

# 기상 및 영농투입재 비용이 양파 단수에 미치는 영향: 패널회귀분석과 베이지안네트워크모형을 중심으로\*

이승인\*\*    손찬수\*\*\*    이혜림\*\*\*\*

## Keywords

기상 요인(meteorological factors), 영농투입재 비용 요인(farm input cost factors), 양파 단수(onion yield), 패널회귀분석(panel regression analysis), 베이지안네트워크 모형(Bayesian network model)

## Abstract

This study examined the effects of meteorological and farm input cost factors on the onion yields in Korea by employing a panel regression analysis. Also, it investigated the variables' interdependencies and their relations to the onion yields by using a Bayesian network model. We collected the panel data from 1991 to 2019 for our analysis. More specifically, we used the panel data of the regional meteorological factors by month (cumulative precipitation, cumulative sunshine duration, average relative humidity, average temperature), farm input cost factors of the onion (the cost of fertilizer and the cost of agricultural medicines), and the onion yield of three regions in Korea. We used STATA 14.0 and Hugin Expert for descriptive analysis, panel regression analysis, and the Bayesian network model. Our analysis can be summarized in two significant ways. First, we chose the fixed-effect model based on the Hausman test. The results based on the fixed-effect model confirmed that the average relative humidity (October, -), the cumulative precipitation (January, -; March, +; May, +), the cumulative sunshine duration (April, +), the average temperature (June, -), the cost of fertilizer(+) and the cost of agricultural medicines(+) were the significant variables of

---

\* 본 논문의 아이디어 구상에 도움을 주신 농촌진흥청 농업빅데이터일자리팀 조용빈 팀장님, 농촌진흥청과 네덜란드 바헤닝언대학 연구센터(WUR) 간 협력사업으로 추진된 'Bayesian Network Modelling & Big Data' 훈련 프로그램에서 베이지안네트워크 분석 방법을 지도해 주신 Hans Marvin 박사님과 Yamine Bouzembrak 박사님, 그리고 본 논문을 심사해 주시고 유익한 조언을 해 주신 익명의 심사위원님들에게 깊은 감사를 드립니다. 더불어 논문 내용 중 있을 수 있는 오류는 저자들의 책임이며, 본 연구 결과는 저자들의 소속기관 공식 의견이 아닌 개인 의견을 밝힙니다.

\*\* 농촌진흥청 농업빅데이터일자리팀 박사후연구원, 교신저자. e-mail: silee79@korea.kr

\*\*\* 농촌진흥청 농업빅데이터일자리팀 농업연구원.

\*\*\*\* 농촌진흥청 농업빅데이터일자리팀 농업연구사.

the model. Second, we analyzed the relationship between meteorological factors and farm input cost factors on the onion yields through the Bayesian network model and showed that the Bayesian network model is a promising analysis method that is useful in the agricultural sector to predict the yield. We also discussed the implications and limitations of this study and future research directions at the end of the paper.

#### 차례

1. 서론
2. 현황 분석
3. 연구 방법
4. 분석 결과
5. 요약 및 제언

## 1. 서론

양파(*Allium cepa* L.)는 백합과 2년생 초본식물로 대한민국의 전역에서 재배가 가능하고, 고추, 마늘과 함께 우리나라의 대표적인 양념 채소 중 하나이다. 대부분 노지에서 재배되기 때문에 양파는 기온, 강수량, 일조량 등 기상 요인에 의해 단수가 크게 영향을 받는 특징을 지닌다(오승원·김민수 2017). 최근 빈번한 기상이변, 농가의 불확실한 재배면적 관련 의사결정 등으로 양파 생산량 예측의 어려움은 더욱 커지고 있다. 일반적으로 양파 생산량 파악을 위해서는 재배면적과 단수 정보가 중요하나, 대체로 단수보다는 재배면적이 생산량에 더 큰 영향을 주는 것으로 알려져 있다. 재배면적은 수확 전 조사를 통하여 산출이 가능한 반면, 단수는 상대적으로 예측이 어려운 실정이다. 따라서 정확한 단수 추정이 가능하다면 재배면적과 함께 정확한 생산량 예측이 가능해 양파의 가격안정, 수급조절에 유용하게 활용될 수 있을 것이다(권영석 외 2015).

현재 농산물 생산량 예측과 관련하여 단수에 대한 관심과 중요성이 더욱 높아지고 있으며, 기상 요인과 생육 요인 등 단수에 영향을 미치는 주요 요인과 관련해서 다양한 연구들이 진행되고 있다. 기상 요인과 양파 생육, 생산량과의 관계를 살펴본 연구로는 이혜진 외(2019), 나명환 외(2019), 오승원·김민수(2017), 최성천·백장선(2016), 권영석 외(2015), 이중웅·이영석(1995) 등이 있다. 이혜진 외(2019)는 양파 생산량 증가 원인을 밝히기 위하여 생육 데이터와 기상·토양환경 데이터를 수집·분석하였는데, 구의 비대가 양파 생산량 증대의 직접적인 원인이 아니며, 결주율 등을 비롯하여 재식주수, 재배기술, 냉해발생 정도 등 다양한 요인에 대한 고려가 필요하다는 결과를 제시하였다. 나명환 외(2019)는 생육재배 기간을 절기를 기준으로 파종기, 육묘기, 정식기, 월동기, 경엽신장기, 구 비대기로 구분하고, 생육단계별 기상 요인의 주효과와 교호작용을 고려한 양파 단수 예측 모델을 개발하였다. 오승원·김민수(2017)는 양파의 생산량 예측과 관련해서 노지에서 재배되는 양파의 특성으로 인해 지역의 거리를 고려한 공간정보를 반영한 모형에 비해서 기상정보와 인접한 지역의 정보를 결합시킨 모형의 적합도가 우수하다는 연구 결과를 제시하였다. 최성천·백장선(2016)은 양파는 기상 변화에 민감하고, 이상기후 발생으로 인해서 평년과의 생산량 차이 및 지역별 편차가 크게 발생하기 때문에 수급량을 조절하기 위해서는 기후변화를 고려한 생산량 예측이 중요하다고 하였다. 권영석 외(2015)는 양파는 월동작물로 1차적으로 월동기 기상조건에 의해 영향을 받고, 구 비대기의 기온이

수량의 증감에 절대적인 영향을 미치나, 구 비대기에 고온이 장기간 지속될 경우 구 비대 불량으로 인하여 전체적인 수량이 감소하게 된다고 하였다. 끝으로 이중웅·이영석(1995)은 기상 요인 중 양과 작황에 영향을 주는 요인은 기온과 강수량이라고 하였다. 이렇듯 기상 요인과 양과 생산량과의 관계를 살펴본 연구는 지속적으로 진행되어 왔다. 그러나 양과 생산량 관련 예측력 향상을 위해서는 기상 요인과 더불어 양과 단수에 영향을 미칠 수 있는 새로운 요인들을 반영한 확장된 연구가 필요하나 이와 관련한 연구는 아직 부족한 실정이다.

양과 영농투입재 비용 조사는 양과 생산비 조사 차원에서 매년 이루어지고 있으며, 영농투입재 비용은 기상 요인에 영향을 받는 동시에 단수에 영향을 미치는 주요 요인이다. 기상 요인과 영농 투입재 비용 요인을 함께 이용할 경우, 기상 요인만을 고려한 기존 단수추정모형에 비해서 예측력 향상에 큰 도움을 줄 수 있을 것으로 판단되나, 이와 관련한 연구는 시도조차 없는 실정이다. 따라서 본 연구에서는 양과 단수추정과 관련하여 기상 요인과 영농투입재 비용 요인을 결합함으로써 기존 연구모형의 확장을 시도하고자 한다. 즉, 양과의 생육단계별 주요 기상 요인과 영농투입재 비용 요인, 양과 단수의 관계를 검증하고, 기상 요인과 영농투입재 비용 요인, 양과 단수의 상호의존성에 대한 정보를 바탕으로 특정 변수의 변화에 따른 다른 변수들의 변화에 대한 다양한 시나리오 정보를 제공하고자 한다. 이를 위하여 본 연구에서는 국가통계포털<sup>1</sup>과 기상자료개방포털<sup>2</sup>을 통해서 기상, 영농 투입재 비용 및 양과 단수 자료를 수집하고, 양과 단수와 연관성이 뛰어난 특성변수를 선정한 후 패널회귀분석과 베이지안네트워크모형을 적용하여 분석을 진행하였다. 양과 단수에 영향을 미치는 기상 요인과 영농투입재 비용 요인에 대해 주요 정보를 도출코자 패널회귀분석을, 기상 요인과 영농 투입재 비용 요인, 그리고 양과 단수 사이의 관계를 탐색해 특정 변수가 일어날 확률분포를 모형화 하고 변수들의 복잡한 상호의존성 파악과 특정 변수 관련 최적 자원 배분 및 조합에 대한 방안을 제시 하기 위해서 베이지안네트워크모형을 활용하였다. 이를 통해 양과 수급을 위한 유용한 정보를 제공 함으로써 재배농업인과 정책담당자의 합리적인 의사결정을 지원할 수 있을 것이다.

1 국가통계포털(<http://kosis.kr>).

2 기상자료개방포털(<http://data.kma.go.kr>).

## 2. 현황 분석

양파의 생육과정은 ‘육묘기’, ‘활착기’, ‘월동기’, ‘경엽신장기’, ‘구 비대기’로 분류할 수 있다. 양파의 발아적온은 15~20℃ 이고, 생육적온은 17℃ 전후이며, 구 비대 적온은 15~20℃ 이다. 25℃ 이상이 될 경우에는 고온장애가 발생하며, 구 비대기에는 일정시간 이상의 일조량이 중요하다. 생육 시기별 기상재해와 예상되는 문제점은 <그림 1>과 같다.

