

면적조사 및 생산량조사 표본설계  
A Study on the Sample Design for Crop  
Area Survey and Product Survey in Korea

박 홍 래\*

Abstract

This paper describes an outline of the sampling design for crop area survey and product survey in Korea. The design attempts to form a double stratification, to obtain an efficient allocation of the sample and to reduce the sampling error by establishing crop concentrated strata. The optimum numbers of sample field and sample plot are investigated. The design is made it possible to reduce the sampling errors as well as to reduce the sample size further than the present survey.

1. 연구목적

이 연구는 우리나라 작물별 식부면적과 생산량 추정을 위하여 효율적이고 포괄적인 표본설계를 작성하는데 목적이 있다.

농수산부에서 현재 사용하고 있는 작물별 식부면적 및 생산량조사 표본설계는 1974년에 작성되었는데 그후 10년간 동일표본 단위구를 줄곧 사용하여 왔다. 식부면적조사는 층화 1단 추출방법을 사용하였고, 여기서 1차 추출단위는 단위구로 하였다. 작물별 생산량조사는 면적조사 표본에서 다시 부차표본을 추출하는 층화 3단 추출법을 적용하였다.

그간 농업 기반의 변화도 심하였고, 주요작물 및 작부체계의 변화도 컸기 때문에 모집단의 크기와 작목구성도 크게 변하였을 것으로 예상된다. 그밖에 행정구역의 변경도 여러번 있었기 때문에 과거의 모집단을 가지고 현재의 식부면적 및 생산량을 추정하면 정확성이 떨어진다.

이러한 이유 때문에 작물별, 시도별 식부면적 및 생산량을 보다 효율적이고, 정확도를 높게 추정할 수 있도록 표본설계를 하는것이 본 연구의 목적이다.

---

\* Department of Computer Science and Statistics, Seoul National University, Seoul, 151, Korea

## 2. 조사내용

작물별, 시도별 식부면적 및 생산량조사에 있어서 조사대상 작물은 다음 21개를 포함한다. 즉, 수도, 옥도, 길보리, 쌀보리, 과맥, 맥주맥, 소맥, 대두, 소두, 고구마, 감자, 고추, 마늘, 양파, 김장무우, 김장배추, 참깨, 땅콩, 유채, 사과, 배 등 21개이다.

면적조사는 경지면적과 식부면적을 조사하며 추정항목은 다음과 같다. 경지면적은 전답면적조사, 증감사유별 면적조사, 관개시설 및 경지처리 상태등 재배 여건별 면적조사 등이다. 식부면적은 작물별 식부면적조사, 요인별 식부면적조사 등이다. 생산량 조사 대상작물은 위 면적조사와 같다. 생산량 조사는 작황조사, 예상수확량조사, 실수확량 조사로 나누어 조사한다.

## 3. 표본설계의 기본 사항과 특성

표본설계에 있어서 하나의 원칙은 목표 정도를 설정하고 이것을 만족할 수 있는 범위내에서 표본의 크기를 되도록 작게 하는 것이다. 본 설계의 특징을 요약하면 다음과 같다.

1) 다목적표본의 설계는 하나의 표본으로 두 가지의 조사 즉, 면적조사와 생산량조사를 할 수 있도록 설계한다.

2) 주요작물선정·면적조사와 생산량조사의 조사 대상 작물이 21개나 된다. 이 모든 항목에 대하여 만족한 표본을 설계한다는 것은 쉬운일이 아니기 때문에 주요작물조사에 초점을 맞추어 이들의 조사에 정확성을 기우이도록 표본을 설계하였다. 주요작물은 쌀, 대맥, 무우, 배추, 고추, 마늘, 양파, 특작, 과일을 말한다.

3) 총화·전국 모집단의 단위구수는 1,125,000개이다. 단위구의 총화는 지목구분, 답리작 구분, 재배작물구분에 따라 설정하였다.

본 설계에서는 이중총화를 고려하였다. 1차계층은 성격층이라 하여 위에서 말한 총화지표에 의한 구분이고, 2차층은 1차계층내에서 단위구를 20개내지 50개로 묶은 구분을 말한다.

본 설계에서는 1차계층을 14계층으로하고, 각 계층에 따라 2차계층을 설정하였다.

4) 표본배정·표본배정은 최적배정의 원리에 입각하여 1차계층별로 논비율에 따라 추출률을 정하였다. 논비율이 큰 계층은 추출률을 적게하고 밭비율이 큰 계층은 추출률을 크게 하였다. 이것은 논비율이 큰 계층은 단위구간의 면적변동이 적고, 밭비율이 큰 계층은 그 변동이 크기 때문이며 이러한 배정에 의한 결과 밭 작물의 오차를 감소시킬 수 있었다.

5) 주산층·작물간의 정도의 차가 크기때문에 이를 줄일 필요가 있다. 특히, 경제성 작물 즉 고추, 마늘, 양파에 대하여는 표본오차를 더 감소시킬 필요가 있다. 그리하여 경제

성 작물의 주산지를 택하여 주산단지층을 형성하고 여기에는 표본을 더 증가 배정하였다.

시도별 추정오차도 작물에 따라서는 큰 차이가 있기 때문에 시도간의 오차도 조절할 필요가 있다. 이 설계에서는 전체적으로 18개의 주산층을 만들었다.

6) 분산분석·수량조사에 있어서 단위구내 표본 필지수, 표본 필지내 표본 포구수의 합리적인 배정을 위하여 필지간, 포구간 분산을 산출한 결과 한 단위구에서 한 필지를 선정하고, 한 필지내에서 한 포구를 선정하는 것이 합리적이라는 결론을 얻었다. 이것은 수량조사에서 현행 표본의 1/4 표본을 사용하여도 수량조사의 오차는 거의 같다는 사실이다.

7) 추정식·오차를 감소시키는 방법으로 추정식의 선택은 중요하다. 본 설계에서는 선형추정과 비추정을 계산하여 효율을 검토한 결과 비추정이 효율이 크다는 것을 입증하였다.

#### 4. 현행 표본의 고찰

현행 면적조사 및 생산량조사 표본설계에서 표본의 크기, 층화, 표본배정, 추정식에 관하여 고찰한다.

현행 면적조사 표본의 크기는 43,822개이며, 이것은 전체 모집단의 1/25이다. 21개의 작물 식부면적을 조사하는 다목적 표본으로서 43,822개가 너무 큰 표본이라 할수는 없으나 적은 표본은 아니다. 표본추정치는 전국 추정오차도 작아야하고, 시·도별 작물별 오차도 작아야하기 때문에 표본이 커질수 밖에 없으나, 표본크기 43,822개는 조사업무량과 비용 그리고 조사 관리, 비표본오차의 관점에서 더욱 감소시켜야 한다.

현행 표본의 설계에서 층화지표는 지목구분, 모작구분, 수리구분, 작물구분으로 되어 있다. 지목구분은 전, 답, 혼합단위구, 모작구분은 1모작, 2모작 단위구, 수리구분은 수리안진, 수리불안진, 작물구분은 보통작물, 채소, 특용, 과수 단위구를 가리킨다. 이러한 층화지표에 따라서 형성되는 계층은 72계층이나되며 너무 많다. 그리고 72계층으로 나누고 보니까 그 성격이 뚜렷하지 못하여 층화의 의의가 약하다.

현행 표본설계는 각 읍·면별로 72개의 성격계층(1차계층)을 형성하고 다시 각계층내에서 단위구를 50개씩 묶어서 2차계층을 만들었다. 이것은 단순히 표본추출을 하기 위하여 50개중 2개를 임의 추출하는데 사용된다. 2차계층은 표본추출에도 이용되지만 표본오차를 감소하는데 2차계층을 활용하여야 한다.

