

분위수 회귀모형을 이용한 심리검사의 기준 추정*

장 승 민[†]

강 연 옥

한림대학교 심리학과

검사점수의 정규성이 가정될 수 없는 심리검사의 기준은 평균과 표준편차보다는 특정 분위수로 추정된다. 또한 검사점수의 기준은 검사가 측정하는 속성이 관련 변인과 일정한 함수적 관계를 갖는다고 가정될 수 있는 경우 통계적 모형을 이용하여 추정될 수 있다. 본 연구는 검사점수의 정규성이 가정되지 않지만 검사가 측정하는 속성이 관련 변인과 함수적 관계를 갖는다고 가정될 수 있는 경우, 분위수 회귀모형을 이용해 검사점수의 분위수 기준을 추정하는 절차를 소개하였다. 분위수 회귀모형에 대한 기본적 개념을 설명하고 이것이 검사 기준을 추정하는데 어떻게 적용될 수 있는지를 예시하였다. 정상 장/노년 1060명을 대상으로 실시된 12점 만점의 계산능력 검사 점수에 대해 연령과 교육 수준에 따라 구분된 하위 집단의 제5, 제10, 제25 백분위수를 분위수 회귀모형을 이용하여 추정하였다. 모형에 의해 추정된 분위수들은 각 하위 집단별로 계산된 표본 분위수와는 달리 연령과 교육수준에 따른 계산능력의 차이에 대한 이론적 모형과 부합하는 기준을 제공하였다.

주요어 : 분위수 회귀모형, 심리검사 기준, 모형 기반 표준화, 정규성 가정

* 이 논문은 한림대학교 특성화사업비(HRF-S-21)에 의하여 연구되었음.

[†] 교신저자(Corresponding Author) : 장승민 / 한림대학교 심리학과 / (200-702) 강원도 춘천시 한림대학길 39

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

심리검사의 점수는 검사가 측정하는 속성에서 사람들이 갖는 상대적 위치를 나타낼 때 분명한 의미를 갖는다(Furr & Bacharach, 2014). 검사점수의 기준(norm)은 개인이 속한 집단 전체의 점수에서 개인의 검사점수가 갖는 상대적 위치를 가늠할 수 있는 참조 체계를 제공한다. 이 때 검사점수 기준이 제공하는 핵심 정보는 기준 집단에서 개별 검사점수보다 높은 점수를 얻은 사람들과 낮은 점수를 얻은 사람들의 비율, 즉 개별 검사점수의 분위 정보이다. 많은 심리검사들은 검사점수의 분위 정보를 나타내기 위해 원 점수 대신 이를 변환한 변환 점수를 사용한다.

검사 기준에서 분위 정보를 얻기 위해 가장 널리 사용되는 변환 점수로는 표준점수와 백분위 점수가 있다. 표준점수는 검사점수의 모집이 정규분포를 따른다고 가정할 수 있는 경우 분위 점수로 해석될 수 있다. 예를 들어 검사점수가 정규분포를 따른다고 할 때 특정 표준점수를 얻은 사람보다 점수가 낮은 사람들의 비율은 표준점수가 -1일 때 약 15.8%, -1.5일 때 약 6.7%, -2일 때 약 2.3%가 된다. 따라서 이 경우 검사 기준의 분위 추정치는 표준점수를 산출하는데 필요한 모집의 평균과 표준편차를 추정함으로써 가능해진다. 반면 검사점수의 정규성을 가정하기 어려운 경우에는 검사점수의 해석에 필요한 분위수(quantile), 예컨대 제5 백분위수나 제15 백분위수 등을 직접 추정하여 검사점수와 비교할 수 있다. 즉, 추정된 제5 백분위수나 제15 백분위수 등을 절단점으로 하여 개별 검사점수가 어느 구간(예, 제5 백분위수 이하)에 속하는지를 판단하여 검사점수를 해석할 수 있다. 따라서 평균과 표준편차를 기준으로 제공하는 심리검사들은 검사점수의 분포가 정규분포를 따른다고

가정할 수 있는지에 대한 근거를 제공하는 것이 필요하며, 이 가정이 적절하지 않다고 판단될 경우 분위수에 대한 직접 추정 등의 대안적인 기준을 제공하는 것을 고려해야 한다.

검사 기준의 개발에서 고려되어야 하는 또 다른 중요한 요소는 기준 산출에 사용되는 참조 표본(reference sample) 또는 기준 표본(normative sample)의 크기이다. 이는 검사점수의 기준이 표본 자료로부터 추정되는 모집 분포의 특성에 대한 기술이기 때문이다. 평균의 경우 신뢰할 수 있는 추정을 위해서는 각 기준 집단별로 적어도 50 이상의, 모집을 잘 대표하는 표본이 필요하며 제10 백분위수, 제5 백분위수와 같이 평균에서 멀리 떨어진 값의 추정을 위해서는 더 큰 크기의 표본이 필요하다(Bridges & Holler, 2007; Crawford & Garthwaite, 2008; Garsd, Ford, Waring, & Rosenblatt, 1983; Guo, et al., 2000). 참조 표본이 여러 하위 집단들, 예컨대 연령이나 교육 수준, 성별 등에 의해 구분되는 집단들로 구성되고 이 하위 집단들이 서로 다른 분포를 갖는다고 가정되면, 각 하위 집단들은 저마다의 기준이 필요하다. 따라서 각 하위 집단의 기준을 집단마다 독립적으로 추정하는 경우, 하위 집단이 많아질수록 기준 추정에 필요한 표본의 크기도 커지게 된다(장승민, 강연옥, 2012).

그러나 많은 심리검사의 기준 개발 연구들은 현실적인 제약 등의 이유로 신뢰할 수 있는 기준의 추정에 필요한 충분한 크기의 표본을 사용하지 못하고 있다. 이러한 연구들에서는 추정된 평균 또는 분위수의 집단 간 차이가 해석하기 어려운 불규칙한 형태를 보이는 경우가 흔히 발생한다.¹⁾ 이미 국내에서 수행

1) 집단간 차이의 불규칙한 형태는 표본의 크기와

된 많은 심리검사 기준 연구들이 이와 같이 불규칙한 형태의 평균과 표준편차 기준들을 제공하고 있음이 보고된 바 있다(장승민, 강연옥, 2012).

분위수를 기준으로 제공하는 기준 연구에서도 유사한 경우를 쉽게 발견할 수 있다. 치매 등을 선별하기 위해 국내에서 널리 사용되어 온 신경심리검사 배터리 중 하나인 한국판 CERAD 평가집(CERAD-K, 우종인 등, 2003)도 그 중 하나이다. CERAD-K에 포함되어 있는 검사 중 언어유창성 검사는 언어 산출 장애 또는 언어 능력을 평가하기 위해 사용된다. 이 검사는 피검자에게 ‘동물’에 속하는 단어를 1분 동안 말하게 하여 그 개수를 검사 점수로 간주한다. 기준표(우종인 등, 2003, pp.25-28, 표 15-표 18)에 보고된 이 검사 점수의 제5 백분위수를 살펴보면 교육연한이 7년 이상인 남성의 경우 가장 연령이 높은 집단인 75세에서 90세 집단의 점수가 가장 높게 나타났다(그림 1). 반면 같은 교육연한을 가진 여성 집단에서는 연령이 높을수록 낮은 점수를 보였다. 이 결과를 7년 이상의 교육을 받은 전체 남성들의 언어 능력 하위 5%의 수준이 더 젊은 연령 집단에 비해 75세 이상 노인들에서 더 높고, 여성들의 경우에는 그렇지 않다고 해석할 수도 있다. 그러나 이러한 해석은 성별에 관계없이 연령이 증가함에 따라 언어 능력을 포함한 인지적 능력이 점차 감소한다는 기존의 연구 결과들(예, Glisky, 2007)과 배치된다. 오히려 이러한 결과는 해당 집단의 표본 크기가 작아 표집오차가 커서 나타난 것으로 보는 것이 더 적절하다. 실제로 CERAD-K는 기준 산출에 사용되는 표본 크기를 크게 하기 위해

더불어 표본이 모집을 잘 대표하도록 추출되었느냐와도 관련된다.

