

A study on collecting representative food samples for the 10th Korean standard foods composition table

Jinheum Kim^a · Hae-Won Hwang^a · Yu Jung Cho^a · Jinwoo Park^{a,1}

^aDepartment of Applied Statistics, University of Suwon

(Received February 12, 2020; Revised March 13, 2020; Accepted March 23, 2020)

Abstract

Under Article 19, Paragraph 1 of the Food Industry Promotion Act, Rural Development Administration renews the Korean foods composition table every five years. Before the publication of the tenth revision of the Korean foods composition table in 2021, this paper suggests methods for collecting representative samples of 182 highly consumed foods in Korea. Food markets are categorized by their distribution channels, which are supermarkets and local markets. Eight samples are collected from each category by applying the National Food and Nutrient Analysis Program (NFNAP)'s stratified multi-stage sampling. The NFNAP was implemented in 1997 as a collaborative food composition research effort between the National Institute of Health (NIH) and the US Department of Agriculture (USDA) to secure reliable estimates for the nutrient content of food and beverages consumed by the US population. Selected supermarkets for selecting representative food samples are Emart Kayang, Homeplus Siheung, Lottemart Dongducheon, Emart Suwon, Lottemart Dunsan, Lottemart Yeosu, Emart Ulsan, and Hanaroclub Ulsan. Selected local markets also are Dokсандongsijang in Geumcheon-gu and Pungnapsijang in Songpa-gu, Seoul, Ilsansijang in Ilsanseo-gu, Goyang, Unamsijang in Buk-gu, Gwangju, Beopdongsijang in Daedeok-gu, Daejeon, Bongnaesijang in Yeongdo-gu and Jwadongjaeraesijang in Haeundae-gu, Busan, and Jungangsijang in Jinhae-gu, Changwon.

Keywords: food composition table, Kolmogorov-Smirnov goodness-of-fit test, local markets, probability minimum replacement probability proportional to size, supermarkets

1. 서론

농촌진흥청에서는 식품산업진흥법 제19조(식품성분 조사 등) 제1항 ‘정부는 제18조 제1항 및 제2항에 따른 식품정보 제공 등을 위하여 농수산물을 포함한 식품의 영양성분을 분석하여 식품성분표의 발간 및 식품영양학적 품질 특성 등에 대한 조사·연구를 실시할 수 있다’(<http://www.law.go.kr/법령/식품산업진흥법/>)에 의거하여 국가표준식품성분표(Korean standard foods composition table)를 5년 주기로 발표하고 있다. 국가표준식품성분표 데이터베이스(DB)에 수록되는 식품은 크게 두 가지 범주로 구분된다. 음식을 만들기 위한 농축수산물 원재료(쌀, 시금치, 콩, 계란, 소고기, 고등어 등)와 가공식품(햇

This work was carried out with the support of “Cooperative Research Program for Agriculture Science and Technology Development (Project No. PJ013397052018)” Rural Development Administration, Republic of Korea.

¹Corresponding author: Department of Applied Statistics, University of Suwon, 17, Wauan-gil, Bongdam-eup, Hwaseong-si, Gyeonggi-do 18323, Korea. E-mail: jwpark@suwon.ac.kr

반, 고등어 통조림, 단무지, 햄 등)으로 단순 형태의 식품을 의미하며 여러 재료가 섞여 조리된 음식은 해당되지 않는다. 국가표준식품성분 DB는 여러 행정부처에서 활용되고 있는데 가령 농림축산식품부에서는 식품수급 조절, 보건복지부에서는 한국인의 영양섭취 기준 설정, 교육부에서는 학생들의 단체급식 식단 작성, 식품의약품안전청에서는 제품 영양표시 등에 사용하고 있다. 기존에는 국가표준식품성분 DB 구축을 위해 최대 생산지역, 대표품종, 시장점유율 등을 고려하여 대표시료를 선정하기는 했지만 통계적 표본설계라는 과학적 방법을 적용하지 않은 채 선정하였기 때문에 대표성 문제가 제기되고 있다. 반면, 미국의 경우는 통계적 표본설계를 통해 대표성 있는 표본을 선정하여 국가표준식품성분 DB를 구축하고 있다 (Perry 등, 2000, 2003; Pehrsson 등, 2000, 2006, 2013). 우리나라에서도 합리적인 표본설계를 통해 대표성 있는 표본을 기초로 국가표준식품성분 DB를 구축할 필요가 절실하다. 본 논문에서는 1997년 미국 농무성의 Nutrient Data Laboratory (NDL)와 National Institutes of Health (NIH)가 공동으로 개발한 후 2001년과 2011년에 개정한 National Food and Nutrient Analysis Program (NFNAP)을 참고하여 우리나라 상황에 맞고 대표성을 담보할 수 있는 통계적 표본설계 방안을 제시한다.

식품영양성분분석의 국내외 기술 수준과 시장 현황은 다음과 같다. 먼저 우리나라의 경우는 농촌진흥청에서 1970년에 ‘국가표준식품성분표’ 초판을 발간한 이후로 약 5년 주기로 개정판을 발간하고 있으며 2016년까지 제9개정판이 발간되었다. ‘국가표준식품성분표 (제9개정판)’에는 한국인이 많이 소비하는 식품 3,000종에 대해 43종의 영양성분이 수록되어 있으며, 이를 누구나 편리하게 이용할 수 있도록 현재 농촌진흥청 국립농업과학원 농식품종합정보시스템(<http://koreanfood.go.kr/>)을 통해 제공하고 있다. 또한 성분분석의 신뢰도를 높이기 위해 분석품질관리 프로그램을 상시 운영하고 있고 매년 200종의 식품을 선정, 분석하여 식품성분 DB를 갱신하고 있다. 미국의 경우는 US Department of Agriculture (USDA) Nutrient Data Laboratory (NDL)에서 USDA National Nutrient DB for Standard Reference (SR) (<https://data.nal.usda.gov/>)를 제공하고 있으며, SR에는 식품 8,176종에 대해 146종의 식품성분이 수록되어 있다. 캐나다의 경우는 연방보건부에서 ‘The Canadian Nutrient File’(<https://food-nutrition.canada.ca/>)을 통해 5,807종의 식품에 대해 150종의 영양성분 정보를 제공하고 있다. 일본의 경우는 문부과학성에서 1878종의 식품에 대해 36종의 영양소를 수록한 ‘일본표준 식품성분표’를 발간하고 있다. 한편 국제적으로는 식품농업기구(Food and Agriculture Organization of the United Nations; FAO)와 유엔 대학교(United Nations University; UNU)가 양해각서를 맺고 국제식품 영양성분 데이터시스템(International Network of Food Data Systems; INFOODS)를 설립하여 양질의 식품성분 데이터 생산을 위해 노력하고 있다. 또한 INFOODS에서는 지역데이터 센터의 설립을 통해 식품성분 데이터의 국제교류를 촉진시키고 있다. 2002년(중국), 2004년(대한민국), 2007년(몽골)에서 심포지엄을 개최하여 식품성분 데이터의 질적 향상 방안을 모색하였다.

NFNAP의 주요 목적은 미국인들이 소비하는 식료품의 영양성분에 대해 신뢰성 있는 측정값을 제공하는 것이다. 이를 위해 미국 인구주택총조사의 자료와 AC Nielson의 소매점(retail outlets) 식료품 매출액 자료, 식료품 브랜드별 시장점유율 자료 등에 층화다단계추출법(stratified multi-stage sampling)을 적용하여 표본을 설계하고 그에 따라 대표시료를 추출하였다. 본 논문에서는 먼저 우리나라 식료품점을 유통형태에 따라 대형마트와 전통(재래)시장으로 구분하고 각각 층화다단계추출법을 적용하여 대표시료를 추출한다. 2절에서는 1단계 추출에서 적용할 크로미(Chromy)의 최소복원확률비례추출법(probability minimum replacement probability proportional to size; PMRPPS)에 대해 소개하고, 3절에서는 대형마트와 전통시장에 기초한 표본추출 결과를 제시한다. 4절에서는 연구결과를 요약하고 논문의 제한점에 대해 토의한다.