현 표본은 추출률이 1/25이며 50개씩 묶은 2차계층에서 2개를 임의 추출한다. 이것은 면적 비례배정과 같다. 층화추출에 있어서 표본배정은 층의 크기와 층내분산에 비례하는 것이 좋다.

추정식의 선정은 표본오차를 감소시키는데 기여한다. 추정식으로 선형추정식과 비추정식을사용하는게 일반적으로 비추정식이 보다 유효추정치로 알려져 있다. 현행조사 분석은 비

추정식을 사용하고 있다. 현재 사용하고 있는 식부면적추정치  $\hat{A}_d$ ,  $\hat{A}_z$ 와 그 상대분산  $\hat{V}(\hat{A}_d)$ 은 다음과 같다.

$$\hat{A}_d = X \sum \sum \frac{Y_{hi}}{X_{hi}}, \text{ 도별추정치}$$

$$\hat{A}_z = \sum \hat{A}_d, \text{ 전국추정치}$$

$$\hat{V}(\hat{A}_d) = \frac{1}{m} \left( \frac{S_x^2}{\bar{X}^2} + \frac{S_y^2}{\bar{Y}^2} - 2 \frac{S_{xy}}{\bar{X}\bar{Y}} \right)$$

여기에서

$m$  : 도내표본 단위구수

$$S_x^2 = \frac{1}{m-1} \sum (X_{di} - \bar{X}_d)^2$$

$X_{hi}$  : 경지면적,  $Y_{hi}$  : 작물별 식부면적

상대분산을 보면 단순임의 추출에 의한 비추정공식을 사용하고 있다. 실제로 총화 추출을 하고 단순 추출 공식을 적용하면 오차가 실제보다 크게 계산됨으로 개선되어야 한다.

## 5. 총 화

신표본 설계는 면적조사와 수량조사를 동일표본에서 같이 하는데 면적조사 표본설계가 수량조사 표본설계보다 중요한 비중을 차지한다.

다음 표 1은 작물별 생산량의 상대오차를 가리킨다. 이표를 보면 생산량  $CV$ 는 식부면적  $CV$ 와 크기가 거의 같으며 10a 당 수량  $CV$ 의 영향은 아주 적다. 이것은 생산량 오차 ( $CV$ )를 줄이려면 면적오차를 줄여야 한다는 것을 뜻한다. 따라서 면적조사 표본 설계의 총화를 자세히 분석하였다.

### 1) 단위구의 구성

총화를 잘하려면 모집단의 구성과 특징을 잘 파악하여야 한다. 논비율이 큰 단위구에는 작물수가 적고, 밭비율이 큰 단위구에는 작물수가 많은데 과연 얼마나 되는지? 밭비율이 커지면서 어떤 작물이 식부되는지? 고추, 마늘, 양파와 같은 작물은 논비율과 관계가 있는지? 이러한 점을 알아보기 위하여 분석한 것이 표 2, 표 3이다.

표 2는 논비율별 단위구당 평균 식부 작물수이다. 이것을 보면 밭비율이 커질수록 작물

표 1. 작물별 생산량  $CV$ (현행) 전국추정치의  $CV$

|          | 통일   | 일반   | 고추   | 마늘   | 양파   | 대맥   | 무우   | 배추   | 참깨   |
|----------|------|------|------|------|------|------|------|------|------|
| 식 부 면 적  | 0.65 | 0.36 | 1.01 | 1.74 | 5.35 | 1.22 | 1.84 | 1.80 | 1.09 |
| 10 a 수 량 | 0.01 | 0.01 | 0.04 | 0.03 | 0.00 | 0.01 | 0.01 | 0.00 | 0.03 |
| 생 산 량    | 0.65 | 0.36 | 1.08 | 1.92 | 5.46 | 1.21 | 1.97 | 1.79 | 1.13 |

표 2. 논비율별 단위구당 평균 작물수

| 논 비 율 |   | 100% | 70~100% | 30~70% | 30% 미만 |
|-------|---|------|---------|--------|--------|
| 진     | 국 | 4.5  | 15.9    | 22.0   | 18.6   |
| 경     | 기 | 2.9  | 15.8    | 19.2   | 23.7   |
| 강     | 원 | 3.5  | 14.2    | 20.8   | 18.5   |
| 충     | 북 | 4.6  | 14.8    | 20.9   | 20.8   |
| 충     | 남 | 4.3  | 17.5    | 25.9   | 25.8   |
| 진     | 북 | 4.2  | 18.5    | 27.0   | 36.3   |
| 진     | 남 | 4.4  | 15.4    | 21.9   | 3.5    |
| 경     | 북 | 5.5  | 11.9    | 17.5   | 15.7   |
| 경     | 남 | 6.1  | 18.5    | 25.2   | 23.9   |
| 계     | 주 | 3.9  | 8.5     | 12.0   | 10.6   |

수가 많아진다. 따라서 면적조사에서 발 비율이 큰 계층에 표본을 더 배정하여야 한다는 것을 알 수 있다.

표 3은 논 비율별 작물별 단위구 분포이다. 논 100% 단위구가 31.8%, 논·발 혼합단위구는 56.8%, 발 100% 단위구는 11.4%이다. 논비율 100% 단위구에는 고추, 마늘이 거의 없고, 양파가 1,527개 양파 단위구중 약 17%가 식부되었다.

표 3. 표본조사 단위구 분포

| 논 비 율 |       | 100%   | 70~100% | 69~50% | 49~30% | 29~0% | 0%    | 계      |
|-------|-------|--------|---------|--------|--------|-------|-------|--------|
| 단     | 위 구 수 | 13,944 | 10,380  | 4,894  | 4,313  | 5,270 | 5,007 | 43,808 |
|       | %     | 31.8   | 23.6    | 11.2   | 9.8    | 12.0  | 11.4  |        |
| 1     | 모 작   | 11,621 | 8,772   | 4,232  | 3,747  | 3,234 |       |        |
|       | %     | 83.3   | 254.7   | 173    | 173.8  | 171.8 |       |        |
| 과     | 수     | —      | 429     | 427    | 382    | 464   | 716   | 2,418  |
| 특     | 작     | —      | 293     | 166    | 110    | 104   | 159   | 832    |
| 채     | 소     | —      | 1,094   | 796    | 650    | 748   | 814   | 4,012  |
| 보     | 통     | —      | 3,985   | 3,506  | 3,171  | 3,956 | 3,318 | 17,936 |
|       | 밭     | 13,916 | 10,351  | 4,851  | 4,258  | 4,899 | 177   | 38,452 |
| 대     | 맥     | 1,981  | 2,340   | 1,413  | 1,237  | 1,396 | 736   | 9,103  |
| 배     | 추, 무  | 1,620  | 7,319   | 4,183  | 3,837  | 4,656 | 3,920 | 25,545 |
| 고     | 추     | 420    | 6,476   | 4,181  | 3,865  | 4,715 | 3,820 | 23,477 |
| 마     | 늘     | 1,440  | 4,279   | 2,495  | 2,397  | 2,976 | 2,273 | 15,860 |
| 양     | 파     | 262    | 369     | 171    | 220    | 252   | 253   | 1,527  |

## 2) 층화구분

층화는 이중층화로하며 1차층은 성격층이라 하여 층화지표로서 논비율, 작물구분, 답리 작 구분을 사용하였고, 2차층은 1차층내에서 단위구분을 20개~50개씩 묶은 것이다. 1차층화는 다음 4가지 층화구분을 설정하여 이들의 효율을 비교하였다.

총구분 1에는 다음 18 층이 있다. 여기서 비율은 논비율을 말한다.

