

Estimation of Hard-to-Measure Measurements in Anthropometric Survey¹⁾

Jong Hoo Choi²⁾, Ryu Jin Kim³⁾

Abstract

Anthropometric survey is important as a basis for human engineering fields. According to our experiences, there are difficulties in obtaining the measurements of some body parts because respondents are reluctant to expose. In order to overcome these difficulties, we propose a method for estimating such hard-to-measure measurements by using easy-to-measure measurements those are closely related to them. Multiple Regression Model, Feedforward Neural Network(FNN) Model and Projection Pursuit Regression(PPR) Model will be used as analytical tools for this purpose. The method we propose will be illustrated with real data from the 1992 Korea national anthropometric survey.

Keywords : Anthropometric survey, hard-to-measure measurements, Feedforward Neural Network Model, Projection Pursuit Regression Model

1. 서론

정부는 1979년, 1986년, 1992년, 1997년 네차례에 걸쳐 '국민인체측정조사'를 수행하였다. 이 조사의 목적은 제반 산업설계의 기반이 되는 인체측정자료를 산업체에 보급함으로써 의류, 가구류, 신발, 설비 등 산업제품이 보다 국민생활에 편리하게 만들어지게 하고자 하는데 있다(박경수, 1993; 공업진흥청, 1992).

인체측정의 측정항목(부록 참조)의 선정은 KS A 7003(인체측정용어정의 한국표준)에 따랐으며, 측정방법은 마틴(Martin)자를 이용하여 직접 측정법을 사용하였고 KS A 7004(인체측정방법 한국표준)에 따랐다(국립기술품질원, 1997). 이러한 직접 측정방법은 비용이 적게 들며, 이동 측

1) 이 연구는 고려대학교 특별연구비에 의하여 수행되었음.

2) Associate Professor, Department of Informational Statistics, Korea University, Chochiwon, 339-700, Korea.

E-mail : jchoi@tiger.korea.ac.kr

3) Division of Statistics for Environmental Studies, Department of Environmental and Mathematical Sciences, Okayama University, Okayama, 700-8530, Japan.

E-mail : jacobian@nownuri.net

정이 용이하나 측정과정에서 피측정자와 측정자 간에 신체접촉이 이루어지기 때문에 측정오차가 수반되는 문제가 발생된다. 특히 젓꼭지 간격(양쪽 젓꼭지점 사이의 직선거리), 밑가슴둘레(유방 밑 유곽선을 기준으로 하는 가슴 수평 둘레), 회음높이(좌우 좌골결절의 앞부분을 연결하는 선과 정중선과의 만나는 점인 회음점에서 바닥까지의 수직거리) 등과 같은 측정항목의 경우 측정시 피측정자의 수치심을 유발시키며 정확하게 측정하기에 어려움이 많다. 본 연구의 목적은 이러한 점에 착안하여 상대적으로 측정이 용이한 측정항목을 이용하여 측정이 곤란한 부위의 측정항목을 간접적으로 추정하는 통계적 모형을 구축하고 그 효용을 비교하려는 데 있다.

본 연구에서 고려되는 측정이 곤란한 측정항목은 ‘젓꼭지 간격’과 ‘밑가슴둘레’이다. ‘젓꼭지 간격’의 측정방법은 피측정자가 서있는 자세에서 피측정자가 자연스럽게 숨을 들이마신 후 숨을 멈추듯이 할 때, 양쪽 젓꼭지점 사이의 직선거리를 앞쪽에서 측정한다. 또한 ‘밑가슴둘레’의 측정방법은 피측정자가 동일한 자세를 취하게 하고 좌우 유방 밑 유곽선을 지나도록 하는 둘레를 앞쪽에서 수평되게 측정한다(공업진흥청, 1992).

