

國民營養調查를 위한 標本設計 小考

全 泰 允* · 鄭 基 惠**

*釜山 東亞大學校 商經大學 應用總計學科

**서울大學校 保健大學院

A Sample Design for National Nutrition Survey

Tae Yoon Jun* · Kee Hey Chung**

**Department of Applied Statistics, College of Business and Economics,
Dong-A University*

***School of Public Health, Seoul National University*

= ABSTRACT =

In order to make clear the relationship between sample design and sample survey in community, it was conducted research on sample design for National Nutrition Survey in 1983. In this paper it was tried to analyze the data based on The Report of a Settled Population, 1981 conducted by National Bureau of Statistics Economic Planning Board.

The sample was basically using stratified two-stage sampling with systematic sampling of Ban or Li as administrative unit. The population represents the whole nation excluding Jeju-do because of budget. The selection of sampling unit and sampling procedure was as follows.

- 1) Stratify the nation-wide area in 20 sections according to administrative districts.
- 2) Determine the sample size in each section according to equal proportional rate(1 / 8040) and to about 1,000 households in the sample.
- 3) Select the 25 sampling units by section according to households proportion.
- 4) Select the 10 households at random from each Ban or Li according to equal probability proportion as the final sampling unit.

Using the procedure, it was sampled 1,000 households for National Nutrition Survey in 1983.

緒 論

보건사회부는 행정에 必要한 각종 자료를 수집하는데 標本抽出方法을 도입하여 統計調査를 실시하여왔다.

家口 또는 家口員을 추출단위로 하는 調査로 國民營養調査, 國民健康調査, 結核實態調査 등이 있으며 이들 調査는 全國을 推定하는 대규모 표본조사이므로 표본설계계획이 調査結果에 미치는 影響은 매우 크다고 하겠다.

國民營養調査는 1969년 이후 매년 실시되어 왔는데 그 標本設計를 보면 처음에는 전국을 山間, 平野, 山麓, 海岸島嶼, 都市의 5개지역으로 분할하여 각 地域에서 동일크기의 표본을 任意로 선정할 것이었다¹⁾. 이러한 지역분할은 각 지역집단의 가구수를 알 수 없는 것이어서 全國値는 물론 地域別推計値도 구할 수가 없는 것이었다.

그후 經濟企劃院 調査統計국이 國勢調査區를 위해서 설정한 調査區를 표본추출계획의 틀(tool)로 사용하여, 여기서 다시 家口를 추출하는 層化二段抽出計劃으로 바뀌었다²⁾. 이 경우에는 營養調査가 갖는 調査內容의 성격상 標本誤差를 엄청나게 크게하는 결과를 초래하였다.

1978년 이후에는 全國의 행정단위를 層化하여 多段抽出하는 표본추출계획으로 바뀌었다³⁾. 이와같이 標本設計의 기본틀 자체가 일관성이 없는 것이어서 그동안 實施 혹은 計劃되어온 국민영양조사의 調査結果는 時系列的인 分析이나 그 결과의 多面的活用이 불가능하였다.

美國의 경우 해마다 國民健康調査와 國民營養調査에 대한 보고서⁴⁾와 調査結果를 分野別로 分析한 小報告書⁵⁻⁸⁾를 함께 발표하고 있으며 日本⁹⁾도 마찬가지이다.

標本調査에 있어서는 標本調査計劃이 調査의 計劃, 調査項目의 선정, 調査方法, 調査結果등 조사의 전측면에 커다란 영향을 미치는 것이다. 이에 본 研究는 몇가지의 조사설계, 특히 國民營養調査와 국민건강조사를 검토하여 조사의 効率化를 위해서 最적의 標本抽出計劃을 設計하는 것을 目的으로 하였다.

구체적인 목적은 다음과 같다. 첫째, 家口關係調査의 標本抽出體系를 국민건강조사를 중심으로 동일한 틀을 구축하여 제시하고 둘째, 1983년도 國民營養調査를 위한 표본설계를 작성한다.

研究 方法

保健社會部가 실시하고 있는 여러 조사중 國民營養調査는 비교적 역사가 긴 統計調査이지만 方法論으로써 체계화되어 있다고 할 수 없으며, 營養調査를 통해서 확실한 食生活實態나 國民營養狀態의 實像을 파악할 수 없다는 비평도 받고 있다.

