한국심리학회지 : 임상
The Korean Journal of Clinical Psychology
2012, Vol. 31, No. 4, 923-944

정규분포가 가정된 심리검사의 규준추정을 위한 모형 기반 접근*

장 승 민† 강 연 욱

한림대학교 심리학과

객관적이고 신뢰할 수 있는 심리검사의 규준추정은 모집의 분포 특성을 잘 대표하는 충분한 크기의 표본에 의존한다. 많은 심리검사 규준연구들은 연령과 교육수준 등에 따라 개별적으로 규준을 추정해야 하는 여러 하위집단을 갖는 반면 각 집단의 안정적 규준추정에 필요한 표본크기를 확보하지 못하는 경우가 흔하다. 본 연구는 심리검사 규준추정에 있어 표본크기의 중요성을 구체적으로 살펴보고, 각 집단별 표본크기가 충분하지 않을 때 일반적인 규준추정 절차보다 안정적이고 신뢰할 수 있는 규준을 제공하는 모형 기반 규준추정 절차를 소개하였다. 구체적으로는 검사점수의 정규성을 가정할 수 있는 경우에 평균과 분산에 대한 회귀모형을 이용하여 각 규준집단의 평균과 표준편차를 추정하는 절차를 소개하였다. 또한 정상 장/노년 1067명의 한국판 보스톤 이름대기 검사 자료를 이용하여 모형 기반 규준추정 과정을 예시하여 이 절차가 실제 규준추정에 어떻게 적용될 수 있는지를 소개하였다. 끝으로 모형 기반 규준추정을 적용할 때의 고려사항들을 논의하였다.

주요어 : 심리검사, 규준추정, 모형 기반 접근, 정규성 가정

---

* 이 논문은 2010년도 한림대학교 교비 학술연구비[HRF-2010-007(2)]의 지원을 받아 연구되었음.

† 교신저자(Corresponding Author) : 장승민 / 한림대학교 심리학과 / (200-702) 강원도 춘천시 한림대학길 39
Fax : 033-256-3424 / Email : jahngs@hallym.ac.kr

심리검사 규준 제작의 기본적 목적은 개인이 얻은 심리검사점수가 그가 속한 집단의 전체 검사점수 분포에서 갖는 상대적 위치를 가늠할 수 있도록 검사점수 모집의 분포를 추정하는 것이다.[1] 일반적으로 검사점수의 모집이 정규분포를 따른다고 가정할 수 있는 경우는 표본자료로부터 해당 모집의 평균과 표준편차를 추정함으로써 모집 분포를 추정한다. 모집의 평균과 표준편차가 추정되면 검사점수는 모집 분포에서 갖는 상대적 위치, 즉 백분위에 대한 정보를 담고 있는 표준점수로 변환될 수 있으며 이를 통해 검사점수의 상대적 위치를 추정할 수 있다. 반면 검사점수 모집의 정규성을 가정하기 어렵거나 불가능하다고 판단되는 경우에는 표준점수 추정치로 백분위를 추정하는 것이 적절치 않다. 이러한 경우에는 일반적으로 모집의 평균과 표준편차를 추정하기 보다는 검사점수의 해석에 필요하다고 판단되는 백분위수를 직접 추정하여 검사점수와 비교할 수 있도록 한다.

검사를 통해 측정되는 심리적 특성의 분포가 연령이나 성별, 교육수준 등의 인구통계학적 변인에 따라 달라진다면 검사점수의 상대적 위치는 이들 변인에 의해 구분되는 하위집단별로 달라진다. 이와 같은 경우에는 하위집단에 따라 서로 다른 규준이 필요하며 따라서 집단별로 서로 다른 평균과 표준편차, 또는 백분위수를 추정해야 한다. 하위집단별로 서로 다른 규준이 필요한지 그렇지 않은지를 판단하기 위해서는 일차적으로 검사점수가 측정하는 심리적 구성개념에 대한 연구결과들을 종합한 경험적, 이론적 근거가 필요하다. 이러한 근거가 명확하지 않은 경우라면 각 하위집단별로 얻은 표본자료의 분석으로부터 경험적 근거를 찾아야 한다.

검사 규준은 표본자료로부터 추정되기 때문에 신뢰할 수 있는 규준추정을 위해서는 모집의 특성을 잘 대표하는 충분한 크기의 표본이 필요하다. 서로 다른 분포를 갖는 여러 하위집단들의 규준을 추정하는 경우에는 일반적으로 각 하위집단 마다 독립적으로 규준을 추정하며, 따라서 하위집단이 많아질수록 규준추정에 필요한 표본의 수도 증가한다. 예컨대, 만약 한 집단의 규준을 비교적 정확히 추정하기 위해 100명의 표본이 필요하고, 어떤 심리검사의 규준을 여덟 개의 연령집단에 대해 각각 추정해야 한다면, 필요한 전체 표본수는 800이 될 것이다. 각 연령집단별로 남녀 성별에 따라 규준이 달라야 한다면 1600명이 필요할 것이고, 이를 다시 네 개의 학력집단별로 세분화하여 규준을 구해야 한다면 필요한 표본수는 6400이 될 것이다.

필요한 크기보다 작은 표본으로부터 추정된 규준을 사용하게 되면 검사점수의 상대적 위치에 대한 추정이 부정확하게 되고 검사점수의 의미를 해석하는데 오류를 낳게 되며, 따라서 이 규준을 이용한 피검자에 대한 평가나 진단도 정확하지 않게 된다. 그러나 여러 하위집단의 심리검사 규준을 추정하는 연구들에서 규준추정에 필요한 충분한 표본수가 사용되는 경우는 흔치 않다. 이는 규준개발 연구자들이 규준추정에 필요한 표본수를 확보하는 것의 중요성을 충분히 인식하지 못하는 데에

1) 심리검사는 검사점수를 점수들 간의 상대적 위치, 즉 순위 정보를 담고 있는 규준을 참조하여 해석하는 규준참조검사와 검사가 평가하고자 하는 특정 영역이나 내용 등의 준거를 참조하여 해석하는 준거참조검사로 구분할 수 있다. 본 논문에서는 규준참조검사에서의 규준추정에 관한 내용을 다룬다.

도 일부 기인하겠지만, 여러 가지 현실적인 이유로 인해 규준이 추정되는 모든 하위집단에 필요한 충분한 크기의 표본을 확보하는 것이 어렵기 때문이기도 하다. 심리검사 규준개발 연구자들은 일종의 딜레마에 빠져 있는 셈이다. 학문적 엄격성을 위해 충분한 크기의 표본 없이는 규준추정을 하지 않을 것인가? 아니면 현실적 어려움을 이유로 오용의 가능성이 큰 부정확한 규준추정을 계속할 것인가?

본 연구는 심리검사 규준추정에 있어 표본크기 문제의 중요성을 구체적으로 살펴보고, 각 집단별 표본크기가 충분하지 않을 때 일반적인 규준추정 절차보다 안정적이고 신뢰할 수 있는 규준을 산출하는 대안적인 규준추정 방법을 제안한다. 모형 기반 접근법이라 부를 수 있는 이 방법은 하위집단을 구성하는 인구통계학적 변인에 따른 검사점수 분포의 차이가 어떤 체계적인 규칙을 갖는다는 가정이 가능할 때, 이 체계적 차이를 수학적 함수로 표현하는 통계 모형을 사용하여 규준을 추정한다. 본 연구에서는 검사점수의 정규성을 가정할 수 있는 경우에 적용할 수 있는 회귀모형을 이용하여 각 규준집단의 평균과 표준편차를 추정하는 절차를 제안한다. 또한 실제 심리검사 규준 자료를 이용해 모형 기반 접근법을 적용한 규준추정의 과정을 예시함으로써 연구자들이 규준개발 연구에 실제로 이 접근을 적용할 수 있도록 안내하고자 한다.

분포의 정규성을 가정하는 심리검사 규준추정에서 표본크기의 문제

평균 추정에서 표본크기의 문제

검사점수 규준의 추정은 표본자료에 의존하기 때문에 규준추정에 사용된 표본이 달라진다면 추정된 규준도 달라질 수 있다. 또한 표본의 크기가 작은 경우 추정값의 신뢰구간이 넓어지기 때문에 추정된 규준의 정확성을 자신할 수 없게 된다. 따라서 보다 정확한 규준의 추정을 위해서는 표집오차를 줄일 수 있도록 규준집단으로부터 충분한 수의 표본을 얻는 것이 필요하다. 평균의 경우 규준집단별로 적어도 50명 이상의 표본점수가 있어야 신뢰할 수 있는 추정이 가능하며 제10 백분위수와 같은 다른 값의 경우에는 더 많은 표본수가 필요하다. Bridges와 Holler(2007)는 모집의 검사점수가 정규분포할 때 평균 추정값의 신뢰구간의 폭이 수용 가능한 수준에 들기 위해서는 각 규준집단별로 적어도 50에서 70 정도의 표본수를 사용할 것을 제안하였으며, Crawford와 Garthwaite(2002, 2008)는 같은 조건에서 평균으로부터 멀리 떨어진 값을 추정할수록 더 많은 표본이 필요하다고 지적하였다. 예를 들어, 평균으로부터 1 표준편차 아래($z = -1.0$, 백분위=15.8)나 1.5 표준편차 아래($z = -1.5$, 백분위=6.7) 지점은 분포의 정규성을 가정하는 심리검사 규준에서 집단을 구분하기 위한 지점으로 흔히 사용되는데 이들 값에 대한 추정이 수용 가능한 범위의 신뢰구간을 갖기 위해서는 각 규준집단별로 100 이상의 표본수가 필요하다(Crawford & Garthwaite, 2008).