100% : 1모작, 2모작

70~100% : 1모—보통, 채소, 특작, 과수, 2모—보통, 채소, 특작, 과수

30~70% : 보통, 채소, 특작, 과수

30%이하 : 보통, 채소, 특작, 과수

총구분 2는 구분 1과 논비율을 달리하여 14 층이 된다.

90%이상 : 1모작, 2모작

40~90% : 1모—보통, 채소, 특작, 과수

40%이하 : 보통, 채소, 특작, 과수

총구분 3은 논비율 또는 밭비율에 따라서 층화구분을 하고, 작물구분은 고려하지 않는다. 이것은 현행 논비율 구분과 같으며 모두 7개 층이 있다.

100%, 80~100%, 60~80%, 40~60%, 20~40%, 0~20%, 0%

총구분 4는 총구분 1과 같으며 층수를 줄여서 14개 층을 만든다.

100% : 1모작, 2모작

70~100% : 보통, 채소, 특작, 과수

30~70% : 보통, 채소, 특작, 과수

30%이하 : 보통, 채소, 특작, 과수

이 층화구분은 층의 특성을 뚜렷이 부각시키고, 층수를 작게하여 층별 추정을 시도하고 표본오차를 줄이는데 목적이 있다.

표 4는 층화구분별 작물별, 식부면적 표본오차이다. 이 표에서 전국 추정오차를 보면 층화구분 네개중 구분 4가 가장 좋다는 것을 알 수 있다. 특히 무우, 배추, 고추, 특작, 과수작물에 있어서 구분 4가 좋은 것을 볼 수 있다.

현행표본은 72개 1차층을 사용하고 있으나, 실제 계산은 무계층단순추출에 의한 공식으로 계산하고 있다. 이표에서 무계층과 총구분 4를 비교하면 무계층 추정에 비하여, 구분 4의 비추정이 양과작물만 제외하면 전체적으로 오차가 크게 감소한다.

마찬가지로 무계층 단순추정과 비교하면 구분 4의 단순추정의 오차가 훨씬 작다. 이것은 구분 4에 의한 층화의 효과가 좋다는 것을 가리킨다.

다음은 2차층화의 효과를 보기 위하여 무계층 2차 비추정과 무계층 비추정을 비교하면 2차층화의 효과가 크다. 특히 쌀과 특작, 과수의 오차가 크게 줄어들었다. 이상의 분석에서 구분 4가 효율적이며, 특히 무계층과 비교할때, 오차가 크게 감소함을 알 수 있다. 표 4를 보면 쌀의 경우는 무계층에 비하여 총구분 4의 오차가 69%감소하고 배추, 무우는 16.9%, 고추는 23.7%, 마늘은 7%, 양파는 1.86%, 특작은 33.7%, 과수는 46.7%의 오차감소를 볼 수 있다.

표 4. 증구분별, 작물별, 식부면적 표본오차

| 시도   | 계층구분 | CV    | 쌀     | 대맥   | 배추무우 | 고추   | 마늘   | 양파   | 특용   | 과수   |
|------|------|-------|-------|------|------|------|------|------|------|------|
| 전    | 무계층  | 2차비추정 | 0.066 | 1.07 | 1.30 | 0.85 | 1.54 | 4.88 | 0.73 | 1.56 |
|      |      | 비추정   | 0.267 | 1.22 | 1.42 | 1.01 | 1.70 | 5.39 | 0.89 | 1.80 |
|      |      | 선형추정  | 0.280 | 1.22 | 1.42 | 1.00 | 1.70 | 5.39 | 0.89 | 1.80 |
|      | 구분1  | 비추정   | 0.083 | 1.11 | 1.18 | 0.77 | 1.57 | 5.29 | 0.59 | 1.64 |
|      |      | 선형추정  | 0.100 | 1.12 | 1.19 | 0.77 | 1.57 | 5.29 | 0.60 | 1.65 |
|      | 구분2  | 비추정   | 0.105 | 1.12 | 1.20 | 0.79 | 1.58 | 5.37 | 0.68 | 1.67 |
| 선형추정 |      | 0.121 | 1.13  | 1.21 | 0.79 | 1.58 | 5.38 | 0.68 | 1.68 |      |
| 국    | 구분3  | 비추정   | 0.067 | 1.20 | 1.34 | 0.87 | 1.65 | 5.36 | 0.76 | 1.77 |
|      |      | 선형추정  | 0.087 | 1.20 | 1.35 | 0.87 | 1.65 | 5.36 | 0.77 | 1.77 |
|      | 구분4  | 비추정   | 0.083 | 1.14 | 1.18 | 0.77 | 1.58 | 5.29 | 0.59 | 0.96 |
|      |      | 선형추정  | 0.100 | 1.14 | 1.19 | 0.78 | 1.58 | 5.30 | 0.60 | 1.00 |

2차층의 총화효율을 보면 쌀의 경우 75.2%, 대맥은 12.2%, 배추·무우는 8.4%, 고추 15.8%, 마늘 9.4%, 양파 9.46%, 특작 17.9%, 과수 13.3%의 오차 감소율을 볼 수 있다. 요약하면 면적조사에서 총화의 효과가 무총화에 비하여 크고 2차총화에서 그 효과가 더욱 증가하여 표본오차가 감소함을 알 수 있다.

### 3) 주산층

이상의 분석에서 총화의 효과를 보았다. 그런데 작물별 식부면적오차를 분석한 결과 몇 가지 문제점을 볼 수 있다. 첫째는 마늘, 양파의 오차가 다른 작물에 비하여 크고, 둘째 시도별 오차의 차가 작물에 따라 지역에 따라 차이가 크다. 작물별로 보면 양파의 오차가 현저히 크다. 시도별 오차를 보면, 서울, 부산, 대구, 인천, 제주도의 오차가 모든 작물에 걸쳐서 크고 전남 대맥의 오차가 특히 크다. 이런 점을 보완하기 위하여 고추, 마늘, 양파, 전남 대맥에 대하여는 주산층을 만들었다. 이 주산층은 각 작물의 주산지 중에서 오차에 크게 영향을 미치는 18개시군을 선출한 것이며, 주산층에 표본을 특별히 더 많이 배정하여 고추, 마늘, 양파, 대맥, 참깨의 면적오차를 감소한 것이다. 주산층에 속하는 시군을 보면 고추는 7개군, 마늘은 3개군, 양파는 10개군, 대맥은 1개군, 참깨는 1개군으로 되어있다.

### 4) 특수층

특수층은 도시근교와 오지, 낙도를 대상으로 설정한다. 근교층은 도시 근교에서 행정구역의 변경이나, 구획정리, 공장이전등의 이유로 토지이용변경이 자주 있고, 작물변경 때문에 지목 또는 작물구분이 바뀌는 경우가 비교적 많다.

이런 경우는 모집단의 단위구수가 달라지고, 단위구의 증구분이 달라진다. 근교층은 도시 근교에서 지목이나 작물구분이 심한 단위구를 묶어서 층을 형성한다. 오지 및 낙도는 거리가 멀고 교통이 불편한 지역을 묶어서 낙도층을 만든다. 낙도층에 대해서는 작물별 식부

면적, 수량조사에 실측조사방법을 사용하는 것보다는 면적조사나 행정조사로 하는 방법이 더 합리적이고 편리하며 때로는 더 정확할 수도 있다.

## 6. 표본 배정

표본 배정을 합리적으로 하면 또한 오차를 줄이는데 기여한다. 논 비율이 큰 층에는 표본을 적게 배정하고 밭 비율이 큰 층에는 표본을 많게 배정한다면 보다 작은 표본으로 표본의 정도를 높일 수 있다.

