한국인 인체치수조사를 위한 표본설계 연구

A Sampling Design of the Korean Anthropometric Survey

박진우* · 김진호** · 황인극***

The Korean Anthropometric Survey (Size Korea) is a sample survey which estimates on percentiles of several dimensional measurements of the human body and its component parts. The purpose of this study is to design a sample, which is designed on the base of 1997 survey database. Two different methods are considered to get the sample size for estimating the 5th and 95th percentile of body dimensions of Korean age range 0-80 years.

Ⅰ. 서 론

국민을 대상으로 하는 인체측정은 시간이 경과함에 따라 변하는 국민의 체형 및 신체치수에 적합한 각종 산업제품의 인체 적합도를 높여 기능성, 편리성, 안전성 등을 향상시키고 생산 활동을 향상시키기 위해 이루어진다. 특히 한국인 인체치수조사(Size Korea) 사업은 체형 및 체격 변화에 따른 신체치수를 제품설계에 적시 반영하키고, 다양한 인체 특성치를 수용하여 제품설계에 보다 효율적이고 합리적으로 적용하고자 하는 조사이다.

산업설계의 기초 자료로 사용되고 있는 인체측정 자료는 인종과 생활방식 등에 따라 상이하기 때문에 외국의 자료를 이용하는 데에는 한계가 있다. 특히 우리나라의 경우 국민들의 평균키나 몸무게 등은 매년 증가하는 경향을 보이고 있다. 그러므로 이러한 체위의 변화를 반영하지 못하는 과거의 자료를 이용할 때 문제가 발생할 수 있다. 변화된 국민의 체형 및 신체치수를 보다 정확히 파악하기 위해서 국가적 차원의 조사 사업이 필요하다. 미국, 일본, EU를 비롯한 많은 나라들에서는 국민들의 인체치수조사를 체계적으로 실시해오고 있으며 ISO (2003)에서는 인체치수 데이터베이스 구축을 위해 필요한 요구사항들에 대한 국제적인 표준을 제시하고 있다.

우리나라에서는 과거 네 차례에 걸쳐 국민인체치수조사를 실시한 바 있는데, 최근의 조사는 1997년 한국표준과학연구원에 의해 전국 13,062 명의 국민을 대상으로 이루어진 바

* 수원대학교 통계정보학과 부교수
** 공주대학교 산업시스템공학과 부교수
*** 공주대학교 산업시스템공학과 부교수

있다. 산업자원부 기술표준원에서는 2003년부터 새로이 제5차 한국인 인체치수조사를 수행하고자 한다. 본 연구는 2003년 조사를 위한 새로운 표본을 설계하는 것을 목적으로 하고 있다.

국가단위의 인체치수조사는 하나의 표본에 의해 거의 200 군데에 달하는 신체부위를 측정하는 다변량 표본조사이다. 대부분의 표본조사에서의 관심모수가 모평균이나 모총계인데 반해 인체치수조사의 관심모수는 백분위수(percentiles)라는 특성을 지니고 있다. 또한 성별, 연령그룹별로 모수를 추정해야 하는 관계로 성별, 연령별 신체특성을 고려하여 설계하는 것이 필수적이다. 아울러 한 명의 개인을 대상으로 많은 부분을 정확히 측정해야 하는 관계로 조사에 소요되는 시간이 길고, 조사비용도 많이 들며 이로 인해 조사대상을 섭외하는 것 자체가 매우 까다롭다는 것을 염두에 두는 것이 필요하다.

II. 1997년 국민인체치수조사 자료분석

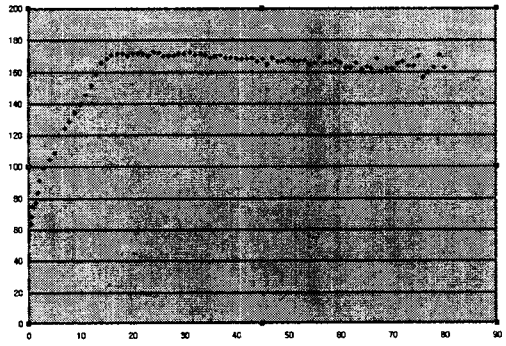
1. 연령그룹 구분

1997년 국민표준체위조사 자료를 기반으로 전국 13,062명에 대해 120개 항목을 조사한 자료를 가지고 일차 자료점검을 했는데 그 중 극소수의 이상점이 발견되었다. 전체 표본크기에 비해 미미한 정도이었으므로 이상점을 제거시킨 원자료(source data)를 기초로 하여 분석하였다.

