

## 쌀 예상 생산량 추정방법에 대한 연구\*

강창완<sup>1)</sup> 김대학<sup>2)</sup>

### 요약

한국의 농업에서 쌀 생산량조사는 매우 중요한 조사로 알려져 있다. 특히 쌀 예상 생산량조사는 농업정책결정과 관련하여 유용한 기초자료를 제공한다는 점에서 가능한 한 정확한 예측을 필요로 한다. 본 논문에서는 통계적 모형을 이용한 쌀 예상 생산량 추정방법을 제안하고 기존의 주관적 추정방법인 달관추정과 비교함으로써 쌀 예상 생산량 추정에서 통계적방법의 응용 가능성과 타당성을 제시하고 있다.

주요용어: 쌀 예상 생산량조사, 달관추정, 회귀모형.

### 1. 서론

우리나라의 쌀 생산량조사는 국가 차원에서 매우 중요한 일이라 할 수 있다. 왜냐하면 쌀을 주식으로 하는 국가에 있어 식량자급화 문제는 우선적으로 중요시되고 있고 우리나라도 예외는 아니기 때문이다. 특히 쌀과 관련된 농업정책의 결정은 대부분 통계자료에 근거하여 이루어질 수 밖에 없고 이러한 이유로 쌀 생산량 조사는 농림부에서 실시하는 농업통계조사중 가장 중요한 조사의 하나로 알려져 있다.

현재 농림부에서 실시하는 쌀 생산량조사는 크게 세가지로 구분되는데 그것은 매년 8월 15일 시점에서 조사하는 작황조사, 9월 15일 시점에서의 예상 생산량조사 그리고 최종 수확기에 조사하는 실 수확량조사이다. 본 연구에서 다루고자하는 주제는 그 중 쌀 예상 생산량조사인데 이는 농업정책 결정이 쌀 예상 생산량조사 결과의 이용시기와 매우 밀접하기 때문이고 이러한 쌀 예상 생산량조사는 통계적인 방법이 적용되어야만 하기 때문이다.

한편 강창완(1998)은 1996년도 쌀 예상 생산량에 대하여 군집분석을 이용한 추정을 제안한 바 있다. 그러나 이 방법은 지역별 특성을 고려하지 못하였다는 한계를 지녔다. 따라서 본 논문에서는 1996년에 시행되었던 쌀 예상 생산량 조사를 이용하여 지역별 특성을 고려한 통계적인 추정기법을 제안하고 이에 따른 적용 결과를 1996, 1997, 1998년도에 걸쳐 실제 결과와 비교하여 그에 따른 문제점을 살펴보기로 한다.

\* 본 연구는 1999년 농의대학교 교내연구비에 의해 수행되었다.

1) (614-714) 부산시 부산진구 가야동 산24, 동의대학교 수확·컴퓨터통계학부, 조교수

E-mail: cwkang@hyonin.dongueui.ac.kr

2) (712-702) 경북 경산시 하양읍 금락리 330, 대구가톨릭대학교 정보통계학과, 부교수

E-mail: dhkim@cuth.cataegu.ac.kr

## 2. 쌀 생산량 조사

### 2.1. 쌀 생산량 조사방법

쌀 생산량조사는 1965년 쌀에 대한 단위당(10a) 수량조사를 처음으로 행정조사에서 표본조사로 전환되어 시행되어 왔는데 이러한 쌀 생산량조사는 다음과 같이 이루어진다. 조사표본은 면적조사 표본조사 단위구(2ha)의 부차표본이고 면적조사 단위구에서 3단추출을 하게 된다. 1차 추출단위는 단위구, 2차 추출단위는 필지, 3차 추출단위는 포구(3m<sup>2</sup>)이다.

이에 대한 추출과정을 살펴보면 논벼에 대한 생산량조사에서 각 시군에 배정된 표본단위구를 재배면적에 비례하여 계통추출하는데 즉, 논벼가 재배된 표본단위구들만 각 시군 내에서 나열한 다음 표본단위구별 누계면적에 의거해 임의 계통추출한다.(1차) 또한 표본 필지는 추출된 표본단위구내의 필지들 중에서 2개 필지를 임의 계통 추출하고(2차), 마지막으로 각 표본필지에서 2개의 포구를 임의 추출한다(전중우, 1995). 이렇게 추출된 표본에서 시군당 10a당 평균 수량을 추정하고 여기에다 시군의 논벼 재배면적을 곱하여 다음과 같이 시군별 생산량을 계산한다.

$$\text{쌀 생산량} = \text{재배면적} \times 10\text{a당수량}$$

즉, 논벼 생산량은 먼저 각 시군단위에서 추계된 단위면적(10a)당 생산량에 해당 시군의 논벼 재배면적을 곱하여 각 시군의 생산량을 구하고, 각 도별 생산량은 구해진 시군별 생산량의 합으로 구하며 전국의 논벼 생산량은 각 도의 생산량의 합으로 구한다.