위에서 층구분 4가 효율적이라는 사실을 확인하였으므로 표본배정은 층구분 4에 대하여 고찰한다. 층구분 4에 의한 14계층에 순서에 따라서 일련번호를 붙인다. 즉, 층번호 1은 논비율 100% 1모, 층번호 2는 100% 2모, 층번호 3은 70~100% 보통과 같이 하여 층번호 14는 30%미만 과수층이 된다.

현행표본은 면적비례 표본배정을 하였다. 그러므로 논 비율이 큰 층에는 표본이 오차에 비하여 많이 배정되고, 밭 작물에 대해서는 표본이 작게 배정되었기 때문에 정도가 떨어진다.

표본배정을 위하여 우선 표본의 크기를 정하여야 한다. 이것은 각 도내 계층별 표본의 크기를 산정하고, 이를 합하여 도별, 전국의 표본의 크기를 구한다.

각 도내 층별 표본의 크기를 구하는데 1984년 면적조사 자료를 이용하여 다음 공식을 적용한다.

$$n_0 = n \left( \frac{CV}{CV_0} \right)^2,$$

다만

$n_0$ ; 각 계층별 표본의 크기

$n$ ; 현행표본에서 각 계층별 표본 크기

$CV$ ; 현행 각 계층별 추정치의 상대오차

$CV_0$ ; 목표정도

이와같이 구한 계층별 크기를 합한 결과가 다음 표5 층별표본배정과 같다. 그런데 도별 층별로 표본추출률을 각각 다르게 하면 실제 여러가지 작업이 곤란하기 때문에는 비율 100%, 70~100%, 30~70%, 30%미만의 4개 구분에 따라서 구분간 추출률은 달리하고 구분안의 층의 추출률은 같도록 하였다.

이러한 점을 생각하여 표본의 크기는 표5와 같이 4개구분으로 산출한 결과, 논 100% 층은 6,517개, 70~100% 층은 5,616개, 30~70% 층은 10,338개, 30%미만 층은 13,939개가 된다. 그리고 전국의 표본은 36,670개가 된다.



표 5. 층별 표본배정 분석

| 논 비 율          | 100%            | 70~100%         | 30~70%          | 30%미만           | 계      |
|----------------|-----------------|-----------------|-----------------|-----------------|--------|
| 단위 구수<br>현 행   | 13,943<br>31.9% | 10,117<br>23.1% | 9,263<br>21.1%  | 10,488<br>29.9% | 43,822 |
| 단위 구수<br>신 배 정 | 6,517<br>17.9%  | 5,616<br>15.4%  | 10,538<br>28.7% | 13,938<br>38.0% | 36,670 |
| 추 출 률          | 1/53            | 1/45            | 1/22            | 1/19            |        |

4개구분에 따른 표본의 크기를 각각의 모집단의 크기로 나누어 표본추출률을 산출한 것이 표 5와 같다. 즉 논 비 율 100%층은 1/53, 70~100%층은 1/45, 30~70%층은 1/22, 30% 미만층은 1/19 이 된다. 이 추출률을 기준하여 다음 4 가지 표본배정을 비교하였다.

4 가지 표본배정

|         | 100% | 70~100% | 30~70% | 30% 미만 |
|---------|------|---------|--------|--------|
| 현행 배정 1 | 1/25 | 1/25    | 1/25   | 1/25   |
| 배 정 2   | 1/50 | 1/45    | 1/25   | 1/20   |
| 배 정 3   | 1/50 | 1/40    | 1/25   | 1/20   |
| 배 정 4   | 1/45 | 1/40    | 1/30   | 1/25   |

배정 1은 현행 표본배정이고 배정 2와 3은 같으나 70~100%층이 다르다. 배정 4는 배정 2,3 보다는 비 율이 큰 층에 더 비중을 두었다. 배정 2,3은 다 같이 현행배정 보다는 논 지역은 추출률을 작게하고 밭지역은 다소 크게 하였다.

표 6은 표본배정에 따른 작물별 오차이다. 이것은 위 4 가지 배정의 효율을 비교한 것이며 배정 2,3,4 중에서 배정 3은 표본크기가 배정 2나 배정 4보다 약간 크기는 하지만 오차

표 6. 표본배정에 따른 작물별 오차(전국)

| 배정구분 | 표본수    | CV     | 쌀      | 대맥    | 배추무우  | 고추    | 마늘    | 양파    | 특용    | 과수    |
|------|--------|--------|--------|-------|-------|-------|-------|-------|-------|-------|
| 1    | 43,822 | 비추정    | 0.083  | 1.137 | 1.178 | 0.767 | 1.576 | 5.294 | 0.595 | 0.959 |
|      |        | 단순임의추정 | 0.1002 | 1.140 | 1.191 | 0.775 | 1.577 | 5.296 | 0.605 | 0.997 |
| 2    | 34,994 | 비추정    | 0.0979 | 1.409 | 1.250 | 0.746 | 1.769 | 6.327 | 0.569 | 0.892 |
|      |        | 단순임의추정 | 0.1227 | 1.413 | 1.261 | 0.751 | 1.770 | 6.329 | 0.578 | 0.925 |
| 3    | 35,678 | 비추정    | 0.0960 | 1.384 | 1.241 | 0.741 | 1.744 | 6.234 | 0.566 | 0.889 |
|      |        | 단순임의추정 | 0.1202 | 1.388 | 1.251 | 0.747 | 1.745 | 6.236 | 0.575 | 0.922 |
| 4    | 32,289 | 비추정    | 0.0985 | 1.385 | 1.307 | 0.808 | 1.902 | 6.296 | 0.621 | 0.984 |
|      |        | 단순임의추정 | 0.1215 | 1.389 | 1.319 | 0.814 | 1.803 | 6.298 | 0.631 | 1.022 |

가 모든 작물에 대해서 작다. 배정 3과 배정 1은 표본 크기가 같지 않기 때문에 직접 비교가 안된다. 배정 3은 표본이 35,678이고 배정 1은 43,822로서 배정 1이 23%나 크다.

그렇다하더라도 고추, 특작, 과수는 배정 3이 좋으며 다만 양파는 배정 3이 좋지 않다. 배정 3과 배정 1을 같은 표본크기 35,678로 한다면 다음표 7과 같이 배정 3이 배정 1보다 발작물에 있어서 양파만 제외하고 좋다. 그리고 배정 1은 논 작물과 양파에 있어 효율이 약간 높다.

표 7. 배정 1과 배정 3의 오차  
(표본크기는 다같이 35,678일 때)

|       | 쌀     | 대 맥  | 배추 무우 | 고 추  | 마 늘  | 양 파  | 특 작  | 과 수  |
|-------|-------|------|-------|------|------|------|------|------|
| 배 정 1 | 0.092 | 1.26 | 1.31  | 0.85 | 1.75 | 5.87 | 0.66 | 1.06 |
| 배 정 3 | 0.096 | 1.38 | 1.24  | 0.74 | 1.74 | 6.23 | 0.57 | 0.89 |

이상 고찰에서 4가지 배정방법중 배정 3을 적용하기로 한다. 표 8은 층구분 4에 따른 배정 3을 적용하였을 때의 전국 작물별 표본오차이다.