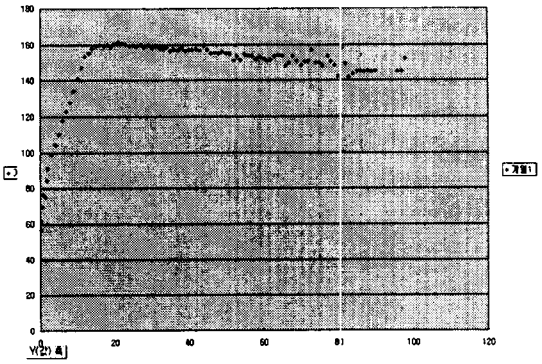
<그림 1>은 키, 가슴둘레, 몸무게 세 변수에 대해 연령별 평균값의 변화추이를 나타내는 그림이다. 세 변수 모두 연령에 따라 급격한 성장을 보이다가 일정 연령을 지나면 눈에 띠게 성장이 정체되는 시기가 있음을 알 수 있다. 또한 남자와 여자의 성장 양상에 약간의 차이가 있음도 알 수 있다.

<그림 1> 연령별 평균값 변화추이

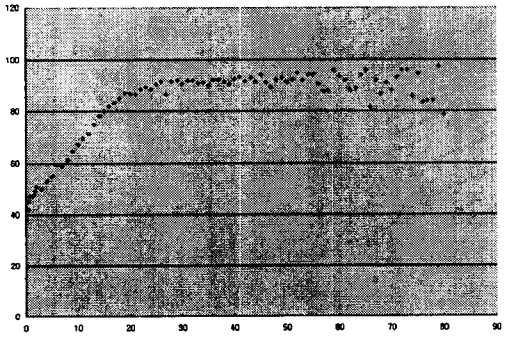
남자 키와 연령별



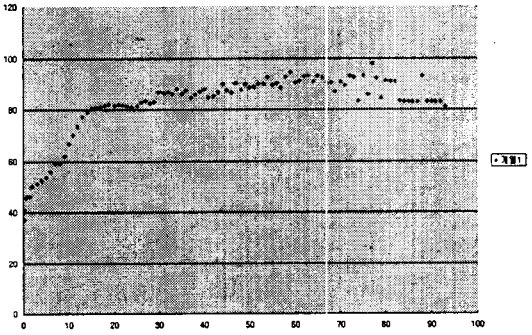
여자 키와 연령별



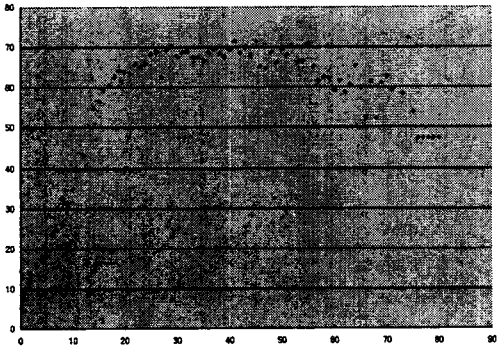
남자 가슴둘레와 연령별



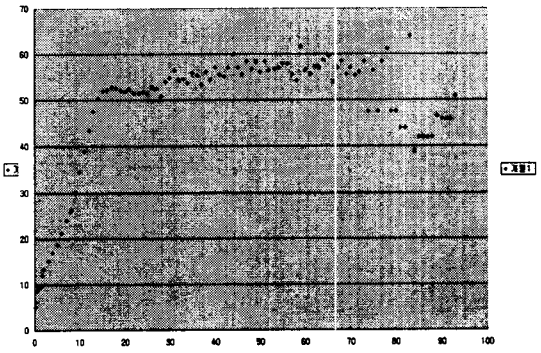
여자 가슴둘레와 연령별



남자 몸무게 연령별



여자 몸무게와 연령별



이 결과로서 성별, 연령그룹별로 각 변수들에 대한 통계량이 추정되어야 된다는 것을 알 수 있다. 그러나 연령그룹의 구분을 너무 세분하면 연령그룹 간의 차이가 발생하지 않

을뿐더러 조사 규모도 너무 크지는 문제점이 있다. 따라서 본 연구에서는 연령그룹의 구분은 성장이 극심한 시기는 구간을 짧게 하고, 성장에 큰 변화가 없는 시기는 구간을 길게 하였다.

2 표본크기와 표본오차

97년 표본설계 시에 표본크기의 결정은, 각 연령그룹별로 5개 중요부위인 키, 몸무게, 가슴둘레, 앉은 키, 팔길이 중 가장 변동계수가 큰 몸무게를 설계변수로 정한 후 연령그룹별 몸무게 평균추정값의 목표상대표준오차가 3.2%가 되도록 하는 방식으로 이루어졌다. 그 결과 남자 6190명, 여자 6310명, 합계 12,500명의 표본을 추출하도록 설계되었다. 다음의 표 1은 성별, 연령그룹별 목표 표본크기를 나타내고 있다. 참고로 실제 조사에서의 표본크기는 13,062명이 되었는데 조사과정에서 각 연령그룹별로 약간씩 추가되었기 때문이다.

