

분위 로지스틱 성장 곡선 추정 절차의 비교*

장 승 민†

한림대학교 심리학과

배 소 영

한림대학교 언어청각학부

발달 과정에서 나타나는 개인차 혹은 개인의 상대적 위치를 기술하는 한 가지 방법은 상위 10%, 하위 10% 등에 해당하는 개인들의 분위 성장 궤적들을 추정하여 각 개인이 어느 분위의 성장 곡선에 위치하는지를 기술하는 것이다. 본 연구는 분위 로지스틱 성장 곡선을 추정하는 두 가지 방법인 2단계 비선형 회귀모형과 로지스틱 함수의 역함수로 변환한 값에 대한 분위수 회귀모형을 소개하였다. 실제 자료인 563명의 영아들의 낱말 이해 검사 점수에 대하여 두 방법을 적용한 결과를 비교하여 어느 방법이 성장 곡선 추정에서 더 좋은 수행을 보이는지를 확인하였다. 분석 결과 2단계 비선형 회귀모형은 추정 분위수 역전 현상을 보였지만 분위수 회귀모형은 이러한 현상을 보이지 않았다. 또한 분위수 회귀모형은 비선형 회귀모형에 비해 더 높은 분류 정확도를 보였다. 분위 성장 곡선을 추정하기 위해 통계 모형을 적용하는 것의 필요성을 논의하였다.

주요어 : 성장 곡선, 분위수 회귀모형, 비선형 회귀모형, 로지스틱 성장 함수

* 이 논문은 한국연구재단 신진연구자 지원사업(과제번호:2013S1A5A8024788)과 2013년도 한림대학교 교비 학술연구비(HRF-201301-002)의 지원을 받아 연구되었음.

† 교신저자: 장승민, 한림대학교 심리학과, (200-702) 강원도 춘천시 한림대학길 1

Fax: 033-256-3424, E-mail: jahngs@hallym.ac.kr

신체적, 심리 행동적 특성들의 성장 곡선을 추정하는 것은 인간의 발달과정, 특히 영유아 및 아동기의 성장 과정이나 노년기의 노화 과정을 이해하는데 매우 중요하다. 평균적인 유아나 노인들의 시간에 따른 발달 궤적의 추정은 정상적이고 자연적인 변화를 이해하는데 중요할 뿐만 아니라 비정상적인 변화 유형을 정의하고 그 유발요인을 탐색하는데 기초적인 지식을 제공한다. 생물학적으로나 심리학적으로 인간은 개체 간에 상이한 개인차를 갖고 있으며 발달 과정의 맥락에서는 성장 궤적에서 또한 개인차를 갖는다. 어떤 유아는 다른 유아보다 이른 시기에 사용 어휘가 매우 빠르게 늘어나는가 하면 어떤 유아는 뒤늦게 이와 같은 성장을 경험하며 또 어떤 유아는 상대적으로 완만한 속도의 어휘 성장을 보이기도 한다.

성장에서 나타나는 개인차 혹은 개인의 상대적 위치를 기술하는 한 가지 방법은 평균적인 성장 곡선 또는 상위 10%, 하위 10% 등에 해당하는 분위(quantile) 성장 궤적들을 추정하여 각 개인이 어느 분위의 성장 곡선에 위치하는지를 기술하는 것이다. 예컨대 영유아들의 생물학적 성장이 정상적으로 이루어지고 있는지를 판단하기 위해 소아과 의사는 일정한 기간마다 병원을 방문하여 정기검진을 받는 아이의 신장과 체중이 분위 성장 곡선(궤적)의 어디쯤에 위치하는지를 확인한다. 따라서 심리 행동적 발달 과정의 건전성을 판단하거나 진단하는데 필요한 참조 준거를 마련하기 위해서는 다양한 심리 행동적 특성들의 분위 성장 궤적을 추정하는 것이 필요하다.

심리 행동적 발달 과정을 측정하는 심리검사들은 흔히 검사 점수의 범위가 제한되어 있어 획득 가능한 최소 점수와 최대 점수가 정

해져 있으며, 경우에 따라서는 다수의 피검자들이 최소 점수나 최대 점수를 얻을 수 있도록 고안되기도 한다. 예를 들어 아동의 어휘력을 측정하는 검사가 수십 개의 단어 목록으로 구성되어 있다고 할 때 정상적인 발달과정을 거치는 아동은 영아기에는 0점에 가까운 점수를 얻지만 연령이 증가하여 학령기에 접어들고 입학 후 학년이 올라감에 따라 만점에 가까운 점수를 얻게 된다. 인지 능력의 노화 과정을 측정하는 검사의 경우 노화가 본격화되기 이전에는 만점에 가까운 점수가 기대되고 노화가 진행됨에 따라 점수가 점차 낮아질 것으로 기대된다. 이처럼 최소 점수나 최대 점수를 얻는 것이 자연스러운 심리검사에 의해 측정되는 특성들의 성장 곡선은 일반적인 회귀분석에서 사용되는 직선의 형태나 다항함수의 형태로 표현되기 어려우며 따라서 선형 회귀 모형으로 추정되기에 적절하지 않다. 평균적인 발달의 과정이 수행의 최소값과 최대값으로 수렴하는 경우 발달의 궤적은 다양한 형태의 성장 곡선¹⁾ 중 로지스틱 성장 곡선으로 표현될 수 있으며 따라서 로지스틱 함수에 기반한 회귀모형을 이용하여 적절히 추정될 수 있다. 또한 평균적인 성장 곡선이 로지스틱 함수의 형태를 띠는 발달적 특성의 경우 분위 성장 곡선을 추정하고자 할 때에도 로지스틱 함수에 기반한 회귀모형이 사용될 수 있다.

본 연구는 분위별 발달 성장 곡선을 추정하

1) 선형 함수, 다항 함수, 로지스틱 함수 이외에도 성장 곡선을 추정하기 위해 지수 함수, 고펜트 함수 등이 사용될 수 있으며 특정한 함수를 가정하지 않은 성장 프로파일을 추정하기 위해 요인분석 등이 사용될 수도 있다(Burchinal & Appelbaum, 1991; McArdle & Epstein, 1987).

기 위해 로지스틱 함수를 적용하는 두 가지 방법을 소개하였다. 두 방법은 (1) 첫 단계에서 측정 시기별로 표본 분위수를 구하고 두 번째 단계에서 이 분위수들의 로지스틱 성장 곡선을 추정하는 2단계 비선형 회귀모형과 (2) 로지스틱 함수를 이용한 변환 점수의 분위 회귀선을 추정하는 선형 분위수 회귀모형을 지칭한다. 또한 로지스틱 함수를 적용한 분위별 성장 곡선 추정에 적합한 실제 자료를 이용해 각 방법이 어떻게 적용될 수 있는지 예시하고 각 접근법을 이용하여 분석한 결과를 비교하였다. 특히 각 접근법에 의해 추정된 분위별 성장 곡선에서 추정 분위수 역전 현상의 문제가 없는지 확인하고 두 방법의 결과가 서로 얼마나 일치하는 성장 곡선을 산출하는지 비교하였다.

