0.974 & 1.00 \end{bmatrix}$$

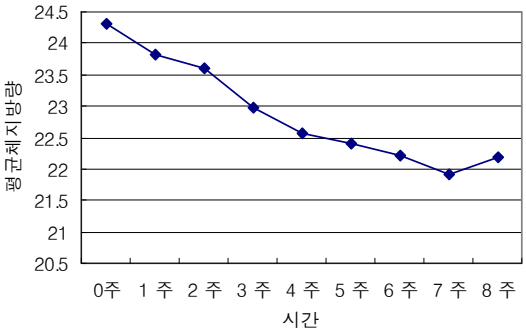

그림 2.1 평균체지방량의 변화추이

이 행렬에서 알 수 있듯이 시점들 간의 상관계수들은 대부분 0.97 이상으로 대단히 높은 상관관계를 보여준다. 따라서 반응변수에 해당하는 체지방량들의 반복측정자료가 독립적이지 않고 매우 밀접한 관련을 보여주므로 혼합모형의 일종인 변량계수모형을 이용하여 최적모형을 산출하고 그에 따르는 모형분석을 실시하고자 한다.

2.2. 변량계수모형 설정

이러한 실험자료를 이용하여 변량계수모형을 설정할 때, 처리가 실시되기 전의 사전 체지방값 (pre), 허리엉덩이비율 (WHR), 비만지수 (BMI), 수축기혈압 (SBP), 확장기혈압 (DBP), 중성지방 (TRG), 총콜레스테롤 (TC) 및 시간은 고정인자가 되며, 개체 및 개체와 관련된 인자들은 변량인자가 된다. 모형설정 과정은 먼저 유의한 고정인자들을 도출하고 난 후, 시간에 관련된 고정인자와 개체와 시간과의 교호작용의 효과를 고려한 변량인자들을 시간에 대하여 축차적으로 설정하였다. 다음은 가능한 경우의 변량계수모형 들이며, 단순한 모형부터 시작하여 단계적으로 각 모형마다 해당 모형에 포함된 변수들에 대한 유의성 검정을 실시하는 과정을 거쳐 최종적으로 적합한 모형을 찾고자 한다.

(모형 1) 고정인자만 포함한 모형

$$\begin{aligned} y_{ij} =& \mu + b(pre)_i + c_1(WHR)_i + c_2(BMI)_i + c_3(SBP)_i + c_4(DBP)_i \\ &+ c_5(TRG)_i + c_6(TC)_i + m_l(time)_{ij} + e_{ij} \end{aligned}$$

여기서 y_{ij}는 i번째 개체의 j번째 시점에서의 반응값, $(pre)_i$는 i번째 개체의 처리 전의 체지방량이며, b는 그에 대한 회귀계수이다. WHR, BMI, SBP, DBP, TRG, TC 는 고정인자들로 취급하며, c_1에서 c_6은 이러한 고정인자들이 반응값에 미치는 회귀계수들이다. $(time)_{ij}$는 i번째 개체의 j번째 시점의 고정인자이며, m_l은 $(time)_{ij}$에 대한 선형 회귀계수이다. 모형 1은 독립변수들이 고정인자들만 포함하므로 고정인자효과모형 (fixed effect model)에서 추정 및 검정을 할 수 있다. 이 모형의 특징은 개별적인 개체들의 효과를 고려하지 않고 전체적으로 시간에 대한 반응값의 변화를 알고자 할 때 이용한다.

모형분석 결과 중성지방 (TRG)은 반응변수인 체지방과는 거의 관련이 없는 것으로 나타났으므로 중성지방을 모형에서 제거시키고, 변량인자인 개체인자를 포함시킨 모형을 채택하여 유의성 여부를 검정하고자 한다.