<표 1> 성별 연령그룹별 목표 표본크기(1997년 조사)

나이그룹 (단위:세)	남자	여자	나이그룹 (단위:세)	남자	여자
	기존크기	기존크기		기존크기	기존크기
0.2	33	34	12	292	252
0.4	51	43	13	273	293
0.6	37	36	14	291	315
0.8	20	18	15	338	326
1.4	44	49	16	278	321
1.8	67	55	17	318	310
2	189	153	18	343	290
3	200	194	19	242	214
4	199	217	20-24	487	550
5	148	195	25-29	280	241
6	172	189	30-34	162	173
7	241	217	35-39	218	180
8	220	252	40-49	188	198
9	206	233	50-59	89	129
10	233	245	60이상	66	145
11	265	243			
합 계				6190	6310

그러나 97년 표본크기 결정 과정에서 드러나는 문제점 중의 하나는 추정하고자 하는

모수와 표본크기 결정시 고려한 통계량이 서로 다르다는 점이다. 즉, 관심모수는 백분위수인데 반해 설계시 고려한 모수는 모평균이었다는 사실을 지적할 수 있다.

다음으로 97년 데이터를 가지고 주요변수들인 키, 가슴둘레, 몸무게의 평균추정량과 제95 분위수에 대한 상대표준오차를 계산하였는데 <표 2>는 그 결과를 나타낸다. 전체적으로 평균추정량의 상대표준오차에 비해 백분위수의 상대표준오차가 작게는 2배, 크게는 3-4배 이상 큰 것을 알 수 있다.

<표 2> 주요변수들의 평균추정량과 제 95 분위수에 대한 상대표준오차(CV, %)

(1) 남자(1997년 조사자료 기초)

연령그룹	키		몸무게		가슴둘레	
	평균추정량	95분위수	평균추정량	95분위수	평균추정량	95분위수
0.2	1.1526	2.1094	2.9066	5.6452	0.5707	1.0625
0.4	0.8314	2.1439	2.0039	4.3103	1.1324	2.5532
0.6	0.5990	1.6598	1.6294	2.5000	0.7465	1.6146
0.8	1.9798	3.8922	4.5402	8.5417	0.4657	2.7941
1.4	0.9049	2.6238	1.9665	4.1667	0.3730	1.6537
1.8	0.5649	2.2141	1.5456	5.1786	0.2857	1.7720
2	0.4219	2.2660	1.0783	6.3235	4.3173	2.0833
3	0.2578	1.4915	0.7697	3.9474	1.2886	1.5681
4	0.2647	1.3016	0.7803	2.8750	1.0066	1.7075
5	0.3044	0.4829	1.0061	5.2885	0.9518	1.8750
6	0.3108	1.5552	1.0667	5.2459	1.1585	2.1773
7	0.2665	1.8059	0.9326	5.3279	0.9505	1.7134
8	0.3098	1.7329	1.0718	4.8220	0.9750	2.6012
9	0.2863	1.7014	1.2921	6.1765	0.9453	2.8378
10	0.2735	1.7091	1.1730	5.7692	0.9696	2.9004
11	0.2762	1.8051	1.2072	6.1927	0.8324	2.7301
12	0.2961	2.0006	1.2052	6.5126	0.7792	3.3677
13	0.3368	1.6735	1.1810	3.5448	0.7377	3.4078
14	0.2511	1.4972	0.9595	3.9007	0.7347	2.6578
15	0.1865	1.2004	0.8603	4.8973	0.6642	2.4590
16	0.1815	1.2465	0.8040	3.8851	0.6910	2.1335
17	0.1737	1.1383	0.7040	4.1667	0.6789	2.4332
18	0.1747	1.2077	0.7010	4.0323	0.6237	2.5052
19	0.2133	1.2933	0.8829	4.9383	0.7333	2.7124
20	0.1418	1.3092	0.5181	3.3871	0.5495	2.0256
25	0.1840	1.1615	0.7553	3.8690	0.6397	2.0669
30	0.1740	0.8989	0.8288	3.9634	0.9310	2.6160
35	0.2109	1.4358	0.7925	4.8485	0.8974	2.4378
45	0.2183	0.9086	0.8821	3.8690	0.7895	2.5540
55	0.2989	1.2956	1.2131	4.3750	1.4875	2.1019
65	0.3867	1.3905	1.7207	3.2095	1.7903	2.5495

