발 간 등 록 번 호
11-1543000-000328-01

젖소 이력관리 데이터를 활용한 원유생산 예측 모형 개발

농림축산식품부 축산경영과

서울대학교 산학협력단

요 약

본 연구는 기관별로 산재한 젖소사육 관련 통계자료를 이용하여 월별 산유량을 예측하는 모형을 개발하는 것을 목표로 한다. 월별 산유량은 월별 착유우 두수와 두당 착유량에 영향을 미치는 기후 요소 및 쿼터와 같은 환경변수에 의해 결정된다. 정확한 산유량을 예측하기 위해서 착유우 두수에 기반한 예측력 높은 산유량 모형을 개발하고, 미래의 착유우 두수를 예측하는 생존모형을 개발하는 것이 필요하다. 이 두 작업을 성공적으로 수행하기 위해서 반드시 필요한 작업이 특정시점의 착유우 두수를 정확하게 파악하는 것이다.

본 연구에서 착유우 두수를 추정하는 방법으로 축산물품질평가원에서 관리하고 있는 젖소/육우 이력데이터를 활용하는 방법을 제안한다. 젖소/육우 이력데이터는 특정 개체에 대해 출생, 양도/양수, 폐사/도축에 대한 사건을 발생시점별로 기록한 것이다. 이력데이터를 통해 1) 과거 특정시점에서 월별 착유우 두수를 파악하고 2) 미래 기대 착유우 두수에 대한 모형을 개발하였다.

미래 산유량을 예측하기 위해서 착유우 두수를 추정하고 그것에 기반한 산유량을 추정하는 이 단계 방법을 사용하였다. 먼저 산유량 모형은 2010년 6월부터 2013년 8월까지 젖소 이력데이터로부터 파악한 산차별 착유우 두수와 낙농진흥회에서 제공하는 월별 일평균 산유량-쿼터량 데이터를 이용하여 개발하였다. 그리고 젖소 이력데이터로를 이용하여 착유우 두수 추정모형을 개발하였다. 산유량의 추정치는 이력데이터로부터 젖소의 미래 기대 착유우 두수를 계산하여 산유량 모형에 대입하여 예측치를 얻는다.

현재 개발된 모형은 2011년 10월 이후부터 2013년 8월까지 12개월 원유 총생산량에 대해서 2.5%의 평균 오차율을 보였으며 2013년 9월~11월까지 6개월 평균 생산량에 대해 2.4%의 평균 오차율을 보였다. 과거 데이터가 구제역, 쿼터 제한 해제 등 원유 수급이 불안정했던 기간을 포함한 데이터를 사용함으로써 모형의 성능이 떨어진 것으로 판단되며 향후 더 많은 데이터가 축적되고 새로운 모형의 적합이 이루어진다면 이 예측오차율은 더 낮아질 것으로 기대하고 있다. 본 연구에서 제안한 젖소 이력데이터를 활용한 원유 생산 예측 모형에 따르면 2013년 9월~11월 예측 결과 50만 1천톤으로 실측치 51만 3천톤에 근접하였고 2014년 원유 생산량은 214만톤으로 예측된다.

이번 연구에서 사용한 데이터의 정제 및 모형적합 및 산유량 예측과정은 통계 프로그램인 R을 이용하여 구현하였으며 그 소스 코드를 보고서와 함께 제공하였다. 이 보고서에는 젖소 이력데이터를 정제하는 방법과 모형을 적합하는 과정에 대한 자세한 설명을 포함하고 있으며 몇 가지 가능한 시나리오 대해 산유량의 변화를 파악할 수 있게 하였다. 본 보고서의 내용을 참고하면 유업체에서 관리하고자 하는 농가집단에 대해서 예측모형을 타 프로그램을 통해 구현할 수 있을 것이다.

향후 소규모 지역단위의 산유량 예측을 위해서 개별적인 산유량 모형 및 생존 모형이 개발되어야 할 것이며 이를 위해서는 축산물품질평가원의 이력데이터와 낙농진흥회의 원유생산량-쿼터량 데이터의 매칭작업이 이루어져야 할 것이다.

목 차

1. 연구 개요
- 목적
- 연구배경
- 연구내용
  기존연구동향
  이력 데이터를 이용한 생산 예측 방법

2. 데이터 설명 및 데이터 정리작업
- 데이터 설명
- 데이터 변환작업

3. 모형개발
- 원유 생산량에 대한 회귀분석 모형
- 생존모형 분석
- 원유 수요량에 대한 회귀분석 모형

4. 생산/수요량 예측결과
- 생산량 예측결과
- 수요량 예측결과

5. 프로그램 설명 및 운용방법

1. 연구개요

1.1 목적

본 연구의 목적은 기관별로 산재된 낙농관련 전산자료를 융합, 젖소의 사육두수를 기반으로 한 원유생산 예측 모델을 구축하는 것으로 낙농관련 전산자료의 정보이용효율을 극대화하는데 있다. 원유생산 예측 모델에 사용될 전산자료는 도축이후에 이용될 목적으로 구축된 이력관리시스템의 젖소 사육정보, 젖소 검정 성적 관련 자료, 농가 생산량-쿼터량 자료가 있다.

낙농산업 선진화 대책 중 전국단위 수급 조절제의 효율적인 수급관리를 위하여 단기 및 중기 원유생산 전망에 대한 요구는 꾸준히 증가해왔다. 반면 원유 생산 예측은 계절성 및 기후면화의 민감성 등 원유수급 관련 변수들이 다양하여 정확한 중기 수급전망이 어려워 원유생산에 대한 지속적인 예측 모형 개발 및 관련 연구가 필요하다. 생산 예측 결과에 따라 원유 수급의 불안정을 예상하고 선제적 수급 안정 대책을 추진함으로써 산업의 피해를 최소화하고자, 젖소 이력관리 개체정보를 활용한 정확한 생산예측 모델을 개발하고 과학적인 원유수급 관리 체계를 구축하고자 한다.

본 연구는 이력관리 정보에 기반한 원유 생산/수요량 예측모형을 개발하고 프로그램으로 구현하는 것을 목적으로 한다. 프로그램 개발은 통계분석 프로그램인 R[1)]환경 하에서 이루어졌으며, 이 프로그램은 http://www.r-project.org/ 무료로 설치할 수 있다.

1.2 연구배경

현재 축산물물질평가원과 농협젖소개량부, 낙농진흥회 등 여러 기관이 산유량 예측과 관련된 정보를 수집하고 정리하고 있다. 이러한 정보의 축적으로 데이터에 의한 원유 생산 모델을 개발할 수 있는 환경 마련되고 있으며, 이 정보를 이용할 경우 기존 표본 추출 방법을 통한 원유 생산 예측보다 더 정확한 예측치를 얻을 것으로 기대하고 있다. 이를 위해서는 관련 정산정보를 통합하고, 통합된 데이터에 기반하여 통계적 예측 모형을 개발하는 것이 필요하다. 통합하고자 하는 전산 데이터는 젖소의 이력정보와 생산량/쿼터 데이터, 비유곡선 데이터이다. 뿐만 아니라 기후, 가격 등 여러 환경변수를 고려하여 유용한 통계적 모형을 개발하여 원유생산에 대한 보다 정확한 미래의 중/단기 예측을 하고자 한다.

1.3 연구내용

1.3.1 기존연구 동향

원유생산량에 대한 예측은 일본의 경우 지정단체 블록별 원유 생산량 예측치를 합산하여 산출한다. 지정단체의 블록별 예측치는 과거 원유 생산 실적 데이터 동향 패턴을 기본으로 기온이나 젖소 사육두수를 포함시킨 예측모델 (ARIMA model)에 의한 추계를 기본으로 산출한다. 미래 예측치를 추정하기 위한 기온과 같은 변수들은 평년 수준으로 설정한다. 한편 우리나라의 경우 젖소 사육두수 및 통계청에서 제공하는 착유우 두수 비율 추계자료를 이용하여 분기별 원유 생산량을 예측한다.

---

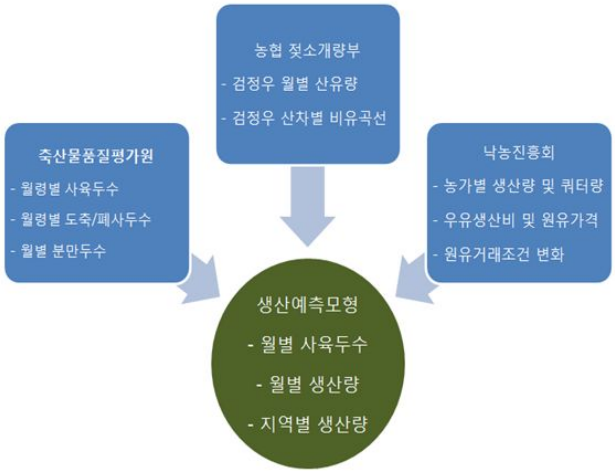
1) http://www.r-project.org/

1.3.2 이력데이터를 이용한 생산 예측방법

이력관리 시스템은 소의 원산지 및 사육농장을 파악할 수 있도록 소의 출생 및 양수, 양도, 도축 기록을 저장 및 관리하는 체계를 말한다. 이력관리 시스템은 소비자에게 제공되는 쇠고기의 추적관리를 통해 안심 먹거리를 제공하고자 개발한 것으로, 국내에서 사육되는 모든 소의 정보를 가지고 있다. 특별히 한우의 경우 품종 추적을 위해 출생 시 모개체 고유번호를 이력관리 시스템에 기록한다. 젖소 및 육우의 경우도 출생시 모개체번호를 기록하기도 하는데, 이 모개체번호가 젖소의 착유 시작 시점을 알 수 있는 정보다. 뿐만 아니라 출생 시 기록된 모개체번호로부터 알 수 있는 과거의 착유 시작 기록을 이용하면 특정 시점에서 젖소의 산차까지 알 수 있다.