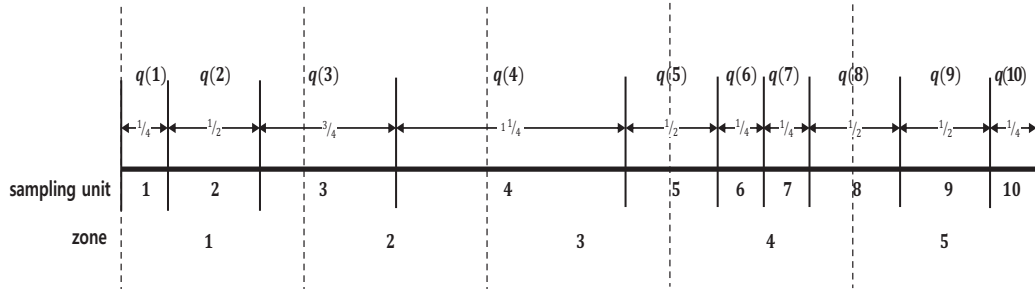


Figure 2.1. Ordered frame listing with selection zones.

Table 2.1. Cumulative sequential probabilities for cumulative sample counts

Case	$P\left(\sum_{j=1}^i n(j) = I(i) + 1 \mid \sum_{j=1}^{i-1} n(j)\right)$	
	$\sum_{j=1}^{i-1} n(j) = I(i-1)$	$\sum_{j=1}^{i-1} n(j) = I(i-1) + 1$
$F(i) = 0$	0	0
$F(i) > F(i-1) \geq 0$	$\frac{F(i)-F(i-1)}{1-F(i-1)}$	1
$F(i-1) \geq F(i) > 0$	0	$\frac{F(i)}{F(i-1)}$

2. 크로미의 PMRPPS 소개

크로미의 PMRPPS (Chromy, 1979, 1980, 1981; Williams와 Chromy)를 정의하기 위해 먼저 몇 가지 기호를 소개하자. 추출틀(sampling frame)의 크기를 N , i ($i = 1, 2, \dots, N$)번째 추출단위(sampling unit)의 보조변수 값(예, 인구수 등)을 $S(i)$, i 번째 추출단위가 표본으로 추출된 횟수를 확률변수 $n(i)$ 라고 하자. 이때 $n(i)$ 는 0 이상의 자연수를 값으로 취한다. 모든 추출단위의 보조변수 값의 총합을 $S(+)=\sum_{i=1}^N S(i)$, 총표본수를 n 이라고 하자. i 번째 추출단위가 표본으로 추출될 기대횟수를 $q(i)$ 라고 하면 $q(i)$ 는 다음과 같다.

$$q(i) = E(n(i)) = n \frac{S(i)}{S(+)}$$

크로미는 추출틀을 n 개 구역(zone)으로 나누었는데 각 구역의 크기는 동일하게 $S(+)/n$ 이 되도록 나누었다. 크로미의 PMRPPS는 원칙적으로 각 구역에서 $S(i)$ 에 비례하여 표본을 하나씩 추출하는 방법이다. 이 추출방법의 특징은 인접한 두 구역에 걸쳐 있는 추출단위는 두 구역 중에서 한 구역에서만 표본으로 추출되는데 $q(i)$ 값이 1보다 크면 인접한 두 구역에서 모두 표본으로 추출될 수 있다. Figure 2.1에서 첫 번째, 두 번째, 세 번째 추출단위는 구역 1에서 표본으로 추출될 수 있다. 세 번째 추출단위가 구역 1에서 표본으로 추출되었다면 구역 2에서는 표본으로 추출되지 않고, 구역 1에서 추출되지 않았다면 구역 2에서는 추출될 수 있다. 네 번째 추출단위는 구역 2나 3 중 적어도 한 구역에서 표본으로 추출된다. 이와 같은 크로미의 표본추출 과정을 알고리즘으로 요약하면 Table 2.1과 같다. Table 2.1에서 $I(i)$ 와 $F(i)$ 는 각각 $\sum_{j=1}^i q(j)$ 의 정수 부분과 소수 부분을 의미한다. $I(0) = F(0) = 0$ 으로 정의하자. 또한 확률 $P(\sum_{j=1}^i n(j) = I(i) \mid \sum_{j=1}^{i-1} n(j))$ 도 Table 2.1로부터 계산되는데 ‘ $\sum_{j=1}^i n(j) = I(i)$ ’인 경우는 ‘ $\sum_{j=1}^i n(j) = I(i) + 1$ ’인 경우의 여사건이기 때문이다. 크로미의 표본추출 알고리즘을 써서 Figure 2.1의 추출단위들의 누적 표본수에 대한 조건부확률을 계산하면 Table 2.2와 같다. Table 2.2를 써서 누적 표본수의 경로와 전이확률을 나타내면 Figure 2.2와 같다. 가능한 경로수는 총 79개인데 그 중 한 예로 누적 표본수의 경로가 다음과 같을 때,

sampling unit(i)	0	1	2	3	4	5	6	7	8	9	10												
cumulative sample count ($\sum_{j=1}^i n(j)$)	0	$\frac{3}{4}$	0	$\frac{1}{3}$	0	$\frac{1}{2}$	1	$\frac{1}{2}$	2	$\frac{1}{3}$	3	$\frac{2}{3}$	3	$\frac{1}{2}$	3	$\frac{1}{4}$	3	$\frac{1}{2}$	4	$\frac{1}{3}$	4	$\frac{1}{2}$	5

Table 2.2. Cumulative sequential probabilities for cumulative sample counts of sampling units in Figure 2.1

i	$q(i)$	$\sum_{j=1}^i q(j)$	$I(i)$	$F(i)$	$P\left(\sum_{j=1}^i n(j) = I(i) + 1 \mid \sum_{j=1}^{i-1} n(j)\right)$	
					$\sum_{j=1}^{i-1} n(j) = I(i-1)$	$\sum_{j=1}^{i-1} n(j) = I(i-1) + 1$
1	0.25	0.25	0	0.25	0.25	-
2	0.50	0.75	0	0.75	2/3	1
3	0.75	1.50	1	0.50	0	2/3
4	1.25	2.75	2	0.75	0.50	1
5	0.50	3.25	3	0.25	0	1/3
6	0.25	3.50	3	0.50	1/3	1
7	0.25	3.75	3	0.75	0.50	1
8	0.50	4.25	4	0.25	0	1/3
9	0.50	4.75	4	0.75	2/3	1
10	0.25	5.00	5	0.00	0	0

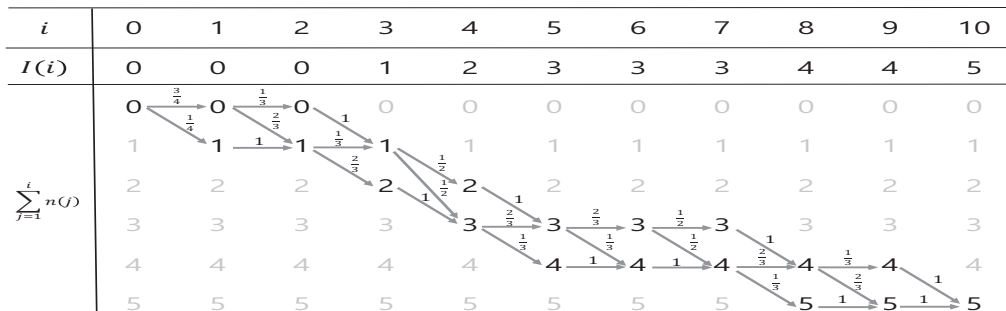


Figure 2.2. All possible paths for cumulative sample counts of sampling units in Figure 2.1.

