

임의계수모형을 이용한 지속성에 대한 연구

강형근¹⁾, 김병수²⁾, 남정모³⁾

포천중문 의과대학교 의예과¹⁾
연세대학교 상경대학 응용통계학과²⁾
연세대학교 의과대학 예방의학교실³⁾

I. 序 論

고혈압이나 고지혈증은 관상동맥질환의 중요한 위험요인으로 알려져 있다. 유아기나 소아기에 혈압 또는 콜레스테롤 수치가 높은 대상이 성인이 되어도 계속적으로 높은 수준을 유지할 경우 혈압 또는 콜레스테롤이 지속성 현상(tracking phenomenon)을 갖는다고 한다. 만성질환의 중요한 위험요인이 지속성 현상을 갖는다면 소아기나 청소년기에 이와같은 위험요인을 갖는 대상을 파악하여 필요한 조치를 하므로써 만성질환의 예방을 가능케 하기 때문에 만성질환의 위험요인이 지속성 현상을 갖는가를 밝히는 것은 역학 연구(epidemiologic study)에서 매우 중요한 문제이다.

이와같은 지속성 현상을 파악하기 위한 많은 방법론적 연구가 진행되어 왔다. 이와같은 연구들을 살펴보면 다음과 같이 크게 두가지로 나눌 수 있다. 첫째는 어떤 특성의 상대적인 순위(relative rank)가 시간의 경과에 따라 계속적으로 일정하게 유지되는 상대적 유지도(relative maintenance)로 지속성을 정의한 연구이고 둘째는 주어진 관찰치로부터 미래의 값을 예측하는 예측도(predictability)로 지속성을 정의한 연구이다.

상대적 유지도로 지속성을 정의한 연구들을 살펴보면 다음과 같다. McMahan(1981)은 성장곡선모형

(growth curve model)을 이용하여 모평균에 대한 개인의 상대적 편차가 시간의 경과에 따라 어느 정도 유지되는가로 지속성 현상을 설명하였다. Foulkes와 Davis(1981)는 다항회귀식(polynomial regression)을 이용하여 각 개인의 회귀곡선을 추정하고 추정된 개인의 회귀곡선이 교차하는 정도로 지속성 현상을 설명하였다. Segar와 Tager(1993)는 장기간의 관찰연구에서 회귀계도(regression tree)를 이용하여 추정된 성장곡선이 동질성이 있는가로 지속성 현상을 설명하였다. Taylor 등(1994)은 반복측정된 자료의 변화율에 대한 상대적인 순위가 시간의 경과에 따라 일정하게 유지되는 정도로 지속성 현상을 설명하였다.

한편 예측도로 지속성을 정의한 연구들을 살펴보면 다음과 같다. Ware와 Wu(1981)는 직교다항식(orthogonal polynomial)을 추정한 뒤, 추정된 직교다항식으로부터 미래의 값을 예측하였다. Rao(1987)는 성장곡선모형을 이용하여 미래의 값을 예측하였다. Tian 등(1994)은 성장곡선모형 대신에 제한적 3차 스피라인 함수(restricted cubic spline function)를 이용하여 미래의 값을 예측하였다.

이 연구에서는 지속성 현상을 상대적 유지도로 정의하고 반복측정된 자료에 적용이 가능한 임의계수모형(random coefficient model)을 이용하여 지속성 현상을 적절하게 설명할 수 있는 새로운 지속성 지수를 제안하고자 한다. 이 연구의 구체적인 연구목적은 다음과 같다.

* 이 논문은 1995년도 한국학술진흥재단의 신진연구인력지원연구비에 의하여 연구되었음.

첫째, 임의계수모형(random coefficient model)을 이용하여 지속성 현상을 적절하게 설명할 수 있는 새로운 지속성 지수를 제안하였다.

둘째, 모의실험(simulation)을 통하여 기존의 지속성 지수와 검정력을 비교하였다.

셋째, 제안된 지속성 지수가 지속성 현상을 적절하게 설명하는가를 알아보기 위하여 실제 자료에 적용하여 그 활용성을 알아보았다.

II. 研究對象 및 方法

이 연구에서는 반복측정된 자료로부터 임의계수모형을 이용하여 개인의 각 시점에서의 추정치를 구하고, 구해진 추정치의 순위를 이용하여 지속성 현상을 적절하게 설명할 수 있는 새로운 지속성 지수를 제안하였다. 반복측정된 자료에 임의계수모형을 적용한 것은 대부분 연속형으로 측정된 만성질환의 위험요인은 측정오차(random measurement error)를 동반하기 때문이다. 또한 제안된 새로운 지속성 지수의 분포를 구하였다. 제안된 새로운 지속성 지수를 기존의 지속성 지수와 모의실험을 통하여 비교하였다. 이 연구에서 비교할 기존의 지속성 지수는 상대적 유지도로 지속성 현상을 정의한 McMahan(1981)의 지수이다. 모의실험은 시점에 대한 특정 상관계수 행렬을 가정하고 여러시점에서 반복측정된 자료를 발생시킨 후 추정된 지속성 지수의 검

정력(power)을 비교하였다. 모의실험 과정은 Fig. 1과 같으며 GAUSS를 이용하여 자료를 발생하였고 프로그래밍하였다.