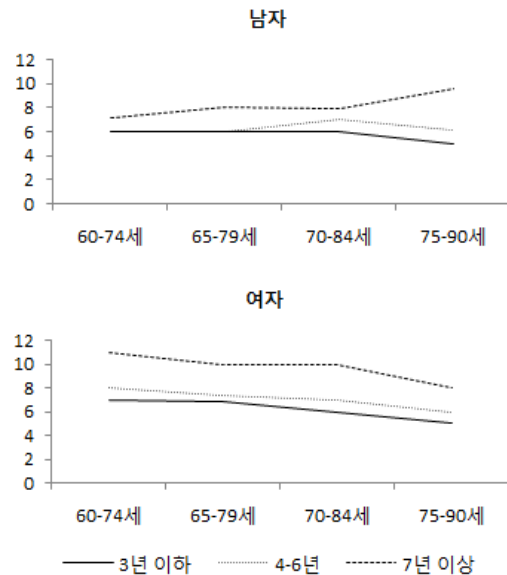


그림 1. 교육연한과 연령에 따른 남녀별 언어유창성 검사의 제5 백분위수(자료출처: 우종인 등, 2003, 표 4)

중복연령기준법을 사용하였음에도 불구하고, 교육연한 7년 이상인 75세 이상 남성의 표본 크기($n=31$)가 같은 교육 수준의 다른 연령 집단들의 표본 수($n=56-81$)에 비해 낮은 수준이었다.

또 다른 신경심리검사 배터리인 서울신경심리검사 초판(SNSB, 강연옥, 나덕렬, 2003)에서도 이와 같은 사례를 발견할 수 있다. SNSB의 경우 일부 검사에 대해 하위 16%에 해당하는 절단점수를 기준으로 제시하고 있다. 이 검사 배터리에 포함된 한국형 간이 정신상태검사(K-MMSE)의 장소지남력 점수의 기준표(p. 64, 표 A-2)를 보면 60세에서 64세에 해당하는 문맹자($n=4$)의 절단점수(3)가 65세에서 69세 사이의 문맹집단($n=13$)과 70세에서 74세 사이의 문맹집단($n=12$)의 절단점수(4)보다 오히려 낮게 나타났다. 또 다른 예로 서은현 등(2007)이

보고한 벤톤 시각기억검사 실시 A의 정반응 점수에 대한 기준(p. 754, 표 4)에서는 교육연한이 3년 이하인 남성 집단의 경우 최고령 집단인 75-90세 집단에서 제95 백분위수가 가장 높게 나타난 반면 같은 학력의 여성 집단의 경우는 연령이 높을수록 제95 백분위수가 점차 낮아지는 형태를 보였다. 이 연구에서 중복연령기준법을 사용한 각 연령 집단별 표본의 크기는 교육연한 3년 이하 남성 집단의 경우 13(60세에서 74세), 17(65세에서 79세), 15(70세에서 84세), 11(75세에서 90세)이었다. 반면 교육연한 3년 이하 여성의 해당 연령 집단별 표본 크기는 각각 80, 92, 77, 43이었다.

인지능력은 장년기 이후 연령이 증가할수록 감소하고(Glisky, 2007; Salthouse, Atkinson, & Berish, 2003) 학력이 높을수록 높다고(Barone, & van de Werfhorst, 2011; van Hooren et al., 2007) 알려져 있다. 상기한 예에서와 같이 인지능력을 측정하는 심리검사의 표본 점수가 기준에 알려진 체계적 차이에서 벗어난 형태를 보인다면 이는 표집오차에 의해 영향을 받은 것으로 이해할 수 있다. 이러한 경우는, 예를 들어 검사점수의 평균 혹은 특정 분위수가 연령이 높을수록 단조 감소하고 학력이 높을수록 단조 증가한다는, 이론에 부합하는 일정한 체계(모형)를 상정할 수 있으며 연구자는 그 체계의 규칙성을 반영하는 함수를 이용하여 기준을 모형에 맞게 추정할 수 있다.

예를 들어 기준 표본에서 각 연령별로 얻은 지능검사 점수의 평균이 들쭉날쭉한 형태를 보인다면, 이 평균들에 대해 연령을 예측변수로 하는 회귀분석을 실시한 후 여기에서 추정된 예측값을 체계적인 차이를 갖는 기준으로 사용할 수 있다(Taylor, 1998; Zachary &

Gorsuch, 1985). 또는 월령별로 계산된 유아의 단어 이해 검사 점수의 제5 백분위수들이 단조 증가하지 않는 형태를 보인다면 이 분위수들의 예측값을 월령을 예측변수로 하여 로지스틱 함수를 이용한 비선형 회귀분석으로 추정할 수 있다(배소영, 곽금주, 2011; Fenson, et al., 1994). 이와 같이 기준 점수의 차이가 하위 집단에 따라 일정한 규칙성을 갖도록 하는 절차를 연속규준화(continuous norming)라고 부른다.

앞서 설명한 연속규준화 절차는 첫 단계에서 하위 집단별로 평균 또는 분위수를 얻은 후 다음 단계에서 이 통계량들에 대한 회귀선을 추정하는 2단계 추정 절차를 사용한다. 그러나 2단계 추정 절차는 하위 집단별 표본 수의 차이가 고려되지 않고, 하위 집단을 구분하는 변인(예, 연령)과의 최적 함수 형태를 결정할 때 참고해야 하는 각 예측변수의 통계적 유의성과 모형 적합도를 적절히 평가할 수 없다는 단점이 있다(Jahng, Wood, & Trull, 2008).²⁾

기준 점수가 체계적 형태를 갖도록 추정하기 위해 2단계 절차를 사용하지 않고 표본 자료 전체를 이용할 수도 있다. 장승민과 강연옥(2012)은 검사점수가 정규분포를 따른다고 가정할 수 있을 때 전체 표본 자료에 대해 회귀모형을 적용하여 하위 집단별 평균과 분산

2) 2단계 접근은 1단계에서 하위집단별로 얻어진 통계량(추정치)을 관찰값으로 하여 2단계에서 모형을 추정한다. 이 때 관찰값으로 사용되는 통계량은 이미 추정오차를 갖고 있는데 2단계 추정법은 이를 무시함으로써 최종 모형의 추정에서 표준오차를 과소평가하게 되고 결과적으로 예측변수의 통계적 유의성과 모형의 적합성을 과대평가하게 된다. 이와 관련된 수리적 예시는 Jahng 등(2008)의 각주 9(p. 363)를 참고할 수 있다.