(2) 여자

연령그룹	키		몸무게		가슴둘레	
	평균추정량	95분위수	평균추정량	95분위수	평균추정량	95분위수
0.2	1.42674	5.0553	3.1284	5.5556	1.3254	3.0303
0.4	0.74024	1.9115	2.5240	6.7073	1.1621	3.6413
0.6	1.37797	3.5896	3.0007	7.0833	1.0955	2.5828
0.8	2.44879	5.3684	4.8938	7.6786	0.4692	1.1856
1.4	1.09750	3.9518	2.2485	5.2174	1.0207	5.9454
1.8	0.71273	2.1939	2.0852	6.2500	0.7108	2.0036
2	0.46406	2.1794	1.1830	5.1471	3.1780	2.2936
3	0.31815	1.7177	0.7960	4.8611	1.2030	1.7336
4	0.28751	1.8336	0.7974	4.7619	1.1044	1.7287
5	0.31903	1.7227	0.9295	4.3478	1.0664	1.9393
6	0.33057	1.8254	1.0798	6.9231	1.0846	2.4180
7	0.28473	1.5221	1.0156	4.5082	0.9080	1.6975
8	0.29795	1.7017	1.0770	6.6176	1.0025	2.5226
9	0.31531	1.6871	1.2246	6.5476	0.9872	2.6786
10	0.32834	2.3934	1.2580	5.8511	0.9817	3.8317
11	0.28540	1.5873	1.3122	5.4545	0.9796	3.3030
12	0.22657	1.2067	1.0528	5.5804	0.8683	3.4706
13	0.20994	1.3956	0.9784	5.7540	0.6933	3.0590
14	0.17887	1.0974	0.8032	5.0781	0.6985	3.0033
15	0.17732	1.1330	0.7544	4.2308	0.7568	2.2727
16	0.17314	1.2224	0.7094	4.4574	0.7478	2.2376
17	0.18506	1.3228	0.7155	4.5038	0.7534	2.5946
18	0.18309	1.2641	0.7287	4.8828	0.7767	2.4096
19	0.21670	1.2388	0.7682	4.3089	0.9217	2.6902
20	0.13074	1.1438	0.4827	3.2000	0.5993	2.2854
25	0.19725	1.2232	0.6673	4.0650	0.8699	1.8293
30	0.21378	1.0492	0.9176	4.8580	0.8694	1.2698
35	0.24490	1.3174	0.8059	3.6538	0.9369	2.3511
45	0.20906	1.2035	0.8402	3.4420	0.9504	2.5301
55	0.26263	1.2804	1.0408	4.2254	1.3133	3.3607
65	0.24651	1.1471	1.0848	4.2279	1.3378	2.0957

III. 국민인체 치수조사에서의 표본설계

1. 모집단과 추출틀

본 조사의 조사대상 모집단은 우리나라에 거주하는 모든 0세부터 90세까지의 대한민국 국민으로 하였다. 또 표본틀로는 우리나라 국민의 현황을 파악할 수 있는 가장 최신의 총조사 결과인 2000년 인구주택총조사 (통계청, 2002) 결과를 이용하였다.

2 설계변수의 결정

1997년 조사는 무려 120개 변수를 하나의 표본을 통해 측정, 추정하는 다변량 표본설계이다. 다변량 표본설계를 위해 모든 조사변수를 다 고려하는 것은 방대하고 비효율적이므로 가장 관심의 대상이 되는 조사변수들을 선택하여 표본설계를 하였다. 이때 어떤 조사변수들을 설계변수로 선택할 것인가를 결정하는 일은 매우 중요한 일이다.

1997년 조사보고서(1997)에는 남자 28개 항목, 여자 32개 항목의 변수들을 각각 연령, 키, 가슴둘레를 이용하여 적합시킨 회귀식이 나와 있다. 각 회귀식들의 결정계수(R^2)값이 모두 0.9를 넘는 것을 통해 일차적으로 연령, 키, 가슴둘레는 매우 중요한 설계변수임을 알 수 있다. 다음으로는 전체 측정 부위들 중 상대적으로 변동계수가 큰 부위로서 앉은 다리두께, 진동길이, 손두께 등이 있는데 이들을 추가적인 설계변수로 고려하였다. 왜냐하면 ISO기준(2002)과 같이 변동계수가 큰 변수를 사용하여 표본크기를 구하면 다른 변수에서 필요한 요구 조건을 만족하기 때문이다.

3. 백분위수 추정을 위한 표본크기 결정

1) 표본크기 결정방법

방법 I

백분위수 추정을 위한 표본크기 결정 공식으로 널리 알려진 것이 두 가지 있는데 하나는 ISO (2003)에서 제안한 식이고, 다른 하나는 유한모집단에서 백분위수 추정과 관련한 대표적인 연구결과인 Sedransk와 Smith (1988)의 추정법을 사용하는 것이다.