지속성이 없다는 귀무가설하에서 상관계수 행렬은 단위행렬(Identity matrix)이 된다. 즉 모든 시점간의 상관계수는 0이 된다(시점 $j = j'$ 이면 $\rho_{jj} = 1$, $j \neq j'$ 이면 $\rho_{jj'} = 0$). 한편 지속성 현상이 있다는 대립가설 하에서는 특정 상관계수 행렬(specific correlation matrix)을 가정하고 특정 상관계수 행렬을 갖는 다변량정규분포(multivariate normal distribution)로부터 표본수가 30, 시점이 6인 자료를 발생하였다. 임의계수모형 추정에는 다항회귀모형(polynomial regression model)을 이용하였고 곡선의 차수는 5차로 하였다. 유의수준(significance level, α) 0.05하에서 단측검정을 실시하였고 이와같은 시행을 1,000번 반복시행하여 검정력(power)을 비교하였다.

III. 研究成績

1. McMahan의 지속성 지수

McMahan(1981)은 지속성 현상을 어떤 특성에 대한 상대적인 순위(relative rank)가 시간의 경과에 따라서 계속적으로 유지되는 상대적 유지도로 정의하고, 시간의 경과에 따른 어떤 특성의 자연적인 증가를 성장곡선모형으로 모형화하고, 이 모형으로부터 지속성 지수를 구하였다. 어떤 특성을 갖는 모집단으로부터 n 명의 관찰대상자를 시간의 경과에 따라 m 번 관찰한다고 하자. 이때 개인 i 를 m 번 관찰한 벡터를 X_i , q 개의 모수($q \leq m$)를 갖는 성장곡선 모형의 모수벡터를 β_i , 개인내 계획행렬(within individual design matrix)을 A 라 하면 성장곡선 모형은 β_i 가 주어졌을 때 X_i 의 조건분포(conditional distribution)와 β_i 의 주변분포(marginal distribution)로써 다음 식 (1)과 같이 나타낼 수 있다.

$$(X_i | \beta_i) \sim N(A\beta_i, \sigma^2 I) \dots \dots \dots (\text{식 } 1)$$

$$\text{여기서, } \beta_i \sim N(\beta, \Sigma)$$

따라서 X_i 의 주변분포는 다음 식 (2)와 같다.

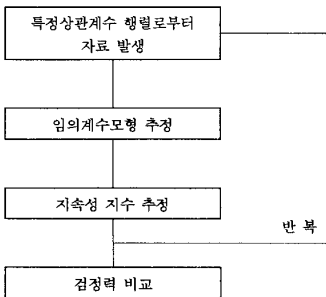


Fig. 1. 모의실험 과정.

$$X_i \sim N(A\beta_j, A\Sigma A' + \sigma^2 I) \dots\dots\dots(\text{식 } 2)$$

$$\tau = \frac{1' \Delta (A\Sigma A' - \Delta^2) \Delta \underline{1}}{1' \Delta (\Delta \underline{1} \underline{1}' \Delta - \Delta^2) \Delta \underline{1}} \dots\dots\dots(\text{식 } 7)$$

McMahan은 다음 식 (3)과 같이 전체 관찰대상자의 평균으로부터 개인 i의 상대적 편차가 시간의 경과에 따라 일정하게 유지되는 정도로 지속성을 정의하였다.

$$E(X_{ij}|\beta_j) = \mu_j + \kappa_j \sigma_j \quad (i=1, 2, \dots, n, \\ j = 1, 2, \dots, m) \dots\dots(\text{식 } 3)$$

여기서, μ_j, σ_j : 시점 j에서의 전체 관찰대상자의 평균과 표준편차.

$$\kappa_j : \text{개인 } i \text{의 상대적 순위로서 } E(\kappa_j) = 0, \\ \text{var}(\kappa_j) = 1.$$

식 (3)을 행렬로 표시하면 다음 식 (4)와 같다.

$$A\beta_j = A\hat{\beta}_j + \Delta \underline{1} \dots\dots\dots(\text{식 } 4)$$

여기서, Δ : 대각항이 $\Delta_{ii} = [A\Sigma A']^{-1/2}$ 인 $m \times m$ 대각행렬.
 $\underline{1}$: $m \times 1$ 단위벡터.