표 2.2 고정인자효과모형 1의 추정 결과

고정인자	회귀계수(표준오차)	유의확률	G 행렬
INT	9.434(3.649)	0.010	
PRE	0.865(0.018)	<0.001	
WHR	-18.97(5.74)	0.001	
BMI	0.401(0.069)	<0.001	
SBP	-0.029(0.009)	0.002	해당 없음
DBP	-0.087(0.012)	<0.001	
TRG	0.0001(0.0015)	0.924	
TC	0.047(0.004)	<0.001	
TIME	-0.264(0.038)	<0.001	

(모형 2) 유의하지 않은 고정인자는 제거하고 개체인자를 포함한 모형

$$\begin{aligned} y_{ij} =& \mu + b(pre)_i + c_1(WHR)_i + c_2(BMI)_i + c_3(SBP)_i + c_4(DBP)_i + c_6\,(TC) \\ &+ m_l(time)_{ij} + p_i + (pm_l)_i(time)_{ij} + e_{ij} \end{aligned}$$

여기서 p_i는 i번째 개체효과, $(pm_l)_i$는 개체와 시간과의 교호작용효과로 i번째 개체의 시간에 대한 기울기 값을 의미한다. p_i와 $(pm_l)_i$는 변량인자효과 (randon factor effect)의 계수로 서로 독립이 아니며 이변량 정규분포를 따른다. 즉

$$\begin{pmatrix} p_i \\ (pm_l)_i \end{pmatrix} \sim N\,(0, G)$$

이며, 분산공분산행렬 G는 다음과 같다.

$$G = \begin{bmatrix} \sigma_p^2 & \sigma_{p,pm_l} \\ \sigma_{p,pm_l} & \sigma_{pm_l}^2 \end{bmatrix}$$

여기서 σ_p^2, $\sigma_{pm_l}^2$은 각각 p_i, $(pm_l)_i$의 분산성분이며, σ_{p,pm_l}은 p_i, $(pm_l)_i$의 공분산이다. 분산성분이 0 보다 클 때는 의미가 있지만, 0이거나 음수로 추정될 때는 의미가 없으므로 해당 변수는 모형에서 제외시킨다. 분산성분이 0보다 크다는 것은 각 개체들의 회귀선들 사이에 변동이 있다는 것을 의미한다. 따라서 분산성분의 추정치 값이 유의하지 않더라도 양의 값으로 추정되면 해당 인자는 모형에 포함된다 (Brown 과 Prescott, 1999).

모형 2는 고정인자와 변량인자들로 구성되어 있으므로 혼합모형에 해당하는 변량계수모형을 이용하여 분석한다. 모형분석 결과, 허리엉덩이비율 (WHR), 수축기혈압 (SBP)의 유의확률이 각각 0.242, 0.474로 유의하지 않으며, 변량인자 효과들의 분산값은 0보다 크므로 의미가 있다고 판단된다.

표 2.3 변량계수모형 2의 추정 결과

고정인자	회귀계수(표준오차)	유의확률	G 행렬
INT	3.410(6.572)	0.608	
PRE	0.909(0.031)	<0.001	
WHR	-12.34(10.34)	0.242	
BMI	0.349(0.125)	0.009	$\begin{bmatrix} 0.582 & -0.027 \\ -0.027 & 0.060 \end{bmatrix}$
SBP	-0.012(0.017)	0.474	
DBP	-0.050(0.022)	0.031	
TC	0.027(0.008)	0.002	
TIME	-0.264(0.047)	<0.001	

그러므로 유의하지 않은 허리엉덩이비율, 수축기혈압을 제외한 모형을 다시 설정한다.

(모형 3) 유의하지 않은 고정인자는 다시 제거하고 개체인자를 포함한 모형

$$\begin{aligned} y_{ij} =& \mu + b(pre)_i + c_2(BMI)_i + c_4(DBP)_i + c_6(TC)_i + m_l(time)_{ij} \\ &+ p_i + (pm_l)_i(time)_{ij} + e_{ij} \end{aligned}$$

모형 3에서 고정인자계수들의 유의확률이 모두 0.05 이하로 유의하며, 변량인자들의 분산값이 0.529, 0.60으로 0보다 크므로 의미가 있다고 판단되며, 고정인자인 시간에 대하여 일차함수 관계가 성립되며, 개체별 시간과의 관계도 일차함수 관계가 성립됨을 보여준다.