표 8. 층구분 4, 배정 3에 의한 작물별 면적 오차(전국추정)

|        | 쌀     | 대 맥   | 배추 무우 | 고 추   | 마 늘   | 양 파   | 특 작   | 과 수   |
|--------|-------|-------|-------|-------|-------|-------|-------|-------|
| 비 추 정  | 0.096 | 1.384 | 1.241 | 0.741 | 1.744 | 6.234 | 0.566 | 0.889 |
| 단순임의추정 | 0.120 | 1.388 | 1.251 | 1.747 | 1.745 | 6.236 | 0.575 | 0.922 |

## 7. 면적추정

면적 추정은 경지면적 추정과 식부면적 추정이 있다. 경지면적은 표본조사와 전수조사를 병행하되 전수조사는 경지면적이 2ha 이상 집단적으로 증감이 생긴 경우에 해당한다. 그러므로 경지면적은 표본추정치와 전수조사치의 합이 된다. 식부면적은 표본에 의한 추정치가 된다.

추정식은 선형추정과 비추정이 있다. 일반적으로 비추정이 선형추정보다 정도가 높다. 표 4에서 2차비는 2차층을 설정하여 비추정을 한것이고, 비추정은 1차층만을 고려한 것이다. 선형추정도 1차층만을 고려한 것이다. 이 중 2차비추정이 가장 높고 비추정과 선형추정은 별로 차이가 없다. 층구분 4에서 2차 비추정을 산출하지 못한것은 자료가 없기 때문이다. 그러나 여기서 층구분 4에서 2차층을 만들어 2차 비추정을 한다고 가정하면 이 표 4의 비추정보다 오차가 평균 10~20% 줄어든다. 앞으로 층화 또는 층이라면 층구분 4를 가리키고, 표본배정은 배정 3을 가리킨다.

이중 층화를 하였을 때의 추정식은 다음과 같다.

$d$  도의 면적 추정치  $\hat{A}_d$ , 전국 추정치  $\hat{A}_z$ ;

$$\hat{A}_{dh} = X_h \cdot \frac{\sum_i \sum_j Y_{hij}}{\sum_i \sum_j X_{hij}} \quad i=1 \cdots n_h, \quad j=1 \cdots m_{hi}, \quad h=1 \cdots 14,$$

$$\hat{A}_d = \sum_h \hat{A}_{dh}, \quad \hat{A}_z = \sum_d \hat{A}_d, \quad d=1, \cdots, 13, \quad \text{시도}$$

여기서  $h$  는 1 차층이고  $i$  는 2 차층이다. 상대분산은 다음과 같다.

$$\hat{V}(\hat{A}_{dh}) = \frac{\sum_i \Delta X_{hi}^2}{(\sum_{ij} X_{hij})^2} + \frac{\sum_i \Delta Y_{hi}^2}{(\sum_{ij} Y_{hij})^2} - 2 \frac{\sum \Delta X_{hi} \Delta Y_{hi}}{\sum \sum X_{hij} \cdot \sum \sum Y_{hij}},$$

$$\hat{V}(\hat{A}_d) = \frac{\sum \hat{V}(\hat{A}_{dh}) \cdot (\hat{A}_{dh})^2}{\hat{A}_d^2},$$

$$\hat{V}(\hat{A}_z) = \frac{\sum \hat{V}(\hat{A}_d) \cdot (\hat{A}_d)^2}{\hat{A}_z^2},$$

다만,

$n_h$  : 1 차층  $h$  내의 2 차층 수

$m_{hi}$  :  $h$  층  $i$  2 차층내 표본단위구 수

$X_{hij}$  : 1 차층  $h$  내  $i$  2 차층내  $j$  표본단위구의 전년도 경지면적

$Y_{hij}$  : 1 차층  $h$  내  $i$  2 차층내  $j$  표본단위구의 금년도 경지면적

$X_h$  : 1 차층  $h$  의 전년도 총경지면적

$\Delta X_{hi} = X_{hi1} - X_{hi2} = hi$  차층  $j$  2 차층내 1 번조사구와 2 번 조사구의 차.

작물별 식부면적은 위 공식에서  $Y$  가 작물별 식부면적이 되고  $X$  는 금년도 경지면적이 된다.

## 8. 평 가

지금까지 논의한 것을 요약하면 14 개 1 차층을 만들고 다시 그안에서 2 차층을 만들었다. 표본배정은 층별배정을 하여 논 비율이 큰 층은 추출률을 적게하고 밭비율이 큰 층은 추출률을 크게하였다. 추정식은 2 차층을 고려한 비추정을 검토하였다. 그리고 주산층을 설정하여 작물별 오차의 차를 감소시키고 도별 오차를 감소시키는데 기여하였다. 결국은 추정치의 오차를 감소시키고 도별 작물별 오차의 차도 줄이면서 동시에 전체 표본수를 48,822 개에서 39,697 개로 약 10% 줄이는 데 노력하였다. 이러한 여러가지 시도를 평가한 것이 표 9이다.

표 9 는 현행표본과 신표본과 주산지설정에 따른 표본오차를 나타낸다. 현표본은 크기가 43,822 개 표본이고 신표본은 크기가 39,697 개이다. 표 9 에서 첫째줄은 현행표본, 둘째줄은 신표본, 셋째줄은 주산층에 의하여 감소된 표본오차를 표시한 것이다.

전국오차를 보면 대맥을 제하고 신표본이 현행표본보다 오차가 작다. 특히 쌀, 고추, 특작, 과일이 그렇다. 대맥은 신표본이 약간 크지만 이것은 표본이 작기 때문이다. 대맥은 전국에 보급돼 있기 때문에 주산층을 만들 수도 없다.

다만 전남 강진에 표본을 더 많이 주어 오차를 줄였다. 고추는 오차를 많이 줄일 수 있었다. 마늘, 양파는 주산지를 설정하여 크게 줄인 것이 현행조사와 비슷하다. 그러나 2차층 비추정을 하면 이것보다 약 10%의 오차를 더 줄일 수 있다.