그림 1. 양파의 생육과정

생육과정(주요 농작업)												
8월	9월	10월	11월	12월	1월	2월	3월	4월	5월	6월	7월	
육묘기(조.중.만생)			활착기		월동기		경엽신장기		구 비대기			
파종 및 묘상관리			정식	배수구 정비		배수	1회 웃거름	2회 웃거름	병해충방제, 관수		수확	
기상재해 및 예상되는 문제점												
8월	9월	10월	11월	12월	1월	2월	3월	4월	5월	6월	7월	
가뭄			저온	혹한	저온	잦은비	가뭄		장마			
발아불량, 묘생육불량, 정식지연			초기생육 불량	동해	생육불량	습해발생 병다발생	구비대불량, 잎끝고사 고자리파리		수확지연 병다발생			
병충해 방제												
8월	9월	10월	11월	12월	1월	2월	3월	4월	5월	6월	7월	
			혹색썩음균핵병, 고자리파리, 파충채벌레			노균병, 잣빛곰팡이병, 세균병, 고자리파리			혹반병(채종재배) 총채벌레			

자료: 농촌진흥청 농사로(www.nongsaro.go.kr).

한편, 한국농촌경제연구원(2020)에 따르면, 조생종 양파 재배면적은 2000년 1,642ha에서 2019년 2,855ha로 연평균 약 4% 증가하였으며, 중만생종 양파 재배면적은 2000년 1만 5,131ha에서 2019년 2만 1,777ha로 연평균 약 1% 증가한 것으로 조사되었다. 양파 재배면적은 양파의 대체작물인 마늘 등의 수확기 가격에 따라서 증가와 감소를 반복하는 경향을 보이는 것으로 나타났다. 양파 생산량은 재배면적과 단수 증가로 인하여 2000년 이후 연평균 약 4% 증가하였으며, 2019년에는 역대 최고인 159만 4천 톤을 기록하였다. 양파의 단수는 2000년 이후 재배기술 향상과 기상 호조로 인하여 연평균 약 2% 증가하는 추세를 보이고 있으나, 연작에 의한 병해 발생 및 빈번한 기상이변으로 인하여 과거와 비교해서 단수의 변동성은 한층 증가한 것으로 나타났다(한국농촌경제연구원 2020).

<표 1>은 2019년도 지역별 양파 생산 현황으로 전라남도, 경상남도, 경상북도가 전체 재배면적의 약 76%를 차지하고 있으며, 재배면적과 생산량은 전라남도가, 10a당 생산량은 경상남도가 가장 많은 것으로 나타났다.

한편, 2019년도 양파 생산비<sup>3</sup>는 10a당 253만 원으로 조사되었으며, 양파 작목의 노동집약적인 성격으로 인해 양파 생산비에서 노동비가 차지하는 비중이 가장 높은 것으로 나타났다<표 2>.

표 1. 지역별 양파 생산 현황(2019년)

구분	면적(ha)	10a당 생산량(kg)	생산량(톤)
경기도	493	6,545	32,282
강원도	102	6,014	6,112
충청북도	155	4,353	6,729
충청남도	944	5,500	51,922
전라북도	2,373	7,957	188,857
전라남도	8,467	6,873	581,958
경상북도	3,309	7,471	247,231
경상남도	4,330	8,170	353,752
제주도	1,130	7,875	89,017

자료: 국가통계포털(KOSIS), 「농작물생산조사」.

표 2. 양파 생산비 현황(2019년)

구분	10a당 생산비(원)
생산비 합계	2,533,348
직접생산비	2,304,559
종묘비	240,134
비료비	257,493
농약비	127,549
기타 재료비	69,943
수도광열비	7,378
농구비	20,032
영농시설비	4,634
자동차비	9,018
노동비	1,369,916
위탁영농비	150,036
기타비용	48,029
생산관리비	398
간접생산비	228,789
토지용역비	195,603
자본용역비	33,186
부산물생산비	530
부산물공제생산비	2,532,818

자료: 국가통계포털(KOSIS), 「농축산물생산비조사」.

3 본 연구에서 활용된 주요 변수 중 하나인 ‘영농투입제 비용’은 양파를 생산하기 위하여 투입되는 원료비와 재료비(농약비, 비료비, 종묘비 등)를 의미함. 본 연구에서 영농투입제 비용(농약비, 비료비)은 양파 생산비 현황 자료를 활용하여 도출하였음.

### 3. 연구 방법

#### 3.1. 자료의 수집

양파의 단수에 영향을 미치는 기상 요인과 영농투입재 비용 요인을 파악하기 위해 과거 29년간('91~'19년)의 자료<sup>4</sup>를 수집·활용하였다. 국가통계포털을 통해 양파의 연도별·지역별 단수 자료와 연도별 영농투입재 비용 자료를 수집하였으며, 기상자료개방포털을 통하여 기상 자료를 수집, 이를 통계분석에 활용하였다. 기상 요인으로는 「종관기상관측」의 평균상대습도, 월강수량, 평균기온, 합계 일조시간 등을 이용하였으며, 전체 양파 생산량의 약 76%를 차지하는 전남·경북·경남 지역에 속해 있는 관측지점들의 평균 기상값을 각 도를 대표하는 값(기본값)으로 활용하였다.

<표 3>은 지역별 양파 단수 관련 기초통계량이다. 양파 3대 주산지인 전라남도(최솟값 4932.0kg, 최댓값 7395.0kg), 경상북도(최솟값 5592.0kg, 최댓값 7955.0kg), 그리고 경상남도(최솟값 5264.0kg, 최댓값 8386.0kg)에서 생산되는 생산량(10a 기준)은 지역별로 상당한 차이를 보이는 것으로 나타났다.

표 3. 지역별 양파 단수 기초통계량('92~'19년)

(kg/10a)

구분	평균	표준편차	최솟값	최댓값
전라남도	5952.5	541.6	4932.0	7395.0
경상북도	6627.4	608.1	5592.0	7955.0
경상남도	6527.3	824.7	5264.0	8386.0
3개도 전체	6369.0	732.3	4932.0	8386.0

자료: 국가통계포털(KOSIS).

<표 4>는 생육시기별 주요 기상 요인(10월 평균상대습도, 11월 강수량, 12월 평균기온, 1월 강수량, 2월 평균상대습도, 3월 강수량, 4월 합계일조시간, 5월 강수량, 6월 평균기온)과 영농투입재 비용 요인(비료비, 농약비)에 대한 기초통계량을 분석한 결과<sup>5</sup>이다.

4 양파는 9~10월에 파종하여 다음 해 6월 이후 수확되는 작물이기 때문에, 기상 요인은 '91년 10월의 자료부터 '19년 6월의 자료를 이용하였으며, 영농투입재 비용과 단수(연간 통계)는 '92년 자료부터 '19년 자료를 활용하였음.

표 4. 생육시기별 주요 기상 요인 및 영농투입재 비용 요인 관련 기초통계량('91~'19년)

구분	단위	평균	표준편차	최솟값	최댓값
10월 평균상대습도	%	69.84063	3.94957	59.4	80
11월 강수량	mm	46.59811	41.94124	0.175	197.089
12월 평균기온	°C	2.50659	1.74540	-2.682	5.523
1월 강수량	mm	25.00109	15.66230	0.33	78.109
2월 평균상대습도	%	58.20492	6.62467	42.111	69.5
3월 강수량	mm	69.39895	40.10870	9.16	172.136
4월 합계일조시간	hr	213.9098	25.88338	154.338	292.17
5월 강수량	mm	118.0057	56.69360	22.569	289.911
6월 평균기온	°C	21.31610	0.71600	19.46	22.671
비료비	원	153568.4	62790.71	80550	257493
농약비	원	55991.43	37567.83	8189	131374

자료: 국가통계포털; 기상자료개방포털.

### 3.2. 분석모형

#### 3.2.1. 패널회귀분석

본 연구에서는 김석 외(2019), 정학균 외(2013), 조재환 외(2012) 등의 선행연구를 바탕으로 시간의 변화에 따른 그룹별 영향력 파악을 위하여 패널회귀분석을 실시하였다. 즉, t년도 i지역의 양파 단수는 t년도에 i지역에서 재배되고 있는 양파 단수가 t년도 i지역의 기상 요인과 영농투입재 비용 요인 등에 의해 영향을 받게 되므로<sup>6</sup> 이를 고려하여 다음과 같이 단수결정모형을 설정<sup>7</sup>하였다.

$$y_{i,t} = \alpha + \beta x_{i,t} + \varepsilon_{i,t} \quad (\text{단, } \varepsilon_{i,t} = \mu_i + \lambda_t + e_{i,t})$$

5 영농투입재 중 비료와 농약은 생육부진 방지 및 병충해 발생 억제 등을 통해서 생산량 증대에 도움을 주는 대표적인 영농투입재이므로, 본 연구에서는 비료비와 농약비를 영농투입재 비용 요인으로 채택하여 분석에 활용하였음. 한편, 양파 영농투입재 비용(비료비, 농약비)은 생산자물가지수(2015=100)를 반영, 명목가격을 실질가격으로 변환하여 분석에 활용하였음.

6 양파는 고추와 더불어 대부분 노지에서 재배되어 기상 요인의 영향을 많이 받는 대표적인 농작물임. 한편, 조재환 외(2012)는 기상요인을 활용하여 단수결정모형을 제안하였는데, 본 연구에서는 영농투입재 비용(비료비, 농약비 등) 요인을 추가하여 기존 단수결정모형의 확장을 시도하였음.