농협 젖소개량사업소에서는 평균적인 착유기간, 건유기간, 그리고 착유기간 내 착유량의 변화에 대한 데이터를 제공한다. 이력데이터로부터 착유시작 시점이 파악된 개체에 대해서 평균적인 착유기간을 적용하면 특정 시점에서의 착유 유무를 알 수 있다. 만약 모든 소의 착유시작 시점을 파악할 수 있다면 평균적인 착유-건유 패턴에 의해 특정 시점의 산차별 착유우 두수를 알 수 있는 것이다.

과거시점에서 산차별 착유두 두수와 실제 원유 생산량의 관계를 모형화 하면 산유량 모형을 추정할 수 있다. 추정한 산유량 모형에 미래의 착유우 두수를 대입하면 예측 산유량을 얻을 수 있다. 산유량 모형의 추정은 3.1에서 미래 착유우 두수의 추정모형은 3.2에서 소개한다.



다음 장에서는 분석 시 사용하고자 하는 데이터에 관한 설명을 다루겠다.

2. 데이터 설명 및 데이터 정리

2.1 데이터 설명

본 연구에 사용하고자 하는 데이터는 젖소이력관리데이터, 육우관리데이터, 농가별 산유량-쿼터량 데이터, 유우군 능력검정 자료다.

2.1.1 젖소이력관리 데이터

젖소이력관리 데이터[2]는 축산물품질 평가원에서 제공하는 데이터로 농장번호(farm_id), 개체번호(cow_id), 사건(event), 날짜(DATE2), 모개체번호(M_cow_id) 필드를 포함하고 있다. 농장번호는 농장주소 및 소유주 정보를 알 수 있는 키값이다. 개체번호는 개체 식별 번호를 나타낸다. 사건은 소의 출생, 양수, 양도, 폐사 및 도축출하를 기록한다. 날짜는 사건의 발생 시점을 기록한다. 모개체번호는 소의 출생 시 어미소의 개체번호를 기록한다. 이 데이터를 통해서 특정시점의 젖소의 상태 (생존 여부, 소속농가, 월령)와 출생시 모개체의 번호를 알 수 있다.

farm_id	cow_id	event	DATE2	M_cow_id
000001	00000000012	출생	20010201	000000000010
000001	00000000012	양도	20010303	
000002	00000000012	양수	20010303	
000002	00000000012	도축출하	20010310	

2.1.2 육우관리데이터

육우관리 데이터[3]는 축산물품질평가원에서 제공하는 데이터로 개체번호(cow_id), 종류(type), 성별(sex), 현재상태(event_1), 출생일(date2), 모개체번호(M_cow_id) 필드로 구성되어 있다. 이 데이터를 통해서 젖소의 상태(생존여부)와 출생시 모개체번호를 알 수 있다.

cow_id	type	sex	event	DATE2	M_cow_id
41000000111	육우	수	사육	20020201	000000000015
41000000112	육우	수	폐사	20020203	000000000016
41000000113	육우	수	도축	20020203	000000000017
41000000114	육우	수	폐사	20020210	000000000018

2.1.3 농가별 산유량 및 쿼터 데이터

낙농진흥회에서 제공한 농가별 산유량 및 쿼터 데이터는 농가번호, 농가별 월별 일평균 산유량, 월별 일평균 쿼터량이 기록되어 있다. 이 데이터에 포함된 농가번호는 젖소 이력 데이터의 농가번호와 다르다. 즉 두 데이터의 농가 코드가 일치하지 않으므로 농가를 대

2) 본 연구에서는 2010년 6월~2013년 8월 생존소를 대상으로 분석하였다.
3) 본 연구에서는 2010년 6월~2013년 8월 생존소를 대상으로 분석하였다.

응시켜주는 추가적인 데이터가 필요하다. 이 두 데이터에서 농가코드에 대한 농가 대응 정보는 낙농진흥회에서 정리하였다.

2.1.4 한국 유우군 능력검정 자료

비유곡선에 대한 자료로서 산차별로 월별 산유량에 대한 정보를 가지고 있다. 본 자료는 한국 유우군 능력검정 사업보고서 자료를 이용하였다.

2.2 데이터 변환

젖소이력관리데이터와 육우관리데이터는 특정개체의 출생시점과 모개체번호를 추적할 수 있는데, 이 정보를 이용하여 모개체(젖소)의 착유시점을 파악할 수 있다. 그리고 유우군 능력검정 자료의 평균적인 초산연령, 착유기간과 건유기간을 활용하여 특정 개체의 상태(착유여부, 산차)를 추정할 수 있다. 특정 시점에서 젖소의 상태를 추정하기 위해서 데이터를 변환하는 작업을 설명한다.

2.2.1 착유사건의 기록

젖소의 출생기록에 모개체번호가 기록된 경우 그 모개체의 착유 시작을 기록할 수 있다. 출생신고는 송아지 출생 이후 1주일 이내로 이루어지기 때문에 착유시작 시점은 1주일의 오차가 발생할 수 있다. 착유일 400일로 하여, 착유기간을 착유 시작시점부터 이후 400일로 정하였다.

젖소 이력관리 데이터와 육우 관리 데이터에는 착유사건이 기록되어 있지 않기 때문에 착유사건은 새로 기록해야한다. 먼저 젖소 이력관리데이터의 출생사건을 고른 다음 모개체번호(M_cow_id)를 개체번호(cow_id)로 기록한다. 다음 사건(event) 필드에 착유를 기록하고 날짜(DATE2) 필드에 개체 출생날짜를 기록한다. 그리고 그 모개체가 그 시점에 어떤 농가에 있는지 추적하여 농가번호(farm_id)를 기록한다. 다음으로 육우관리 데이터를 변환한다. 육우관리 데이터 역시 모개체 번호를 이용하여 젖소 이력관리 데이터에서 착유사건을 기록하는 방법과 동일하게 착유사건을 기록한다

예) 다음과 같은 이력데이터의 기록이 있다고 하자

farm_id	cow_id	event	DATE2	M_cow_id
000001	00000000012	출생	20010201	000000000010

모개체번호(M_cow_id)가 000000000010이므로 cow_id에 기록한다. farm_id는 그 개체가 있는 농장번호를 찾아 기록한다. 여기서는 000002라고 가정하자 사건(event)필드에는 착유라 기록한다. 그리고 날짜(DATE2)는 위 출생일을 기록한다. 새로 생성된 데이터는 아래와 같다.

farm_id	cow_id	event	DATE2	M_cow_id
000002	00000000010	착유	20010201	.

2.2.2 건유사건의 기록

건유사건은 이력데이터를 통해 직접 파악할 수 없다. 건유는 다음 착유를 위해서 젖소를 휴지시키는 사건으로, 착유종료 시점을 통해 정의할 수 있다. 이력데이터는 착유 시작만 기록되어 있으므로 착유의 종료는 데이터 분석에서 정의해야하는 사건이다. 본 연구에서는 착유일을 착유 시작시점 이후 400일로 정의했고 건유시작 시점을 착유일 400일 이후 시점으로 하였다.

건유사건은 착유일 400일 이후 생존한 소를 대상으로 기록한다. 건유사건을 기록하기 위해서는 이력데이터에 착유사건이 기록되어 있어야 한다. 이력데이터를 개체번호와 날짜로 정렬한다. 이 정렬을 통해 데이터는 개체별로 시간 순으로 정리한다. 착유시작 400일 이후에 생존한 경우(폐사, 도축출하 사건이 착유시작 400일 이후에 있는 경우) 착유시작일 + 400일 시점을 건유시작 시점으로 날짜(DATE2) 필드에 기록한다.

예) 2001년 2월 1일 착유기록이 있는 00000000010에 대해 400일 후 2002년 3월 8일에 생존해있다고 가정한다. 그러면 다음과 같은 건유기록이 추가된다.

farm_id	cow_id	event	DATE2	M_cow_id
000002	00000000010	착유	20010201	.
000002	00000000010	건유	20020308	.

2.3 데이터 변환 시 고려사항

2.3.1 모개체번호 결측치

젖소 이력관리 데이터 및 육우관리데이터의 모개체번호의 결측치가 많다. 출생신고 시 모개체번호 기록에 강제성이 없기 때문에 많은 송아지 출생기록에 모개체번호가 누락되어 있다. 모개체번호의 결측은 데이터 변환과정에서 다음 문제를 발생시킨다.

