

반복요인이 3개인 반복측정자료에 대한 통계적 분석방법 * -양평 주민 혈압자료를 이용하여-

강성현¹⁾ 박태성²⁾ 이성곤³⁾ 김창훈⁴⁾ 김명희⁵⁾ 최보율⁶⁾

요약

본 논문에서는 2001년 경기도 양평 지역의 주민 880여명을 대상으로 반복측정되어 얻어진 혈압자료를 이용하여 반복요인이 세 개인 경우에 적절한 공분산구조를 선택하는 방법에 대해서 알아보고 혼합선형모형(mixed linear model)을 이용한 분석을 통해서 혈압에 영향을 주는 공변량(covariates)에 대해서 알아보았다.

주요용어: 반복측정자료, 공분산구조, 반복요인, 혼합선형모형.

1. 개요(Introduction)

반복측정자료(repeated measures data)는 같은 실험단위(unit) 혹은 개체(subject)로 부터 여러 관측 시간이나 실험조건 하에서 반복적으로 측정하여 얻어진 자료를 말한다. 이러한 반복측정자료는 동일한 개체를 대상으로 하여 얻어진 값들이므로 처리(treatment)간 수치들은 서로 어느 정도 상관관계가 있으며, 이것은 각 처리의 오차항간의 공분산행렬(covariance matrix)로 표현되어진다. 최근에 널리 사용되는 분석모형은 Harville, Laird와 Ware에 의해 제안된 랜덤효과(random subject effects)를 가지는 다변량 혼합선형모형(multivariate linear mixed models)과 Jennrich와 Schluchter에 의해 제안된 특정한 공분산구조모형을 들 수 있다(Harville DA, 1977; Laird NM and Ware JH, 1982; Jennrich RI and Schluchter MD, 1986). 반복측정자료의 분석은 흔히 각 처리에서 얻어진 측정치들 사이에 서로 차이가 있는지 없는지를 검정하거나 반응변수에 유의한 영향은 주는 공변량이 어떤 것인지를 알아보하고자 하는데 있다. 여기서, 유의한 변수를 찾는 데 있어서 적절한 공분산구조의 선택이 중

* 이 논문은 2001년 한국학술진흥재단의 지원에 의하여 연구되었음. (KRF-2001-015-DP0069)

- 1) (151-747) 서울시 관악구 신림9동, 서울대학교 통계학과 대학원, 박사수료
E-mail: gadin@biostats.snu.ac.kr
- 2) (151-747) 서울시 관악구 신림9동, 서울대학교 통계학과 대학원, 교수
E-mail: tspark@stats.snu.ac.kr
- 3) (151-747) 서울시 관악구 신림9동, 서울대학교 통계학과 대학원, 박사수료
E-mail: skon@biostats.snu.ac.kr
- 4) (133-791) 서울시 성동구 행당1동, 한양대학교 의과대학 예방의학교실, 전공의
E-mail: kchprev@hotmail.com
- 5) (301-832) 대전시 중구 용두동, 을지외과대학 예방의학교실, 전임강사
E-mail: mhkim@emc.eulji.ac.kr
- 6) (133-791) 서울시 성동구 행당1동, 한양대학교 의과대학 예방의학교실, 교수
E-mail: bychoi@hanyang.ac.kr

표 2.1: 양평주민 혈압측정자료

ID	SEX	AGE	BMI	DOCTOR	BP	TBP	SIDE	GRP	SICK	SICKF	REPE
1	0	59	28.5	5	148	1	1	1	1	1	0
1	0	59	28.5	5	82	0	1	1	1	1	0
1	0	59	28.5	5	144	1	0	1	1	1	0
1	0	59	28.5	5	82	0	0	1	1	1	0
1	0	59	28.5	5	150	1	1	2	1	1	0
1	0	59	28.5	5	86	0	1	2	1	1	0
1	0	59	28.5	5	148	1	0	2	1	1	0
1	0	59	28.5	5	85	0	0	2	1	1	0
1	0	59	28.5	5	.	1	1	3	1	1	0
1	0	59	28.5	5	.	0	1	3	1	1	0
1	0	59	28.5	5	.	1	0	3	1	1	0
1	0	59	28.5	5	.	0	0	3	1	1	0
2	1	53	25.8	5	140	1	1	1	0	1	0
2	1	53	25.8	5	110	0	1	1	0	1	0
2	1	53	25.8	5	142	1	0	1	0	1	0
2	1	53	25.8	5	108	0	0	1	0	1	0
2	1	53	25.8	5	138	1	1	2	0	1	0
2	1	53	25.8	5	110	0	1	2	0	1	0
2	1	53	25.8	5	140	1	0	2	0	1	0
2	1	53	25.8	5	110	0	0	2	0	1	0
2	1	53	25.8	5	.	1	1	3	0	1	0
2	1	53	25.8	5	.	0	1	3	0	1	0
2	1	53	25.8	5	.	1	0	3	0	1	0
2	1	53	25.8	5	.	0	0	3	0	1	0

