

도서관 성과 측정을 위한 잠재성장모형 적용에 관한 연구*

A Study on the Application of Latent Growth Model for Measuring the Outcomes of Library

박 성 재 (Sung-jae Park)**

한 상 우 (Sang-woo Han)***

조 세 흥 (Sae-hong Cho)****

목 차

- | | |
|-----------|----------------------|
| 1. 서 론 | 4. 도서 대출량 변화 분석 및 논의 |
| 2. 이론적 배경 | 5. 결 론 |
| 3. 연구방법론 | |

초 록

본 연구의 목적은 도서관의 성과를 측정하기 위해서 잠재성장모형을 적용하고 그 활용가능성을 논하는 것이다. 도서관의 성과를 측정하기 위하여 도서관 활동에 대한 데이터, 즉 이용자의 대출 데이터를 이용하였다. 시간에 따른 변화를 담고 있는 종단자료를 분석하기 위한 통계적인 모형으로 잠재성장모형을 이용하였다. 서울 소재 공공도서관의 2010년부터 2014년까지의 95,962명의 이용자의 대출데이터를 무조건모형, 조건모형, 혼합성장모형을 적용하여 대출의 특성을 분석하였다. 분석결과, 대출량은 절편요인이 4.19, 기울기요인이 0.24의 선형성장을 보였다. 성별에 따른 차이분석에서 큰 차이를 보이지 않았으나 4개의 그룹으로 나누었을 때, 10세 미만의 어린이의 대출 패턴이 급속히 증가하는 추세를 보였다. 향후 문헌정보학 분야에서 종단연구자료를 분석할 때, 잠재성장모형이 활용될 것으로 기대된다.

ABSTRACT

The purpose of this study is to discuss the application of the Latent Growth Model to measure the outcomes of public library. For outcome measurements, library circulation data were collected to identify longitudinal changes of library users' reading habit. The latent growth model was applied to statistically test the changes over time. The circulation data of 95,962 users registered in some public libraries in Seoul, ranged between 2010 and 2015, were analyzed using unconditional model, conditional model, and growth mixture model which all are called the latent growth model. The results show that the intercept of the model is 4.19 and the slope is 0.24 in the linear growth model. The gender difference in two latent variables including intercept and slope was a shade difference. The result from the growth mixture model analysis, additionally indicates that the number of books checked out by children under age 10 is rapidly increased. The application of the latent growth model in library fields is expected to widely spread out for the longitudinal data analysis.

키워드: 도서관 성과, 대출, 종단 자료, 잠재성장모형, 무조건모형, 조건모형, 성장혼합모형

Library Outcomes, Circulation, Longitudinal Data, Latent Growth Model, Unconditional Model, Conditional Model, Mixture Growth Model

* 이 논문은 2014년 대한민국 교육부와 한국연구재단의 지원을 받아 수행된 연구임 (NRF-2014S1A5A2A03066128).

** 한성대학교 도서관정보문화트랙 부교수(spark@hansung.ac.kr) (제1저자)

*** 광주대학교 문헌정보학과 조교수(swhan@gwangju.ac.kr) (교신저자)

**** 한성대학교 컴퓨터공학부 교수(chosh@hansung.ac.kr) (공동저자)

논문접수일자: 2018년 11월 15일 최초심사일자: 2018년 11월 15일 게재확정일자: 2018년 11월 27일
한국문헌정보학회지, 52(4): 179-194, 2018. [http://dx.doi.org/10.4275/KSLIS.2018.52.4.179]

1. 서론

최근 도서관의 성과에 대한 연구가 다수 진행되었다. 성과를 어떻게 정의할 것인지, 측정을 위한 도구와 측정지표 등에 대해서 각각 달리 제시하고 있다. 성과가 서비스 이용자에게 있어서의 변화라는 점에서 도서관 이용자에게 변화내용이나 도서관을 통해 얻게 된 혜택에 대해서 질문하고 그 응답 결과를 분석하는 방법으로 진행되었다. 이러한 연구들은 설문에 참여한 사람들의 심리적인 변화까지도 측정이 가능하다는 점에서 의미가 있다. 그러나 한편으로는 자기보고 방식의 설문지를 통해 측정함으로써 발생하는 제한점, 예를 들면 자기 자신에 대해 긍정적으로 말하고자 하는 경향성으로 인한 오류(Mischel 2013), 질문을 수정하면 응답 결과가 달라지는 것(McDonald 2008) 등의 문제가 발생할 수 있다. 심지어는 자기 자신에 대해서 제대로 이해하고 있는 응답자는 없다는 점을 문제점으로 지적하기도 한다(McDonald 2008). 따라서 이용자들이 직접 응답하지 않고 이용자들의 활동을 기록하고 있는 도서관 데이터를 활용한다면 이러한 자기보고 방식의 설문지법이 가진 한계점을 극복할 수 있을 것이다.

또한 성과는 변화를 측정한다는 점에서 특정 시점에서 응답자의 변화를 묻기보다는 여러 시점에서의 응답자의 상태를 묻고 시점 간의 차이를 측정하는 것이 정확한 측정이 될 수 있다. 전자는 횡단자료로 특정 시점에서의 이용자 상태를 깊이있게 볼 수 있다는 장점이 있는 반면에 시간적으로 어떻게 변화되었는지를 정확히 파악하기 쉽지 않다는 문제점이 있다. 변화정

도에 대해서도 이용자가 보고하는 방식이기 때문에 앞서 지적한 문제점이 동일하게 발생할 수 있다. 따라서 종단자료를 활용하는 것이 성과를 측정하는데 있어서 정확한 변화를 제시해 줄 수 있을 것이다. 특히, 도서관 데이터는 활동 기록이 시점별로 기록이 된다는 점에서 시간에 의한 변화상을 그려낼 수 있다는 점에서 성과 분석에 활용될 수 있다.

일반적으로 시간의 변화에 따른 분석을 수행하는 종단연구에서는 전체의 변화정도를 나타내는 기울기를 측정할 때 전체 평균의 기울기를 사용한다. 전체의 경향성을 파악할 수 있다는 점에서 의미있는 방법이 될 수 있지만 전체에 포함된 개인의 변화와 차이를 설명할 수 없다는 문제점을 가지고 있다. 시간에 따른 변화의 양상은 다양하게 나타나고 이는 개인에 따라서 전체에서 구분된 집단에 따라서 달리 나타날 수 있다는 점에서 개인 혹은 집단의 기울기를 산출하고 이들 기울기의 분산을 통계적으로 검증할 필요가 있다(Singer and Willett 2003). 이러한 분석을 가능하게 하는 것이 잠재성장모형이다. 즉, 잠재성장모형은 관측되지 않은 잠재적인 성장요인을 추정하는 것을 목적으로 한다(신탉수 2014).