國民營養調査는 國民의 營養狀態 및 食品攝取實態와 食생활태도의 3개분야에 걸친 폭넓은 조사이다. 身體計測, 消費食品의 計測등에 必要한 장비도 있어야하고 상당량의 人的·物的資源도 必要하다. 그런데 신체계측이나 식품소비는 數量的으로 파악해야 하는 것인 반면에 食品攝取態度는 設問에 의한 質的記述로서 파악해야 하는 것이어서 調査에 따르는 方法과 기술도 다양하다고 하겠다. 따라서 國民營養調査를 위한 標本設計에 있어서는 細部的인 調査內容과 計劃, 事前에 책정된 조사에 산의 규모, 調査의 時期, 장비의 이동운반용이성, 調査戰略을 감안해야 할 것이다.

調査區가 크고 調査對象家口가 산재해 있으면 實查에서 어려움을 겪게 되는 것이 과거의 例였다¹⁰⁾. 그래서 本研究에서는 調査單位區의 규모를 적게 하므로서 자칫하면 誤差가 많아질 수 있는 食品消費量등 非標本誤差를 줄이도록 하였다.

1) 國民健康調査와 동일한 標本抽出體系

國民營養調査의 표본설계는 家口 또는 家口員을 대상으로 하는 조사의 체계화를 시도한다는 意味에서 國民健康調査¹¹⁾와 동일한 틀을 사용하였다.

國民健康調査는 行政區域單位를 標本틀로 하여 全國을 20개로 層化하여 각종에서 가구수에 比例하여 等確率로 邑面洞을 추출한 다음 이를 조사지구로 하고 里班을 단위로 하는 조사단위구를 추출한다는 體系로 되어 있다.

國民健康調査와 國民營養調査의 標本抽出體系는 Fig. 1에 제시한 바와 같이 동일한 統計標本地區(母標本)에서 第一次抽出單位地區인 읍·면·동을 國民健康調査에서는 200個地區를, 國民營養調査에서는 25個地區를 추출한다. 단 營養調査에서는 조사지구수가 적으므로 각종에서 等確率로 추출하지 못한다. 그래서 第二次抽出單位인 調査單位區(里·班)의 抽出比가 다르도록 조정되었다.

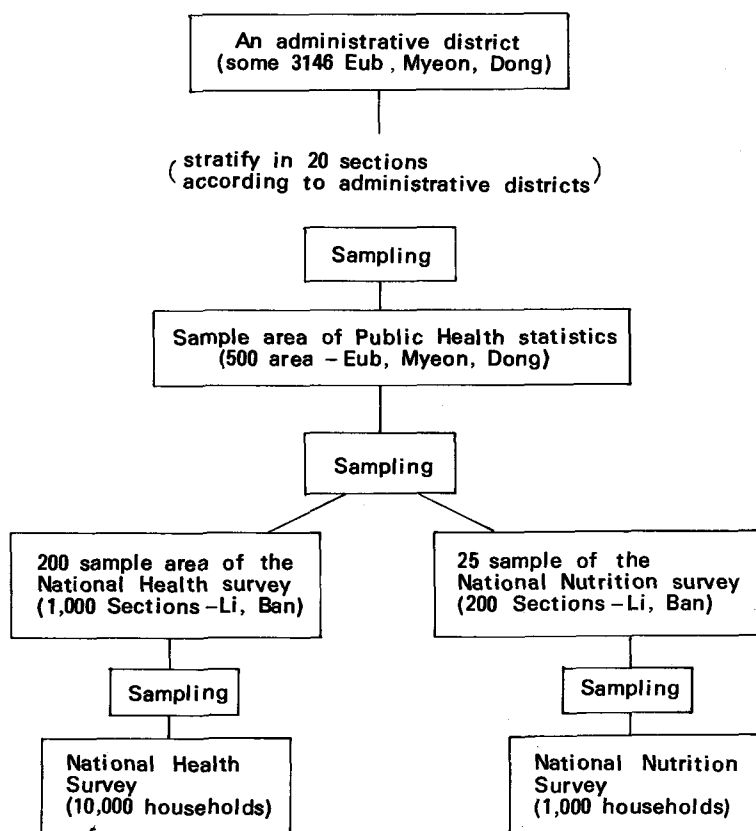


Fig. 1. A system of sample in National Health Survey and National Nutrition Survey.

2) 行政區域單位를 標本틀로 使用

邑面洞을 第一次抽出單位(P.S.U)로 하고 里班을 第二次抽出單位(S.S.U)로 하여 다시 가구를 抽出하였다. 이의 장점으로는 다음과 같은 것을 들 수 있다.