요하다. 그러나, 일반적으로 적절한 공변량구조를 선택하는 것이 명백하지는 않다. 보통의 경우, 관측시점이 작고 자료가 균형(balanced)되어 있으며 결측치가 없는 완전한(complete) 자료인 경우에는 공분산구조에 특별한 가정이 필요 없는 비구조공분산모형(unstructured covariance models)을 사용한다. 하지만, 각 연구대상자에 대해 관측치의 수가 많은 경우에는 비구조공분산모형의 사용은 모수추정에 있어서 사용되는 반복 알고리즘이 수렴하지 않게 되는 문제점이 생기게 된다.

본 논문에서는 2001년 경기도 양평 지역의 주민 880여명을 대상으로 얻어진 혈압자료를 이용하여 반복요인이 세 개인 경우에 적절한 공분산구조를 적절한 공분산구조에 대해 알아보고 고정효과(fixed effects)를 가지는 혼합선형모형(mixed linear models)과 랜덤효과(random effects)를 가지는 혼합선형모형을 이용하여 혈압에 영향을 주는 공변량에 대해서 알아보려 한다. 2절에서는 자료의 수집과정 등에 대한 설명과 기초적인 분석을 통하여 자료의 특성을 알아보았고, 3절에서는 혈압자료를 분석하기 위한 모형(models)과 공분산구조에 대해서 설명을 하였다. 분석결과는 4절에서 설명을 하였다.

2. 자료구조 및 특성(Data and Characteristics)

[표 2.1]은 2001년 한양대 의과대학 예방의학교실에서 조사한 자료 중 2명의 연구대상자에 대한 자료만을 나타낸 것이다.

표 2.2: 성별에 따른 연구대상자의 특성

구분		남자(Male)		여자(Female)		p-value (t-test)
		Mean ¹	Std ²	Mean	Std	
3회 측정	AGE	56.0	12.7	60.7	10.7	0.016
	BMI	24.0	3.2	26.0	3.4	0.004
	SBP	129.5	16.5	135.0	20.9	0.054
	DBP	82.0	10.3	84.8	11.8	0.096
	N ³	296		405		-
2회 측정	AGE	55.7	13.7	52.8	14.3	0.006
	BMI	24.0	4.4	25.1	4.7	0.002
	SBP	124.8	19.5	121.1	19.6	0.012
	DBP	81.4	11.0	78.3	11.4	0.002
	N	53		112		-

¹ Mean : mean² Std : standard deviation³ N : number of subjects

혈압측정방법은 5명의 측정자(의사)가 연구대상자에 대해서 왼쪽과 오른쪽에 대한 수축기혈압(SBP)과 이완기혈압(DBP)을 각각 2회씩 측정하게 되고 2회 측정된 SBP, DBP가 각각 5mmHg이상 차이가 나는 경우에는 1회를 추가 측정하였다(JNC 6th Report, 1997). [표 2.1]을 통해서 조사된 각 변수를 살펴보면 성별(SEX), 나이(AGE), 체질량지수(BMI)가 조사되었고, DOCTOR는 조사를 실시한 의사(5명)를 나타내며 BP는 측정된 혈압수치이다. TBP(1=SBP, 0=DBP)는 수축기혈압과 이완기혈압을 나타내는 변수이다. 그리고, 각 연구대상자에 대해 왼쪽과 오른쪽을 구분하는 변수는 SIDE(1=왼쪽, 0=오른쪽)이며 REPE(1=2회 측정군, 0=3회 측정군)는 각 연구대상자가 2회 측정되었는지 3회 측정되었는지를 나타내는 변수이다. SICK(1=고혈압, 0=정상군)는 혈압을 측정하기 전에 고혈압에 대한 사전병력을 나타내는 변수이며 고혈압과 정상군의 최종 판정은 사전병력을 물어본 정보와 측정된 혈압자료를 이용하여 최종적으로 고혈압/정상군의 여부를 판정하게 되는데 SICKF(1=고혈압, 0=정상군)가 각 조사대상자의 최종적인 고혈압 판정결과를 나타내는 변수이다. 마지막으로 GRP변수는 측정회수(1회, 2회, 3회)를 나타내는 변수이다. [표 2.1]을 보면 첫 번째와 두 번째 연구대상자는 왼쪽 팔과 오른쪽 팔에 대해 수축기, 이완기혈압을 각각 2회 측정하였고 최종적으로 고혈압환자임을 알 수 있다.