사회과학 분야에서는 이러한 잠재성장모형을 다양하게 활용하고 있는데 반해 문헌정보학 분야에서는 그 활용을 찾아보기 쉽지 않다. 문헌정보학 분야, 특히 도서관은 이용자들의 도서 대출기록이 시스템에 자동으로 쌓이게 되면서 종단자료가 쉽게 획득이 가능하다는 점에서 잠재성장모형의 적용이 쉽고 매우 유용할 것으로 기대된다. 따라서 본 논문은 잠재성장모형에 대한 소개와 함께 도서관 데이터를 이용하여 도서

대출량이 어떻게 변화하는 지에 대해 분석하였다. 도서 대출은 일반적으로 도서관의 중요활동인 대출과정의 산출물이라고 볼 수 있지만 본 연구는 단순히 수치에 집중하기보다는 대출량이 어떻게 변화하는지, 변화의 정도와 특징은 무엇인지를 분석함으로써 성과를 보이하고자 한다. 따라서 도서관의 기본적인 서비스인 도서 대출 서비스의 성과를 측정하는 방법론의 하나로 잠재성장모형을 제안하고자 한다.

잠재성장모형은 분석 데이터의 형태에 따라 다양한 분석방법이 적용가능한데, 본 연구에서는 다음 연구문제를 해결하기 위해 무조건모형과 조건모형, 혼합성장모형을 적용하였다.

- 연구문제 1. 도서 대출량은 시간이 지남에 따라 어떠한 변화를 보이는가?
- 연구문제 2. 도서 대출량은 성별에 따라 다른 차이가 나타나는가?
- 연구문제 3. 도서 대출량은 변화 양상에 따라 다양한 집단으로 구분가능한가? 집단의 연령별 특징은 무엇인가?

2. 이론적 배경

2.1 잠재성장모형

종단자료는 변화를 측정하기 위해 수집되는 자료로 시간의 변화에 따라 최소한 3회 이상 측정된 자료이다(Willett 1989). 일부 t-test에서 진행되는 사전/사후 분석은 특정 시점을 기준으로 그 전과 그 후의 차이를 분석한다는 점에서 시간흐름의 영향을 고려했다고 할 수 있지만 변

화는 특정 시점이 아닌 지속성을 띠는 점에서 두 번의 측정만으로는 충분하지 않을 수 있기 때문이다(Polyhart and Vanderberg 2010). 몇 번의 측정이 가장 적합한가와 관련하여 측정시점이 많아질수록 모형의 관별과 추정이 실패할 가능성이 낮아진다는 점에서 보다 많은 시점을 사용하는 것이 바람직하다(김수영 2017).

그러나 측정시점을 3회 이상으로 늘린다는 것은 변화의 다양성을 측정할 수 있다는 장점이 있지만 통계분석의 독립성 가정을 만족시키지 못한다는 다른 문제점을 발생시킨다. 종단자료는 한 사람에 대해서 시점을 달리하여 여러 번 측정을 했다는 점에서 종속변수 값이 서로 독립적이라는 독립성 가정을 위배하게 된다. 한 사람의 종속변수 값들은 다른 사람의 종속변수 값들보다 상관성이 높게 나타나면서 문제가 발생하는 것이다. 결국은 독립성 가정 문제를 해결할 수 있어야만 종단자료가 제공하는 다양한 변화상을 분석할 수 있는 것이다.

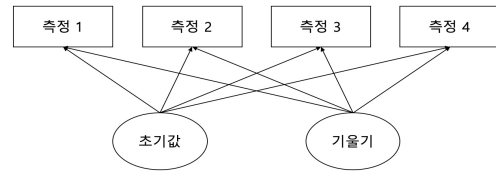
독립성 가정을 만족시키기 위한 방법으로 종단자료에서의 종속변수 값을 시점에 따라 부여함으로써 서로 다른 변수를 만드는 것이 하나의 방법이 될 수 있다. 예를 들면, 4개의 시점에서 측정된 종속변수는 시점을 달리하여 y_1, y_2, y_3, y_4 로 구분하여 자료를 구축한다. y 라는 한 변수가 4개의 값을 가지지 않고 서로 다른 변수로 분리됨으로써 독립성 가정을 만족하게 되는 것이다. 이러한 다변량 자료를 이용하여 여러 개의 시점에서 측정한 자료를 분석하는데 사용하는 방법이 중의 대표적인 방법이 잠재성장모형이다.

Singer와 Willett(2006)은 1982년부터 2005년까지 연구 분야별로 종단연구를 적용한 연구에 대한 분석을 실시했다. 분석 대상으로 삼은

약학, 경영학, 생물학, 심리학, 사회학, 농학, 동물학, 경제학 분야에서 종단 연구는 꾸준히 증가하고 있는 경향을 발견하였다. 그러나 이러한 종단연구들은 기존의 통계학에서 사용하던 반복측정에 의한 ANOVA나 회귀분석을 기반으로 분석이 된다는 점을 지적하고 종단자료의 특성을 반영하여 정확한 변화를 측정하기 위한 통계적 분석방법, 예를 들면 잠재성장모형 등의 활용이 요구됨을 지적하고 있다.

잠재성장모형은 종단자료를 바탕으로 분석하는 통계 모형으로 시간에 따른 변화를 분석하는데 있어서 전체 샘플의 특성을 분석하기 보다는 개인의 특성을 분석한다(김수영 2017). 즉, 변화에 있어서 개인들 간의 차이가 있는지를 분석하고 차이가 있다면 이러한 차이를 설명할 수 있는 변수는 무엇인지를 찾는 분석방법이라 할 수 있다. 또한 전체 샘플에서 변화의 양상에 따라 집단을 구분하고 집단간의 변화의 차이를 분석하는데도 잠재성장모형은 활용이 되고 있다.

잠재성장모형은 McArdle과 Epstein이 1987년 제안한 연구방법으로 구조방정식모형(Structural Equation Model: SEM)의 한 유형이다. 구조방정식 모형에서 사용하는 잠재변인(Latent Variable)과 관측변인(Observed Variable)으로 구성되는데 전자는 직접적인 측정이나 관찰이 어려운 변인으로 변인을 계량화할 수 있는 구성요인, 즉 관측변인의 측정을 통해 추정된다. <그림 1>은 잠재성장모형의 일반적인 경로 모형으로 총 4회에 걸쳐 측정된 관측변인이 있다. 이들 변인을 통해서 초기값과 기울기라는 잠재변인을 추정하게 되고 집단의 변화를 측정하게 된다.



<그림 1> 잠재성장모형의 경로모형

잠재성장모형에서 일반적으로 사용하는 공식은 <공식 1>과 같다.

$$\begin{aligned}
 y_{ti} &= \pi_{0i} \times 1 + \pi_{1i} \times (t-1) + \epsilon_{ti} \\
 \pi_{0i} &= \beta_{00} + \zeta_{0i} \\
 \pi_{1i} &= \beta_{10} + \zeta_{1i}
 \end{aligned}
 \tag{공식 1}$$

y_{ti} 는 i 번째 사람의 t 시점에서의 관측치를 의미한다. π_{0i} 는 첫 번째 시점, 즉 $t=1$ 인 경우를, π_{1i} 는 성장기울기를 나타낸다. 이는 일차함수의 형태를 나타내는 것으로 시간이 흐름에 따라 $\pi_{0i} + \epsilon_{1i}$ 값으로부터 π_{1i} 값 만큼 변화함을 의미한다. 성장기울기에 추가적으로 $(t-1)$ 의 성장요인부하량을 사용하는데 선형성장의 경우에는 시점이 증가할 때마다 값이 1씩 증가하는 고정부하량(fixed growth loading)이 사용된다. 이는 선형성장을 의미하고 비선형성장을 위한 공식에는 이차함수로 $\pi_{2i} \times (t-1)^2$ 값이 추가된다. ϵ_{ti} 는 관측치에 대한 오차로 정규분포를 이룬다. 첫 번째 시점과 성장기울기에 대한 값은 ζ_{0i} , ζ_{1i} 라는 분산값을 갖고 이를 통해 개인 간의 차이를 분석할 수 있다. 앞서 언급했듯이, 분산값이 크면 개인 간의 차이가 크다는 점에서 개인 간의 성장이 전체 평균을 통해 분석되기 보다는 개인 혹은 집단을 구분하여 성장을 추정할 필요가 있다는 점을 확인할 수 있다.