분위 로지스틱 성장 곡선의 추정

분위 성장 곡선을 추정하는 가장 단순한 절차는 일종의 2단계 추정법을 이용하는 것이다. 이 절차는 연구가 목적으로 하는 연령대의 각 시기별로 검사점수의 표본 분위수를 계산한 후 이 분위수들의 성장 궤적을 추정하는 것이다. 예를 들어 70세 이후의 인지적 기능 저하에 관한 연구를 위해 71세에서 90세까지의 노인들을 대상으로 20개의 연령집단에 대해 각 연령별로 50명씩 모두 1000명의 표본(횡단 자료)에 대해 인지기능 검사를 수행했다고 하자. 분위 성장 곡선의 2단계 추정법은 먼저 각 연령별 표본, 즉 71세 표본 50명, 72세 표본 50명 등에 대해 각기 특정 분위수(예, 제5 백분위수)를 구한 후 71세에서 90세까지의 각 표본에서 얻어진 20개의 특정 분위수를 이용하여 해당 분위의 성장 곡선을 추정하고 이를

다른 분위수들에도 반복하여 적용한다. 즉, 각 연령 표본에서 얻어진 20개의 제5 백분위수 값들을 사용하여 제5 백분위 성장 곡선을 추정하고, 다시 20개의 제10 백분위수 값들을 사용하여 제10 백분위 성장 곡선을 추정하고, 나머지 분위수에 대해서도 같은 절차를 반복한다. 이렇게 해서 추정된 각 분위별 성장 곡선으로 연구자가 목표하는 분위 전체를 포함하는 발달 분위 곡선 차트를 구성할 수 있다. 성장 곡선의 형태가 로지스틱 함수로 잘 표현될 수 있는 경우는 표본 분위수에 대해 로지스틱 함수를 이용한 비선형 회귀분석을 적용하여 성장 곡선을 추정할 수 있다.

Fenson 등(1994)은 영유아의 의사소통능력을 측정하는 맥아더 의사소통발달목록(MacArthur Communicative Development Inventory: M-CDI)을 이용하여 영유아기 의사소통능력의 분위별 발달 곡선을 추정하기 위해 전술한 2단계 추정법을 사용하였다. 예를 들어 이들은 8-16개월의 영유아 659명(남아 325명, 여아 334명)의 단어 이해 검사 점수를 얻은 후, 33명에서 86명 사이의 각 월령별 표본에서 백분위수를 구한 후 각 월령별 특정 백분위수(예, 제90 백분위수)를 로지스틱 성장 모형을 이용해 추정하여 분위별 성장 곡선을 구하였다.

이와 같은 2단계 방법은 절차가 단순하고 일반적인 성장 곡선 모형을 다룰 수 있는 연구자들이 쉽게 적용할 수 있다는 장점이 있는 반면 몇 가지 문제점을 안고 있다. 첫 번째 문제는 각 분위별 성장 곡선을 추정하는데 사용되는 자료의 수가 매우 적다는 것이다. 예를 들어 Fenson 등(1994)의 연구에서 각 분위별 성장 곡선은 8개월에서 16개월 사이의 표본에서 얻은 9개의 분위 점수들을 바탕으로 추정되었는데 이 9개의 개별 분위 점수들은 각각

33명에서 86명 사이의 표본으로부터 계산되었다. 2단계 추정법은 각 시기별 표본수가 충분히 크지 않다면 표집오차에 민감하고, 최종적인 단계인 성장 곡선 추정과정이 1단계에서 추정된 점수를 바탕으로 하기 때문에 모형의 적합성에 대해 적절히 평가하기 어렵다. 뿐만 아니라 각 분위별 성장 곡선의 추정은 다른 분위수의 값에 의해 영향을 받지 않기 때문에 각각의 분위수에 대해 추정된 분위별 기대값이 분위 서열과 불일치하는 결과를 얻을 수 있다. 예를 들어 모형에 의해 추정된 제90 백분위 성장 곡선이 제95 백분위 성장 곡선과 관찰된 연령 시기 내에서 교차하여 특정 연령대에서 추정된 제90 백분위수가 추정된 제95 백분위수보다 큰 값을 갖는 추정 분위수 역전 현상이 나타날 수 있다.²⁾ 이러한 문제는 전체 표본수가 충분히 크더라도 각 시기별 표본수가 적을 때 발생할 가능성이 크며 남녀 등의 하위 집단별 분석이 추가되는 경우 더 쉽게 나타난다.

분위별 성장 곡선을 추정하는 또 다른 방법은 분위수 회귀모형을 이용하는 것이다. 분위수 회귀모형은 Edgeworth와 Laplace 등에 의해 제안된 중위수 회귀법(median regression)을 Koehler와 Bassett(1978)가 확장하여 제안한 이후 최근 경제학, 생태학, 의학을 포함한 다양한 학문 분야에서 그 사용이 증가하고 있는 모형이다(Cade & Noon, 2003; Koehler & Hallock, 2001; Yu, Lu, & Stander, 2003). 일반적으로

인 회귀모형이 결과변인의 예측변인에 대한 조건 평균의 함수를 최소제곱법 등을 사용하여 추정하는 반면, 분위수 회귀법은 예측변인에 대한 조건 중위수나 조건 제90 백분위수 등의 특정 분위수의 함수를 관찰값과 기대값의 차이의 절대값을 바탕으로 하는 특정한 손실함수를 이용하여 추정한다(Hao & Naiman, 2007; Koehler & Hallock, 2001).³⁾ 또한 분위수 회귀법은 오차의 분포에 대한 가정이 필요 없어 관심 변인의 모집분포에 대한 가정 없이도 사용할 수 있다. 분위수 회귀법은 비선형 성장모형을 추정할 수 있는 알고리즘이 제안되기도 하였지만(Koehler & Park, 1996), 일반적으로 상용 프로그램(예, SAS, SPSS)을 이용해 분석할 수 있는 선형 모형으로 사용된다. 그러나 로지스틱 성장모형과 같은 단조변화 모형의 경우 분위수 회귀법이 단조변화함수의 변환에 의해 결과가 달라지지 않기 때문에 결과변인에 해당하는 관찰값을 로지스틱 함수의 역함수를 이용해 변환한 후 선형모형을 적용해 추정할 수 있다(Koehler & Hallock, 2001). 또한 분위수 회귀법은 모형의 추정에 표본자료 전체를 사용한다는 점에서 2단계 추정법의 단점을 피할 수 있다.