- 착유우 두수의 과소평가: 모개체번호의 누락되면서 착유시작 사건이 누락된다.
- 건유기간의 과대평가: 건유시작 이후 착유시작 사건이 누락되면서 건유기간의 길어진다.
- 산차의 과소평가: 착유시작 사건이 누락되면서 산차가 과소평가 된다.

착유우 두수 및 건유기간의 수정은 이력데이터에서 수정하지 않는다. 기록된 사건간의 시간간격을 통해 누락된 사건을 보정하는 작업을 고려할 수 있으나 착유우 두수의 예측 및 산유량 예측모형에 심각한 편이를 발생시키는 것으로 파악되어 개체별 사건 수정은 하지 않는다. 대신 시점별로 파악한 착유우 두수, 건유우 두수를 직접 조정하는 방법을 적용한다. 이는 5장 프로그램 설명 부분에서 자세히 다룬다. 산차의 과소평가 문제는 연령을 통한 산차기록으로 해결할 수 있다.

2.3.2 다태분만

착유기록은 모개체번호가 기록된 출생정보로부터 파생되는 것으로 다태분만의 경우 같은 한 모개체의 착유시작 사건이 중복되어 기록된다. 이력데이터를 통해 착유사건을 기록한 후 한 개체의 착유시작사건이 한 시점에 중복되어 기록되어 있는지 혹은 인접한 날에 기록되어 있는지 확인하고 제거하여야 한다

2.3.3 지연신고

개체의 출생신고가 지연되고 그 개체의 모개체가 분만이후 1주일 이내에 도축출하 혹은 폐사 신고된 경우, 개체 사망이후에 착유가 기록된다. 개체 사망이후 모든 기록을 제거하는 작업을 통해 착유사건을 지운다.

2.3.4 산차의 기록

산차(phase)는 개체별로 착유사건과 함께 기록한다. 착유기록에 의해 산차를 기록할 경우 결측 문제로 항상 과소평가된 개체별 산차를 기록하게 되므로 본 연구에서는 평균적인 젖소의 생애주기에 따라 산차를 기록한다. 평균적인 젖소의 생애주기는 다음과 같다.

- 1산 : 출생후 870일
- 착유간격: 460일

그리고 건유 기록이 있는 경우 직전 착유의 산차를 사용하여 데이터를 채워 넣는다.

예) 00000000010개체는 1997년 10월 3일에 출생하였고 2001년 2월 1일에 착유한 기록이 있다. 출생후 1217일 지난 후 착유이로 870일 1산 시작 1330(870+460)일 2산 시작일 중 2산 시작일에 가깝다. 이 착유기록의 산차(phase)에 '2'를 기록한다. 한편 건유사건이 있으므로 직전 착유에 기록된 산차 '2'를 기록한다.

farm_id	cow_id	event	DATE2	M_cow_id	phase
000001	00000000010	출생	19971003		
000001	00000000010	착유	20010201		2
000001	00000000010	건유	20020308		2

3. 모형의 개발

3.1 산유량 회귀분석 모형

본 연구에서는 일평균 산유량에 대한 선형 회귀 모형을 제안한다. 먼저 다음 변수를 정의한다.

· y_t: t 시점의 일평균 산유량

· $phase1Cow_t$: t 시점의 1산차 착유우 두수

· $phase2Cow_t$: t 시점의 2산차 착유우 두수

· $phase3Cow_t$: t 시점의 3산차 이상 착유우 두수

· $temp_t$: t 시점의 월평균 기온

· hum_t: t 시점의 월평균 습도

· $quota_t$: t 시점의 일평균 쿼터량

· ϵ_t: 오차

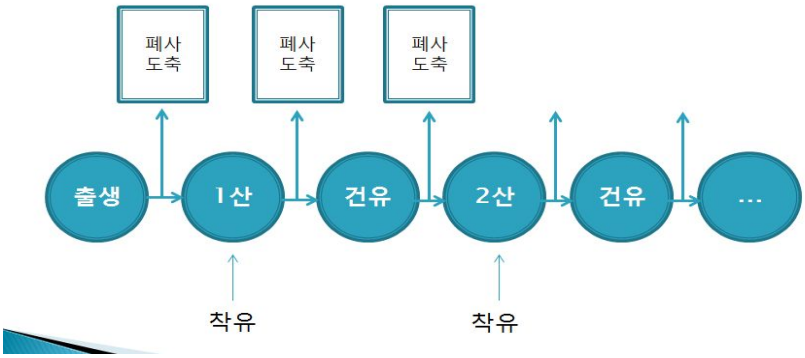
이 변수로 표현된 일평균 산유량에 대한 회귀모형 식은 다음과 같다.

$$y_t = \beta_0 + \beta_1 phase1Cow_t + \beta_2 phase2Cow_t + \beta_3 phase3Cow_t + \beta_4 temp_t + \beta_5 temp_t + \beta_6 quota_t + \epsilon_t$$

회귀 계수는 2010년 6월부터 2013년 8월까지 데이터를 이용하여 최소제곱법을 통해 추정하였다

3.2 생존 모형 분석

산유량 회귀모형을 통해 미래의 산유량을 예측하기 위해서는 미래의 산차별 착유우 두수가 필요하다. 미래의 산차별 착유우 두수를 얻기 위해서 젖소의 산차별 생존모형을 이용한다. 먼저 젖소의 생존확률 구하기 위해서 젖소의 생애주기를 출생, 착유, 건유, 폐사 및 도축출하의 사건 열로 정의하고 각 사건간의 전이확률을 생존모형을 통해 분석한다. 미래의 기대 착유우의 두수는 특정 시점에서 어떠한 젖소가 미래시점에 착유상태에 생존해있을 기대 착유우 두수라고 정의한다.



3.2.1 Cox 비례 위험 모형

젖소 생애주기에서 각 사건간 생존 확률은 다양한 공변량에 의해 영향을 받는다. 온도, 습도, 유대가격, 퀴터량등 다양한 변수에 의해 생존확률이 변화하는 모형을 고려하고자 Cox 비례 위험모형을 사용하였다. Cox 비례 위험 모형에서 생존함수 $S(t)$는 다음과 같이 정의된다.

$$S(t|Z) = \exp(-\int_0^t \exp(\gamma^T Z(u))\lambda_0(u)du)$$

$Z(t)$: 공변량 벡터

γ: Cox 비례 위험 모형의 회귀 계수

$\lambda_0(t)$: 기저 위험함수

본 연구에서는 월별 생존 시간이라고 가정하였므로 생존 함수 $S(t)$를 다음과 같이 정의하였다.

$$S(t|Z) = \prod_{u \le t}(1 - dA(u))^{\exp(\gamma^T Z(u))}$$

$$dA(t) = \frac{S(t) - S(t-)}{S(t)}$$

만약 어떤 젖소가 t_0 시점까지 살아 있다는 것이 관측되었다면 그 젖소가 t 시점까지 생존할 확률은 $S(t|Z)/S(t_0|Z)$이다. 본 연구에서는 산차별, 사건별로 생존모형을 구하였으며 5장과 부록에 Cox 비례 위험모형을 통해 추정한 계수 $\hat{\gamma}$ 와 $dA(t)$에 대한 정보가 수록되어 있다. 젖소의 생존확률을 $\hat{\gamma}$와 $dA(t)$를 이용해 계산할 수 있다.

3.2.2 예측 착유두수의 보정

미래 시점의 예측착유우 두수는 특정시점에 생존해 있는 소에 생존확률을 곱하여 만들어진 기대생존 착유우 두수다. 모형적합 결과 착유우의 과대/과소 추정을 피하기 위해서 과거 시점에서 추정한 기대 생존 착유우의 두수와 관측착유우 두수의 회귀분석으로 착유수 수를 보정한다. 이에 대한 추가적인 연구가 필요하다.

- 보정 착유우 두수(1산) = 4.705×10^4 + 0.346×기대착유우 두수(1산)
- 보정 착유우 수(2산) = 2.202×10^4 + 0.651×기대착유우 두수(2산)
- 보정 착유우 수(3산이상) = -1.658×10^4 + 1.413×기대착유우 두수(3산이상)

3.3 원유 수요 모형

본 연구에서는 제품별 원유의 연간 수요량에 대한 회귀모형을 도출하였으며 모형식은 [참고문헌]을 이용하였다. 모든 제품에 대한 가격정보를 얻을 수 없어 제품의 성격 및 가격의 시계열 변화를 고려하여 [표 7] 같이 제품군에 대한 회귀모형을 고려하였다.