무조건 모형이 시점변화에 따른 개인 간의 변

화의 차이를 검증한다면 이러한 변화를 만들어 내는 요인에 대한 분석은 조건 모형을 적용함으로써 가능하다. 조건 모형은 기존의 무조건 모형(공식 1)에 관심매락변인(X)을 추가함으로써 가능하다. 만약 관심매락변인이 성별이라면 성별에 따라 변화가 차이가 있는지, 즉 초기값과 기울기가 어떻게 다른지에 대한 분석이 가능해진다.

$$\begin{aligned} \pi_{0i} &= \beta_{00} + \gamma_{00}X_i + \zeta_{0i} \\ \pi_{1i} &= \beta_{10} + \gamma_{01}X_i + \zeta_{1i} \end{aligned} \quad \langle \text{공식 2} \rangle$$

〈공식 2〉의 γ_{00} , γ_{01} 은 변인 X가 초기값과 기울기에 미치는 영향을 나타낸다. 영향정도가 클수록 높은 값을 갖는다는 점에서 회귀계수로 해석이 가능하다.

모형에 대한 추정 이후에 모형이 실재를 잘 설명하고 있는지에 대한 모형적합도 검증이 요구된다. 모형적합도 검증을 위해서는 기본적으로 카이제곱 검정을 실시한다(김수영 2017). 그러나 카이제곱 검정은 관찰되는 사례에 민감하다는 점에서 적합도를 정확하게 측정하기 위한 다양한 방법들과 지수들이 제안되었다. 일반적으로 많이 사용하는 적합도 검정으로 TLI, CFI, SRMR, RMSEA 등이 있다. Hu와 Bentler(1999)는 구조방정식 모형의 적합도를 검증하기 위한 각 지수의 기준점을 〈표 1〉과 같이 제시하고 있다.

〈표 1〉 모형적합도 기준

적합도 지수	기준값과 범위
TLI, CFI	> 0.90
SRMR	< 0.08
RMSEA	< 0.06

조건모형과 무조건모형 이외에도 잠재성장 모형은 다양한 모형으로 확장이 가능하다. 앞서 언급한 두 모형은 한 영향변인의 변화만을 측정하는 반면에 사회 현상은 다양한 변인의 변화가 동시에 일어나는 경우가 많다. 도서관 현장에 있어서도 도서관 대출이 많은 사람은 도서관 프로그램에 참여하는 경우가 많다는 점에서 대출량과 프로그램 참여율은 상호 영향을 미치는 변수가 될 수 있다. 이러한 변인 간의 변화를 측정하기 위한 다변량 잠재성장모형(multivariate latent growth model)이 개발되었다.

또한 데이터에 존재하는 위계구조를 반영하여 성장의 차이를 분석할 수 있는 다층잠재성장 모형(multi-level latent growth model)이 제안되기도 하였다. 도서관 현장에서 이용자들은 주로 특정 도서관, 예를 들면 집에서 가장 가까운 도서관을 주로 이용한다. 도서관 이용은 이용자의 개인적 특성뿐만 아니라 도서관의 특성도 영향을 미친다는 점에서 도서관이라는 상위 계층과 그 도서관을 이용하는 이용자라는 하위 계층으로 나누어 분석할 때 이용에 있어서 변화 차이를 보다 깊이 있게 분석할 수 있다.

집단 내에서 동일한 특성이 아닌 서로 다른 특성을 가진 집단이 있다면 이들 집단들과 관심 변인간의 영향관계를 분석했을 때, 변수와 집단 간의 보다 명확한 분석이 가능하다. 이렇게 하위집단과 관심 변수 간의 관계를 추적하는 잠재계층분석(Latent class analysis)과 잠재성장모형이 혼합된 모형이 성장혼합모형(Growth mixture model)이다. 즉, 집단 내에서 서로 다른 성장을 보이는 집단을 구분하고 각각 집단과 관심 변인 간의 영향 관계를 분석

하는 것이다. 성장 모형에서 일반적으로 성장기울기의 변화가 없는 집단, 성장기울기가 상승하는 집단, 성장기울기가 하락하는 집단 등으로 집단을 구분하여 분석이 가능하다. 데이터의 특성에 따라 집단의 수는 달리 나타나고 모형 적합도에 따라 최적의 하위집단수를 정하게 된다.

이외에도 2차요인 잠재성장모형이나 비연속 사건사모형 등의 다양한 잠재성장모형이 제안되고 있다.

2.2 문헌정보학 분야에서의 잠재성장모형의 활용

국내의 문헌정보학 분야에서 종단자료를 이용한 연구는 다수 진행되었다. 최근 빅데이터 분석이 화두로 등장하면서 시계열 정보를 포함하고 있는 데이터 세트가 생산되면서 종단자료를 분석하는 것이 하나의 트렌드가 되었다. 또한 기존의 연구주제 중의 하나였던 인용분석이나 기술동향 분석 등은 넓게는 이러한 종단자료를 이용한 연구의 한 유형이라 할 수 있다. 종단연구의 특징이 특정 기간 동안 관심이 있는 요인에 대해서 반복적으로 관찰하는 것이라는 점에서 이용자의 행태에 대한 연구가 주요한 영역 중의 하나이다. 대표적인 예로 Kuhlthau (1988)는 이용자들의 정보탐색 과정을 관찰하여 정보탐색모형(Information Search Process Model)을 제안하였다. 첫 번째 연구에 참여한 고등학생을 대상으로 정보탐색행위를 조사하고 이들의 대학교 졸업 후의 정보탐색행위와 비교하여 연구하였다. 개인 저널이나 질문지, 인터뷰 등 다양한 방법을 적용했지만 질적인 분석방법을 통해 모형이 개발되고 구체화되었다.

이외에도 양적 분석방법을 적용한 종단연구가 다수 진행되었지만 국내외 문헌정보학계에서 잠재성장모형을 적용한 연구는 찾아보기 쉽지 않다. 최근 종단연구데이터를 이용한 대표적인 분석방법으로 인식되고 있는 잠재성장모형은 통계학 분야뿐만 아니라 사회과학 영역에서도 많은 관심을 받고 있다. 따라서 잠재성장모형에 대한 소개와 그 활용은 문헌정보학계에 새로운 연구 분석방법을 제안하고 이용자에 대한 다른 시각에서의 이해를 가능하게 할 것이라는 점에서 의미가 있을 것이다.