위의 식 (4)의 공분산 행렬(covariance matrix)을 구하면 다음 식 (5)와 같다.

$$A\Sigma A' = \Delta \underline{1} \underline{1}' \Delta \dots\dots\dots(\text{식 } 5)$$

McMahan은 총분산에 대하여 지속성으로 설명할 수 있는 분산의 비로 다음 식 (6)과 같이 지속성 지수를 제안하였다.

$$\xi = \frac{1' \Delta A\Sigma A' \Delta \underline{1}}{(1' \Delta \Delta \underline{1})^2}, \quad (0 \leq \xi \leq 1) \dots\dots(\text{식 } 6)$$

ξ 는 지속성 현상이 없는 경우(즉, $A\Sigma A'$ 이 대각행렬인 경우)에도 陽數를 가질 수 있기 때문에 McMahan은 ξ 를 수정한 지속성 지수 τ 를 다음 식 (7)과 같이 제안하였다.

τ 는 완전한 지속성 현상이 있는 경우(즉, $A\Sigma A' = \Delta \underline{1} \underline{1}' \Delta$)에 1이 되고, 지속성이 없는 경우(즉, $A\Sigma A'$ 이 대각행렬인 경우)에 0이 된다. 즉 지속성 지수 τ 는 $0 \leq \tau \leq 1$ 의 범위를 가지며 값이 클 수록 지속성 현상이 있는 것으로 해석할 수 있다. 한편 지속성 지수 τ 를 Δ '에 의해 표준화 시킬 경우 τ 는 모든 시점간의 상관계수의 평균($\bar{\rho}$)이 된다.

McMahan은 τ 의 근사적 신뢰구간을 구하기 위하여 잭나이프 방법(jackknife method)을 이용하였다. 전체 관찰대상자에 대한 τ 의 추정량을 $\bar{\tau}$, 개인 i를 제외한 (n-1)명의 관찰대상자에 대한 τ 의 추정량을 $\hat{\tau}_i$ 라고 하면 τ 에 대한 의사값(pseudo value) $\bar{\tau}_i$ 는 다음과 같다.

$$\bar{\tau}_i = n \hat{\tau} - (n-1) \hat{\tau}_i \quad (i = 1, 2, \dots, n)$$

따라서, τ 의 잭나이프 추정량은

$$\bar{\tau} = \frac{1}{n} \sum_{i=1}^n \bar{\tau}_i \text{ 이 되고 } n \text{이 커짐에 따라 다음식 (8)}$$

이 성립한다.

$$\frac{n^2 (\bar{\tau} - \tau)}{\left[\frac{1}{(n-1)} \sum_{i=1}^n (\bar{\tau}_i - \bar{\tau})^2 \right]^{1/2}} \rightarrow N(0, 1) \dots\dots\dots(\text{식 } 8)$$

2 새로운 持續性 指數

이 연구에서는 상대적 유지도로 지속성을 정의하고 다음과 같이 새로운 지속성 지수를 제안하였다. 개인 i의 시점 j에서의 관찰치를 y_{ij} 라고 하면 개인 i의 임의계수모형(random coefficient model)은 다음 식 (9)와 같다.

$$y_{ij} = x_{ij} \cdot \beta_i + \epsilon_{ij} \\ (i = 1, 2, \dots, n \quad j = 1, 2, \dots, m) \dots\dots(\text{식 } 9)$$

$$\begin{aligned} \text{여기서, } \beta_i &\sim N(\beta, \Sigma) \\ \epsilon_{ij} &\sim N(0, \sigma^2 I) \end{aligned}$$

식 (9)로부터 시점 j에서 개인 i의 추정치는 $\hat{y}_{ij} = x_{ij} \cdot \hat{\beta}_j$ 가 된다. 이때 시점 j에서 개인 i의 추정치 \hat{y}_{ij} 에 대한 순위를 $r_{ij}(r_{ij}=1, 2, \dots, n)$ 라고 정의하자.

i \ j	1	2	...	m	계
1	r_{11}	r_{12}	...	r_{1m}	$r_{1.}$
2	r_{21}	r_{22}	...	r_{2m}	$r_{2.}$
.
.
n	r_{n1}	r_{n2}	...	r_{nm}	$r_{n.}$
계	$r_{.1}$	$r_{.2}$...	$r_{.m}$	$r_{..}$

이때 다음 식 (10)가 성립하면 완전한 지속성이 있다고 말할 수 있다.

$$r_{i1} = r_{i2} = \dots = r_{im} \quad (\forall i=1, 2, \dots, n) \quad \text{.....(식 10)}$$

이와같은 개념으로 다음 식 (11)과 같이 지속성 지수를 제안하였다.

$$\gamma = \frac{1}{c} \sum_{i=1}^n \left(\sum_{j=1}^m r_{ij} \right)^2 \quad \text{.....(식 11)}$$

$$\text{여기서, } c = \max \sum_{i=1}^n \left(\sum_{j=1}^m r_{ij} \right)^2 = m^2 \frac{n(n+1)(2n+1)}{6}$$

만약 주어진 관찰치가 완전한 지속성 현상을 갖는다면 $\gamma=1$ 이 된다.