다음 절들에서는 실제 자료를 이용하여 2단계 비선형 회귀모형과 변환점수에 대한 분위수 회귀모형이 로지스틱 성장 곡선을 추정하기 위해 어떻게 적용될 수 있는지를 구체적으로 예시하고 이 두 절차의 분석 결과를 비교하였다.

2) 예컨대, 제 90백분위와 제 95백분위 성장 곡선이 각기 서로 다른 기울기를 갖는 선형 함수로 추정된다면 이 두 직선은 특정 시기에 교차할 수 있고 어떤 연령에서는 선형 함수에 의해 추정된 제 90백분위수가 다른 선형 함수에 의해 추정된 제 95백분위수보다 높은 값을 가질 수 있다.

3) 국내 문헌에서 분위수 회귀모형의 추정 절차에 대한 보다 구체적인 설명을 얻고자 하는 경우 장승민과 강연옥(심사중)을 참조할 수 있다.

방 법

표본 및 변인

본 분석에는 Fenson 등(1994)이 개발한 맥아더 의사소통발달목록의 개정판인 맥아더-베이트스 의사소통발달목록(MacArthur-Bates CDI, Fenson et al., 2007)의 한국판 표준화 자료가 사용되었다(배소영, 곽금주, 2011). 이 자료는 8개월에서 17개월의 영아 563명과 18개월에서 36개월의 유아 1,138명 등, 총 1701명을 대상으로 수집되었다. CDI는 영유아의 의사소통능력을 나타내는 제스처, 어휘, 문법 등을 측정하며 대부분의 하위 검사들이 정상적인 발달 과정에서 연령이 증가함에 따라 최고점에 수렴하는 특징을 가지고 있다. 본 연구의 분석에는 이 중 8개월에서 17개월 영아 563명(남아 278명, 여아 285명)의 자료가 사용되었다.⁴⁾ 각 월령별 남녀 표본수는 표 1에 나타난 바와 같다.

분석에는 영아용 낱말 이해 검사의 점수가 사용되었다. 영아용 낱말 이해 검사는 18가지 범주에서 영아들의 생활을 반영하여 선택된 284개의 낱말 항목(내용 낱말 및 문법 형태소)이 체크리스트로 제공되면 영아가 해당 낱말을 이해하고 있는지에 대해 영아의 주 양육자가 판단하여 표시하도록 한다. 따라서 가능한 검사 점수의 범위는 0점에서 284점이 된다. 표 2에는 8개월에서 17개월의 월령별 영아들의 검사 점수 평균과 표준편차가 남녀별로 제시되어 있다.

4) 영아와 유아는 각 하위검사에서 서로 다른 목록으로 검사를 받기 때문에 각기 다른 성장 곡선을 갖는다. 본 예시에서는 영아의 성장 곡선을 다루었다.

분석절차

영아용 낱말 이해 검사 점수의 분위 성장 곡선을 추정하기 위해 Fenson 등(1994)이 제안한 2단계 절차와 분위수 회귀모형을 이용한 단일 단계 절차가 사용되었다. 두 모형에서 가정된 검사 점수의 성장 곡선은 아래의 로지스틱 함수로 추정되었다.

$$\hat{Y}_i = d / [1 + e^{-(\alpha + \beta T_i)}] \quad (1)$$

여기에서 \hat{Y}_i 는 낱말 이해 검사 점수의 각 분위에 대한 성장 곡선을 의미하며 T_i 는 8개월에서 17개월 사이의 월령을 의미한다. d 는 검사점수의 가능한 점근 최대 점수이며 본 검사의 경우 284가 된다. 성장 곡선, 즉 \hat{Y}_i 의 가능한 최저 점수는 0이다. 식 (1)의 성장 함수를 이용하여 7개의 분위수에 대해 분위 성장 곡선을 추정하였다. 7개의 분위수는 제5 백분위수, 제10 백분위수, 제25 백분위수, 제50 백분위수, 제75 백분위수, 제90 백분위수, 및 제95 백분위수가 선정되었다. 각 분위 성장 곡선은 남녀별로 각각 추정되었다.

먼저 각 월령별로 7개의 표본 분위수를 계산하여 얻었다. 각 분위수에 대한 남녀별 표본 통계량이 표 3에 제시되어 있다. 그림 1은 표 3의 표본 분위수를 선도표로 나타낸 것이다. 남아의 경우 15개월 이후의 표본 분위수가 다소 불규칙한 형태를 보였지만 전반적으로 연령에 따른 점수의 차이가 로지스틱 함수의 형태를 띠고 있음을 확인할 수 있다.

2단계 분위 성장 곡선 추정을 위해 표 3에 제시되어 있는 각 표본 분위수에 대해 식 (1)

표 1. 월령별 남녀 표본수

성별	월령(개월)										합계
	8	9	10	11	12	13	14	15	16	17	
남자	16	29	35	30	27	29	27	27	32	26	278
여자	29	27	24	31	28	30	31	29	27	29	285
합계	45	56	59	61	55	59	58	56	59	55	563

표 2. 월령별 낱말 이해 검사 점수의 평균과 표준편차

성별	구분	월령(개월)										전체
		8	9	10	11	12	13	14	15	16	17	
남자	평균	33.8	29.9	31.1	38.8	62.2	77.3	91.6	87.8	141.8	129.2	73.1
	표준편차	32.0	25.1	24.9	38.0	37.5	48.8	45.2	40.5	80.8	72.6	61.8
여자	평균	21.6	23.9	52.8	56.8	74.4	96.6	107.2	132.0	132.7	145.5	85.0
	표준편차	21.1	22.7	36.7	41.2	51.7	58.2	61.9	64.3	68.7	67.3	67.1
전체	평균	25.9	27.0	39.9	47.9	68.4	87.2	99.9	110.7	137.6	137.8	79.1
	표준편차	25.8	24.0	31.8	40.4	45.3	54.2	54.9	58.1	75.0	69.7	64.8

표 3. 남녀별 낱말 이해 검사 점수의 표본 분위수

성별	백분위	월령(개월)									
		8	9	10	11	12	13	14	15	16	17
남자	5	0	3	5	2	7	4	36	36	31	0
	10	0	4	8	4	7	4	37	44	54	34
	25	6.5	7	14	14	32	42	48	58	66.5	76
	50	25.5	27	22	34.5	57	86	83	83	127	135.5
	75	58	50	48	49	81	121	135	111	217.5	164
	90	70	69	74	74	117	148	152	154	259	220
	95	109	79	79	94	125	150	165	161	283	260
	여자	5	1	3	8	10	11	18	5	16	37
여자	10	1	4	15	15	13	26.5	13	50	42	59
	25	7	8	28	29	29	50	54	95	76	95
	50	17	15	49	45	69	86	124	128	139	133
	75	25	36	70	78	108	146	157	177	175	185
	90	57	58	97	120	158	170	178	220	211	248
	95	72	74	97	147	171	192	192	232	263	256