최종후 외(1998)는 동일한 문제에 대하여 다중회귀모형과 신경망모형(Neural Network Model)의 적용을 시도하였다. 여기서 목표변수인 측정곤란부위를 설명하기 위한 설명변수로서 측정곤란부위를 제거한 나머지 부위와 피측정자의 연령을 고려한 바 있는데 본 연구에서는 피측정자의 측정부위만을 설명변수로 고려한다. 본 연구의 분석 모형에서 연령을 제외한 이유는 측정이 용이한 측정항목을 이용하여 측정이 곤란한 부위의 측정항목을 간접적으로 추정하는 통계적 모형을 구축하고자 하는 본 연구결과의 효용성 판단을 위하여 본 연구에서 이차적 변수라고 생각되는 연령을 배제시킨 것이다. 본 연구에서 고려되는 모형으로는 선형모형으로서 다중회귀모형을, 그리고 비선형모형으로서 신경망모형과 PPR(Projection Pursuit Regression)모형을 적용하여 각각 측정곤란항목의 추정모형을 구축하고, 구축된 세 모형의 타당성 평가와 비교를 시도한다.

2절에서는 본 연구에서 활용된 분석자료를 설명하며, 3절에서는 신경망모형에 의한 분석결과를, 4절에서는 PPR모형에 의한 분석결과를 각각 설명한다. 마지막으로 5절에서는 모형에 대한 평가결과를 정리한다.

2. 분석자료

1992년 국민인체측정조사는 6세부터 50세까지의 전 국민을 모집단으로 하여 표본설계에 따른 6,647개의 표본에 대하여 인체 84개 부위를 측정하였다(공업진흥청, 1992). 본 연구에서 활용하는 자료는 1992년 ‘국민인체측정조사’ 자료 중에서 피측정자의 나이가 만 12세 이상(중학생 이상)인 여성에 대한 것이다. 이 자료 중에서 각 측정항목의 평균을 중심으로 $\pm 3 \times$ 표준편차를 벗어나는 자료를 제거시킨 결과 분석에 이용되는 자료는 3,144개이다. 또한 분석자료에서 랜덤추출한 300개의 자료는 분석결과의 타당성 평가를 위하여 사용하였고 나머지 2,844개 자료는 모형구축을 위하여 이용하였다.

연구의 목표변수가 되는 ‘밑가슴둘레’, ‘젓꼭지 간격’을 설명하는 중요 측정항목을 선별하기 위하여 단계적 변수선택법에 의한 다중회귀분석을 이용하였다. 이때 진입유의수준(Significance Level Entry)은 0.1이고, 잔류유의수준(Significance Level Stay)은 0.005로 하였다. 단계적 변수선택법에 의해 선택된 설명변수는 밑가슴둘레에 대해서는 8개, 젓꼭지간격에 대해서는 9개가 각각 선택되었으며 구체적인 내용은 <표 1>과 같다.

<표 1> 단계적 변수선택법으로 선택된 설명변수

| 밑가슴둘레 | 젓꼭지간격 |
|---|--|
| 가슴너비, 목둘레, 윗가슴둘레, 가슴둘레, 허리둘레, 손너비, 어깨길이, 앞뎡 | 몸통너비, 가슴너비, 엉덩이너비, 가슴둘레, 앞뎡 팔꿈치 높이, 손너비, 앞뎡, 뒤뎡, 안소매길이 |

3. 신경망모형

3.1 신경망모형에 의한 탐색적 평가

본 연구에서 적용되는 신경망모형은 전방향 신경망모형(Feedforward Neural Network Model)인 다계층 퍼셉트론(Multi Layer Perceptron: MLP)모형이다. 이는 입력층(input layer), 은닉마디로 구성된 은닉층(hidden layer), 그리고 출력층(output layer)으로 구성된다(Berry and Linoff, 1997 ; Kohonen, 2001).