### 9. 현행 생산량조사 고찰

생산량조사는 면적조사와 수량조사가 결합하여 이루어진다. 면적조사의 층화에서 언급한

표 9. 현행, 신표본, 주산층에 따른 표본오차 및 표본크기

|     | 쌀     | 대 맥    | 배추, 무우 | 고 추   | 마 늘   | 양 파   | 특 용   | 과 일   |
|-----|-------|--------|--------|-------|-------|-------|-------|-------|
| 전 국 | 0.726 | 1.22   | 1.42   | 1.01  | 1.70  | 5.39  | 0.89  | 1.80  |
|     | 0.10  | 1.38   | 1.24   | 0.74  | 1.74  | 6.23  | 0.57  | 0.89  |
|     | —     | —      | —      | 0.708 | 1.67  | 5.31  | —     | —     |
| 서 울 | 6.85  | 0      | 14.25  | 45.50 | 38.57 | 0     | 20.16 | 63.81 |
|     | 4.21  | 0      | 9.16   | 40.34 | 40.80 | 0     | 19.77 | 73.68 |
|     | 2.98  | 0      | 6.48   | 28.53 | 28.85 | 0     | 13.98 | 52.10 |
| 부 산 | 6.00  | 0      | 12.45  | 41.03 | 27.13 | 0     | 37.85 | 51.11 |
|     | 5.37  | 0      | 16.15  | 55.79 | 26.83 | 0     | 53.14 | 46.93 |
|     | 3.79  | 0      | 11.42  | 39.45 | 18.98 | 0     | 37.57 | 33.18 |
| 대 구 | 5.17  | 9.74   | 15.78  | 17.66 | 21.72 | 53.88 | 17.42 | 12.72 |
|     | 2.28  | 9.96   | 13.74  | 18.16 | 28.36 | 66.69 | 15.67 | 3.95  |
|     | 1.61  | 7.04   | 9.72   | 12.84 | 20.06 | 47.16 | 11.08 | 2.79  |
| 인 천 | 6.16  | 49.70  | 14.82  | 17.68 | 33.63 | 0     | 15.51 | 24.32 |
|     | 2.08  | 56.65  | 9.90   | 14.42 | 30.13 | 0     | 11.37 | 19.13 |
|     | 1.47  | 40.06  | 7.00   | 10.20 | 21.31 | 0     | 8.04  | 13.52 |
| 경 기 | 0.67  | 6.62   | 3.59   | 2.66  | 5.30  | 81.43 | 3.00  | 6.06  |
|     | 0.24  | 6.36   | 1.95   | 2.15  | 5.93  | 73.03 | 1.72  | 3.91  |
|     |       |        |        |       |       | 49.48 |       |       |
| 강 원 | 1.73  | 7.84   | 4.08   | 2.96  | 5.08  | 31.31 | 3.01  | 8.94  |
|     | 0.51  | 7.25   | 2.70   | 2.49  | 4.63  | 28.99 | 2.34  | 5.37  |
|     |       |        |        |       |       | 25.22 |       |       |
| 충 북 | 1.22  | 4.98   | 4.85   | 2.45  | 4.80  | 0     | 2.45  | 5.59  |
|     | 0.40  | 5.82   | 4.4    | 1.49  | 5.41  | 0     | 1.47  | 2.61  |
|     |       |        | 1.39   |       |       |       |       |       |
| 충 남 | 0.65  | 3.31   | 3.76   | 2.60  | 3.55  | 19.16 | 2.25  | 5.67  |
|     | 0.23  | 3.30   | 3.69   | 2.14  | 3.05  | 18.85 | 1.35  | 2.93  |
|     |       |        |        | 2.87  |       |       |       |       |
| 전 북 | 0.62  | 11.87  | 4.05   | 2.84  | 5.45  | 24.46 | 2.99  | 8.12  |
|     | 0.22  | 14.27  | 3.64   | 2.03  | 6.09  | 27.73 | 1.77  | 4.28  |
|     |       |        | 2.01   |       |       | 23.93 |       |       |
| 전 남 | 0.64  | 98.48  | 3.98   | 2.82  | 4.33  | 8.73  | 2.21  | 11.04 |
|     | 0.18  | 139.31 | 3.53   | 2.83  | 3.46  | 9.06  | 1.35  | 8.36  |
|     |       | 44.66  |        |       | 3.29  | 6.73  |       |       |
| 경 북 | 0.72  | 1.78   | 3.96   | 2.45  | 4.00  | 12.04 | 2.20  | 2.78  |
|     | 0.26  | 2.01   | 3.91   | 1.68  | 5.07  | 16.08 | 1.49  | 1.23  |
|     |       |        |        | 1.51  | 4.91  | 15.47 |       |       |
| 경 남 | 0.72  | 2.24   | 4.92   | 3.21  | 4.96  | 10.42 | 2.67  | 7.67  |
|     | 0.32  | 2.65   | 4.24   | 3.11  | 5.46  | 13.24 | 1.87  | 5.70  |
|     |       |        |        |       | 11.82 |       |       |       |
| 제 주 | 25.00 | 0      | 11.64  | 14.56 | 14.20 | 19.76 | 4.44  | 5.66  |
|     | 9.29  | 0      | 9.99   | 12.99 | 11.90 | 17.21 | 2.64  | 1.85  |
|     | 6.57  | 0      | 7.06   | 9.18  |       |       |       |       |

것과 같이 생산량조사의 정도는 면적조사의 정도에 따라서 크게 좌우된다. 표 1에서 작물별 생산량오차의 크기가 식부면적의 오차와 거의 같다는 것을 볼 수 있다. 생산량오차에서 10a 당 수량의 오차가 차지하는 비율을 보면 쌀의 경우 2.7%, 고추는 3.7%, 마늘은 1.6%, 대맥, 무우, 배추는 모두 1%미만이다. 이와같이 생산량조사에서 면적조사가 차지하는 비중이 절대적으로 크다는 것을 알 수 있다.

현행 생산량조사는 면적조사 표본단위구를 이용하여 수량조사 표본을 추출한다. 즉 해당 작물이 있는 면적조사 단위구에서 3단추출을 한다. 1차 추출단위는 단위구, 2차 추출단위는 필지, 3차 추출단위는 포구로 한다. 단위구는 해당작물 식부면적에 비례하여 임의 계통 추출한다. 필지는 표본단위구내의 필지중에서 표본 필지를 식부면적 비례로 임의 계통 추출한다. 미곡, 고추는 2개 필지, 기타작물은 1개 필지를 추출한다. 포구는 표본필지를 답사하여 3m<sup>2</sup> 크기의 포구를 2개 임의 추출한다.

전국의 표본의 크기는 개별작물이 갖는 경제적 중요도에 따라서 허용오차를 다르게 책정한 후 각 작물별 전국 표본수를 결정한다. 전국의 표본수를 정한다음 시, 도별 표본수를 배정한다.

도별 생산량  $Y_d$ 와 상대분산  $CV_d^2$ 의 추정은 다음과 같다.

$$\hat{Y}_d = \hat{A}_d \cdot \bar{X}_d$$

$$CV^2(\hat{Y}_d) = \left( \frac{V(\hat{A}_d)}{\hat{A}_d} + \frac{V(\bar{X}_d)}{\bar{X}_d} + \frac{Cov(\hat{A}_d, \bar{X}_d)}{\hat{A}_d \bar{X}_d} \right)$$

## 10. 수량조사 총화추정과 표본배정

생산량조사 표본은 면적조사 표본의 부차표본이기 때문에 면적조사의 14개 2차 총화법을 적용하여 표본배정과 수량추정을 하도록한다.

수량조사에서 총화추출과 무총화추출의 효율을 비교하면 총화추출이 참깨를 제외하고 모든 작물에서 높다. 다음표 10은 무총화와 총화추출에 의한 작물별 수량오차를 나타낸 것이다.

이 표에서 생산량오차를 보면 총화추출이 무총화보다 모든 작물에서 작다. 10a 당 수량오차도 참깨를 제외하고 작다. 그러므로 수량조사는 총화추출에 의한 추정을 하는 것이 좋

표 10. 무총화, 작물별 수량오차

|          | 통일   | 일반   | 고추   | 마늘   | 양파   | 대맥   | 무우   | 배추   | 참깨   |
|----------|------|------|------|------|------|------|------|------|------|
| 생 산 량    | 0.65 | 0.36 | 1.08 | 1.92 | 5.46 | 1.21 | 1.97 | 1.79 | 1.13 |
| 10 a 수 량 | 0.01 | 0.01 | 0.04 | 0.03 | 0.00 | 0.01 | 0.01 | 0.00 | 0.03 |
| 식 부 면 적  | 0.65 | 0.36 | 1.01 | 1.74 | 5.35 | 1.21 | 1.84 | 1.80 | 1.09 |

총화, 작물별 오차

|          | 통일   | 일반   | 고추   | 마늘   | 양파   | 대맥   | 무우   | 배추   | 참깨   |
|----------|------|------|------|------|------|------|------|------|------|
| 생 산 량    | 0.55 | 0.25 | 0.98 | 1.80 | 5.41 | 1.10 | 1.81 | 1.74 | 0.99 |
| 10 a 수 량 | 0.00 | 0.00 | 0.01 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.04 |
| 식 부 면 적  | 0.55 | 0.25 | 0.77 | 1.57 | 5.28 | 1.09 | 1.63 | 1.57 | 0.80 |

다. 표본배정은 해당작물이 식부되어 있는 2차층에 식부면적비례 배정을 한다. 그리고 해당 작물이 식부된 2차층에는 적어도 1개 이상의 단위구를 배정하도록 한다.