$n(1) = n(2) = n(6) = n(7) = n(9) = 0$ 이고 $n(3) = n(4) = n(5) = n(8) = n(10) = 1$ 이므로 이 경로를 따라 추출되는 표본은 $\{3, 4, 5, 8, 10\}$ 이며 표본으로 추출될 확률은 $1/72$ 이다. 이 추출표본은 추출단위가 한 개씩 추출된 경우에 해당한다. 총 79개 가능한 표본 중에서 표본 $\{3, 4, 4, 8, 10\}$ 와 같이 추출단위 4가 중복하여 추출되는 표본은 33개이다. 이 표본은 1구역에서 추출단위 3, 2, 3구역에서 추출단위 4, 4구역에서 추출단위 8, 5구역에서 추출단위 10이 각각 추출된 것이다.

3. 유통형태에 따른 대표시료 추출

NFNAP에서는 1차 추출단위인 카운티를 추출하기에 앞서 generalized Core Based Statistical Areas (gCBSA)라는 개념을 도입했는데 CBSA에 포함된 카운티들은 gCBSA를 CBSA로 정의했으며, 그렇지 않은 카운티들은 gCBSA를 카운티 그 자체로 정의했다 (Pehrsson 등, 2000, 2013). 먼저 카운티들을 센서스 권역(region), 센서스 권역내 디비전(division), 디비전내 주(state) 순으로 나눈 뒤, 주내에서는 gCBSA의 총인구수, gCBSA 내에서는 도시성(unbarnicity)을 기준으로 정렬하였다. 센서스 권역은 west, midwest, south, northeast이며 각 권역은 다시 디비전, 주 순으로 세분된다. west 권역은 각각 8, 5개 주로 이루어진 mountain, pacific 디비전으로, midwest 권역은 각각 5, 7개 주로 이루어진 east north central, west north central 디비전으로, south 권역은 각각 9, 4, 4개 주로 이루어진 south atlantic, east south central, east south central 디비전으로, northeast 권역은 각각 6, 3개 주로 이루어진 new England, middle atlantic 디비전으로 나뉘어진다. 본 논문에서는 1차 추출단위인 시/군/구를

Table 3.1. Counts of survey foods by food groups

Food groups	cereals	potatos & starches	sugars & sweeteners	pulses	nuts & seeds	vegetables	mushrooms	fruits	meats	eggs
Count	19	3	5	8	9	37	1	25	11	3
Food groups	fishes	seaweeds	milks & milk products	oils & fats	teas	beverages	alcohols	seasoning	prepared foods	others
Count	11	5	5	2	5	3	4	13	10	3

추출하기 위해 17개 시/도의 지리적 위치와 총인구수를 고려하여 전국을 4개 권역 즉, 서울특별시와 인천광역시를 ‘서울인천권’(1권역), 경기도를 ‘경기권’(2권역), 강원도, 충청남북도, 대전광역시, 전라남북도, 광주광역시, 제주특별자치도를 ‘강원충청전라제주권’(3권역), 경상남북도, 대구광역시, 부산광역시, 울산광역시를 ‘경상권’(4권역)으로 나누었다. 이후로 권역내 시/도, 시/도내에서는 시/군/구 총인구수를 기준으로 정렬하였다. 한편 NFNAP에서는 카운티 인구수와 소매점의 식료품 매출액, 식료품 브랜드별 시장점유율로 각각 추출틀을 만들고 층화3단추출방법을 제안했는데 (Pehrsson 등, 2013) 본 논문에서는 식료품 브랜드별 시장점유율에 대한 자료 확보가 어려워 층화2단추출방법을 제안한다. Table 3.1에 정리된 것처럼 2021년 개정예정인 ‘국가표준식품성분표’ 작성을 위해 조사대상이 된 식품수는 총 182개이며 식품군으로는 총 20개이다. 본 논문에서는 식품점점을 유통형태에 따라 마트와 시장으로 구분하고자 한다. 그 이유는 조사대상 식품 중에는 마트나 시장 어느 곳에서도 채취가 가능한 식품이 있지만 어느 한 곳에서만 채취가 가능한 식품도 있기 때문이며, 동일한 식품이라 할지라도 국민들이 마트와 시장 중 어느 한 곳에서만 구매를 하지 않기 때문이다. 다만 조사 및 분석을 시행하는 농촌진흥청의 여건(시간, 예산, 성분분석기기 용량 등) 상 추출된 점포에서 다양한 식품을 채취해야 하기 때문에 대형마트와 50개 이상의 점포가 있는 전통시장으로 제한하여 조사하고자 한다.

3.1. 대형마트에 기초한 표본추출

1단계에서는 시/군/구를 추출하고, 2단계에서는 1단계에서 추출된 시/군/구 내에서 대형마트를 추출한다. 1단계에서는 대형마트가 1개 이상 있는 16개 시/도내 161개 시/군/구 총인구수(외국인 포함)로 추출틀을 구성하고, 2단계에서는 대형마트 상호와 매장면적으로 추출틀을 구성한다. 총인구수는 2015년 인구주택총조사결과를 보정한 2017년 총인구수로 정의한다(<http://www.census.go.kr/>). 2단계에서 추출단위를 대형마트로 제한한 이유는 Table 3.1에 정리된 것처럼 식품성분분석의 대상이 된 식품수가 많고 식품군도 다양하기 때문이다. 본 논문에서는 대형마트를 전국 홈플러스, 이마트, 롯데마트, 농협하나로클럽 지점으로 정의하였으며, 각 매장의 매장면적은 국토교통부 부동산정보조회시스템(<http://kras.seoul.go.kr/>)을 열람하여 수집하였다. 전국 17개 시/도별 대형마트수는 Table 3.2와 같고, 대형마트수별 시/군/구수는 Table 3.3과 같다. 세종특별자치시를 포함하여 대형마트가 없는 시/군/구를 제외했을 때 총인구수의 분포는 1권역부터 4권역까지 각각 26.4, 26.8, 22.1, 24.7%로 나타났다.

우리나라 국민들이 소비하고 있는 식품을 전국적으로 골고루 추출하기 위해 1단계에서는 USDA NDL에서 제안한 NFNAP (Pehrsson 등, 2000, 2013)를 참고하여 다음 세 가지 조건을 만족하는 표본을 추출한다.

1. 권역 및 시/도에서 골고루 추출된 표본이어야 한다.
2. 권역내 시/도 구분 없이 시/군/구 총인구수를 기준으로 정렬했을 때에도 골고루 추출된 표본이어야 한다.

Table 3.2. Counts of supermarkets (Homeplus, Emart, Lottemart, Hanaroclub) by si/do

Brand name	si/do							
	Seoul	Busan	Daegu	Incheon	Gwangju	Daejeon	Ulsan	Sejong
Homeplus	19	13	9	11	3	7	4	0
Emart	29	5	6	5	5	2	1	0
Lotemart	15	8	2	9	4	3	2	0
Hanaroclub	5	6	3	0	0	1	2	0
Total	68	32	20	25	12	13	9	0

Brand name	si/do								
	Gyeonggi	Gangwon	Chungbuk	Chungnam	Jeonbuk	Jeonnam	Gyeongbuk	Gyeongnam	Jeju
Homeplus	33	4	3	5	5	4	8	8	1
Emart	45	6	3	7	3	3	9	7	3
Lotemart	34	3	5	7	6	5	3	12	1
Hanaroclub	4	1	2	1	1	2	2	2	0
Total	116	14	13	20	15	14	22	29	5

Table 3.3. Distribution of the number of supermarkets (NoS) based on 161 si/gun/gu

NoS	1	2	3	4	5	6	7	Total
Count	42	37	40	24	14	3	1	161

3. 권역 및 시/도 구분 없이 시/군/구 총인구수를 기준으로 정렬했을 때에도 골고루 추출된 표본이어야 한다.