[표 2.2]는 성별과 2회, 3회 측정군에 따른 연구대상자의 특성을 나타내는 표이다. 결과를 살펴보면, 여성의 경우 3회 측정군이 상대적으로 2회 측정군보다는 고연령이고 수축기/이완기혈압도 높게 나타났다. 3회 측정군에서는 남성보다는 여성이 이완기/수축기 혈압이 더 높게 나타나고 있다.

[표 2.3]은 측정자에 따라 3회 측정한 연구대상자와 2회 측정한 연구대상자의 특성을 나타내는 표이다. 우선 2회 측정군과 3회 측정군을 비교해 보면 2회 측정군보다 3회 측정군

표 2.3: 측정자에 따른 연구대상자의 특성

구분		측정자(Doctor)					p-value (ANOVA)
		A	B	C	D	E	
3회 측정군	N^1	21	34	50	50	10	-
	AGE(std)	55.6(11.1)	61.8(9.3)	60.5(11.0)	56.9(13.8)	62.5(4.2)	0.13
	BMI(std)	25.6(3.2)	26.0(3.5)	24.3(2.8)	25.5(4.0)	27.3(2.2)	0.06
	SBP(std)	134.8(21.7)	132.0(15.4)	130.7(19.0)	136.8(21.9)	128.1(16.8)	0.41
	DBP(std)	84.6(11.5)	84.6(9.2)	83.4(12.8)	83.3(11.1)	85.0(9.5)	0.95
	SEX(male %)	14.3	20.6	42	42	10	-
2회 측정군	N	138	273	1400	76	74	-
	AGE(std)	53.7(14.3)	52.9(14.1)	54.8(13.7)	54.6(14.7)	56.6(13.8)	0.24
	BMI(std)	24.4(4.4)	24.4(4.4)	24.8(3.3)	25.7(7.5)	24.8(3.4)	0.24
	SBP(std)	122.8(17.2)	122.5(21.1)	119.1(17.0)	126.8(17.2)	125.7(23.8)	0.04
	DBP(std)	79.0(10.5)	79.8(11.8)	77.7(10.4)	81.2(10.2)	82.5(13.3)	0.02
	SEX(male %)	47.8	39.2	43.6	35.5	47.8	-

¹ N=number of subjects

표 2.4: 왼쪽/오른쪽에 따른 고혈압군과 정상군의 분류

빈도		오른쪽 측정		Total
		정상군	고혈압군	
왼쪽 측정	정상군	595	14	601
	고혈압군	8	250	278
Total		611	268	879

• McNemar's Test : 0.16(P-value), Kappa= 0.88

에서 혈압, 나이, 체질량지수가 높게 나타남을 알 수 있다. 2회 측정군에서 SBP와 DBP에 대한 ANOVA 결과를 보면 p-value가 유의하게 나타났는데 이는 각 측정자별 연구대상자들 간의 차이라고 할 수 있다.

[표 2.4]는 왼쪽 혈압자료만 이용한 경우와 오른쪽 혈압자료만 이용한 경우로 나누었을 때 최종적인 고혈압판정여부를 나타내는 표이다. 이 표에서 대각부분은 왼쪽 자료와 오른쪽 자료에서 공통된 고혈압판정결과를 나타내는 부분이며 비대각부분이 서로 다른 결과를 나타내는 부분이다. 결과를 보면, 왼쪽 자료만을 이용할 경우 정상군이지만 오른쪽 자료만을 이용한 경우에는 고혈압군으로 판정된 대상자가 14명이고 반대로 왼쪽 자료만을 이용한 경우 고혈압군이지만 오른쪽 자료만을 이용한 경우에는 정상군으로 판정된 대상자가 8명이라는 것을 알 수 있다. 즉, 왼쪽 또는 오른쪽에 따라 상반된 결과를 보여주고 있다