### 11. 표본필지수와 표본포구수에 관한 고찰

수량조사에 있어서 표본단위구내 표본필지수의 결정과 표본필지내 표본포구의 결정은 중요하다. 왜냐하면 수량조사의 비용구성을 보면 단위구조사, 필지별 작물면적조사는 식부면적 조사에서 이미 끝났기 때문에 추가비용은 별로 들지 않는다. 그러나 포구당 조사비용은 수량조사 비용의 대부분을 차지할만큼 큰 비중을 차지한다. 그러므로 표본 필지수와 포구수를 보다 합리적으로 결정하여야 한다. 앞에서 말한바와 같이 현행조사는 모든 작물에 대하여 한필지에서 2개포구를 선정하고 있다.

#### 1) 분산 분석

단위구내 표본필지수의 적정치와 필지내 표본단위구수의 적정치를 구하기 위하여 다음 분산분석을 하였다. 전국에서 10개군을 뽑아 통일벼, 일반벼, 고추, 마늘, 양파, 대맥, 무우, 배추, 참깨등에 대하여 분석하였다.

쌀, 고추의 경우는 다음과 같다.

분 산 분 석

| 변 동           | $d \cdot f$ | $MS$  | $E(MS)$                                  |
|---------------|-------------|-------|--|
| 단 위 구 ( $D$ ) | $n-1$       | $MSP$ | $\sigma_3^2 + 2\sigma_2^2 + 4\sigma_1^2$ |
| 필 지 ( $S$ )   | $n(2-1)$    | $MSS$ | $\sigma_3^2 + 2\sigma_2^2$               |
| 포 구 ( $T$ )   | $2n(2-1)$   | $MST$ | $\sigma_3^2$                             |

$$\hat{\sigma}_3^2 = MST, \hat{\sigma}_2^2 = \frac{MSS - MST}{2}, \hat{\sigma}_1^2 = \frac{MSP - MSS}{4}$$

$$\bar{n}_{opt} = \sqrt{\hat{\sigma}_3^2} / \sqrt{\hat{\sigma}_2^2}, \bar{m}_{opt} = \sqrt{\hat{\sigma}_2^2} / \sqrt{\hat{\sigma}_1^2}$$

기타 작물에서는 1 단위구 1 필지 2 포구를 선정하므로 분산분석에서 필지는 고려할 필요없이 최적 포구수를 정하면 된다. 이와같이 하여 얻은 최적포구수와 필지수는 표 11과 같다. 이것을 보면 적정 포구수는 거의 모두 1보다 작거나 1과 2 사이에 있다. 그보다 큰

것은 마늘, 양파에서 1개군이 있다. 최적필지수도 모두 1보다 작고 2보다 큰곳은 일반미에서 1개군이 있다. 이것은 필지당, 포구당조사 비용을 무시한 최적표본수이므로 만약 비용함수를 고려하면 최적표본수는 단위당 비용의 평방근에 반비례하기 때문에 모두 1보다 작아야 한다.

표 11. 작물별 최적필지수, 최적 포구수

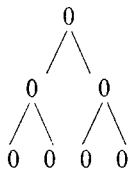
|   |   | 최적필지수(m) |     |     | 최적포구수(n) |     |     |     |
|---|---|----------|-----|-----|----------|-----|-----|-----|
|   |   | 1 이하     | 1~2 | 2~3 | 1 이하     | 1~2 | 2~3 | 3~4 |
| 통 | 일 | 4        | 1   |     | 3        | 1   |     |     |
| 일 | 반 | 3        | 3   | 1   | 9        | 0   |     |     |
| 고 | 추 | 7        | 0   |     | 9        | 0   |     |     |
| 마 | 늘 |          |     |     | 4        | 2   | 1   |     |
| 참 | 깨 |          |     |     | 9        | 1   |     |     |
| 양 | 파 |          |     |     | 1        | 0   |     | 1   |
| 무 | 우 |          |     |     | 7        | 0   |     |     |
| 배 | 추 |          |     |     | 9        | 0   |     |     |

이 분석에서 표본필지수와 표본포구수의 적정치는 1이라는 것을 알았다. 즉 표본단위구에서 1개 필지를 추출하고 1개 필지에서 1개 포구를 선정한다는 뜻이다.

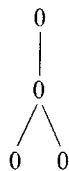
2) 표본오차에 의한 적정치 고찰

위에서 고찰한 최적 필지수와 포구수에 대한 결론을 보다 확실히 파악하기 위하여 다음과 같은 세가지 모형을 설정하였다. 모형 1은 현행조사의 표본수이다. 즉 1단위구에서 2필지, 1필지에서 2포구를 말한다. 모형 2는 1단위구에서 1필지, 1필지에서 2포구를 말한다. 모형 3은 1단위구에서 1필지, 1필지에서 1포구를 말한다. 이를 도시하면 다음과 같다.

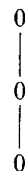
모형 1 (현행)



모형 2 (1/2 현행)



모형 3 (1/4 현행)



다음표 12는 현행, 1/2 표본, 1/4 표본에 의한 10a 당 수량의 표본오차를 계산한 것이다. 이표에서 현행표본이나 1/2 표본이나 1/4 표본이나 오차가 같다는 것을 확인할 수 있다.

같은 분석을 생산량추정에서 시도하면 표 13과 같다. 이 표는 이 세가지 모형을 가지고 생산량을 추정하였을때의 표본오차를 작물별, 모형별로 작성한 것이다.

이상은 무층화 표본을 가지고 고찰한 것이다. 층화추출을 가지고 같은 분석을 한 결과

표 12. 현행, 1/2, 1/4, 10a 당 수량조사 표본오차(무층화추출)

|        | 통일   | 일반   | 고추   | 마늘   | 양파   | 대맥   | 무우   | 배추   | 참깨   |
|--------|------|------|------|------|------|------|------|------|------|
| 현행     | 0.01 | 0.01 | 0.04 | 0.03 | 0.00 | 0.01 | 0.01 | 0.00 | 0.03 |
| 1/2 표본 | 0.01 | 0.01 | 0.04 | 0.03 | 0.00 | 0.01 | 0.01 | 0.00 | 0.03 |
| 1/4 표본 | 0.01 | 0.01 | 0.04 | 0.03 | 0.00 | 0.01 | 0.01 | 0.00 | 0.03 |

현행, 1/2, 1/4, 10a 당 수량조사 표본오차(층화추출)

|        | 통일   | 일반   | 고추   | 마늘   | 양파   | 대맥   | 무우   | 배추   | 참깨   |
|--------|------|------|------|------|------|------|------|------|------|
| 현행     | 0.00 | 0.00 | 0.01 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.04 |
| 1/2 표본 | 0.00 | 0.00 | 0.01 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.04 |
| 1/4 표본 | 0.00 | 0.00 | 0.01 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.04 |

표 13. 현행, 1/2, 1/4 표본 생산량 표본오차(층화추출)

|        | 통일   | 일반   | 고추   | 마늘   | 양파   | 대맥   | 무우   | 배추   | 참깨   |
|--------|------|------|------|------|------|------|------|------|------|
| 현행     | 0.55 | 0.25 | 0.98 | 1.82 | 5.41 | 1.10 | 1.81 | 1.74 | 0.99 |
| 1/2 표본 | 0.55 | 0.25 | 1.00 | 1.82 | 5.41 | 1.10 | 1.82 | 1.75 | 1.00 |
| 1/4 표본 | 0.56 | 0.25 | 1.05 | 1.82 | 5.39 | 1.10 | 1.81 | 1.76 | 1.00 |

같은 결론을 얻었다. 결론적으로 수량조사에서 표본의 적당수는 단위구에서 한필지, 표본 필지에서 한포구를 선정한다는 것을 보았다. 이와같이 하면 신표본의 크기는 현행표본보다 1/4 크기로 줄어드는 셈이다. 표본포구수가 1/4로 감소되는 대신 표본단위구수를 현재보다 다소 늘리는 것이 바람직하며, 단위구안에서 포구수를 늘리는 것은 좋지 않다.