첫 번째 조건을 만족하도록 하기 위해 1단계 추출틀을 권역별로 정렬한 후 (1-4권역 순으로 정렬) 권역 내에서 시/도에 대한 행정구역 대분류 코드를 기준으로 시/도를 정렬한다. 시/도 내에서는 시/군/구 총인구수를 기준으로 서편타인 정렬(serpentine sort)한다 (Chromy, 1979, 1981; Williams와 Chromy, 1980). 서편타인 정렬이란 오름차순과 내림차순 정렬을 교대로 하는 것을 말한다. 따라서 서편타인 정렬은 인접한 시/도 내에서 총인구수가 가장 많거나 혹은 가장 적은 시/군/구가 서로 이웃하게 정렬되는 특징이 있다. 16개 시/도를 서울특별시(대분류 코드: 11), 인천광역시(23)(이상 1권역), 경기도(31)(2권역), 광주광역시(24), 대전광역시(25), 강원도(32), 충청북도(33), 충청남도(34), 전라북도(35), 전라남도(36), 제주특별자치도(39)(이상 3권역), 부산광역시(21), 대구광역시(22), 울산광역시(26), 경상북도(37), 경상남도(38)(이상 4권역)의 순서로 나열한 후, 총인구수를 기준으로 서울특별시내 23개 시/군/구를 오름차순, 인천광역시내 7개 시/군/구를 내림차순으로 정렬한다. 이와 같은 요령으로 정렬을 계속 진행하여 맨 마지막으로 경상남도내 12개 시/군/구를 내림차순으로 정렬하여 Table 3.4와 같은 1단계 추출틀(‘§’로 표시)을 얻었다. 크로미의 PMRPPS를 써서 1단계 추출틀로부터 후보표본(candidate sample)을 추출한다. 이때 각 권역에서 표본이 2개씩 추출되도록 표본의 크기 n 은 8로 설정한다.

한편 첫 번째 조건을 만족하는 후보표본이 두 번째 조건을 얼마나 잘 만족하는지를 알아보기 위해 먼저 시/도를 구분하지 않고 권역 내에서 시/군/구 총인구수를 기준으로 서편타인 정렬을 한 후 누적 총인구수의 비율을 구하고 이를 161개 시/군/구 이상적인 모집단 분포함수로 정의한다. 또한 후보표본을 이상적인 모집단의 시/군/구 순서대로 정렬한 후 누적 총인구수의 비율을 후보표본의 표본 분포함수로 정의한다. 1단계 후보표본의 분포가 두 번째 조건에 따라 정의된 이상적인 모집단의 분포와 동일 한지를 검정하기 위해 일표본 콜모고로프-스미르노프(Kolmogorov-Smirnov) 적합도 통계량을 계산한다 (Massey, 1951). 마찬가지로 1단계 후보표본이 세 번째 조건을 얼마나 잘 만족하는지를 알아보기 위

Table 3.4. Sampling frame in the first stage when the secondary sampling units are supermarkets (§) or local markets (†)

si/do	si/gun/gu	popula- tion	si/do	si/gun/gu	popla- tion	si/do	si/gun/gu	popula- tion	
Seoul	Jung ^{§,†}	127896		Jangan ^{§,†}	295436		Sebuk ^{§,†}	382455	
	Jongno [†]	157277		Dongan ^{§,†}	322590		Chungnam	Asan ^{§,†}	338473
	Yongsan ^{§,†}	223898		Gwangmyeong ^{§,†}	328253			Dongnam ^{§,†}	273839
	Geumcheon ^{§,†}	249930		Suji [§]	338993			Seosan ^{§,†}	171364
	Seongdong ^{§,†}	302367		Gwangju [§]	341046			Dangjin [§]	168508
	Gangbuk ^{§,†}	313698		Yeongtong ^{§,†}	350818			Nonsan ^{§,†}	123774
	Dobong ^{§,†}	332586		Danwon [§]	354452			Hongseong ^{§,†}	103048
	Dongdaemun ^{§,†}	357380		Gwonseon ^{§,†}	364650			Boryeong ^{§,†}	100922
	Gwangjin ^{§,†}	363934		Sangnok ^{§,†}	379209			Yesan [†]	78678
	Mapo ^{§,†}	368841		Gimpo ^{§,†}	392438			Buyeo [†]	67028
	Yeongdeungpo ^{§,†}	393560		Giheung [§]	416746			Geumsan [†]	55672
	Jungnang ^{§,†}	396892		Uijeongbu ^{§,†}	426530			Seocheon [†]	53321
	Dongjak ^{§,†}	400236		Paju ^{§,†}	430650			Gyeryong [§]	41858
	Seocho ^{§,†}	414550		Deogyang ^{§,†}	433606		Jeonbuk	Buan [†]	51249
	Gangdong ^{§,†}	423978		Siheung ^{§,†}	446369			Gochang [†]	54947
	Guro ^{§,†}	436869		Bundang ^{§,†}	475778			Namwon ^{§,†}	79282
	Seongbuk ^{§,†}	445417		Pyeongtaek ^{§,†}	486077			Gimje [§]	82358
	Yangcheon ^{§,†}	452111		Namyangju [§]	644936			Jeongeup ^{§,†}	109263
Eunpyeong ^{§,†}	466243	Hwaseong [§]	708569		Gunsan ^{§,†}	273649			
Gwanak ^{§,†}	511222	Bucheon ^{§,†}	845514		Deokjin ^{§,†}	294824			
Gangnam ^{§,†}	522514	Gwangju	Buk ^{§,†}	450323		Iksan ^{§,†}	298928		
Nowon ^{§,†}	543499		Gwangsan ^{§,†}	419625		Wansan ^{§,†}	362132		
Gangseo ^{§,†}	581675		Seo ^{§,†}	306054	Jeonnam	Yeosu ^{§,†}	271741		
Songpa ^{§,†}	633953		Nam ^{§,†}	219256		Suncheon [§]	266809		
Incheon	Bupyeong ^{§,†}	538236		Dong ^{§,†}	100914		Mokpo ^{§,†}	233948	
	Namdong ^{§,†}	534050	Daejeon	Daedeok [†]	188469		Gwangyang ^{§,†}	144010	
	Seo ^{§,†}	509776		Dong ^{§,†}	239752		Naju [§]	105579	
	Nam ^{§,†}	421851		Jung ^{§,†}	243540		Muan [§]	81420	
	Yeonsu ^{§,†}	337595		Yuseong ^{§,†}	366639		Goheung [†]	60984	
	Gyeyang ^{§,†}	318104		Seo ^{§,†}	487449		Wando [†]	48332	
	Jung ^{§,†}	115153	Gangwon	Wonju ^{§,†}	341338		Boseong [†]	39942	
Gyeonggi	Yeoncheon [§]	43990		Chuncheon ^{§,†}	283742	Jeju	Seogwipo ^{§,†}	169358	
	Gwacheon [†]	53583		Gangneung ^{§,†}	215914		Jeju ^{§,†}	472399	
	Gapyeong [†]	59548		Donghae ^{§,†}	89151	Busan	Haeundae ^{§,†}	399363	
	Dongducheon ^{§,†}	97373		Sokcho ^{§,†}	78618		Busanjin ^{§,†}	366492	
	Yangpyeong ^{§,†}	109359		Samcheok ^{§,†}	67860		Saha ^{§,†}	330293	
	Yeoju ^{§,†}	111476		Hongcheon [†]	67168		Buk ^{§,†}	296945	
	Uiwang ^{§,†}	151852		Taebaek ^{§,†}	44733		Nam ^{§,†}	275941	
	Pocheon ^{§,†}	161239		Pyeongchang [†]	40264		Dongnae ^{§,†}	261671	
	Guri ^{§,†}	189741		Yeongwol [†]	36934		Geumjeong ^{§,†}	246891	
	Anseong ^{§,†}	198832		Jeongseon [†]	36002		Sasang ^{§,†}	231587	
	Paldal ^{§,†}	206092		Goseong [†]	28621		Yeonje ^{§,†}	201053	
	Yangju [§]	212631		Yangyang [†]	25556		Suyeong [†]	173059	
	Icheon [§]	218798		Hwacheon [†]	24625		Gijang ^{§,†}	156561	
	Hanam ^{§,†}	218994		Yanggu [†]	22355		Yeongdo ^{§,†}	123531	
	Osan [§]	219877	Chungbuk	Jeungpyeong [†]	37625		Gangseo [†]	116702	
	Jungwon [†]	226657		Jecheon ^{§,†}	136929		Seo [†]	106924	
	Cheoin [§]	240521		Sangdang ^{§,†}	166664		Dong [†]	86106	
	Sujeong ^{§,†}	241862		Cheongwon ^{§,†}	196192		Jung ^{§,†}	43799	
	Gunpo [§]	280913		Chungju ^{§,†}	212274	Daegu	Jung [†]	75324	
	Ilsandong [§]	284692		Seowon ^{§,†}	221147		Nam ^{§,†}	152785	
Ilsanseo ^{§,†}	286804		Heungdeok ^{§,†}	258308		Seo ^{§,†}	191414		