[표 2.5]는 3회 측정군 165명에 대해서 2회만 측정한 자료와 3회 측정한 자료를 이용하여 고혈압군과 정상군의 판정여부를 나타내는 표이다. 결과를 보면 2회 측정한 자료를 이

표 2.5: 2회/3회에 따른 고혈압군과 정상군의 분류

빈도		3회 측정		Total
		정상군	고혈압군	
2회 측정	정상군	80	8	86
	고혈압군	0	79	79
Total		80	85	165

• McNemar's Test : 0.01(P-value), Kappa= 0.93

용한 경우에는 정상이지만 3회 측정한 자료를 모두 이용한 경우에 6명이 고혈압군으로 판정됨을 알 수 있다.

지금까지 자료소개 및 특성에 대해서 알아보았으며 3절부터는 반복요인이 3개인 반복 측정자료에서의 공분산구조의 선택과 모형적합을 통해서 혈압의 차이에 영향을 주는 공변량에 대해서 알아보도록 하겠다.

3. 모형(Models)

반응변수인 혈압(BP)은 수축기혈압(SBP)과 이완기혈압(DBP)으로 되어있지만 수축기혈압(SBP)과 이완기혈압(DBP)에 대해서 따로 모형을 설정할 필요가 없다. 교호작용항(interaction term)을 모형에 포함하여 각각 혈압의 차이에 영향을 주는 공변량의 효과를 볼 수 있기 때문에 동시에 모형을 적합하여 통해서 분석을 하였다.

분석에 사용한 혼합선형모형은 반복측정 분산분석(repeated measures ANOVA)모형을 벡터(vector)를 사용하여 다변량모형으로 확장시킨 형태이며, 다양한 형태의 공분산행렬을 사용할 수 있다. 또한 개체들간의 변동이 심할 때에 개체효과를 나타내는 랜덤효과항을 포함할 수 있는 가장 일반적인 형태의 모형이다. 모형의 형태는 다음과 같다.

$$y = X\beta + \epsilon, \epsilon \sim N(0, \Sigma)$$

단, y : 반응변수(response)

β : 모수 벡터(vector of unknown parameters)

X : 계획행렬(design matrix)

Σ : 공분산행렬(covariance matrix)

오차항이 또 다른 랜덤효과항을 포함하고 있는 경우($\epsilon = Zb + U$)는

$$y = X\beta + Zb + U$$

로 나타내어진다. 이 모형은 개체내의 변동을 효과적으로 설명할 수 있게 된다. 연구에서 명의 연구대상자가 있다고 가정을 하자. $i(1, \dots, n)$ 는 i 번째 연구대상자를 의미한다. n 번째 연구대상자에 대한 반응변수는 y_{ij} 로 표현할 수 있다.

여기서,

$$i = \text{연구대상자}(1, \dots, 883)$$

$$j = \text{각 개체의 } j\text{번째 측정}(1, \dots, 12)$$

을 나타낸다.

동일한 연구대상자에 대해 반복측정된 혈압자료는 서로 상관되어 있으며 각 연구대상자간 혈압자료는 그렇지 않다.

본 논문에서 적합한 모형에서 고정효과(fixed effects)만을 가지는 모형을 구체적으로 명시하면 아래와 같다.

$$\begin{aligned} y_{ij} = & \beta_0 + \beta_1(\text{age})_i + \beta_2(\text{bmi})_i + \beta_3(\text{side})_{ij} + \beta_4(\text{tbp})_{ij} + \beta_5^*(\text{grp})_{ij} \\ & + \beta_6(\text{sex})_i + \beta_7^*(\text{doctor})_i + \beta_8(\text{repe})_i + \beta_9(\text{sex} * \text{tbp})_{ij} + \beta_{10}(\text{side} * \text{tbp})_{ij} \\ & + \beta_{11}^*(\text{doctor} * \text{tbp})_{ij} + \beta_{12}^*(\text{grp} * \text{tbp})_{ij} + \beta_{13}(\text{repe} * \text{tbp})_{ij} + \epsilon_{ij} \end{aligned}$$

β_0 : 절편(intercept)

β_1 : 나이(age)의 효과(연속형 변수)

β_2 : 체질량지수(bmi)의 효과(연속형 변수)

β_3 : 팔(side)의 효과(side=1,오른쪽 ; 0,왼쪽)