### 2.2. 쌀 예상 생산량 조사

매년 9월 15일을 시점으로 조사하는 쌀 예상 생산량 조사는 다음 연도의 농업정책과 관련하여 기초 자료를 제공한다는 점에서 매우 중요한 의미를 갖는데 이 조사의 내용은 다음과 같다. 먼저 작황조사(매년 8월 15일)를 실시했던 A포구(표본필지당 2개의 포구에서 임의로 선정된 1개 포구)에서 아래의 조사항목들을 조사하게 된다.

포기당 이삭수, 이삭당 낱알수, 유효이삭비율, 품종, 피해상황조사

포기당 유효이삭수는 포기당 이삭수 × 유효이삭비율로 계산되고 이삭당 완전낱알수는 이삭당 낱알수 × 완전낱알비율로 계산된다. 이를 토대로 1m<sup>2</sup>당 유효완전낱알수는

$$1\text{m}^2\text{당포기수} \times \text{포기당유효이삭수} \times \text{이삭당완전낱알수} \quad (2.1)$$

로 되고 이를 이용하여 실측조사한 결과를 토대로 10a당 예상수량을 추정하게 되며 이는

$$10\text{a당 예상수량(kg)} = 1\text{m}^2\text{당 유효완전낱알수} \times \text{천립중(g)} \div 1000 \quad (2.2)$$

으로 표현할 수 있다. 여기서 천립중이란 천 개의 낱알의 무게(g)를 말한다. 한편 실수확량 조사에서 10a당 수량은 실제 표본포구내의 벼를 직접 베어 탈곡, 건조시킨후 중량을 실측한다. 그러나 예상 생산량조사시 10a당 수량 추정에서는 완전낱알수와 천립중이 시기적으

표 3.1: 벼 생육단계별 기상자료

연도	87	88	89	90	91	92	93	94	95	96
$X_1$	167.2	155.6	233.6	336.8	280.8	76.6	203.9	148.8	219.7	352.4
$X_2$	549.6	282.5	332.7	258.4	471.8	194.7	468.4	163.4	295.0	200.7
$X_3$	292.7	94.4	190.9	144	162.8	229.3	211.8	103.2	340.0	127.1
$X_4$	33.9	55.1	199.8	245.8	84	77.1	52	47.7	24.6	24.2
$X_5$	300.0	251	255.7	177	227.4	283.7	210.9	265.4	229.1	180.8
$X_6$	179.7	220.4	239.1	242.2	197.2	251.2	157.8	292.0	34.8	265.0
$X_7$	92.6	119	121	141.1	152.9	90.2	104.7	157.4	117.9	116.7
$X_8$	128	112.8	82.2	91.8	123.8	45.1	78.6	71.4	35.5	67.4
$X_9$	22.4	28.1	26.2	22.2	22.7	21.7	21.8	23.2	21.5	21.7
$X_{10}$	24.3	25.2	24.7	26.3	24.3	25.5	22.9	28.6	25.8	25.6
$X_{11}$	24.9	25.0	24.7	26.2	24.8	24.8	22.7	27.1	26.4	25.1
$X_{12}$	20.6	22.0	21.6	22.1	22.7	22.5	21.8	24.8	22.6	21.8

로 결정이 되지 않은 상태이기 때문에 이 두개의 수량구성요소를 예측하여만 한다. 이러한 미 결정된 수량구성요소를 예측하는 방법은 여러 가지가 있을 수 있겠으나 본 고에서는 지역별 회귀모형을 이용한 방법을 제안하고 1996년 농림부에서 적용했던 추정방법과 이를 비교하기로 한다.

### 3. 유사연도를 이용한 쌀 예상 생산량 추정

1996년 이전까지의 예상 생산량 추정은 각 시도의 농업통계조사 공무원의 달관추정을 이용하였다. 이는 실측 조사한 이삭수, 낱알수와 현지의 재배여건(품종, 관리상태, 피해상황 등) 등을 종합하여 표본 필지별 10a당 예상수량을 해당 지역의 숙련된 통계조사공무원이 경험적(주관적)으로 추정하는 것을 의미한다. 그러나 이러한 방법은 매우 놀라울 정도로 오차가 크지 않았음에도 불구하고 추정방법의 객관성 문제를 야기시켰고 이로 인해 1996년에는 통계적 방법을 적용하게 되었다. 이제 1996년 당시 시도되었던 예상 생산량 추정에 대해 살펴보기로 하자.

예상 생산량추정은 좁은 의미에서 10a당 수량을 추정하는 문제로 볼 수 있으며 그러므로 예상 생산량조사시에는 관측이 불가능한 수량구성요소 -완전낱알수(완전낱알비율)와 친립중-의 예측이 주요 문제가 된다. 한편 완전낱알수와 친립중에 영향을 주는 요인으로 일조시간, 강수량, 그리고 기온을 고려하게 되었고 이러한 관점에서 주어진 벼의 생육별 기상자료를 토대로 과거의 유사연도를 구하여 미 관측 수량구성요소를 대치하는 방법을 고려하였다. 유사연도를 구하기 위한 방법으로 다변량통계기법인 주성분분석과 군집분석을 한 결과를 살펴보기로 하자.