제품군	시유	조제분유	치즈군	발표유군	기타
유제품 항목	시유	전지분유	가공치즈 자연치즈	발효유, 연유,전지분유, 탈지분유	버터

[표 8] 제품군 분류

예측 모형은 다음과 같다.

$$\log(y_t) = \beta_0 + \beta_1 \log(\mathrm{pop_t}) + \beta_2 \log(\mathrm{popYoung_t}) + \beta_3 \log(\frac{\mathrm{price_t}}{\mathrm{conv_t}}) + \beta_4 \log(\mathrm{gdp_t})$$

y_t: 수요량

pop_t: 인구

$popYoung_t$: 19세 이하 인구

$price_t$: 제품 가격

$conv_t$: 소비자물가지수

gdp_t: 국민총생산

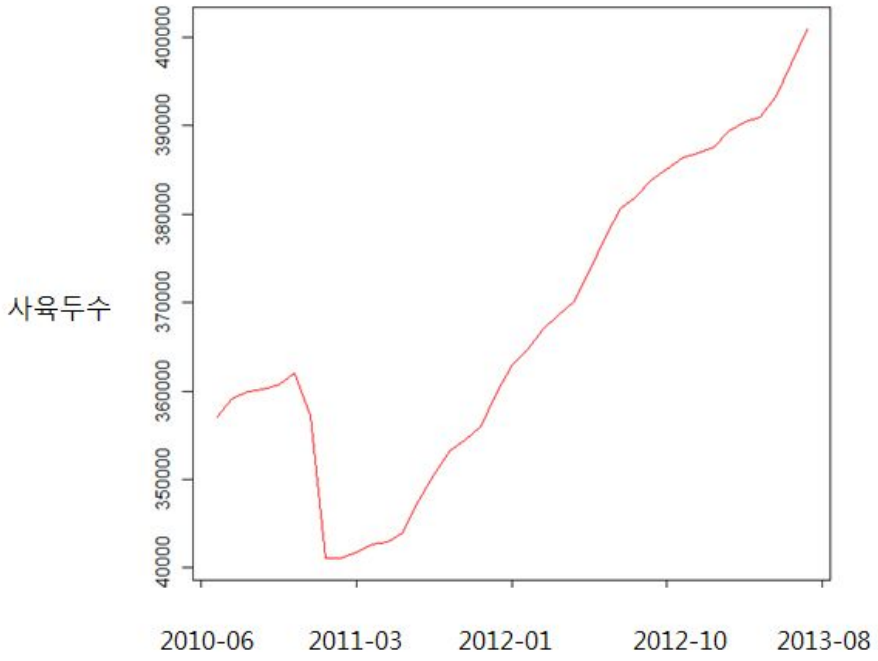
회귀계수의 추정은 2001년~2013년 자료를 이용하여 최소제곱법으로 제품군별로 추정하였다.

4. 원유생산량/수요량 예측 결과

4.1 생산량 예측결과

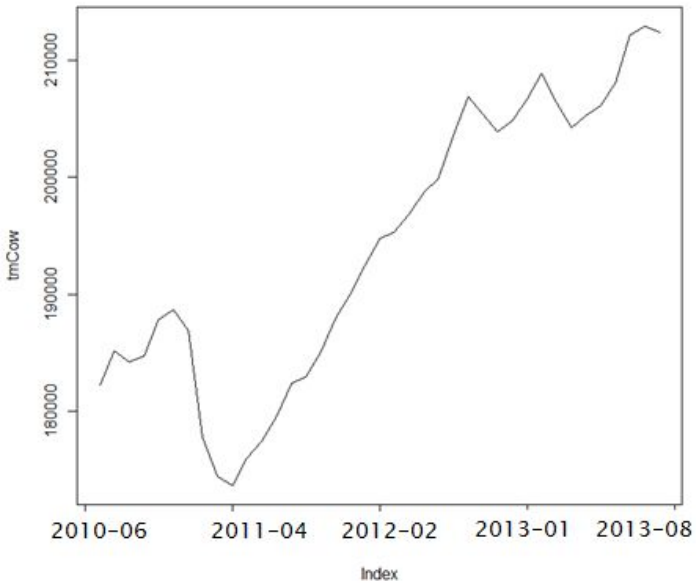
4.1.1 탐색적 자료분석

젖소 이력데이터를 통해 총사육두수의 변화를 파악할 수 있다. 2010년 12~2012년 2월 총사육두수가 34만두까지 급감한 이후 꾸준히 증가하고 있음을 알 수 있다.



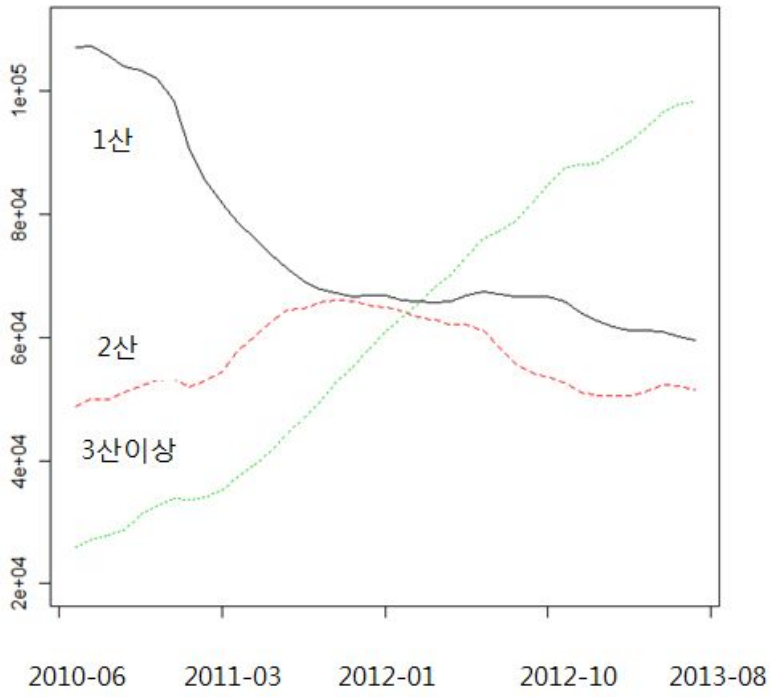
[그림 3] 총 사육두수

1산 착유시작일 29개월, 착유기간 400일, 건유기간 60일로 가정하여 젖소이력데이터와 육우이력데이터를 이용하여 착유우 두수를 계산할 수 있다. 2011년 4월 착유우 두수가 최저에 도달할 이후 2013년 8월까지 착유우 두수가 증가하고 있음을 알 수 있다. 증가추세는 2013년 1월 이후 감소하였다.



[그림 4] 착유우 두수

연령 기준에 의한 산차별 착유우 두수를 파악할 수도 있다. 1산차 착유우 두수는 2010년 6월 이후 지속적으로 감소하고 있으며 2산차 착유우 두수는 2012년 최고점에 도달한 이후 감소한다. 3산차 이상의 착유우 두수는 2010년 6월 이후 증가하고 있다.



[그림 5] 산차별 착유우 두수

4.1.2 모형적합 결과

회귀 모형 적합결과는 [표 8]과 같다.

회귀 모형

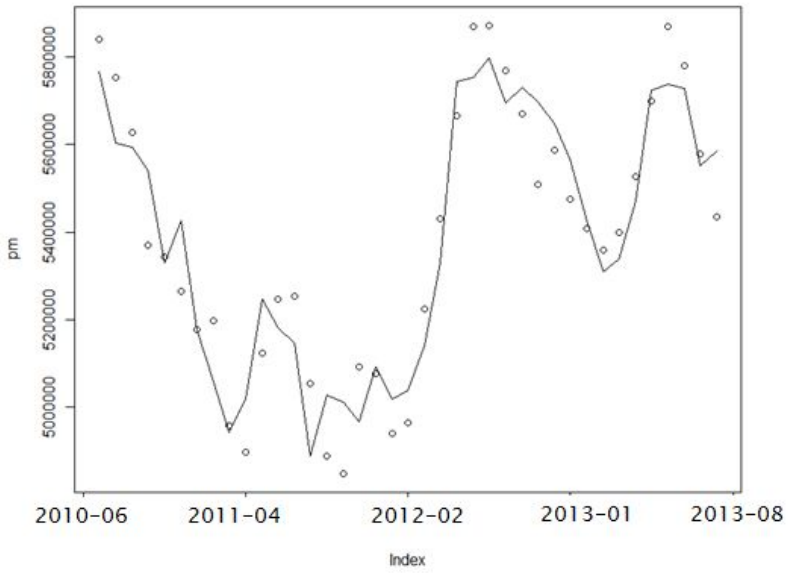
상수	phase1Cow	phase2Cow	phase3Cow	temp	hum	quota	temp3
-3.55×10^{7}	4.34×10^{1}	2.013×10^{1}	1.76×10^{1}	2.879×10^{4}	-5.08×10^{4}	5.45×10^{6}	-4.71×10^{4}

[표 9] 회귀모형계수

- phase1Cow: 1산차 착유우 두수
- phase2Cow: 2산차 착유우 두수
- phase3Cow: 3산차 이상 착유우 두수
- temp: 기온
- hum: 습도
- quota: 일평균 쿼터량
- temp3: 더미변수(평균기온 5도이하 또는 25도 이상)

회귀모형 계수는 산차별 착유우 두수에 대해서 양의 값을 가지고 기온과 쿼터에 대해서 양의 값을 갖고 습도와 더미변수 temp3에 대해서는 음의 값을 갖는다. 착유우 두수가 증가하거나 쿼터량이 증가하면 총 산유량이 증가하며, 기온은 5도에서 25도 사이에서는 기온 상승이 산유량을 증가시키지만 저온이나 고온에서 산유량을 떨어뜨리는 것으로 나타난다. 그리고 습도는 건조할수록 산유량이 증가하는 것으로 나타났다.

