Journal of the Ergonomics Society of Korea
Vol. 28, No. 4 pp.117-124, November 2009

3차원 인체치수 조사 자료의 품질 개선을 위한 연구

박선미[1] · 남윤자[2] · 박진우[3]

[1]아이오와주립대학교 의류학과 / [2]서울대학교 의류학과 / [3]수원대학교 통계정보학과

A Study for Quality Improvement of Three-dimensional Body Measurement Data

Sun Mi Park[1], Yun Ja Nam[2], Jin Woo Park[3]

[1]Department of Apparel, Educational Studies, and Hospitality Management, Iowa State University, Iowa, 50011
[2]Department of Clothing Textile, Seoul National University, Seoul, 151-742
[3]Department of Applied Statistics, University of Suwon, Whasung-Si, 445-743

ABSTRACT

To inspect the quality of data collected from a large-scale body measurement and investigation project, it is necessary to establish a proper data editing process. The three-dimensional body measurement may have measuring errors caused from measurer's proficiency or changes in the subject's posture. And it may also have errors caused in the process of algorithm expressing the information obtained from the three-dimensional scanner into numerical values, and in the course of data-processing dealing with numerous data for individuals. When those errors are found, the quality of the measured data is deteriorated, and they consequently reduce the quality of statistics which was conducted on the basis of it. Therefore this study intends to suggest a new way to improve the quality of the data collected from the three-dimensional body measurement by proposing a working procedure identifying data errors and correcting them from the whole data processing procedure-collecting, processing, and analyzing- of the 2004 Size Korea Three-dimensional Body Measurement Project. This study was carried out into three stages: Firstly, we detected erroneous data by examining of logical relations among variables under each edit rule. Secondly, we detected suspicious data through independent examination of individual variable value by sex and age. Finally, we examined scatter-plot matrix of many variables to consider the relationships among them. This simple graphical tool helps us to find out whether some suspicious data exist in the data set or not. As a result of this study, we detected some erroneous data included in the raw data. We figured out that the main errors are not because of the system errors that the three-dimensional body measurement system has but because of the subject's original three-dimensional shape data. Therefore by correcting some erroneous data, we have enhanced data quality.

Keyword: Data editing, Editing rule, 3D body measurement, Outlier, Scatter-plot Matrix

*본 연구는 2003-2004 산업자원부 기술표준원 주관 Size Korea 사업의 결과로 도출되었음.
교신저자: 남윤자
주 소: 135-535 서울시 강남구 도곡동 467-17 타워팰리스 E동 1103호, 전화: 010-5686-6844, E-mail: yunja@snu.ac.kr

1. 서 론

인체치수 조사 자료는 산업별 제품 설계 시 기초 자료가 되며, 의류, 신발, 자동차, 가구, 전자 업계 등 다양한 산업 분야에서 제품의 디자인을 위한 기본적인 자료로 활용되고 있다. 각 나라에서는 이러한 산업계의 인체치수 통계 자료의 요구를 수용하고 반영하기 위해 자국민을 대상으로 인체치수 조사사업을 실시하고 있는데, 우리나라에서는 정부 주도하에 1979년부터 약 6년마다 다섯 차례에 걸쳐 국민표준체위조사를 실시한 바 있다(산업자원부 기술표준원, 2004). 2004년에 실시된 '제 5차 한국인 인체치수 조사사업'은 Size Korea라는 사업명으로 실시되었으며, 직접측정방법에 의한 인체치수의 계측뿐 아니라 처음으로 3차원 인체측정이 병행 실시되었다. 3차원 인체측정은 3차원 전신 스캐너를 이용하여 인체의 표면 데이터를 획득함으로서 3차원 체형에 관한 데이터를 얻는 방법이다. 이렇게 얻어진 정보를 데이터베이스화하여 산업계에서 요구하는 각종 인체측정 관련 정보를 제공하게 된다(남윤자 외, 2004).