이용자료는 1987년부터 1996년까지 벼의 생육단계별 기상자료 표 3.1이며 이 표에서 사용된 변수들의 구체적 내용은 다음과 같다.

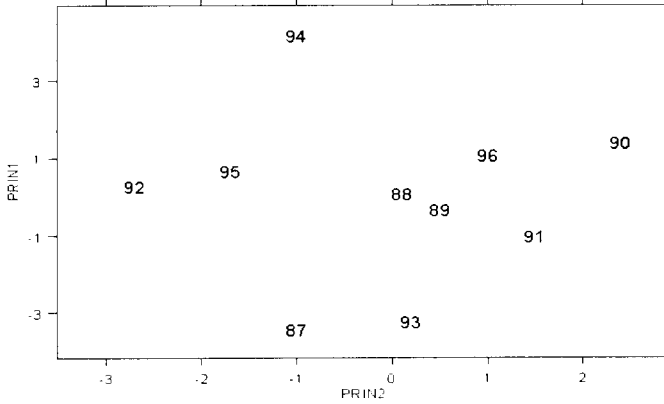


그림 3.1: 주성분 점수의 이차원 플롯

$X_1$ : 분얼기,  $X_2$ : 유수형성기,  $X_3$ : 출수 및 개화기,  $X_4$  등숙기 의 강수량  
 $X_5$ : 분얼기,  $X_6$ : 유수형성기,  $X_7$ : 출수 및 개화기,  $X_8$  등숙기 의 일조시간  
 $X_9$ : 분얼기,  $X_{10}$ : 유수형성기,  $X_{11}$ : 출수 및 개화기,  $X_{12}$  등숙기 의 평균기온

이때 분얼기란 6월 상순-7월 상순, 유수형성기는 7월 상순-8월 상순, 출수 및 개화기는 8월 상순-8월 하순 그리고 등숙기는 9월상순-9월 중순을 의미하고 강수량, 일조시간 그리고 평균기온의 단위는 각각  $mm$ , 시간, 섭씨온도를 사용하였다.

### 3.1. 주 성분 분석을 이용한 유사연도 추출

주성분분석(principal component analysis)은 여러 개의 변수에 대하여 얻어진 다변량 자료를 분석의 대상으로 하며 그 목적은 다차원적인 변수들을 축소, 요약하는 차원의 단순화와 더불어 서로 상관되어있는 변수들간의 복잡한 구조를 분석하는데 있다(김기영,전명식,1989). 이제 쌀 예상 생산량을 추정하기 위하여 표 3.1자료에 대한 주성분분석 결과를 보자. 여기서는 2차원으로 요약된 주성분점수를 이용하여 유사한 개체(연도)를 탐색하였다. 먼저 부록의 표A.1를 보면 고유값 기준으로 4개의 주성분수가, 누적비율로는 2-3개가 적당하게 보인다. 여기서는 주성분수를 2개로 하여 주성분점수를 구하고 2차분석으로 개체들을 그룹화하였다. 그림 3.1은 첫 두 개의 주성분 점수를 이차원 평면상에 도시한 그림이며 88년, 90년은 96년과 유사한 그룹으로 묶이는 것을 볼 수 있다.

### 3.2. 군집분석을 이용한 유사연도 추출

군집분석(cluster analysis)이란  $N$ 개의 개체에 대하여 각각 여러변수를 측정한 자료에 대하여 객관적인 기준에 근거하여 유사한 개체(연도)를 찾는 탐색적인 통계방법 (김기영,전

표 3.2: 최단연결법에 의한 집락결과

Number of Clusters	Clusters Joined		Frequency of New Cluster		Normalized Minimum Distance
	9	90	96	2	
8	91	93	2		0.470606
7	88	89	2		0.541523
6	CL7	94	3		0.590629
5	CL6	CL9	5		0.618524
4	87	CL8	3		0.623533
3	CL5	95	6		0.639112
2	CL4	CL3	9		0.649707
1	CL2	92	10		0.668659

표 3.3: Ward방법에 의한 집락결과

Number of Clusters	Clusters Joined		Frequency of New Cluster		Semipartial R-Squared	R-Squared
	9	90	96	2		0.008516
8	91	93	2		0.022147	0.969337
7	88	89	2		0.029325	0.940012
6	92	94	2		0.044711	0.895301
5	87	CL8	3		0.064348	0.830953
4	CL7	CL6	4		0.084095	0.746858
3	CL4	95	5		0.109177	0.637681
2	CL3	CL9	7		0.181288	0.456393
1	CL5	CL2	10		0.456393	0.000000

명식,1990)을 말하며 개체들 사이의 유사성(similarity) 또는 유클리드 거리를 계산하여, 최소거리를 갖는 개체를 찾는 방법이다. 여기서는 최단연결법과 Ward의 방법을 적용하여 유사연도를 구하였다. 표3.2는 최단연결법에 의한 집락결과이며 표3.3은 Ward의 방법으로 구한 결과이다. 두 결과 모두 96년과 90년이 가장 가까운 개체로 묶이는 것을 볼 수 있다.