3. 연구방법론

3.1 데이터 수집

본 연구는 공공도서관의 성과를 분석하기 위한 프로젝트의 일환으로 진행이 되었으며 성과를 측정하기 위한 방법으로 잠재성장모형을 적용한다는 점에서 그에 맞는 종단자료의 수집이 요구된다. 모형의 적용을 위해 서울 소재 지자체 소속 8개 도서관이 소장하고 있는 이용자와 대출기록을 활용하였다. 국가도서관통계시스템을 통해 연구 대상 지역 도서관의 특성을 분석한 결과는 <표 2>와 같다.

도서관 데이터베이스에는 도서관에 소장되어 있는 장서에 대한 정보와 이용자에 대한 정보, 그리고 이용자가 장서를 이용한 대출에 대한 정보를 소장하고 있다. 대출과 관련된 사항들, 즉 대출일시 및 대출도서관, 대출자 정보, 장서 정보, 반납 정보 등이 포함되어 있다. 수집된 도서관 데이터는 2002년 3월부터 2015년 5월까지

〈표 2〉 연구 대상 도서관 통계

	2010	2011	2012	2013	2014
도서관수(개)	4	5	5	6	8
장서수(권)	284,694	338,241	382,374	431,900	516,361
대출자수(명)	138,016	189,877	49,321	39,822	59,703
대출권수(권)	832,306	855,623	812,887	853,159	888,843

출처: 국가도서관통계시스템(<http://www.libsta.go.kr>)

지의 도서관 대출데이터로 208,812명의 이용자가 이용한 대출기록을 포함하고 있다. 도서관 대출데이터에서 이용자를 식별할 수 있는 이용자 ID 등의 개인정보가 포함되어 있다는 점에서 개인정보보호를 위한 조치가 먼저 진행되었다. 이용자 ID는 연구자에게 식별이 되지 않도록 도서관에서 다른 값을 부여하였고 연구자들에게는 새로운 값이 부여된 상태로 제공되었다. 이용자 데이터에서 대출량 이외에도 이용자의 특징인 성별, 나이 등의 정보를 분석에 활용하였다. 연구기간을 5년으로 설정하고 1월부터 12월까지 일년 단위의 데이터를 모두 가지고 있으며 최근의 대출기록을 보여주는 2010년부터 2014년 데이터만을 분석에 사용하였다. 또한 5년 동안 대출을 한번도 하지 않은 이용자는 대출량 변화를 분석할 수 없다는 점에서 분석 대상 데이터에서 삭제하였다. 최종적으로 도서관에 등록은 되어있지만 연구기간 내에 대출을 하지 않은 112,850명을 제외한 95,962명의 대출데이터가 분석에 사용되었다.

3.2 데이터 분석

대출량의 성장궤적은 조사 초기의 개인별 대출량이 시간이 흐름에 따라 어떻게 변화하는지

를 분석하는 방법이다. 잠재성장모형의 기본적인 분석모형으로 관심매락변수가 포함되지 않은 무조건모형을 적용하였다. 무조건모형은 시간의 변화에 따른 관측변인과 잠재변인 간의 경로계수를 설정하는데 있어서 일정한 비율로 증가하는 선형성을 기반으로 한다. 이차함수모형은 경로계수를 제공함으로써 선형이 아닌 비선형의 증가나 감소로 변화를 나타낸다. 두 모형을 적용하여 어떠한 모형이 이용자들의 대출량의 변화를 잘 보여주는지를 측정하고 카이제곱 검정을 통해 통계적 검증을 진행했다.

무조건모형에 관심매락변인(x)으로 성별을 포함하여 성별에 따른 대출량의 차이를 분석하는 조건모형을 이용하여 분석하였다. 이외에도 이질적인 데이터에서 동일한 특성을 갖는 잠재계층을 식별하는데 사용하는 분석방법인 성장혼합모형을 통해 도서 대출량의 변화에서 차이를 보이는 집단을 식별하고 각 집단별 특징을 분석했다.

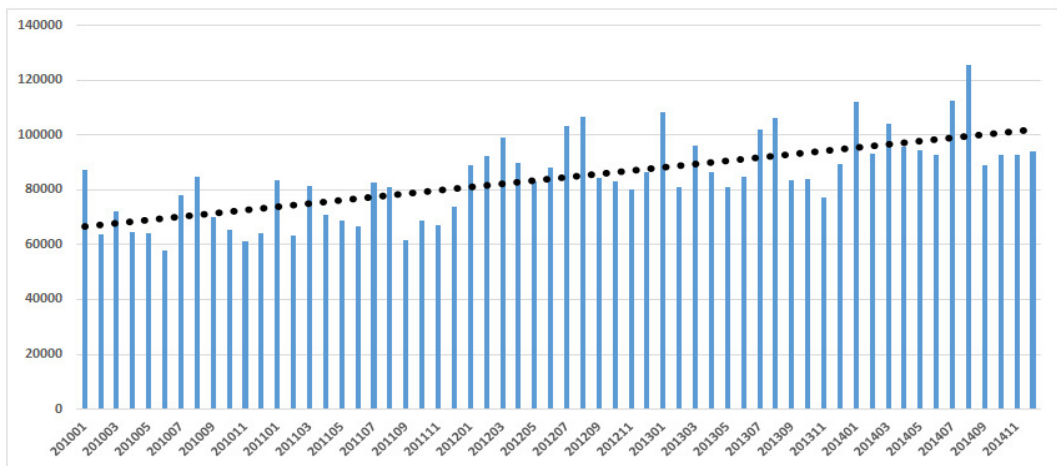
도서관 데이터를 정제하는 전처리작업은 엑셀을 이용하였고 잠재성장모형을 적용한 도서 대출량 변화분석을 위해 SPSS 23.0과 Mplus 7.0을 사용하였다. Mplus에서 모형을 추정하기 위해서 최대우도추정기법(Maximum likelihood estimation)을 적용하였다.