<표 2>는 ‘젓꼭지 간격’과 ‘밑가슴둘레’에 대한 하나의 회귀모형과 두 개의 신경망모형에 대한 분석 결과 중 MSE를 보여주고 있다. <표 2>에서 신경망모형1은 <부록>의 측정항목 84개 중에서 목표변수인 ‘젓꼭지 간격’과 ‘밑가슴둘레’를 제외한 입력변수 82개에 대한 모형이고, 신경망모형 2는 <표1>의 단계적 변수선택법을 이용한 회귀분석에 의하여 선택된 설명변수를 사용한 모형이다. 이 때 적용된 신경망모형은 탐색적 평가를 위하여 3개의 마디를 갖는 1개의 은닉계층으로 구성된 것이다. 이 모형은 통계패키지(SAS E-Miner)에서 디폴트로 제공된다.

<표 2>에서 MSE값이 가장 작은 모형은 젓꼭지 간격의 경우 신경망모형2, 그리고 밑가슴둘레의 경우 신경망모형1로 나타났다.

<표 2> 각 모형에 대한 평가

| | 젓꼭지 간격에 대한 모형의 MSE | 밑가슴둘레에 대한 모형의 MSE |
|--------|--------------------|-------------------|
| 신경망모형1 | 1.41 | 6.29 |
| 신경망모형2 | 1.25 | 6.53 |
| 회귀모형 | 1.28 | 6.64 |

3.2 최적화된 신경망모형

이제 단계적 변수선택방법을 이용하여 선택된 변수만을 이용하여 목표변수를 추정할 때, 신경망모형을 몇 개의 계층과 몇 개의 마디로 구성하는 것이 최적인지를 탐색하고자 한다.

신경망모형을 최적화하기 위해 1개의 은닉계층에 대하여 각각 1개~10개의 마디를 갖는 10가지의 모형, 그리고 2개의 은닉계층에 대하여 각각 1개~10개의 마디를 갖는 100가지의 모형을 비교한다. 이때 신경망모형의 타당성 평가를 위하여 2,884개의 자료를 분석용 자료(training data) 60%,

타당성 평가용 자료(validation data) 40%로 나누어 분석한다.

<표 3>과 <표 4>에서 밀가슴돌레와 젓꼭지간격에 대하여 총 110개의 신경망모형 중에서 분석용 자료의 MSE가 가장 작은 순서에 따른 신경망모형을 나타내었다. 여기서 MSE(V)는 평가용 자료에 대한 MSE 결과이다. <표 3> 밀가슴돌레의 경우 1개의 은닉계층, 5개의 마디를 갖는 신경망모형에서 MSE가 가장 작게 나타났으며(평가용 자료의 경우는 1개의 은닉계층, 4개의 마디를 갖는 신경망모형), <표 4> 젓꼭지간격의 경우에서 밀가슴돌레의 경우 2개의 은닉계층, 각각 2, 6개의 마디를 갖는 신경망모형에서 MSE값이 가장 작게 나타났다(평가용 자료의 경우는 2개의 은닉계층, 각각 2개의 마디를 갖는 신경망모형).

<표 3> 밀가슴돌레에 대한 신경망모형

| | 신경망모형 | | MSE | MSE(V) |
|-----------------------|--------|---------|--------|--------|
| | Layers | Neurons | | |
| 밀 가 슴 돌 레 | 1 | 5 | 6.8868 | 5.7076 |
| | 1 | 4 | 6.9528 | 5.6696 |
| | 1 | 8 | 7.0262 | 5.7604 |
| | 1 | 9 | 7.0379 | 5.8426 |
| | 2 | 1, 8 | 7.0480 | 5.9313 |
| | 2 | 3, 4 | 7.1443 | 5.9300 |
| | 2 | 3, 6 | 7.1488 | 5.9820 |
| | 2 | 2, 5 | 7.1523 | 5.9802 |
| | 2 | 2, 8 | 7.1668 | 5.9638 |
| | 2 | 3, 3 | 7.1868 | 5.8467 |