(1) 최근의 家口數와 人口를 알 수 있으므로 確率比例抽出法을 사용하여 P.S.U를 추출하고 里班을 抽出한 다음 최종 추출단위인 가구를 추출한다면 효율 높은 推定值를 얻을 수 있다. 그리고 이 경우 邑面洞은 中央에서 一覽적으로 추출하고 里班의 抽出은 基本 틀만을 제시하여 邑面洞事務所와 管轄보건소의 협조를 얻어 內容的인 정리도 할 수 있게 된다.

(2) 1個班은 約 20家口로 구성토록 최근에 재정리되었다. 郡部에 있어서는 1個面에는 10~20個의 里 또는 自然部落이 있고 1個部落은 약 20~30家口로 구성되어 있다. 따라서 이와 같은 集落을 S.S.U로 이용하는 것은 다음과 같은 이점이 따른다.

① 집락의 크기에 변동이 적어 標本誤差를 줄일 수 있

다.

② 調査員이 現地에서 조사구확인이 쉬울 뿐 아니라 地域을 대표하는 統班 및 里長의 협조를 구하기 쉽다.

③ 調査區內的 家口數가 적기 때문에 調査區 內의 調査效率를 높일 수 있다.

3) 母集團의 선정과 總抽出率의 決定

確率比例抽出에 있어서는 抽出率의 決定이 그 先行課題가 된다. 1983年度의 國民營養調查는 예산이 限定되어 있고 조사대상가구를 1,000家口로 事前에 豫定되어 있어 調査結果에 대한 정도는 調査範圍의 크기와 項目의 총합으로서 조정되어야 할 부분이 많은 것이다. 그래서 總抽出率의 決定은 조사대상가구를 1,000家口로 고정시켜 놓고 다음과 같이 하였다.

우선 常住人口調查¹²⁾의 結果를 사용하여 年間 家口數의 증가율을 算出하여 3%를 얻었다. 이를 근거로 全國의 年間 家口증가율이 3%라는 가정하에서 最近의 常住人口調査時點인 1981년 10월 1일 현재를 기점으로 調

查時點까지의 家口增加率을 계산하면 5%를 얻는다. 이를 바탕으로 逆算하면 1981년 10월 1일 현재의 常住人口調查의 結果를 모집단으로 사용하면 958 가구를 추출하므로써 조사시점에서는 표본의 크기가 약 1,000 가구가 된다. 이상과 같은 假定과 推計를 근거로 하여 本研究에서는 1981년도 상주인구조사결과를 母集團으로 하여 總抽出率을 1/8040로 固定하였다.

4) 層化 및 事後層化

層別은 四大都市(서울, 부산, 대구, 인천)는 각 도시를 각각 1개층으로 하였으며 調査豫算上 제주도를 제외한 各道는 市部와 郡部로 나누어 16個層으로 全國을 20개층으로 만들었다.

직업, 산업분포, 교육수준, 健康水準, 所得等 各種속성에 대해서는 各 邑面洞을 類別하여 副次層을 편성할 것을 고려하였으나 第一次抽出單位數가 25個로 限定되어 있어 層別로 抽出率을 다르게 해야 한다는 난점이 뒤따른다. 따라서 必要하다면 調査後 全國值 및 地域層化單位에서 각종 속성을 표식으로 한 比推定에 의한 事後層化를 하는 것도 하나의 방안이 될 것이다.

例로써 特性 Y를 갖는 性別人口數의 事後層化 또는 比推定은 다음과 같이 할 수 있다.

$$\hat{Y}_i = X_i \frac{y_i}{x_i} = X_i \frac{\sum_{j=1}^h y_{ij}}{\sum_{j=1}^h x_{ij}} = X_i \frac{\sum_{j=1}^h \sum_{k=1}^w y_{ijk}}{\sum_{j=1}^h \sum_{k=1}^w x_{ijk}}$$

- \hat{Y} : 속성 Y를 갖는 推定人口數
- X: 추정인구 또는 既調査人口
- x: 營養調査에서 파악된 人口
- y: x中 속성 Y를 갖는 人口
- i: 性別
- j: 層數(j = 1, 2, 3, ……………, h)
- k: 調査地區數(k = 1, 2, 3, ……………, w)

5) 調査地區의 抽出

最終抽出單位集落의 크기와 總抽出率이 결정되었으므로 제 1차 抽出單位區의 抽出率과 제 2차 抽出單位區의 抽出率은 어떻게 배분할 것인지가 그 다음 작업단계가 된다. 이때 고려되어야 할 것은 層化作業에 반영되지 아니한 産業別就業構造, 教育水準, 生活水準等 各種標識의 지역간 이질성과 지역내 균등성일 것이다. 邑面洞간의 異質度는 同一 邑面洞內에서의 里班간의 이질도 보다는 그 정도가 크다고 본다면, 가능한 한 P.S.U

의 추출율을 높이는 것이 代表性을 제고한다는 의미에서 바람직하다.