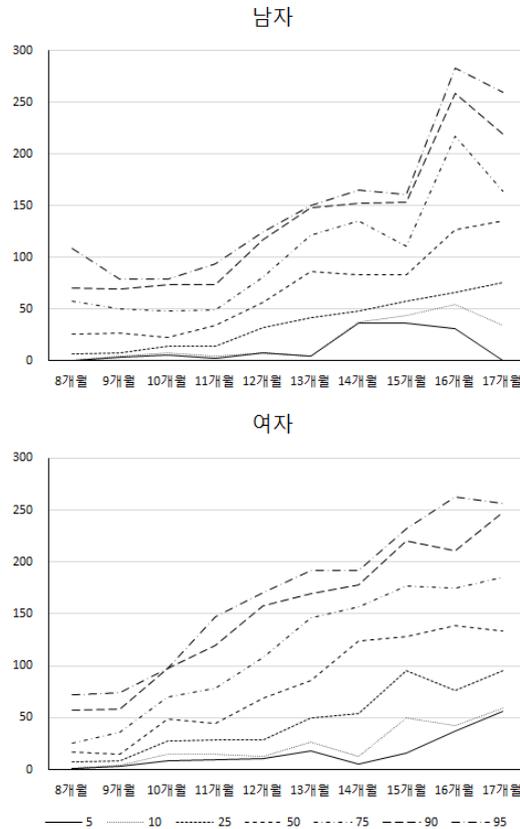


그림 1. 남녀 영아별 낱말 이해 검사 점수의 표본 분위 궤적

의 함수를 이용한 비선형 회귀모형이 적용되었다. 예를 들어 남아의 제50 백분위 성장 곡선의 경우 각 월령별로 얻은 표본 제50 백분위수 10개(8개월-25.5점, 9개월-27점, ..., 17개월-135.5점)를 이용하여 비선형 회귀분석이 실시되었다.

분위수 회귀모형을 이용한 성장 곡선 추정 절차는 다음과 같은 순서로 수행되었다. 먼저 비선형 로지스틱 성장 곡선을 선형 분위수 회귀모형으로 추정하기 위해 검사 점수를 식 (1)의 성장 함수의 역함수를 이용해 변환하였다. 식 (1)의 역함수를 이용한 변환값은 다음과 같이 표현된다.

$$Y_i^* = \log[(d - Y_i) / Y_i] \quad (2)$$

Y_i 값이 커질수록 변환된 값 Y_i^* 는 작아진다는 것에 유의해야 한다. 즉, Y_i^* 의 제5 백분위는 Y_i 의 제95 백분위를 의미한다. 식 (2)를 이용한 변환에서 로그함수 안의 값이 계산되기 위해서는 분모 또는 분자가 0이 되어서는 안 된다. 로그함수 안의 분자는 검사 점수의 가능한 최대 점근 값인 $d(=284)$ 와 검사 점수가 같은 경우, 즉 검사 점수가 284인 경우 0이 된다. 로그함수 안의 분모는 검사 점수가 0인 경우 0이 된다. 따라서 0과 284를 포함한 검사 점수들 전체를 변환하기 위해서는 식 (2)

를 수정하여야 한다. 이와 같은 수정은 다음의 식을 통해 가능하다.

$$Y_i^* = \log[(da - Y_i)/(Y_i + b)] \quad (3)$$

식 (3)에서 a 와 b 는 각각 $(da - Y_i)$ 와 $(Y_i + b)$ 가 가능한 모든 Y_i 에 대하여 0이 아닌 값을 갖게 하면서 각각 $(d - Y_i)$ 와 Y_i 와 실질적으로 차이가 없게 하는 값이어야 한다. 본 연구에서는 $a=1.0001$, $b=.0001$ 이 사용되었다. 식 (3)을 이용해 검사 점수 Y_i 를 Y_i^* 로 변환한 후 다음과 같은 선형 회귀선을 이용하여 성장 곡선을 추정하였다.

$$\hat{Y}_i^* = \alpha^* + \beta^* T_i \quad (4)$$

2단계 추정 절차가 월령별 표본 분위수에 대한 분석인데 반해 분위수 회귀모형은 모든 월령의 자료를 동시에 사용하여 변환 점수의 각 분위수 성장 곡선을 추정한다. 식 (4)를 이용해 각 분위별로 추정된 성장 곡선의 추정값 \hat{Y}_i^* 는 다시 아래의 식 (5)를 이용하여 원 점수의 척도로 역변환되었다.

$$\hat{Y}_i = (da - be^{\hat{Y}_i^*}) / (1 + e^{\hat{Y}_i^*}) \quad (5)$$

결 과⁵⁾

2단계 비선형모형의 추정 결과

표 3의 표본 분위수에 대해 식 (1)을 이용한

2단계 추정 절차의 남녀별 분위 성장 곡선의 모수 추정 결과와 월령별 추정값이 표 4와 표 5에 각각 제시되어 있다. 분석을 위한 프로그램으로는 SAS 9.2의 NLIN 절차가 사용되었다. 표 4의 모든 분위수 성장 곡선의 기울기 추정값이 음수로 나타났다. 이는 월령이 높을수록 추정된 검사 점수의 값이 높게 나타남을 의미한다. 이는 표 5에 제시된 월령별 분위 성장 곡선 추정값을 통해 확인할 수 있다.

한편 2단계 추정 절차에 의해 분위 성장 곡선이 적절히 추정되었다면 표 5에 나타난 낮은 분위의 성장 곡선 추정값은 높은 분위의 성장 곡선 추정값보다 모든 월령에서 낮게 나타나야 한다. 그러나 표 5에서 보듯이 남자의 경우 8개월, 9개월, 10개월에서 추정된 제5 백 분위수가 추정된 제10 백 분위수보다 높게 나타났다. 이는 여러 분위별 성장 곡선이 동시에 추정되지 않고 각각 따로 추정되기 때문에 발생하는 문제로 특히 각 월령별 표본수가 많지 않은 경우에 더 자주 발생한다. 2단계 추정 절차는 각 월령별 표본 분위기를 이용하여 곡선을 추정하기 때문에 본 자료에서와 같이 구분 월령(10개)이 많지 않고 각 월령의 표본수(16에서 32)도 많지 않은 경우 추정 분위수의 역전 현상에 더 취약하다고 볼 수 있다. 그럼 2는 표 5의 분위별 추정 성장 곡선을 남녀별로 도표로 표현한 것이다. 남자의 경우 제5 백분위 성장 곡선과 제10 백분위 성장 곡선이 10개월과 11개월 사이에서 교차함을 확인할 수 있다.