β_4 : tbp의 효과(tbp=1,SBP ; 0,DBP)

β_5^* : 측정회수(grp)의 효과(grp=1, 2, 3)

β_6 : 성별(sex)의 효과(sex=1,남자 ; 0,여자)

β_7^* : 의사(doctor)의 효과(doctor=1, 2, 3, 4, 5)

β_8 : 측정군(repe)의 효과(repe=1,2회 측정군 ; 3,3회 측정군)

β_9 : 성별(sex)과 tbp의 교호작용효과

β_{10} : 팔(side)과 tbp의 교호작용효과

β_{11}^* : 의사(doctor)와 tbp의 교호작용효과

β_{12}^* : 측정회수(grp)와 tbp의 교호작용효과

β_{13} : 측정군(repe)과 tbp의 교호작용효과

이며,

$$\text{Var}(\epsilon_{ij}) = \sigma_e^2$$

이고, $\text{Var}(\epsilon_{ij}) = \sigma_e^2$ 은 가정한 공분산형태에 따라 달라지게 된다.

여기서, *로 표현된 항들은 그룹을 나타내는 모수들이다. 예를 들어 β_7^* 는 의사의 효과를 나타내는 모수로 의사의 수준이 5이므로 4개의 가변수를 사용하여 4개의 모수를 정의할 수 있으나 편의상 한 개의 모수로 표현하였다.

고정효과(fixed effects)와 개체 랜덤효과(random effects)를 가지는 모형은 다음과 같다.

$$\begin{aligned} y_{ij} = & \beta_0 + \beta_1(\text{age})_i + \beta_2(\text{bmi})_i + \beta_3(\text{side})_{ij} + \beta_4(\text{tbp})_{ij} + \beta_5^*(\text{grp})_{ij} \\ & + \beta_6(\text{sex})_i + \beta_7^*(\text{doctor})_i + \beta_8(\text{repe})_i + \beta_9(\text{sex} * \text{tbp})_{ij} + \beta_{10}(\text{side} * \text{tbp})_{ij} \\ & + \beta_{11}^*(\text{doctor} * \text{tbp})_{ij} + \beta_{12}^*(\text{grp} * \text{tbp})_{ij} + \beta_{13}(\text{repe} * \text{tbp})_{ij} + \pi_i + \epsilon_{ij} \end{aligned}$$

이며,

$$\begin{aligned} \pi &\sim i.i.dN(0, \sigma_\pi^2) : \text{개체 } i \text{의 랜덤효과} \\ \text{Var}(\epsilon_{ij}) &= \sigma_\pi^2 + \sigma_e^2 \end{aligned}$$

이다.

반복인자로는 TBP(1=SBP, 0=DBP), GRP(측정회수), SIDE(1=왼쪽, 0=오른쪽)로 3개의 반복인자를 가지고 있다.

혼합선형모형의 모수추정 방법은 Jennrich and Schluchter(1986)와 Laird, Lange, and Stram(1987)이 제안한 ML(maximum likelihood) 추정방법과 REML(Restricted ML) 추정방법이 있다. ML 추정방법은 σ^2 을 추정할 때 자유도 n 으로 나누어주며, REML 추정방법은 σ^2 을 추정할 때 LSE(least square estimate)와 마찬가지로 자유도 $n-p$ 로 나누어주는 차이가 있다.

각 연구대상자에 대해 최대 12회의 혈압을 측정하므로 공분산행렬을 가지게 된다. 모형 $\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\epsilon}$, $\boldsymbol{\epsilon} \sim N(\mathbf{0}, \boldsymbol{\Sigma})$ 에 대해서 반복측정 자료분석에서 사용될 수 있는 여러 형태의 공분산행렬의 구조를 살펴보자.