(continued)

(continued)

si/do	si/gun/gu	popula- tion	si/do	si/gun/gu	popla- tion	si/do	si/gun/gu	popula- tion
	Dalseong ^{§,†}	239240		Sangju [§]	98189		Seongsan ^{§,†}	226726
	Dong ^{§,†}	340396		Yeongju ^{§,†}	108443		Masanhoewon ^{§,†}	202864
	Suseong ^{§,†}	425546		Gimcheon [§]	140765		Jinhae ^{§,†}	186658
	Buk ^{§,†}	445293		Andong ^{§,†}	165704		Masanhappo ^{§,†}	178751
	Dalseo ^{§,†}	583043		Nam ^{§,†}	239044		Tongyeong ^{§,†}	132962
Ulsan	Nam ^{§,†}	333117		Gyeongju [§]	262264		Sacheon ^{§,†}	111655
	Jung ^{§,†}	233343		Buk ^{§,†}	267865		Miryang [§]	103894
	Buk ^{§,†}	196147		Gyeongsan ^{§,†}	282626		Haman [†]	67906
	Dong ^{§,†}	172306		Gumi ^{§,†}	423003		Changnyeong [†]	62544
Gyeongbuk	Gunwi [§]	22230	Gyeongnam	Gimhae ^{§,†}	536456		Geochang [†]	60083
	Yeongdeok [†]	36148		Jinju ^{§,†}	353209		Goseong [†]	52439
	Yecheon [†]	46321		Yangsan ^{§,†}	332596		Hadong [†]	43441
	Mungyeong ^{§,†}	70016		Geoje ^{§,†}	257847		Namhae [†]	42986
	Yeongcheon [§]	97675		Uichang ^{§,†}	250560			

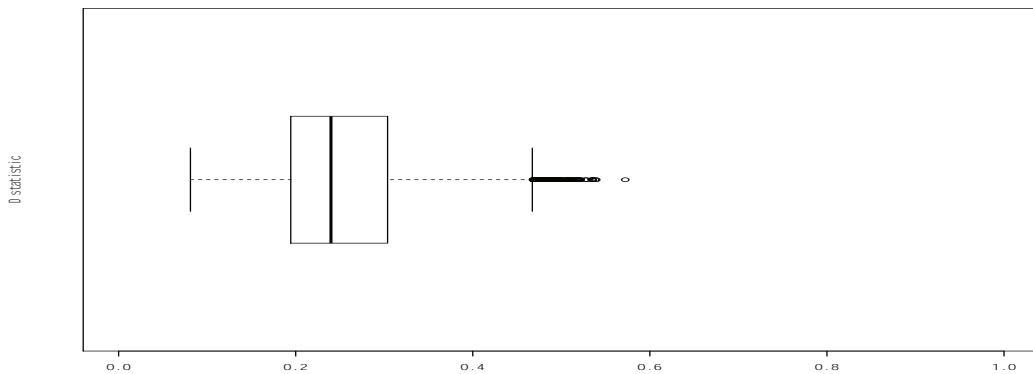


Figure 3.1. Boxplot of D -statistic based on 20,000 candidate samples when the secondary sampling units are supermarkets.

해 권역 및 시/도를 구분하지 않고 시/군/구 총인구수를 기준으로 서편타인 정렬한 후 누적 총인구수의 비율을 구하고 이를 161개 시/군/구 이상적인 모집단 분포함수로 정의한다. 또한 후보표본을 이상적인 모집단의 시/군/구 순서대로 정렬한 후 누적 총인구수의 비율을 후보표본의 표본 분포함수로 정의한다. 1단계 후보표본의 분포가 세 번째 조건에 따라 정의된 이상적인 모집단의 분포와 동일한지를 검정하기 위해 콜모고로프-스미르노프 적합도 통계량을 계산한다. 1단계 후보표본의 통합 통계량 값(D)은 두 콜모고로프-스미르노프 통계량 중에서 더 큰 값으로 정의한다. 첫 번째 조건을 만족하는 1단계 후보표본을 추출하고 상술한 것처럼 D 통계량을 구하는 과정을 반복할 때 1단계 최종표본(final sample)은 가장 작은 D 값을 갖는 후보표본으로 정의한다. 본 논문에서는 20,000번 반복하였으며 D 값의 분포는 Figure 3.1과 같다. Figure 3.2는 두 번째(왼쪽 그림)와 세 번째(오른쪽 그림) 조건 하에서 최종표본의 콜모고로프-스미르노프 통계량 값과 표본 분위수 대 이상적인 모집단의 분위수(sample quantile vs. ideal sample quantile, Q-Q)를 나타낸 그림이다. 20,000개 후보표본 중에서 1단계 최종표본은, 서울특별시 강서구, 금천구(이상 1권역), 동두천시, 수원시 권선구(이상 2권역), 대전광역시 서구, 여수시(이상 3권역), 울산광역시 남구, 북구(이상 4권역)로 나타났다 ($D = 0.0811$). 최종표본을 살펴보면 각 권역에서 추출단위가 2개씩 추출되었는데 Table 3.5에서 볼 수 있듯이 이와 같은 유형의 후보표본은 20,000개 후보표본 중에서 65.3%를 차지하였다.

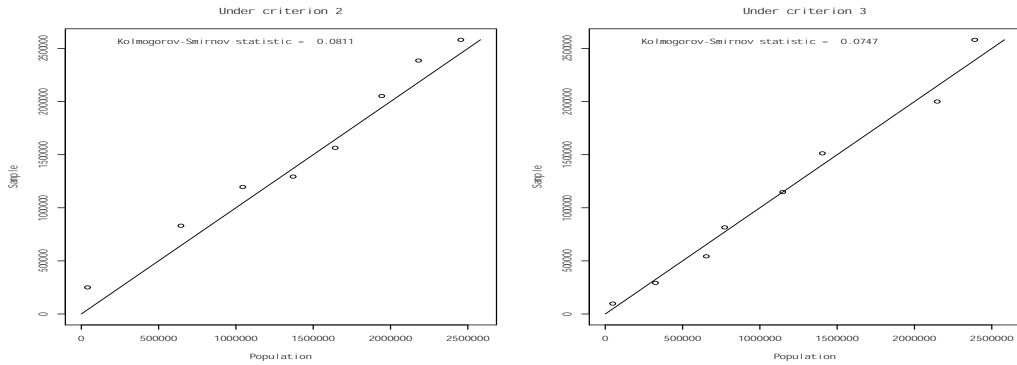


Figure 3.2. Q-Q plot of sample vs. ideal sample for si/gun/gu population size by region (left) and overall si/gun/gu population size Q-Q plot of sample vs. ideal sample (right) when the secondary sampling units are supermarkets.