표 2.4 변량계수모형 3의 추정 결과

고정인자	회귀계수(표준오차)	유의확률	G 행렬
INT	-4.473(2.100)	0.041	
PRE	0.916(0.031)	<0.001	
BMI	0.232(0.083)	0.009	$\begin{bmatrix} 0.529 & -0.011 \\ -0.011 & 0.060 \end{bmatrix}$
DBP	-0.050(0.021)	0.023	
TC	0.022(0.007)	0.004	
TIME	-0.264(0.047)	<0.001	

반응변수와 시간과의 관계가 1차함수의 관계가 성립되므로, 다음 모형으로 시간과의 관계가 2차함수의 관계가 성립되는지를 알기 위해 시간에 대한 2차함수 모형을 설정한다.

(모형 4) 시간에 대한 2차함수 모형

$$\begin{aligned} y_{ij} =& \mu + b(pre)_i + c_2(BMI)_i + c_4(DBP)_i + c_6(TC)_i + m_l(time)_{ij} + m_q\,(time)_{ij}^2 \\ &+ p_i + (pm_l)_i(time)_{ij} + (pm_q)_i\,(time)_{ij}^2 + e_{ij} \end{aligned}$$

여기서 m_q는 전체적으로 시간에 대한 반응값의 이차함수에 해당하는 계수이며, 회귀계수 b, c_2, c_4, c_6 그리고 m_l, m_q는 고정인자 효과들이다. $(pm_q)_i$는 개체 별로 시간과의 관계가 2차함수에 해당하는 변량인자 효과이며, p_i, $(pm_l)_i$, $(pm_q)_i$는 다변량 정규분포를 한다. 즉,

$$\begin{pmatrix} p_i \\ (pm_l)_i \\ (pm_q)_i \end{pmatrix} \sim N\,(0, G)$$

이며, 분산공분산행렬 G 는 다음과 같이 구성된다.

$$G = \begin{bmatrix} \sigma_p^2 & \sigma_{p,pm_l} & \sigma_{p,pm_q} \\ \sigma_{p,pm_l} & \sigma_{pm_l}^2 & \sigma_{pm_l,pm_q} \\ \sigma_{p,pm_q} & \sigma_{pm_l,pm_q} & \sigma_{pm_q}^2 \end{bmatrix}$$

여기서 σ_p^2, $\sigma_{pm_l}^2$, $\sigma_{pm_q}^2$는 각각 p_i, $(pm)_i$, $(pm_q)_i$의 분산성분이며, $\sigma_{p,pm}$, σ_{p,pm_q}, σ_{pm_l,pm_q}는 공분산성분들이다. 시간에 대한 2차함수의 모형분석 결과, m_l과 m_q에 대한 회귀계수의 유의확률은 각각 0.001, 0.0002로서 유의적인 관계를 가진다. m_l에 대한 추정치 값은 -0.641, m_q에 대한 추정치 값은 0.042 로 시간이 지남에 따라서 감소하나 감소 폭은 크지 않음을 알 수 있다. 변량인자의 분산값의 추정치 $\hat{\sigma}_p^2 = 0.674$, $\hat{\sigma}_{pm_l}^2 = 0.137$는 0보다 크기 때문에 의미가 있지만, 2차항에 해당하는 분산성분의 추정치 $\hat{\sigma}_{pm_q}^2$는 값이 거의 0이므로 의미가 없다고 판단된다. 따라서 개체별 시간에 대한 2차함수 관계는 없으며, 2차항 이상의 고차항의 관계는 존재하지 않는다.

표 2.5 변량계수모형 4의 추정 결과

고정인자	회귀계수(표준오차)	유의확률	G 행렬
INT	-3.847(2.107)	0.077	
PRE	0.918(0.031)	<0.001	
BMI	0.230(0.083)	0.010	$\begin{bmatrix} 0.674 & -0.114 & 0.004 \\ -0.114 & 0.137 & -0.004 \\ 0.004 & -0.004 & 0.000 \end{bmatrix}$
DBP	-0.049(0.021)	0.024	
TC	0.022(0.007)	0.004	
TIME	-0.641(0.119)	<0.001	
$(TIME)^2$	0.042(0.011)	0.0002	

그러므로 개체별 시간에 대한 2차함수의 효과 $(pm_q)_i$를 제거한 모형을 선택하여 유의성 검정을 실시한다.