미래의 산유량을 예측하기 위해서는 특정시점의 착유우 두수를 기준으로 미래의 기대 착유우 두수를 계산해야 한다. 기대 착유우 두수는 3.2.1의 생존모형에 의해서 계산되며, 생존 모형은 5장에 자세히 설명되어 있다. 모형적합결과에 대해서 간단히 설명하면, 송아지의 경우 겨울철에 생존확률이 떨어지며, 착유우의 경우 여름에 생존확률이 떨어지는 것으로 나타난다. 그리고 구제역 발생기간에 젖소의 생존률이 급감한 것으로 보인다. 이 생존모형을 통해 계산된 산차별 착유우 두수는 [표 8]의 회귀모형에 대입되어 예측 산유량이 산출된다. 그림 7은 과거에 관측된 착유우 두수를 이용하여 적합한 결과다. 그림 7의 실선 즉 적합된 회귀모형이 산유량의 변화 추이를 잘 반영하고 있음을 알 수 있다. 2012년 1월을 기준으로 산유량을 기대 착유우 두수와 회귀모형으로 예측한 결과 6개월, 1년간 산유량 예측 오차가 약 2.4%, 2.3%로 나타났다. 이 결과는 향후 데이터가 축적되면서 더 개선될 것으로 기대하고 있다.



[그림 6] 일 평균 생산량 그래프

4.1.3 월별 산유량 예측치

원유 생산량은 젖소 산차별 생존모형과 산유량 회귀모형에 의해 월별로 예측된다. 2013년 9월부터 2년간 월별 산유량을 예측한 결과다. 2014년 2월~4월 산유량이 다소 감소하며 2014년 6월부터 12월까지는 산유량이 증가하는 추세할 것으로 예상된다. 모형을 통해 적합된 결과에 의하면 2013년 9월부터 11월 까지 원유생산량이 50만 1천톤으로 실측치는 51만 3천톤으로 추정되었다.

	2013년 9월	2013년 10월	2013년 11월
실측치	16.86	16.65	16.64
예측치	17.05	17.47	16.85

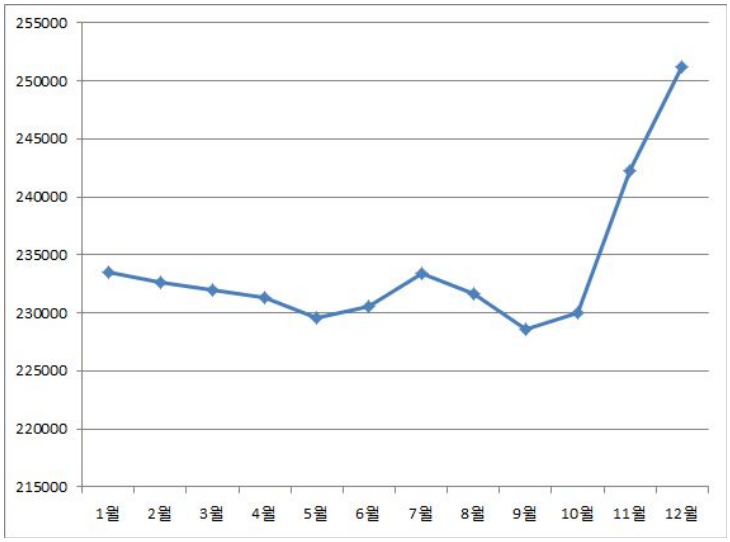
[표 10] 2013년 후반기 실측치와 예측치 비교 (단위: 만톤)

2014년 착유우 두수는 10월까지 감소, 증가, 감소의 추세를 따른다. 본 연구의 모형에 따르면 2014년 말 착유우 두수가 큰 폭으로 증가하는 것으로 보인다[4]. 이는 산차별 착유우 두수가 동시에 증가하는 구간으로 착유우 두수 예측에 주의가 필요하다.

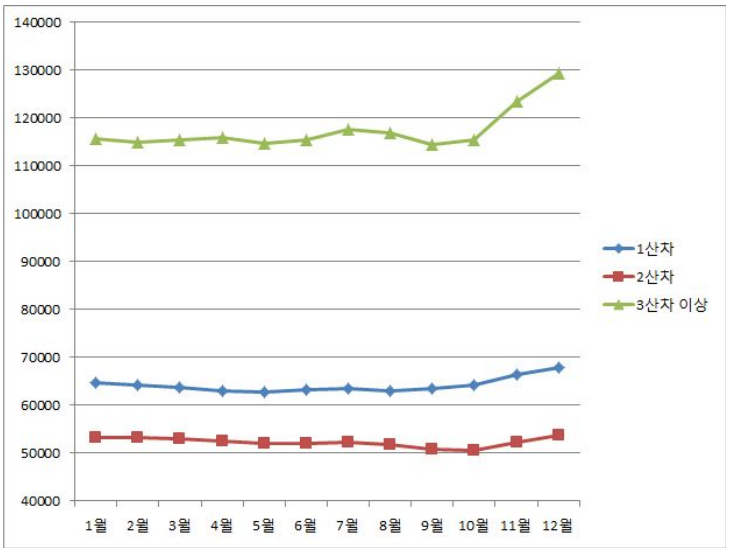
이렇게 예측된 착유우 두수를 이용하여 [표 8]의 회귀모형에 대입하면 예측 생산량이 나온다. 2014년 예측된 월별 일평균 생산량과 월별 총생산량은 [그림 10]과 [그림 11]에 나와 있다. 월별 예측 총생산량은 1월부터 5월까지 산유량이 증가하고 무더운 6월부터 10월까지 산유량이 떨어지며 다시 가을부터 산유량이 증가하는 계절적 요인이 잘 반영되어

4) 본 연구에 사용된 생존모형은 선택적 도태와 같은 목장 경영자의 제약조건 내 의사결정과정을 반영하지 않았다. 따라서 착유우수가 급증하는 구간에서는 착유우 수에 대한 전문가의 보정작업이 필요하다.

있다. 그리고 월 총 생산량의 추세를 보면 하반기로 갈수록 산유량이 증가하는 추세에 있음을 알 수 있다.



[그림 7] 2014년 착유우 두수 예측치

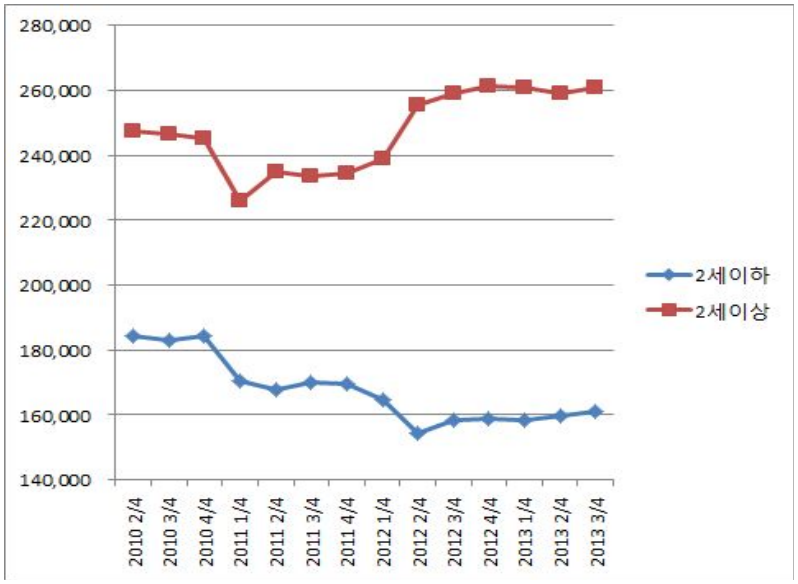


[그림 8] 2014년 산차별 착유우 두수 예측치

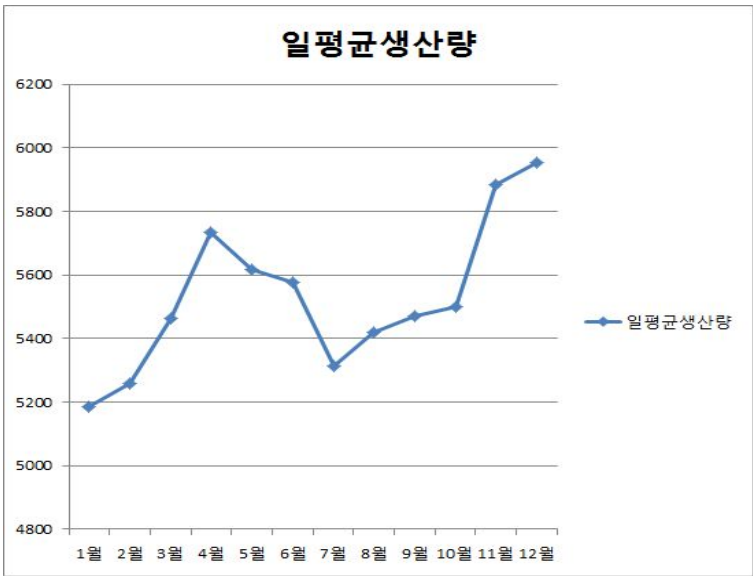
2014년 기온, 습도, 쿼터량이 2013년 수준과 같은 경우, 본 연구에서 제안하는 2014년 원유 예측생산량은 214.2만톤이다.