(표준편차)의 체계적 차이를 추정하는 절차를 소개하였다. 그러나 검사점수의 정규성을 가정하기 어려운 경우에는 평균과 표준편차를 이용하여 검사점수의 상대적 위치를 추정하는 것이 적절치 않기 때문에 검사점수 해석에 참고할 수 있는 분위수를 직접 모형화하는 것이 필요하다. 이와 같이 분위수를 체계적 차이를 갖는 함수로 모형화하기 위한 절차로 분위수 회귀모형이 사용될 수 있다. 본 연구는 검사점수의 정규성이 가정되지 않지만 검사가 측정하는 속성이 관련 변인과 함수적 관계를 갖는다고 가정될 수 있는 경우, 분위수 회귀모형을 이용해 검사점수의 분위수 기준을 추정하는 절차를 소개하였다. 다음 절에서는 분위수 회귀모형의 기본 개념과 이를 이용하여 검사점수의 기준을 추정하는 절차를 소개하였다. 이어서 실제 심리검사 기준 자료를 이용하여 분위수 회귀모형을 적용한 기준 추정의 과정을 예시하였다.

분위수 회귀모형을 이용한 모형 기반 기준 추정 절차

평균과 중위수(median, 중앙값)를 여러 값들을 대표하는 단일 수치, 즉 중심경향성(central tendency)으로서 이해하는 한 가지 방식은 이 수치들이 개별 값과 갖는 거리의 전체 합을 가장 작게 만드는 특정 값이라는 것이다. 전체 자료를 대표하는 대푯값이 전체 값들과 다른 정도(거리의 합)를 손실이라고 표현할 때 대푯값은 손실을 최소화해야 한다. 평균은 관찰값과 갖는 거리를 편차의 제곱($[y_i - \mu]^2$)으로 정의할 때, 중위수는 그 거리를 편차의 절댓값($|y_i - m|$)으로 정의할 때 각각 거리의 합

을 가장 작게 만든다. 이를 식으로 표현하면 임의의 수 c 에 대하여 c 가 평균일 때 손실함수 $\sum (y_i - c)^2$ 이 가장 작아지고 c 가 중위수일 때 손실함수 $\sum |y_i - c|$ 이 가장 작아진다. 다시 말해 평균은 편차의 제곱합을 최소로 만드는 값으로, 중위수는 편차의 절댓값합을 최소로 만드는 값으로 정의할 수 있다.

선형 회귀모형은 관찰값의 평균을 예측변수의 선형함수(예, $\hat{y}_i = \beta_0 + \beta_1 x_i$)로 추정하는데 이 선형함수는 평균에서와 마찬가지로 관찰값과의 편차제곱의 합을 가장 작게 만든다(최소제곱합). 예를 들어 예측변수가 하나인 단순 회귀모형이 최소화하는 손실함수는 다음과 같다.

$$\sum_i (y_i - (\beta_0 + \beta_1 x_i))^2 \quad \text{식 (1)}$$

반면 중위수 회귀모형(median regression model)은 관찰값의 중위수를 예측변수의 선형함수로 추정하며 이 선형함수는 중위수에서와 마찬가지로 관찰값과 갖는 편차의 절댓값합을 최소로 만든다(최소절댓값합). 예를 들어 예측변수가 하나인 중위수 회귀모형은 아래의 손실함수를 가장 작게 만든다.

$$\sum_i |y_i - (\beta_0^* + \beta_1^* x_i)| \quad \text{식 (2)}$$

분위수 회귀모형은 중위수 회귀모형을 Koenker와 Bassett(1978)가 확장하여 제안하였다. 중위수 회귀모형이 관찰값의 중위수를 예측변수의 선형함수로 추정한다면 분위수 회귀모형은 0에서 1사이의 임의의 p 에 대해 관찰값의 p 분위수를 예측변수의 선형함수로 추정한다. 임의의 분위 p 에 대해 예측변수가 하나인

분위수 회귀모형이 최소화하는 손실함수는 다음과 같다(Hao & Naiman, 2007; Koenker & Hallock, 2001).

$$\sum_{y_i \geq \beta_0^{(p)} + \beta_1^{(p)} x_i} p |y_i - (\beta_0^{(p)} + \beta_1^{(p)} x_i)| + \sum_{y_i < \beta_0^{(p)} + \beta_1^{(p)} x_i} (1-p) |y_i - (\beta_0^{(p)} + \beta_1^{(p)} x_i)| \quad \text{식 (3)}$$

분위 $p = .5$ 에 대하여 식 (3)을 최소화 하는 $\beta_0^{(.5)}, \beta_1^{(.5)}$ 과 식 (2)를 최소화하는 β_0^*, β_1^* 는 동일하며 따라서 중위수 회귀모형은 분위수 회귀모형의 특수한 형태로 간주할 수 있다.

분위수 회귀모형은 오차의 분포에 대한 가정이 필요 없고 이상값(outlier)에 영향을 받지 않는다. 따라서 분포의 정규성이 가정되기 어려운 검사점수의 규준을 추정할 때 일반적인 선형 회귀모형에 비해 더 적절한 추정값을 제공할 수 있다. 또한 분위수 회귀모형은 단조 변환에 대해 등변성(equivariance)을 갖는다. 즉, 단조 변환한 자료를 이용한 모형의 추정값과 원자료를 이용한 모형의 추정값이 자료와 동일한 변환을 따른다. 따라서 검사점수와 예측변수가 단조 비선형적인 관계(예, 로그선형, 로지스틱)를 갖는 경우 검사점수의 적절한 단조변환을 통해 선형 모형으로 추정이 가능하다(Hao & Naiman, 2007). 또는 예측변수를 적절히 변환하여 선형 분위수 회귀모형을 적용하여도 동일한 결과를 얻을 수 있다. 예측변수가 여럿이고 이들이 검사점수와 갖는 관계가 서로 다른 함수의 형태를 갖는 경우라면 검사점수를 변환하는 것보다는 예측변수를 변환하는 것이 더 적절할 수 있다.

분위수 회귀모형의 모형 적합도는 산출될 손실함수(또는 목적함수)의 값이 작을수록 좋게 평가된다. 따라서 예측변수를 변환한 모형

이 변환하지 않은 모형보다 더 작은 손실함수 값을 산출한다면 예측변수를 변환하는 것이 더 적절하다고 결론내릴 수 있다. 또한 절편만을 포함하는 모형의 손실함수 값($Q[p1]$)과 예측변수들을 포함한 모형의 손실함수 값($Q[p2]$)을 이용해 선형 회귀모형의 R^2 과 유사한 모형 적합도 지수(pseudo $R^2[p] = 1 - Q[p2]/Q[p1]$)를 구하거나, 서로 다른 두 위계적 모형의 손실함수 값들을 이용해 모형을 비교하고 예측변수들의 통계적 유의성을 판단할 수도 있다(Koenker & Machado, 1999).

정리하자면 검사점수의 정규성을 가정하기 어려워 평균과 표준편차보다는 분위수를 기준으로 제공하는 것이 적절하고 동시에 규준 집단을 구분하는 예측변수들과 검사점수가 체계적인 단조함수로 표현될 수 있는 관계를 갖는다고 가정할 수 있는 경우, 분위수 회귀모형을 이용하여 연속적인 형태를 갖는, 모형에 기반한 규준을 추정할 수 있다. 다음 절에서는 분위수 회귀모형을 이용한 모형 기반 규준 추정 절차를 실제 자료를 이용하여 구체적으로 예시하였다.