4. 도서 대출량 변화 분석 및 논의

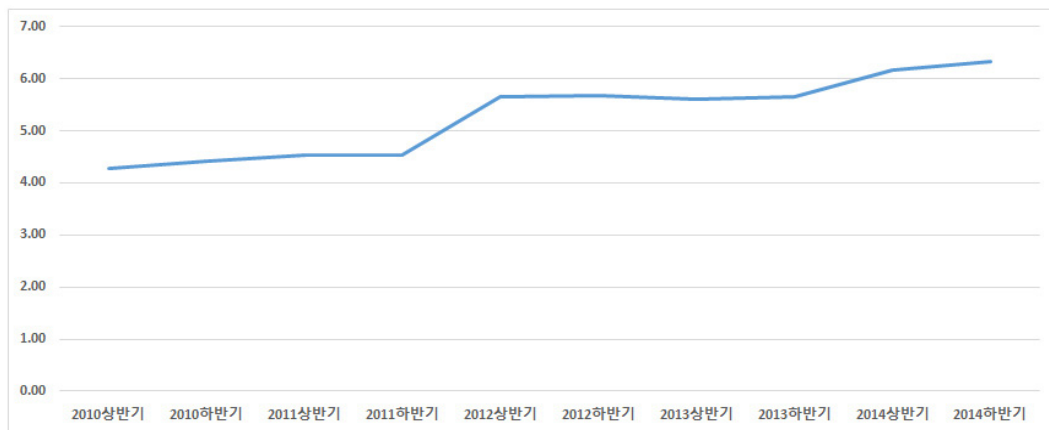
4.1 도서 대출에 대한 기술통계량

연구대상 지역의 도서관 이용자는 도서를 얼마나 대출할까? 월별 대출량에 차이가 있을까? 이러한 질문에 대해서 도서관 데이터는 시기별로 대출량을 계산하여 양의 증감을 확인할 수 있

도록 도와준다. <그림 2>는 2010년부터 2014년까지 도서관의 도서 대출량에 대한 월별 증감을 보여주는 그래프이다. 월별로 대출량은 차이를 보이고 있으며 매년 1월과 7월, 8월에 대출량이 상승하는 것을 확인할 수 있다. 점선은 월별 대출량의 추세를 나타내는 것으로 대출량이 시간이 지남에 따라 상승하는 경향이 있음을 알 수 있다. <그림 3>은 반기별로 평균 대출량의 변화를



<그림 2> 월별 대출량 변화



<그림 3> 반기별 대출량 평균 변화

〈표 3〉 반기별 도서대출량 공분산 행렬

	2010 상반기	2010 하반기	2011 상반기	2011 하반기	2012 상반기	2012 하반기	2013 상반기	2013 하반기	2014 상반기	2014 하반기
2010 상반기	156.31 (1.00)	114.78 (.711)	86.79 (.533)	70.13 (.427)	61.31 (.317)	49.84 (.263)	41.64 (.223)	34.48 (.182)	27.94 (.127)	19.39 (.080)
2010 하반기	114.78 (.711)	166.84 (1.00)	114.40 (.680)	88.09 (.519)	75.96 (.381)	62.60 (.320)	52.83 (.273)	43.76 (.223)	35.85 (.158)	27.07 (.108)
2011 상반기	86.79 (.533)	114.40 (.680)	169.84 (1.00)	122.58 (.716)	106.01 (.527)	83.91 (.425)	70.23 (.360)	59.40 (.300)	51.85 (.227)	39.77 (.157)
2011 하반기	70.13 (.427)	88.09 (.519)	122.58 (.716)	172.64 (1.00)	136.27 (.671)	106.04 (.532)	88.13 (.448)	75.81 (.380)	67.99 (.295)	54.95 (.216)
2012 상반기	61.31 (.317)	75.96 (.381)	106.01 (.527)	136.27 (.671)	238.55 (1.00)	168.40 (.719)	134.37 (.581)	111.56 (.476)	103.66 (.383)	83.36 (.278)
2012 하반기	49.84 (.263)	62.60 (.320)	83.91 (.425)	106.04 (.532)	168.40 (.719)	229.97 (1.00)	160.61 (.708)	127.79 (.555)	118.86 (.447)	94.60 (.322)
2013 상반기	41.64 (.223)	52.83 (.273)	70.23 (.360)	88.13 (.448)	134.37 (.581)	160.61 (.708)	223.87 (1.00)	166.62 (.734)	152.48 (.581)	125.91 (.434)
2013 하반기	34.48 (.182)	43.76 (.223)	59.40 (.300)	75.81 (.380)	111.56 (.476)	127.79 (.555)	166.62 (.734)	230.22 (1.00)	194.98 (.733)	162.58 (.553)
2014 상반기	27.94 (.127)	35.85 (.158)	51.85 (.227)	67.99 (.295)	103.66 (.383)	118.86 (.447)	152.48 (.581)	194.98 (.733)	307.41 (1.00)	243.39 (.716)
2014 하반기	19.39 (.080)	27.07 (.108)	39.77 (.157)	54.95 (.216)	83.36 (.278)	94.60 (.322)	125.91 (.434)	162.58 (.553)	243.39 (.716)	375.54 (1.00)

주. 괄호는 상관계수. 모든 계수가 0.001 수준에서 유의함(양측)

보여준다. 전체적으로 평균 대출량이 지속적으로 상승함을 알 수 있다. 2011년 하반기에서 2012년 상반기로 넘어가면서 대출량이 급격히 증가하였고 분석기간 동안 대출량의 성장이 감소와 증가를 반복하는 것으로 나타났다. 〈표 3〉은 반기별 도서대출량의 공분산행렬을 나타낸다. 분산은 2012년 하반기에 전반기 대비 감소하는 경향이 있었지만 전체적으로 증가하는 것으로 나타났다. 이는 통계적 검증이 필요하지만 분산 증가는 시간이 지남에 따라 도서 대출량의 격차가 점점 커진다는 것을 의미한다.

4.2 도서 대출량의 성장변화 추정

모형적합도 측면에서 선형모형과 비선형모형

은 자료를 잘 설명하고 있는 것으로 나타나지 않았다. 〈표 1〉에 제시된 값과 비교했을 때, TLI와 CFI 값이 두 모형 모두 0.9보다 작고 RMSEA는 0.1보다 크다는 점에서 모형적합도가 우수하다고 할 수 없다.

도서 대출량에 대한 변화를 측정하기 위해 무조건모형에서 선형모형과 비선형모형을 적용한 결과는 〈표 4〉와 같다. 두 모형을 비교했을 때, 모형적합도 면에서는 이차함수 기반의 비선형모형이 데이터를 더 잘 설명하고 있는 것으로 나타났다. 두 모형의 성장계적이 통계적으로 유의한 차이를 보이는지를 검정한 결과는 〈표 5〉와 같다. 두 모형간의 카이제곱의 차이는 54,299.78이고 자유도 차이는 4로 나타났으며 이는 99% 신뢰수준에서 통계적으로 유의한 것으로 나타났다.

〈표 4〉 무조건모형에 의한 선형모형과 비선형모형 분석결과

	선형모형	비선형모형
절편요인 평균	4.19***	4.21***
선형 기울기요인 평균	0.24***	0.24***
비선형 기울기요인 평균		-0.00
절편요인 분산	132.91***	129.89***
선형 기울기요인 분산	4.67***	21.90***
비선형 기울기요인 분산		0.27***
절편요인/선형 기울기요인 공분산	-14.26(-0.57)***	-22.37***
절편요인/비선형 기울기요인 공분산		1.10***
선형기울기요인/비선형 기울기요인 공분산		-2.16***
모형적합도	df=50	df=46
카이제곱	122596.60***	68296.82***
AIC	7365109.13	7310817.35
BIC	7365251.13	7310997.22
TLI	0.82	0.89
CFI	0.80	0.89
SRMR	0.08	0.06
RMSEA	0.16	0.125

주. 괄호는 상관계수. * p<0.05, ** p<0.01, *** p<0.001 (양측).