그러나 調査員數가 한정되어 있을 때 邑面洞數를 늘리면 旅費의 지출이 늘어나고 調査期間이 연장될 수 있을 것이다. 그래서 예산면의 제약과 實查의 효율화를 고려하고, 특히 營養調査에서 채택한 項目의 특성에 맞추어서 二次抽出率에 비중을 더 두도록 배분하였다.

調査地區의 추출은 各層(20個層)에서 邑面洞을 確率比例抽出에 의하여 추출하고 다시 추출된 邑面洞에서 m個班里를 系統的으로 無作為抽出하였다.

즉 이를 記號로 表示하면,

$$f = f_1 \cdot f_2 \cdot f_3$$

$$1/8040 = f_1 \cdot f_2 \cdot 1/Mos(\text{measurement of size})$$

$$1/Mos = f_3 = f \cdot 1/f_1 \cdot f_2$$

단, f_1 : 1차 추출율

f_2 : 2차 추출율

f_3 : 3차 추출율

f: 총 추출율

MOS_j: 근사시점에서의 班(里)의 크기

그리고 j層 k 1次抽出單位에서 추출되는 조사구수, m_{jk}는 즉,

$$m_{jk} = 1/8040 \cdot M_{jk} \cdot B_{jks} \times P_j / P_{jk}$$

단, M_{jk}: j층 k 1次抽出單位內의 2次抽出單位數

B_{jks}: j층 k 1次抽出單位中の 3 2次抽出單位內의 3次抽出單位數

P_j: j層의 크기(家口數)

P_{jk}: j층 k 1次抽出單位의 크기(家口數)

5) 推定值 및 標本誤差計算

(1) 線型推定

총 추출율을 f(f = 1/8040)라 하고 標本으로 선정된 m個의 조사지구 각각에서 얻은 調査變量 y의 總計值를 Y₁, Y₂, ……………, Y_m이라 하고 i 지구내의 최종단위개수를 N_i, 거기에서의 j 추출단위의 變量值를 y_{ij}라 하면

$$Y_i = \sum_{j=1}^{N_i} y_{ij}$$

가 되므로 y의 母集團總計值의 不偏推定值는,

$$Y = 1/f \sum_{i=1}^m Y_i$$

또 最終單位當의 母平均 μ의 不偏推定量은

$$\bar{y} = Y / \sum_{i=1}^M N_i$$

또 分散은 計算을 용이하게 하기 위해서 Paired me-

thod를 사용하면,

$$\widehat{V}(\bar{Y}) = 1/f(f-1) \sum_k \frac{w_k^2}{k} (y_{k1} - y_{k2})^2$$

가 된다.

(2) 比推定

確率變數型의 比推定은 표준오차가 적은 경우에는 不偏推定보다는 전체적으로 平均自乘誤差가 적어진다는 利點이 있다.

人口를 補助變量으로 하는 比推定에 의한 食品攝取量 및 기타 事象數의 推計值와 그 分散은 다음과 같이 계산할 수 있다.

① 特性 Y에 관한 比推定

$$\widehat{Y} = X \cdot \bar{y}_r = X \cdot \frac{\sum_{i=1}^k \sum_{j=1}^k y_{ij}}{\sum_{i=1}^k \sum_{j=1}^k x_{ij}}$$

단, \widehat{Y} : 어떤 特性 Y를 갖는 인구수의 推定值

X : 전국인구 총수

x_{ij} : j층 k 표본지역의 인구수

y_{ij} : j층 k 표본지역내의 Y특성을 갖는 인구수

그리고 \widehat{Y} 에 관한 分散은,

$$\begin{aligned} \text{Var}(\widehat{Y}) &= (1-f)/X^2 [\sum (y_{k1} - y_{k2})^2 + \bar{y}_r \cdot \sum (x_{k1} - x_{k2})^2 - 2\bar{y}_r (y_{k1} - y_{k2}) \cdot (x_{k1} - x_{k2})] \\ &\approx 1/X^2 [(y_{k1} - \bar{y}_r \cdot x_{k1}) - (y_{k2} - x_{k2})]^2 \end{aligned}$$

結果 및 考察

본 研究에서 사용된 標本抽出方法에 의하여 추출된 지역별분포는 Table 1과 같다.