7 영농투입재 비용과 관련하여 t년도 i지역의 영농투입재 비용 요인을 변수로 활용해야 하나, 지역별 영농투입재 비용 자료 확보의 어려움으로 연간 10a당 평균 영농투입재 비용 자료를 지역별 자료로 대체하여 분석을 진행하였음.



여기서  $t$ 와  $i$ 는 연도와 지역을 나타내는 하첨자이며,  $y$ 는  $t$ 년도  $i$ 지역의 양파 단수를 의미하는 종속 변수 벡터이다. 또한,  $x_{i,t}$ 는  $t$ 년도  $i$ 지역의 기상 요인과 영농투입재 비용 요인을 나타내는 설명변수 벡터이며,  $\varepsilon_{i,t}$ 는 오차항,  $\mu_i$ 는 관찰되지 않은 개체 효과 요인,  $\lambda_t$ 는 관찰되지 않은 시간 효과 요인,  $e_{i,t}$ 는 개체와 시간에 따라 변화하는 요인을 의미한다. 여기서  $\beta$ 는 기상 요인과 영농투입재 비용 요인이 양파의 단수에 미치는 영향 정도를 나타내는 모수 벡터이다(조재환 외 2012). 패널회귀분석은 일반 선형회귀모형의 오차에서 개체특성 효과와 시간특성 효과를 분리한 순수한 오차 제시가 가능하며, 이때 개체 효과( $\mu_i$ )는 특정 지역의 지형 또는 재배 방법 등 각 지역의 시간과 관계없이 변하지 않는 특성을 의미하고, 시간 효과( $\lambda_t$ )는 농산물 관련 정책 등 특정 시기 모든 지역에 영향을 미치는 특성이라고 할 수 있다(김용석 외 2014).

한편, 패널회귀분석은 오차항의 형태에 따라 고정효과모형과 확률효과모형으로 구분되며, 둘 중 하나를 선택해서 모수( $\beta$ )를 추정한다. 고정효과모형은 시간불변의 개별특성 효과 독립변수들과 관련이 있다고 가정(고정된 모수인 개체 효과와 시간 효과 외에 나머지 오차는 평균 0, 분산  $\sigma^2$ 인 독립적인 등분산성을 가지는 것으로 가정)하며, 확률효과모형은 시간불변의 개별특성 효과 독립변수들과 관련이 없다고 가정(개체 효과와 시간 효과, 나머지 오차는 평균 0, 분산을  $\sigma^2_\mu$ ,  $\sigma^2_\lambda$ ,  $\sigma^2_\varepsilon$ 인 것으로 가정)한다(김용석 외 2014; 황준성 2010). 고정효과모형과 확률효과모형 중 어느 모형이 분석에 더 적합한지는 Hausman 검정을 통해서 확인할 수 있는데, 귀무가설을 ‘확률효과모형이 더 적합하다’로 설정하고, Hausman 검정 결과가 기각될 경우에는 고정효과모형을, 채택될 경우에는 확률효과모형을 이용한다(김수경·김상혁 2018; Ashenfelter et al. 2003; Hsiao 2003). Hausman 검정 결과를 통해 두 모형 중 적합한 모형이 결정되면, 본 연구에서는 적합한 모형에 근거하여 기상 요인과 영농투입재 비용 요인 등이 양파의 단수에 미치는 영향을 분석하고자 한다. 본 연구에서는 패널회귀분석을 위해서 STATA 14.0 통계패키지를 활용하였다.

### 3.2.2. 베이지안네트워크모형

이산형 변수들 사이의 관계에 대해 사전적인 확률분포가 제시되었을 때 베이지안 정리를 활용하여 사후적인 확률분포로 변경하여 질적 또는 양적으로 분석이 가능한 기법 중 하나가 베이지안네트워크이다(선은정·서영욱 2019). 베이지안네트워크<sup>8</sup>는 국한된 영역에서 축적된 지식을 확률로 나타

내는 대표적인 통계기법으로 변수들 사이의 확률적 의존 관계를 나타내는 그래프와 변수별 조건부 확률로 구성된다(박영은 외 2010). 따라서 베이지안네트워크는 특정 개체의 속성 관련 값이 주어졌을 때, 분류 클래스 노드(Classification Node)의 사후 확률분포(posterior probability distribution)를 제시해줌으로써 개체에 대한 하나의 자동 분류기(classifier) 역할을 수행하게 된다(김정현 2005). 즉, 데이터에 기반한 베이지안네트워크 학습과 관련하여 주어진 문제와 관련한 변수를 ‘노드(node)’라고 하고, 이들 노드 사이에 존재하는 확률적 의존성을 호(arc)로 나타낼 수 있는데, 학습을 통해 각 변수들이 갖는 조건부 확률을 계산하여 변수 사이의 관계를 파악할 수 있는 정보를 제공한다(박현정·김상훈 2011; 이진창·최관 2007). 베이지안네트워크는 불확실성이 높은 분야의 데이터로부터 속성들 사이의 비선형적인 상호의존성을 비교적 정확하게 예측할 수 있는 계량적 도구로 평가받고 있다(이진창 외 2009; Jensen 1996). 이러한 특성으로 인해서 베이지안네트워크는 최근 이미지 처리 등의 공학분야, 질병 발생 확률을 예측하는 의학분야, 유전자 사이의 관계를 규명하는 유전공학분야, 기업의 부도 예측 등을 통한 위험관리를 위한 금융분야, 그리고 전략수립 및 시장분석 등의 경영분야에서 다양하게 활용되고 있다(박영은 외 2010; Heckerman et al. 1995; Morgan and Henrion 1990).

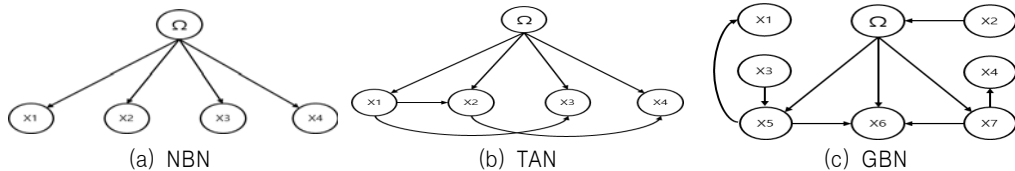
베이지안네트워크는 가장 단순한 형태인 NBN(Naive Bayesian Network), NBN의 확장 형태인 TAN(Tree Augmented Naive Bayesian Network), 가장 일반적인 형태인 GBN(General Bayesian Network)으로 크게 분류할 수 있다(배은찬·이진창 2016). John and Langley(1995)는 NBN이 단순하고 효율적이라는 장점이 있는 반면, 네트워크에 속한 일반 속성 노드들이 주어진 클래스 안에서 서로 독립적이라고 가정하고 있기 때문에 현실 세계의 현상을 제대로 설명하지 못하는 단점이 있다고 주장하였다(이은민·이진창 2017; 배은찬·이진창 2016). TAN은 Friedman et al.(1997)에 의해 소개되었는데, NBN과 달리 일반 속성 노드들 사이에 상호 독립적이라는 가정을 완화하기 위하여 일반 속성 노드들 사이에 트리(tree) 형태의 관계가 있음을 가정하고 있다. TAN은 NBN과 달리 노드 의존성을 추가함으로써 학습능력이 향상되어 분류성능이 향상된다는 특징을 지닌다(김현마·정성환 2013). GBN은 베이지안네트워크의 가장 일반화된 형태로, 다른 베이지안네트워크와 달리 클래스 노드조차 일반 속성 노드와 구분하지 않고 모든 노드 사이의 상호의존성을 하나의

8 베이지안네트워크는 빌리프 네트워크(belief network) 또는 방향성 비순환 그래픽 모델(directed acyclic graphical model)이라고도 불리고 있음. 특정 노드(node)마다 하나의 조건부 확률(conditional probability table)을 지니는 비순환 유향 그래프(Directed Acyclic Graph: DAG)로 정의할 수 있으며, 각각의 노드는 사건을 의미하며 노드를 연결해 주는 DAG는 사건 사이의 관계(relationship)를 의미함.

베이지안네트워크로 표현한다(이건창·최관 2007; Bouckaert 1995). GBN은 특정 제약없이 현실에 가까운 네트워크 구조화가 가능하다는 장점이 있는 반면, 변수가 증가할수록 학습비용이 급격히 증가하는 단점을 지니고 있다(선은정·서영욱 2019; 민병철 2017).

Hugin Expert Software는 베이지안네트워크의 구성요소인 이산형 또는 연속형 확률변수(random variable set), 관측값(사전확률), 방향성 있는 비순환 그래프(directed acyclic graph; DAG), 조건부 확률표(joint conditional probability table) 등을 포함하며 이를 토대로 모델링과 시뮬레이션을 통해 다양한 분석 정보를 제공해 주는 대표적인 베이지안네트워크 분석 프로그램이다(김지현 2006). 본 연구에서는 양파 단수 예측모델 개발을 위한 모델링과 시뮬레이션을 위하여 Hugin Expert Software를 활용하였다. 특히 베이지안네트워크 유형 중 TAN(Tree Augmented Naive Bayesian Network) algorithm을 이용하였는데, 이를 기반으로 Hugin Expert Software를 활용하여 베이지안네트워크의 그래픽 모형을 도출하고, 변수 조건을 다양하게 변화시키면서 기상 요인, 영농투입제 비용 요인, 그리고 양파 단수의 연관성 추론과 관련한 연구를 진행하였다.

그림 2. 베이지안네트워크 유형



자료: Cheng and Greiner(1999).