Table 3.5. Percentage of pairs of (the number of samples from region 1, the number of samples from region 2, the number of samples from region 3, the number of samples from region 4) based on 20,000 candidate samples when the secondary sampling units are supermarkets

Pair type	(2, 2, 2, 2)	(2, 2, 3, 1)	(2, 3, 1, 2)	(2, 3, 2, 1)	(3, 1, 2, 2)	(3, 1, 3, 1)	(3, 2, 1, 2)	(3, 2, 2, 1)
Percentages(%)	65.3	1.5	21.6	0.5	8.0	0.2	2.8	0.1

Table 3.6. Sampling frame in the second stage based on the first sampling units; Kanfseq-gu, Geumcheon-gu (Seoul-si), Dongducheon-si, Gwonseon-gu (Gyeonggi-do), Seo-gu (Daejeon-si), Yeosu-si (Jeollanam-do), Nam-gu, Buk-gu (Ulsan-si)

si/do	si/gun/gu	Area (m ²)	Market name	si/do	si/gun/gu	Area (m ²)	Market name	
Seoul	Kangseo	9895	Homeplus Kangseo	Daejeon	Seo	6610	Homeplus Tanbang	
		6722	Homeplus Kayang			10721	Homeplus Dunsan	
		19644	Lottemart			6560	Emart Dunsan	
			Kimpoairport					
		137754	Emart Kayang					
	Geumcheon	16813	Lottemart Geumcheon	Jeonnam	Yeosu	4207	Lottemart Yecheon	
		9583	Homeplus Geumcheon			3829	Emart Yeosu	
		69423	Homeplus Siheung			8147	Lottemart Yeosu	
	Gyeonggi	Dongducheon	107623	Lottemart Dongducheon	Ulsan	Nam	6610	Lottemart Ulsan
							11575	Homeplus Namgu
Gwonseon		169423	Hanaroclub Suwon	11575			Homeplus Namgu	
		8001	Homeplus Seosuwon	1961			Hanaroclub Okdong	
		32060	Lottemart Suwon	9898			Emart Ulsan	
		6687	Lottemart Gwonseon	Buk		23586	Hanaroclub Ulsan	
		15176	Emart Suwon			8007	Homeplus Bukgu	
		12429	Emart Seosuwon			13747	Lottemart Jinjang	

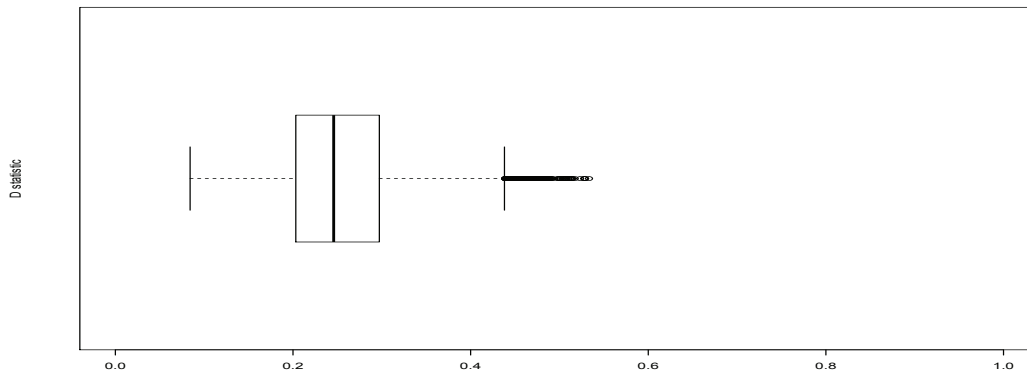
2단계 추출틀은 Table 3.6과 같이 1단계에서 최종표본으로 결정된 시/군/구에 있는 대형마트의 매장면적(단위: m²)으로 구성한다. 1단계에서 최종표본으로 추출되지 않은 시/군/구에 있는 대형마트의 매장면적에 대한 자료도 확보되어 있어 1단계에서 추출된 표본에 무관하게 항상 2단계 추출틀을 구성할

Table 3.7. The total number of local markets with more than 50 stores by si/do

si/do	Seoul	Busan	Daegu	Incheon	Gwangju	Daejeon	Ulsan	Sejong	
Count	136	96	62	28	17	18	20	0	
si/do	Gyeonggi	Gangwon	Chungbuk	Chungnam	Jeonbuk	Jeonnam	Gyeongbuk	Gyeongnam	Jeju
Count	99	37	25	28	18	20	38	68	8

Table 3.8. Distribution of the number of local markets (NoLM) based on 172 si/gun/gu

NoLM	1	2	3	4	5	6	7	8	9	10	11	12	13	14	18	Total
Count	38	34	18	16	14	18	14	6	3	2	2	1	3	1	2	172

**Figure 3.3.** Boxplot of D -statistic based on 20,000 candidate samples when the secondary sampling units are local markets.

수 있다. 2단계에서는 1단계 최종표본에 포함된 시/군/구 추출틀에 각각 PPS를 적용하여 1개씩 총 8개 표본을 추출한다. 따라서 최종적으로 추출된 2차 추출단위는 이마트 가양점, 홈플러스 시흥점(이상 1권역), 롯데마트 동두천점, 이마트 수원점(이상 2권역), 롯데마트 둔산점, 롯데마트 여수점(이상 3권역), 이마트 울산점, 하나로클럽 울산점(이상 4권역)으로 나타났으며, 성분분석을 위한 식료품의 대표시료는 추출된 대형마트내 식료품 코너에서 Table 3.1에 포함된 식료품의 브랜드별로 채취하여 얻었다.

3.2. 전통시장에 기초한 표본추출

1단계에서는 3.1절과 마찬가지로 시/군/구를 추출하고, 2단계에서는 1단계에서 추출된 시/군/구 내에서 전통시장을 추출한다. 1단계에서는 전국 17개 시/도내 250개 시/군/구 중 농수산물을 포함한 식품 등을 판매하는 전통시장이 1개 이상 있는 16개 시/도내 172개 시/군/구 총인구수로 추출틀을 구성하고, 2단계에서는 전통시장 상호와 전통시장내 점포수로 추출틀을 구성한다. 전통시장에 대한 정보(자료 수집 기간: 2015.1.1-2019.10.31)는 공공데이터포털(<http://www.data.go.kr/>)에서 수집하였다. 전국 17개 시/도별 50개 이상 점포로 이루어진 전통시장수는 Table 3.7과 같고, 전통시장수별 시/군/구수는 Table 3.8과 같다. 세종특별자치시를 포함하여 점포수가 50개 이상 있는 전통시장이 없는 시/군/구를 제외했을 때 총인구수의 분포는 1권역부터 4권역까지 각각 27.7, 20.3, 24.5, 27.5%로 나타났다.