분위수 회귀모형을 이용한 규준 추정의 예시: 계산능력 검사

표본 및 도구

예시에 사용된 규준 자료는 서울신경심리검사(SNSB: 강연욱, 나덕렬, 2003)의 개정판 제작을 위해 45세에서 90세 사이의 정상 장노년 1067명(여자: 600명, 남자: 467명)을 대상으로 2009년부터 2011년 사이에 수집되었다. 표집은 자료의 대표성을 확보하기 위해 전국의 대도

표 1. 연령과 교육 수준에 따른 기준 표본의 분포

연령	교육 수준						합계
	문맹	0-3년	4-6년	7-9년	10-12년	13년 이상	
45-49세			12	12	25	22	71
50-54세			23	24	34	27	108
55-59세		9	27	21	27	23	107
60-64세	7	12	31	28	24	25	127
65-69세	13	21	29	23	27	21	134
70-74세	13	27	32	21	20	22	135
75-79세	23	23	34	23	33	20	156
80-84세	20	20	26	20	32	15	133
85-90세	13	17	21	18	10	10	89
합계	89	129	235	190	232	185	1060

시, 중소도시 및 농어촌을 망라하여 이루어졌으며 피검자 선정 과정에서 건강 선별 기준(Christensen, Multhaup, Nordstrom, & Voss, 1991)에 근거하여 신경과 또는 정신과적인 병력이나 두뇌손상이 의심되는 사람들이 제외되었다.³⁾ 전체 기준 표본 중 계산능력 검사를 실시한 1060명에 대해 분석이 수행되었다.

계산능력 검사는 덧셈, 뺄셈, 곱셈과 나눗셈 각각 3문제씩 모두 12문제로 구성되어 있다. 각 문제 당 1점씩이 배점되어 있으며 획득 가능한 점수의 범위는 0-12점이다. 심리검사를 통해 측정되는 계산능력은 연령이 높을수록 낮고 교육 수준이 높을수록 높은 것으로 알려져 있다(강연옥, 나덕렬, 2003; 황순택, 김지혜, 박광배, 최진영, 홍상환, 2011). 계산능력 검사는 치매 등의 인지적 장애를 선별하기 위한 목적으로 제작되어 있어 일반적인 성인이 쉽

게 만점을 얻을 수 있으나 연령이 높을수록, 교육 수준이 낮을수록 만점자의 비율이 낮게 나타난다. 따라서 본 분석에서는 연령대와 교육 수준에 대해 기준 집단을 구분하여 기준을 추정하였다. 연령 집단은 45세에서 90세까지의 연령을 다섯 살 단위로 모두 아홉 개의 범주로 구분하였다(45~49세, 50~54세, 55~59세, 60~64세, 65~69세, 70~74세, 75~79세, 80~84세, 85~90세).⁴⁾ 교육 수준은 교육 연수에 따라 “문맹”, “무학이나 문맹이 아님~3년”, “4~6년”, “7~9년” 및 “10~12년”, “13년 이상”의 총 6개의 범주로 구분하였다. 나이와 교육 수준에 따른 기준 표본의 분포가 표 1에 제시되어 있다.

예비분석

먼저 각 연령 집단과 학력 집단에 따른 검

3) 이 자료의 표본 특성과 자료 수집 절차 등에 대한 자세한 내용은 서울신경심리검사 2판(강연옥, 장승민, 나덕렬, 2012)을 참고.

4) 마지막 연령집단은 85~90세로 연령 범위가 6년이었다.

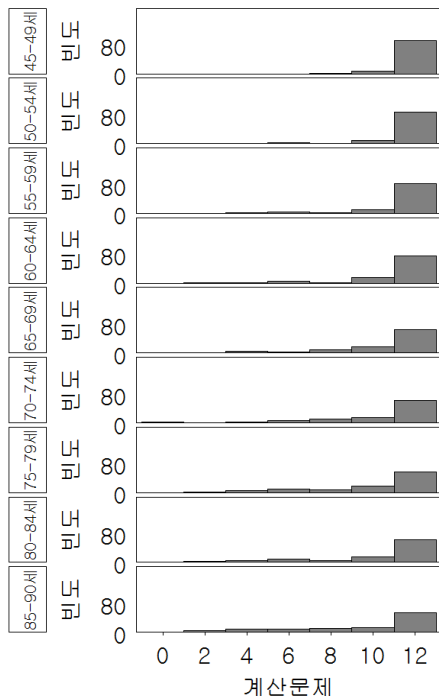


그림 2. 연령에 따른 각 하위 집단의 계산문제 점수의 히스토그램

사점수의 분포를 확인하였다. 표 1에 나타난 바와 같이 총 54개의 범주(하위 집단) 중 49개의 범주가 하나 이상의 표본을 갖고 있었다. 계산점수의 연령 집단별 히스토그램과 학력 집단별 히스토그램이 각각 그림 2와 그림 3에 제시되어 있다. 그림에서 보듯이 각 하위 집단의 검사점수의 분포는 ‘문맹’ 집단을 제외하고 부적 왜도가 높은 분포를 보이고 있으며 정규분포를 따른다고 보기 어렵다. 정규성에 대한 검정(Kolmogorov-Smirnov test) 결과도 모든 집단에서 유의수준 .05에서 정규성 가설이 기각되었다. 연령과 학력을 동시에 고려한 49개 하위 집단에 대한 정규성 검정 결과도 하나 이상의 표본을 갖는 대부분의 집단에서 정규성이 기각되었다.⁵⁾ 따라서 계산능력 검사점수

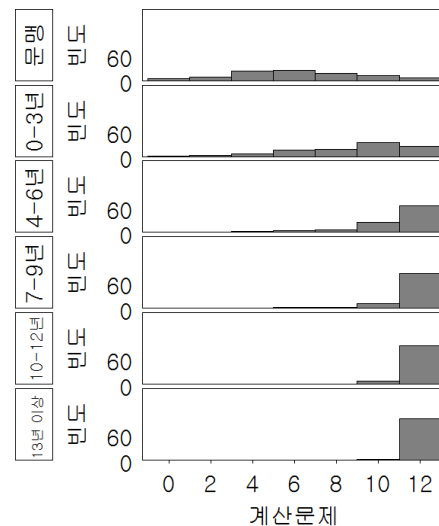


그림 3. 교육 수준에 따른 각 하위 집단의 계산문제 점수의 히스토그램

는 평균과 표준편차를 통해 상대적 위치를 추정하는 것이 적절치 않으며 기준을 제공하기 위해 분위수를 직접 추정하는 것이 더 바람직하다고 할 수 있다.

다음으로 연령과 교육 수준이 검사점수의 분포와 어떤 체계적인 관련을 갖는지를 살펴 보기 위하여 연령 집단과 학력 집단 각각에 따라 제5, 제10 및 제25 백분위수를 확인하였다. 그림 4는 연령 집단에 따른 계산능력 검사점수의 제5, 제10 및 제25 백분위수를 나타낸 것이다. 각 분위수는 연령이 높을수록 점진적으로 낮아지며 동시에 제5 백분위수와 제25 백분위수의 차이가 커지는 것을 확인할 수 있다. 검사점수의 각 분위수는 연령대가 높아질수록 일정하게 단조 감소하는 것으로 보인

5) 정규성이 기각되지 않은 집단은 문맹이면서 75-79세인 집단을 제외한 나머지 문맹 집단과 교육연한이 0-3년인 55-59세, 60-64세, 85-90세 집단이었으며 모두 표본 수가 20이하였다.