그러나, 식 (11)에서 제안한 지속성 지수는 상대적 유지도의 개념으로 지속성을 설명하는데 다음과 같은 문제점이 있다. 개인의 수가 2(x : 개인 1, o : 개인 2)이고 관찰 시점이 5인 경우에 자료의 형태가 Fig. 2와 같다고 하자.

x	x	o	o	o
o	o	x	x	x
1	2	3	4	5

(a) 시점,

x	o	x	o	x
o	x	o	x	o
1	2	3	4	5

(b) 시점

Fig. 2. 지속성의 개념.

Fig. 2에서 (a)가 (b)의 경우에 비하여 상대적으로 높은 지속성 현상이 갖는다고 할 수 있다. (a)의 경우 개인 1의 순위는 시점 1, 2에서 2, 시점 3, 4, 5에서 1로 $(\sum r_{1j})^2=49$, 개인 2의 순위는 시점 1, 2에서 1, 시점 3, 4, 5에서 2이므로 $(\sum r_{2j})^2=64$ 이고 $c=125$ 이므로 $\gamma=0.904$ 가 된다. 마찬가지로 (b)의 경우에 개인 1의 순위의 합은 $(\sum r_{1j})^2=64$, 개인 2의 순위의 합은 $(\sum r_{2j})^2=49$ 로 $\gamma=0.904$ 로 동일한 결과를 나타낸다.

따라서 이와같은 문제를 해결하기 위하여 개인 i에 대하여 다음 식 (12)와 같이 가중치를 주었다.

$$w_i = \frac{1}{m-1} \sum_{j=2}^m \left\{ 1 - \left(\frac{r_{ij} - r_{i1}}{n-1} \right)^2 \right\} \quad \text{.....(식 12)}$$

이때 w_i 는 $r_{ij}=r_{i1}(\forall j, j')$ 이면 1이 되고 $|r_{ij} - r_{i1}| = n-1(\forall j, j')$ 이면 0이 된다 ($0 \leq w_i \leq 1$). 이 연구에서는 개인의 가중치 w_i 를 고려하여 다음 식 (13)과 같이 새로운 지속성 지수를 제안하였다.

$$\gamma_w = \frac{1}{c} \sum_{i=1}^n w_i \left(\sum_{j=1}^m r_{ij} \right)^2 \quad \text{.....(식 13)}$$

$$= \frac{1}{c} \sum_{i=1}^n \left[\frac{1}{m-1} \sum_{j=2}^m \left(\frac{r_{ij} - r_{i1}}{n-1} \right)^2 \right] \left[\sum_{j=1}^m r_{ij} \right]^2$$

$$\text{여기서, } c = m^2 \frac{n(n+1)(2n+1)}{6}$$

식 (13)에서 새로 제안한 지속성 지수를 Fig. 2에 적용한 결과는 다음과 같다. Fig. 2에서 (a)의 경우 개인에 대한 가중치는 $w_1=3/4$, $w_2=3/4$ 이므로 $\gamma_c=0.677$ 이고 (b)의 경우 개인에 대한 가중치는 $w_1=3/4$, $w_2=3/4$ 이므로 $\gamma_c=0$ 이 된다.

지속성이 없다는 귀무가설하에서 γ_c 의 평균을 계산하면 다음식 (14)와 같다.

$$E(\gamma_c) = \frac{1}{60m^2(n-1)(2n+1)} \cdot [15(5n^2-2n-7)m^2+5(5n^2-12n+5)m+6(n^2+1)] \dots\dots\dots(\text{식 } 14)$$

한편 γ_c 의 분산을 추정하기 위하여 부스트랩 방법(bootstrap method, Efron 1982)을 이용하였다. 표본수가 m , 시점이 n 인 관찰치로부터 P 개의 부스트랩 표본을 반복추출(with replacement)하여 P 개의 γ_c 의 부스트랩추정량 γ_c^{*p} ($p=1, 2, \dots, P$)를 계산하고 계산된 γ_c^{*p} 로부터 다음 식 (15)와 같이 분산을 추정하였다.