(모형 5) 시간에 대한 고정인자가 2차함수, 변량인자가 1차함수인 모형

$$\begin{aligned} y_{ij} =& \mu + b(pre)_i + c_2(BMI)_i + c_4(DBP)_i + c_6(TC)_i + m_l(time)_{ij} + m_q(time)^2 \\ &+ p_i + (pm_l)_i(time)_{ij} + e_{ij} \end{aligned}$$

고정인자의 시간에 대한 2차항을 포함하고, 개체별 시간에 대하여 1차함수의 효과 $(pm)_i$만 포함한 모형으로 모든 고정인자계수들이 유의하며, 변량인자의 공분산행렬 G 에서 분산성분이 각각 0.557, 0.061로 의미가 있다.

표 2.6 변량계수모형 5의 추정 결과

고정인자	회귀계수(표준오차)	유의확률	G 행렬
INT	-3.846(2.107)	0.077	
PRE	0.916(0.031)	<0.001	
BMI	0.232(0.083)	0.009	$\begin{bmatrix} 0.557 & -0.016 \\ -0.016 & 0.061 \end{bmatrix}$
DBP	-0.050(0.021)	0.023	
TC	0.022(0.007)	0.004	
TIME	-0.641(0.109)	<0.001	
$(TIME)^2$	0.042(0.011)	0.0002	

다음으로 반응변수와 고정인자의 시간과의 관계가 3차함수 관계를 포함하는 모형을 선택하여 추정과 검정을 실시하고자 한다.

(모형 6) 시간에 대한 고정인자가 3차함수, 변량인자가 일차함수인 모형

$$\begin{aligned} y_{ij} =& \mu + b(pre)_i + c_2(BMI)_i + c_4(DBP)_i + c_6(TC)_i + m_l(time)_{ij} + m_q\,(time)_{ij}^2 \\ &+ m_c\,(time)_{ij}^3 + p_i + (pm_l)_i(time)_{ij} + e_{ij} \end{aligned}$$

표 2.7 변량계수모형 6의 추정 결과

고정인자	회귀계수(표준오차)	유의확률	G 행렬
INT	-4.236(2.127)	0.055	
PRE	0.916(0.031)	<0.001	
BMI	0.232(0.083)	0.009	
DBP	-0.050(0.021)	0.023	$\begin{bmatrix} 0.559 & -0.016 \\ -0.016 & 0.061 \end{bmatrix}$
TC	0.022(0.007)	0.004	
TIME	-0.235(0.319)	0.463	
$(TIME)^2$	-0.0646(0.079)	0.416	
$(TIME)^3$	0.0079(0.0058)	0.177	

모형 6은 모형 5에서 $(TIME)^3$ 변수를 추가시킨 모형이다. 추정 결과 시간에 대한 3차함수 관계를 의미하는 회귀계수 m_c가 유의하지 않을 뿐만 아니라, 오히려 m_l, m_q의 회귀계수에도 유의하지 않게 나오도록 영향을 미쳤다. 따라서 모형 6은 모형 5보다 좋은 모형이 아니다.

전반적으로 각 모형에 대한 유의성 검정결과, 모형 5가 가장 적절한 모형으로 선택된다.

2.3. 변량계수모형의 적합성 검정

모형적합성 판정 기준으로 우도값에 관련된 $-2\log(L)$, AIC (Akaike's information criterion)와 오차분산을 채택하였다. $-2\log(L)$과 오차분산은 적을수록 좋은 모형이며, AIC는 자유도를 고려한 우도함수에 관련된 값으로 클수록 좋은 모형이다. 각 모형별로 $-2\log(L)$, AIC, 오차분산을 표 2.8에 적시하였다.