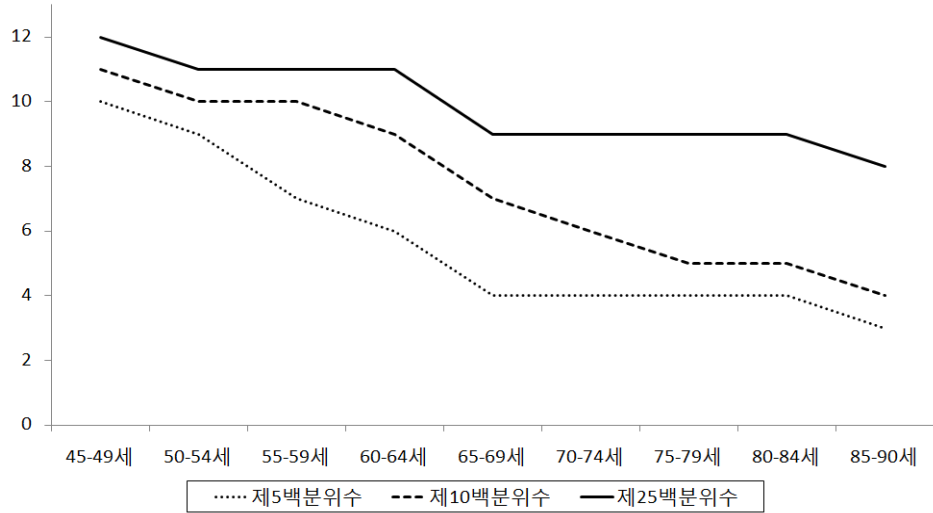


그림 4. 계산능력 점수의 연령 집단별 제5, 제10, 제25 백분위수

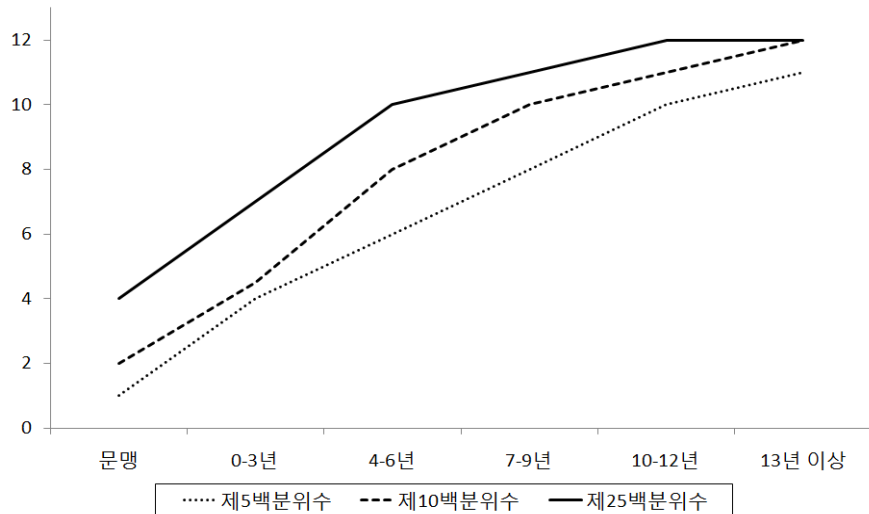


그림 5. 계산능력 점수의 학력 집단별 제5, 제10, 제25 백분위수

다. 그림 5는 학력 집단에 따른 검사점수의 제5, 제10 및 제25 백분위수를 나타낸 것이다. 각 분위수는 학력이 높을수록 단조 증가하며 큰 값을 나타냈다. 또한 상대적으로 학력이 낮은 집단들 사이에서의 차이가 학력이 높은 집단들 사이에서의 차이보다 크게 나타났다.

분석 및 결과

연령과 학력에 따른 각 하위 집단의 제5, 제10, 제25 백분위수를 추정하기 위하여 분위수 회귀모형을 사용하였다. 모형에 포함될 예측변수로는 연령(x_1)과 교육 수준(x_2) 및 이 둘

표 2. 계산능력 점수의 분위수 회귀모형의 모수 추정값

변인	추정값	95% 신뢰구간		유사 R ²
		하한값	상한값	
제5 백분위수				
절편	-0.52	-2.76	2.35	.45
문맹	2.16	0.16	5.07	
log(교육 수준)	6.69	5.64	8.48	
연령 ²	-0.01	-0.03	-0.00	
제10 백분위수				
절편	2.81	1.80	3.69	.48
문맹	1.92	0.12	2.96	
log(교육 수준)	5.13	4.05	5.64	
연령 ²	-0.05	-0.07	-0.02	
log(교육 수준)*연령 ²	0.03	0.01	0.04	
제25 백분위수				
절편	8.57	7.79	10.28	.43
log(교육 수준)	1.91	1.54	2.27	
연령	-0.62	-0.92	-0.40	
log(교육 수준)*연령	0.35	0.24	0.50	

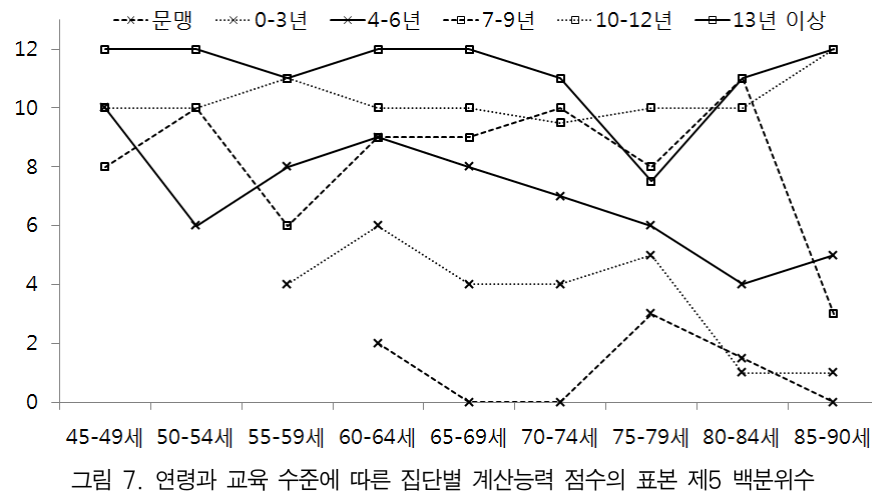
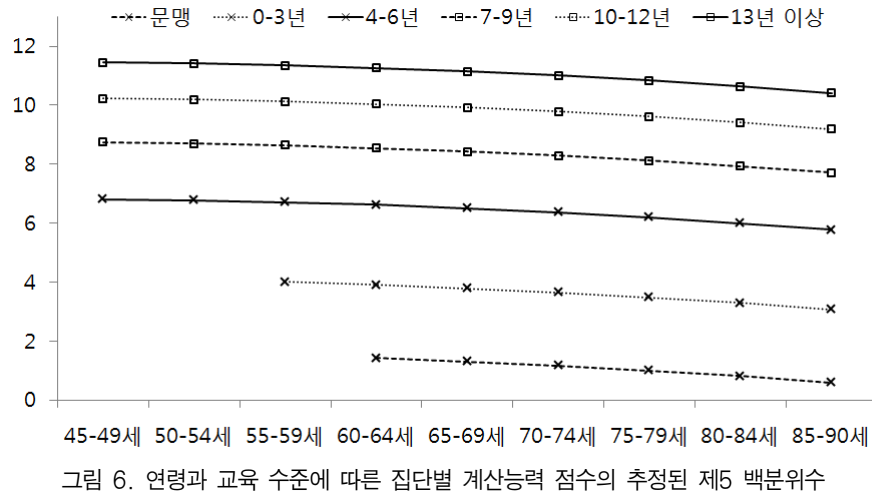
의 상호작용 항이 검토되었다. 연령 변인 x_1 에는 연령 집단에 따라 “45-49세” 집단부터 “85-90세” 집단까지 1에서 9까지의 숫자를 부여하였다.⁶⁾ 학력 변인 x_2 에는 학력 집단에 따라 “문맹” 집단부터 “13년 이상” 집단까지 1에서 6까지의 숫자를 부여하였다. 문맹집단의