하나의 규준집단의 규준추정에 필요한 표본수를 감안할 때, 검사점수와 관련된 인구통계학적 변인들에 따라 고려해야 할 규준집단의 수가 많아진다면 필요한 전체 표본수는 수천에서 수만이 될 수도 있다. 물론 인구통계학적 변인에 따라 집단 평균이 달라질 것이라는 가정이 이론적으로나 경험적으로 타당하지 않

거나 주어진 표본자료를 이용한 집단 평균차에 대한 통계적 검증의 결과가 이를 지지하지 않는다면, 전체집단에 대해서 하나의 평균만 추정할 수도 있다. 여러 하위집단별로 서로 다른 규준을 추정해야 함에도 각 집단별로 필요한 표본수를 확보하지 못한 연구들에서는 추정된 규준평균의 집단 간 차이가 이론적으로 설명되기 어려운 불규칙한 형태를 보이는 경우가 흔하다. 그리고 이러한 불규칙성은 대개 각 집단별로 적은 수의 표본을 사용함으로써 증가한 표집오차에 크게 기인한다.

실제 국내에서 수행된 여러 규준연구들에서도 이와 같은 결과를 확인할 수 있다. 최근 국내에서는 치매나 혈관성 질환으로 인한 인지능력의 비정상적 저하를 평가하기 위한 다양한 종류의 신경심리검사들의 규준연구가 활발히 이루어지고 있다. 안타깝게도 이 연구들이 보고한 검사 규준들 대부분에서 적어도 부분적으로는 실제 검사현장에 적용하여 검사점수를 해석하는 것을 주저하게 만드는 측면들이 발견된다. 예를 들어, 서은현 등(2007)이 554명의 60세에서 90세 사이의 정상노인 표본을 이용해 추정한 벤톤 시각기억검사(Benton Visual Retention Test: BVRT)의 집단별 규준추정값을 보자. 이들이 보고한 BVRT 실시 A의 정반응 점수 규준(754 쪽, 표 4)에는 교육연한이 3년 이하인 남성 집단의 경우, 다른 교육수준 집단이나 동일한 교육연한을 갖는 여성 집단과는 달리, 연령이 높을수록 검사점수의 평균이 높은 것으로 나타났다(그림 1 참조). 이 결과가 실제 저학력 남성 집단의 경우 연령이 높을수록 BVRT 점수로 반영되는 시각기억능력의 평균수준이 높고 따라서 더 높은 규준평균을 적용해야 한다는 것을 의미한다고 보기는 어렵다. 오히려 이러한 결과는 저학력 남성 집단의 표본수가 적기 때문에 규준추정의 오차가 커져 나타난 것으로 보는 것이 타당하다. 이 연구에서는 집단별 규준추정에 사용되는 표본수를 더 많이 확보하기 위해 중복연령규준법을 사용하였음에도 불구하고, 교육연한 3년 이하 남성 집단의 각 연령집단별 표본수가 60세에서 74세 집단은 13, 65세에서 79세 집단은 17, 70세에서 84세 집단은 15, 75세에서 90세 집단은 11에 불과하였다.

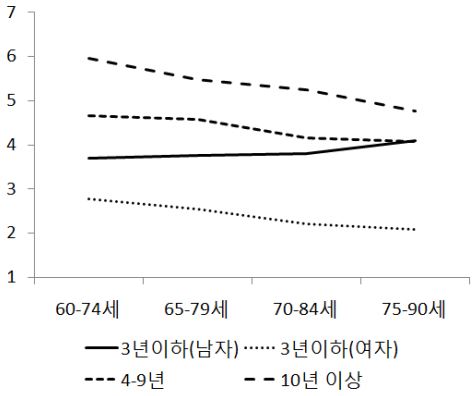

그림 1. BVRT 실시 A 정반응 점수의 연령과 학력에 따른 평균 (자료출처: 서은현 등, 2007, 표 4)

또 다른 예로, 45세에서 90세 사이의 장/노년 1,582명으로 구성된 표본으로부터 한국형 간이 정신상태검사(Korean-Mini Mental State Examination: K-MMSE)의 규준을 추정한 강연욱(2006)의 연구에서는 교육연한이 7년 이상인 85세에서 90세 집단의 K-MMSE 평균점수(28.50)가 같은 교육수준의, 연령대가 낮은 다른 집단들의 평균(26.41-27.77)보다 높게 나타났다(7 쪽, 표 3). 교육연한이 6년 이하인 다른 집단들의 경우 연령이 높을수록 K-MMSE 평균점수가 일관적으로 낮게 나타난 것으로 볼 때, 실제로 해당 학력의 85세 이상 연령집단이 85세 미만의 연령집단들보다 평균적으로 더 높은 인지능력을 갖기 때문에 이와 같은 결과가

얻어졌다고 해석하기는 어렵다. 오히려 이와 같은 결과는 해당 표본집단의 표집오차가 크기 때문에 나타났을 가능성이 더 큰 것으로 판단된다. 실제 이 연구에서 사용된 교육연한 7년 이상의 85세 이상 집단의 표본수는 2에 불과하였으며 강연욱(2006)도 85세 이상 고령자들의 규준에 대한 제한점을 지적하고 있다.

앞서 언급한 두 연구 외에도 최근 국내에서 수행된 대부분의 신경심리검사 규준연구에서 이와 유사한 결과들을 확인할 수 있으며 그 원인은 각 연구에서 사용된 표본의 크기가 규준집단별로 필요한 정도를 충족하지 못하였기 때문으로 판단된다. 몇 가지 예를 들면, 전산화된 치매선별 검사(최진영, 박미선, 조비룡, 양동원, 김상윤, 2002, 표 6), 아동용 언어기억 검사(강연욱, 2003, 표 3), 노인용 이야기 회상 검사(안효정, 최진영, 2004, 표 7), 숫자폭 따라하기 검사(송호정, 최진영, 2006, 표 7-표 10), K-MoCA(강연욱, 박재설, 유경호, 이병철, 2009, 표 2), 한국판 치매 평가 검사(석정서, 최진영, 김호영, 2010, 표 2), 위스콘신 카드 분류 검사(지연경, 조민경, 한지원, 김태희, 김기웅, 2011, 표 2) 등에서 부분적으로 유사한 문제점이 확인되었다. 이 보고들에 나타난, 설명하기 어려운 집단 간 규준평균 차이에는 예외 없이 표본수가 적은 규준집단이 관련되어 있었다.

표준편차 추정에서 표본크기의 문제

앞서 언급한 대로 검사점수 모집분포의 정규성을 가정하면 표준점수가 검사점수의 백분위 정보를 제공할 수 있다. 검사점수를 표준점수로 변환하기 위해서는 각 규준집단의 평균뿐 아니라 표준편차의 값도 필요하며, 따라서 표준점수를 이용한 규준추정은 평균의 추정뿐 아니라 표준편차의 추정도 포함한다.

두 집단의 평균이 동일하더라도 한 규준집단(A 집단)의 표준편차가 다른 규준집단(B 집단)의 표준편차보다 더 크다면, 평균보다 낮은 검사점수의 백분위는 표준편차가 더 큰 집단(A 집단)에서 더 높게 평가된다. 예를 들어, 평균, 두 집단의 표준편차, 검사점수가 각각 $M_A=M_B=10$, $SD_A=2.5$, $SD_B=2$, $X=8$이라면, 각 집단에 대한 X의 표준점수는 $Z_A=(8-10)/2.5=-0.8$, $Z_B=(8-10)/2=-1$이 되어, 검사점수 8점의 백분위는 A 집단에서는 21.2, B 집단에서는 15.8이 된다. 두 집단의 표준편차의 차이가 큰 경우라면, A 집단의 평균이 B 집단의 평균보다 높고 A 집단에 속한 갑의 검사점수가 B 집단에 속한 을의 검사점수보다 낮더라도, 즉 갑이 을보다 자신이 속한 규준집단의 평균에 비해 수행이 더 많이 떨어지더라도, 갑의 백분위가 을의 백분위보다 더 높게 평가될 수도 있다(예, $M_A=11$, $M_B=10$, $SD_A=5$, $SD_B=2$, $X_{갑}=7$, $X_{을}=8 \Rightarrow Z_{갑}=(7-11)/5=-0.8$, $Z_{을}=(8-10)/2=-1$). 따라서 규준을 추정할 때 각 집단의 표준편차가 정확하게 추정되지 않는다면 이를 사용하여 계산된 표준점수와 백분위 추정 역시 신뢰할 수 없게 된다. 즉, 정규분포를 따르는 검사점수의 규준추정에서 표준편차의 정확한 추정은 평균의 정확한 추정만큼 중요하다고 할 수 있다.