결국 1996년도 농림부에서 채택한 유사연도를 이용한 쌀 예상 생산량은 다음과 같다. 먼저 주성분분석으로 얻어진 유사연도는 88년과 90년이고 군집분석으로 얻어진 유사연도는 90년이므로 공통의 결과를 보인 90년도를 최종적인 유사연도로 간주하였다. 즉 1996년 쌀 예상 생산량은 10a당 수량공식에 두 개의 미 결정된 수량구성요소를 1990년 자료를 대입하여 구할 수 있다. 즉 식2.1과 식2.2를 이용하여 10a당 예상 생산량은

$$\begin{aligned}
 10a\text{당 예상생산량} &= 1m^2\text{당포기수} \times \text{포기당유효이삭수} \times \text{이삭당완전날알수} \times \text{천립중} \\
 &= 22.3 \times 20.0 \times 62.6\text{개} \times 17.3g \\
 &= 483\text{kg}
 \end{aligned}$$

로 계산된다. 그러므로 96년도 쌀 예상 생산량은

$$\begin{aligned} 96년도\ 쌀\ 예상\ 생산량 &= 96년\ 벼\ 재\ 배\ 면\ 적(1,048,987ha) \times 10a\ 당\ 수\ 량(483kg) \\ &= 5,066,607M/T \end{aligned}$$

으로 추정되었다. 한편 1996년 실제 10a당 수량은 507kg, 수확량은 5,321,958 M/T 이었다.

#### 4. 회귀모형을 이용한 96년 쌀 예상 생산량 추정

1996년에 사용하였던 유사연도를 이용한 쌀 예상 생산량 추정은 기존의 주관적 추정에서 객관적 추정으로 전환시켰다는 데는 의미가 있지만 세 가지의 중요한 문제점을 포함하고 있다. 첫째는 유사연도를 구하는데 있어 지역별 특성을 고려하지 않은 전국기상자료를 근거로 추정하였고 또한 이에 근거해 전국 10a 당 수량을 구하여 쌀 예상 생산량을 추정한다는 점이다. 이는 각 지역별로 벼 재배면적과 기상조건이 다르다는 특성을 전혀 고려하지 못한 단점이 존재한다. 둘째로 기존 연도의 자료의 제한성, 즉 유사연도가 선택되면 해당연도 수치를 그대로 이용해야만 하는 점이고 그리고 마지막으로 도별 예상 생산량을 구할 수 없다는 점이다. 이러한 점에서 본 절에서는 미결정 수량구성요소에 직접적으로 영향을 줄 수 있는 요인으로 시기적으로 관측가능한 포기당 이삭수, 유효 이삭수, 이삭당알수(농림부, 1985-1998)와 등숙기 기상자료(기상청, 1986-1998)중 9월 상순의 일조시간을 고려하여 지역별로 회귀모형을 고려하였다. 여기서 특별히 등숙기 기상자료를 고려하게 된 이유는 일반적으로 벼의 생육단계상 벼 낱알의 충실도 혹은 무게가 출수기 이후의 일조시간에 가장 큰 영향을 받는다고 알려져 있기 때문이다. 이제 지역별 회귀모형에 의한 추정 과정은 다음과 같이 요약할 수 있다.

단계 1) 도별로 완전날알비율 및 친립층을 각각 종속변수로 두고 설명변수로는 시기적으로 관측 가능한 수량구성요소와 등숙기(9월 상순)의 일조시간을 후보 변수 대상으로 하여 변수선택을 실시한다.

단계 2) 선택되어진 설명변수들을 이용하여 회귀분석을 통한 각도별 예상 10a당 수량을 계산하고 쌀 예상 생산량을 추정한다.

단계 3) 각 도별로 구해진 쌀 예상 생산량을 모두 합하여 전국 생산량을 예측하고 벼 재배면적으로 나누어 전국 예상 10a당 수량을 구한다.

표 4.1는 부록에 수록된 각 도별 자료(1985년-1996년)에 대하여 변수증감법으로 적합한 회귀모형이며 표 4.2는 구해진 회귀식을 적용한 후 1996년도 수량구성요소를 예측한 결과이다. 이때  $T_1$  는 유효이삭비율,  $T_2$  는 포기수  $T_3$ 는 이삭당 낱알수 그리고  $T_4$ 는 유효이삭수를 나타낸다. 또  $T_5$ 는 일조시간을 나타낸다.

이제 표 4.2의 결과를 이용하여 전국 및 각 도별 쌀 예상 생산량을 구하면 표 4.3과 같다. 이 결과를 보면 1996년도 실제 생산량 5,321,958 M/T과 비교하였을 때 지역별 회귀모형을 적용한 경우 예상 생산량이 5,194,434 M/T, 3절의 유사연도를 이용하였을 때 5,066,607 M/T로서 회귀방법이 좀 더 오차를 줄이고 있음을 알 수 있다. 이에 대한 타당성을 좀 더 알아보기 위하여 1997년, 1998년의 경우를 비교하기로 한다.