$$\text{Var}(\gamma_c) = \frac{1}{P-1} \sum_{p=1}^P (\gamma_c^{*p} - \gamma_c^*)^2 \dots\dots\dots(\text{식 } 15)$$

$$\text{여기서, } \gamma_c^* = \frac{1}{P} \sum_{p=1}^P \gamma_c^{*p}$$

이때 표본수 $n \rightarrow \infty$ 에 따라 다음 식 (16)이 성립한다.

$$\frac{\gamma_c - E(\gamma_c)}{\sqrt{\text{var}(\gamma_c)}} \rightarrow N(0,1) \dots\dots\dots(\text{식 } 16)$$

3. 모의실험 결과

이 연구에서는 특정 상관계수 행렬을 갖는 다변량 정규분포로부터 시점이 6, 관찰치의 수가 30인 자료

를 발생하였고 제안한 지속성 지수와 McMahan의 지속성 지수를 구하기 위한 임의계수모형과 성장곡선모형의 차수는 5차로 하였다. 여러시점에서 반복 측정된 자료의 시점간 상관계수 행렬이 R_1 과 같은 경우 γ_c 와 τ 의 검정력을 비교한 결과는 Table 1과 같다.

$$R_1 = \begin{bmatrix} 1 & r & r & r & r & r \\ r & 1 & r & r & r & r \\ r & r & 1 & r & r & r \\ r & r & r & 1 & r & r \\ r & r & r & r & 1 & r \\ r & r & r & r & r & 1 \end{bmatrix}$$

Table 1. 상관계수 행렬이 R_1 일 경우 γ_c 와 τ 의 검정력 비교

r	γ_c	τ
0.0	0.058	0.052
0.1	0.331	0.534
0.2	0.696	0.899
0.5	0.999	1.000
0.9	1.000	1.000

Table 1에서 τ 는 귀무가설하에서의 검정력이 0.05보다 크기 때문에 비교를 위하여 τ 의 귀무가설하에서의 검정력이 0.05가 되도록 값을 조정하였다. 시점간의 상관계수가 커질수록 γ_c 와 τ 모두 검정력이 증가함을 볼 수 있었다. 특히 시점간의 상관계수가 0.5이상일 경우에 두 지속성 지수는 검정력에 차이를 보이지 않음을 볼 수 있었다.

Table 2는 여러시점에서 반복측정된 자료에서 시점간 상관계수 행렬이 R_2 와 같은 경우 γ_c 와 τ 의 검정력을 비교한 결과이다.

Table 2를 살펴보면 상관계수 행렬 R_2 하에서 시점간 상관계수가 커질수록 γ_c 와 τ 모두 검정력이 증가함을 볼 수 있으며 γ_c 가 τ 보다 높은 검정력을 보이고 있다.

Table 3은 여러시점에서 반복측정된 자료에서 시점간 상관계수 행렬이 R_3 와 같은 경우 γ_c 와 τ 의 검정력을 비교한 결과이다.

$$R_2 = \begin{bmatrix} 1 & -r & -r & -r & -r & -r \\ -r & 1 & r & r & r & r \\ -r & r & 1 & r & r & r \\ -r & r & r & 1 & r & r \\ -r & r & r & r & 1 & r \\ -r & r & r & r & r & 1 \end{bmatrix}$$

Table 2. 상관계수 행렬이 R_2 일 경우 γ_c 와 r 의 검정력 비교

r	γ_c	r
0.1	0.176	0.122
0.2	0.356	0.240
0.5	0.969	0.745
0.9	1.000	1.000

$$R_3 = \begin{bmatrix} 1 & -r & r & r & r & r \\ -r & 1 & -r & -r & -r & -r \\ r & -r & 1 & r & r & r \\ r & -r & r & 1 & r & r \\ r & -r & r & r & 1 & r \\ r & -r & r & r & r & 1 \end{bmatrix}$$

Table 3. 상관계수 행렬이 R_3 일 경우 γ_c 와 r 의 검정력 비교

r	γ_c	r
0.1	0.112	0.114
0.2	0.178	0.264
0.5	0.718	0.754
0.9	1.000	1.000

Table 3에서 보면 상관계수 행렬 R_3 하에서는 γ_c 와 r 에 비하여 더 높은 검정력을 보임을 알 수 있다. 그러나 자료의 형태상 상관계수 행렬이 R_2 일 경우가 R_3 일 경우에 비하여 지속성 현상이 강함을 알 수 있다. 본 연구에서 제안한 γ_c 는 R_2 일 경우가 R_3 일 경우에 비하여 지속성 현상이 강함을 잘 설명하고 있다. 그러나 r 는 상관계수 행렬이 R_2 일 경우와 R_3 일 경우에 비슷한 검정력을 보이고 있다. McMahan의 지속성 지수는 반복측정된 자료에서 시점간의 상관계수의 평균으로 표현되므로 자료의 형태와는 관계없이 상관계수의 평균이 같을 경우 지속성 정도가 같게 표현됨을 알 수 있다.