측정도구를 사용하여 인체치수 자료를 수집하는 경우, 측정 환경의 변화나 측정자의 숙련도 또는 피험자의 자세 변화 등에 의해 측정오차가 발생할 수 있다. '제 5차 한국인 인체치수 조사사업'의 3차원 인체측정에 사용된 전신 스캐너는 Cyberware사의 WB4이며, 현재 상용화된 인체 스캐너 중 가장 높은 신뢰도를 갖는 것으로 알려져 있다(산업자원부 기술표준원, 2002). 그러나, 3차원 인체측정은 시행한지 얼마되지 않고, 근본적으로 기존의 직접측정과는 다른 프로토콜에 의해 치수를 추출하기 때문에 발생하는 오차의 종류와 성격이 다르다(ISO 20686, 2005). 이러한 오차가 커질 경우 측정 자료의 품질에 영향을 미쳐 자료의 활용이 제한된다. 따라서 인체치수 자료의 품질과 유용성을 보장하기 위해서는 인체치수 데이터에 대한 체계적인 편집 과정이 필요하다. 데이터 편집의 측면에서 3차원 인체측정은 피험자의 3차원 형상 및 치수 자료를 동시에 데이터베이스화 할 수 있기 때문에 필요할 때 언제든지 재측정할 수 있고, 데이터의 검증을 용이하게 하여 신뢰성 높은 데이터를 제공할 수 있는 특징이 있다.

데이터 편집(data editing)이란 데이터 수집 및 처리, 분석 과정에서 데이터의 오류를 식별하여 찾아내고 이를 수정하는 일련의 과정을 뜻한다(Granquist, 1995). 오류가 내포된 데이터를 방치한 채 데이터를 분석할 경우 결과적으로 작성되는 통계의 품질 및 신뢰성의 손상을 가져오므로 통계 작성 이전에 수집된 데이터를 편집하는 것은 필수적인 작업이라고 할 수 있다(Granquist, 1995; Waal, T. and Quere, R, 2003). 그러나 우리나라의 경우 데이터 편집에 관한 논의나 연구가 상대적으로 매우 빈약한 편이며(박진우 외, 2005), 특히 대규모 인체치수 조사사업에서 수집된 자료를 대상으로 신뢰성 높은 자료를 제공하기 위한 데이터 편집에 관한 연구는 미흡한 실정이다.

따라서 본 연구에서는 '제 5차 한국인 인체치수 조사사업'에서 수집된 3차원 인체측정 자료를 대상으로 데이터의 수집 및 처리, 분석 과정에서 데이터의 오류를 식별하고 이를 수정하는 체계적인 작업 과정을 제시하여 대규모 3차원 인체치수 데이터 통계의 품질 및 신뢰성을 향상시키고자 한다. 또한 데이터 편집에 관한 기존의 연구들이 대부분 질적 변수나 이산형 변수를 주로 다루고 연속형 변수인 경우 일변량 또는 이변량 정도에 국한하여 처리하였는데 본 연구의 경우 다변량 연속 변수에 관한 처리 과정을 보여주고 있으므로 데이터 편집에 대한 새로운 사례를 제공할 수 있다. 이는 추후 3차원 인체치수 조사사업의 데이터 편집을 위한 시스템 개발에 기초 자료로 활용될 수 있을 것이다.

2. 연구 방법

2.1 3차원 인체측정

'제 5차 한국인 인체치수 조사사업'의 3차원 인체측정은 2003년 6월부터 2004년 11월까지 총18개월에 걸쳐 전국 3개 권역 5개 지역(서울, 대전/광주, 대구/부산)에서 총 5,000명(8~75세)을 대상으로 실시되었다(남윤자 외, 2004). 본 연구는 그 중 서울, 대전 지역 3,118명(여자 1,570명, 남자 1,548명)의 데이터를 연구대상으로 하였다<표 1>.

표 1. 연구대상(단위: 명)

연령	여자	남자	계
8~10세	92	91	183
11~13세	123	122	245
14~16세	137	123	260
17~19세	120	121	241
20~29세	319	319	638
30~39세	316	310	626
40~49세	155	155	310
50~59세	152	151	303
60세 이상	156	156	312
계	1570	1548	3118