〈표 5〉 모형간 카이제곱 차이 검정

	x^2	df	Δx^2	Δdf	p
선형모형	122,596.60	50	54,299.78	4	< 0.001
비선형모형	68,296.82	46			

선형모형에서 절편요인의 평균은 4.19, 기울기요인의 평균은 0.24로 나타났고 이는 모두 통계적으로 유의했다. 즉, 대출량은 그 변화가 작지만 점차 증가하는 것으로 나타났다. 절편과 기울기의 공분산이 -14.26으로 나타났고 두 요인 간의 상관은 통계적으로 유의했다. 표준화된 추정치에서의 두 요인의 상관계수는 -0.57로 강한 상관관계를 나타냈다. 또한 부적 상관관계를 보였는데 이는 초기에 많은 대출을 한 사람들은 감소하는 경향을 보이고 적은 대출을 한 사람들은 성장하는 경향을 보임으로써 그 격차가 줄어들고 있음을 알 수 있다.

또한 절편요인의 분산과 기울기요인의 분산을 비교했을 때, 기울기요인의 분산이 상대적으로 매우 작다는 점은 대출량 변화의 차이가 시작점 대출량의 차이보다 크지 않다는 것을 의미한다.

도서관 성과의 관점에서 분석했을 때, 지속적으로 증가하는 대출량은 연구 대상 지역 도서관의 이용자들이 다양한 도서관의 서비스를 통해 발생하는 연구 대상 도서관의 성과라 할 수 있다. 이용자들이 대출한 도서의 수를 지표로 그 변화추이를 보여주는 국가도서관통계시

시스템의 '1관당 대출도서 수'는 증가와 감소를 반복하는데 이와는 다른 결과를 보이고 있는 것이다(〈표 1〉 참조). 이러한 결과의 차이를 만드는 것은 사용된 지표의 차이로 발생하는 것으로 추측된다. 본 연구에서 사용된 대출량은 대출된 한 권의 도서가 여러 번 대출되었을 때 그 각각을 지표의 측정값으로 산출하는 반면에 대출도서의 수에 있어서는 이를 반영하지 못한다. 따라서 국가도서관통계시스템의 평가지표가 도서관에서의 대출의 양상을 제대로 반영하고 있는지에 대해서 추가적인 연구가 진행되어야 할 것이다.

대출량의 변화에 있어서 성별에 따른 차이를 분석하였고 그 결과는 〈표 6〉과 같다. 여성과 남성의 절편요인의 평균추정치는 각각 4.23, 4.18로, 여성이 남성보다 조금 높게 나타났다. 기울기요인의 평균추정치는 여성이 0.23, 남성이 0.24로, 남성이 여성보다 조금 높게 나타났다. 이는 초기에 여성의 대출량이 높지만 남성이 대출량이 조금 더 빠르게 증가한다는 것을 알 수 있으나 그 차이는 크지 않은 것으로 나타났다.

모형적합도에 있어서 TLI와 CLI의 값이 0.9보다 작고 RMSEA는 0.1보다 크다는 점에서 성별에 따른 구분이 도서 대출량을 설명하는데 적합한 모형이라고 보기 어렵다는 것을 보여주고 있다.

4.3 성장혼합모형을 이용한 잠재계층분석

도서대출량의 변화에 있어서 이질적인 변화를 보이는 집단을 찾고 그 특징을 분석하기 위하여 성장혼합모형을 적용하였다. 성장혼합모형을 적용할 때, 첫 번째는 잠재집단의 수를 추정하는 것이다. 추정을 위해서 다양한 방법들이 제안되었는데, 대표적으로 BIC, Lo-Mendell-Rubin(LMR), Bootstrap Likelihood Ratio 검증(BLRT) 등의 적합도지수가 집단 수 추정의 정확성이 높다고 보고되고 있다(Shin 2014). BIC 적합도지수를 적용하는 경우, 집단 수가 늘어남에 따라 BIC값이 감소하다가 늘어나는 지점에서의 최적의 집단 수가 결정된다. 본 연구에서는 집단의 개수가 10개 이상인데도 불구하고

〈표 6〉 조건모형에 의한 성별차이 분석

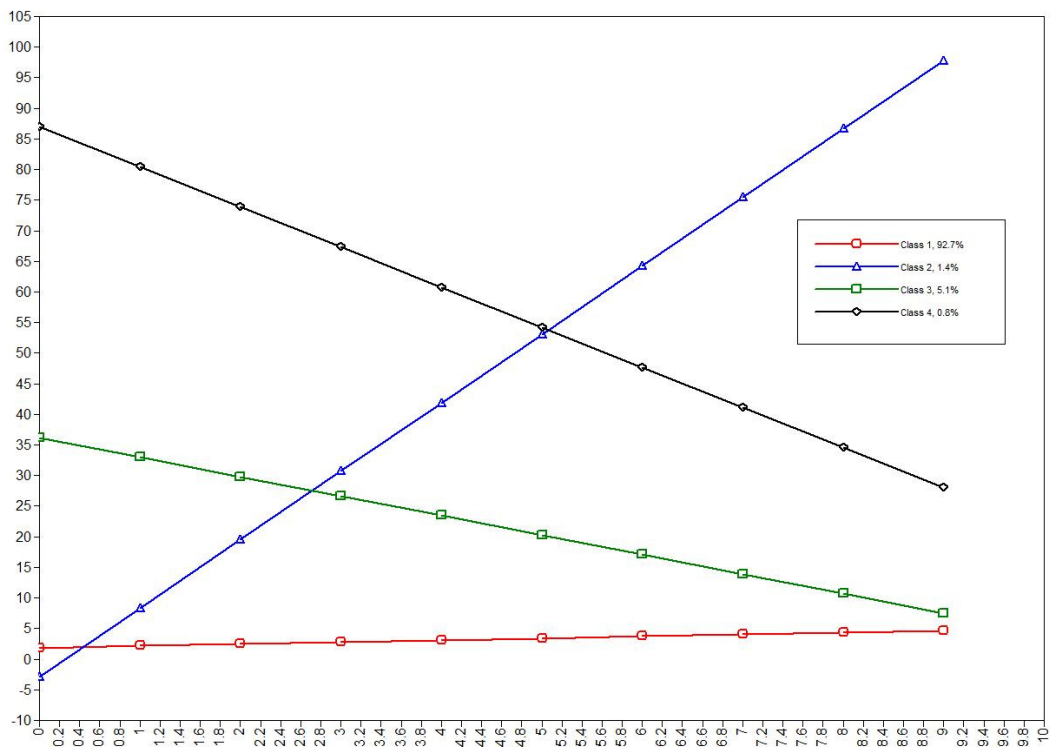
	여성	남성
절편요인 평균	4.23***	4.18***
선형 기울기요인 평균	0.23***	0.24***
절편요인/선형 기울기요인 공분산	-14.25(-0.57)***	
모형적합도	df = 58	
카이제곱	122,966.81***	
AIC	7,365,076.98	
BIC	7,365,237.92	
TLI	0.81	
CFI	0.80	
SRMR	0.08	
RMSEA	0.16	

BIC값이 계속적으로 감소하는 추세를 보였다. 10개 이상으로 집단을 구분하는 것은 역설적이게도 집단간 차이를 모호하게 한다는 점에서 BIC를 적용하기 보다는 집단을 구분하고 데이터의 특성을 보여줄 수 있는 집단의 개수를 임의로 선정하는 것이 적합하다고 판단하였다. 집단의 개수를 3개, 4개, 5개로 순차적으로 늘려가면서 집단의 특성을 분석했을 때, 4개의 경우에서 공통적인 특성, 증가의 기울기가 큰 경우와 작은 경우, 감소의 기울기가 큰 경우와 작은 경우가 발생하였다. 따라서 집단의 수를 4개로 결정하였고 그 결과는 <그림 4>, <표 7>과 같다.