표 2.8 모형 적합성에 관한 기준값 결과

모형	$-2\log(L)$	AIC	오차분산
모형1	1084.9	-1086.9	2.208
모형2	896.5	-904.5	0.773
모형3	898.2	-906.2	0.773
모형4	889.6	-901.6	0.728
모형 5	891.2	-899.2	0.728
모형 6	897.9	-905.9	0.725

고정인자만 포함한 모형 1은 변량인자를 함께 고려한 다른 모형보다 적합성 검정 측면에서 대단히 적절하지 않다. 자유도를 고려한 AIC 기준에서는 모형 5가 가장 큰 값으로 나타났으며, $-2\log(L)$ 기준으로는 모형 4의 값이 가장 작지만, 모형 4와 모형 5의$-2\log(L)$ 값의 차이는 단지 1.6 차이로 극히 미미하다. 따라서 우도함수 기준으로 모형5를 가장 적절한 모형으로 채택한다. 오차분산 기준으로도 적절하지 않은 모형 6을 제외한 모형들 중에서는 모형 4와 모형 5의 오차분산이 0.728로 가장 작게 나타났다. 이러한 사항들을 종합적으로 분석한 결과 모형 5가 가장 적합한 모형으로 채택된다.

2.4. 최적모형 분석

유의하지 않은 허리엉덩이비율, 수축기혈압, 중성지방을 제외한 고정인자들의 분산분석표는 표 2.9와 같다. 즉 처리 전 체지방값, 비만지수, 확장기혈압, 총콜레스테롤 및 시간 인자들은 대단히 유의한 것으로 판명되었다. 고정인자들의 체지방에 미치는 효과는 표 2.6에서 보듯이 처리전 체지방값, 비만지수, 총콜레스테롤은 증가할수록 체지방량은 많아지며, 확장기혈압은 적을수록 체지방량은 감소하는 경향을 보여주며, 시간과 체지방량은 2차함수의 관계가 성립됨을 보여준다.

표 2.9 변량계수모형 5의 고정인자에 대한 분산분석표

요인	분자의 자유도	분모의 자유도	F 값	유의확률
PRE	1	31	868.8	<0.001
BMI	1	31	7.74	0.009
LBP	1	31	5.77	0.023
TC	1	31	9.95	0.004
TIME	1	246	34.39	<0.001
$(TIME)^2$	1	215	14.52	0.0002

표 2.1에서도 시간에 따른 체지방 감소는 이차함수의 관계가 성립됨을 보여준다. 즉, 처리 후 7주까지 시간이 지남에 따라 체지방값은 감소하나 대체로 감소의 폭은 점점 작아지다가 8주째에는 오히려 다소

증가하는 경향을 보인다. 마지막으로 변량인자효과 p_i, $(pm_l)_i$의 분산공분산행렬 G의 추정값은 다음과 같다.

$$G = \begin{bmatrix} 0.557 & -0.016 \\ -0.016 & 0.061 \end{bmatrix}$$

개체효과 및 개체별 시간에 대한 선형 효과는 유의하지만, 개체별 시간에 대한 2차항 이상의 고차항의 효과는 없는 것으로 나타났다.

3. 결론

체지방에 관한 실험에 참여한 36명의 여대생들은 매일 섭취하는 식사종류 및 식사량에 대하여 식사일지를 작성하며, 또한 식사종류 및 식사량을 카메라폰으로 찍어서 실험관리자에 보내며, 실험관리자는 이를 바탕으로 실험참여자들이 하루에 섭취하는 총 칼로리를 계산하고, 매주마다 8주간에 걸쳐서 상담을 실시하여 8주간에 걸친 체지방에 관한 반복측정자료를 구하였다. 이 실험자료를 이용하여 혼합모형의 일종인 변량계수모형을 적용하여 분석한 결과 회귀계수들의 유의성 검정 및 우도함수를 이용한 $-2\log(L)$, AIC 기준을 적용한 결과 모형 5가 가장 적합한 모형으로 선택되었다. 오차분산 기준으로도 적합하지 않은 모형 6을 제외하고, 모형 5의 오차분산이 0.728로 가장 적은 것으로 나타났다.