1단계 추출에서 얻어지는 표본이 3.1절에서 제시한 첫 번째 조건을 만족하도록 하기 위해 3.1절처럼 1-4권역 순서로 나열하고 권역 내에서 행정구역 대분류 코드를 기준으로 시/도를 오름차순으로 나열한 후 시/군/구 총인구수를 기준으로 서편타인 정렬하여 Table 3.4와 같은 1단계 추출틀(‘1’로 표시)을 얻었

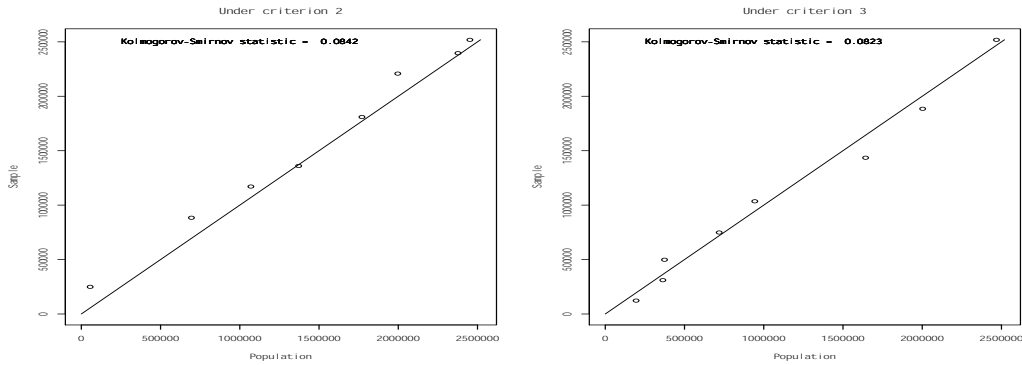


Figure 3.4. Q-Q plot of sample vs. ideal sample for si/gun/gu population size by region (left) and overall si/gun/gu population size Q-Q plot of sample vs. ideal sample (right) when the secondary sampling units are local markets.

Table 3.9. Percentage of pairs of (the number of samples from region 1, the number of samples from region 2, the number of samples from region 3, the number of samples from region 4) based on 20,000 candidate samples when the secondary sampling units are local markets

Pair type	(2, 1, 2, 3)	(2, 1, 3, 2)	(2, 2, 1, 3)	(2, 2, 2, 2)	(3, 0, 2, 3)	(3, 0, 3, 2)	(3, 1, 1, 3)	(3, 1, 2, 2)
Percentage(%)	2.5	10.2	12.9	52.9	0.7	2.6	3.6	14.6

Table 3.10. Sampling frame in the second stage based on the first sampling units; Geumcheon-gu, Songpa-gu (Seoul-si), Ilsanseo-gu (Gyeongg-do), Buk-gu (Gwangju-si), Daedeok-gu (Daejeon-si), Haeundae-gu, Yeongdo-gu (Busan-si), Jinhae-gu (Gyeongsangnam-do)

si/do	si/gun/gu	No. of stores	Sijang name	si/do	si/gun/gu	No. of stores	Sijang name
Seoul	Geumcheon	118	Nammun	Daejeon	Daedeok	62	Beopdong
		244	Hyundai			122	Sintanjin
		233	Daemyungyeolbit street	Busan	Haeundae	143	Banyeo
		179	Doksan Dongwoo			232	Jaesong
	64	Eunhangnamu	104			Haeundae	
	72	Macheon	125			Jwadongjaerae	
	Songpa	136	Macheonjungang	66	Bansonggolmok		
		135	Bangi	198	Bansong2dong		
		116	Saemaaul	60	Banyeo3donggolmok		
		70	Seokchon	70	Banyeo2dongsangga		
165		Pungnap	Yeongdo	272	Namhang		
Gyeonggi	Ilsanseo	88	Ilsan			140	Bongrae
Gwangju	Buk	54	Unam			101	Cheonghak
Daejeon	Daedeok	73	Songchon	Gyeongnam	Jinhae	279	Jungang
		213	Jungri				

다. 이 추출틀에 크로미의 PMRPPS를 적용하여 표본의 크기 $n = 8$ 인 후보표본을 추출한다. 이 후보표본이 3.1절에서 제시한 두 번째와 세 번째 조건을 얼마나 잘 만족하는지를 알아보기 위해 3.1절 처럼 후보표본에 대한 D 통계량 값을 계산한다. 이와 같은 과정을 20,000번 반복하여 얻은 D 값의 분포는 Figure 3.3과 같으며 가장 작은 D 값에 해당하는 후보표본을 1단계 최종표본으로 결정한다.

Figure 3.4는 3.1절에서 언급한 두 번째(왼쪽 그림)와 세 번째(오른쪽 그림) 조건 하에서 최종표본의 콜모고로프-스미르노프 통계량 값과 Q-Q 그림을 나타낸 것이다. 20,000개 후보표본 중에서 1단계 최종표본은, 서울특별시 금천구, 송파구(이상 1권역), 고양시 일산서구(2권역), 광주광역시 북구, 대전광역시 대덕구(이상 3권역), 부산광역시 해운대구, 영도구, 창원시 진해구(이상 4권역)로 결정하였다 ($D = 0.0842$). 1, 3권역에서는 기대했던 대로 시/군/구가 2개씩 추출되었지만 2권역에서는 기대했던 것보다 1개가 적고 4권역에서는 1개가 많았는데 이는 각 권역의 총인구수 비율이 동일하지 않기 때문이다. Table 3.10에 정리된 것처럼 20,000개 후보표본 중에서 권역별로 2개씩 고르게 추출된 후보표본의 비율은 52.9%로 가장 높았으며, 최종표본과 같은 유형의 후보표본은 12.9%를 차지하였다.

2단계 추출틀은 Table 3.4와 같이 1단계에서 최종표본으로 결정된 시/군/구에 있는 전통시장의 이름과 전통시장내 점포수로 구성한다. 1단계에서 최종표본으로 추출되지 않은 시/군/구에 있는 전통시장에 대한 정보도 확보되어 있어 1단계에서 추출된 표본에 무관하게 항상 2단계 추출틀을 구성할 수 있다. 2단계에서는 1단계 최종표본에 포함된 시/군/구에서 PPS를 적용하여 표본을 1개씩 총 8개 표본을 추출한다. 따라서 최종적으로 추출된 2차 추출단위는, 서울특별시 금천구 독산동우시장과 송파구 풍납시장(이상 1권역), 고양시 일산서구 일산시장(2권역), 광주광역시 북구 운암시장, 대전광역시 대덕구 법동시장(이상 3권역), 부산광역시 영도구 봉래시장과 해운대구 좌동재래시장, 창원시 진해구 중앙시장(이상 4권역)으로 나타났으며, 성분분석을 위한 식료품의 대표시료는 추출된 전통시장내 식료품을 취급하는 모든 소매점에서 Table 3.1에 포함된 식료품의 브랜드별로 채취하여 얻었다.

4. 결론

농촌진흥청에서는 식품산업진흥법 제19조 제1항에 의거하여 국가표준식품성분표를 5년 주기로 발표하고 있다. 본 논문에서는 2021년 국가표준식품성분표 제10개정판 발간을 앞두고 우리나라 국민들이 많이 소비하고 있는 식품의 대표시료를 합리적이고 과학적으로 추출하기 위한 방안을 제시하였다. 제10개정판 작성을 위해 대상이 된 식품은 총 182개이며 식품군으로는 총 20개이다. 식료품점을 유통형태에 따라 대형마트와 전통시장으로 구분한 후 NFNAP에서 제안한 층화다단계추출법을 각각 적용하여 표본을 추출하였다. 대형마트에 기초한 표본추출에서는 161개 시/군/구 총인구수로 이루어진 1차 추출틀에 크로미의 PMRPPS 방법을 적용하여 시/군/구를 8개 추출한 후, 시/군/구에 있는 대형마트의 매장면적으로 이루어진 2차 추출틀에 각각 PPS를 적용하여 1개씩 총 8개 표본을 추출하였다. 최종적으로 추출된 대형마트는 이마트 가양점, 홈플러스 시흥점, 롯데마트 동두천점, 이마트 수원점, 롯데마트 둔산점, 롯데마트 여수점, 이마트 울산점, 하나로클럽 울산점으로 나타났다. 또한 전통시장에 기초한 표본추출에서는 172개 시/군/구 총인구수로 이루어진 1차 추출틀에 크로미의 PMRPPS 방법을 적용하여 시/군/구를 8개 추출한 후, 시/군/구에 있는 전통시장의 점포수로 이루어진 2차 추출틀에 각각 PPS를 적용하여 1개씩 총 8개 표본을 추출하였다. 최종적으로 추출된 전통시장은 서울특별시 금천구 독산동우시장과 송파구 풍납시장, 고양시 일산서구 일산시장, 광주광역시 북구 운암시장, 대전광역시 대덕구 법동시장, 부산광역시 영도구 봉래시장과 해운대구 좌동재래시장, 창원시 진해구 중앙시장으로 나타났다. 성분분석을 위한 대표시료는 추출된 대형마트내 식료품 코너와 전통시장내 식료품을 취급하는 모든 소매점에서 Table 3.1에 포함된 식료품의 브랜드별로 채취하여 얻었다. 1차 추출에서 적용한 PMRPPS는 PPS의 일종인데 조사변수인 총인구수를 기준으로 권역별, 시/도별로 추출단위를 서번타인 정렬한 후 추출틀을 균등하게 분할하여 각 구역에서 표본을 하나씩 추출하기 때문에 PPS보다 골고루 표본을 얻을 수 있는 장점이 있다. 본 논문에 수록하지는 않았지만 시/군/구 인구수를 조사변수로 두고 20,000개 후보표본을 각각 PPS와 PMRPPS로 추출하여 평균제곱오차를 비교해본 결과, 2차 추출단위가 대형마트의 경우는 16개 시/도 중 10개 시/도에서 PMRPPS가 더 우수하였으며, 전통시