분위수 회귀모형의 추정 결과

월령별 표본 분위기를 이용한 2단계 추정 절차를 사용하지 않고 모든 월령의 자료를 동

5) 본 예시에서 사용된 SAS 명령어를 부록에 제시하였다.

표 4. 2단계 절차를 이용한 분위 성장 곡선 모수 추정 결과

성별	백분위	모수	추정값	표준오차	95% 신뢰구간	
					하한값	상한값
남자	5	α (절편)	5.16	2.16	0.18	10.14
		β (기울기)	-0.16	0.15	-0.50	0.17
	10	a	6.46	1.48	3.04	9.87
		β	-0.29	0.10	-0.51	-0.07
	25	a	5.39	0.38	4.52	6.27
		β	-0.26	0.03	-0.32	-0.21
	50	a	4.79	0.45	3.77	5.82
		β	-0.28	0.03	-0.35	-0.21
	75	a	4.28	0.83	2.38	6.18
		β	-0.29	0.06	-0.43	-0.15
	90	a	4.29	0.73	2.60	5.99
		β	-0.33	0.06	-0.46	-0.20
	95	a	4.04	0.66	2.66	5.42
		β	-0.33	0.05	-0.44	-0.22
여자	5	a	9.50	1.46	6.14	12.85
		β	-0.47	0.09	-0.68	-0.26
	10	a	6.35	0.89	4.31	8.39
		β	-0.29	0.06	-0.43	-0.16
	25	a	5.13	0.60	3.73	6.52
		β	-0.27	0.04	-0.36	-0.17
	50	a	4.48	0.54	3.25	5.72
		β	-0.28	0.04	-0.36	-0.19
	75	a	4.27	0.45	3.23	5.32
		β	-0.31	0.03	-0.39	-0.23
	90	a	4.21	0.33	3.45	4.97
		β	-0.35	0.03	-0.41	-0.29
	95	a	4.46	0.25	3.93	4.98
		β	-0.40	0.02	-0.44	-0.35

표 5. 2단계 절차를 이용한 분위 성장 곡선의 월령별 추정값

성별	백분위	월령(개월)									
		8	9	10	11	12	13	14	15	16	17
남자	5	5.8	6.8	8.0	9.4	10.9	12.8	14.9	17.4	20.2	23.5
	10	4.5	6.0	7.9	10.5	13.8	18.2	23.8	30.9	39.9	50.9
	25	10.2	13.2	16.9	21.6	27.5	34.8	43.7	54.4	66.9	81.3
	50	20.3	26.2	33.6	42.7	53.9	67.1	82.4	99.5	118.2	137.7
	75	35.0	44.9	57.0	71.3	87.8	106.3	126.2	146.7	167.0	186.3
	90	45.5	59.6	76.6	96.3	118.3	141.5	164.7	186.7	206.6	223.7
	95	56.0	72.3	91.4	112.9	135.8	159.1	181.5	201.9	219.7	234.6
여자	5	0.9	1.5	2.4	3.8	6.1	9.6	15.1	23.4	35.8	53.3
	10	5.1	6.9	9.1	12.2	16.1	21.2	27.7	36.1	46.4	59.0
	25	13.6	17.5	22.5	28.7	36.4	45.7	56.9	70.1	85.1	101.9
	50	26.5	33.9	43.1	54.2	67.3	82.5	99.6	118.1	137.5	157.0
	75	40.2	52.1	66.5	83.5	102.8	123.8	145.6	167.3	187.8	206.4
	90	56.1	73.7	94.3	117.6	142.3	167.0	190.2	210.8	228.2	242.3
	95	61.7	83.0	108.1	135.6	163.7	190.0	213.2	232.1	246.9	257.9

시에 사용하여 각 분위수 성장 곡선을 추정할 분위수 회귀모형의 분석결과는 표 6과 표 7에 각각 제시되어 있다. 분석은 SAS 9.2의 QUANTREG 절차를 통해 이루어졌다. 표 6의 분위수 성장 곡선 모수 추정값은 낱말 이해의 원점수 Y_i 에 대한 비선형 회귀모형의 추정값이 아니라 로지스틱 함수의 역함수로 변환된 Y_i^* 에 대한 선형 회귀모형 추정값이다. 여기에서도 기울기 추정값은 모두 음수로 나타났으며 이는 월령이 높을수록 추정된 검사 점수의 값이 높게 나타남을 의미한다. 표 7에 제시된 월령별 분위 성장 곡선 추정값에서 이와 같은 특징을 확인할 수 있다.

한편 2단계 추정 절차에서와는 달리 변환 점수에 대해 분위수 회귀모형을 적용한 분위

성장 곡선 추정값은 분위수 역전 현상이 나타나지 않았다. 표 7의 분위별 성장 곡선 추정값을 보면 앞서 2단계 추정법에서와는 달리 8개월, 9개월, 10개월을 포함하여 남녀 모든 월령에서 추정 분위수 역전 현상이 나타나지 않았다. 분위수 회귀모형 역시 여러 분위별 성장 곡선이 동시에 추정되지 않고 따로 추정되기 때문에 분위수 역전 현상이 발생할 수 있지만 본 자료에서는 구분 월령과 각 월령별 표본수가 많지 않음에도 불구하고 추정 분위수의 역전 현상이 나타나지 않았다. 그림 3은 표 7의 분위별 추정 성장 곡선을 남녀별로 도표로 표현한 것이다. 남자의 제5 백분위 성장 곡선과 제10 백분위 성장 곡선을 포함하여 모든 분위 성장 곡선이 서로 교차하지 않았음을

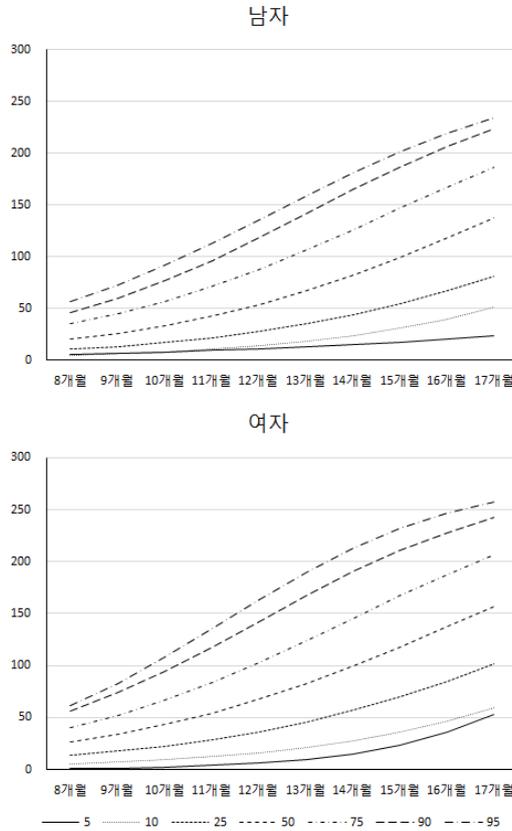


그림 2. 2단계 절차를 이용한 남녀 영아의 분위별 추정 성장 곡선

확인할 수 있다.