## 4. 분석 결과

### 4.1. 패널회귀분석 결과

기상 요인은 양파 단수와 밀접하게 관련이 있으므로 문헌 고찰을 통해 생육단계별 주요 기상 요인을 고려하여 연구모형을 설정하고 분석에 활용하였다. 양파의 생육과정은 일반적으로 육묘기(9~10월),

활착기(11~12월 중순), 월동기(12월 말~2월 초순), 경엽신장기(2월 중순~4월 중순), 구 비대기(4월 말~7월 초순)로 구분되는데, 육묘기에는 가뭄, 활착기에는 저온과 폭한, 월동기에는 저온과 잦은 비, 경엽신장기에는 잦은 비와 가뭄, 그리고 구 비대기에는 가뭄·장마와 관련한 기상 요인이 양과 생육에 영향을 미치는 것으로 보고되고 있으며, 이러한 생육단계별 기상 요인은 양과 단수에 영향을 미칠 것으로 예상할 수 있다. 또한 예상치 못한 기상 환경 변화로 인한 생육부진, 병충해 발생 등과 관련이 있는 다양한 양과 영농투입재 비용 요인 역시 양과 단수와 밀접한 관계가 있을 것으로 판단된다. 본 연구에서는 양과의 생육단계를 크게 정식기(10~11월), 월동기(12~2월), 구 비대기(3~6월)로 구분하였으며, 생육단계별 주요 기상 요인으로는 10월 평균상대습도(arh\_10), 11월 강수량(aorf\_11), 12월 평균기온(temp\_12), 1월 강수량(aorf\_1), 2월 평균상대습도(arh\_2), 3월 강수량(aorf\_3), 4월 합계일조시간(sdos\_4), 5월 강수량(aorf\_5), 6월 평균기온(temp\_6)을 선정하였다. 한편, 비료와 농약은 방제 및 생산량 증대에 도움을 주는 대표적인 영농투입재 요소로, 기상 요인에 따라 양과의 생육부진 및 병충해 발생이 예상될 때, 이로 인한 피해를 최소화하기 위해 활용되어 단수에 영향을 미치는 주요 요인이라고 할 수 있다. 이에 본 연구에서는 영농투입재 중 농약과 비료의 영향이 중요하다고 판단되어 농약과 비료의 대리변수로 비료비(cof)와 농약비(amd)를 분석에 활용하였다. 그리고, 종속변수(양과 단수: quantity)와 설명변수(기상 요인, 영농투입재 비용 요인)는 모두 로그(log) 값으로 변환하여 패널회귀분석을 실시하였다.

앞에서 제시한 양과 단수결정모형 추정 시 고정효과모형과 확률효과모형으로 나누어 모수( $\beta$ )를 추정한 결과는 <표 5>와 같다. 먼저 고정효과모형과 확률효과모형 중 어느 모형이 더 적합한지에 대한 Hausman 검정을 진행하였다. Hausman 검정은 p-value  $\geq 0.1$ 로 가정하여 귀무가설이 기각될 경우, 고정효과모형을 선택하고, 채택될 경우 확률효과모형을 사용하게 된다(오재균·여기태 2019). <표 5>에 제시된 Hausman 검정 결과, p-value가 1% 유의수준 내에서 귀무가설이 기각되어 고정효과모형을 선택하는 것이 더 적합한 것으로 나타났다. 다음으로 양과 단수결정모형에 투입된 설명변수에서 기상 요인 중 10월 평균상대습도, 1월 강수량, 3월 강수량, 4월 합계일조시간, 5월 강수량, 6월 평균기온과 관련된 모수( $\beta$ ) 추정치 부호는 (-), (-), (+), (+), (+), (-)를 가지며, 통계적으로 의미가 있는 것으로 나타났다. 반면, 11월 강수량, 12월 평균기온, 2월 평균상대습도와 관련된 모수( $\beta$ ) 추정치 부호는 (-), (-), (+)로 나타났으나, 통계적으로는 유효하지 않은 것으로 분석되었다.

표 5. 패널회귀분석을 활용한 양파 단수 추정 결과

변수명	고정효과모형		확률효과모형	
	추정치(Estimate)	t-값(value)	추정치(Estimate)	z-값(value)
ln(10월 평균상대습도)	-.5653679	-2.67**	-.7591492	-3.32***
ln(11월 강수량)	-.0048575	-0.45	-.0088260	-0.74
ln(12월 평균기온)	-.0082556	-0.26	-.0571577	-1.99**
ln(1월 강수량)	-.0242594	-1.82*	-.0414455	-2.96***
ln(2월 평균상대습도)	.0493679	0.47	-.1758089	-1.80*
ln(3월 강수량)	.0313633	2.23**	.0251991	1.63
ln(4월 합계일조시간)	.2064642	2.03**	.2000898	1.81*
ln(5월 강수량)	.0348864	1.72*	.0270844	1.25
ln(6월 평균기온)	-1.0690780	-2.66**	-1.3406610	-3.26***
ln(비료비)	.2682443	3.59***	.3115462	3.85***
ln(농약비)	.1018719	3.66***	.0901967	3.02***
constant	8.5604890	5.31***	10.976090	6.74***
sigma_u	.06385092		0	
sigma_e	.07849253		.07849253	
rho	.39821572 (fraction of variance due to u_i)		0 (fraction of variance due to u_i)	
Test of Goodness of fit	F(11, 70) = 6.54***		X <sup>2</sup> (11) = 68.87***	
H0 : cov(xi,t,ui) = 0 X <sup>2</sup> (2), (P>X <sup>2</sup> )	14.61(.0007)			

주: \*, \*\* and \*\*\* indicate significance at the 10%, 5% and 1% level.

자료: 저자 작성.

분석 결과를 활용한 양파 단수 추정 결과를 제시하면 다음과 같다. 즉, 10월 평균상대습도가 1% 상승함으로써 단수에 미치는 순수한 한계 효과는 -51.5578kg(10a 기준)이었고, 월 강수량이 10mm 증가함으로써 단수에 미치는 순수한 한계 효과는 1월 강수량은 -61.8005kg(10a 기준), 3월 강수량은 28.78325kg(10a 기준), 5월 강수량은 18.82886kg(10a 기준)로 나타났으며, 4월 합계일조 시간이 1시간 증가함으로써 단수에 미치는 순수한 한계 효과는 6.14731kg(10a 기준), 6월 평균 기온이 1°C 상승함으로써 단수에 미치는 순수한 한계 효과는 -319.427kg(10a 기준)으로 분석되었다. 한편, 양파 단수결정모형에 투입된 설명변수에서 영농투입제 비용 요인(비료비, 농약비)은 모두 통계적으로 유효했으며, 이들 변수의 모수( $\beta$ ) 추정치 부호는 (+), (+)로 나타났다. 이는 비료비가 1,000원 상승하여 단수에 미치는 순수한 한계 효과가 11.1250kg(10a 기준), 농약비가 1,000원 상승하여 단수에 미치는 순수한 한계 효과가 11.5879kg(10a 기준)임을 나타낸다.

## 4.2. 베이지안네트워크 분석 결과

### 4.2.1. 분석 의의 및 절차

기상 요인과 영농투입재 비용 요인은 양파의 단수에 영향을 미칠 뿐만 아니라 기상 요인, 영농 투입재 비용 요인 사이에서도 서로 영향을 주고받는 상호의존성을 지니고 있다. 그러나 선행연구들은 설명요인들의 상호작용은 간과한 채 변수들이 독립적으로 특정 성과에 영향을 미칠 것이라고 가정한 상황에서 대부분 연구를 진행해 왔다(이동기 외 2010). 본 연구 역시 1차적으로 기상 요인과 영농 투입재 비용 요인(설명변수)이 양파 단수(종속변수)에 미치는 영향을 살펴보기 위해서 패널회귀 분석을 진행하였으며, 이와 더불어 양파 단수에 영향을 주는 요인들 사이에 존재하는 연관성을 살펴 보고 단수를 예측하기 위해 TAN algorithm의 베이지안네트워크 방법론을 적용하였다. 이를 통해 기상 요인, 영농투입재 비용 요인 등 설명변수 사이의 상호 연관성을 살펴볼 수 있으며, 양파 단수 예측과 관련하여 최적 자원 배분 및 조합을 통한 예상 시나리오를 도출할 수 있을 것이다.

그림 3. 베이지안네트워크 분석 단계



자료: 저자 작성.

먼저, 베이지안네트워크의 일반적인 분석 단계는 <그림 3>과 같이 크게 네 가지로 구분될 수 있다. 전문가 의견 또는 문헌 고찰 등을 통해 주요 변수를 도출하고 자료를 수집하는 단계, 베이지안네트워크 기법을 활용한 모델링 단계, 베이지안네트워크모형의 타당성 검증 단계, 그리고 마지막으로 민감도 분석 단계의 4단계로 베이지안네트워크 분석은 진행된다.

본 연구는 양파의 단수와 기상 요인, 영농투입재 비용 요인들의 상호의존성을 분석하기 위해 베이지안네트워크모형을 이용하였다. 분석에 이용된 변수는 기상 요인으로 12월 평균기온, 1월 강수량, 2월 평균상대습도, 3월 강수량, 4월 합계일조시간, 5월 강수량, 6월 평균기온을, 영농투입재 비용 요인으로는 비료비와 농약비이다. 베이지안네트워크 분석에는 이진형(binary)이나 명목형(norminal)

자료가 적합(Jensen 2001)하므로, 본 연구에서는 선행연구를 검토하여 기상 요인<sup>9</sup>, 영농투입제 비용 요인, 그리고 양파 단수 등의 수치를 평균을 중심으로 4개의 집단(Very Low, Low, High, Very High)으로 사전 범주화하여 분석하였으며, 활용된 자료는 <표 6>과 같다.