먼저 ISO에서 제안한 식은 다음의 (1) 식과 같이 표현된다:

$$n=\left(\frac{1.96\times CV}{a}\right)^2\cdot(1.534)^2 \qquad (1)$$

위 식에서 $CV = SD/\overline{Y}$ 이며 모평균 추정량의 변동계수의 추정값이다. 이 값은 평균 추정에서 표본크기를 결정하는 식에 $(1.534)^2$을 곱해준 식이다. 이 식은 일종의 근사식으로서 정규분포 하에서 95분위수 추정량의 분산이 평균추정량의 분산에 비해 $(1.534)^2$배 효율이 떨어진다는 사실을 이용하여 구한 근사식이다. 이 식은 평균의 추정량에 관한 기존의 이론을 그대로 이용한다는 면에서 편리하다는 장점을 지닌다. 그러나 표본의 크기가 작거나, 자료 중에 이상값 (outlier)이 존재하거나, 모집단의 분포가 대칭형이 아니라 한쪽으로 치우친 분포인 경우 부적절할 위험이 있다.

모집단의 분포가 정규분포 형태가 아니고 약간 치우친 분포를 띠게 될 경우에는 모평균보다는 백분위수를 추정하는 것이 더 바람직한 것으로 알려져 있다. 실제로 몸무게, 가슴둘레 등 둘레부위들은 정규분포보다 대수정규분포(log-normal distribution)에 더 가까운 것으로 알려져 있다.

방법 II

유한모집단에서 백분위수 추정과 관련한 대표적인 연구결과인 Sedransk와 Smith의 추정법(1988)이 있다. 그 추정법에 따라 제95분위수 추정량의 신뢰구간을 계산한 후 그것을 변동계수의 식으로 환산한 후 표본의 크기를 결정하는 것이다. 즉, 신뢰구간의 폭을 l 로 표현한다면 변동계수를 다음의 (2) 식과 같이 계산한다.

$$CV(\widehat{Y}_{95}) = \frac{l}{2 \cdot \widehat{Y}_{95}} \tag{2}$$

여기서 $\widehat{Y}_{95}$는 제 95분위수의 추정량을 나타낸다.

위 식에 의해 변동계수가 계산될 경우 새로운 표본크기는 다음의 식 (3)으로 계산되어진다.

$$n = n_0 \cdot \left(\frac{CV(\widehat{Y}_{95})_{old}}{CV(\widehat{Y}_{95})_{new}}\right)^2 \tag{3}$$

2) 새로운 표본크기 안

다음의 <표 3>은 앞에서 소개한 두 가지 방법에 의해 남자의 진동길이, 손두께, 몸무게 세 변수에 대한 표본크기를 계산한 결과를 나타내고 있다. 먼저 과거 데이터로 각 연령그룹별 변동계수를 계산한 후 그에 따라 새로운 목표정도를 정하였다. 가능한 각 연령그룹별 목표정도가 균일한 것이 바람직하지만 그렇게 할 경우 변수에 따라 특정 연령그룹은

너무 많은 표본을 필요로 하는 그룹이 생긴다. 가급적 균일한 목표정도를 유지하되 표본의 규모가 현실적으로 감당할 만한 규모가 될 수 있도록 목표정도를 정하였다.

표3의 결과를 살펴보면 대부분의 나이그룹, 대부분의 변수에 대해 전반적으로 ISO에서 제안하는 방안이 데이터에 근거한 유한모집단 비모수추정법에 의한 방안에 비해 작은 표본크기를 요구하는 것을 알 수 있다. 작더라도 어느 정도 비슷한 것이 아니라 경우에 따라서는 거의 1/10 정도의 표본크기를 요구하는 경우도 있다. 단순히 ISO 방안이 더 효율적이어서 이런 결과가 나온 것이라기보다는 인체치수 데이터가 ISO 방안에서 전제하는 가정에 맞지 않기 때문에 나타난 결과이다. 따라서 ISO 방안을 기준으로 표본크기를 결정하는 것은 바람직하다고 판단되어 유한모집단 비모수추정법에 의한 방안을 따르기로 하였다.