Size Korea 인체치수 조사사업을 위한 3차원 인체측정은 '산업기술기반조성에 관한 보고서(2002)'에 따른 프로토

콜에 근거하여 다음과 같은 과정으로 진행되었다. 피험자는 자신의 체형에 맞는 측정복 및 측정모를 착용한 후 총 66개의 랜드마크를 몸에 표시한다. 3차원 인체 스캔은 기본선 자세, 응용 선 자세, 응용 앉은 자세, 머리스캔 자세의 총 4자세로 실시되었으며, 인체치수측정 프로그램인 3DM을 사용하여 총 128개의 인체치수를 추출하였다. 3DM은 3차원 인체 스캔 데이터에서 치수를 측정하기 위해 개발된 인체치수 측정 전문 프로그램이다. 반자동 측정방식으로 운영되는데, 기준점은 수동으로 지정해주는 반면 측정 경로 설정 및 측정치 계산은 자동으로 이루어지며, 다양한 체형의 인체 형상에 대한 보다 정확한 치수측정을 위해서 환경 변수의 값을 조절 할 수 있도록 설계되어 있다(산업자원부 기술표준원, 2002). 3DM 소프트웨어의 사용법은 다음과 같다. 측정자는 3차원 스캐너로부터 스캔된 인체 형상을 불러온 후, 자세를 결정하고, 각 자세별로 정해진 측정 기준점을 Scene 화면을 보면서 지정한다. 자세별 랜드마크의 지정이 완료되면 미리 정해진 알고리즘에 의해 각 자세별 측정 항목이 자동으로 계산된다.

2.2 데이터 편집규칙

대규모 인체치수 조사사업에서 입력이 완료된 데이터들의 품질을 점검하기 위해서는 적절한 편집규칙 과정이 필요하다. 본 연구에서는 '제 5차 한국인 인체치수 조사사업'에서 수집된 3차원 인체측정 데이터의 신뢰도를 높이기 위해 데이터 편집의 과정을 크게 두 단계로 나누어 진행하였다. 첫째 단계는 측정된 변수들 사이의 논리적 관계를 점검함으로써 오류 데이터의 존재 여부를 파악하는 논리적 편집의 단계이다. 논리적 편집을 위해서는 사전에 적절한 논리적 편집규칙들(logical edit rules)을 마련하는 것이 필요하다. 두 번째 단계는 논리적 편집을 마친 데이터 내에 통계적 이상치(outlier)의 존재 여부를 파악하는 통계적 데이터 편집(statistical data editing)의 단계이다. 통계적 데이터 편집은 다시 두 단계를 거친다. 먼저, 각각의 변수들에 대해 독립적으로 일변량 이상치의 여부를 점검하고, 다음 단계는 산점도 행렬(scatter-plot matrix)을 통해 여러 변수들 사이의 관계 측면에서 이상치 존재 여부를 확인한다. 통계적 데이터 편집을 위해서는 SAS와 R을 활용하였다.

2.2.1 논리적 편집규칙

논리적 편집이란 변수들 간의 논리적인 관계를 검토하여 데이터의 오류 여부를 판정하는 과정으로 이를 위해 사전에 적절한 편집규칙을 마련하는 것이 필요하다. 본 연구에서는 측정 항목들 간의 논리적 관계를 검토하기 위해 <표 2>에 나온 바와 같은 총 54개의 논리적 편집규칙을 마련하여 데이터를 점검하였다. 오류로 판정된 데이터의 경우 3차원 형상자료를 확인하여 오류 발생의 원인 및 재측정 여부를 판단하게 된다.

표 2. 논리적 편집규칙

구분	번호	논리적 편집규칙	구분	번호	논리적 편집규칙
높이 (25)	1	키 > 눈높이	높이 (25)	18	위앞엉덩뼈가시높이 > 넙다리높이
	2	눈높이 > 어깨가쪽높이		19	샅높이 > 다리벌린 샅높이
	3	어깨가쪽높이 > 어깨높이		20	손귄끝높이 > 손끝높이
	4	어깨높이 > 겨드랑높이		21	앉은키 > 앉은눈높이
	5	겨드랑높이 > 젖가슴높이		22	앉은눈높이 > 앉은목뒤높이
	6	젖가슴높이 > 젖가슴아래높이		23	앉은목뒤높이 > 앉은어깨높이
	7	젖가슴아래높이 > 허리높이		24	앉은어깨높이 > 앉은팔꿈치높이
	8	허리높이 > 배꼽수준허리높이		25	앉은무릎높이 > 앉은오금높이
	9	배꼽수준허리높이 > 엉덩이높이	둘레 (4)	26	목밑둘레 > 목둘레
	10	엉덩이높이 > 샅높이		27	목밑둘레 > 목뒤둘레
	11	볼기고랑높이 > 무릎뼈가운데높이		28	발목최대둘레 > 종아리최소둘레
	12	무릎뼈가운데높이 > 무릎높이		29	젖가슴둘레 > 젖가슴아래둘레 (여자)
	13	무릎높이 > 장딴지높이	길이 (12)	30	배꼽수준앞중심길이 > 앞중심길이
	14	장딴지높이 > 가쪽복사높이		31	목뒤어깨가쪽길이 > 어깨길이
	15	목뒤높이 > 목앞높이		32	어깨가쪽사이길이 > 어깨가쪽사이수평길이
	16	허리높이 > 엉덩뼈능선높이		33	다리가쪽길이 > 허리옆가쪽복사길이
	17	엉덩뼈능선높이 > 위앞엉덩뼈가시높이		34	총길이 > 목뒤오금길이