표 6. 분석에 활용된 범주화 자료

변수명	단위	평균	구간			
			Very Low(1)	Low(2)	High(3)	Very High(4)
양파 단수	kg/10a	6,369	5,369 미만	5,369 ~ 6,369	6,369 ~ 7,369	7,369 이상
12월 평균기온	°C	2.5	0.5 미만	0.5 ~ 2.5	2.5 ~ 4.5	4.5 이상
1월 강수량	mm	25.0	15.0 미만	15.0 ~ 25.0	25.0 ~ 35.0	35.0 이상
2월 평균상대습도	%	58.2	53.2 미만	53.2 ~ 58.2	58.2 ~ 63.2	63.2 이상
3월 강수량	mm	69.4	39.4 미만	39.4 ~ 69.4	69.4 ~ 99.4	99.4 이상
4월 합계일조시간	hr	213.9	183.9 미만	183.9 ~ 213.9	213.9 ~ 243.9	243.9 이상
5월 강수량	mm	118.0	88.0 미만	88.0 ~ 118.0	11.80 ~ 148.0	148.0 이상
6월 평균기온	°C	21.3	20.3 미만	20.3 ~ 21.3	21.3 ~ 22.3	22.3 이상
비료비	원	190,445.3	170,445.3 미만	170,445.3 ~ 190,445.3	190,445.3 ~ 210,445.3	210,445.3 이상
농약비	원	63,863.0	43,863.0 미만	43,863.0 ~ 63,863.0	63,863.0 ~ 83,863.0	83,863.0 이상

자료: 저자 작성.

#### 4.2.2. 변수들 사이의 연관성 도출

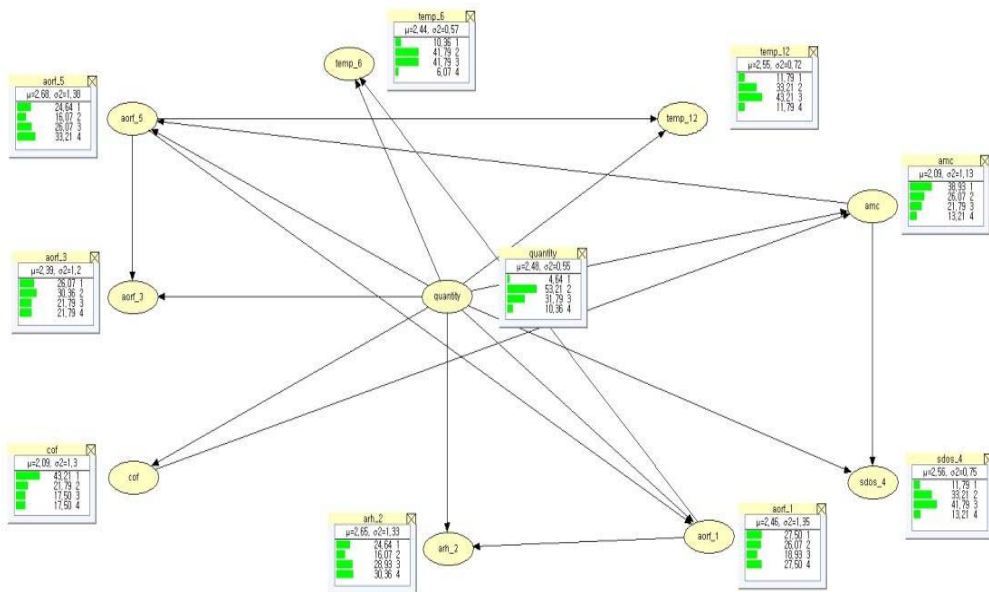
베이지안네트워크모형<sup>10</sup>은 설명변수 사이의 관련성을 함께 고려할 수 있기 때문에 현실 상황을 더 잘 설명할 수 있는 장점이 있다(Neapolitan 2004). 연구를 위해 베이지안네트워크 구성과 조건부 확률 테이블 계산을 위하여 우선 수집된 자료 중 약 82%(’91~’15년 자료)를 활용하였고, Hugin Expert(<http://www.hugin.com>) 통계 패키지의 TAN algorithm을 적용하여 변수 사이에 상호 통계적으로 어떻게 영향을 주고받는지에 대한 연관성 분석을 진행하였다.

9 모형의 간명성을 고려하여 본 베이지안네트워크 분석에서는 양파의 생육단계 중 정식기(10~11월)를 제외한 월동기(12~2월), 구비대기(3~6월)의 기상 요인을 변수로 활용하였음.

10 베이지안네트워크모형에서 변수들 사이의 관계는 가족관계 용어로 표현될 수 있음. 즉, 양파 단수 요인은 기상 요인(12월 평균온도, 1월 평균상대습도, 2월 평균온도, 3월 강수량, 4월 합계일조시간, 5월 강수량, 6월 평균온도), 영농투입제 비용 요인(비료비, 농약비)의 부모 노드가 되며, 양파 단수와 4월 합계일조시간이 연결되어 있을 경우, 4월 합계일조시간은 양파 단수의 자식 노드가 됨 (Bouzembrak and Marvin 2016).

<그림 4>는 베이지안네트워크 분석을 통해 도출된 변수들 사이의 의존성과 사전확률을 보여주고 있다. 즉, 국가통계포털, 기상자료개방포털을 통해 확보된 양파 단수와 기상 요인, 영농투입재 비용 요인 사이의 직·간접적인 관계 및 이때의 확률을 보여주고 있다. 예를 들어, 베이지안네트워크모형은 양파 단수가 Low(2)의 확률이 53.21%로 가장 높게 나타났으며, High(3)일 확률이 31.79%, Very High(4)일 확률이 10.36%, 그리고 Very Low(1)일 확률이 4.64%라는 것을 보여주고 있다.

그림 4. 변수 사이의 의존성과 사전확률을 보여주는 베이지안네트워크모형



자료: 저자 작성.

한편, 베이지안네트워크모형의 타당성을 검증하는 방법에는 두 가지가 있다(Marvin et al. 2016). 첫째, 선행연구 결과와 본 연구의 결과물을 비교해서 결과의 유사성을 확인하는 것이다. 예를 들어, 사후확률을 단수가 가장 낮은 수준의 특정값(100%)과 가장 높은 수준의 특정값(100%)으로 설정할 때, 다른 변수들의 변화 방향을 확인하고 선행연구 결과와 비슷한 방향성을 따르는지를 비교하는 방법이다. 베이지안 추론 결과, 기상 요인, 영농투입재 비용 요인과 양파 단수의 관계가 선행 연구와 유사하게 나타나 연구모형의 타당성은 확보되었다고 판단하였다. 둘째, 데이터 셋을 학습용과

11 <표 7>은 양파의 단수 수준을 가장 낮은 수준과 가장 높은 수준으로 변화할 경우, 다른 변수의 변화 정도를 제시하고 있음. 예를 들어 비료비와 5월 강수량에 대해서, 단수 수준을 변화할 경우 함께 연동하여 변화하는 것을 확인할 수 있는데, 단수 변화에 따른 비료비,



검증용으로 구분하고, 학습 데이터 셋으로 베이지안네트워크를 구성한 후, 검증 데이터 셋을 사용해 분류기의 정확도를 시험하는 방법이다. 모형의 타당성 검증을 위해 데이터 셋을 학습 데이터 ('92~'14년, 69개)와 검증 데이터('15~'19년, 15개)로 구분하여 분석한 결과, 15개 케이스 중 7개 케이스를 정확(46.67%)하게, 8개 케이스는 정확하지 않게(53.33%) 예측하는 것으로 분석되어서 모형의 검증 정확도는 약 46.67%인 것으로 나타났다. 배은찬·이건창(2016)은 정확도가 50%를 넘지 못할 경우, 실무에서 사용하는 데 제한이 있을 수도 있다고 하였다. 그러나, 본 연구는 베이지안 네트워크모형을 양파 단수 예측에 적용가능한지에 대한 탐색적 연구라는 점, 그리고 향후 실제 모델 개발 및 적용 상황에서 생육정보, 기술수준 등을 비롯하여 양파 단수와 연관성이 높은 다양한 데이터를 활용할 경우 정확도는 더욱 높아질 것이라는 점, 그리고 선행연구와 비슷한 결과물을 제시하고 있다는 점에서 본 연구에서 도출된 베이지안네트워크모형의 타당성에는 문제가 없는 것으로 판단하였다. 따라서 제안된 베이지안네트워크모형을 기반으로 다음 단계인 민감도 분석(what-if)을 진행하였다.

---

5월 강수량 변화 방향성은 비료는 농작물의 건전 생육과 품질 균일 등을 통해 단수 증대에 영향을 미치며, 구 비대기 수분 부족은 수확량 감소의 원인이라는 양파 관련 선행연구 결과를 반영하고 있으며, 따라서 도출 결과에 대한 타당성이 확보된 것으로 판단함.

### 4.2.3. 민감도 분석(what-if)

베이지안네트워크는 변수들 사이의 관계를 확률적으로 나타낼 수 있어 특정 변수값을 입력할 경우, 특정 변수 변화에 따른 다른 변수의 변화가 시각적으로 표현되기 때문에 민감도 분석(what-if)을 통해서 다양한 시나리오별 분석 결과 제시가 가능하다(선은정·서영욱 2019).