4. 실제 자료에 대한 적용

지속성 현상을 보이는가를 알아보기 위하여 제안

된 새로운 지속성 지수가 실제자료에서도 활용될 수 있는가를 알아보기 위하여 강화아동혈압연구(서 등 1993) 자료에 적용하여 보았다.

강화아동혈압연구는 1986년에 인천광역시 강화군 강화읍의 초등학교 1학년에 재학중인 학생 430명(남 211명, 여 219명)을 대상으로 1991년까지 6년동안 매년 수축기 및 이완기 혈압과 이와 관련된 요인(신장, 체중 등)들을 측정한 자료이다. 이중 1991년까지 6년동안 계속적으로 추적관찰된 학생은 304명(남 145명, 여 159명)이었다. 따라서 본 연구에서는 304명의 자료를 이용하여 혈압이 지속성 현상을 보이는가를 알아보려고 한다.

6년동안 계속적으로 추적관찰된 304명을 남녀별로 나누어 수축기 및 이완기 혈압에 대한 6년간의 평균의 변화를 살펴보면 다음 Table 4와 같다.

Table 4. 성별 학년별 수축기 및 이완기 혈압의 평균과 표준편차 (단위 : mmHg)

학년	남		여	
	수축기 혈압 (평균±표준편차)	이완기 혈압 (평균±표준편차)	수축기 혈압 (평균±표준편차)	이완기 혈압 (평균±표준편차)
1	92.28±9.34	59.79±5.55	96.40±9.92	60.77±8.14
2	98.98±7.54	62.78±6.80	99.99±8.99	65.21±7.07
3	102.57±8.51	64.04±9.85	103.75±10.00	66.81±9.14
4	103.57±9.10	70.28±8.31	105.19±9.63	71.18±8.08
5	107.77±10.81	71.68±7.87	110.08±10.66	73.21±8.21
6	108.79±9.91	72.87±7.20	112.11±11.15	73.74±7.67

Table 5. 성별 수축기(하위 대각행렬, lower diagonal) 및 이완기 혈압(상위 대각행렬, upper diagonal)에 대한 시점간 상관계수 행렬

성	학년	1	2	3	4	5	6
남	1	-	0.472	0.308	0.284	0.417	0.241
	2	0.620	-	0.452	0.443	0.437	0.311
	3	0.560	0.643	-	0.411	0.321	0.264
	4	0.524	0.626	0.596	-	0.428	0.231
	5	0.544	0.574	0.553	0.643	-	0.407
	6	0.427	0.497	0.488	0.608	0.681	-
여	1	-	0.383	0.324	0.265	0.255	0.273
	2	0.624	-	0.372	0.383	0.221	0.266
	3	0.518	0.646	-	0.451	0.311	0.205
	4	0.421	0.645	0.703	-	0.419	0.297
	5	0.317	0.492	0.476	0.583	-	0.416
	6	0.348	0.519	0.507	0.613	0.683	-

Table 6. 성별 수축기 및 이완기 혈압에 대한 새로운 지속성 지수와 95% 신뢰구간

성	구분	γ	95% 신뢰구간	z
남	수축기 혈압	0.8543	0.8309-0.8777	16.24
	이완기 혈압	0.7796	0.7535-0.8057	8.91
여	수축기 혈압	0.8531	0.8323-0.8739	18.11
	이완기 혈압	0.7855	0.7618-0.8092	10.28

Table 4를 살펴보면 남녀 모두 수축기 및 이완기 혈압이 학년이 올라갈수록 증가하는 양상을 보임을 알 수 있다.

한편 6년동안 계속적으로 추적관찰된 304명을 남녀별로 나누어 수축기 및 이완기 혈압에 대한 여섯 시점간의 상관계수 행렬을 구해보면 다음 Table 5와 같다.

위의 Table 5를 살펴보면 남자의 경우 수축기 혈압의 시점간의 상관계수는 0.427부터 0.681이고 이완기 혈압은 0.231부터 0.472, 여자의 경우 수축기 혈압의 시점간의 상관계수는 0.317부터 0.703이고 이완기 혈압은 0.205부터 0.451로 높은 상관성을 지니고 있음을 알 수 있다.

위 자료를 이용하여 본 연구에서 제안한 새로운 지속성 지수 γ 와 95% 신뢰구간(confidence interval)을 구하면 다음 Table 6과 같다.