표 4.1: 각 도별 최적 회귀모형

지역	완전남알비율( $Y_1$ )	예측모형	천립중( $Y_2$ )	예측모형
경기	$Y_1 = 0.904 \times T_1$	$r^2 = 0.99$	$Y_2 = 0.77 \times T_2$	$r^2 = 0.99$
강원	$Y_1 = 0.838 \times T_1$	$r^2 = 0.99$	$Y_2 = 22.35 - 0.34 \times T_3$ $+ 0.005 \times T_5$	$r^2 = 0.72$
충북	$Y_1 = 0.910 \times T_1$	$r^2 = 0.99$	$Y_2 = 25.02 - 0.42 \times T_2$	$r^2 = 0.34$
충남	$Y_1 = -5.9 + 6.88 \times T_1$	$r^2 = 0.47$	$Y_2 = 21.64 - 0.23 \times T_4$	$r^2 = 0.40$
전북	$Y_1 = -5.76 + 6.73 \times T_1$	$r^2 = 0.72$	$Y_2 = 33.11 - 0.622 \times T_2$	$r^2 = 0.36$
전남	$Y_1 = 0.908 \times T_1$	$r^2 = 0.99$	$Y_2 = 43.16 - 1.07 \times T_2$	$r^2 = 0.84$
경북	$Y_1 = -3.28 + 4.24 \times T_1$	$r^2 = 0.92$	$Y_2 = 36.38 - 0.87 \times T_2$	$r^2 = 0.48$
경남	$Y_1 = -1.9 + 2.84 \times T_1$	$r^2 = 0.81$	$Y_2 = 42.18 - 1.13 \times T_2$	$r^2 = 0.46$
제주	$Y_1 = -5.19 + 6.28 \times T_1$	$r^2 = 0.54$	$Y_2 = 22.93 - 0.12 \times T_3$	$r^2 = 0.53$

표 4.2: 96년도 수량구성요소 예측결과

지역	$1m^2$ 당 포기수	유효 이삭수	이삭당 남알수	완전남알수 (예측)	천립중 (예측)
경기	21.9	20.8	68.8	61.7	16.9
강원	23.5	19.1	69.0	57.3	17.8
충북	22.0	21.1	70.2	63.5	15.9
충남	21.3	20.1	73.6	68.2	17.1
전북	23.2	19.5	68.2	62.2	18.7
전남	22.9	19.3	67.9	60.7	18.7
경북	22.4	20.4	71.3	65.4	16.9
경남	22.1	20.1	68.9	62.4	17.3
제주	20.8	18.3	74.1	69.1	14.1

표 4.4의 회귀모형에 의한 예상 생산량은 1997년의 경우는 1996년까지의 자료를 이용하여 위에서 설명한 추정단계를 거쳐 추정되어진 결과이고 98년 예상 생산량은 97년까지의 자료를 이용하여 구한 결과이다. 농림부 예상 생산량의 경우는 97년부터 다시 달관추정으로 예상 생산량을 예측하고 있으며 비록 주관적 추정이기는 하나 매우 정확한 예측을 하고 있음을 볼 수 있다.

한편 유사연도에 의한 예상 생산량은 1997년, 1998년 모두 1990년, 1991년 그리고 1996년과 유사한 연도로 나타나 유사그룹연도에 해당하는 수량구성요소 평균치를 대입하여 예

표 4.3: 1996년 도별 쌀 예상 생산량

지역	재배면적 (천 ha)	10a 당수량 (kg)	예상생산량 (M/T)	실수확량 (M/T)	절대값오차 ( B - A )
경기	144.2	474	683969	717160	33191
강원	46.8	458	214076	213153	923
충북	60.2	469	282122	30408	18286
충남	175.1	499	874059	914097	40038
전북	152.8	526	803930	817370	13440
전남	208.9	502	1048186	1071974	23788
경북	140.8	505	711291	693547	17744
경남	120.1	480	576181	593711	17530
제주	0.2	371	620	538	82
전국	1049.1	495	5194434	5321958	165022

표 4.4: 실수확량과 예상 생산량과의 비교(단위: 천톤)

	1996년	1997년	1998년
농림부 예상생산량(A)	5066(483):군집	5353(509):달관	5133 (485):달관
회귀모형 예상 생산량(B)	5194(495)	5442(517)	5009(474)
유사연도 예상생산량(C)	5066(483)	5093(484)	4975(470)
실수확량(D)	5322(507)	5448(518)	5089(482)
절대값오차  D - A	256	95	44
절대값오차  D - B	128	6	80
절대값오차  D - C	256	355	114

상 생산량을 구하였다. 이들 세가지 방법에 대하여 1996년부터 1998년까지의 평균절대값 오차(mean of absolute error) 를 계산하여보면 농림부 예측의 경우는 131.7천톤이고 지역별 회귀모형에 근거한 예측의 경우는 71.3천톤, 유사연도를 이용한 경우는 241.7천톤으로 지역별 회귀모형에 근거한 예측이 가장 오차가 적음을 알 수 있다. 참고로 농림부 예상 생산량 자료는 매년 농림부에서 발표하는 보도자료에서 인용하였음을 밝혀둔다.