그림 5-1. 민감도 분석(quantity=Very High(4))

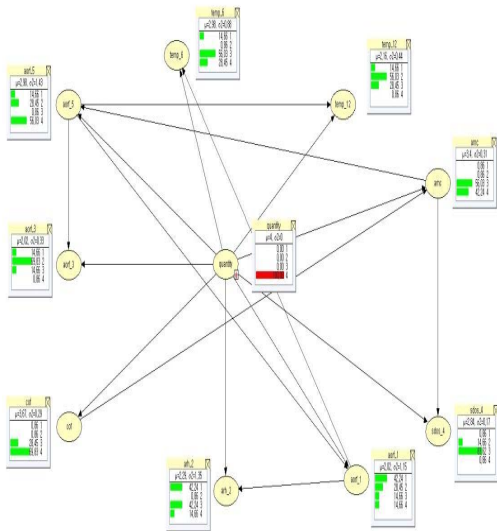
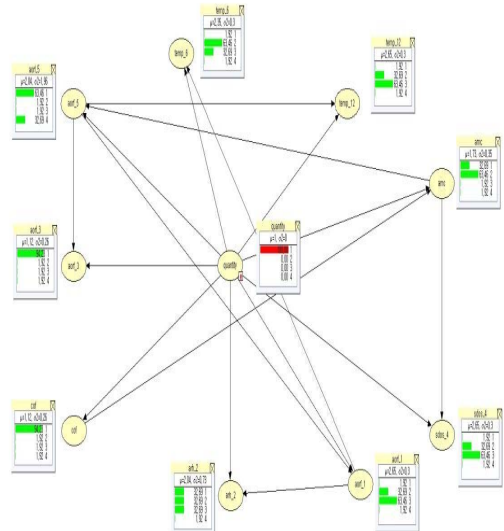


그림 5-2. 민감도 분석(quantity=Very Low(1))



자료: 저자 작성.

<그림 5-1>과 <그림 5-2>, <표 7>은 양과 단수(quantity)가 가장 높은 구간과 가장 낮은 구간에 대한 특정값을 갖는 경우, 기상 요인과 영농투입재 비용 요인에 대한 사후확률의 변화를 제시해 주고 있다. 예를 들어, 양과 단수(quantity)가 가장 높은 구간(Very High: 4)의 경우, 비료비(cof)가 가장 높은 구간(Very High: 4)의 확률은 17.50%에서 69.83%로 52.33% 정도 크게 증가한 것으로 나타났으며, 5월 강수량(aorf\_5)이 가장 높은 구간(Very High: 4)의 확률은 33.21%에서 56.03%로 22.82% 상승한 것으로 분석되었다.

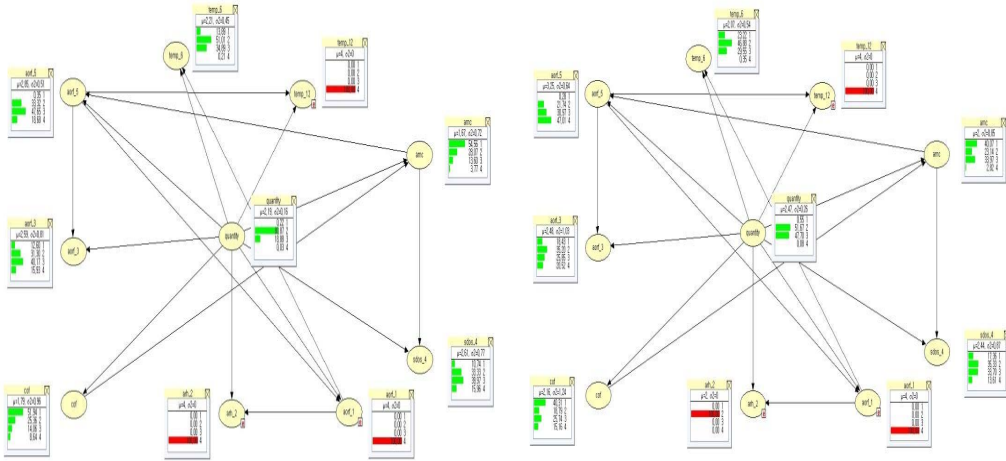
표 7. 양파 단수(quantity=Very High(4)/Very Low(1)) 변화에 따른 관련 변수의 확률분포 변화

설명변수	사전확률	사후확률	
		quantity=Very High(4)	quantity=Very Low(1)
12월 평균기온	Very Low(1)	11.79	14.66
	Low(2)	33.21	56.03
	High(3)	43.21	28.45
	Very High(4)	11.79	0.86
1월 강수량	Very Low(1)	27.50	42.24
	Low(2)	26.07	28.45
	High(3)	18.93	14.66
	Very High(4)	27.50	14.66
2월 평균상대 습도	Very Low(1)	26.64	42.24
	Low(2)	16.07	0.86
	High(3)	28.93	42.24
	Very High(4)	30.36	14.66
3월 강수량	Very Low(1)	26.07	14.66
	Low(2)	30.36	69.83
	High(3)	21.79	14.66
	Very High(4)	21.79	0.86
4월 합계 일조시간	Very Low(1)	11.79	0.86
	Low(2)	33.21	14.66
	High(3)	41.79	83.62
	Very High(4)	13.21	0.86
5월 강수량	Very Low(1)	24.64	14.66
	Low(2)	16.07	28.45
	High(3)	26.07	0.86
	Very High(4)	33.21	56.03
6월 평균기온	Very Low(1)	10.36	14.66
	Low(2)	41.79	0.86
	High(3)	41.79	56.03
	Very High(4)	6.07	28.45
비료비	Very Low(1)	43.21	0.86
	Low(2)	21.79	0.86
	High(3)	17.50	28.45
	Very High(4)	17.50	69.83
농약비	Very Low(1)	38.93	0.86
	Low(2)	26.07	0.86
	High(3)	21.79	56.03
	Very High(4)	13.21	42.24

자료: 저자 작성.

<그림 6-1>과<그림 6-2>, <표 8>은 시나리오 1, 2 상황에서 양파 단수의 사후확률 변화 정도를 제시하고 있다. 시나리오 1, 2 상황은 조건 1, 2는 동일하고 조건 3에서 차이가 나는데, 시나리오 1의 상황에서는 양파 단수의 사전확률에 비해서 낮은 구간(Low)의 양파 단수 사후확률이 증가하는 반면, 시나리오 2의 상황에서는 양파 단수의 사전확률에 비해서 높은(High) 구간의 양파 단수 사후확률이 증가하는 것으로 분석되었다.

그림 6-1. 기상 요인 변화에 따른 양파 단수의 확률분포 변화(시나리오 1)      그림 6-2. 기상 요인 변화에 따른 양파 단수의 확률분포 변화(시나리오 2)



자료: 저자 작성.

표 8. 기상 요인 변화에 따른 양파 단수의 확률분포 변화

시나리오	조건 1	조건 2	조건 3	양파 단수(quantity)		
				구간	사전확률	사후확률
시나리오 1	12월 평균기온 = Very High(4)	1월 강수량 = Very High(4)	2월 평균상대습도 = Very High(4)	quantity=Very Low(1)	4.64	0.22
				quantity=Low(2)	53.21	80.87
				quantity=High(3)	31.79	18.88
				quantity=Very High(4)	10.36	0.03
시나리오 2	12월 평균기온 = Very High(4)	1월 강수량 = Very High(4)	2월 평균상대습도 = Low(2)	quantity=Very Low(1)	4.64	0.55
				quantity=Low(2)	53.21	51.67
				quantity=High(3)	31.79	47.70
				quantity=Very High(4)	10.36	0.08

자료: 저자 작성.



한편, 엔트로피(Entropy)는 베이지안네트워크의 민감도 분석과 관련하여 주목받아 온 개념이다. 정보이론의 엔트로피에 의하면 주어진 정보가 적을수록 엔트로피가 증가하는데, 베이지안 기반 환경에 무질서의 정도를 적용시켜 새롭게 수식화함으로써 엔트로피를 최소화시킬 수 있는 방안을 도출하여 확률적으로 더 정확한 방법을 유도할 수 있다(허고은·정용규 2009). 즉, 엔트로피는 무질서 또는 불확실성에 대한 값을 의미하며, 불확실성에 대해 큰 영향을 줄 수 있는 변수를 결정하는 데 유용하다. Cover and Thomas(2006), Kjørulff and Madsen(2013) 등이 제시한 것처럼 노드 X의 엔트로피 함수  $H(X)$ 는 다음과 같다.

$$H(X) = - \sum_X P(X) \log P(X), P(X) \text{ 는 } X \text{의 확률분포 함수}$$

엔트로피 함수<sup>12</sup>에 기반하여 모형 내에서 양파의 단수 구간에 영향을 끼치는 변수의 영향력을 살펴 보면, 농약비(0.24)와 비료비(0.21)가 가장 많은 영향을 미치고 있으며, 다음으로 3월 강수량(0.13), 6월 평균온도(0.10), 5월 강수량(0.09), 1월 강수량(0.09), 4월 합계일조시간(0.08) 순으로 나타났다. 이러한 결과는 Marvin et al.(2016)의 연구 결과에서도 언급되었듯이 제안된 베이지안네트워크 모형에 활용된 변수들 중에서 엔트로피 함수 결과가 높은 변수들에 대한 자료 수집 노력 및 활용을 우선시해야 한다는 중요한 정보를 제공한다.

## 5. 요약 및 제언

### 5.1. 연구 결과 요약 및 결론

본 연구는 기상 요인과 영농투입재 비용 요인, 양파 단수의 영향 관계를 살펴보고, 특정 요인의 변화에 따른 다른 요인의 변동과 관련한 다양한 시나리오 결과 제시를 통하여 양파 재배 농민 및 관련 정책

<sup>12</sup> Hugin Expert 통계 패키지를 활용한 모형의 타당도 분석, 민감도 분석, 엔트로피 산출 방법은 홈페이지(<http://www.hugin.com>)의 Package user manual 참조.