새로운 지속성 지수를 이용하여 강화아동혈압연구의 자료를 분석해 본 결과 남녀 모두 수축기 및 이완기 혈압이 지속성 현상을 보임을 알 수 있다.

IV. 考察 및 結論

만성질환의 위험요인이 지속성 현상을 보이는가를 밝히는 것은 역학연구에서 매우 중요한 문제이다. 만성질환의 위험요인이 지속성 현상을 있는가에 대한 많은 연구가 있어 왔다. 심혈관 질환의 위험요인인 혈압(Suh 등 1994 : 이순영 1990 : Lauer 등 1984 : Andre 등 1986 : Lauer 등 1986 : Palti 등 1988), 콜레스테롤(Freedman 등 1985 : Stuhldreher 등 1991), 비만(Rolland-Cachera 등 1987 : Cronk 등 1983 : Baumgartner 등 1988 : Casey 등 1992)이 지속성 현상을 보이는가에 대한 연구가 있어 왔다. 만성질환의 위험요인 뿐만 아니라 신장, 체중 등 신체지수(Rolland-Cachera 등 1987 : Casey 등 1992), 식이습관(Stein 등 1991 : Garry 등 1992)이나 운동(Lee 등 1992 : Beunen 등 1992) 등의 지속성 현상을 보고자 하는 연구들도 있어 왔다.

이와같이 만성질환의 위험요인이 지속성 현상을 갖는가를 알아보기 위한 많은 연구가 진행되어 왔지만 아직까지도 지속성 현상에 대한 일차된 정의는

없는 형편이다. 그러나 지속성 현상을 파악하기 위한 방법론적인 연구들을 살펴보면 상대적으로 정의된 연구(McMahan 1981 : Foulkes와 Davis 1981 : Segar와 Tager 1993 : Taylor 등 1994)와 예측도로 정의된 연구(Ware와 Wu 1981 : Rao 1987 : Tian 등 1994)로 크게 나눌 수 있다. 이밖에 도 초기값과 변화율 사이의 관계로 지속성 현상을 설명하고자 하는 연구들도 있어 왔다(Blomqvist 1977 : Zerbe 등 1994 : Zucker 등 1995).

이 연구에서는 지속성 현상을 상대적 유지도로 정의하였고 임의계수모형을 이용하여 지속성 현상을 파악할 수 있는 새로운 지속성 지수를 제안하였고 모의실험을 통하여 기존의 McMahan의 지수와 비교하였다. 모의실험 결과 McMahan의 지속성 지수는 자료의 형태와는 관계없이 시간간의 상관계수의 평균으로 지속성의 정도를 표현하는 문제가 있음을 보였다. 이 연구에서 제안한 지속성 지수는 자료의 형태에 따른 지속성 현상을 적절히 표현함을 볼 수 있었다. 또한 이 연구에서 제안한 지속성 지수를 실제 자료에 적용해 본 결과 실제 자료에서도 그 활용성을 볼 수 있었다.

이 연구에서 제안된 지속성 지수는 대부분의 연구와 마찬가지로 반복측정된 자료에 임의계수모형을 적용할 때 임의계수모형의 차수에 따라 값의 변화가 있으므로 적절한 모형의 차수를 결정해야 하는 어려움이 있다. 그러나 적절한 임의계수모형의 차수를 선정한다면 이 연구에서 제안한 지속성 지수는 자료의 형태에 따른 지속성 현상을 파악하는데 적절한 지표가 될 수 있다. 결론적으로 이 연구에는 지속성 현상을 적절히 파악할 수 있는 새로운 지속성 지수를 제안함으로써 만성질환의 예방 등의 의학연구에 도움을 줄 수 있다고 생각한다.

REFERENCES

서 일, 남정모, 강형근, 혈압의 역학적 연구와 지속성에 대한 통계학적 분석, 응용통계연구 1993;6:191-200
이승연, 아동혈압의 지속성에 관한 시계열분석, 1990; 연세대학교 석사학위논문
Andre JL, Deschamps JP, Petit JC, et al. Change of blood pressure over five years in

childhood and adolescence. Clin Exp Hypertens A 1986;8:539-45

Baumgartner RN, Roshe AF. Tracking of fat pattern indices in childhood: the Melbourne Growth Study. Hum Biol 1988;60:549-67

Beunen G, Lefevre J, Claessens AL, et al. Age-specific correlation analysis of longitudinal physical fitness levels in men. Eur J Appl Physiol 1992;64:538-45

Blomqvist, N. On the relation between change and initial value. Journal of American Statistical Association 1977;72:746-49

Casey VA, Dwyer JT, Coleman KA, et al. Body mass index from childhood to middle age: a 50-y follow-up. Am J Clin Nutr 1992;56:14-8