6) 1단계 연속규준화를 사용할 때 연령과 같은 변인은 연령 집단을 구분하지 않고 연령을 직접 연속 변수로 사용할 수도 있으며 이렇게 할 때 1단계 절차의 이점을 더 잘 살릴 수 있다. 본 연구에서는 규준 집단의 구성과 하위 집단별로 규준을 따로 구하는 전통적일 접근과 비교하기 쉽도록 연령을 집단으로 변환한 모형을 제시하였다.

경우 다른 학력 집단들 간의 차이와 구분되는 추가적인 차이가 있을 가능성을 검토하기 위해 문맹집단 여부를 나타내는 별도의 가변수 (x_3 : 문맹=1, 문해=0)를 모형에 포함하였다. 이상의 내용을 반영하여 분위 $p = .05, .10, .25$ 에 대하여 다음과 같은 분위수 회귀모형을 구성하였다.

$$y_i = \beta_0 + \beta_1^{(p)}x_{1i}^* + \beta_2^{(p)}x_{2i}^* + \beta_3^{(p)}x_{1i}^*x_{2i}^* + \beta_4^{(p)}x_{3i}^* + \epsilon_i^{(p)}$$

자료에 가장 잘 부합하는 x_1^* 와 x_2^* 의 형태를 결정하기 위해 x_1^* 의 자리에는 x_1 과 x_1^2 을,



x_2^* 의 자리에는 x_2 와 $\log_e(x_2)$ 를 번갈아가며 넣어 모형을 분석하였다. 모형의 선택은 각 분위수별로 독립적으로 실시되었다. 분석에는 SAS 프로그램의 QUANTREG 절차가 사용되었다. 모형에 포함된 상호작용 항의 계수가 통계적으로 유의미하지 않은 경우에는 모형에서 누락하였으며 상호작용 항의 계수가 유의미한 경우에는 상호작용 항을 구성하는 각 1차 항, 즉 x_1^* , x_2^* 의 계수가 유의미하지 않은 경우도

모두 모형에 포함시켜 최종 모형을 구성하였다.⁷⁾ 각 분위수별 최종 모형과 분석 결과가 표 2에 제시되어 있다.⁸⁾

7) 예를 들어 제5 백분위수 모형을 위해 사용된 SAS 코드는 다음과 같았다.

```
PROC QUANTREG DATA=DATANAME;
MODEL CALC = ILLIT LN_EDU AGE_2 /QUANTILE
= 0.05;
RUN;
```

8) 모형의 모수는 심플렉스(simplex) 알고리즘에 의

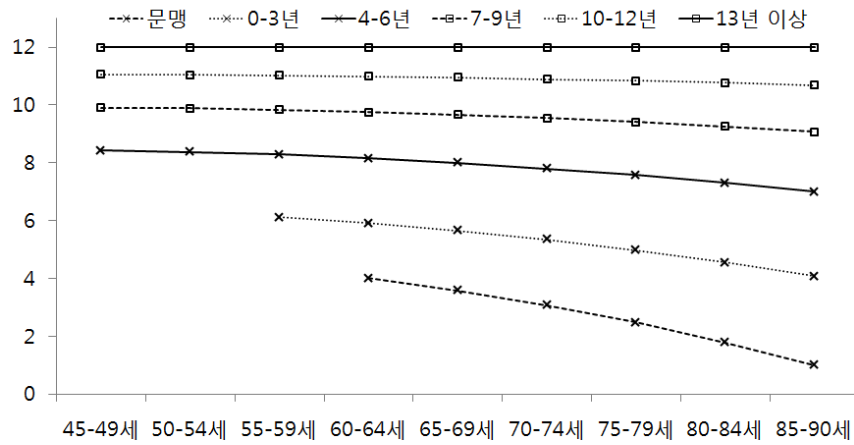


그림 8. 연령과 교육 수준에 따른 집단별 계산능력 점수의 추정된 제10 백분위수

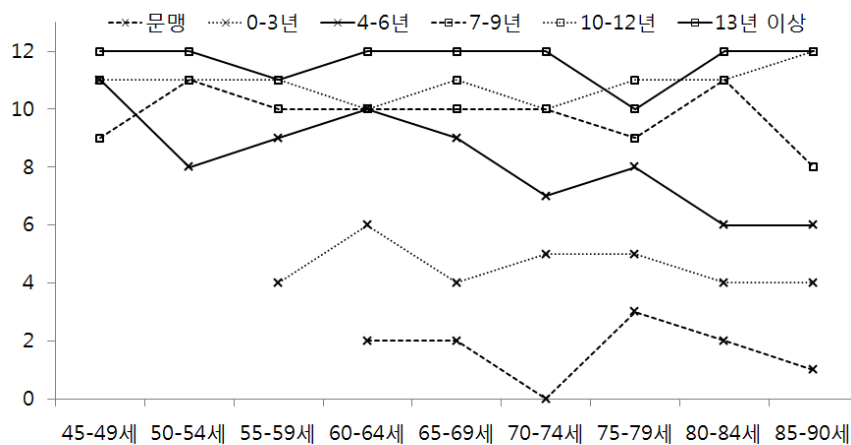


그림 9. 연령과 교육 수준에 따른 집단별 계산능력 점수의 표본 제10 백분위수

제5 백분위수의 경우 연령의 제곱(x_1^2)과 로그변환된 교육 수준($\log_e(x_2)$)이 선택되었으며 이 둘의 상호작용은 통계적으로 유의미하지 않아 제외되었다. 문맹집단의 효과는 모형에 포함되었다. 제10 백분위수의 경우에도 연령

해 추정되었고 신뢰구간은 역순위검정(inversion of rank test)에 의해 추정되었다. 예측변수의 형태는 손실함수를 더 작게 만드는 것으로 선택되었다.