## 5. 결론 및 제언

쌀 예상 생산량추정은 농업정책 결정에 있어서 중요한 부분이며 이에 대한 통계적 기법의 응용은 필수적이라 할 수 있다. 위에서 1996년도 쌀 예상 생산량추정을 다변량통계기법과 지역별 회귀모형을 비교함으로써 지역별 회귀모형이 다소 상대적으로 큰 장점이 있음을 알 수 있었다. 또한 농림부에서 주로 적용해 오던 기존의 달관추정이라는 주관적 추정 방법의 대안으로 통계적 예측모형이 가능하다는 것을 보일 수 있었다.

한편 우리가 제안하였던 추정에서 완전날알수의 예측은 완전날알비율을 종속변수로 두고 유효이삭비율을 설명변수로 한 회귀모형이 적절하게 나타났으나 친립중에 대한 예측모



형은 본 논문에서 고려한 설명변수로는 여전히 만족할 만한 설명력을 얻지 못했다고 할 수 있다. 그러므로 본 논문에서 제시한 통계모형이 최적모형이 아니며 단지 위와 같은 통계적 방법으로 쌀 예상 생산량추정에 대해 객관화할 수 있는 대안을 제시한다는데 의미를 부여하고자 한다. 이와 관련하여 다음 몇 가지를 제안하고자 한다.

첫째, 우리가 제안했던 모형에서 천립중에 영향을 주는 요인들에 대한 전문적 연구의 필요성이다. 특히 지역별로 천립중과 관련된 요인에 대한 연구가 필요하며 이를 통해 더 좋은 예측이 가능하리라 기대된다. 둘째, 본 고에서 고려하지 못한 여러 가지 예측모형에 대한 검토가 필요하며 물론 이를 위해서는 지역별 농업통계 조사기관에서의 통계전문가의 활용이 필요하다. 즉, 쌀 예상 생산량추정은 지역별로 지역적 특성을 고려한 자료에 근거하여 최적의 예측모형을 구해야한다. 왜냐하면 본 고에서는 자료의 제한으로 도별 수량구성 요소와 기상자료를 이용하였으나 좀 더 나은 추정을 위해서는 시군별 자료등에 근거한 시군별 예상 생산량을 구하여 전국 예상 생산량을 구하여야 하며 이는 한 개인의 연구에 의해 해결될 성질의 것은 아니며 각 지역별 농업전문가와 통계전문가의 공동연구에 의해 해결할 수 있을 것이다.

## 참고문헌

- [1] 강창완 (1998). 다변량통계기법을 이용한 쌀 예상 생산량추정 <동의논집> 29,61-66.
- [2] 기상청 (1986-1998). <기상연보>.
- [3] 김기영, 전명식 (1989). <SAS 주성분분석>, 자유아카데미, 서울.
- [4] 김기영, 전명식 (1990). <SAS 군집분석>, 자유아카데미, 서울.
- [5] 농림부 (1985-1998). <작물통계>.
- [6] 전종우 (1995). <면적 및 생산량조사 표본설계>., 서울대학교 통계연구소.

[ 2000년 6월 접수, 2000년 8월 채택 ]

## 부록

쌀 예상 생산량 추정을 위한 표3.1의 비 생육단계별 기상자료를 주성분 분석한 결과를 표A.1과 표A.2에 그리고 쌀 예상 생산량 추정에 필요한 각 도별(지역별) 수량구성요소 및 기상자료를 표A.3과 표A.4에 수록하였다. 표A.3과 표A.4에서 사용한 기호는  $D_1$ =포기수  $D_2$ =유효이삭수  $D_3$ =이삭당 낱알수  $D_4$ =유효이삭비율  $D_5$ =완전낱알비율  $D_6$ =천립중  $D_7$ =일조시간 을 의미한다.

표 A.1: 주성분분석 결과 -고유값

	Eigenvalue	Difference	Proportion	Cumulative
PRIN1	4.71977	2.22849	0.393314	0.39331
PRIN2	2.49128	0.63620	0.207607	0.60092
PRIN3	1.85508	0.78317	0.154590	0.75551
PRIN4	1.07191	0.20276	0.089326	0.84484
PRIN5	0.86915	0.33048	0.072429	0.91727
PRIN6	0.53868	0.21888	0.044890	0.96216
PRIN7	0.31980	0.20454	0.026650	0.98881
PRIN12	0.00000		0.000000	1.00000

표 A.2: 주성분분석 결과 -고유벡터

	PRIN1	PRIN2	PRIN3	PRIN4	PRIN5	PRIN6
$X_1$	0.046409	0.518418	-0.352542	0.053227	0.063358	0.331542
$X_2$	-0.396513	0.109935	0.078794	0.410942	0.160122	-0.076148
$X_3$	-0.227511	-0.338188	-0.318531	0.153603	0.477132	-0.019333
$X_4$	0.070013	0.356670	0.042862	-0.382531	0.619447	-0.533410
$X_5$	-0.089961	-0.449580	0.435276	0.085915	0.255685	-0.019538
$X_6$	0.426104	-0.094464	-0.021571	-0.218001	0.060993	0.211041
$X_7$	0.299296	0.334356	0.126595	0.436850	0.008479	-0.262460
$X_8$	-0.191718	0.324983	0.438848	0.317844	0.102858	0.275668
$X_9$	0.011839	0.128187	0.598357	-0.366284	-0.030075	0.143170
$X_{10}$	0.433340	-0.114536	0.063248	0.101776	0.154021	0.170675
$X_{11}$	0.376451	-0.075521	0.012544	0.238460	0.439773	0.351414
$X_{12}$	0.371926	-0.117755	0.073981	0.331988	-0.247393	-0.486275