### 12. 수량조사 추정

추정식은 층화추출 공식을 적용한다. 층화는 1차층 14개에서 다시 2차층으로 구성된다.  $h1$  차층,  $i2$  차층의 수량  $\hat{Y}_{hi}$  와 그 분산은 다음과 같다.

$$\hat{Y}_{hi} = \hat{A}_{hi} \bar{X}_{hi}$$

$$\hat{V}(\hat{Y}_{hi}) = (\hat{A}_{hi} \bar{X}_{hi})^2 \left( \frac{V(\hat{A}_{hi})}{\hat{A}_{hi}^2} + \frac{V(\bar{X}_{hi})}{\bar{X}_{hi}^2} + 2 \frac{Cov(\hat{A}_{hi} \bar{X}_{hi})}{\hat{A}_{hi} \bar{X}_{hi}} \right)$$

다만  $\hat{V}(\hat{A}_{hi})$ 와  $\hat{A}_{hi}^2$ 은 면적조사에서 얻는다. 또한  $\hat{V}(\bar{X}_{hi})$ 는 다음과 같다.

$$\hat{V}(\bar{X}_{hi}) = \frac{\sum (X_{hij} - \bar{X}_{hi})^2}{m_i(m_j - 1)}$$

다음  $d$  도내  $h$  층의 수량  $\hat{Y}_h$  와 그 분산은 아래와 같다. 도별, 전국추정치와 분산도 이와 같은 방법으로 얻게 된다.

$$\hat{Y}_h = \sum_i \hat{A}_{hi} \bar{X}_{hi}, \quad 10a \text{ 당 수량 } \hat{Y}_h = \frac{\hat{Y}_h}{\hat{A}_h}$$



$$V(\hat{Y}_h) = \sum_i V(\hat{Y}_{hi}) = \sum_i (\hat{A}_{hi} \bar{X}_{hi})^2 \left( \frac{V(A_{hi})}{A_{hi}} + \frac{V(\bar{X}_{hi})}{\bar{X}_{hi}} + 2 \frac{Cov(A_{hi}, \bar{X}_{hi})}{A_{hi} \bar{X}_{hi}} \right)$$

$$\hat{Y}_h \text{의 상대분산} = CV^2 = \frac{V(\hat{Y}_h)}{\hat{Y}_h^2}$$

### 13. 표본포구의 크기

현행 수량조사포구의 크기는 쌀의 경우 3조형 3m<sup>2</sup>, 약 1평에 해당한다. 포구의 형(type)과 크기(size)는 수량 추정치의 편의(bias)의 방향과 크기를 좌우한다. 일반적으로 포구의 크기가 작으면 수량추정에서 과다 편의가 일어나고, 포구의 크기가 커지면 이 편의가 줄어든다. 그러므로 포구의 크기는 어느정도 이상이 되어야 한다. 1962년 저자가 경기도 화성군에서 10개면을 택하여 포구의 형태와 크기에 관한 연구를 한 적이 있다. 다음 표 14는 그 결과의 일부분이다. 이 표에서 3조형은 크기가 1.431평에서 0.96평, 0.50평으로 작아질수록 편의는 minus 쪽으로 커진다. 그리고 CV도 커진다. 사각형, 삼각형의 경우는 포구의 크기가 작을수록 편의는 plus 쪽으로 커지는 것을 볼 수 있다. 포구의 크기를 현행의 3m<sup>2</sup>(1평)으로 할 것인가 또는 1.5평으로 크게 할 것인가가 앞으로 생산량조사에서 연구할 중요과제의 하나다.

표 14. 포구의 형태와 크기, 편의

| 포구<br>크기<br>(평) | 포구<br>수 | 사 각 형     |      |      | 삼 각 형     |      |      | 원 형       |      |      | 포구<br>크기<br>(평) | 3 조 형     |      |       |
|-----------------|---------|-----------|------|------|-----------|------|------|-----------|------|------|-----------------|-----------|------|-------|
|                 |         | $\bar{X}$ | CV   | 편의%  | $\bar{X}$ | CV   | 편의   | $\bar{X}$ | CV   | 편의   |                 | $\bar{X}$ | CV   | 편의    |
| 0.5             | 80      | 27.02     | 2.98 | 3.53 | 26.90     | 3.53 | 3.07 | 26.13     | 3.07 | 2.70 | 0.498           | 24.24     | 3.16 | -7.13 |
| 1.0             | 80      | 26.74     | 3.15 | 2.45 | 26.85     | 3.09 | 2.87 | 27.13     | 2.74 | 3.95 | 0.966           | 25.57     | 2.74 | -1.26 |
| 2.0             | 80      | 26.58     | 3.07 | 1.84 | 26.76     | 3.06 | 2.54 | 26.64     | 3.08 | 2.07 |                 |           |      |       |
| 3.0             | 80      | 26.59     | 2.86 | 1.88 | 26.58     | 3.24 | 1.84 | 26.69     | 3.14 | 2.26 | 1.431           | 26.17     | 2.56 | 0.26  |

### 14. 요 약

표본설계의 기본사항은 목표정도를 충족하는 범위내에서 표본크기를 되도록 적게하고 정확도를 높이는 것이다. 본 설계는 면적조사와 수량조사를 다같이 대상으로 하여 전체 표본을 약 44,000개에서 40,000개로 감소하였다. 작물별 면적및 수량오차를 일률적으로 줄일 수 있었고 동시에 시도별 정확도를 보다 균등하도록 하였다. 수량조사 표본단위구수를 현재의 1/4로 줄였을 때 표본오차가 커지지 않는다는 것을 확인하였다.

이 표본설계의 특징을 다음과 같이 요약할 수 있다.

- 1) 층화는 이중층화를 하여 1차층은 14개로 형성하고 2차층은 20개 내지 50개 단위구

로 묶었다. 이러한 층화에 의하여 2%~70%의 표본오차를 감소시킬 수 있었다.

2) 표본배정은 논 비율이 큰 층은 적게하고 밭 비율이 큰 층은 크게하여 효율을 높였다. 이러한 배정은 특히 밭 비율이 큰 층에서 그 효과를 볼 수 있다.

3) 주산층을 설정하여 작은 표본을 가지고 경계작물 면적추정치의 정도를 높이는데 효과가 있었다.

4) 근교층, 낙도층을 설정하므로써 조사를 보다 정확히 하는 방안을 고찰하였다.

5) 수량조사에서 필지간 오차, 포구간 오차를 분석하여 단위구내 필지나 포구를 증가하여도 표본의 정도를 높이지 못한다는 사실을 확인하였다. 이것은 현재의 표본포구수를 1/4로 줄이고 그 대신 표본 단위구수를 증가하는 것이 보다 효과적이라는 것을 뜻한다.

6) 수량조사에서 표본포구의 크기와 추정치의 편의를 고찰하였다.

7) 면적조사와 생산량조사에서 층화추출에 의한 비추정치를 적용하였다. 비추정이 단순임의 추정보다 오차를 크게 감소시키지는 못하였으나 전체적으로 약간씩 줄일 수 있었다.

이러한 모든 방법이 종합적으로 작용하여 표본을 적게 하면서 오차를 줄이는 데 기여하게 된다.

## 참 고 문 헌

- (1) Hong Nai Park, A Study on Sampling Methods in Rice Yield Survey 통계학연구, 제 4 권 1 호
- (2) 농수산부; 농업통계개요, 1980 년