수립자들의 효율적인 의사결정을 지원하기 위해서 수행되었다. 이와 같은 연구목적을 달성하기 위하여 본 연구에서는 국가통계포털과 기상자료개방포털을 통해서 기상 자료와 영농투입재 비용 자료, 양과 단수 자료를 수집하고, 양과 단수와 연관성이 뛰어난 특성변수를 선정한 후 패널회귀 분석과 베이지안네트워크모형을 적용한 분석을 진행하였다. 본 연구에 대한 분석 결과를 요약하면 다음과 같다.

첫째, 양과 단수결정모형 추정과 관련하여 패널회귀분석을 실시하였으며, Hausman 검정 결과 확률효과모형보다 고정효과모형이 더 적합한 것으로 밝혀졌다. 고정효과모형에 기초하여 생육 단계별 기상 요인과 양과 단수 사이의 관계를 살펴보면, 3월 강수량, 4월 합계일조시간, 5월 강수량이 증가하면 양과 단수가 통계적으로 유의하게 증가하는 것으로 나타났으며, 10월 평균상대습도, 1월 강수량, 6월 평균기온이 증가하면 양과 단수는 통계적으로 유의하게 감소하는 것으로 나타났다. 한편, 양과 영농투입재 비용 요인과 양과 단수의 관계에서는 비료비와 농약비 증가는 양과 단수를 통계적으로 유의하게 상승시키는 것으로 확인되었다.

둘째, 기상 요인과 영농투입재 비용 요인, 그리고 양과 단수 변수에 대해서 베이지안네트워크의 TAN algorithm을 활용하여 변수 사이의 상호의존성을 살펴보았다. 그 결과, 기상 요인과 영농투입재 비용 요인, 양과 단수 사이의 직·간접적인 관계를 나타내는 베이지안네트워크모형을 도출하였으며, 비순환 유향 그래프(DAG)와 조건부확률표(CPT)에 근거한 민감도 분석(what-if)을 통해서 양과 단수 구간 변화에 따른 설명변수의 확률분포 변화, 기상 요인 구간 변화에 따른 양과 단수의 확률분포 변화, 그리고 기상 요인과 영농투입재 비용 요인의 변화에 따른 양과 단수의 확률분포 변화 분석을 통해서 다양한 상황별 결과를 확보하였다. 특히, 엔트로피 함수를 활용하여 양과 단수와 상호의존성이 높은 요인으로 농약비(0.24), 비료비(0.21), 3월 강수량(0.13), 6월 평균온도(0.10), 5월 강수량(0.09), 1월 강수량(0.09), 4월 합계일조시간(0.08) 등으로 확인되었다.

이러한 분석 결과를 토대로 본 연구의 학문적 시사점을 제시하면 다음과 같다. 첫째, 본 연구에서는 기존 선행연구를 기반으로, 양과 관련 영농투입재 비용 요인 중 양과 단수와 기상 요인과 밀접한 관계가 있을 것으로 예상되는 비료비와 농약비를 설명변수로 추가하고, 생육단계별 밀접한 관계가 있을 것으로 예상되는 월별 기상요인을 세분화함으로써 기존 양과 단수 추정 모델에 대한 정교화를 추진했다는 점에서 기존 선행연구와 차별화된 학문적 의의를 지닌다. 둘째, 본 연구는 농업분야에서 데이터 마이닝 기법 중 베이지안네트워크모형을 활용해서 기상 요인과 영농투입재 비용 요인,

양과 단수의 상호의존성을 살펴본 초기 탐색적 연구라고 할 수 있다. 기상이변과 소비패턴의 변화 등으로 인하여 농산물 생산·소비에 대한 불확실성이 증가하고 있는 상황에서, 생산자인 농민과 정책수립자 등 다양한 이해관계자들이 발생 가능한 시나리오별 맞춤형 전략 수립과 관련하여 합리적이고 효율적인 의사결정에 도움을 줄 수 있는 방법론의 적용가능성을 제시함으로써 농산물 단수 예측과 관련한 연구에 새로운 이정표를 제시했다는 점에서 기존 연구와 구별되는 특징을 지닌다. 셋째, 본 연구는 패널회귀분석과 베이지안네트워크모형을 활용하여 양과 단수를 추정한 연구로, 다양한 연구방법론을 활용한 융복합 연구라는 점에서 선행연구와 차별점을 지닌다.

본 연구의 실무적 시사점은 다음과 같다. 첫째, 본 논문은 양과 단수를 중심으로 이루어진 기존 연구에서 상대적으로 관심을 갖지 않은 영농투입재 비용 요인을 모델에 새롭게 추가하여 영농투입재 비용 요인과 양과 단수의 관계에 대한 연구자들의 관심을 환기시키고, 새로운 시각에서 양과 단수와 관련한 문제를 이해하고 해결 방안을 모색할 수 있는 기회를 제공했다는 점에서 실무적 의의를 지닌다고 할 수 있다. 둘째, 베이지안네트워크모형은 특정 변수값을 조정하였을 경우, 베이지안네트워크모형 안에서 상호의존성 관계가 있는 변수들의 변화 정도를 직관적이고 알기 쉬운 그래프로 표시해 주기 때문에 실무자 측면에서 쉽게 이해하고 의사결정을 내리기에 유용한 방법이다. 따라서, 베이지안네트워크모형의 민감도 분석(what-if), 엔트로피(Entropy) 함수를 활용하여 양과 단수 최적화와 관련한 변수의 집합 또는 범위(수준)와 관련한 시각화 정보를 용이하게 도출할 수 있으며, 이를 통해 양과 수급과 관련한 의사결정 및 시나리오별 맞춤형 대응전략 수립에 상당히 유용하게 활용될 수 있을 것이다.

## 5.2. 연구의 한계 및 향후 연구 방향

본 연구는 학문적으로나 실무적으로 다양한 시사점을 지니고 있음에도 불구하고 몇 가지 한계를 지니고 있다. 첫째, 본 연구는 양과 생산비 중 영농투입재 비용과 관련해서 연도별·지역별 자료를 활용할 계획이었으나, 자료 확보의 어려움으로 인해서 연간 10a당 전국 평균 영농투입재 비용 자료를 지역별 자료로 대체하여 활용하였다. 향후 연구에서는 통계청 등 관련 기관 사이의 상호협력을 통해서 지역별 양과 영농투입재 비용에 대한 구체적이고 정확한 자료를 확보하고 이 자료를 활용할 수 있게 된다면, 자료의 타당성 확보 및 예측력 정확도 향상 관점에서 큰 도움이 될 것이라고 판단된다.



더불어 본 연구에서는 비료비, 농약비 등 생산비 요인 중 영농투입재 비용을 새로운 변수로 활용하였는데, 단수에 대한 영농투입재 비용 변수가 지니는 한계를 보완하고 단수에 영향을 주는 새로운 변수의 적용 가능성 검토 측면에서 향후 연구에서는 영농투입재 비용과 더불어 비료사용량, 농약사용량 등 영농투입재 사용량 변수를 활용한 연구의 진행이 필요하다고 생각된다.

둘째, 베이지안네트워크모형 분석과정에서 입력값으로 범주값(categorical value)을 적용하였는데, 선행연구들에서는 범주값 활용으로 인한 정확도 문제를 한계점으로 제시하고 있으며, 본 연구도 비슷한 상황이라고 할 수 있다. 따라서 향후 연구에서는 다양한 양과 작목 관련 전문가들과의 다면적 협의 과정을 통해서 범주 설정에 대한 정밀화 과정과 이를 통한 정확도 향상 노력이 추가되어야 한다고 생각된다. 한편 본 연구에서는 베이지안네트워크의 TAN algorithm을 활용하였다. 향후 연구에서는 NBN, GBN 등 다른 베이지안네트워크 algorithm을 비롯하여 C4.5 algorithm 및 CART algorithm 등을 비롯하여 이들을 활용한 앙상블 방법을 통해서 양과 단수 예측과 관련한 성능 비교 연구가 필요하다고 판단된다.

셋째, 본 연구에서 제시한 베이지안네트워크모형은 농업분야의 단수 예측과 관련해서 새로운 시도라고 할 수 있지만, 실무 활용도를 위해서는 모델의 정확도를 좀 더 향상시킬 필요가 있다고 판단된다. 이를 위해 향후 연구에서는 양과 단수와 관련한 좀 더 정확하고 다양한 요인(예: 생육정보 등)에 대한 빅 데이터를 수집하고, 체계화 및 고도화 과정을 통하여 농산물 단수 관련 베이지안네트워크 모형 분석에 반영해야 할 것이다.

마지막으로, 영농투입재 비용 요인의 주체인 농약, 비료는 기술적 진보의 결과라고 할 수 있는데, 본 연구에서는 기술변수가 포함되어 있지 않은 한계를 지니고 있다. 향후 연구에서는 단수를 설명하는 주요 변수로 기술변수를 연도변수로 설정하여 본 연구의 단수추정모형을 확장하는 연구로 진행할 필요가 있다고 생각된다.