표 A.3: 도별 수량구성요소 및 기상자료

지역	연도	$D_1$	$D_2$	$D_3$	$D_4$	$D_5$	$D_6$	$D_7$	연도	$D_1$	$D_2$	$D_3$	$D_4$	$D_5$	$D_6$	$D_7$
경기	85	20.8	21.2	66.0	0.990	0.884	16.3	.	86	20.9	22.1	67.9	0.990	0.883	16.1	541.7
	87	21.2	21.5	62.6	0.991	0.889	16.1	703.3	88	21.7	21.0	64.3	0.991	0.916	16.9	680.7
	89	21.8	22.5	63.4	0.991	0.905	16.9	646.7	90	21.9	18.2	67.2	0.990	0.893	16.6	311.0
	91	21.9	20.7	62.4	0.991	0.894	16.7	506.7	92	21.9	21.9	62.5	0.994	0.911	17.1	477.3
	93	22.0	20.6	69.4	0.975	0.867	16.6	947.0	94	21.8	19.5	66.9	0.989	0.919	17.3	650.7
	95	21.7	19.8	68.7	0.989	0.878	16.4	373.7	96	21.9	20.8	68.8	0.991	.	.	763.7
강원	85	26.1	16.0	73.2	0.991	0.821	16.6	.	86	25.9	16.6	73.7	0.986	0.806	15.8	510.8
	87	25.7	17.8	68.0	0.985	0.812	16.2	580.5	88	25.8	17.2	67.9	0.987	0.789	16.5	527.0
	89	25.2	19.4	65.9	0.991	0.835	16.9	494.0	90	24.3	16.5	69.9	0.992	0.845	16.4	299.6
	91	23.6	19.0	65.0	0.990	0.865	16.1	496.6	92	23.6	20.1	63.2	0.989	0.843	16.5	431.8
	93	23.3	16.3	63.4	0.831	0.663	18.6	723.5	94	23.2	17.7	66.3	0.993	0.867	17.9	648.7
	95	23.2	17.5	70.7	0.989	0.819	15.9	378.7	96	23.5	19.1	69.0	0.990	.	.	666.7
충북	85	23.4	18.6	78.2	0.990	0.876	15.7	.	86	23.1	19.6	78.1	0.991	0.875	14.8	467.7
	87	23.4	20.3	69.6	0.988	0.899	15.9	491.7	88	23.4	19.2	76.4	0.992	0.914	15.4	567.3
	89	23.2	21.4	69.9	0.992	0.902	14.9	663.7	90	22.9	18.0	73.8	0.993	0.904	15.7	380.7
	91	22.5	20.3	65.2	0.987	0.917	15.8	534.0	92	22.5	22.4	61.9	0.993	0.911	15.8	396.7
	93	22.0	20.4	69.6	0.984	0.863	15.6	835.7	94	21.8	19.2	71.9	0.995	0.928	16.1	619.3
	95	21.9	19.7	71.6	0.992	0.904	16.0	289.0	96	22.0	21.1	70.2	0.993	.	.	692.3
충남	85	21.2	20.4	73.9	0.991	0.917	17.0	.	86	21.1	20.6	73.4	0.991	0.902	17.0	467.5
	87	21.4	20.7	66.8	0.987	0.903	16.8	545.0	88	21.8	20.1	72.2	0.991	0.934	17.4	620.5
	89	21.9	21.8	70.1	0.990	0.911	16.5	706.0	90	21.9	18.1	75.0	0.991	0.921	17.5	366.5
	91	21.9	19.7	68.0	0.992	0.940	17.4	525.0	92	21.7	21.6	65.2	0.992	0.931	17.0	454.5
	93	21.7	20.5	68.8	0.988	0.902	16.8	967.5	94	21.3	19.8	70.3	0.993	0.924	17.5	695.5
	95	21.0	19.7	73.1	0.989	0.877	16.6	321.5	96	21.3	20.1	73.6	0.992	.	.	703.5
전북	85	24.1	18.2	69.3	0.990	0.903	18.1	.	86	24.2	18.6	71.7	0.992	0.912	17.4	357.0
	87	24.5	19.3	63.4	0.992	0.904	17.8	538.0	88	24.4	19.4	67.8	0.994	0.931	18.1	521.5
	89	24.6	19.7	67.8	0.993	0.902	17.7	569.5	90	24.4	17.1	73.2	0.993	0.920	18.0	348.5
	91	24.2	18.3	65.1	0.990	0.925	18.8	541.5	92	24.0	20.8	63.2	0.991	0.908	18.0	406.0
	93	23.7	19.1	67.1	0.985	0.862	18.6	922.0	94	23.4	19.6	63.6	0.992	0.916	18.7	749.5
	95	23.4	19.2	67.1	0.988	0.876	18.3	299.5	96	23.2	19.5	68.2	0.992	.	.	623.5
전남	85	25.1	15.6	82.2	0.980	0.878	16.7	.	86	25.0	16.4	80.2	0.980	0.853	15.6	354.3
	87	25.3	16.4	71.8	0.979	0.868	16.1	468.0	88	24.9	17.0	77.1	0.983	0.912	16.9	414.8
	89	24.8	17.2	73.4	0.978	0.870	16.4	391.5	90	24.5	15.9	76.6	0.984	0.891	17.4	524.0
	91	24.0	16.3	68.5	0.984	0.913	18.1	646.8	92	23.6	19.3	60.3	0.983	0.900	18.1	462.5
	93	23.2	18.0	63.1	0.965	0.862	18.5	717.8	94	23.1	19.3	59.7	0.981	0.922	18.5	806.4
	95	22.8	18.8	63.7	0.982	0.902	18.4	319.0	96	22.9	19.3	67.9	0.985	.	.	624.8