※ 통계청 가축동향 조사의 연령별 젖소분포 현황에 따르면 2012년 2분기(6월)에서 2012년 4분기(12월) 사이 24개월령 이상 젖소의 수가 22000여두 증가하였다. 이 증가분은 2012년 1산차 및 2산차 착유우 두수의 증가에 영향을 주며 2014년 하반기에는 3산차이상 착유우 두수에 영향을 주게 된다. 한편 2세 이하의 젖소가 2012년 2분기부터 소폭 증가추세에 있어 2014년 1산 착유 상태로 진입하게 된다. 젖소개체가 전반적으로 증가하는 경우, 농장의 젖소 사육능력에 따라 선택적 도태를 하는 것이 일반적인데, 본 연구에서

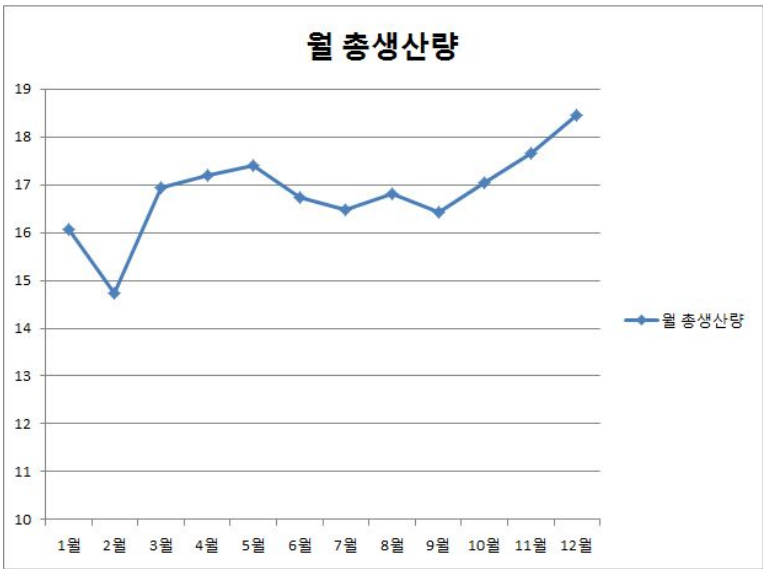
제안한 생존모형은 이 현상을 반영하지 않고 있다. 즉, 현재 추정한 생존확률 모형에 의해 계산된 기대 착유우 두수가 2014년 하반기에 과대산출될 가능성이 있으며, 이는 차후 연구를 통해 확인하거나 개선해야 할 부분이다.



[그림 9] 연령별 젖소 분포 현황 (자료: 통계청 가축동향조사)



[그림 10] 2014년 일평균 생산량 (단위:톤)



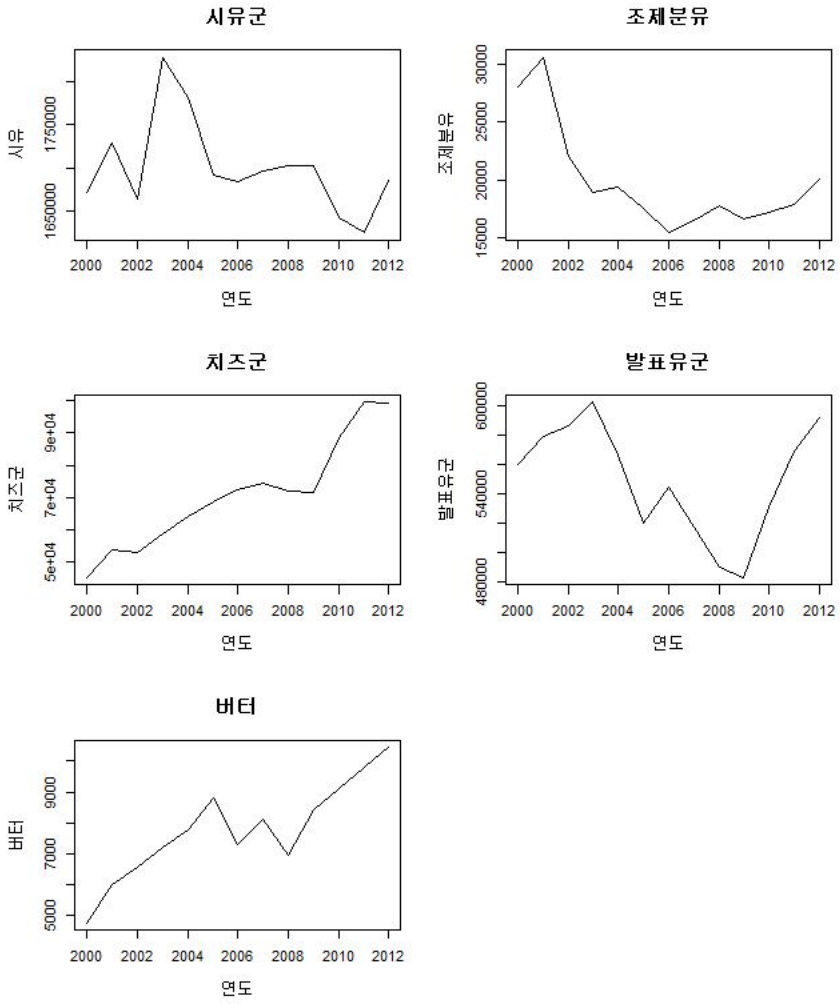
[그림 11] 2014년 월 총생산량 (단위: 만톤)

4.2. 수요량 예측 결과

4.2.1 탐색적 자료분석

제품군별로 연간 국내 수요량을 조사한 결과 시유, 조제분유는 2000년대 후반 수요가 감소하고 있으며, 발효유, 치즈 및 버터는 수요가 증가하고 있는 것으로 파악되었다.

4.2.2 수요량 예측모형



제품군은 3장 수요모형에서 설명한 바와 같이 시유, 조제분유, 치즈군, 발효유군, 기타군으로 나누어 적합하였다. 모형적합 결과는 아래 [표10]과 같다.

제품군	상수	log(인구)	log(20세미만인구)	log(가격/물가지수)	log(GDP)
시유	32.746	0	−0.9545	−0.3010	−0.2590
조제분유	−824.105	49.1904	0	0.2636	−4.6539
치즈군	8.5273	−0.2385	0	−0.4746	0.6997
발효유군	66.891	−2.7659	0	0.1364	−1.1504
버터	−10.48	0	0.8130	−0.9688	1.2140

[표 11] 제품 군별 수요량 예측 모형

2014년 인구는 통계청 2014년 추계인구 자료를 이용하고 2013년 수준으로 가격과 물가지수가 유지되며 2013년 1인당 GDP가 2% 성장한다는 가정 하에 각 제품군의 수요예측치는 [표 11] 와 같다.

- 2014년 추계인구 : 50,423,955명
- 2014년 20세 미만 추계인구 10,463,466명
- 2014년 추계 GDP :25810 달러

제품군	시유군	분유군	치즈군	발효유군	기타군	수요총계
수요량	171.54	2.171	10.176	60.52	106.43	350.83

[표 12] 2014년 제품군별 국내수요량 예측결과

단위: 만톤

5. 프로그램 설명 및 운영방법

이번 장에서는 개발한 프로그램을 구성하는 세부 모듈의 기능을 기술하고 각 세부 모듈이 처리하는 입력 데이터의 출력 데이터의 형식에 대해 설명한다. C, C++, JAVA 등 다른 프로그램으로 아래에서 설명하고 있는 데이터의 입출력 과정을 구현하고, 입력데이터를 형식에 맞게 관리한다면 이력정보를 이용한 산유량 예측 시스템을 자체개발, 운용할 수 있을 것이다.

본 연구에서 개발한 프로그램은 1) 젖소 이력데이터, 육우의 데이터의 변환 및 저장, 2) 월별 데이터의 생성, 3) 산유량에 대한 회귀모형의 추정 및 회귀모형의 저장, 4) 생존함수의 추정 및 생존모형의 저장, 5) 변환된 이력데이터, 저장된 생존모형, 회귀모형을 이용한 예측치 생성, 네 가지 작업을 하는 모듈로 구성된다.

5.1 젖소 이력데이터, 육우 데이터의 변환 및 저장

5.1.1 입력 데이터 형식

입력데이터는 각 기관으로부터 받은 원본 데이터를 기준으로 설명한 것으로 현재 개발한 프로그램을 수정 없이 활용하기 위해서는 반드시 데이터의 필드형식을 지켜 데이터를 관리해야 한다.