표 2. 논리적 편집규칙(계속)

구분	번호	논리적 편집규칙	구분	번호	논리적 편집규칙
길이 (12)	35	목뒤오금길이 > 배꼽수준등길이	두께 (4)	45	앉은엉덩이배두께 > 앉은배두께
	36	배꼽수준등길이 > 등길이		46	벽면몸통두께 > 엉덩이두께
	37	등길이 > 목뒤등뼈위겨드랑수준길이		47	벽면몸통두께 > 젖가슴두께
	38	목뒤손목안쪽길이 > 팔길이		48	젖가슴두께 > 젖가슴아래두께 (여자)
	39	팔길이 > 팔안쪽길이	계산식 (6)	49	허리높이−샅높이=엉덩이수직길이
	40	팔꿈치손끝수평길이 > 아래팔수평길이		50	목뒤높이−샅높이=몸통수직길이
	41	앉은엉덩이무릎수평길이 > 앉은엉덩이오금수평길이		51	목옆허리둘레선길이−목옆젖꼭지길이= 젖꼭지허리둘레선길이
너비 (3)	42	위팔사이너비 > 어깨너비		52	목뒤젖꼭지허리둘레선길이−목뒤젖꼭지길이= 젖꼭지허리둘레선길이
	43	어깨너비 > 가슴너비		53	팔길이−위팔길이=아래팔길이
	44	가슴너비 > 목밑너비		54	팔굽힌목뒤손목안쪽길이−목뒤어깨가쪽길이=굽힌팔길이

<표 2>의 논리적 편집규칙 중 번호 1의 경우를 예로 들어보면, 개인의 눈높이가 키보다 클 수 없으므로 각 조사 대상자의 키와 눈높이 값을 서로 비교하여 이 규칙을 위반하는 경우 오류로 판정하게 된다.

2.2.2 통계적 편집규칙

일변량(univariate) 편집규칙

일변량 편집이란 개별 변수들을 독립적으로 생각하여 각 변수별로 이상치가 존재하는지의 여부를 파악하는 과정이다. 본 연구에서는 이상치 여부를 파악하기 위해 품질관리 분야의 관리도에서 널리 사용하는 3σ법을 사용하였다. 즉, 성별, 연령대별 표본 데이터를 이용하여 각 변수별로 평균과 표준편차를 계산한 후, 평균을 중심으로 $\pm 3\sigma$ 이상을 벗어난 값일 경우 이상치로 의심하여 검출하도록 하였다. 이상치로 검출된 데이터의 경우 피험자의 3차원 형상 및 치수 데이터로부터 오류 여부를 확인하여 오류가 명확한 경우 재측정을 실시하였다. 한편 이상치로 검출되긴 했지만 오류가 아닌 것으로 확인되면 원래 관측 데이터를 그대로 유지시킨다.

다변량(multivariate) 편집규칙

인체치수 조사에서 측정되는 여러 변수들은 독립적이지 않고 서로 밀접한 상관관계를 지니는 것이 일반적이다. 그러므로 일변량 편집에서는 이상이 없는 것으로 판단되지만 변수들 사이의 관계를 감안하면 오류인 데이터가 존재할 수 있다. 예를 들어 어느 피험자의 키가 185cm, 몸무게 45kg으로 관측되었을 경우, 이 데이터는 키나 몸무게 각각의 변수에 대한 일변량 편집에서는 이상치가 아닌 것으로 나타난다. 하지만 두 변수의 관계를 고려한다면 이상치로 파악된다. 따라서 본 연구에서는 일변량 편집을 거친 데이터에 대해 추가적인 편집을 실시하였다.