평균의 신뢰구간과 마찬가지로 표준편차의 신뢰구간 역시 표본수가 커질수록 작아지기 때문에 효율적인 표준편차의 추정을 위해서도 충분한 크기의 표본이 필요하다. 그림 2는 앞서 언급한 서은현 등(2007)의 연구에서 보고된 하위집단별 BVRT 실시 A의 정반응 점수 표준편차 추정값을 보여준다. 앞서 그림 1에서 교육연한 3년 이하 남성 집단의 연령별 평균 차

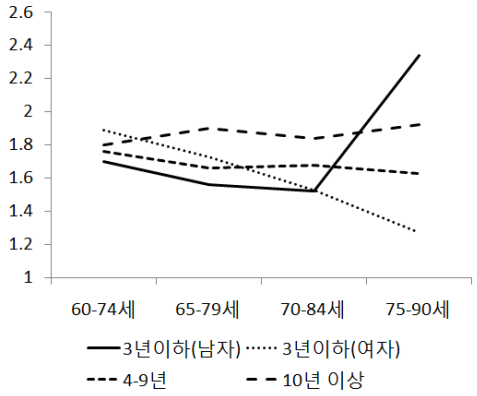

그림 2. BVRT 실시 A 정반응 점수의 연령과 학력에 따른 표준편차 (자료출처: 서은현 등, 2007, 표 4)

이의 형태가 다른 집단과 다르게 나타났던 것과 마찬가지로 해당 집단의 연령별 표준편차의 차이의 형태도 다른 집단과 매우 다르게 나타나고 있다. 특히, 교육연한 3년 이하 남성 집단의 표준편차가 75세에서 90세 연령대에서 다른 연령대에 비해 매우 큰 값을 보이고 있다. 이 집단의 평균과 표준편차의 값이 다른 집단을 고려했을 때 일정한 패턴을 벗어나 큰 값을 갖는다는 점과 이 집단의 표본수(11)가 매우 적다는 점을 감안하면, 이 집단에는 집단내의 다른 사람들에 비해 매우 높은 점수를 받은 일부의 사람들이 섞여 있음을 의심할 수 있다. 실제 보고된 해당집단의 95백분위수는 모든 집단 중에서 가장 높은 10점이었으며 10점은 BVRT 실시 A 정반응 점수의 만점이었다(서은현 등, 2007, 754 쪽, 표 4). 다시 말해, 이 집단의 표준편차 추정값이 다른 집단에 비해 매우 높게 나타난 것은 모집의 실제 표준편차를 반영한 것이라기보다는 표본수가 적어 표집오차가 커졌기 때문으로 보는 것이 타당하다. 따라서 이 규준집단에서 얻은 평균 및 표준편차의 추정치를 이용하여 해당 집단에 속한 개인의 검사점수를 해석하는 것은 적절치 않다. 정리하자면, 여러 하위집단에 따라 검사점수의 분포가 달라져 각 하위집단별로 서로 다른 평균과 표준편차를 각각 추정해야 하는 경우에는 각 하위집단별로 일정 수준 이상의 표본수가 필요하며, 충분한 표본수에 근거하지 않은 규준추정값을 근거로 검사점수를 해석하는 것은 위험한 일이다.

규준연구에서 필요한 표본수 확보의 어려움

안타깝게도 대부분의 규준연구들에서 하위집단별로 규준추정에 필요한 표본수가 확보되지 못하는 것은 현실적으로 불가피한 측면이 있다. 규준집단의 수가 많은 경우에는 각 집단별로 100명 이상의 표본을 얻으려면 전체 표본수가 수천 이상에 이르기도 한다. 그러나 규준연구를 위해 수천의 표본을 확보하는 데는 시간과 비용에 대한 현실적인 제약이 매우 크다. 강연욱(2006)의 연구에서는 1500명 이상의 표본이 사용되었음에도 연구에서 네 개의 학력집단과 다섯 개의 연령집단이 고려됨으로써 모두 20개 집단의 규준을 추정하였기 때문에 각 집단별 평균 표본수가 100명에 미치지 못하였다.

또한 특정 규준집단의 표본을 확보하기가 쉽지 않은 경우도 있다. 예를 들어 40대 문맹자나 80세 이상, 그 중에서도 대졸학력자와 같은 경우는 규준집단에 해당하는 표본을 구하기가 쉽지 않기 때문에 필요한 표본수를 확보하는 것이 매우 어렵다. 강연욱(2006)의 연구에서는 문맹의 경우 연령에 따라 표본수가 15에서 97이었으며, 85세 이상의 경우 학력에 따라 2개에서 15개의 표본이 사용되었다.

규준자료를 충분히 확보하는 데 따르는 현실적 제약은 특히 검사를 집단으로 실시하지

않는 일대일 대면 검사의 경우 두드러진다. 전형적으로 임상심리평가사 등의 전문적인 교육을 받은 검사자에 의해 한 번에 한 피검자에 대해 실시되는 일대일 대면 검사는, 집단으로 실시할 수 있는 검사들에 비해 검사 실시에 더 많은 시간과 비용을 요구하기 때문에 규준추정에 필요한 충분한 표본수를 확보하는 것이 더욱 어렵다. 반면 일대일 대면을 통해 실시되는 검사들은 일반적으로 검사실시의 정확성과 검사결과 해석의 타당성을 더욱 요구하기 때문에 객관적이고 신뢰할 수 있는 규준의 필요성이 더 절실히 요구된다.

정규성을 가정한 모형 기반 규준추정 절차

규준추정을 위한 모형 기반 접근의 개념

검사점수 분포의 차이가 규준집단 간에 체계적인 규칙성을 갖는 경우라면, 보다 정확한 규준추정을 위하여 충분한 표본수를 확보해야 하는 필요성과 이 필요성을 달성하기 어려운 현실적인 제약 사이의 딜레마를 줄일 수 있는 한 가지 가능한 접근이 있다. 예를 들어 어떤 심리검사가 측정하는 특정 인지 능력이 연령이 높아질수록 일정하게 감소하는 것으로 알려져 있다고 하자. 이 경우 각 연령 규준집단의 표본수가 충분하지 않다면 각 표본으로부터 추정된 평균이 표집오차로 인해 연령에 따라 일정하게 감소하지 않고 들쭉날쭉하게 나타날 수 있다. 그러나 연령에 따른 검사점수 평균의 일정한 감소를 표현하는 적절한 모형, 예컨대 선형 회귀모형을 사용하여 연령집단별 평균을 추정한다면, 모형에 의해 추정된 평균은 각 연령집단 표본에서 얻은 불규칙적인 표본평균보다는 실제 연령에 따른 검사점수 평균의 변화 또는 차이를 더 잘 반영하게 될 것이다.

모형을 이용하여 각 연령집단의 평균을 추정할 때는 해당 연령집단의 표본뿐 아니라 모형에 포함된 전체 표본자료가 사용되므로 결과적으로 해당 연령집단의 표본만을 사용해 얻은 평균보다 신뢰구간이 더 좁은, 다시 말해 더 신뢰할 수 있는 평균의 추정치를 얻을 수 있다. 마찬가지로 만일 연령집단에 따라 표준편차가 다르고 이 차이에 일정한 규칙성이 있다면, 평균과 유사한 방식으로 표준편차를 모형화함으로써 적은 표본수로 인한 추정의 부정확성을 보완할 수 있다. 다시 말해 규준집단 분포 간의 체계적 관계를 반영하는 적절한 통계 모형을 이용한 모형 기반 접근을 사용하면, 규준추정시 각 규준집단의 표본수가 적어 발생할 수 있는 표집오차의 영향을 최소화할 수 있다.

규준추정을 위한 모형의 설정

모형 기반 규준추정 절차를 적용하기 위해서는 분포의 하위집단별 차이를 적절하게 기술할 수 있는 이상적인 규칙의 형태를 결정하는 것이 중요하다. 예컨대 연령과 교육수준에 따라 나뉜 각 집단들에서 얻어진 표본평균과 표본표준편차와 잘 부합하면서 비교적 단순히 표현될 수 있는 규칙을 설정해야 한다. 인지능력을 측정하는 대부분의 신경심리검사의 경우에는 학력이나 연령이 높을수록 검사점수의 평균이 단조 증가하거나 단조 감소하는 형태가 일반적이며 따라서 이러한 규칙성은 1차 선형 함수나 로그 등으로 이를 변환한 함수를

이용하여 표현할 수 있다.

검사점수 분포의 정규성이 타당하다고 판단되는 경우 규준추정은 각 규준집단의 평균과 표준편차를 추정하는 문제로 단순화 된다.[2] 따라서 검사점수의 정규성을 가정할 수 있는 경우, 모형 기반 접근은 결국 여러 하위집단의 평균과 표준편차를 일정한 규칙을 따르는 함수로 모형화하는 것을 의미한다.

평균과 표준편차를 함수를 통해 모형화하는 한 가지 방법은 규준집단을 구분하는 변인, 예를 들어 연령집단과 교육수준집단을 범주형 변인이 아닌 연속형 변인으로 간주하여 집단 간 차이가 점진적인 형태를 나타내도록 하는 것이다. 이와 같은 방식의 한 예는 신경평가집(Neurological Assessment Battery) 중 Visual Discrimination Test(Stern & White, 2009) 등에서 사용된 연속규준화(continuous norming) 절차이다.

연속규준화 절차는 먼저 하위집단별로 표본평균과 표본표준편차를 얻은 후 이 값들을 각각 회귀모형으로 추정하는 2단계를 거쳐 하위집단별 분포의 점진적인 차이를 추정한다. Zachary와 Gorsuch(1985)와 Taylor(1998)는 WAIS-R(Wechsler, 1981)의 규준표에 나타난 연령에 따른 평균과 표준편차의 불규칙적 차이를 2단계 연속규준화 절차를 이용하여 조정하였다. 그러나 이와 같은 2단계 추정법은 각 하위집단별로 표본수의 차이가 크게 되면 적은 수의 표본에 의해 추정된 집단 평균(또는 표준편차)과 많은 수의 표본에 의해 추정된 집단 평균의 추정 신뢰성의 차이를 반영하지 못하게 된다. 더욱이 연령, 학력, 성별 등의 변인 중 어떤 변인을 모형에 포함시킬 것인가와 각 변인을 어떤 형태의 함수로 모형화할 것인가를 결정하기 위해서는 규준자료에 대한 모형적합도를 비교하고 변인 효과의 통계적 유의성을 판단해야 하는데, 1단계에서 얻은 소수의 집단별 평균(또는 표준편차)을 이용해 모형을 구성하는 방식으로는 전체자료에 대한 모형의 통계적 유의성을 적절히 평가할 수 없다.[3] 그러나 연속규준화가 반드시 2단계를 거칠 필요는 없다.