- 젖소 이력데이터
  · 필드: 이력제_농가코드, 이력제_농가명, 이력제_농장명, 이력제_사육지주소, · 기준일자, 개체식별번호, 모개체번호, 신고구분
  · 필드 구분자: |
  · 필드 세부사항
  · 이력제_농가코드: 젖소 이력데이터의 농가코드
  · 기준일자: 8자리 ex) 20111020
  · 개체식별번호: 12자리 식별번호
  · 모개체식별번호: 12자리 식별번호
  · 신고구분: 출생신고, 폐사, 도축출하, 양도, 양수

- 육우 데이터
  · 필드: 개체식별번호, 소의종류, 성별, 현재상태, 출생일자, 모개체번호
  · 필드 구분자: |
  · 필드세부사항
  · 개체식별번호: 15자리
  · 출생일자: 10자리 ex) 2011-10-20
  · 모개체번호: 15자리, 앞 세자리를 제거하면 젖소이력데이터의 개체식별번호 코드가 됨. ex) 육우 데이터의 모개체번호가 410000187207766 기록되어 있으면 이 소는 젖소 이력데이터의 개체번호는 000187207766로 표시된다.

5.1.2 데이터의 변환

· 착유사건을 기록한다.
· 출생신고가 안된 소를 제거한다.
· 다태출생으로 중복된 착유기록을 제거한다.
· 착유가 출생신고 보다 먼저 기록되거나 폐사, 도축출하 이후에 착유기록된 것을 지운다.
· 착유에 대한 산차를 기록한다.
· 건유를 기록한다.
· 건유에 대한 산차를 기록한다.
· 데이터를 개체에 대해 시간 순으로 정렬한다.

※ 착유/건유기록, 데이터 정리 및 산차 기록은 2장 데이터 설명 및 데이터 정리작업의 데이터 변환 부분을 참고한다.

5.1.3 변환 결과

변환된 파일을 farm_id, cow_id, DATE2, event, phase, tmp 필드로 구성되어 있다. farm_id는 젖소 이력데이터의 농가코드, cow_id는 젖소개체번호, DATE2는 날짜, event는 젖소 이력데이터의 신고구분, phase는 산차, tmp는 신고구분과 산차를 이용해 만든 데이터다.

예)

farm_id	cow_id	DATE2	event	phase	tmp
347413	201525	17682	출생신고	0	출생신고
347413	201525	18647	폐사	0	폐사
257036	234865	16931	출생신고	0	출생신고
257036	234865	18892	착유	2	착유02
257036	234865	19292	건유시작	0	건유02
257036	234865	19471	착유	4	착유04
101232	330887	17689	출생신고	0	출생신고
101232	330887	18332	착유	1	착유01
101232	330887	18640	폐사	0	폐사

5.1.4 운용방법

입력데이터는 c:\lactation\input 위치에 젖소이력데이터는 젖소.txt, 육우데이터는 육우.txt로 저장한다. 데이터의 변환은 c:\lactation\source 위치에 있는 convertFile.r, dryPhase.r을 실행한다. 코드 실행 후 c:\lactation\output\historyFile 내에 history.csv로 저장되어 있다.

5.2 월별 데이터의 생성

젖소 이력데이터, 육우 데이터의 변환하여 history.csv를 생성하였다. 이 파일은 시점별로 젖소의 상태 정보를 모두 가지고 있다. 이를 이용하여 월별로 생존해 있는 소를 대상으로 상태를 기록하여 파일로 저장한다.

5.2.1 입력데이터 형식

c:\lactation\output\historyFile 내에 history.csv를 이용한다.

5.2.2 데이터의 변환

2010년 6월부터 2013년 8월까지 월별로 살아 있는 소를 대상으로 소의 상태를 기록한다. 그리고 출생일을 이용하여 월령을 계산하여 age 필드에 입력하고 이력데이터에 기록된 가장최근 기록으로부터 몇 개월 지났는지를 milkAge 필드에 입력한다. 그리고 status 필드에 history.csv의 tmp 필드정보를 사용하여 소의 상태를 입력한다.

farm_id	cow_id	status	age	phase	milkAge
347413	201525	출생신고	22	0	0
257036	234865	출생신고	43	0	0
101232	330887	착유01	22	1	5
327874	837200	출생신고	27	0	0
269489	2747187	출생신고	37	0	0
297111	3215458	출생신고	26	0	0
296127	6212728	출생신고	46	0	0
287178	6238613	착유01	26	1	7

※ 주의: 월별 데이터를 생성하면 건유우의 수와 건유기간이 과대 추정되어 산출된다. 이는 착유기록의 결측치로 발생하는 문제다. 본 연구에서는 월별 데이터를 생성할 때 건유상태에 있는 젖소에 대해서 기간별 보정을 하였다. 건유기간이 3개월 이상 기록된 젖소에 대해서 평균적인 착유-건유 기간을 이용하여 상태를 수정하였다.

예) 2산차 건유시작이 5개월 지난 젖소는 건유시작 2개월 후 건유를 종료하면서 다음 산차인 3산차에 진입하고 그 이후 3개월 지난 소로 수정한다.

5.2.3 변환결과

월별로 생성한 변환결과는 C:\lactation\output\monthlyFile 에 날짜.csv 형식으로 기록한다.

예) 2010-06-01.csv

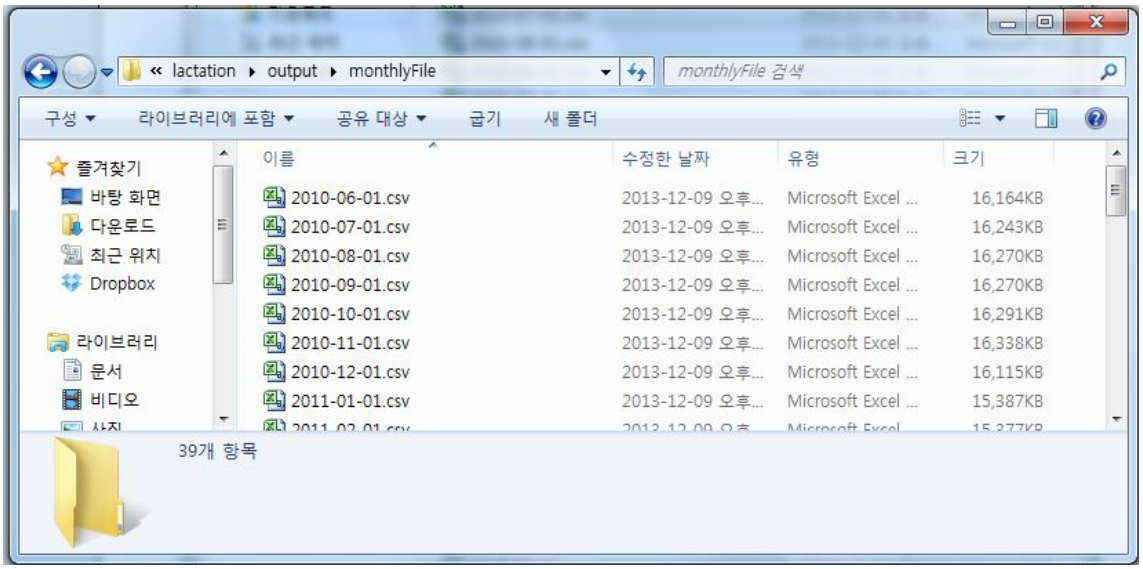
5.2.4 운용방법

입력데이터는 c:\lactation\output\historyFile 위치에 history.csv 다. 데이터의 변환은 c:\lactation\source 위치에 있는 monthlyFile.r을 실행한다. 코드 실행 후 c:\lactation\output\monthlyFile 내에 csv 파일로 저장되어 있다.

5.3 산유량에 대한 회귀모형의 추정 및 회귀모형의 저장

산차별 착유우 두수와 쿼터량 및 기타 공변량을 이용하여 일평균 산유량을 예측하는 회귀모형을 생성한다.

※ 주의: 회귀 모형을 새로 적합해야 하는 경우 프로그램 코드를 반드시 수정해야 한다. 적합된 모형을 이용하기 위해서는 저장 모형결과만 필요하므로 모형적합은 필요없다.



5.3.1 입력데이터 형식

– 산유량–쿼터 데이터

C:\lactation\input에 quota1.csv 파일이다. 산유량–쿼터 데이터는 농가별로 농가번호(farm_id), 주소(address), 2007년 10월 일평균 쿼터량(Q0710), 2007년 10월 일평균 산유량, ..., 2013년 8월 일평균 쿼터량(Q1308), 2013년 8월 일평균 산유량(P1308) 필드로 이루어져있다.