## 참고 문헌

- 권영석, 김철우,곽정호, 최민선, 배상경, 황업지, 최인후. 2015. “양파 생육 특성 및 기상 모니터링을 통한 연차별 수량 예측.” 『한국원예학회 학술발표요지』, pp. 70-71. 한국원예학회.
- 김경현. 2005. “베이지안 네트워크에 기초한 백혈병 유전자데이터의 분석.” 서울대학교 공과대학 컴퓨터공학부 공학 학사학위논문.
- 김석, 박성훈, 양태현, 여기태. 2019. “패널회귀분석을 이용한 내항 화물운송사업체의 경영특성분석에 관한 연구.” 『디지털융복합연구』 제17권 제3호, pp. 79-92. 한국디지털정책학회. <http://doi.org/10.14400/JDC.2019.17.3.079>
- 김수경, 김상혁. 2018. “패널데이터 분석을 이용한 지역관광성과에 영향을 미치는 투자요인에 관한 연구.” 『관광연구논총』 제30권 제3호, pp. 23-44. 한양대학교 관광연구소. <http://doi.org/10.21581/jts.2018.08.30.3.23>
- 김용석, 심교문, 정명표, 최인태. 2014. “기온을 이용한 무 생산량 패널분석.” *Korean Journal of Agricultural Science* 제41권 제4호, pp. 481-485. 농업과학연구소. <http://doi.org/10.7744/cnujas.2014.41.4.481>
- 김지현. 2006. “베이지안 네트워크를 이용한 유료노인요양시설의 경영성과 예측모델 개발.” 이화여자대학교 대학원 간호학과 박사학위논문.
- 김현미, 정성환. 2013. “망막 질환 진단을 위한 베이지안 네트워크에 기초한 데이터 분석.” 『멀티미디어학회논문지』 제16권 제3호, pp. 269-280. 한국멀티미디어학회. <http://doi.org/10.9717/kmms.2013.16.3.269>
- 나명환, 임기안, 조용빈, 손찬수. 2019. “생육 재배 기간별 환경 인자의 주효과와 교호작용을 고려한 양파 단위면적당 생산량 예측 모델 개발.” 『한국품질경영학회 추계학술발표논문집』, pp. 162-162. 한국품질경영학회.
- 민병철. 2017. “베이지안 네트워크를 이용한 카셰어링 산업의 미사용 이탈요인 분석: 앱 내 행동 데이터를 중심으로.” 연세대학교 정보대학원 빅데이터전공 석사학위논문.
- 박영은, 김상훈, 박현정, 이동기. 2010. “한국영화의 수익성 영향요인에 대한 탐색적 연구.” 『경영학연구』 제39권 제2호, pp. 459-488. 한국경영학회. UCI:G704-000126.2010.39.2.003
- 박현정, 김상훈. 2011. “영화파일의 불법 다운로드를 결정하는 요인에 대한 연구.” 『소비문화연구』 제14권 제4호, pp. 1-20. 한국소비문화학회. <http://doi.org/10.17053/jcc.2011.14.4.001>
- 배은찬, 이건창. 2016. “앙상블 데이터마이닝 기법을 이용한 주식유동성 예측성과에 관한 실증연구.” 『한국컴퓨터정보학회논문지』 제21권 제6호, pp. 9-19. 한국컴퓨터정보학회. UCI:G704-001619.2016.21.6.006
- 선은정, 서영욱. 2019. “방어적·전략적 사회적책임활동과 기업의 재무적 특성변수와의 관계 분석.” 『기업경영연구』 제26호 제2권, pp. 79-98. 한국기업경영학회.
- 오승원, 김민수. 2017. “공간 및 기상정보 시계열 모형을 이용한 양파 생산량 예측.” *Journal of the Korean Data Analysis Society*. 제19권 제5호, pp. 2447-2456. 한국자료분석학회. <http://doi.org/10.37727/jkdas.2017.19.5.2447>
- 오재균, 여기태. 2019. “패널회귀분석을 이용한 글로벌 선사의 재무요인 특성분석에 관한 연구.” 『디지털융복합연구』 제17권 제4호, pp. 65-73. 한국디지털정책학회. <http://doi.org/10.14400/JDC.2019.17.4.065>
- 이건창, 최관. 2007. “감리지적기업의 분류적 특성에 관한 연구: 베이지안 망과 C5.0, 그리고 앙상블 방법간의 비교를 중심으로.” 『경영학연구』 제36권 제3호, pp. 705-737. 한국경영학회. UCI:G704-000126.2007.36.3.001
- 이건창, 이근영, 조남용. 2009. “베이지안 네트워크 분류와 비즈니스 프로세스 모델링을 통한 신용카드 회원 이탈에 관한 연구.” 『지식경영연구』 제10권 제4호, pp. 1-15. 한국지식경영학회. <http://doi.org/10.15813/kmr.2009.10.4.001>
- 이동기, 김상훈, 박영은, 박현정. 2010. “한국영화의 수익성 영향요인에 대한 탐색적 연구.” 『경영학연구』 제39권 제2호, pp. 459-488. 한국경영학회. UCI:G704-000126.2010.39.2.003

- 이은민, 이견창. 2017. “데이터마닝 분류기를 활용한 사용자의 과거 장소예측에 관한 연구.” 『한국컴퓨터정보학회 논문지』 제22권 제11호. pp. 97-104. 한국컴퓨터정보학회.
- 이중웅, 이영석. 1995. 『주요 양념채소의 식부면적과 작황 결정요인 분석에 관한 연구』. 한국농촌경제연구원.
- 이혜진, 홍우진, 조연진, 손찬수, 조용빈. 2019. “전남 지역 ‘18년·‘19년 작기 양파 생육상태 및 기상상황 비교.” 『한국작물학회 학술발표대회 논문집』. pp. 103-103. 한국작물학회.
- 정학근, 김창길, 문동현. 2013. “패널분석-확률효과모형에 의한 등숙기 이상기상이 쌀 단수에 미치는 영향 분석.” 『한국기후변화학회지』 제4권 제4호. pp. 317-330.
- 조재환, 서정민, 강점순, 홍창오, 김진호, 이상규. 2012. “기온 및 강수량 변동이 노지 건고추 단수에 미치는 영향.” 『한국환경과학회지』 제21권 제12호. pp. 1503-1510. 한국환경과학회. <http://doi.org/10.5322/JES.2012.21.12.1503>
- 최성천, 백장선. 2016. “공간 패널 회귀모형을 이용한 양파 생산량 추정.” 『응용통계연구』 제29권 제5호. pp. 873-885. 한국통계학회. UCI:G704-000408.2016.29.5.002
- 한국농촌경제연구원. 2020. 『농업전망 2020: 농업·농촌 포용과 혁신, 그리고 지속가능한 미래』. 한국농촌경제연구원 보고서.
- 황준성. 2010. “패널회귀분석을 이용한 국내 발전산업의 투입과 산출의 관계에 대한 연구.” 『글로벌경영학회지』 제7권 제2호. pp. 151-168. 글로벌경영학회.
- 허고은, 정용규. 2009. “효율적인 베이지안망 학습을 위한 엔트로피 적용.” 『한국인터넷방송통신TV학회 논문지』 제9권 제3호. pp. 31-36. 한국인터넷방송통신TV학회.
- Ashenfelter, O., D. Zimmerman, and P. B. Levine. 2003. *Statistics and Econometrics: Methods and Applications*. New York: John Wiley & Sons.
- Bouckaert, R. 1995. *Bayesian Belief Networks: From Construction to Inference, Doctorial Dissertation*. University of Utrecht, The Netherlands.
- Bouzembrak, Y. and Hans J. P. Marvin. 2016. “Prediction of Food Fraud Type using Data from Rapid Alert System for Food and Feed (RASFF) Bayesian Network Modelling.” *Food Control*. vol. 61, pp. 180-187. <https://doi.org/10.1016/j.foodcont.2015.09.026>
- Cheng, J. and R. Greiner. 1999. “Comparing Bayesian Network Classifiers.” Proc. of the 15th Conf. on Uncertainty in Artificial Intelligence, Morgan Kaufmann Publishers. pp. 101-107.
- Cover, T. M. and J. A. Thomas. 2006. *Elements of Information Theory(Wiley Series in Telecommunications and Signal Processing)*. USA: Wiley-Interscience.
- Hsiao, C. 2003. *Analysis of Panel Data*. New York: Cambridge University Press.
- Friedman, N., M. Geiger, and M. Goldszmidt. 1997. “Bayesian Network Classifiers.” *Machine Learning*. vol. 29(2-3), pp. 131-163.
- Heckerman, D., D. Geiger, and D. M. Chickering. 1995. “Learning Bayesian Networks: The Combination of Knowledge and Statistical Data.” *Machine Learning*. vol. 20, pp. 197-243. <https://doi.org/10.1007/bf00994016>
- Jensen, F. V. 1996. *An Introduction to Bayesian Networks*. UCL Press, London.
- Jensen, F. V. 2001. *Bayesian Networks and Decision Graphs*. NY: Springer-Verlag.
- John, G. H. and P. Langley. 1995. “Estimating Continuous Distributions in Bayesian Classifiers.” Proceedings of the Eleventh Conf. on Uncertainty in Artificial Intelligence(UAI), pp. 338-345. Morgan Kaufmann, Quebec, Canada.
- Kjærulff, U. B. and A. L. Madsen. 2013. *Sensitivity Analysis, In: Bayesian Networks and Influence Diagrams: A Guide*

to Construction and Analysis. New York, NY: Springer New York, pp. 303-325.

Marvin, Hans J. P., Y. Bouzembrak, E. M. Janssen, H.J. van der Fels-Klerx, E. D. van Asselt, and G. A. Kleter. 2016. "A Holistic Approach to Food Safety Risks: Food Fraud as an Example." *Food Research International*. vol. 89, pp. 463-470. <https://doi.org/10.1016/j.foodres.2016.08.028>

Morgan, M. G. and M. Henrion. 1990. *Uncertainty: A Guide to Dealing with Uncertainty in Quantitative Risk and Policy analysis*. Cambridge: Cambridge University Press.

Neapolitan, R. E. 2004. *Learning Bayesian Networks*. Pearson Prentice Hall.

국가통계포털. <<http://kosis.kr>>.

기상자료개발포털. <<http://data.kma.go.kr>>.

농촌진흥청 농사로. <[www.nongsaro.go.kr](http://www.nongsaro.go.kr)>.

Hugin Expert 통계패키지. <<http://www.hugin.com>>.

원고 접수일: 2020년 05월 11일
원고 심사일: 2020년 05월 22일
심사 완료일: 2020년 06월 25일