<표 4> 젓꼭지간격에 대한 신경망모형

| | 신경망모형 | | MSE | MSE(V) |
|-----------------------|--------|---------|--------|--------|
| | Layers | Neurons | | |
| 젓 꼭 지 간 격 | 2 | 2, 6 | 1.2402 | 1.3041 |
| | 2 | 2, 3 | 1.2410 | 1.3031 |
| | 2 | 2, 5 | 1.2479 | 1.3137 |
| | 2 | 1, 1 | 1.2486 | 1.2991 |
| | 2 | 3, 6 | 1.2520 | 1.3380 |
| | 2 | 2, 4 | 1.2590 | 1.2965 |
| | 2 | 3, 3 | 1.2594 | 1.3051 |
| | 1 | 5 | 1.2601 | 1.3221 |
| | 2 | 1, 2 | 1.2618 | 1.3078 |
| | 2 | 2, 2 | 1.2623 | 1.2910 |

4. PPR 모형

PPR(Projection Pursuit Regression) 모형은 다음과 같다.

$$Y_{ij} = \beta_{i0} + \sum_{m=1}^M \beta_{im} f_m(\alpha_m^T X_j) + \varepsilon_{ij}, \quad i = 1, \dots, q, \quad j = 1, \dots, n$$

여기서 Y_{ij} 는 j 번째 벡터 Y_j 의 i 번째 값이고, ε_{ij} 는 오차항이다. 또한 β_{i0} 와 β_{im} 는 미지의 모수이고, α_m 은 $\alpha_m^T \alpha_m = 1$ 이 되는 p -차원의 모수 벡터이다. $k=1, \dots, M$ 인 β_k 는 q -차원 벡터이다(Klinke and Grassmann, 1998).

f_m 은 $E(f_m(\alpha_m^T X_j)) = 0$ 과 $E(f_m^2(\alpha_m^T X_j)) = 1$ 을 충족하는 미지의 함수이다. 모형에서는 α 와 β 그리고 f_m 이 추정되며 다음의 L_2 를 최소화함으로써 얻어진다.

$$L_2 = \sum_{i=1}^q W_i \sum_{j=1}^n w_j (Y_{ij} - \beta_{i0} - \sum_{m=1}^M \beta_{im} f_m(\alpha_m^T X_j))^2$$

α_m 이 주어질 때 f_m 의 추정값, \hat{f}_m 은 초평활(super smoothing)이나 헤르미트(Hermite) 함수를 통해 얻어진다. $f_m(\alpha_m^T X_j)$ 의 평활값은 가중평균이 0 이고 가중분산이 1로 표준화된다. $m=1$ 즉, 항(term)이 하나라고 할 때, 표준화함수가 \hat{f}_1 이라면, β_1 에 대하여 그 추정값은 $\hat{f}_1(\alpha_1^T X_j)$ 를 갖는 Y_j 의 회귀에서 얻어지며 다음의 식과 같다.

$$\hat{\beta}_1 = \sum_{j=1}^n w_j (Y_j - \bar{Y}) \hat{f}_1(\alpha_1^T X_j)$$

따라서 $\hat{f}_1(\alpha_1^T X_j)$ 의 가중분산은 1이다. 그리고 α_1 에 대한 추정값은 가우스-뉴턴(Gauss-Newton) 방법으로 얻어진다.

$$g_{ij}(\alpha_1) = Y_{ij} - (\bar{Y})_i - \beta_{1j} f_1(\alpha_1^T X_j)$$

이 때 L_2 를 최소화시키는 α_1 의 추정값은 다음의 순환식에 의해서 얻어진다.

$$\hat{\alpha}_1^{(l+1)} = \hat{\alpha}_1^{(l)} + d^{(l)}$$

PPR분석은 단계적 변수선택방법에 의해 선택된 변수인 <표 1>에서 제시된 설명변수를 이용하여 각각 밀가슴둘레와 젖꼭지간격 변수에 대하여 실시하였다.