수많은 양적 변수들 사이의 상관관계를 파악하기 위해 가장 간편한 통계적 그래프로는 산점도 행렬(scatter−plot matrix)이 대표적이다. 산점도는 두 변수를 각각 X−Y 좌표 상에 표시하는 그래프이며, 산점도 행렬은 여러 변수들을 각각 둘씩 짝을 지어 작성한 산점도를 행렬 형식으로 나타낸 그래프이다. 다음의 <그림 1>은 4개 변수들의 산점도 행렬의 예를 나타내고 있다. 변수들 사이의 연관성을 다각적으로 검토하여 반복적으로 약한 상관관계를 보이고 있는 점들을 검출하였다. 이상치로 의심되는 점에 대해서는 원 자료를 확인하여 오류 여부를 확인하도록 하였다.

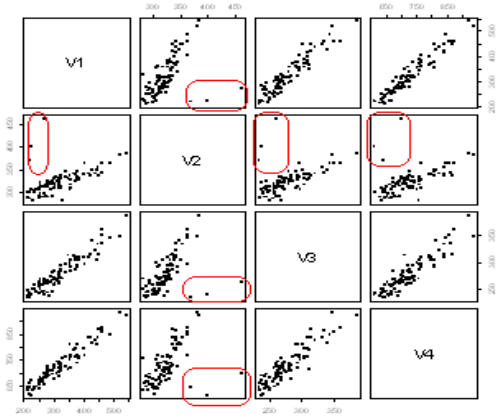

그림 1. 4개 변수들 사이의 산점도 행렬

'R'을 활용하여 <표 3>에서 제시한 총 16개의 주요 변수들을 대상으로 상호 연관성을 살펴보았다. 통계분석 프로그램 'R'은 두 변수들 사이의 관계를 산점도 행렬로 표현할 수

있어 변수들 간의 연관성에 의한 이상치를 살피는데 매우 효과적이다.

표 3. 다변량 편집규칙에 사용된 주요 항목

키, 목뒤높이, 샅높이, 목밑둘레, 겨드랑둘레, 가슴둘레, 젖가슴둘레, 젖가슴아래둘레, 엉덩이둘레, 허리둘레, 배둘레, 등길이, 발직선길이, 팔길이, 어깨가쪽사이길이, 몸무게

3. 연구 결과

3.1 논리적 편집규칙 적용 결과

<표 2>에서 제시한 54개의 논리적 편집규칙을 사용하여 성별, 연령별로 데이터를 분석한 결과 총 22개의 논리적 편집규칙에서 오류가 발생하였다. 논리적 편집규칙 적용 결과 나타난 오류의 원인은 크게 랜드마크 입력 오류, 체형에 의한 요인, 측정 자세에 의한 요인 등으로 나눌 수 있었다 <표 4>.

<표 4>의 랜드마크 입력 오류 중 관계식 1번에서 가장 많은 오류가 발생하고 있는데, 그 원인은 엉덩이수직길이는 허리뒤점을 기준으로 하고 허리높이는 허리옆점을 기준으로 치수를 추출하고 있기 때문으로 파악된다. 즉, 허리뒤점과 허리옆점은 기본적으로 같은 높이에 위치해야 하지만, 3차원 스캔을 실시하기 전 피험자의 몸에 랜드마크를 부착하는 과정에서 허리옆점의 수준을 허리뒤점으로 옮길 때 오차가 발생한 것이다. 이 경우 피험자의 3차원 형상을 확인해 보아 그 차이가 20mm가 넘어가는 경우에 허리뒤점의 높이를 허리옆점의 수준으로 수정하여 주었다. 관계식 2번의 경우를 살펴보면, 대부분의 피험자는 어깨점보다 어깨가쪽점이 더 높지만 그 차이가 미세하기 때문에 오류가 발생한 것으로 파악된다. 이 경우 피험자의 3차원 형상을 확인해 어깨점과 어깨가쪽점의 랜드마크 위치의 정확성을 확인한 후 수정해주었다. 그 외 나머지 관계식에서 발생한 오류는 모두 재측정하여 수정하여 주었다.