그러나 조사를 원활히 실시하고 조사결과가 충분히

Table 1. The number of Sampling area and household

	Total	Metro- polis	City	Country Eub* Myeon*	
Eub* Myeon Dong	25	9	8	4	4
Li(Ban)*	100	41	21	23	15
The number of household	1,000	406	216	230	148

* The administrative unit

이용될 수 있도록 信賴度를 높이기 위해서는 標本の 抽出方法에 아직 검토되어야 할 점이 많다. 그리고 각종 統計調查의 體系化를 위한 母標本の 설정 등 행정수요의 다양화에 대처하는 標本管理體系도 개발되어야 할 과제로 남아 있으며 統計調查의 체계화에 대한 必要性을 다음과 같이 要約할 수 있겠다.

1) 각 통계조사에서 調查自體는 독립적으로 실시되고 있으나 각 통계조사대상은 相互關聯. 結合되어 있기에 각 통계조사도 상호관련·결합·종합되게끔 하나의 體系속에서 이루어져야 한다.

2) 調查의 設計, 實施 및 分析을 종합함으로써 統計의 質的向上을 도모할 수 있고 경비의 절약 및 人的資源活用的 極大化를 기할 수 있다.

3) 營養事業의 多樣化에 맞추어 統計調查의 目的도 달라져야 할 것이며, 營養事業이 복잡할수록 健全한 統計制度가 이를 뒷받침하여야 한다.

結 論

標本調查에 있어서는 標本調查計劃이 調查의 全側面에 커다란 영향을 미치는 것이다. 이에 本 研究은 1981년 경제기획원이 발간한 常住人口調查의 結果를 이용하여 1983년도 國民營養調查 實施를 위한 최적의 標本抽出計劃을 설계하였다.

調查豫算상 제주도를 除外한 全國의 家口 및 家口員을 調查의 대상으로 하여 層化多段 無作為抽出法에 의하여 선정된 調查單位區의 家口 및 家口員을 調查토록 하였다. 調查單位區 및 家口の 抽出은 다음 순서로 하였다.

1) 被調查地域을 行政區劃과 인구집중도에 의하여 20개 地域으로 層化한다.

2) 各층마다 동일추출율(1/8040)로 被調查家口數를 決定한다.

3) 被調查家口數를 荷重으로 하여 各층마다 調查區(邑面洞) 25個를 無作為로 抽出한다.

4) 抽出된 조사구마다 班을 10家口로 형성된 集落으로 보고 確率比例抽出한다. 이와같이 抽出된 班을 調查單位區로 한다.

이상의 結果로 총 對象家口 1,000 家口를 抽出하였다.

REFERENCES

1) 허 금·유정렬·이기열·성낙응·채법석·차철환

- : 국민영양보고 (1969년). 한국영양학회지 3(1): 2-43, 1970.
- 2) 보건사회부: 국민영양조사 보고. 보건사회부, 1972.
- 3) 보건사회부: 국민영양조사 보고서. 보건사회부, 1979.
- 4) National Center for Health Statistics: *Data from the National Health Examination survey and the National Health and Nutrition Examination survey(1971-1974)*. U.S. Department of Health and Human Services 11, 1982.
- 5) National Center for Health Statistics: *Total serum cholesterol Levels of Adults 18-74 Years*. U.S. Department of Health and Human Services 11(205) : 31, 1982.
- 6) National Center for Health Statistics: *Caloric and Selected Nutrient Values for Persons 1-74 Years of Age, First Health and Nutrition Examination Survey(1971-1974)*. U.S. Department of Health and Human Services 11(209): 88, 1982.
- 7) National Center for Health Statistics: *Food Consumption Profiles of White and Black Persons Aged 1-74 Years*. U.S. Department of Health and Human Services 11(210) : 103, 1980.
- 8) National Center for Health Statistics: *Weight and Height of Adults 18-74 Years of Age*. U.S. Department of Health and Human Services 11(211) : 49, 1982.
- 9) 豊川裕之·三宅由子: 日本國 食品攝取에 關한 研究. 日本公衛誌 22(10) : 571-578, 1975.
- 10) 김정근·안소영: 보건분야 통계조사의 실태와 개선책. 인구문제논집 21: 177-268, 1977.
- 11) 김정근: 건강조사의 효율화를 위한 표본설계에 관한 연구. 한국보건통계학회지 8(1) : 1-8, 1983.
- 12) 경제기획원: 상주인구조사보고서. 경제기획원, 1981