비구조(Unstructured)의 경우 공분산행렬 $\boldsymbol{\Sigma}$ 의 원소 $\sigma_{ij}(i = 1, \dots, t; j = 1, \dots, t)$ 를 일정하게 부여하지 않은 즉, 모든 요소를 모수로 생각을 해서 모두 추정하는 방법으로 이 때 추정해야 할 모수의 수는 $t(t+1)/2$ 개 이다. 복합대칭성(Compound Symmetry)은 동일분산, 동일공분산의 형태를 갖는 공분산행렬을 의미한다. 즉, 각 처리의 분산은 σ^2 으로 모두 동일하고 여러 처리간의 공분산은 $\rho\sigma^2$ 으로 일정하다. 따라서, 추정해야 할 모수의 수는 2개이다. 그리고, 제1차 자기상관회귀분석(First-order autoregressive)의 경우 분산은 σ^2 으로 일정하며 공분산은 $\sigma^2\rho^{|i-j|}$ 로 두 변수간의 시점의 차이가 크면 공분산이 줄어들고 차이가 작으면 공분산이 늘어나게 되는 형태의 공분산행렬이다. 따라서, 복합대칭성인 경우와 마찬가지로 추정해야 할 모수의 수는 2개이다. 마지막으로 복합구조(Composite Structure)모형은 각 반복요인의 공분산구조를 크로네커곱(Kronecker product, \otimes)을 이용하여 지정한 형태로 반복요인이 두 개인 반복측정자료분석에 사용하게 된다.

양평주민 혈압자료에 사용한 공분산구조는 비구조(UN), 복합대칭성(CS)과 복합구조모형을 사용하였으며 시점에 따라 관측된 자료가 아니기 때문에 제1차 자기상관($AR(1)$)의 경우는 타당하지 않아 사용하지 않았다. 앞서 언급한 것처럼 3개의 반복인자를 가지고 있다. 따라서, 공분산구조로 복합구조를 사용할 때 각각의 반복인자를 $UN@UN@UN$ 등으로 정의하여 모형을 분석할 수 있겠지만 Mixed procedure(SAS)에서는 반복인자로 2개까지만 지정할 수 있기 때문에 본 논문에서는 공분산 구조가 $UN@UN$ 와 $UN@CS$ 인 경우는 첫 번째 반복인자로 GRP(측정 회수)를 지정하고 두 번째 반복인자로는 SIDE와 TBP를 하나의 변수(SIDE_TBP)로 표현하여 반복인자로 지정하였다. 그리고 $CS@UN$ 는 위와 반대로 반복인자를 지정하였다.

표 4.1: 공분산구조에 따른 혼합선형모형 적합 결과

Effects	df (num)	df (denum)	공분산행렬구조					Random
			UN	CS	UN@UN	UN@CS	CS@UN	
age	1	857		0.0001	0.0001	0.0001		0.0001
sex	1	857		0.0065	0.0001	0.0007		0.0064
bmi	1	857		0.0001	0.0001	0.0001		0.0001
side	1	862		0.0330	0.0480	0.1692		0.0328
tbp	1	859		0.0001	0.0001	0.0001		0.0001
grp	2	1029	적합	0.0001	0.00531	0.0003	적합	0.0001
doctor	4	857	되지	0.0099	0.0003	0.0003	되지	0.0096
repe	1	857	없음	0.0001	0.0001	0.0001	없음	0.0001
tbp*sex	1	859		0.7018	0.9946	0.9152		0.7017
tbp*side	1	862		0.0421	0.0053	0.1884		0.0418
tbp*doctor	4	859		0.0001	0.0001	0.0001		0.0001
tbp*grp	2	1029		0.0117	0.0001	0.0001		0.0115
tbp*repe	1	859		0.0001	0.0001	0.0001		0.0001

4. 분석결과(Results)

[표4.1]은 3절에서 언급한 공분산행렬을 이용하여 혼합선형모형을 적합한 결과이다. 공분산행렬을 비구조(unstructured)로 지정한 경우는 모형이 적합되지 않았으며, 복합구조 중에서는 $CS@UN$ 인 경우가 적합되지 않았다. 랜덤효과모형은 랜덤효과로 개체(subject)효과만을 모형에 포함하는 경우만 적합이 되었으며, 이 경우는 공분산행렬을 복합대칭성(composite symmetry)으로 지정한 경우와 같기 때문에 [표 4.1]을 통해서 동일한 결과를 보여주고 있음을 알 수 있다. 각 공분산구조에 비슷한 결과를 보여주고 있지만 주된 관심사 중에 하나인 혈압을 왼쪽 또는 오른쪽 측정여부에 따라 혈압의 차이가 있는가를 나타내는 교호작용항($tbp * side$)의 경우는 공분산구조에 따라 다른 결과를 보여주고 있기 때문에 적절한 공분산구조의 선택이 중요하다고 할 수 있다.