체형 요인은 표준에서 벗어나는 특이 체형인 경우에 많이 발생하였다. 특히 목밑둘레 · 목둘레와 관련된 관계식 15의 경우 비만하거나 목이 짧은 체형에서 많이 발생하였으며, 관

표 4. 논리적 편집규칙 적용 결과 오류 원인 분석(단위: 명)

분류	번호	논리적 편집규칙	여자	남자
랜드마크 입력오류	1	허리높이−샅높이=엉덩이수직길이	1483	1451
	2	어깨가쪽높이 > 어깨높이	413	425
	3	겨드랑높이 > 젖가슴높이	3	1
	4	젖가슴높이 > 젖가슴아래높이	13	6
	5	목뒤높이 > 목앞높이	3	1
	6	허리높이 > 엉덩뼈능선높이	5	26
	7	엉덩뼈능선높이 > 위앞엉덩뼈가시높이	1	1
	8	위앞엉덩뼈가시높이 > 넙다리높이	49	10
	9	손귄끝높이 > 손끝높이	1	0
	10	넙다리둘레 > 넙다리중간둘레	4	0
	11	젖가슴둘레(여) > 젖가슴아래둘레(여)	1	0
	12	등길이 > 목뒤등뼈위겨드랑수준길이	1	0
	13	위팔사이너비 > 어깨너비	1	0
	14	무릎뼈가운데높이 > 무릎높이	433	319
체형 요인	15	목밑둘레 > 목둘레	16	14
	16	젖가슴두께(여) > 젖가슴아래두께(여)	30	−
	17	허리높이 > 배꼽수준허리높이	0	7
	18	엉덩이높이 > 샅높이	0	1
	19	어깨너비 > 가슴너비	132	81
측정자세오류	20	샅높이 > 다리벌린샅높이	185	136
	21	팔굽힌목뒤손목안쪽길이−목뒤어깨가쪽길이=굽힌팔길이	1441	1452
	22	앉은엉덩이배두께 > 앉은배두께	306	537

계식 16번 젖가슴두께와 젖가슴아래두께의 경우 주로 20세 이하의 여성에서 많이 발생하였다. 이 경우에도 피험자의 3차원 형상 데이터로부터 오류 여부를 확인한 후 오류가 명확하다고 판단되는 경우 재측정을 실시하였다. 그 외 나머지 관계식에서 발생한 오류는 모두 재측정하여 수정하여 주었다.

측정 자세 요인은 스캔 시 피험자의 자세에 의해 발생한 오차이다. 관계식 20번의 샅높이와 다리벌린샅높이의 경우 피험자의 측정 자세 변동 및 측정복 위에 부착된 랜드마크의 유동성 때문에 발생한 것으로 확인되었으며, 관계식 21번의 경우, 팔굽힌 목뒤손목안쪽길이와 굽힌 팔길이는 2자세에서 측정된 반면, 목뒤어깨가쪽길이는 3자세에서 측정이 되었기 때문에 발생한 오류로 확인되었다. 이 경우 형상 데이터의 오류 여부를 확인하여 오류가 명확한 경우 재측정하였다. 관계식 22번의 논리적 오류는 앉은자세에서 피험자의 무게 중심이 뒤로 기울게 되면서 앉은 배두께가 앉은 엉덩이 배두께보다 더 커지게 되면서 발생한 오류이다. 이 경우 형상 데이터로부터 오류의 원인만 확인하고, 치수 데이터의 수정은 하지 않았다.

3.2 통계적 편집규칙 적용 결과

3.2.1 일변량 데이터 편집 결과

성별, 연령대별 표본 데이터의 128개 변수값을 대상으로 $\pm 3\sigma$ 이상을 벗어난 값을 검출한 결과 총 3118명 중 461명(14.78%)이 검출되었으며, 이들의 3차원 형상 및 치수 데이터를 확인하여 측정 오류 여부를 확인하였다<표 5>.

특이점으로 검출된 461명의 오류 원인을 파악한 결과는 <표 6>과 같다.

3차원 형상 데이터로부터 인체치수를 추출하는 과정에는 이상이 없으나 피험자가 평균에서 많이 벗어난 특이 체형이기 때문에 검출된 305명의 경우에는 수정하지 않고 그대로 두었다.