여섯 가지의 관심변수에 따라 표본의 크기가 서로 다르게 나타나는데 특정 나이그룹에 대해 표본의 크기가 지나치게 커지는 것을 피하면서 가급적 표본크기가 가장 크게 나타나는 변수의 값을 취하는 방법으로 표본의 크기를 최종 결정하였다. 최종표본의 크기와 그에 따른 기대 변동계수를 계산한 결과가 다음의 표 4이다. 가능한 한 구간의 폭이 같은 그룹에 대해서는 표본의 크기가 같게 하는 것이 바람직하므로 표 4에 나타난 최종적인 표본의 크기를 정할 때에는 이러한 점을 반영하였다. 또한 나중에 실제 조사를 마친 후 사후분석을 할 때 직업군별 차이에 따른 특성의 차이를 살펴볼 목적으로 가능한 한 20세 이상의 성인들의 표본의 크기를 크게 하였다. 이렇게 하여 구해진 최종 표본의 크기는 남자 6,650명, 여자 6,750명하여 총 13,400명이다. 이는 지난 1997년 표본크기인 남자 6,190명, 여자 6,310명 총 12,500명에 비해 약 900명 정도 많아진 수이다. 한편 이렇게 했을 때 기대 변동계수는 키의 경우 0.55-1.8% 내외로 표본오차가 매우 작다. 변동계수가 가장 큰 변수들로 손두께, 몸무게를 들 수 있는데 그 경우도 대부분 5% 이내이므로 새로운 표본크기에 의해 표본조사를 할 경우 각 연령그룹별로 모든 조사변수들에 대해 상당히 정도가 높은 통계를 구할 수 있을 것으로 기대된다.

<표 3> 새 표본설계를 위한 두 가지 표본크기 안(남자)

나이그룹	기존크기	진동길이			손두께			몸무게		
		목표정도	비모수	ISO	목표정도	비모수	ISO	목표정도	비모수	ISO
0.2	33							4	8	41
0.4	51							4	32	30
0.6	37							4	2	14
0.8	20							4	24	61
1.4	44							4	28	25
1.8	67							4	31	24
2	189	3	17	921	4	9	157	5	184	21
3	200	3	121	87	4	156	27	5	129	11
4	199	3.7	205	35	5	193	20	5	65	11
5	148	4	233	20	4	176	17	5	204	14
6	172	3	197	60	4	212	34	5	158	18
7	241	3.7	197	37	4	217	25	5	223	20
8	220	4	249	31	4.5	227	22	5	115	24
9	206	4	213	27	4	131	22	5	225	32
10	233	4	229	32	4	245	29	5	269	30
11	265	4	245	27	5.3	271	16	5	180	36
12	292	4	211	26	5	244	16	5	367	40
13	273	3.5	172	29	4.5	204	24	5	120	36
14	291	4	239	23	4	215	20	5	158	25
15	338	4	243	22	4.7	231	19	5	253	24
16	278	3	205	35	4.3	210	25	5	137	17
17	318	3	235	38	4	208	25	5	168	15
18	343	4	232	20	4.5	243	19	5	138	16
19	242	3	111	34	4	156	20	5	132	18
20-24	487	5	248	14	4	218	15	5	179	12
25-29	280	3	148	30	4.3	205	16	5	146	15
30-34	162	4	194	21	4	81	15	5	74	10
35-39	218	3	167	46	4.3	184	19	5	123	13
40-49	188	3	118	31	4	104	18	5	74	14
50-59	89	3	112	51	4	47	17	5	25	12
60이상	66	3	74	55	4	44	15	5	8	18
	6190									

<표 4> 새로운 표본크기에 따른 95분위수 추정값의 목표오차 (CV, %)

1) 남자(95분위수)

나이	표본크기	95분위수 추정값의 목표오차 (CV, %)					
		키	앉은다리 두께	가슴둘레	진동두께	손두께	몸무게
0.2	50	1.41		0.64			1.60
0.4	50	1.44		1.90			3.20
0.6	50	0.90		1.08			0.80
0.8	50	1.63		0.95			2.77
1.4	50	1.47		1.02			2.99
1.8	50	0.82		1.18			3.15
2	200	1.79	0.97	1.36	0.87	0.85	4.80
3	200	1.19	3.35	1.78	2.33	3.53	4.02
4	200	0.76	1.89	2.73	3.75	4.91	2.85
5	250	0.34	4.18	3.61	3.86	3.36	4.52
6	250	1.03	2.50	0.95	2.66	3.68	3.97
7	250	0.96	3.69	1.17	3.28	3.73	4.72
8	250	1.24	3.60	2.19	3.99	4.29	3.39
9	250	1.09	2.69	2.24	3.69	2.90	4.74
10	250	0.89	3.15	2.61	3.83	3.96	5.19
11	250	1.53	3.47	2.46	3.96	5.52	4.24
12	250	1.56	3.88	3.39	3.67	4.94	6.06
13	250	1.38	4.08	3.09	2.90	4.06	3.46
14	250	1.22	3.30	2.47	3.91	3.71	3.97
15	250	1.20	4.33	2.68	3.94	4.52	5.03
16	250	0.92	2.93	2.02	2.72	3.94	3.70
17	250	0.82	2.70	2.09	2.91	3.65	4.10
18	250	0.93	2.91	2.54	3.85	4.44	3.71
19	250	1.07	2.87	1.67	2.00	3.16	3.63
20-24	300	0.84	4.39	2.23	4.55	3.41	3.86
25-29	300	1.07	2.97	1.61	2.11	3.55	3.49
30-34	300	0.49	2.34	1.37	3.22	2.08	2.48
35-39	300	0.74	2.28	1.51	2.24	3.37	3.20
40-49	300	0.54	2.49	1.44	1.88	2.36	2.48
50-59	300	0.47	1.46	0.78	1.83	1.58	1.44
60-	200	0.51	2.01	0.97	1.82	1.88	1.00