본 연구는 Zachary와 Gorsuch(1985)와 Taylor(1998)가 사용한 2단계 연속규준화가 갖는 단점을 보완하기 위하여 전체 표본자료에 대한 회귀분석을 이용하는 단일 단계의 연속규준화를 제안한다. 표본자료 전체를 이용한 회귀분석을 사용하면 연속적인 형태를 띠는 집단별 평균(또는 표준편차)을 한 번에 추정할 수 있고, 회귀식의 추정과정에 집단별 표본수가 가중되어 계산됨으로써 하위 집단별 추정 신뢰성의 차이가 반영되며, 전체자료에 대한 모형의 적합성을 평가할 수 있고, 집단을 구분하는 변인 효과의 통계적 유의성을 적절히 판단할 수 있다. 예를 들어 41세에서 80세까지 5년 단위로 구분된 8개의 연령집단에 대해 어떤 기억능력검사의 규준을 추정한다고 하자. 기억능력의 노화에 대한 연구들이 공통적으로 40세 이후 연령이 증가할수록 기억능력이 점진적으로 감퇴한다고 보고하였다면 기억능력

2) 검사점수 분포의 정규성 가정은 흔히 표본을 통해 경험적으로 확인된다. 표본자료의 분포를 히스토그램이나 Q-Q plot 등으로 시각화하여 정규성을 판단하는 것이 일반적이며 Kolmogorov-Smirnov 검정 등을 사용하여 통계적으로 검증하기도 한다. 정규성 가정과 같은 모집분포에 대한 가정은 모형을 적용하기 이전에 확인되기 보다는 모형을 적합시킨 이후에 얻어진 잔차를 이용하여 확인된다.

3) 모수 추정 시 2단계 접근법의 단점에 대해서는 Jahng, Wood, Trull(2008)이 다층모형의 맥락에서 제시한 개략적 설명을 참고(p. 363).

검사점수에 대해 다음과 같이 회귀모형을 사용하여 각 연령집단의 규준을 추정할 수 있다.

$$y_i = f(x_i) + \varepsilon_i$$

여기에서 y_i는 검사점수를, x_i는 각 연령집단에 부여된 연속적인 수치, 예를 들면 0(41세에서 45세)에서 7(76세에서 80세)을 의미하고, $f(x)$는 주어진 x값이 의미하는 연령집단의 검사점수 기댓값, 즉 평균을 나타내며, ε_i는 검사점수 y_i가 집단평균 $f(x_i)$에 대해 갖는 편차, 즉 모형에서의 오차를 의미한다. 정규성을 가정한 모형에서 ε_i는 평균 0, 표준편차 σ를 갖는 정규분포를 따른다.

규준집단 간 평균의 차이를 표현하는 함수 $f(x)$의 가장 단순한 형태는 $f(x)=\beta_0+\beta_1 x$, 즉 일차 선형함수로 표현되며 이 경우 위의 수식은 다음과 같은 단순회귀모형이 된다.

$$y_i = \beta_0 + \beta_1 x_i + \varepsilon_i$$

만일 규준자료로부터 추정된 β_0와 β_1이 각각 9와 -0.5였다면 41세에서 45세의 연령집단(x=0)에 대한 검사점수의 평균 추정값은 9, 46세에서 50세의 연령집단(x=1)의 평균 추정값은 8.5, ... , 76세에서 80세의 연령집단(x=7)의 평균 추정값은 5.5가 된다. 기억능력의 평균이 연령집단에 따라 일정하게 감소하지 않고 감소의 폭이 일정하게 달라진다면 연령집단을 나타내는 변수 x를, 예컨대 x^2이나 (자연로그) $\ln(x)$와 같이 적절한 형태로 변환하여 자료에 부합하는 함수관계를 표현할 수 있다.[4]

검사점수에 영향을 주는 변인이 연령, 학력, 성별 등과 같이 여럿인 경우는 여러 변인들이 검사점수를 동시에 예언하는 중다회귀모형을 설정하여 규준을 추정할 수 있다. 또한 평균함수의 형태가 제곱이나 로그 등의 변환을 통해서도 자료의 특성과 잘 부합하지 않는다면 평균을 다른 형태의 비선형 함수로 나타내는 비선형 회귀모형을 이용해 규준을 추정할 수 있다.[5] 구체적인 함수를 결정하기 전에 평활법(smoothing) 등의 비모수적 분석 절차를 통해 대략적인 함수의 형태를 유추할 수도 있다. 물론 연령, 학력, 성별 등의 변인들이 검사점수의 평균에 유의미한 영향을 주지 않는 것으로 확인된다면 집단의 구분 없이 전체집단에 대해 하나의 평균만을 규준으로 사용할 수도 있다.

표준편차의 경우도 집단에 따라 그 값이 달라지고 그 차이에 일정한 규칙이 존재한다면 이 또한 회귀모형을 이용하여 추정할 수 있다. 일반적으로 표준편차의 모형화는 분산의 모형화를 통해 표현된다. 변인 x의 기댓값이 $f(x)$를 따르는 회귀모형에서, 다른 변인 w에 의해 영향을 받는 오차[$\varepsilon_i = y_i - f(x_i)$] 분산의 회귀모형은 다음과 같이 표현될 수 있다(Carroll & Ruppert, 1988).

$$\varepsilon_i^2 = \sigma^2 g(w_i) + \delta_i$$

위의 식이 분산의 회귀모형인 이유는 오차

4) 연령집단 변수 x대신 x^2을 이용해 모형을 설정하는 경우 주의할 점이 있는데, 만약 x와 x^2 둘 모두를 모형에 포함시키는 경우 평균의 변화가 단조증가 혹은 단조감소하지 않고 증가하다가 감소하거나 감소하다가 증가하는 형태를 나타낼 수 있다.

5) 연령에 따른 다양한 성장곡선을 표현하는 여러 가지 선형 및 비선형 함수(예, 로지스틱 함수)들과 자료에 적합한 성장곡선함수의 선택 절차에 대한 설명은 Burchinal과 Appelbaum(1991)을 참고.

제곱의 기댓값, 즉 $\sigma^2 g(w_i)$가 분산의 근사값이라는 사실에 기초하고 있다. 여기에서 σ^2은 상수로 기본분산[$g(w)$가 1일 때의 분산]을 의미하며 평균 회귀모형에서 절편과 같은 역할을 한다. 분산 $\sigma^2 g(w)$는 w의 값에 따라 달라지며, 이는 w의 값이 집단을 나타내는 경우 집단에 따라 분산(또는 표준편차)이 함수 g에 의해 달라짐을 의미한다. 변인 w에 따른 분산의 값을 결정하는 함수 $g(w)$의 형태로는 $g(w)=\exp(\theta w)$가 널리 사용된다. 여기에서 exp는 상수 e를 밑으로 하는 지수함수를 의미하며 θ는 변인 w의 비선형 회귀계수 이다[$\sigma^2 g(w)=\sigma^2 e^{\theta w}$]. 평균에서와 마찬가지로 분산의 경우도 설명변인(w)이 하나로 한정될 필요는 없다. 또한 평균과 분산이 같은 공변인에 의해 영향을 받을 수도 있는데 이 경우에는 분산함수 g가 w가 아닌 x의 함수, 즉 $g(x)$로 표현된다(Jahng, 2008).

분산함수가 $g(w)=e^{\theta w}$인 경우, 분산 회귀모형은 대표적으로 다음의 세 가지 중 하나의 절차를 통해 추정될 수 있다. 첫 번째 방법은 먼저 평균회귀모형에서 구한 편차(검사점수에서 예측값을 뺀 값)의 제곱을 자연로그로 변환한 뒤 이를 결과변인으로, w를 설명변인으로 하는 선형회귀모형을 적용하는 것이다.[6] 두 번째는 편차제곱의 오차 δ_i가 감마분포를 따르는 것으로 가정하는 일반화 선형모형(generalized linear model)을 적용할 수 있다. 이때 ε_i^2의 기댓값($\sigma^2 e^{\theta w_i}$)과 w_i를 연결하는 연결함수는 자연로그를 사용한다. 세 번째는 평균과 분산을 동시에 모형화하는 방법으로, 이 경우에는 분산 회귀모형의 오차항 없이 분산함수인 $\sigma^2 g(w)$만을 평균 회귀모형의 오차분산으로 모형화한다. 이 방법은 편차제곱을 계산하는 절차 없이, 평균의 회귀모형을 추정하면서 이 모형의 분산이 w에 따라 달라질 수 있도록 허용한다(Jahng, 2008).[7]

정리하면, 검사점수의 정규성이 가정되고, 평균이나 표준편차, 또는 둘 모두의 하위 규준집단들 사이의 차이가 함수로 표현되는 체계적 규칙을 따른다고 가정할 수 있을 때, 평균에 대한 회귀모형과 분산에 대한 회귀모형을 이용하여 각 하위집단의 규준을 추정할 수 있다. 이 경우 각 하위집단의 표본수가 상대적으로 적은 경우에도 이론적으로 타당하고, 해석이 자연스러우며, 표집오차의 영향이 덜한 검사점수의 규준을 얻을 수 있다. 모형 기반 규준추정 절차의 구체적인 적용은 다음 절에서 예시한다.

모형기반 접근을 이용한 규준추정의 예: 한국판 보스톤 이름대기 검사

표본 및 방법

규준자료 및 검사도구

이 절에서는 45세에서 90세 사이의 정상 장노년 1067명(남자: 600명, 여자: 467명)을 대상

6) 분산함수 $\sigma^2 e^{\theta w_i}$에 자연로그를 취하면 $\ln(\sigma^2 e^{\theta w_i})=\ln(\sigma^2)+\ln(e^{\theta w_i})=2\ln(\sigma)+\theta w_i$가 된다. 따라서 편차제곱에 자연로그를 취한 값의 회귀모형은 $\ln(\varepsilon_i^2)=2\ln(\sigma)+\theta w_i+\xi_i$가 되어 $a_i=c+\theta w_i+d_i$의 형태를 갖게 되고, 따라서 절편 c[$=2\ln(\sigma)$]와 기울기 θ를 갖는 선형회귀모형을 적용할 수 있다. 이때 d_i는 정규분포를 따르는 것으로 가정된다.