모형 5를 바탕으로 처리 후 체지방과 관련된 변수들과의 관계는 다음과 같다. 고정인자들인 허리엉덩이비율, 수축기혈압, 중성지방은 거의 관련이 없으며, 처리 전 체지방값, 비만지수, 확장기혈압, 총 콜레스테롤, 및 시간 의 고정인자들은 체지방량과 대단히 밀접한 관련을 보였다. 구체적으로 처리전 체지방값, 비만지수, 총콜레스테롤은 증가할수록 체지방량은 많아지며, 확장기혈압은 적을수록 체지방량은 감소하는 경향을 보여주었다. 처리실시 후 시간에 따른 체지방감소의 효과는 2차함수의 관계가 성립됨을 알 수 있다. 변량인자인 개체효과는 존재하며, 개체와 시간과의 교호작용에서 1차함수 관계가 있으며, 더 이상의 고차항의 관계는 존재하지 않는다. 처리 후 1주부터 7주까지는 시간이 지남에 따라 체지방량이 점점 감소하였으며, 8주에는 다소 증가하였지만, 실험이 끝난 8주 후에는 평균 2.1kg 감소한 효과를 보여주었다.

참고문헌

조진남 (2009). 체중감량자료에 대한 적정 공분산형태모형 산출에 관한 실증연구. <한국데이터정보과학회지>, **20**, 377-385.

조진남, 백재욱 (2009). 변량계수모형의 식이요법 실험자료에 관한 사례연구. <한국데이터정보과학회지>, **20**, 787-796.

조진남, 백재욱 (2010). 콜레스테롤 자료에 대한 적정 공분산행렬 형태 산출에 관한 통계적 분석.<한국데이터정보과학회지>, **21**, 1263-1270.

Brown, H. and Kempton, R. A. (1994). The application of REML in 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.

Choi, J. (2008). A marginal logit mixed-effects model for repeated binary response data. *Journal of the Korean Data & Information Science Society*, **19**, 413-420.

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.

Frees, E. D. (2006). *Longitudinal and panel data,* Cambridge University Press, New York.

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

Hay, C. R. M., Ludlam, C. A., Lowe, G. D. O., Mayne, E. E., Lee, R. J., Prescott, R. J. and Lee, C. A. (1998). The effect of monoclonal or ion-exchange purified factor VIII concentrate on HIV disease progression: A prospective cohort comparison. *British Journal of Haematology*, **101**, 632-637.

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

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

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

Smyth, J. F., Brown, A., Perren, T., Wilkinson, P., Prescott, R. J., Quinn, K. J. and Tedeschi, M. (1997). Glutathione reduces the toxicity and improves quality of the life of women diagnosed with ovarian cancer treated with cisplatin: Results of a double blind, randomized trial. *Annals of Oncology*, **8**, 569-5873.

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

Journal of the Korean Data & Information Science Society
2011, **22**(2), 287–296

한국데이터정보과학회지

A statistical analysis of the fat mass experimental data using random coefficient model†

Jinnam Jo[1]

[1]Department of Information & Statistics, Dongduk Women's University

Received 20 January 2011, revised 15 March 2011, accepted 19 March 2011

Abstract

Thirty six female students participated in the experiment of the fat mass weight loss. they kept diary for foods they ate every day, took a picture of the foods, transmitted the picture to the experimenter by the camera phone, and consulted him about fat mass loss once a week for 8 weeks period. Fat mass weight and its related factors of the students had been measured repeatedly every week during 8 weeks, The repeated measurement data were used for applying various random coefficient models. And hence optimal random coefficient model was selected. From the optimal model, the baseline, body mass index, diastolic blood pressure, total cholesterol and time of the fixed factors were very significant. The fixed quadratic time effect existed. The variance components corresponding to the subject effect, linear time effect of the random coefficients were all positive. Thus random coefficients up to the linear terms were considered as the optimal model. The treatment effect reduced the weight loss to an average of 2.1 kg at the end of the period.

Keywords: Fat mass, random coefficient model, subject effect, variance-covariance matrix.

† This research was supported by the Dongduk Women's University Research Grants 2010.
[1] Professor, Department of Information & Statistics, Dongduk Women's University, Seoul 136-714, Korea. E-mail:jinnam@dongduk.ac.kr