의 제곱과 로그변환된 교육 수준이 더 적절한 변인으로 선택되었다. 두 변인의 상호작용과 문맹집단 효과도 유의미하게 나타나 모형에 모두 포함되었다. 제25 백분위수의 경우에는 연령(x_1)과 로그변환된 교육 수준, 그리고 이들의 상호작용이 모형에 포함되었으며 문맹집단 효과는 통계적으로 유의미하지 않아 제외되었다. 그림 6, 그림 8, 그림 10은 표 2의 최종 모형으로 추정된 계산능력 점수의 각 집단별 제

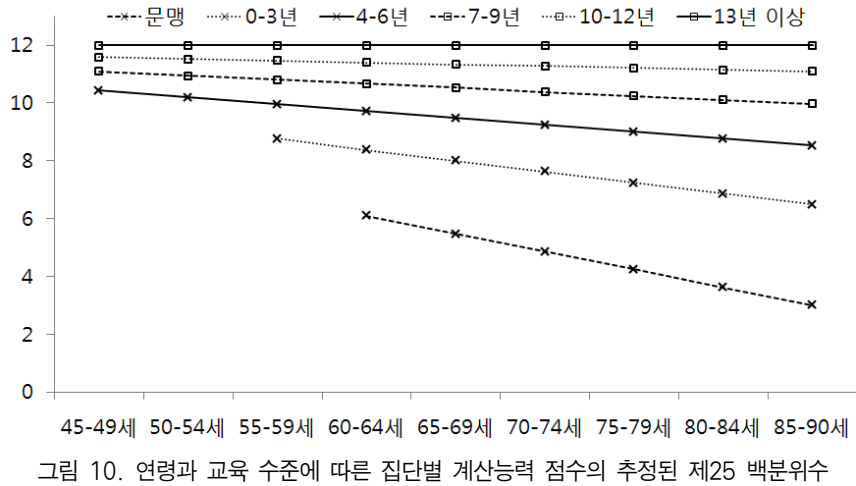


그림 10. 연령과 교육 수준에 따른 집단별 계산능력 점수의 추정된 제25 백분위수

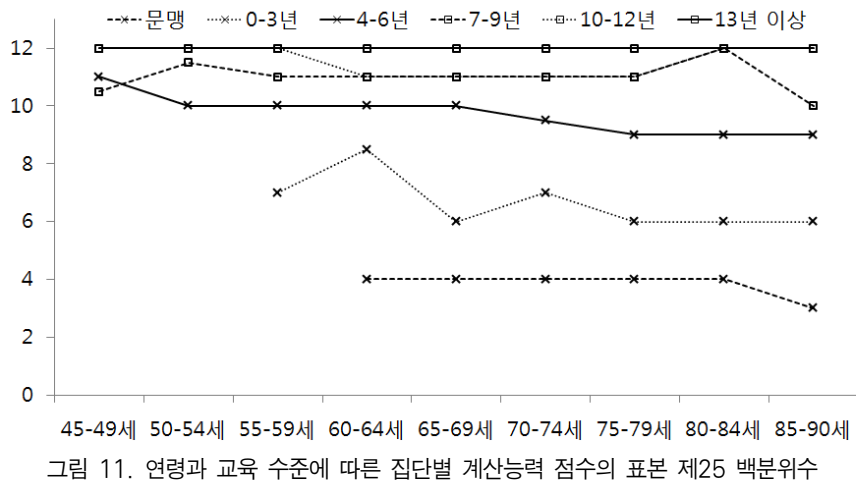


그림 11. 연령과 교육 수준에 따른 집단별 계산능력 점수의 표본 제25 백분위수

5, 제10, 제25 백분위수를 각각 나타낸다. 그림 7, 그림 9, 그림 11은 표본에서 계산된 각 집단별 제5, 제10, 제25 백분위수를 각각 나타낸다. 표본 분위수(그림 7, 그림 9, 그림 11)는 그림 4, 그림 5와 달리 연령이 높아짐에 따라 단조 감소하지 않고 학력이 높아짐에 따라 단조 증가하지 않는 형태를 보였다. 이는 표 1에서 확인한 바와 같이 49개의 각 하위 집단의 표본 크기가 안정적인 분위수 추정에 필요

한 만큼 충분하지 않기 때문으로 판단되며 따라서 개별 검사점수의 상대적 위치를 해석하기 위한 기준으로 표본 분위수를 사용하는 것은 적절치 않아 보인다. 반면 분위수 회귀모형을 통해 추정된 분위수(그림 6, 그림 8, 그림 10)는 계산능력의 검사점수가 연령이 증가함에 따라 단조 감소하고 학력이 높아짐에 따라 단조 증가하는 체계적 차이를 보이고 있어 검사점수에 의해 측정되는 인지능력의 차이에

대한 이론적 이해와 부합하는 기준을 제공하고 있다.

논 의

본 연구는 검사점수의 정규성을 가정하지 않고 이론에 의해 뒷받침되는 체계적인 차이를 갖는 기준을 추정하기 위해 분위수 회귀모형이 어떻게 적용될 수 있는지를 소개하였고, 지능 검사나 신경심리검사에서 사용되는 계산 능력 검사의 기준 자료를 이용하여 이 절차가 검사점수를 해석하는데 필요한 적절한 기준을 제공할 수 있음을 확인하였다. 검사 개발자의 의도에 의해 또는 다른 불가피한 이유로 많은 심리검사들이 정규분포와 다르게 분포하는 점수를 갖는다는 점에서, 검사점수의 기준으로 흔히 사용되는 평균과 표준편차가 아닌 분위수를 추정하는 절차에 더 많은 관심이 요구된다. 이러한 점에서 심리검사의 기준 추정에 분위수 회귀모형이 더 넓게 적용될 필요가 있다.

모형에 기반한 기준 추정 절차는 검사점수와(검사점수의 분포와 관련이 될 것이라고 여겨지는) 변인이 함수적 관계를 갖는다는 가정이 이론적으로나 경험적으로 충분한 근거를 가질 때 사용할 수 있다. 본 연구에서는 함수적 관계의 가장 단순한 형태인 선형적 관계나(예측변수 또는 검사점수를 변환함으로써 적용할 수 있는) 단조 함수적 관계를 갖는 경우에 분위수 회귀모형을 적용할 수 있음을 예시하였다. 그러나 검사점수와 관련 변인의 관계가 특정한 함수적 관계를 갖는다는 가정의 근거가 충분치 않을 때는 모형에 기반한 추정 절차를 적용하여서는 안 될 것이다. 즉, 본 연

구에서 예시된 인지기능 검사와 연령과 같이, 그 함수적 관계와 형태가 충분한 근거를 갖는 분야 외에는 모형에 기반한 기준 추정 절차를 적용하는 것에 신중해야 한다. 이와 같은 가정의 근거가 부족하다면 변수들의 관계가 복잡한 비선형 함수를 따르는 경우나(Koenker & Park, 1996) 함수적 관계를 따르지 않아 비모수적 절차가 필요한 경우에(Bollaerts, Eillers, & Aerts, 2006; Wei, Pere, Koenker, & He, 2006) 적용할 수 있는 보다 복잡한 분위수 회귀법의 적용을 검토해 볼 수 있다.

검사점수가 정규분포하지 않는 경우 박스-콕스(Box-Cox) 변환(Box & Cox, 1964)이나 Johnson 변환(Johnson, 1949)과 같은 변환절차를 통해 검사점수를 정규분포에 가깝게 변환하여 정규분포를 가정하는 기준을 사용할 수도 있다(예, Wechsler, 2008; WHO Multicentre Growth Reference Study Group, 2006). 그러나 본 예시에서와 같이 검사점수의 상/하한 값의 빈도가 높은 분포(그림 2)는 이러한 변환을 통해 정규분포와 유사한 형태로 변환되지 않으므로 분위수 회귀모형을 적용하는 것이 더 나은 선택이 될 것이다.