5. 모형에 대한 평가

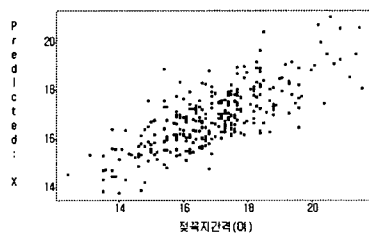
자료로부터 랜덤추출한 300개의 타당성 평가용 자료를 이용하여 회귀모형, 신경망모형, 그리고 PPR모형의 타당성을 평가를 시도하였다. <표 5>에는 두 개의 측정곤란부위에 대하여 각각의 추정모형을 적용하여 실제값과 추정값 간의 상관계수와 MSE를 구한 결과가 정리되어 있다. 이때 신경망모형은 MSE가 가장 작은 신경망모형 즉, 밀가슴둘레에 대하여는 1개의 은닉계층 5개의 마디를 갖는 모형에, 그리고 젓꼭지 간격에 대하여는 2개의 은닉계층 각각 2개와 6개의 마디를 갖는 모형에 적용하였다.

분석결과를 보면 ‘젓꼭지 간격’이나 ‘밀가슴둘레’의 경우 모두 신경망모형에 의한 상관계수가 가장 크게 나타났으며, MSE가 가장 작게 나타났다. 결국 두 경우 모두 신경망 모형이 회귀모형이나 PPR모형보다 나은 추정결과를 보여주고 있다.

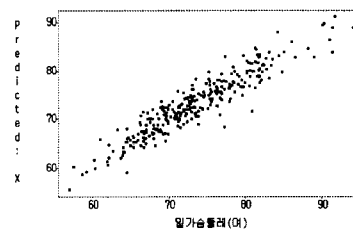
<표 5> 모형에 대한 타당성평가

| 측정곤란부위 | 모형의 종류 | 상관계수 | MSE |
|--------|--------|--------|--------|
| 젓꼭지간격 | 회귀모형 | 0.7585 | 0.6685 |
| | 신경망모형 | 0.7664 | 0.5359 |
| | PPR모형 | 0.7075 | 0.6588 |
| 밀가슴둘레 | 회귀모형 | 0.9362 | 4.5852 |
| | 신경망모형 | 0.9452 | 4.1863 |
| | PPR모형 | 0.9216 | 5.6830 |

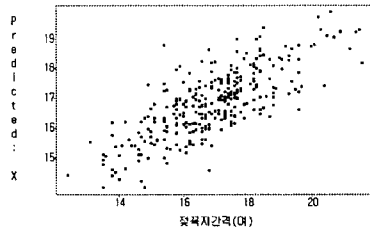
<그림 2>, ... ,<그림 6>은 각각 ‘젓꼭지 간격’, ‘밀가슴둘레’에 대한 회귀모형과 신경망모형, PPR모형에 의한 타당성 평가용자료의 실제값과 추정값 간의 산점도이다.



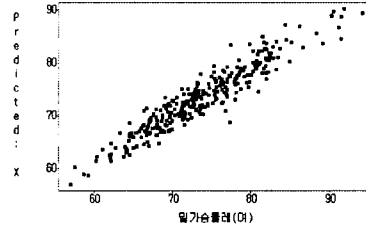
<그림 1> 젓꼭지간격에 대한 회귀모형



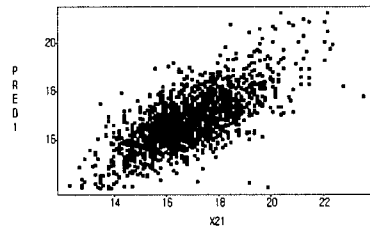
<그림 2> 밀가슴둘레에 대한 회귀모형



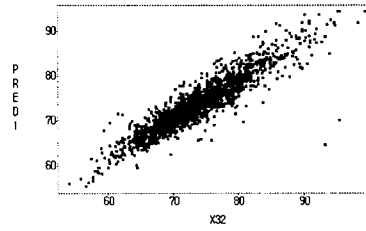
<그림 3> 젖꼭지간격에 대한 신경망모형



<그림 4> 밑가슴둘레에 대한 신경망모형



<그림 5> 젖꼭지간격에 대한 PPR모형



<그림 6> 밑가슴둘레에 대한 PPR모형

참고문헌

[1] 국립기술품질원 (1997). KRISS-97-114-IR, 산업제품의 표준치 설정을 위한 국민표준체위조사 보고서, 9.