표 6. 일변량 편집규칙 검증 결과 오류원인 분석

번호	오류원인	인원수	설 명
1	특이체형	305명	인체치수에는 이상이 없고 특이체형이라서 검출된 경우
2	측정오류	119명	① 랜드마크 입력 오류(61명) ② 측정경로 이상(48명) ③ 랜드마크 입력 누락(6명) ④ 기타(4명)
3	측정 경로 보완 필요	37명	편위팔둘레, 팔굽힌손목안쪽길이, 목밑둘레, 굽힌팔길이 등

측정 과정에 문제가 발생하여 재측정을 실시한 119명의 원인을 분석해 보면, 측정자의 랜드마크 입력 오류가 61명으로 가장 많았으며, 측정 경로에 문제가 있어 수정한 경우가 48명으로 그 뒤를 이었다. 랜드마크 입력 누락이 6명이었고, 기타의 경우로 저장된 랜드마크가 불러와지지 않아 새로 측정한 경우는 4명이 검출되었다.

반자동 측정 프로그램의 보완이 필요한 항목(37명)은 주로 스캔 결과 3차원 형상 데이터의 표면이 거칠어 측정 오류가 발생한 경우로 편위팔둘레, 팔굽힌손목안쪽길이, 목밑둘레, 굽힌팔길이 등의 항목에서 많이 발생하였다. 이 경우 3차원 스캔 데이터의 거친 표면을 부드럽게 처리해준 후 재측정을 실시하였다. 추후 3차원 스캔 데이터의 품질이 좋지 않은 경우를 대비하여 측정 프로그램의 보완이 필요한 것으로 파악된다.

3.2.2 다변량 데이터 편집 결과

본 연구에서는 주요 변수들 간의 연관성을 파악하여 특이

표 5. 연령대별 일변량 편집규칙 검증 결과(단위: 명)

결과	연령 / 분류	8~10세	11~13세	14~16세	17~19세	20대	30대	40대	50대	60대	총합
측정 인원	여자	92	123	137	120	319	316	155	152	156	1570
	남자	91	122	123	121	319	310	155	151	156	1548
	소계	183	245	260	241	638	626	310	303	312	3118
오류 발생 인원	여자	9	12	23	14	45	42	39	32	34	250
	남자	13	7	18	14	44	41	23	26	25	211
	소계	22	19	41	28	89	83	62	58	59	461
수정 인원	여자	4	5	8	3	14	11	14	14	12	85
	남자	3	3	1	2	8	9	6	4	4	40
	소계	7	8	9	5	22	20	20	18	16	125

표 7. 연령대별 다변량 편집규칙 검증 결과(단위: 명)

결과	분류 \ 연령	8~10세	11~13세	14~16세	17~19세	20대	30대	40대	50대	60대	총합
측정 인원	여자	92	123	137	120	319	316	155	152	156	1570
	남자	91	122	123	121	319	310	155	151	156	1548
	소계	183	245	260	241	638	626	310	303	312	3118
오류 발생 인원	여자	0	3	2	0	4	3	0	7	0	19
	남자	2	2	0	8	0	0	6	2	6	26
	소계	2	5	2	8	4	3	6	9	6	45
수정 인원	여자	0	1	2	0	0	0	0	1	0	4
	남자	2	0	0	0	0	0	0	0	1	3
	소계	2	1	2	0	0	0	0	1	1	7

점을 찾아내기 위해 통계 분석 프로그램 R을 사용하여 산점도 행렬을 분석하였다. 그 결과 명확하게 이상치로 의심되는 점들에 대해 원 자료를 확인하여 <표 7>에 제시하였다.

<표 7>을 보면, 여자 19명, 남자 26명의 총 45명의 데이터가 산점도 행렬에서 이상치로 검출된 것을 알 수 있다. 이들 이상치가 검출된 피험자의 형상 데이터를 확인해본 결과 목밑둘레와 가슴둘레의 측정 경로에 오류가 발견되어 결과적으로 여자 4명, 남자 3명의 데이터를 재측정하게 되었다.

이상의 연구 결과를 요약하면 다음과 같다.