7) 다음 절의 분석 예시에서는 분산함수의 모형화를 위해 세 번째 방법을 이용하였다. 분산함수 추정을 위한 다양한 접근방법과 각 추정 방법의 수학적 특성에 대한 구체적인 설명은 Davidian과 Carroll(1987)을 참고.

으로 실시한 한국판 보스톤 이름대기 검사의 규준 자료를 이용하여 모형기반 접근을 적용한 규준추정 과정을 예시한다.

분석에 사용된 규준 자료는 서울신경심리검사(SNSB: 강연욱, 나덕렬, 2003)의 개정판 제작을 위해 2009년부터 2011년 사이에 얻은 것이다. 표본 특성, 자료 수집 및 선별 절차 등에 대한 자세한 내용은 서울신경심리검사 2판(강연욱, 장승민, 나덕렬, 2012)을 참조하기 바란다.

한국판 보스톤 이름대기 검사(Korean version-Boston Naming Test: K-BNT, 김향희, 나덕렬, 1997)는 보스톤 이름대기 검사(Boston Naming Test: BNT, Kaplan, Goodglass, & Weintraub, 1983)를 한국인들에 적합한 문항으로 제작, 표준화한 검사로 원판과 마찬가지로 실어증이나 뇌졸중, 퇴행성 치매 등의 신경과적 질환을 지닌 다양한 환자들에서 발견되는 이름대기장애(naming difficulty)를 평가하기 위하여 널리 사용되고 있다. K-BNT는 문항난이도에 따라 배열된 60개의 흑백 그림을 보고 그림의 이름을 말하도록 하며, 항목 당 1점씩의 점수를 모두 더해 0점에서 60점 사이의 총점을 검사점수로 사용한다.

K-BNT의 점수는 나이와 교육수준과 높은 관련이 있지만 성별과는 관련이 없는 것으로 알려져 있다(강연욱, 나덕렬, 2003; 김향희, 나덕렬, 1999). 따라서 본 분석에서는 성별을 고려하지 않고 연령대와 교육수준에 대해서만 규준집단을 구분하여 규준을 추정하였다. 연령집단은 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).[8] 교육수준은 교육 연수에 따라 "문맹", "무학이나 문맹이 아님～3년", "4～6년", "7～9년", "10～12년", "13～16년", "17년 이상"의 총 7개의 범주로 구분하였다. 나이와 교육수준에 따른 규준표본의 분포가 표 1에 제시되어 있다.

8) 마지막 연령집단은 85～90세로 연령 범위가 6년이었다.

예비분석

먼저 검사점수의 분포가 정규분포를 따른다고 가정할 수 있는지에 대해 사전 점검을 실시하였다. 하위집단의 검사점수의 분포가 정규분포를 이루더라도 전체 표본의 분포는 정규분포를 이루지 않을 수 있기 때문에 검사점수의 분포를 하위집단별로 확인하였다.[9] 표 1에 나타난 바와 같이 총 63개의 범주(하위집단) 중 58개의 범주가 하나 이상의 표본을 갖고 있었다. 일부 집단(예, 교육연한이 0에서 3년인 55세에서 59세 사이의 집단)의 Q-Q plot

9) 모형 기반 절차에서 가정하는 정규성은 각 하위집단에 대한 것이다. 그러나 회귀분석에서 오차의 정규성 검증은 하위 집단 간 분산 동일성의 여부와 상관없이 모형을 구한 후 얻어진 스튜던트 잔차(Studentized residual)를 이용해 모든 집단의 자료에 대해 한 번 수행된다. 이는 모든 집단에서 스튜던트 잔차가 평균 0, 표준편차 1을 갖는 정규분포를 따를 것으로 기대되기 때문이다. 정규성 가정은 모형을 구한 후 얻은 잔차를 이용해 확인하지만 여기에서는 예비분석 단계에서의 확인절차를 포함하였다. 실제 최종 모형에서 얻어진 스튜던트 잔차의 Q-Q plot은 약한 부적 편포를 나타냈고 Kolmigorov-Smirnov 검증 결과 정규분포를 유의미하게(p=.01) 벗어난 것으로 확인되었으나 부적 편포의 정도가 크지 않고(skewness= - .56) 적합도 검증에 사용된 표본수(n=1067)가 매우 크다는 점을 고려할 때 정규성을 가정한 모형 기반 규준추정 절차의 사용에 큰 문제가 없다고 판단하였다.

표 1. 연령과 교육수준에 따른 규준표본의 분포

연령	교육수준							합계
	문맹	0-3년	4-6년	7-9년	10-12년	13-16년	17년이상	
45-49세			12	12	25	18	4	71
50-54세			23	24	34	21	6	108
55-59세		9	27	21	27	14	9	107
60-64세	7	12	31	28	24	23	2	127
65-69세	14	21	29	23	27	19	2	135
70-74세	15	27	32	21	20	18	4	137
75-79세	23	23	34	23	33	13	7	156
80-84세	20	21	27	20	32	13	2	135
85-90세	14	17	21	18	11	8	2	91
합계	93	130	236	190	233	147	38	1067

에서 약한 부적 편포가 나타났으나 대부분의 하위집단의 Q-Q plot은 정규성을 지지하는 형태를 보여 모집 검사점수의 정규성을 가정한 모형 기반 규준추정 방법이 적용될 수 있다고 판단하였다.

다음으로 연령과 교육수준이 검사점수의 분포와 어떤 체계적인 관련을 갖는지를 살펴보기 위하여 연령집단별, 학력집단별 평균과 표준편차를 확인하였다. 그림 3은 연령집단에 따른 검사점수 평균과 표준편차를 나타낸 것이다. 검사점수의 평균은 연령이 높을수록 점진적으로 낮아지는 반면 표준편차는 점진적으로 증가함을 확인할 수 있다. 평균은 연령대가 높아질수록 일정하게 감소하는 것으로 보였다. 표준편차는 연령대가 높아질수록 증가의 폭이 점점 작게 나타났다. 그림 4는 학력집단에 따른 검사점수 평균과 표준편차를 나타낸다. 학력수준이 높을수록 평균은 높게 나타났으며 학력집단 간 차이의 폭은 학력이 높아질수록 점점 작아졌다. 표준편차의 경우 학력이 높아질수록 작아졌으며 감소의 폭이 점점 줄어들었다.

그림 3과 그림 4는 K-BNT 검사점수의 평균과 표준편차가 연령과 학력에 의해 체계적으로 영향을 받는다는 것을 보여준다. 검사점수의 평균은 연령이 증가할수록 선형적 감소 또는 약한 수준의 차이의 가속(2차 변화)을 보이며, 학력이 증가할수록 급격히 증가하다가 증가폭이 서서히 완만해지는 로그함수적 증가를 보인다. 검사점수의 표준편차는 연령의 증가에 따라 로그 증가를, 학력의 증가에 따라 로그 감소를 보여준다고 할 수 있다.

분석 및 결과

예비분석 결과를 바탕으로 연령대(x_1)와 교육수준(x_2) 및 둘의 상호작용 항을 예측변인으로 하는 다음과 같은 회귀모형을 구성하였다.

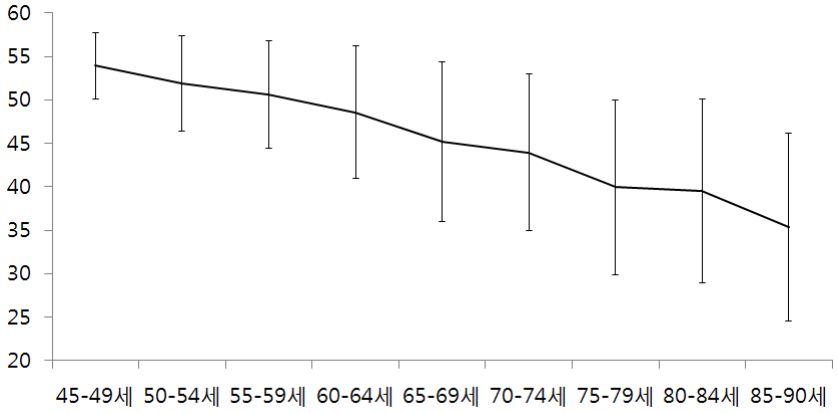

그림 3. K-BNT의 연령집단별 평균과 표준편차

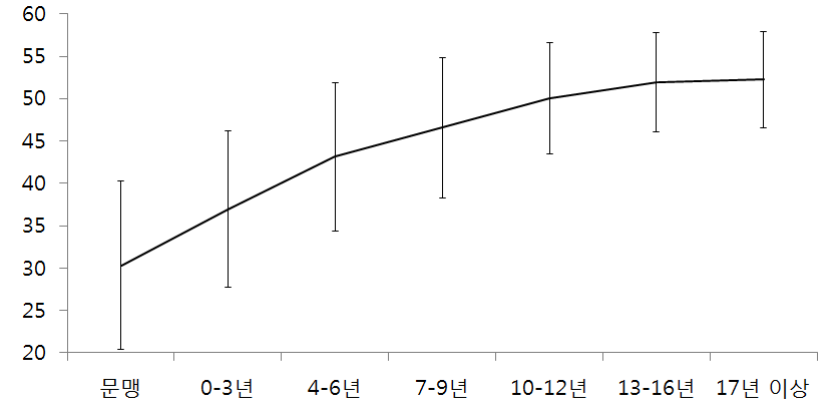

그림 4. K-BNT의 학력집단별 평균과 표준편차

$$y=\beta_0+\beta_1 x_1^*+\beta_2 x_2^*+\beta_3 x_1^* x_2^*+\varepsilon_i$$

연령변인 x_1에는 연령집단에 따라 "45-49세" 집단부터 "85-90세" 집단까지 1에서 9까지의 숫자를 부여하였다. 학력변인 x_2에는 학력집단에 따라 "문맹" 집단부터 "17년 이상" 집단까지 1에서 7까지의 숫자를 부여하였다. 자료에 가장 잘 부합하는 x_1^*과 x_2^*의 형태를 결정하기 위해 x_1^*의 자리에는 x_1과 x_1^2을, x_2^*의 자리에는 x_2와 $\ln(x_2)$를 번갈아가며 넣어 모형을 분석하였으며 그 결과 모형적합도가 가장 좋은 쌍으로 x_1^2과 $\ln(x_2)$가 선정되었다.[10] 최종적으로 선정된 평균의 회귀모형은 다음과 같다.

$$y=\beta_0+\beta_1 x_1^2+\beta_2 \ln(x_2)+\beta_3 x_1^2 \ln(x_2)+\varepsilon_i$$

분산의 모형화는 평균 회귀모형과 동시에 이루어졌으며 분산의 차이를 설명하는 공변인으로 평균과 마찬가지로 연령대(x_1)와 교육수준(x_2) 및 둘의 상호작용 항이 사용되었다. 최종

10) 모형 간의 적합도는 로그우도(log likelihood)와 AIC를 이용하여 비교하였다.