전체 이용자의 92.7%는 1.84의 평균 절편에서 시작하여 상대적으로 변화 수준이 낮은 0.31

의 평균 기울기로 대출량이 증가하는 집단이다. 이에 비해 전체의 1.4%만을 차지하는 집단 2는 11.35의 매우 높은 수준의 성장변화를 보이고 있다. 다음으로 5.1%를 차지하는 집단 3은 완만하게 감소하는 집단으로 36.15의 절편요인에서 시작하여 점차 대출량이 감소하는 집단이다. 대출량이 감소하지만 집단 3에 비해 기울기요인이 큰 집단 4는 전체의 0.8%를 차지하는 것으로 나타났다.

각 집단을 구성하는 구성원의 특징을 분석하기 위해 SPSS에서 집단별 연령별 특성을 분석하였고 그 결과는 <표 8>과 같다. 집단 1은 30대의 비율이 가장 높았으며 30대는 전체 이용자 중에서 차지하는 비율도 가장 높았다. 집단 2에



<그림 4> 하위계층별 평균 성장궤적

〈표 7〉 집단별 요인 추정치

모수 \ 집단	집단 1 (n=89,049: 92.7%)	집단 2 (n=1,331: 1.4%)	집단 3 (n=4,802: 5.1%)	집단 4 (n=760: 0.8%)
절편요인	1.84	-3.0	36.15	86.885
기울기요인	0.31	11.35	-3.18	-6.56

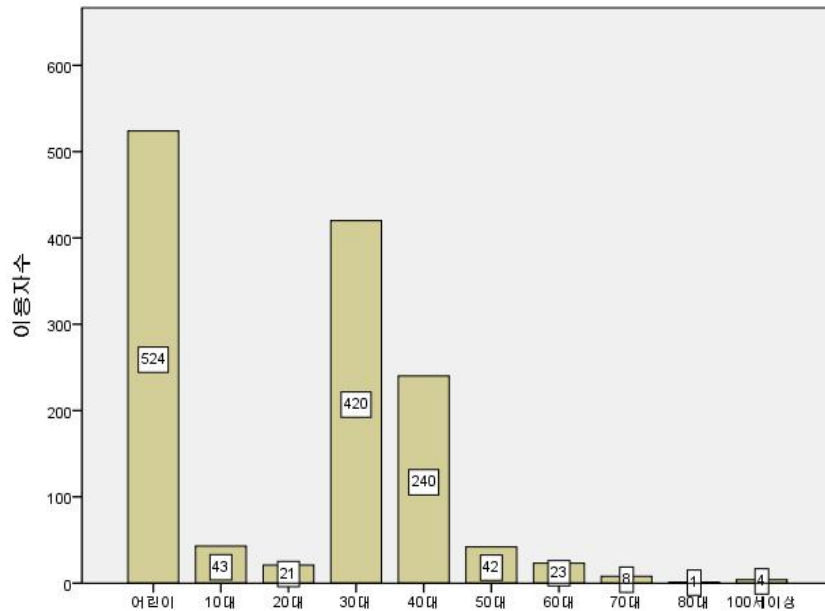
〈표 8〉 하위집단별 연령 분포

	집단 1		집단 2		집단 3		집단 4	
	이용자수	%	이용자수	%	이용자수	%	이용자수	%
10세미만	15,405	17.3%	524	39.5%	1,124	23.4%	241	32.0%
10대	17,448	19.6%	43	3.2%	1,051	21.9%	146	19.4%
20대	15,749	17.7%	21	1.6%	187	3.9%	9	1.2%
30대	22,114	24.8%	420	31.7%	1,041	21.7%	145	19.3%
40대	13,327	15.0%	240	18.1%	1,160	24.2%	176	23.4%
50대	3,355	3.8%	42	3.2%	156	3.3%	20	2.7%
60대	1,086	1.2%	23	1.7%	57	1.2%	14	1.9%
70대	340	.4%	8	.6%	18	.4%	-	0.0%
80대	19	.0%	1	.1%	1	.0%	-	0.0%
90대	2	.0%	-	0.0%	-	0.0%	-	0.0%
100세이상	204	.2%	4	.3%	3	.1%	2	.3%
결측치	13	.0%	-	0.0%	-	0.0%	-	0.0%

나타나는 특징은 10세 미만의 어린이의 비율이 39.5%로 높았고 30대의 비율도 31.7%로 높게 나타났다는 것이다. 10세 미만의 어린이가 읽는 책 중에 그림책이 포함된다는 점에서 1회 대출량이 많을 것으로 예상되고 시간이 흐름에 따라 도서관 방문 빈도도 높아질 것이라는 점에서 대출량이 가파르게 상승하는 집단의 핵심 연령층이라 할 수 있다. 그러나 30대(31.7%)와 40대(18.1%)의 경우에는 이러한 특징을 보이기 어렵다는 점에서 의외의 결과라 할 수 있다. 이는 도서관에서의 도서 대출 행태를 일정부분 반영한 것으로 추측된다. 30대, 40대의 아이들은 대부분 10세 미만이 될 확률이 높다는 점에서 부

모세대인 3,40대가 자녀들을 위한 도서를 대출하는 모습이 도서관 현장에서 종종 발견된다는 점이 이를 뒷받침할 수 있을 것이다.

집단 2와 함께 주목할 점은 집단 4에서 나타나는 특징이다. 집단 4는 시작점에서 높은 대출량을 보이지만 대출량이 상대적으로 가파르게 하락하는 집단이다. 이들 집단에서 가장 높은 비율을 차지하는 연령대는 10세 미만의 어린이(32.0%)로 나타났다는 점에서 앞서 집단 2에서 나타난 결과와 반대의 경향을 보이고 있다. 이를 어느 정도 설명해줄 수 있는 데이터가 집단 4에 포함되어 있는 10대 학생들의 비율이다. 집단 2에서는 3.2%의 낮은 비율을 차지했는데



〈그림 5〉 집단 2에 속하는 이용자의 연령 분포

비해 집단 4에서는 19.4%로 높게 나타났다. 또한 집단 4만큼의 급격한 대출감소량을 보이지는 않지만 눈에 띄는 감소량을 보이는 집단 3에서도 21.9%의 높은 비율을 차지하고 있다. 연령의 기준점이 연구대상 기준의 시작점인 2010년이라는 점에서 10세미만 중 초등학생들이 성장하면서 중학생이 되고 대출량이 급격히 떨어지는 것이라 판단된다. 그러나 이 결과는 분석 결과를 바탕으로 한 또 다른 가설이라는 점에서 데이터를 통한 검증이 요구된다.