장의 경우는 12개 시/도에서 PMRPPS가 더 우수하였다. 본 논문의 1차 추출틀을 보면 전국 250개 시/군/구 중 대형마트의 경우는 161개 시/군/구, 전통시장의 경우는 172개 시/군/구만을 포함하고 있어 각각 89개, 78개 시/군/구가 제외되어 미포함(undercoverage) 문제가 생길 수 있다. 하지만 두 추출틀을 통합하면 197개 시/군/구를 포함하며 인구수 기준으로는 포함률(coverage rate)이 94%가 되기 때문에 다양한 식품종을 구입하기 어려운 소규모 마트나 시장을 조사대상으로 삼을 수 없다는 현실적인 제약을 극복하면서도 대표성을 지닌 추출틀이라고 생각된다.

본 논문은 다음과 같은 제한점이 있다. 첫째, 2017년도에 실시한 전국사업체조사(<http://kosis.kr>)에서는 전국 250개 시/군/구 슈퍼마켓수와 종사자수에 대한 정보만 제공하고 있어 일반 슈퍼마켓이나 중형급 마트의 명부(상호명, 소재지, 매장면적 등)는 확보하기가 어려워 식료품의 유통형태가 마트인 경우 추출틀을 대형마트로만 제한하였다. 둘째, 식품 및 영양성분별로 대표성을 충분히 확보하기 위해서는 유통형태와 지역뿐만 아니라 NFNAP (Pehrsson 등, 2000, 2013)처럼 식료품점 매출액, 식료품 브랜드별 시장점유율 등으로 더욱 자세히 나누어 시료를 확보하는 것이 바람직하지만 시간, 예산, 성분분석 여건 등의 제약으로 인해 제한된 범위 내에서의 대표성을 담보하기 위한 방안을 제시하였다. 셋째, 심사위원이 제안한 것처럼 향후 연구에서는 Stevens와 Olsen (2004)이 제안한 공간균등추출(spatially balanced sampling) 방법과 PMRPPS 방법의 효율을 비교하고자 한다.

References

- Chromy, J. R. (1979). Sequential sample selection methods, *Proceedings of the American Statistical Association, Section on Survey Research Methods*, 401-406.
- Chromy, J. R. (1981). Variance estimators for a sequential sample selection procedure, *Current Topics in Survey Sampling (edited by D. Krewski, R. Platek, and J.N.K. Rao)*, *Proceedings of the International Symposium on Survey Sampling held in Ottawa*, Academic Press, New York, 329-347.
- Massey, F. J. (1951). The Kolmogorov-Smirnov test for goodness of fit, *Journal of the American Statistical Association*, **6**, 68-78.
- Pehrsson, P. R., Haytowitz, D. B., Holden, J. M., Perry, C. R., and Beckler, D. G. (2000). USDA's national food and nutrient analysis program: food sampling, *Journal of Food Composition and Analysis*, **12**, 379-389.
- Pehrsson, P. R., Perry, C. R., Cutrufelli, R. C., et al. (2006). Sampling and initial findings for a national study of fluoride in drinking water, *Journal of Food Composition and Analysis*, **19**, S45-S52.
- Pehrsson, P. R., Perry, C. R., and Daniel, M. (2013). ARS, USDA updates for food sampling strategies to keep pace with demographic shifts, *Procedia Food Science*, **2**, 52-59.
- Perry, C. R., Beckler, D. G., Pehrsson, P., and Holden, J. (2000). A national sampling plan for obtaining food products for nutrient analysis, *Proceedings of the American Statistical Association, Section on Survey Research Methods*, 267-272.
- Perry, C. R., Pehrsson, P. R., and Holden, J. M. (2003). A revised sampling plan for obtaining food products for nutrient analysis. In *Proceedings of the American Statistical Association, Section on Survey Research Methods*, CD-ROM.
- Stevens, D. L. and Olsen, A. R. (2004). Spatially balanced sampling of natural resources, *Journal of the American Statistical Association*, **99**, 262-278.
- Williams, R. L. and Chromy, J. R. (1980). SAS sample selection macros. In *Proceedings of the Fifth Annual SAS Users Group International Conference*, 392-396.

국가표준식품성분 데이터베이스 대표시료 선정을 위한 표본설계

김진흠^a · 황해원^a · 조유정^a · 박진우^{a,1}

^a수원대학교 데이터과학부

(Received February 12, 2020; Revised March 13, 2020; Accepted March 23, 2020)

요약

농촌진흥청에서는 식품산업진흥법 제19조 제1항에 의거하여 국가표준식품성분표를 5년 주기로 발표하고 있다. 본 논문에서는 2021년 국가표준식품성분표 제10개정판 발간을 앞두고 우리나라 국민들이 많이 소비하고 있는 식품으로 선정된 182개 식품의 대표시료를 합리적이고 과학적으로 추출하기 위한 방안을 제안하였다. 농수산물에 포함된 식품을 판매하는 식료품점을 유통형태에 따라 대형마트와 전통시장으로 구분한 후 NFNAP에서 제안한 층화다단계추출법을 각각 적용하여 8개 표본을 추출하였다. NFNAP는 미국인들이 소비하고 있는 식료품의 성분표에 대한 신뢰성 있는 추정을 확보하기 위해 미국 농무성과 국립보건원이 1977년에 협약을 맺고 공동 연구로 개발한 국가식품 영양분석 프로그램이다. 대형마트에 기초한 표본추출에서는 이마트 가양점, 홈플러스 시흥점, 롯데마트 동두천점, 이마트 수원점, 롯데마트 둔산점, 롯데마트 여수점, 이마트 울산점, 하나로클럽 울산점이 표본으로 추출되었고, 전통시장에 기초한 표본추출에서는 서울시 금천구 독산동우시장과 송파구 풍납시장, 고양시 일산서구 일산시장, 광주광역시 북구 운암시장, 대전광역시 대덕구 범동시장, 부산광역시 영도구 봉래시장과 해운대구 좌동재래시장, 창원시 진해구 중앙시장이 표본으로 추출되었다.

주요용어: 식품성분표, 대형마트, 전통시장, 최소복원확률비례추출법, 콜모고로프-스미르노프 적합도 검정

본 논문은 농촌진흥청 연구사업(세부과제번호: PJ013397052018)의 지원에 의해 이루어진 것임.

¹교신저자: 경기도 화성시 봉담읍 와우안길 17, 수원대학교 데이터과학부. E-mail: jwpark@suwon.ac.kr