두 모형의 추정 분위수의 표본 점수 구분 비율 비교

두 분석 절차가 추정한 분위 성장 곡선이 전체 표본을 목표 분위에 맞게 구분했는지 확인하기 위해 각 분위 성장 곡선 상의 추정값보다 낮은 점수의 비율을 확인해 보았다. 표 8은 두 분석 절차가 추정한 분위수보다 낮은 값을 갖는 점수의 빈도와 백분율을 나타낸다. 표에서 알 수 있듯이 분위수 회귀모형에 의해 추정된 분위수가 2단계 비선형모형에 의해 추

정된 분위수보다 전체 월령 구간에서 검사 점수를 더 잘 구분하는 것으로 나타났다. 2단계 추정법은 남자의 경우 제5 백분위, 제25 백분위, 제50 백분위에서, 여자의 경우 제25 백분위, 제50 백분위에서 추정 분위수보다 낮은 점수의 비율이 해당 백분위와 3% 이상 차이를 보였다. 이는 1단계에서 각 분위수가 계산될 때 월령별로 표본수에서 차이가 있었기 때문에 발생한 것으로 볼 수 있다. 반면 분위수 회귀모형은 거의 모든 백분위에서 추정 분위수보다 낮은 점수의 비율이 해당 백분위와 1% 미만의 차이를 보였다(예외, 남자 제10 백분위). 이는 분위수 회귀모형이 추정 과정에서

표 6. 분위수 회귀모형을 이용한 분위 성장 곡선 모수 추정 결과

성별	백분위	모수	추정값	신뢰구간	
				하한값	상한값
남자	5	α (절편)	9.32	5.43	20.67
		β (기울기)	-0.43	-0.79	0.02
	10	a	7.73	6.30	10.32
		β	-0.39	-0.52	-0.25
	25	a	6.59	5.64	7.47
		β	-0.35	-0.40	-0.27
	50	a	5.16	4.45	5.92
		β	-0.30	-0.36	-0.24
	75	a	4.12	3.33	5.62
		β	-0.27	-0.39	-0.22
	90	a	3.98	3.19	4.89
		β	-0.31	-0.41	-0.25
	95	a	4.43	4.07	5.14
		β	-0.39	-0.50	-0.32
여자	5	a	8.74	6.56	10.94
		β	-0.43	-0.48	-0.26
	10	a	7.13	5.85	8.54
		β	-0.34	-0.45	-0.25
	25	a	6.23	5.35	6.88
		β	-0.34	-0.39	-0.28
	50	a	5.29	4.74	6.26
		β	-0.33	-0.41	-0.30
	75	a	4.90	3.64	5.60
		β	-0.36	-0.42	-0.26
	90	a	4.24	3.06	5.19
		β	-0.36	-0.44	-0.26
	95	a	4.01	3.29	5.72
		β	-0.37	-0.52	-0.31

표 7. 분위수 회귀모형을 이용한 분위 성장 곡선의 월령별 추정값

성별	백분위	월령(개월)									
		8	9	10	11	12	13	14	15	16	17
남자	5	0.8	1.3	1.9	3.0	4.6	7.1	10.8	16.3	24.4	36.0
	10	2.7	4.0	5.8	8.5	12.4	17.8	25.5	36.0	50.0	67.9
	25	6.2	8.7	12.2	17.0	23.5	32.2	43.6	58.0	75.7	96.6
	50	17.0	22.4	29.5	38.4	49.5	62.9	78.8	97.0	117.0	138.0
	75	36.0	45.5	57.0	70.5	85.9	103.2	121.7	141.0	160.3	179.0
	90	51.0	65.2	81.9	100.9	121.7	143.4	165.0	185.6	204.4	220.8
	95	58.9	79.0	102.8	129.2	156.5	182.8	206.4	226.2	242.0	254.0
여자	5	1.4	2.1	3.2	4.9	7.5	11.3	17.0	25.3	37.0	53.1
	10	3.4	4.8	6.7	9.4	13.0	18.0	24.7	33.5	45.0	59.5
	25	8.3	11.6	16.0	22.0	30.0	40.5	53.8	70.3	89.9	112.0
	50	19.0	25.8	34.7	46.2	60.5	77.9	98.0	120.2	143.7	167.0
	75	32.3	44.1	59.0	77.4	99.0	123.0	148.2	173.0	196.0	216.1
	90	56.2	74.0	95.0	118.7	143.7	168.7	192.0	212.6	229.9	243.8
	95	72.0	93.3	117.5	143.2	168.9	192.8	213.9	231.4	245.3	256.0

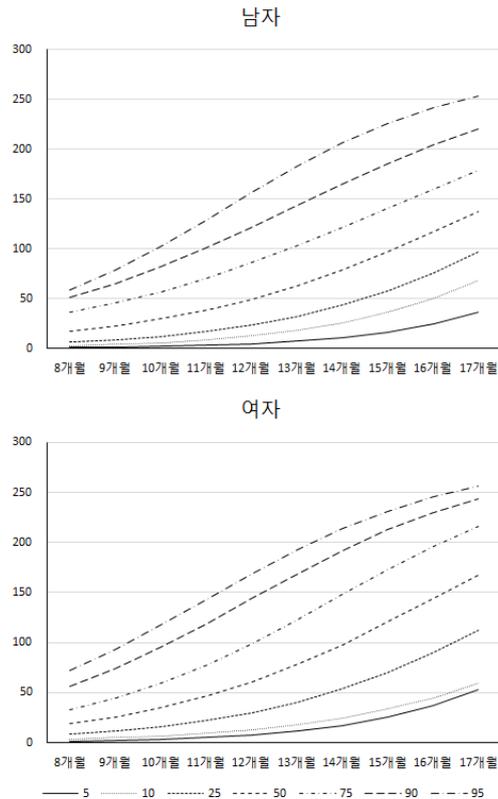


그림 3. 분위수 회귀모형을 이용한 남녀 영아의 분위별 추정 성장 곡선

표 8. 추정 분위수보다 낮은 점수의 빈도와 백분율

성별	추정법		백분위						
			5	10	25	50	75	90	95
남자 (N=278)	2단계 비선형모형	빈도	28	33	78	149	209	248	259
		백분율(%)	10.1	11.9	28.1	53.6	75.2	89.2	93.2
	분위수 회귀모형	빈도	13	31	69	140	208	249	262
		백분율(%)	4.7	11.2	24.8	50.4	74.8	89.6	94.2
여자 (N=285)	2단계 비선형모형	빈도	11	35	85	156	217	253	271
		백분율(%)	3.9	12.3	29.8	54.7	76.1	88.8	95.1
	분위수 회귀모형	빈도	16	30	70	145	214	257	270
		백분율(%)	5.6	10.5	24.6	50.9	75.1	90.2	94.7