혼합선형모형에서는 최대우도추정법(MLE)을 모수추정방법으로 사용할 때 우도비검정(Likelihood ratio test)을 실시하여 모형의 적합도(goodness of fit)검정을 실시할 수 있다. 모수의 수가 많을수록 모형의 적합도는 높아지며 -2 Log Likelihood의 값은 작아진다. 적합도 검정은 모수의 수가 가장 많은 모형을 기준으로 모수의 수를 줄여나가면서 축소된 모형이 기준 모형에 비해 적합한지 아닌지를 검정한다. [표 4.2]에서 $UN@UN$ 구조가 모수의 수가 가장 많으며 -2 Log Likelihood가 가장 작으므로 이를 기준으로 모수의 수를 줄여가면서 적합도 검정을 실시하여 모형의 축소가능 여부를 확인할 수 있다. $UN@CS$ 모형에 대해서는 우도비 검정통계량이 2003.1(자유도=9, p-value<0.0001)이므로 $UN@CS$ 모형으로 모형축소(reduction)가 불가능함을 알 수 있다. CS 모형에 대해서도 우도비 검정통계량이 5119.6(자유도=13, p-value<0.0001)이므로 축소가 불가능하다. 랜덤효과모형에 대해서도

표 4.2: 각 모형별 적합결과 비교

공분산 구조	모형별 적합도검정			-2 log			# of parameters
	df	chi-square	p-value	likelihood	AIC	BIC	
<i>UN</i>				적합되지 않음			
<i>CS</i>	1	7051.27	< .001	54102.6	54150.6	54264.9	24
<i>UN@UN</i>	14	12122.86	< .001	49031.0	49105.0	49281.3	37
<i>CS@UN</i>				적합되지 않음			
<i>UN@CS</i>	6	10119.80	< .001	51034.1	51092.3	51230.2	29
Random	-	-	-	54102.6	54150.6	54264.9	24

마찬가지로 우도비검정(LRT) 결과 p-value<0.0001로 축소가 불가능하다. 따라서, Akaike's Information Criterion(AIC), Schwarz's Bayesian Information Criterion(BIC)과 우도비검정을 통해서 최종모형은 공분산행렬의 구조가 *UN@UN*인 모형이다.

지금까지의 분석결과를 종합해보면, 반복요인이 3개인 반복측정자료를 분석하기 위해서 공분산구조가 복합구조인 경우에 각 반복요인에 대해 *UN@UN@UN*등을 생각할 수 있으나 분석 프로그램(SAS)에서 3개의 반복요인을 지정할 수가 없기 때문에 2개의 반복요인을 하나의 변수로 지정하여 2개의 반복요인을 가지는 형태로 바꾸어 공분산행렬을 정의하였다. 적합된 결과에서 교호작용항에 대한 해석결과 5명의 측정자들(의사)간에 차이가 있는 것으로 나타났다. 이는 측정자간 차이로 보기보다는 측정자별 연구대상자들 간의 차이로 해석하는 것이 타당할 것으로 생각된다. 그리고, 측정회수(GRP)에 따라서도 혈압의 차이가 있는 것으로 나타났으며 2회 측정군과 3회 측정군 사이에도 혈압의 유의한 차이가 있었고 왼쪽을 측정했을 때와 오른쪽을 측정했을 때 혈압의 차이가 나타나고 있음을 보았다. 일반적으로 한번 측정한 혈압으로 대상자를 평가하는 경우에도 상당부분 고혈압을 평가하여 이후의 합병증을 예측할 수 있으나 개인의 혈압은 변이정도가 크기 때문에 정확한 평가를 할 수 없다고 알려져 있다. Hypertension detection and follow-up cooperative group (1978)에서도 한 번 검진을 시행할 때 적어도 한쪽 팔에서 두 번 이상을 측정할 것을 권하고 있고 다른 팔에서도 함께 측정을 하여야 한다고 알려져 있다. 그리고 첫 번째의 검진에서 양쪽 팔에서 혈압을 두 번 이상씩 측정한 후 이후 검진에서 측정할 때에는 높은 혈압이 측정된 쪽의 혈압을 사용하여 대상자를 평가하여야 한다고 권하고 있다.

본 연구의 결과에서 살펴보면 최종모형(*UN@UN*)에 나타난 결과의 해석을 통해서 측정회수와 측정 부위에 따라 혈압의 차이가 나타나고 있어 타 연구 결과와 일치함을 알 수 있었다. 따라서, 한쪽 팔에서만 1회의 혈압을 측정하게 되면 최종적으로 고혈압 여부를 판정이 다르게 나타날 수 있으므로 2번 이상, 양쪽 팔을 모두 측정하여 고혈압여부를 판단해야 함을 확인할 수 있었다.