적으로 사용된 분산함수의 식은 다음과 같다.

$$\sigma^2 g(x_1, x_2) = \sigma^2 \exp(\theta_1 x_1 + \theta_2 x_2 + \theta_3 x_1 x_2)$$

평균 회귀모형과 분산함수모형을 동시에 추정하기 위해서 SAS 프로그램의 MIXED 절차를 사용하였다. 표 2는 최종적으로 선정된 평균과 분산의 모형 추정값이다.

연령과 교육수준은 상호작용을 포함하여 모두 유의미하게 K-BNT 검사점수 평균에 영향을 미쳤으며 분산함수의 경우도 교육수준효과와 상호작용효과가 유의미했다. 그림 5는 모형 기반 접근을 통해 추정된 표 2의 모수 추정치를 바탕으로 계산된 각 집단별 K-BNT의 추정평균을 나타내며 그림 6은 각 집단별 표본평균을 나타낸다.

일반적인 규준추정 절차는 그림 6에 나타난 표본평균 값을 각 집단에 대한 규준으로 사용하는 반면 모형 기반 접근은 그림 5의 추정평균을 규준으로 사용한다. 이 두 그림의 비교는 모형 기반 규준추정의 장점을 잘 드러내준다. 그림 6을 보면 연령과 교육수준에 따른 표본평균 차이가 전체적으로는 규칙성을 보이고 있지만 개별 하위집단의 평균이 이러한 규칙성에서 벗어나 있는 것을 확인할 수 있다. 예를 들어 교육 연수가 17년 이상이면서 60~64세에 해당하는 집단(n=2)은 같은 연령대의 다른 교육수준 집단들에 비해서 매우 높은 평균을 보이고 있는데 반해 교육 연수 17년 이상이면서 70~74세에 해당하는 집단(n=4)의 평균은 교육 연수 "10~12년"과 "13~16년"인 같은 연령대의 두 집단의 평균보다 낮게 나타났다. 개별 하위집단의 평균들이 전반적 평균 차이의 기대된 형태와 완전히 일치하지 않고 그 궤적에서 벗어나기도 하는 이유는 규준화의 대상인 모집단의 특성이 일정한 규칙성에서 벗어나 있기 때문일 수도 있지만, 그보다는 적은 표본수로부터 추정되는 과정에서 발생한 표집오차 때문일 가능성이 더 큰 것으로 판단된다. 따라서 이 경우 각 하위집단별로

표 2. K-BNT 점수의 평균 및 분산 모형의 모수 추정값

변인	추정값	표준오차	검증통계량
		평균함수	
절편	40.35	1.23	32.84***
log(교육수준)	8.25	0.77	10.70***
연령대^2	-0.23	0.03	-8.26***
log(교육수준)*연령대^2	0.05	0.02	2.70**
		분산함수	
기본분산	108.9	37.35	2.92**
교육수준	-0.43	0.07	-5.90***
연령대	-0.01	0.05	-0.18
교육수준*연령대	0.04	0.01	3.31***

주. 검증통계량은 평균함수의 경우 t를 분산함수의 경우 z를 의미. **p<.01. ***p<.001

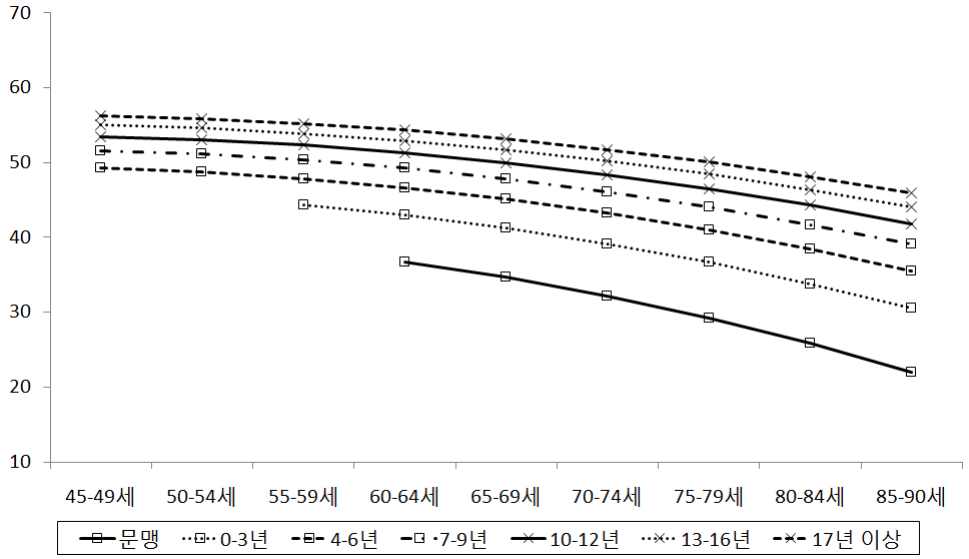

그림 5. 연령과 교육수준에 따른 집단별 K-BNT 추정평균

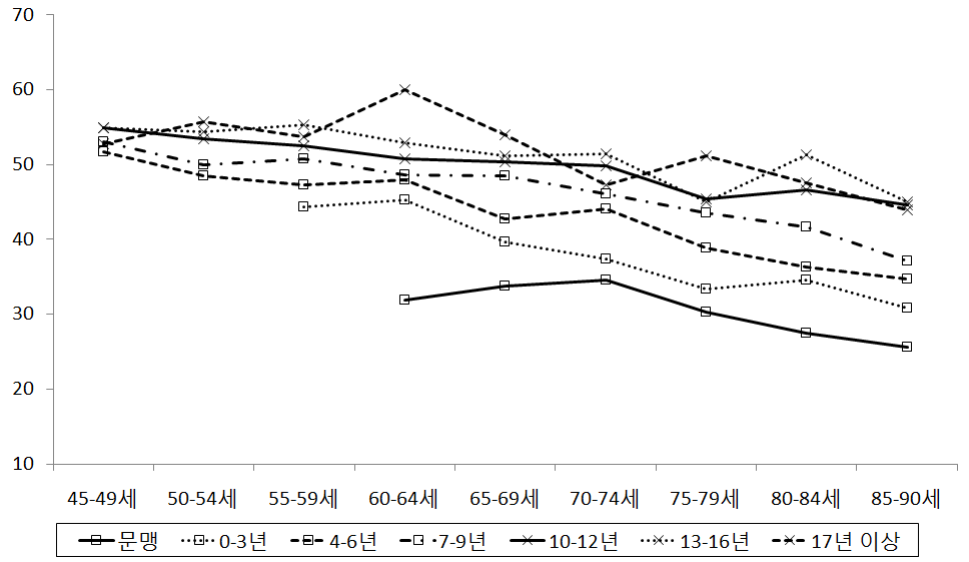

그림 6. 연령과 교육수준에 따른 집단별 K-BNT 표본평균

적은 수의 표본에서 얻어진 평균값들(그림 6)보다는 표본 전체를 이용하여 전반적인 평균 차이의 형태를 추정한 결과(그림 5)가 모집분포의 특성을 더 정확히 반영한다고 보는 것이 타당할 것이다. 그림 5를 보면 집단별 추정평균의 차이가 집단별 표본평균 차이의 전반적인 형태를 잘 드러내면서도 각 하위집단에서 표집오차로 인해 나타난 불규칙성이 모형을

통해 적절히 조정되었음을 볼 수 있다. 특히 연령이 높을수록 낮고 학력이 높을수록 높은, 그림 3과 그림 4에서 나타난 집단 간 평균 차이의 특성이 잘 드러나 있다. 또한 그림 5에는 연령과 교육수준의 상호작용(표 2 참조)으로 인해 학력이 낮을수록 연령증가에 따른 K-BNT 평균점수의 감소가 더 크다는 것이 추가적으로 드러나 있다. 그림 5에 나타난 집단