Cronk CE, Roche AF, Kent R Jr, et al. Longitudinal trends in subcutaneous fat thickness during adolescence. Am J Phys Anthropol 1983;61:197-204

Efron, B. The jackknife, the bootstrapped and other resampling plans. Society for industrial and applied mathematics. Philadelphia, 1982 pp 27-36

Foulkes, M.A., and Davis, C.E. An index of tracking for longitudinal data. Biometrics 1981;37:439-46

Freedman DS, Shear CL, Strinivasan SR, et al. Tracking of serum lipids and lipoproteins in children over an 8-year period: the Bogosa Heart Study. Prev Med 1985;14:203-16

Garry PJ, Hunt WC, Koehler KM, et al. Longitudinal study of dietary intakes and plasma lipids in health elderly men and women. Am J Clin Nutr 1992;55:682-88

Lauer RM, Anderson AR, Beaglehole R, et al. Factors related to tracking of blood pressure in children. US National Center for Health Statistics Health Examination Surveys Cycles II and III. Hypertension 1984;6:307-14

Lauer RM, Mahoney LT, Clarke WR. Tracking of blood pressure during childhood: the Muscatine Study. Clin Exp Hypertens A 1986;8:515-37

Lee IM, Paffenbarger RS Jr, Hsieh CC. Time trends in physical activity among college allumni, 1962-1988. Am J Epidemiol 1992;135:915-25

McMahan, C.A. An index of tracking. Biometrics 1981;37:447-55

Palti H, Gofin R, Adler B, et al. Tracking of

blood pressure over an eight year period in Jerusalem school children. *J Clin Epidemiol* 1988;41:731-35

Rao, C.R. Prediction of future observations in growth curve models. *Statistical Science* 1987;2:434-71

Rolland-Cachera MF, Deheeger M, Guillaud-Bataille M, et al. Tracking the development of adiposity from one month of age to adulthood. *Ann Hum Biol* 1987;14:219-29

Segal, M.R., and Tager, I.B. Trees and tracking. *Statistics in Medicine* 1993;12:2153-68

Stein AD, Shea S, Basch CE, et al. Variability and tracking of nutrient intakes of preschool children based on multiple administrations of the 24-hour dietary recall. *Am J Epidemiol* 1991;134:1427-37

Suh I., Nam, C.M., Lee, E.S., Kim, I.S., and Lee, S.Y. Blood pressure tracking in Korean schoolchildren. *International Journal of Epidemiology* 1994;23:710-15

Stuhlderher W, Donahue R, Drash A, et al. The Beaver County Lipid Study: sixteen-year cholesterol tracking. *Ann NY Acad Sci* 1991;623:466-68

Taylor, J.M.G., Cumberland, W.G., and Sy, J.P. A stochastic model for analysis of longitudinal AIDS data. *Journal of American Statistical Association* 1994;89:727-36

Tian, J.J., Shukla, R.S., and Buncher, C.R. On prediction of future observation in growth curve model. *Statistics in Medicine* 1994;13:2205-17

Ware, J.H., and Wu, M.C. Tracking: Prediction of future values from serial measurements. *Biometrics* 1981;37:427-37

Zerbe, G.O., Wu, M.C., and Zucker, D.M. Studying the relationship between change and initial value in longitudinal study. *Statistics in Medicine* 1994;13:759-68

Zucker, D.M., Zerbe, G.O., and Wu, M.C. Inference for the association between coefficients in a multivariate growth curve model. *Biometrics* 1995;51:413-24

=Abstract=

**A Study on Tracking by Using
Random Coefficient Model**

**Hyung Gon Kang¹⁾, Byung Soo Kim²⁾,
Chung Mo Nam³⁾**

¹⁾*Dept. of Pre Medicine, College of Medicine,
Pochon CHA University*

²⁾*Dept. of Applied Statistics, Yonsei University*

³⁾*Dept. of Preventive Medicine,
Yonsei University*

There are many epidemiologic studies to find the tracking phenomenon. In repeated measurement data, a population is said to have tracking phenomenon with respect to a particular characteristic if, for each individual, the relative rank of observed value maintained over time. Understanding tracking phenomena in epidemiologic study is quite important. If the risk factors of chronic disease have tracking phenomenon, early detection of these risk factors can lead to the possibility of early treatment. In this study, we defined tracking as relative maintenance and proposed new measure of tracking by using random coefficient model. We compared our measure with McMahan's using simulation study. And we applied ours to real data. We may conclude that our new measure of tracking is adequate in explaining and detecting the tracking phenomenon.

Key Words : Tracking phenomenon, Relative maintenance, Random coefficient model