2) 여자(95분위수)

나이	표본크기	키	앉은다리 두께	가슴둘레	진동두께	손두께	몸무게
0.2	50	1.62		1.98			3.23
0.4	50	1.33		1.74			4.80
0.6	50	1.79		1.65			2.51
0.8	50	1.52		0.28			0.42
1.4	50	1.39		2.02			3.00
1.8	50	1.41		1.77			4.83
2	200	1.4	1.13	1.92	0.94	0.78	4.62
3	200	1.49	2.14	1.08	2.33	2.53	4.06
4	200	1.38	3.19	1.56	4.52	4.80	4.35
5	250	0.93	2.96	1.52	3.55	3.23	3.34
6	250	0.85	2.6	1.47	2.57	3.79	3.55
7	250	0.89	2.76	1.18	3.37	3.1	3.80
8	250	1.35	3.14	1.97	3.62	4.70	5.83
9	250	1.48	3.24	2.05	3.22	3.78	5.41
10	250	1.09	3.04	2.99	4	3.05	4.75
11	250	1.28	3.41	2.83	4.58	3.64	4.73
12	250	0.62	2.98	2.79	4.82	4.68	4.82
13	250	1.15	3.28	2.7	3.49	4.54	5.8
14	250	0.84	3.5	2.16	2.9	5.38	4.31
15	250	0.93	3.48	2.15	3.93	5.68	4.34
16	300	0.72	4.07	2.08	3.37	3.20	3.55
17	300	1.09	2.58	2.28	2.87	5.60	4.16
18	250	1.17	3.46	2.2	3.62	3.92	3.98
19	250	0.72	2.19	1.94	3.73	3.9	3.03
20-24	300	1.11	3.68	2.13	5.54	4.84	4.05
25-29	300	1.02	3.06	1.47	2.72	3.92	3.07
30-34	300	0.66	2.4	0.67	2.69	2.80	2.42
35-39	300	0.68	2.38	1.63	2.21	2.81	2.53
40-49	300	0.55	2.29	1.84	2.79	2.88	2.20
50-59	300	0.5	1.63	1.64	1.46	2.26	2.38
60-	200	0.71	2.23	1.36	3.68	2.55	2.37

4. 표본추출방법

주어진 표본크기에 맞게 전국에서 표본을 추출하는 방법을 고려해야 한다. 표본추출은 층화 2단 집락 추출방식을 기본으로 하였다. 먼저 시, 군, 구를 집락으로 생각하여 1차로 추출하고, 다시 선택된 시군구 내에서 조사대상을 2차 추출하는 방식이다.

표본지역으로 먼저 서울과 6개 광역시(부산, 대구, 인천, 광주, 대전, 울산)를 포함시켰고, 나머지 도시와 군지역은 인구수에 따라 정렬한 후 확률크기비례추출법(probability proportional to size sampling : pps sampling) 에 의해 22개의 시군을 추출하였는데 이 과정에서 가능한 한 각 권역별로 시군이 골고루 추출될 수 있도록 최대한 배려하였다. 서울과 6대 광역시에서는 역시 마찬가지 방법으로 구를 표본으로 추출하였다. 다음의 <표 5>에는 표본으로 뽑힌 지역이 나열되어 있다.

<표 5> 표본지역

대지역	소지역
서울	송파, 관악, 강서, 성북, 중랑, 광진, 마포, 성동
부산	부산진, 사하, 사상, 동래, 영도
대구	달서, 북구, 서구
인천	부평, 남구, 계양
광주	북구, 광산
대전	서구, 대덕
울산	남구, 동구
경기강원(도시)	수원, 성남, 안산, 강릉
충청(도시)	아산, 서산, 청주
전라(도시)	전주, 김제, 여수, 순천
경상(도시)	경주, 김해, 통영
경기강원(군)	양구, 포천
충청(군)	영동, 태안
전라(군)	진도, 북제주
경상(군)	성주, 함양

다음으로는 표본으로 추출된 지역 내에서 조사대상을 추출하는 2차 추출 과정에 대해 소개한다. 한국인 인체치수조사는 표본으로 선정된 개인에 대해 150개가 넘는 부위를 측정하는 조사이므로 조사의 부담이 상당히 큰 조사이다. 또한 연령대별로 일정수의 표본을 확보해야 하는데 이 과정도 매우 어려운 작업이 된다. 이런 점을 고려하여 표본선정을 용

이하게 하기 위해 연령그룹을 나누어 각 연령그룹을 조사할 수 있는 조사기관이나 단체를 선정하기로 하였다. 연령그룹 구분과 조사가능 기관 또는 단체를 나타낸 것이 다음의 <표 6>에 나와 있다.