하위집단을 파악하고 이들 집단의 변화를 분석한 혼합성장모형에서 관심있게 볼 수 있는 또 다른 요소로 집단 2가 가지는 가파른 성장 변화가 전체의 성장변화에 어느 정도 영향을 미치고 있는가일 것이다. 기본 선형성장모형에서 절편 요인은 4.19, 기울기 요인은 0.24로 낮게 나타났다. 특히 기울기 요인의 평균 변화가 0.24로 낮

다는 점에서 집단 2를 제거했을 때의 변화를 측정하였다. 전체 집단 중에서 집단 2를 제거하고 선형성장분석을 다시 실행한 결과, 절편요인은 4.30으로 상승하고 기울기요인은 0.081로 하락하였다. 기울기요인이 전체 변화에서 67%정도 하락한다는 점에서 집단 2의 대출 상승이 전체 이용자의 대출량의 평균 기울기를 받치고 있는 상황임을 짐작할 수 있다.

혼합성장모형을 적용하여 도출된 집단의 구분은 선형모형과 비선형모형을 적용하여 도출된 대출량의 지속적인 증가가 모든 집단에 나타나는 현상이 아님을 보여주고 있다. 일부 집단에서는 급격히 감소하는 특성을 나타내고 있고 완만한 감소, 급격한 증가 등의 다양한 양상을 보이고 있다는 점은 각각의 집단이 가진 특성의 분석이 추가적으로 수행되어야 함을 제시하고 있다. 본 연구에서는 연령을 기준으로 그

특성을 분석하였는데 성별, 사회경제적 지위 등의 요소를 추가하여 분석한다면 대출량의 변화를 이끌 수 있는 이용자 특성별 맞춤형 제안을 할 수 있을 것으로 기대된다.

5. 결론

본 연구는 도서관 활동을 담고 있는 데이터를 활용하여 도서관의 성과를 측정하는데 있어서 잠재성장모형을 적용하는 방법론을 제시하였다. 잠재성장모형은 사회과학 분야에서 그 활용도가 높아지고 있는데 특히, 전체 성장의 관점이 아닌 개인 성장의 관점에서 분석이 가능하고 결측치나 측정시점의 차이 문제를 해결할 수 있다는 점에서 다른 방법론에 비해 강점을 가지고 있다.

따라서 본 연구는 잠재성장모형을 적용하여 종단자료 성격을 가지고 있는 도서관 데이터 중 대출자료를 분석하여 이용자 개인의 대출량의 성장패턴을 분석하였다. 분석의 방법으로 무조건모형을 기반으로 하는 선형성장모형과 비선형성장모형을 적용하였고, 그 결과 절편요인은 4.19, 기울기 요인은 0.24로 나타났다. 이는 2010년 이용자들은 평균 4.19권의 도서를 대출하였고 반기별로 0.24권의 평균 기울기만큼 도서 대출이 증가함을 추정할 수 있다는 의미이다. 성별에 따른 비교에서 여성이 남성보다 절편요인은 조금 높았으나 기울기 요인은 남성이 여성보다

더 높게 나타났다. 덧붙여 5년 동안의 도서관 이용자를 변화의 유형에 따라 4개의 집단으로 구분했을 때, 대부분의 이용자들은 낮은 도서 대출률에서 시작하여 큰 변화없이 낮은 기울기 변화를 보였다. 변화율의 정적인 상승과 관련하여 높은 평균 기울기의 변화를 보이는 집단이 발생했는데 이들 집단은 초기 대출률이 높지 않았으나 시간이 흐름에 따라 대출률이 급속도로 증가하는 집단으로 10세 미만의 어린이와 이들을 자식으로 둔 부모들의 대출 패턴에 기인한 것으로 보인다.

본 연구가 잠재성장모형을 이용하여 도서관 데이터를 활용할 수 있는 가능성을 제공하고 있다는 점에서 가치가 있고 대출 패턴에서의 특성을 제시하고 있다는 점에서 의미가 있다. 그러나 도서관 데이터가 가지고 있는 특징들, 예를 들면 초등학생과 중학생, 고등학생에게 있어서의 대출량 변화를 분석하지 못했다는 아쉬움이 있다. 또한 신택수(2014)의 연구에서 언급되어 있듯이, 비선형모형의 경우에도 해석 방식과 다중공선성 문제로 논란이 있다는 점에서 비구조화모형과 같이 간명한 형태로 귀적을 추정할 수 있는 모형의 적용을 통한 대출행태 분석이 요구된다. 현재 잠재성장모형이 다른 분야에서 많이 적용되고 새로운 기법들이 연구된다는 점에서 이에 대한 문헌정보학계의 적용을 통해서 학문적 발전을 위한 노력이 필요한 시점이다.

참 고 문 헌

- 김수영. 2017. 『구조방정식 모형의 기본과 확장: MPLUS 예제와 함께』. 서울: 학지사.
- 문화체육관광부. 국가도서관통계시스템. 2018. [online] [cited 2018. 11. 27.]
<<http://www.libsta.go.kr>>
- 신태수. 2014. 인간발달연구에서의 종단자료 분석: 잠재성장모형을 중심으로 『인간발달연구』, 21(3): 1-28.
- Hu, L. T. and Bentler, P. M. 1999. "Cutoff Criteria for Fit Indexes in Covariance Structure Analysis: Conventional Criteria versus New Alternatives." *Structural Equation Modeling: a Multidisciplinary Journal*, 6(1): 1-55.
- Kuhlthau, C. C. 1988. "Longitudinal Case Studies of the Information Search Process of Users in Libraries." *Library and Information Science Research*, 10(3): 257-304.
- McDonald, J. D. 2008. "Measuring Personality Constructs: The Advantages and Disadvantages of Self-reports, Informant Reports and Behavioural Assessments." *Enquire*, 1(1): 1-18.
- Mischel, W. 2013. *Personality and Assessment*. New York: Psychology Press.
- Ployhart, R. E. and Vandenberg, R. J. 2010. "Longitudinal Research: The Theory, Design, and Analysis of Change." *Journal of Management*, 36(1): 94-120.
- Singer, J. D. and Willett, J. B. 2003. *Applied Longitudinal Data Analysis: Modeling Change and Event Occurrence*. Oxford: Oxford University Press.
- Willett, J. B. 1989. "Some Results on Reliability for the Longitudinal Measurement of Change: Implications for the Design of Studies of Individual Growth." *Educational and Psychological Measurement*, 49(3): 587-602.

• 국문 참고자료의 영어 표기

(English translation / romanization of references originally written in Korean)

- Kim, Soo-Yung. 2017. *Fundamentals and Extensions of Structural Equation Modeling with MPLUS Examples*. Seoul: Hakjisa.
- Ministry of Culture, Sports & Tourism. 2018. *National Library Statistics System*. [online] [cited 2018. 11. 27.] <<http://www.libsta.go.kr>>
- Shin, Tacksoo. 2014. "An Analysis of Longitudinal Data in Human Development Study: With a Special focus on Latent Growth Model." *The Korean Journal of the Human Development*, 21(3): 1-28.