[2] 공업진흥청 (1992). KRISS-92-144-IR, 산업제품의 표준치 설정을 위한 국민표준체위조사 보고서.

[3] 박경수 (1993). 「인간공학 - 작업경제학-」, 영지문화사, 서울.

[4] 최종후, 서두성, 홍한표 (1999). 신경망모형을 이용한 인체측차료에서의 측정곤란부위의 추정, 「Data Analysis」, Vol. 1, No. 1, 29-36.

[5] Berry, J. A. & Linoff, G. S. (1997). *Data Mining Techniques For Marketing, Sales, and Customer Support*, John Wiley & Sons, Inc. New York.

[6] Klinke, S., & Grassmann, J. (1998) Projection Pursuit Regression and Neural Networks, *Sonderforschungsbereich*, 373, Humboldt Universitaet. Berlin.

[7] Kohonen, T. (2001). *Self-Organizing Maps*, 3rd Ed., Springer-Verlag. Heidelberg

[2001년 9월 접수, 2002년 2월 채택]

<부록> 측정부위별 측정항목

| Part I | |
|--|---|
| (1) 키 (2) 눈높이 (3) 어깨높이 (4) 목뒷높이(C7) (5) 허리높이 (6) 팔굽힌 팔꿈치 높이 (7) 엉덩이 밑높이 (8) 손끝높이 (9) 회음높이 (10) 대퇴돌기높이 (11) 무릎안쪽높이 (12) 머리위로 뺀 Grip 높이 | (13) 옆으로 뺀 손끝길이 (14) 앞으로 뺀 손끝길이 (15) 양팔 뺀 손끝길이 (16) 어깨너비 (17) 몸통너비 (18) 가슴너비 (19) 허리너비 (20) 엉덩이너비 (21) 젖꼭지간격(여) (22) 가슴두께 (23) 배두께 (24) 엉덩이두께 |
| Part II | |
| (25) 목둘레 (26) 목밑둘레 (27) 진동둘레 (28) 윗팔둘레 (29) 아래팔둘레 (30) 윗가슴둘레(여) (31) 가슴둘레 (32) 밑가슴둘레(여) (33) 허리둘레 (34) 배둘레 (35) 엉덩이 둘레 (36) 넓적다리둘레 (37) 무릎둘레 (38) 장딴지둘레 (39) 앉은키 (40) 앉아머리위로 뺀 Grip높이 | (41) 앉은 눈높이 (42) 앉은 어깨높이 (43) 앉은 팔꿈치높이 (44) 대퇴높이 (45) 앉은무릎높이 (46) 앉은 오금높이 (47) 엉덩이 무릎길이 (48) 엉덩이 오금길이 (49) 앉은 엉덩이 너비 (50) 뒤허리 발뒤꿈치길이 (51) 어깨점 팔꿈치길이 (52) 팔꿈치 손끝길이 (53) 머리길이 (54) 얼굴길이 (55) 눈턱길이 |
| Part III | |
| (56) 머리너비 (57) 머리두께 (58) 귀구슬사이 너비 (59) 귀구슬사이 턱밑길이 (60) 귀두슬사이 턱끝길이 (61) 귀구슬사이 머리마루점길이 (62) 머리둘레 (63) 눈동자사이너비 (64) 입너비 (65) 손길이 (66) 손바닥길이 (67) 손너비 (68) 손두께 (69) 손둘레 (70) 어깨길이 | (71) 앞품 (72) 앞중심길이 (73) 뒤품 (74) 등길이 (75) 둔부길이 (76) 소매길이 (77) 안소매길이 (78) 화장 (79) 밑위 앞뒤길이 (80) 발등둘레 (81) 발목둘레 (82) 발길이 (83) 발너비 (84) 몸무게 |