<표 6> 표본추출을 위한 연령그룹 및 조사가능 기관

연령그룹	조사기관
0 - 1세	병원, 가정
2 - 5세	유아원, 유치원, 가정
6 - 11세	초등학교
12 - 14세	중학교
15 - 17세	고등학교
18 - 24세	대학교, 군인, 근로자
25세 - 55세	일반인 근로자(회사), 주부
55세 - 70세	양로원, 복지회관, 노인대학

IV. 추정

유한모집단에서 $100\times p$ 백분위수의 추정값의 신뢰구간을 계산하는 단계를 간단히 요약하면 다음과 같다.

첫째, 표본데이터를 이용하여 모집단의 경험적 누적분포(empirical cdf) $F(y)$를 구한다.

둘째, $100\times p$ 백분위수의 추정값을 다음과 같은 식에 의해 계산한다.

$$\hat{Y}_{(t)} = F^{-1}(p)$$

셋째, 표본데이터 중 $\hat{Y}_{(t)}$ 값보다 작은 것들의 수를 세어 그 비율 $\hat{p}$를 계산한다.

넷째, 다음의 식을 이용하여 $F(Y_{(t)})$의 분산추정량을 계산한다.

$$\hat{\sigma}_p^2 = (1-\frac{n}{N})\cdot\frac{\hat{p}(1-\hat{p})}{n-1}$$

다섯째, 다음의 두 값을 계산한다.

$$\xi = \min\{p + 1.96 \times \hat{\sigma}_p, 1\}, \quad \gamma = \max\{p - 1.96 \times \hat{\sigma}_p, 0\}$$

여섯째, 위의 두 값을 이용하여 다음과 같은 신뢰구간을 계산한다.

$$(F^{-1}(\gamma), F^{-1}(\xi))$$

V. 맺음말

본 연구는 한국인 인체치수조사를 위한 표본설계연구로서 다목적 설계인 동시에 백분위수가 모수인 경우를 다루는 연구이다. 우리나라에서 백분위수 추정을 목적으로 하는 표본설계에 관한 최초의 연구라는 측면에서 본 연구는 나름의 의의를 지닌다고 할 수 있다.

표본설계를 위하여 키, 몸무게, 가슴둘레, 앉은 다리 두께, 손두께, 진동길이를 설계변수로 고려하였다. 표본크기를 결정하기 위해 ISO에서 제안한 근사식과 Sedransk와 Smith (1988)가 제안한 유한모집단 백분위수 추정법을 사용한 방식을 동시에 고려했는데 ISO 방식은 효과적이지 못한 것으로 판명되었다.

각 나이그룹별로 변수에 따라 목표 표준상대오차를 0.5-5% 정도가 되게 하는 표본크기를 구한 결과 남자 6,650명, 여자 6,750명하여 총 13,400명의 표본크기가 정해졌다. 추정하고자 하는 백분위수의 추정식은 Sedransk와 Smith (1988)가 제안한 유한모집단 백분위수 추정법을 사용하도록 제안하였다.

참고문헌

김동우 · 김진호. 1986.『1986년도 국민체위조사』. KSRI-86-56-IR. 국립기술품질원.

김동진 · 이영숙 · 김상철 · 양한수 · 이관석 · 이강철 · 남경희 · 안병덕. 1997.『산업제품의 표준치 설정을 위한 국민표준체위 조사 보고서』. KRISS-97-114-IR. 국립기술품질원.

김철중 · 이남식 · 김진호 · 박수찬 · 이웅호 · 최종후 · 강신철 · 이규금 · 이상도 · 이동춘. 1992.『산업제품의 표준치 설정을 위한 국민표준체위 조사』. KRISS-92-144-IR. 국립기술품질원.

박홍래. 2000.『통계조사론』. 영지문화사.

산업자원부 기술표준원 고분자섬유과. 2002.『인체형상 및 치수 표준화 기반기술 구축 - 산업기술기반조성에 관한 보고서』. 산업자원부.

통계청. 2000. 인구주택총조사(CD).

ISO/DIS 15535. 2003. "General Requirements for Establishing Anthropometric Databases." ISO draft.

Sedransk, J. and Smith, P. J. 1988. "Inference for Finite Population Quantiles." *Sampling, In Handbook of Statistics* eds. by Krishnaiah, P. R. and Rao, C. R., 6: 267-289.