표 9. 성별에 따른 각 분위별 두 추정법의 분류 교차표

백분위	2단계 비선형모형	분위수 회귀모형					
		남자			여자		
		높음	낮음	합계	높음	낮음	합계
5	높음	248	1	249	269	5	274
	낮음	16	12	28	0	11	11
	합계	264	13	277	269	16	285
10	높음	243	2	245	250	0	250
	낮음	4	29	33	5	30	35
	합계	247	31	278	255	30	285
25	높음	197	3	200	199	1	200
	낮음	12	66	78	16	69	85
	합계	209	69	278	215	70	285
50	높음	128	1	129	124	5	129
	낮음	10	139	149	16	140	156
	합계	138	140	278	140	145	285
75	높음	69	0	69	66	2	68
	낮음	1	208	209	5	212	217
	합계	70	208	278	71	214	285
90	높음	28	2	30	28	4	32
	낮음	1	247	248	0	253	253
	합계	29	249	278	28	257	285
95	높음	16	3	19	13	1	14
	낮음	0	259	259	1	269	270
	합계	16	262	278	14	270	284

자료 전체를 사용하기 때문에 월령별 표본수의 차이에 크게 영향을 받지 않았기 때문으로 판단된다. 전체 분위수 구간을 통합했을 때, 2단계 절차에 의한 추정 분위수는 남자의 경우 전체 점수의 51.6%를 추정 분위수보다 낮게 분류했으며 여자의 경우 51.5%를 낮게 분류했다. 분위수 회귀모형은 남자의 경우 전체 점수의 50.0%를, 여자의 경우 전체 점수의 50.3%를 추정 분위수보다 낮게 분류했다. 전체 분위수 구간을 통합한 경우 대칭적인 분위수(예, 제25, 제50, 제75 백분위수)가 적절히 추정되었을 때 전체 점수의 50%가 추정 분위수에 비해 낮게 분류될 것으로 기대된다. 이와 같은 결과는 전반적으로 분위수 회귀모형이 2단계 비선형모형에 비해 분류 정확도가 더 높았음을 의미한다.

다음으로 두 추정법이 각 분위수보다 낮은 점수로 분류한 점수들의 일치도를 살펴보았다. 전체 분위수 구간을 통합했을 때는 남자 97.1%, 여자 97.0%의 일치도를 보였다. 남녀별로 각 분위수 구간을 나누어 본 두 추정법의 분류 결과는 표 9에 제시되어 있다. 남자의 경우 제5, 제25, 제50 백분위수에서 불일치가 높았으며 여자의 경우 제25, 제50 백분위수에서 불일치가 높았다.

이상의 결과를 종합할 때 낱말 이해 검사의 분위 로지스틱 성장 곡선의 추정을 위한 본 연구의 분석에서는 월령별 표본 분위수를 이용한 2단계 비선형 회귀모형보다 전 월령의 자료를 동시에 사용한 분위수 회귀모형이 더 바람직한 성장 곡선을 추정했다고 볼 수 있다.

논 의

본 연구에서는 분위 로지스틱 성장 곡선을 추정하기 위해 사용될 수 있는 두 가지 절차에 대해서 살펴보았다. 로지스틱 비선형 회귀모형을 적용한 2단계 절차는 비선형 회귀모형에 대한 이해를 전제할 때 직관적으로 쉽게 적용할 수 있는 장점이 있는 반면, 본 연구의 분석 예시 결과에서 보듯이 추정 분위수 역전 현상이나 추정 분위수를 기준으로 한 분류 비율의 부정확성을 보이는 약점이 있다. 변환점수에 분위수 회귀모형을 적용한 절차는 심리학 분야에서 상대적으로 덜 알려진 분위수 회귀모형을 이용한다는 점과 점수 변환과 재변환 절차가 다소 복잡할 수 있다는 점을 고려해야 하지만, 2단계 절차에서 나타난 문제점을 보이지 않았다는 점에서 분위 성장 곡선의 추정에 효과적인 절차로 판단된다.

분위수 회귀모형은 일반 회귀모형과 마찬가지로 연령 이외에 성별이나 학력 등의 다른 외생 변수를 추가하여 모형을 확장할 수 있다. 예를 들어 본 연구에서 남녀 각각 따로 추정된 월령별 분위수 성장 곡선은 성별 변수를 가변수로 하여 하나의 모형 안에서 동시에 추정될 수 있다. 실제 본 연구에서 사용된 자료에 대한 분석에서도 동일한 성장 곡선을 얻었다.

본 연구에서 예시한 분석의 결과는 2단계 추정 절차에 비해 분위수 회귀모형을 적용한 절차가 성장 곡선 추정에 더 나은 수행을 보인다는 것을 시사한다. 그러나 하나의 분석 사례를 근거로 어느 한 절차가 다른 절차에 비해 우수하다고 결론 내릴 수는 없다. 본 연구에서 확인된 분석 사례에 대한 결과를 일반화하기 위해서는 더 많은 실제 자료에 대한 분석 결과를 비교하거나 수많은 가상 자료에

대한 결과를 비교해 보는 것이 필요하다. 잘 고안된 가상 실험을 통해 두 절차의 수행을 비교하는 연구가 뒤따라야 할 것이다. 여기에는 로지스틱 성장 곡선 이외의 다른 종류의 비선형 성장 곡선에 대한 검토가 포함되는 것도 필요할 것이다.

그 동안 유아기나 노년기의 심리행동적 발달 정도를 평가하기 위해 많은 심리검사들이 개발되어 왔고 또한 다양한 심리검사가 지속적으로 개발되고 있다. 이미 제작된 심리검사의 기준의 개선이나 새로운 심리검사의 기준 개발에 본 연구에서 제안된 절차들이 효과적으로 적용될 수 있을 것이다. 특히 아동발달이나 노화 과정에서 정상에서 벗어난 변화를 선별하는데 사용되는 심리검사들의 기준을 정교화하는 데 분위 성장 곡선 추정 절차를 적용할 수 있다.

많은 연구들이 심리행동적 특성에 대한 성장 곡선의 추정 또는 기준 추정의 과정에서 자료에 대한 단순 요약을 나열하거나 이론적 설명 체계와의 연계 없이 자료에만 의존하는 등 지나치게 단순한 방법을 적용해 왔던 것이 사실이다. 특히 국내에서 수행된 기준 연구나 발달 연구에서 이와 같은 측면이 많이 발견된다(장승민, 강연옥, 2012). 심리검사의 기준을 추정하거나 발달 성장 곡선을 추정할 때에는 모집을 잘 대표하는 충분한 크기의 표본을 확보함과 동시에 누적된 연구 결과와 이론적 이해의 틀을 반영하는 보다 정교한 모형을 효과적으로 적용할 필요가 있다(장승민, 강연옥, 2012; Goldstein & Pan, 1992). 성장 곡선의 추정과 기준 추정에 적용이 가능한 다양한 모형들을 실험적으로 검토하고 새로 제안된 창의적인 방법을 엄밀하면서도 적극적으로 적용하는 연구들이 많아지길 기대한다.