표 A.4: 도별 수량구성요소 및 기상자료(계속)

지역	연도	$D_1$	$D_2$	$D_3$	$D_4$	$D_5$	$D_6$	$D_7$	연도	$D_1$	$D_2$	$D_3$	$D_4$	$D_5$	$D_6$	$D_7$
경북	85	24.4	17.0	78.1	0.979	0.882	15.6	.	86	24.2	17.8	77.7	0.978	0.870	15.3	365.0
	87	23.8	18.5	72.0	0.981	0.889	15.3	562.0	88	23.9	17.8	74.5	0.980	0.906	15.7	420.8
	89	23.9	19.1	72.6	0.985	0.892	15.6	517.0	90	23.2	17.5	76.2	0.985	0.890	16.1	425.4
	91	23.0	18.8	67.6	0.983	0.899	16.2	598.4	92	22.8	21.4	65.3	0.987	0.894	15.8	433.0
	93	22.7	18.9	68.3	0.947	0.732	15.7	601.6	94	22.4	19.7	67.1	0.986	0.920	16.8	760.0
	95	22.4	19.3	71.4	0.987	0.887	18.5	414.6	96	22.4	20.4	71.3	0.989	.	.	658.2
경남	85	22.5	16.7	75.4	0.978	0.860	16.6	.	86	22.5	17.5	77.8	0.972	0.837	16.0	303.6
	87	22.5	16.7	71.4	0.974	0.876	16.6	422.8	88	22.4	17.2	72.7	0.975	0.908	17.4	304.8
	89	22.9	18.1	69.7	0.980	0.880	16.7	304.8	90	22.7	17.0	75.3	0.987	0.899	16.8	425.4
	91	22.1	17.1	69.6	0.981	0.889	17.7	676.2	92	22.4	20.3	62.7	0.985	0.882	16.8	478.6
	93	21.8	18.1	67.5	0.943	0.777	17.4	599.8	94	22.1	19.8	61.3	0.985	0.918	17.6	772.4
	95	22.0	19.4	67.3	0.984	0.887	17.4	407.4	96	22.1	20.1	68.9	0.987	.	.	605.6
제주	85	28.6	13.3	77.3	0.951	0.807	15.2	211.5	86	28.4	14.2	77.8	0.937	0.557	12.8	250.5
	87	27.4	15.3	72.4	0.957	0.784	14.4	149.0	88	27.4	17.1	75.0	0.959	0.880	13.6	451.3
	89	26.8	17.9	71.5	0.959	0.842	13.7	181.0	90	24.1	16.9	77.6	0.968	0.847	13.8	686.7
	91	23.7	17.0	72.4	0.956	0.861	13.8	186.5	92	24.3	17.9	63.3	0.948	0.872	15.2	486.7
	93	22.1	16.8	64.9	0.964	0.874	15.9	106.0	94	23.0	19.0	62.6	0.968	0.852	15.0	791.0
	95	22.5	19.2	62.9	0.974	0.897	15.8	239.5	96	20.8	18.3	74.1	0.975	.	.	732.7

## A Note On the Rice Production Estimation Methods

Chang Wan Kang<sup>1)</sup> Daehak Kim<sup>2)</sup>

### ABSTRACT

In Korea, rice is the major crop because rice cultivation serves both cultural and socio-economic purpose. For this reason, rice production survey is very important to the Ministry of Agriculture and Forestry. Especially, the result of the rice production estimation is used for the agricultural policies. Then, we suggest the method of estimating rice production by using statistical model.

*Keywords:* Rice production estimation; Subjective estimation; Regression model.

---

1) Assistant Professor, Department of Computer Science and Statistics, Dongeui University.  
E-mail: cwkang@hyomin.donguei.ac.kr

2) Associate Professor, Department of Statistical Information, Catholic University of Taegu.  
E-mail: dhkim@cuth.cataegu.ac.kr