참고문헌

- [1] 박용규, 송혜향. (1998). *반복측정과 교차계획 자료의 분석법*. 자유아카데미, 서울.
- [2] 박태성. (1997). 반복측정 자료를 분석하기 위한 통계 모형의 고찰. <한국보건통계학회지>, 제22권. 23-34쪽.
- [3] 최지윤, 박태성. (2000). 반복요인이 2개인 반복측정자료를 분석하기 위한 통계모형의 고찰. <한국보건통계학회지>, 제25권. 2호. 106-115쪽.
- [4] Crowder, M.J. and Hand, D.J. (1990). *Analysis of Repeated Measures*. Chapman and Hall, London.
- [5] Crowder, M.J. and Hand, D.J. (1996). *Practical Longitudinal Data Analysis*. Chapman and Hall, London
- [6] Davis, C.S.(1996). *Analysis of Repeated Measurements*. Lecture note presented at Korean National Group of the International Biometric Society.
- [7] Diggle, P.J., Liang, K.Y. and Zeger, S.L. (1994). *Analysis of Longitudinal Data*. Clarendon Press, Oxford.
- [8] Fronlich E, Labarthe D, Maxwell M, Perloff D, and Weidman W. (1993). Recommendation for human blood pressure determination by sphygmomanometers : Report of a special task force appointed by the steering committee. *American heart association, circulation*, 88, 2460-2467.
- [9] Harville D.A. (1977). Maximum likelihood approaches to variance component estimation and to related problems. *Journal of the American Statistical Association*, 72, 320-340.
- [10] Hypertension detection and follow-up cooperative group (1978). Variability of blood pressure and the results of screening in the hypertension detection and follow-up program. *Journal of chronic disease*, 31, 661-667.
- [11] Jennrich, R.I. and Schluchter, M.D. (1986). Unbalanced Repeated-Measures Models with Structured Covariance Matrices, *Biometrics*, 42, 805-820.
- [12] Laird, N.M., Lange, N. and Stram, D. (1987). Maximum likelihood computations with repeated measures : Application of the EM algorithm, *Journal of the American Statistical Association*, 82, 97-105.
- [13] Laird, N.M. and Ware J.H. (1982). Random-effects models for longitudinal data. *Biometrics*, 38, 963-974.
- [14] Lindsey, J.K. (1993). Models for Repeated Measurements with Structured Covariance Matrices. *Biometrics*, 42, 805-820.

- [15] Park, T. and Lee, Y.J. (2002). Covariance models for nested repeated measures data : analysis of ovarian steroid data. *Statistics in Medicine*, 21, 143-164.
- [16] The 6th Report of the Joint National Committee (JNC) on Prevention, Detection, Evaluation, and Treatment of High Blood Pressure (1997).

[2002년 6월 접수, 2003년 09월 채택]

Statistical Methods for Repeated Measures Data with Three Repeat Factors

SungHyun Kang ¹⁾ Taesung Park ²⁾ Sungon Yi ³⁾ ChangHoon Kim ⁴⁾
MyeongHee Kim ⁵⁾ BoYool Choi ⁶⁾

ABSTRACT

In this paper, we consider choosing the appropriate covariance structure for analyzing repeated measures data with three repeat factors from a study of blood pressure data, which is collected from the local residents of YangPyeong, Gyeonggi-do (2001) and fitted linear mixed models to find the significant covariates on outcome variable(Blood Pressure).

Keywords: repeated measures data; covariance structure; repeat factor; linear mixed models

1) Doctoral student, Department of Statistics, Seoul National University.

E-mail: gadin@biostats.snu.ac.kr

2) Professor, Department of Statistics, Seoul National University.

E-mail: tspark@stats.snu.ac.kr

3) Doctoral student, Department of Statistics, Seoul National University.

E-mail: skon@biostats.snu.ac.kr

4) Apprentice doctor, Department of Preventive Medicine, Hanyang University.

E-mail: kchprev@hotmail.com

5) Full-time lecturer, Department of Preventive Medicine, Eulji University School of Medicine.

E-mail: mhkim@emc.eulji.ac.kr

6) Professor, Department of Preventive Medicine, Hanyang University.

E-mail: bychoi@hanyang.ac.kr