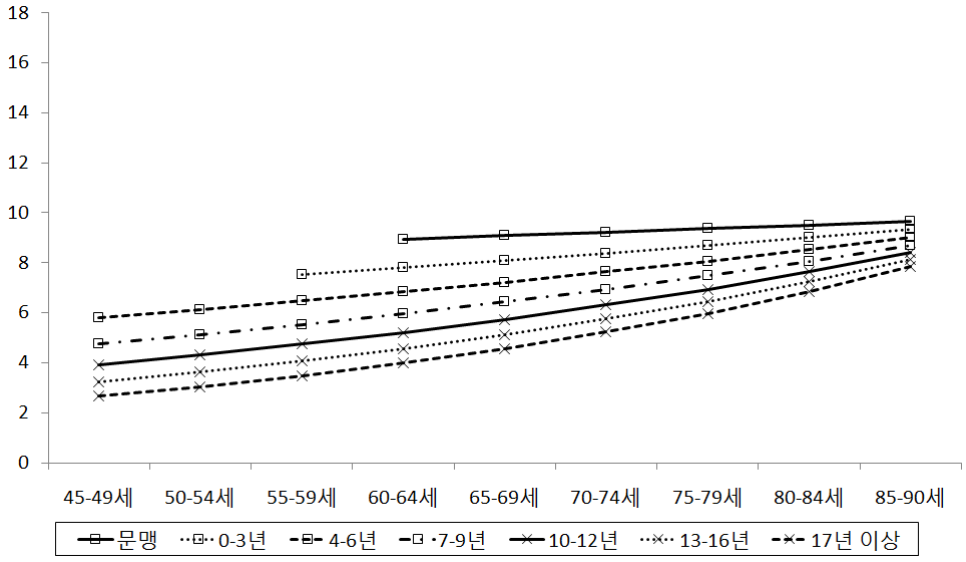

그림 7. 연령과 교육수준에 따른 집단별 K-BNT 추정 표준편차

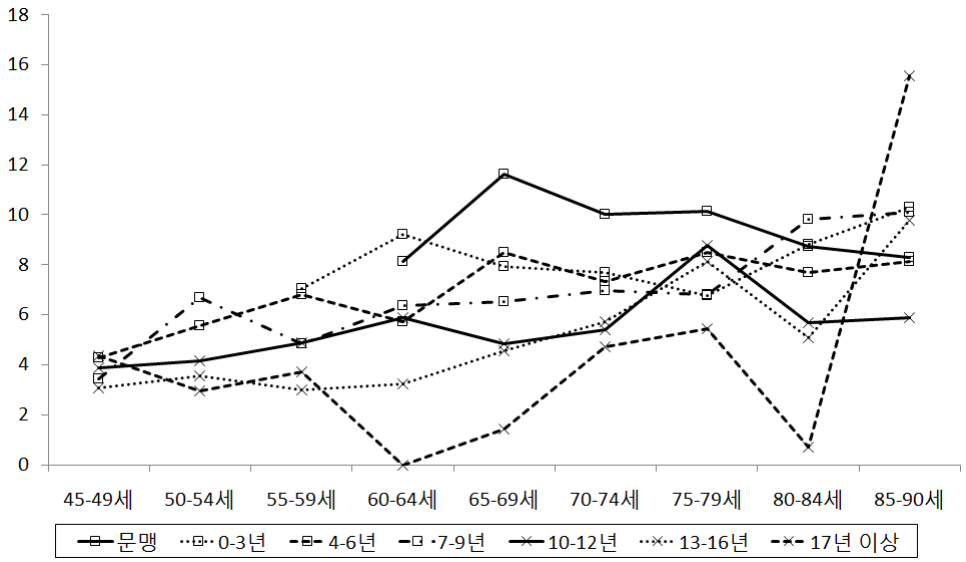

그림 8. 연령과 교육수준에 따른 집단별 K-BNT 표본표준편차

별 추정평균의 차이의 체계성은 연령의 증가나 교육수준의 차이가 K-BNT 검사점수에 미치는 영향에 대해 적절한 이론적 설명을 가능하게 하며, 피검자의 인구통계학적 특성을 고려한 검사점수를 더 의미 있게 해석하고 적용하는 것을 가능하게 해 준다.

그림 7은 모형에 의해 추정된 집단별 표준편차를, 그림 8은 집단별 표본표준편차를 나타낸 것이다. 앞서 연령과 교육수준을 따로 살펴본 그림 3과 그림 4에서는 K-BNT 점수의 표준편차가 연령과 교육수준에 의해 체계적으로 영향을 받는다는 것이 분명하게 드러났다. 이에 반해 각 학력집단에 대해 연령집단을 구분하여 본 그림 8에서는 이 체계성이 분명히 드러나지 않고 불규칙한 형태를 보인다. 특히 교육수준이 17년 이상인 집단의 경우 그 불규칙성이 매우 심하게 나타나 있다. 앞서 언급한대로 이 불규칙성은 모집 분포의 특성이라기보다는 적은 수의 표본에 기인한 증가된 표집오차 때문으로 보아야 할 것이다. 반면 그림 7에서 보는 바와 같이, 모형에 의해 추정된 표준편차의 차이는 그림 3과 그림 4에서 볼 수 있었던, 연령이 높을수록 커지고 학력이 높을수록 작아지는 K-BNT 표준편차의 특성이 잘 반영되어 나타나 있다.

또한 그림 7에서는 연령과 교육수준의 상호작용으로 인해 학력이 낮을수록 연령증가에 따른 표준편차의 증가가 완만해진다는 것도 함께 확인할 수 있다.

논 의

심리검사의 규준은 검사점수의 해석, 특히 정상성(다른 사람들과 유사하다 또는 평균에 가깝다)과 이상성(다른 사람들과 다르다 또는 평균에서 멀다) 평가의 객관적 근거로 사용되기 때문에 가능한 정확하게 추정되어야 한다. 특히, 임상장면에서 많이 사용되는 성격검사나 신경심리검사와 같은, 잠재적인 환자군을 사전에 선별하기 위한 심리검사들의 규준을 정확히 마련하는 것은 정상적 범주에 있는 사람들을 위험군으로 평가하여 불필요한 비용을 낭비하거나 위험군의 사람들을 정상 범주로 평가하여 적절한 치료적 개입을 놓치는 등의 잘못된 임상적 판단을 줄이기 위해 매우 중요하다. 그러나 많은 심리검사들이 검사 규준의 정확성과 유용성에 대해서 의심받고 있으며 학문적 논란에 놓이기도 한다.[11)]

규준은 모집에 대한 기술이지만 표본에 의해서 추정된다는 점에서, 좋은 표본의 확보는 객관적이고 신뢰할 수 있는 규준개발의 가장 중요한 요소이다. 좋은 표본이란 모집의 특성을 잘 대표하는 표본을 의미하며, 표본의 모집 대표성은 유층표집과 같은 잘 설계된 표집절차와 충분한 표본크기에 의해 결정된다. 반면 좋은 표본의 확보를 어렵게 하는 현실적인 걸림돌은 시간과 비용의 한계이다. 검사점수의 해석을 위해 규준을 필요로 하는 심리검사들은 나날이 늘어나는데 반해 각 검사들의 규준개발에 필요한 시간과 비용에 대한 사회적 지원은 부족한 실정이다. 결과적으로 규준개발 연구자들은 주어진 시간과 비용의 한계 내에서 규준개발을 진행하게 되고 이렇게 개발된 규준은 모집의 특성을 정확하게 반영하지 못해 실제 검사점수의 해석에 도움을 주지 못하는 경우가 종종 발생한다.

11) 심리검사 규준에 대한 논란의 일례로 대표적인 지능검사인 K-WAIS의 규준을 둘러싼 김홍근(2004)과 황순택(2006)의 논의를 들 수 있다.

그럼에도 불구하고, 규준추정을 위한 적절한 분석 방법을 사용함으로써 주어진 표본자료의 한계를 부분적으로 보완하는 것은 가능하다. 여기서 소개한 모형 기반 규준추정 절차는, 규준추정이 요구되는 하위집단이 많지만 하위집단 간 분포의 차이에 일정한 규칙성을 가정할 수 있을 때, 각 하위집단의 표본수가 충분히 확보되지 않은 경우에도 전형적인 규준추정 절차보다 상대적으로 안정적이고 신뢰할 수 있는 규준추정값을 제공한다.

모형 기반 규준추정 절차는 표본수와 관련한 이점 이외에도 여러 장점을 갖고 있다. 모형 기반 접근의 가장 큰 장점은 검사점수의 추정된 규준이 이론적으로 설명될 수 있는 체계적 모형과 부합한다는 점이다. 따라서 집단에 따라 상이한 규준을 적용하는 경우 왜 규준을 달리 적용해야 하는지에 대한 설명을 가능하게 한다. 연구자들과 검사 사용자들은 규준을 참조한 검사점수의 해석이 피검자 개인이 속한 집단에 따라 달라져야 하는 이유를 이해하고 설명할 수 있다. 또한 분포에 영향을 미치는 변인이 여럿일 때, 이 변인들을 동시에 고려한 규준을 추정할 수 있으며 나아가 변인들 간의 상호작용을 확인함으로써 검사점수와 설명변인들 간의 보다 역동적인 관련성을 확인할 수 있고 이러한 특성이 반영된 규준을 추정할 수 있다.

모형 기반 규준추정 절차의 여러 장점에도 불구하고 이 절차의 사용시에 유념해야 할 몇 가지 고려사항들이 있다. 먼저 모형 기반 규준추정 절차의 타당성은 일반적인 통계모형 분석에서와 마찬가지로 모형의 타당성에 의존한다. 규준추정 모형에서 모형의 타당성이란 검사점수와 모형에서 사용된 설명변인 간의 관련성(예, K-BNT 평균점수는 연령에 따라 달라지는가?)과 검사점수와 변인 사이에 설정된 함수 관계의 적절성(예, 연령 증가에 따른 K-BNT 평균점수 감소는 일차함수를 따르는가, 로그함수를 따르는가, 아니면 함수적 관계로 표현되지 않는가?)을 모두 포함한다. 모형의 타당성은 검사가 측정하는 심리적 구성개념과 설명변인(예, 연령)의 관계에 대한 이론적 기반, 선행연구로부터 뒷받침되는 경험적 근거, 그리고 규준추정을 위해 사용되는 주어진 표본자료와의 적합성 등을 고려하여 종합적으로 판단되어야 한다. 예를 들어, 어떤 심리적 특성이 특정 연령에서 급격한 증가나 감소를 보이거나 특정 연령을 기점으로 변화의 양상이 달라지는 등의 비연속적 변화를 갖는 경우 연속함수를 가정한 모형 기반 접근은 타당한 규준 추정값을 제공하지 못할 것이다. 이런 경우는 구간분할 회귀분석(piecewise regression analysis)등의 대안적 방법을 사용해야 할 것이다. 모형의 타당성이 의심스러운 경우에는 각 표본집단별로 얻은 규준추정치보다 모형 기반 접근의 결과를 더 신뢰할 근거가 없다.