통계적 모형을 이용한 기준 추정 절차는 기준 집단이 여러 하위 집단으로 구성되어 있지만 각 하위 집단의 표본이 신뢰할 수 있는 기준 추정에 요구되는 크기를 충족하지 못할 때 이를 보완하기 위해 사용될 수도 있다. 그러나 이러한 절차가 모집을 잘 대표하는 충분한 크기의 표본을 대신할 수는 없다(장승민, 강연옥, 2012). 따라서 심리검사의 기준을 개발하는 연구자는 일차적으로 충분한 크기의 표본을 확보하고자 노력한 후 검사가 측정하는 심리적 속성의 특성에 대한 이해를 바탕으로 분위수 회귀모형을 비롯한 통계적 모형

을 이용한 절차의 적용 여부를 검토해야 할 것이다.

참고문헌

- 강연옥, 나덕렬 (2003). 서울신경심리검사 (SNSB). 인천: 휴브알앤씨.
- 강연옥, 장승민, 나덕렬 (2012). 서울신경심리검사 2판(SNSB-II). 인천: 휴브알앤씨.
- 배소영, 박금주 (2011). 한국판 맥아더-베이즈 의사소통발달 평가. 마인드프레스.
- 서은현, 이동영, 추일한, 윤종철, 김기웅, 우종인 (2007). 벤톤 시각 기억 검사(Benton Visual Retention Test)의 한국 노인 정상규준연구. 한국심리학회지: 임상, 26, 745-763.
- 우종인, 김기웅, 김성윤, 김주한, 우성일, 윤종철, 이강욱, 이근희, 이동영, 이정희, 주진형, 한설희 (2003). 한국판 CERAD 평가집 (CERAD-K). 서울대학교출판부
- 장승민, 강연옥 (2012). 정규분포가 가정된 심리검사의 기준추정을 위한 모형 기반 접근. 한국심리학회지: 일반, 31, 923-944.
- 황순택, 김지혜, 박광배, 최진영, 홍상환 (2011). 한국판 웨슬러 성인 지능 검사 4판 (K-WAIS-IV). 대구: 한국심리.
- Barone, C., & van de Werfhorst, H. G. (2011). Education, cognitive skills and earnings in comparative perspective. *International Sociology*, 26, 483-502.
- Bollaerts, K., Eilers, P., & Aerts, M. (2006). Quantile regression with monotonicity restrictions using P-splines and the L1-norm. *Statistical Modelling*, 6, 189-207.
- Box, G. E. P. & Cox, D. R. (1964) An analysis of transformations. *Journal of the Royal Statistical Society, Series B*, 26, 211-252.
- Bridges, A. J., & Holler, K. A. (2007). How many is enough? Determining optimal sample sizes for normative studies in pediatric neuropsychology. *Child Neuropsychology*, 13, 528-538.
- Christensen, K. J., Multhaup, K. S., Nordstrom, S., & Voss, K. (1991). A cognitive battery for dementia: Development and measurement characteristics. *Psychological Assessment*, 3, 168-174.
- Crawford, J. R., & Garthwaite, P. H. (2008). On the “optimal” size for normative samples in neuropsychology: Capturing the uncertainty when normative data are used to quantify the standing of a neuropsychological test score. *Child neuropsychology*, 14, 99-117.
- 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).
- Furr, R. M., & Bacharach, V. R. (2014). *Psychometrics: An Introduction*. Los Angeles: Sage Publications.
- Garsd, A., Ford, G. E., Waring, G. O., & Rosenblatt, L. S. (1983). Sample size for estimating the quantiles of endothelial cell-area distribution. *Biometrics*, 39, 385-394.
- Glisky, E. L. (2007). Changes in cognitive function in human aging. In D. R. Riddle (Ed.), *Brain Aging: Models, Methods, and Mechanisms*. Boca Raton, FL: CRC Press. Retrieved from <http://www.ncbi.nlm.nih.gov/books/NBK3885/>

- Guo, S. S., Roche, A. F., Chumlea, W. C., Johnson, C., Kuczmarski, R. J., & Curtin, R. (2000). Statistical effects of varying sample sizes on the precision of percentile estimates. *American Journal of Human Biology*, 12, 64-74.
- Hao, L., & Naiman, D. Q. (2007). *Quantile Regression*. Thousand Oaks, CA: Sage Publications.
- van Hooren, S. A., Valentijn, A. M., Bosma, H., Ponds, R. W., van Boxtel, M. P., & Jolles, J. (2007). Cognitive functioning in healthy older adults aged 64-81: A cohort study into the effects of age, sex, and education. *Aging, Neuropsychology, and Cognition: A Journal on Normal and Dysfunctional Development*, 14, 40-54.
- Jahng, S., Wood, P. K., & Trull, T. J. (2008). Analysis of affective instability in ecological momentary assessment: Indices using successive difference and group comparison via multilevel modeling. *Psychological Methods*, 13, 354-375.
- Johnson, N. L. (1949). Systems of frequency curves generated by methods of translation. *Biometrika*, 36, 149-176.
- 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., & Machado, J. (1999). Goodness of fit and related inference processes for quantile regression. *Journal of the American Statistical Association*, 94, 1296-1310.
- Koenker, R., & Park, B. J. (1996). An interior point algorithm for nonlinear quantile regression. *Journal of Econometrics*, 71, 265-283.
- Salthouse, T. A., Atkinson, T. M., Berish, D. E. (2003). Executive functioning as a potential mediator of age-related cognitive decline in normal adults. *Journal of Experimental Psychology: General*, 132, 566-594.
- Taylor, R. (1998). Continuous norming: Improved equations for the WAIS-R. *British Journal of Clinical Psychology*, 37, 451-456.
- Wei, Y., Pere, A., Keonker, R., & He, X. (2006). Quantile regression methods for reference growth charts. *Statistics in Medicine*, 25, 1369-1382.
- Wechsler, D. (2008). *Wechsler Adult Intelligence Scale-Fourth Edition: Technical and interpretive manual*. San Antonio, TX: Pearson.
- WHO Multicentre Growth Reference Study Group. (2006). WHO child growth standards based on length/height, weight and age. *Acta Paediatrica Supplementum*, 450, 76-85.
- Zachary, R. A., & Gorsuch, R. L. (1985). Continuous norming: Implications for the WAIS-R. *Journal of Clinical Psychology*, 41, 86-94.
- 원고접수일 : 2014. 11. 26.
수정원고접수일 : 2015. 01. 20.
게재결정일 : 2015. 02. 12.

Estimating psychological test norms using quantile regression models

Seungmin Jahng

Yeonwook Kang

Department of Psychology, Hallym University

Norms of not normally distributed psychological tests could be better estimated by quantiles instead of means and standard deviations. Psychological test norms can also be estimated using statistical models when the measured attribute has a functional relation with related covariates. The current study introduced a normative-quantile-estimation procedure using the quantile regression model when a test score is not expected to have normal distribution but assumed to have a functional relation with related covariates. The idea of the quantile regression model was briefly explained and its application to estimation of a test norm was illustrated. The 5th, 10th, and 25th percentiles of the calculation ability test scores with range of 0-12 were regressed on age and education by the quantile regression model using a normative sample of 1060 normal elderly. The estimated quantiles provided a norm that supports a theoretical model of the relations between calculation ability and the covariates of age and education.

Key words : quantile regression model, psychological test norm, model-based norming, normality assumption