첫째, 54개의 논리적 편집규칙에 의해 데이터를 검증한 결과, 22개의 규칙에서 특이점을 발견하였으며, 오류의 원인을 파악하고 재측정 여부를 판단하여 데이터의 정확성을 높였다. 둘째, 두 단계의 통계적 데이터 편집을 실시하였다. 먼저 성별, 연령대별로 개개의 변수 값들을 독립적으로 검토하여 이상치를 검출한 결과 총 3118명 중 461명을 검출하였고, 그 중 명확한 오류로 판단된 125명을 재측정하였다. 다음으로 산점도 행렬에 의해 변수들 상호간의 관계를 파악하여 이상치의 존재 여부를 파악하였으며, 그 결과 여자 4명, 남자 3명의 데이터를 재측정하게 되었다.

4. 토의 및 결론

본 연구는 '제 5차 한국인 인체치수 조사사업'에서 처음으로 실시한 3차원 인체측정 자료의 품질을 향상시키고자 데이터 편집의 과정을 크게 두 단계로 나누어서 데이터 검증을 실시하였다. 3차원 인체측정은 기존의 직접측정과는 달리 피험자의 3차원 형상 데이터를 데이터베이스화 하여 저장할 수 있다는 장점이 있다. 따라서 언제든지 3차원 형상 데이터의 재측정을 통해 피험자의 인체치수 정보를 추출할 수 있으므로 데이터의 검증이 용이하고 상대적으로 신뢰성 높은 데이터를 제공할 수 있다. 본 연구에서는 이상치로 검출된 피험자의 경우, 3차원 형상 데이터로부터 오류 여부를 확인하여 명확한 오류로 판단된 경우 재측정을 실시하였다.

본 연구에서는 인체치수 데이터베이스 구축 과정에서 합리적인 데이터 편집 과정의 중요성을 제시하고자 하였다. 또한 기존의 연구에서 다루지 않은 다변량 연속 변수의 데이터 편집에 대한 새로운 사례를 보여주는데 의의가 있다. 물론 대규모 다변량 조사의 데이터 편집을 위해서는 GGobi (Swayne et al., 2003)와 같은 전문적인 소프트웨어를 활용하는 방법이 있지만 아직 대부분의 이용자들에게는 익숙하지 않아 일반화하기 어려운 측면이 있다. 본 연구에서는 쉽게 활용할 수 있는 산점도 행렬을 활용하여 다변량 데이터 편집을 수행하였는데, 이는 현실적인 실용성 측면에서 유용하므로 향후 대규모 3차원 인체측정 사업의 데이터 편집을 위한 시스템 개발에 기초자료가 될 수 있을 것으로 기대한다.

참고 문헌

남윤자, 최경미, 정의승, 윤명환. Size Korea 3차원 인체측정 방법 표준화, *패션정보와 기술*, 1, 6-19, 2004.

박진우, 박현주, 김진억. 주택가격동향조사를 위한 데이터 편집 사례연구, *조사연구*, 6(1), 83-98, 2005.

산업자원부 기술표준원, 산업기술기반조성에 관한 보고서, 2002.

산업자원부 기술표준원, 제5차 한국인 인체치수 조사사업 보고서, 2004.

ISO 20686, 3-D Scanning Methodologies for Internationally Compatible Anthropometric Databases, 2005.

Granquist. L., *Improving the Traditional Editing Process, In Business Survey Methods*(eds. Cox et al.), John Wiley &Sons, 177-199, 1995.

Swayne, D. F., Lang, D. T., Buja, A. and Cook, D., GGobi: evolving from XGobi into an extensible framework for interactive data visualization, *CSDA*, 43, 423-444, 2003.

◐ 저자 소개 ◑

❖ 박 선 미 ❖ pjsunmi@paran.com
서울대학교 의류학과 박사
현 재: Iowa State University, Department of Apparel, Educational Studies, and Hospitality Management, Research Scholar
관심분야: 의복구성, 3차원 인체측정

❖ 남 윤 자 ❖ yunja@snu.ac.kr
서울대학교 의류학과 박사
현 재: 서울대학교 의류학과 교수
관심분야: 의복구성, 3차원 인체측정

❖ 박 진 우 ❖ jwpark@suwon.ac.kr
서울대학교 계산통계학과 박사
현 재: 수원대학교 통계정보학과 교수
관심분야: 표본조사, 통계품질

논 문 접 수 일 (Date Received) : 2009년 08월 17일
논 문 수 정 일 (Date Revised) : 2009년 11월 27일
논문게재승인일 (Date Accepted) : 2009년 11월 30일