모형 기반 접근법을 사용할 때 명심해야 하는 또 다른 주의사항은 지나치게 복잡한 모형은 피해야 한다는 것이다. 검사점수에 영향을 미칠 수 있을 것으로 의심해 볼 수 있는 인구통계학적 변인들은 다양하지만 하나의 회귀모형에 여러 개의 설명 변인이 포함되는 경우에는 추정된 모형이 실제 모집 모형의 적절한 반영이기 보다는 우연에 기댄 이득(capitalization on chance)의 결과일 가능성이 높아진다. 따라서 모형에 포함시키는 변인들은 검사점수와의 관련성이 이론적, 경험적으로 근거를 갖는 소수의 변인들로 제한되는 것이 필요하다. 또한 모형에 변인들 간의 상호작용을 포함할 때도 가능한 모든 상호작용을 전부

포함시키기 보다는 이론적으로 설명될 수 있는 타당한 것들로 제한하는 것이 바람직하다.

본 연구에서는 모형 기반 규준추정의 필요성을 설명하고 이 접근법이 각 규준집단에 대해 개별적으로 규준을 추정하는 일반적인 절차에 비해 갖는 이점을 강조하였다. 그러나 모형 기반 접근법이 규준추정의 결과에 영향을 미칠 수 있는 집단별 표본크기, 변인 간 함수의 형태, 모형의 복잡성 등의 조건에 따라 일반적인 규준추정 절차에 비해 얼마나 나은 수행을 보이는지, 또 어떤 조건에서 일반적인 절차에 비해 좋지 않은 수행을 보이는지에 대해서는 탐색되지 않았다. 이는 시뮬레이션 등을 이용한 후속 연구에서 다루어져야 할 것이다.

본 연구에서는 검사점수의 분포가 정규분포하는 경우의 모형 기반 규준추정에 대해서만 다루었다. 심리검사들의 점수 분포가 본질적으로 또는 경험적으로 정규분포를 따르지 않는 경우도 적지 않다. 검사점수의 분포가 정규분포를 따르지 않는 경우에는 분포의 특성이 평균과 표준편차만으로 기술될 수 없기 때문에 제10 백분위수와 같은 특정 순위값에 대한 추정 등 다른 방식의 규준추정 절차를 사용한다. 순위값을 사용하는 규준의 추정에도 모형에 기반한 절차를 적용할 수 있는지는 다른 연구에서 논의할 기회가 있을 것이다.

본 연구에서는 모형 기반 규준추정 절차가 어떻게 하위집단별 표본수가 충분치 않은 경우 규준추정 과정에 대한 표집오차의 영향을 감소시킬 수 있는가를 설명하였다. 그러나 앞서 언급한대로 객관적이고 신뢰할 수 있는 규준개발의 가장 중요한 요소는 모집 분포의 특성을 대표하는 충분한 크기의 표본이다. 설령 모형 기반 접근을 사용한다 하더라도 모집대표성이 떨어지고 표본크기가 작은 집단의 수가 많다면 추정된 규준의 신뢰도도 낮아질 수 밖에 없다. 결국 시간과 비용이 많이 들어가더라도 일정 수준 이상의 크기를 갖는 잘 표집된 표본을 얻는 것이야 말로 신뢰할 수 있는 규준개발의 핵심임은 아무리 강조해도 지나치지 않다. 다양한 종류의 심리검사가 병원, 군대, 기업, 관공서, 학교 등 다양한 장면에서 사람을 선별하고 진단하는데 사용되고 있다는 점에서, 그리고 검사점수의 올바른 적용을 위해 객관적이고 신뢰할 수 있는 검사 규준의 개발이 필수적이라는 점에서 심리검사 규준개발 연구에 연구자들의 더 많은 시간 투자와 정부나 기업체 등의 더 많은 재정적 지원이 요구된다. 모형 기반 규준추정 절차가 많은 표본을 확보하지 않아도 정확한 규준 추정이 가능하다는 근거로 사용되어서는 안 되겠지만, 이 방법이 시간과 비용에 대한 개인적, 사회적 투자가 부족하다는 현실적 제약을 일정 부분 보완할 수 있기를 기대한다.

참고문헌

강연욱 (2003). 아동용 언어기억검사(Seoul Verbal Learning Test-Children's version)의 제작과 표준화 연구. 한국심리학회지: 임상, 22, 435-448.

강연욱 (2006). K-MMSE(Korean-Mini Mental State Examination)의 노인 규준연구. 한국심리학회지: 일반, 25, 1-12.

강연욱, 나덕렬 (2003). 서울신경심리검사(SNSB). 인천: 휴브알엔씨.

강연욱, 박재설, 유경호, 이병철 (2009). 혈관성 인지장애 선별검사로서 Korean-Montreal

Cognitive Assessment(K-MoCA)의 신뢰도, 타당도 및 규준연구. 한국심리학회지: 임상, 28, 549-562.

강연욱, 장승민, 나덕렬 (2012). 서울신경심리검사 2판(SNSB-Ⅱ). 인천: 휴브알엔씨.

김향희, 나덕렬 (1997). 한국판 보스톤 이름대기 검사. 서울: 학지사.

김홍근 (2004). KWIS와 K-WAIS 중 어느 것을 사용할 것인가? 한국심리학회지: 임상, 23, 145-154.

서은현, 이동영, 추일한, 윤종철, 김기웅, 우종인 (2007). 벤톤 시각 기억 검사(Benton Visual Retention Test)의 한국 노인 정상규준연구. 한국심리학회지: 임상, 26, 745-763.

석정서, 최진영, 김호영 (2010). 한국판 치매평가 검사(K-DRS)의 2차 규준연구. 한국심리학회지: 임상, 29, 559-572.

송호정, 최진영 (2006). 한국 노인의 숫자폭 및 시공간폭 검사 표준화 연구. 한국심리학회지: 임상, 25, 505-532.

안효정, 최진영 (2004). 노인용 이야기 회상 검사의 표준화 연구. 한국심리학회지: 임상, 23, 435-454.

지연경, 조민경, 한지원, 김태희, 김기웅 (2011). Wisconsin Card Sorting Test-64 Card Version (WCST-64)의 한국 노인 정상규준연구. 한국심리학회지: 임상, 30, 1037-1046.

최진영, 박미선, 조비룡, 양동원, 김상윤 (2002). 전산화된 치매선별 검사(Computerized Dementia Screening Test: CDST)의 규준연구. 한국심리학회지: 임상, 21, 445-460.

황순택 (2006). K-WAIS는 타당한 지능검사인가?: K-WAIS와 KWIS의 규준 비교. 한국심리학회지: 임상, 25, 849-863.

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.

Burchinal, M., & Appelbaum, M. I. (1991). Estimating individual developmental functions: Methods and their assumptions. *Child Development, 62*, 23-41.

Carroll, R. J., & Ruppert, D. (1988). *Transformation and weighting in regression.* London: Chapman and Hall.

Crawford, J. R., & Garthwaite, P. H. (2002). Investigation of the single case in neuropsychology: Confidence limits on the abnormality of test scores and test score differences. *Neuropsychologia, 40*, 1196-1208.

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.

Davidian, M., & Carroll, R. J. (1987). Variance function estimation. *Journal of the American Statistical Association, 82*, 1079-1091.

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.

Jahng, S. (2008). *Multilevel models for intensive longitudinal data with heterogeneous error structure: Covariance transformation and variance function models*. Unpublished doctoral dissertation,

University of Missouri.

Kaplan, E. F., Goodglass, H., & Wintraub, S. (1983). *Boston Naming Test*. Philadelphia: Lea & Febiger.

Stern, R. A., & White, T. (2009). *Neurological Assessment Battery (NAB) visual discrimination test*. Lutz: Psychological Assessment Resources, Inc.

Taylor, R. (1998). Continuous norming: Improved equations for the WAIS-R. *British Journal of Clinical Psychology, 37*, 451-456.

Wechsler, D. (1981). *Manual for the Wechsler Adult Intelligence Scale-Revised.* New York: Psychological Corporation.

Zachary, R. A., & Gorsuch, R. L. (1985). Continuous norming: Implications for the WAIS-R. *Journal of Clinical Psychology, 41*, 86-94.

원고접수일 : 2012. 7. 18.
1차 수정 원고접수일 : 2012. 8. 20.
게재결정일 : 2012. 9. 27.

한국심리학회지 : 임상
The Korean Journal of Clinical Psychology
2012, Vol. 31, No. 4, 923-944

A model-based approach to estimating psychological test norms under normality assumption

Seungmin Jahng **Yeonwook Kang**

Department of Psychology, Hallym University

Reliable estimation of a psychological test norm requires a large number of samples that well represent characteristics of the population distribution. Samples used in many test norm development studies have been classified into several subgroups, according to age and years of education, for example, whose norms need to be estimated separately. However, in most of the studies, sample sizes of subgroups are not large enough for reliable estimation of norms of the psychological test under investigation. The current study explained why having a large enough sample size for each subgroup is important in test norm estimation and introduced a model-based estimation procedure that provides more reliable norm estimates than a typical norming procedure, especially when the sample size of each subgroup is less than desirable. Specifically, this procedure, under normality assumption of the test scores, uses a regression model that estimates the means and variances of all of the subgroups simultaneously. An example analysis was illustrated in order to demonstrate how to use the suggested procedure using a normative sample of Korean-version Boston Naming Test scores from 1067 normal elderlies. Finally, considerations in application of a model-based norming procedure were discussed.

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