참고문헌

- 배소영, 광금주 (2011). 한국판 맥아더-베이츠의 사소통발달 평가. 마인드프레스.
- 장승민, 강연옥 (2012). 정규분포가 가정된 심리검사의 기준추정을 위한 모형 기반 접근. 한국심리학회지: 일반, 31, 923-944.
- 장승민, 강연옥 (심사중). 분위수 회귀모형을 이용한 심리검사의 기준 추정.
- Burchinal, M., & Appelbaum, M. I. (1991). Estimating individual developmental functions: Methods and their assumptions. *Child Development*, 62, 23-41.
- Cade, B. S., & Noon, B. R. (2003). A gentle introduction to quantile regression for ecologists. *Frontiers in Ecology and the Environment*, 1, 412-420.
- Fenson, L., Dale, P. S., Reznick, J. S., Bates, E., Thal, D. J., & Pethick, S. J. (1994). Variability in early communicative development. *Monograph of Society for Research in Child Development*, 59(5, Series No. 242).
- Fenson, L., Marchman, V. A., Thal, D. J., Bates, E., Dale, P. S., & Reznick, J. S. (2007). *The MacArthur-Bates Communicative Development Inventories: User's Guide and Technical Manual*. Baltimore: Brookes Publishing Co.
- Goldstein, H., & Pan, H. (1992). Percentile smoothing using piecewise polynomials, with covariates. *Biometrics*, 48, 1057-1068.
- Hao, L., & Naiman, D. Q. (2007). *Quantile Regression*. Thousand Oaks, CA: Sage Publications.
- Koenker, R., & Bassett, G. W. (1978). Regression Quantiles. *Econometrica*, 46, 33-50.

- Koenker, R. & Hallock, K. F. (2001). Quantile regression. *Journal of Economic Perspectives*, 15, 143-156.
- Koenker, R., & Park, B. J. (1996). An interior point algorithm for nonlinear quantile regression. *Journal of Econometrics*, 71, 265-283.
- McArdle, J. J., & Epstein, D. (1987). Latent growth curves within developmental structural equation models. *Child Development*, 58, 110-133.
- Yu, K., Lu, Z., & Stander, J. (2003). Quantile regression: Applications and current research areas. *The Statistician*, 52, 331-350.
- 1차원고접수 : 2014. 11. 09.
수정원고접수 : 2014. 12. 09.
최종게재결정 : 2014. 12. 10.

A Comparison of Estimation Procedures for Quantile Logistic Growth Curves

Seungmin Jahng

Dept. of Psychology
Hallym University

Soyeong Pae

Div. of Speech Pathology and Audiology
Hallym University

Identifying an individual's standing on estimated quantile growth curves including lower 10 percentile and upper 10 percentile growth curve, for example, is one way to describe individual differences in developmental process of psychological characteristics. The present study introduced two different approaches for quantile logistic growth curve estimation, i.e., two-step nonlinear regression model and quantile regression model with transformation. An application of the two procedures was illustrated using a real data example of the word understanding test scores obtained from 563 infants. The quantile regression model performed better than the two-step nonlinear regression model in estimation of quantile growth curves. Quantile regression model did not show rank-reversed estimated quantile scores but the two-step nonlinear regression method did. In addition, the former also showed higher accuracy in classification proportion than the latter. It was discussed why we need to use more elaborated statistical models to estimate quantile growth curves.

Key words : growth curve, quantile regression, nonlinear regression, logistic growth function

부록: 예시에서 사용된 SAS 코드

*표본 분위수 생성;

```
PROC UNIVARIATE DATA=RAW_DATA NOPRINT;  
  CLASS GENDER MONTH;  
  VAR DV;  
  OUTPUT OUT=QUANTILE  
    P5=P05_DV  
    P10=P10_DV  
    Q1=P25_DV  
    MEDIAN=P50_DV  
    Q3=P75_DV  
    P90=P90_DV  
    P95=P95_DV;  
RUN;
```

*표본 분위수 자료 형태 변경 및 정렬;

```
DATA QUANTILE_LONG;  
  SET QUANTILE;  
  PTILE_DV=P05_DV;PTILE=05;OUTPUT;  
  PTILE_DV=P10_DV;PTILE=10;OUTPUT;  
  PTILE_DV=P25_DV;PTILE=25;OUTPUT;  
  PTILE_DV=P50_DV;PTILE=50;OUTPUT;  
  PTILE_DV=P75_DV;PTILE=75;OUTPUT;  
  PTILE_DV=P90_DV;PTILE=90;OUTPUT;  
  PTILE_DV=P95_DV;PTILE=95;OUTPUT;  
  DROP P05_DV P10_DV P25_DV P50_DV P75_DV P90_DV P95_DV;  
RUN;  
PROC SORT DATA= QUANTILE_LONG;  
  BY GENDER PTILE MONTH;  
RUN;
```

*표본 분위수에 대한 비선형 회귀모형: 2단계 접근;

```
PROC NLIN DATA=QUANTILE_LONG NOITPRINT;  
  BY GENDER PTILE;  
  PARS ALPHA=4 BETA=-.5;*임의 초기값;  
  MODEL PTILE_DV=284/(1+EXP(ALPHA+BETA*MONTH));  
  OUTPUT OUT=OUT_NLIN P=PRED;  
RUN;
```

한국심리학회지: 일반

*분위수 회귀모형을 위한 자료 변환 및 정렬;

```
DATA RAW_DATA2;  
  SET RAW_DATA;  
  T_DV=LOG((284*1.0001-DV)/(DV+.0001));  
RUN;  
PROC SORT DATA=RAW_DATA2;  
  BY GENDER;  
RUN;
```

*변환 자료에 대한 분위수 회귀모형;

```
PROC QUANTREG DATA=RAW_DATA2;  
  BY GENDER;  
  MODEL T_DV=MONTH/QUANTILE=.05 .1 .25 .5 .75 .9 .95;  
  OUTPUT OUT=OUT_QUANTREG PRED=PRED_T/COLUMNWISE;  
RUN;
```

*예측값을 원 자료 단위로 역변환;

```
DATA OUT_QUANTREG;  
  SET OUT_QUANTREG;  
  PRED_DV=(284*1.0001-.0001*EXP(PRED_T))/(1+EXP(PRED_T));  
RUN;
```