	A	B	C	D	E	F	G	H	I	J	K	L	M	N
1	farm_id	address	Q0710	P0710	Q0711	P0711	Q0712	P0712	Q0801	P0801	Q0802	P0802	Q0803	P0803
2	999999	경기	415.6	258.3	415.6	272.5	415.6	274	415.6	238.4	415.6	244	415.6	22
3	999999	경기	1095.8	989.6	1095.8	1114.8	1095.8	1236.2	1095.8	1335.8	1095.8	1361.2	1095.8	136
4	270743	경기	871.3	746	871.3	717.2	871.3	751.9	871.3	755.8	871.3	873.1	871.3	
5	88439	경기	1786.7	2121.7	1786.7	2062.7	1786.7	1945.5	1786.7	1964.4	1786.7	2003.7	1786.7	203
6	282379	경기	236.4	3.2	236.4	1.7	236.4		236.4		236.4		236.4	
7	999999	경기	711.6	438.9	711.6	389.9	711.6	435	711.6	470.3	711.6	460.7	711.6	38
8	187356	경기	1274.4	1010.8	1274.4	984.6	1274.4	1030.3	1274.4	1077.4	1274.4	1115.5	1274.4	119
9	290557	경기	688.3	601.6	688.3	640.7	688.3	636.9	688.3	714.7	688.3	708.3	688.3	67
10	999999	경기	310.2	291.6	310.2	286.8	310.2	286.7	310.2	280.9	310.2	291.9	310.2	28

– 공변량 데이터

C:\lactation\input에 survX.csv 파일은 날짜(date), 온도(temp), 습도(hum), 구제역유무(disease), 쿼터제한해제 유무(freeQuota), 체세포 2등급 구간 가격조정(optionaMilkPrice), 원유 생산비용(cost), 일평균 쿼터량(quota)의 필드를 가지고 있다. 2010년 6월부터 2013년 8월 까지 데이터가 기록되어 있다.

데이터 예)

date	temp	hum	disease	freeQuota	optional MilkPrice	cost	quota
06월	21.7	72.3	0	0	23.69	641	5876234
07월	24.8	80.1	0	0	23.69	641	5874717
08월	25.6	78.3	0	0	23.69	641	5869199
09월	20.8	74.5	0	0	23.69	641	5868986
10월	14	71	0	0	23.69	641	5844501
11월	6.6	68.6	0	0	23.69	641	5885521
12월	0	66.4	1	0	23.69	641	5886567
01월	−2.9	65.1	1	0	23.69	718	5933247

- 월별데이터

이력데이터의 변환을 통해 생성된 월별 데이터를 이용한다.

C:\lactation\output\monthlyFile 내에 저장되어 있다.

5.3.2 모형적합결과

상수	phase1Cow	phase2Cow	phase3Cow	temp	hum	quota
−3.55×107	4.49×10^{1}	1.19×10^{1}	1.53×10^{1}	2.99×10^{4}	-3.61×10^{4}	6.37×10^{6}

- phase1Cow: 1산차 착유우 수
- phase2Cow: 2산차 착유우 수
- phase3Cow: 3산차 이상 착유우 수
- temp: 기온
- hum: 습도
- quota: 일평균 쿼터량

5.3.3 운용방법

새로운 모형의 적합이 필요한 경우 regression.r 코드를 통해 데이터를 분석하여야 한다.

5.4 생존 모형의 생성

착유우수의 예측을 위해서는 젖소의 생존확률함수를 구해야 한다. 젖소의 생존 데이터와 생존확률에 영향을 미치는 변수를 이용하여 생존확률함수를 구할 수 있다. 그리고 예측된 착유우 수를 보정하는 함수를 생성한다.

※ 주의: 생존모형에 사용되는 공변량을 변경하거나 생존 모형을 새로 적합해야 하는 경우 프로그램 코드를 반드시 수정해야 한다. 적합된 모형을 이용하기 위해서는 저장 모형결과만 필요하므로 모형적합은 필요 없다.

5.4.1 입력데이터 형식

입력데이터는 젖소의 생존시간에 대한 데이터인 history.csv와 생존함수 추정에 사용하는 공변량 데이터인 C:\lactation\input내의 survX.csv 파일이다.

5.4.2 모형적합결과

- 생존모형

젖소의 생존모형은 Cox 비례 위험모형을 사용하였으며 R 패키지인 'survival' 라이브러리의 'coxph' 함수를 이용하였다. 적합된 생존모형 계수는 다음과 같다.

전이사건	temp	hum	disease	freeQuota	optionalMilkPrice	cost	quota
출생>1산	0.0086	−0.0366	0.0000	−0.7662	0.0000	0.0284	−7.5051
1산>1산건유	0.0206	−0.0196	1.0331	−0.1346	0.0000	0.0247	−22.4957
1산건유>2산	0.0026	−0.0408	0.6111	−0.2149	0.0000	0.0000	0.0000
2산>2산건유	0.0132	−0.0186	0.5723	−0.2661	0.0000	0.0168	−14.3030
2산건유>3산	−0.0042	−0.0480	−0.0143	−0.3676	0.0000	0.0000	0.0000
3산>3산건유	0.0192	−0.0164	0.4260	−0.4471	0.0000	0.0252	−22.2237
3산건유>4산	−0.0014	−0.0316	0.3179	−0.3394	0.0000	0.0000	0.0000
4산>4산건유	0.0114	0.0095	0.2810	−0.4369	0.0000	0.0365	−31.9252
4산건유>5산	0.0446	−0.0855	0.6966	−0.1984	0.0000	0.0000	0.0000
5산>5산건유	0.0131	−0.0133	0.0664	−0.5896	0.0000	0.0324	−28.0957
5산건유>6산	0.0246	−0.0698	1.7411	0.2025	0.0000	0.0000	0.0000
6산>6산건유	0.0353	−0.0731	1.0600	−0.3624	0.0000	0.0190	−14.4775

기저생존함수에 대한 정보는 C:\lactation\output\rImage내의 survival.rdata의 survList 객체에 저장되어 있다.

- 착유우 보정 결과

t시점 이후의 기대 착유우 수 = t 시점의 생존확률 × 착유우 수

- phase1Cow 예측치 = 1산차 기대 착유우 수 × 0.379 + 43693
- phase2Cow 예측치 = 2산차 기대 착유우 수 × 0.516 + 27252
- phase3Cow 예측치 = 2산차 기대 착유우 수 × 0.496+ 30242

※ t 시점 이후의 생존확률의 계산

특정시점에서 t시점 이후까지 포함되는 전이 사건에 대해 각 전이 확률의 곱으로 생존확률을 계산한다. 생존확률의 계산은 3장 생존모형을 참조한다.

예) 출생 후 20개월 젖소가 24개월을 생존할 확률

9개월 후 1산차 시작, 그 후 13개월 1산차 건유시작, 그 후 2개월 후 건유종료임. 24개월 생존 확률 = 출생후 20개월 생존한 젖소가 9개월 더 생존할 확률 × 1산차 착유시작우가 13개월 이상 생존할 확률 × 1산차 건유시작우가 2개월 생존할 확률로 정해진다.

5.4.3 운용방법

새로운 모형의 적합이 필요한 경우 survival.r 코드를 통해 데이터를 분석하여야 한다.

5.5 변환된 이력데이터, 저장된 생존모형, 회귀모형을 이용한 예측치 생성

5.5.1 입력데이터의 형식

– 회귀모형

회귀모형은 5.3.2의 회귀모형 계수를 데이터로 불러온다.

– 생존모형

생존함수의 5.4.2의 생존모형함수와 기저생존함수를 데이터로 불러온다.

– 월별데이터

5.3.2의 월별 데이터를 불러온다. 월별데이터는 산차별, 5.2.2에 정의된 milkAge별로 정리해 저장한다.

데이터 형식은 다음과 같다.

데이터 예)

사건	산차	milkAge	마리수	날짜
착유	1	8	2100	2013년8월
착유	1	9	2005	2013년8월
출생	0	19	3121	2013년8월
건유	2	1	1205	2013년8월

위 데이터의 첫행은 2013년 8월 1산차 착유가 8개월 지난 젖소가 2100마리 있음을 알 수 있다.

– 공변량데이터

C:\lactation\input에 survX.csv 파일을 이용한다.

5.5.2 운용방법

① 특정 시점에서 산차별, 월령별 젖소 분포를 월별데이터로부터 불러온다.

② 미래의 공변량 데이터를 입력한다.

③ 미래의 공변량 데이터과 특정시점의 젖소 수를 이용하여 생존모형을 통해 미래시점의

④ 기대 착유우를 예측한다.

⑤ 예측된 산차별 월령별 기대 착유우 수과 미래의 공변량을 이용하여 산유량 예측 회귀모형에 대입한다.

5.6 데이터의 갱신 이후 예측치 생성 방법

데이터를 생신하여 예측치를 생성하고자 할 때 다음 절차를 따른다.

① 축산물품질평가원에 젖소 이력데이터와 육우 데이터를 받아 5.1에 주어진 형식에 맞게 변경하고 파일을 저장한다.

② 낙농진흥회의 생산량 쿼터 데이터를 5.3에 주어진 형식에 맞게 변경하고 파일을 저장한다.

③ 필요한 공변량 데이터를 5.3에 주어진 형식에 맞게 변경한다.

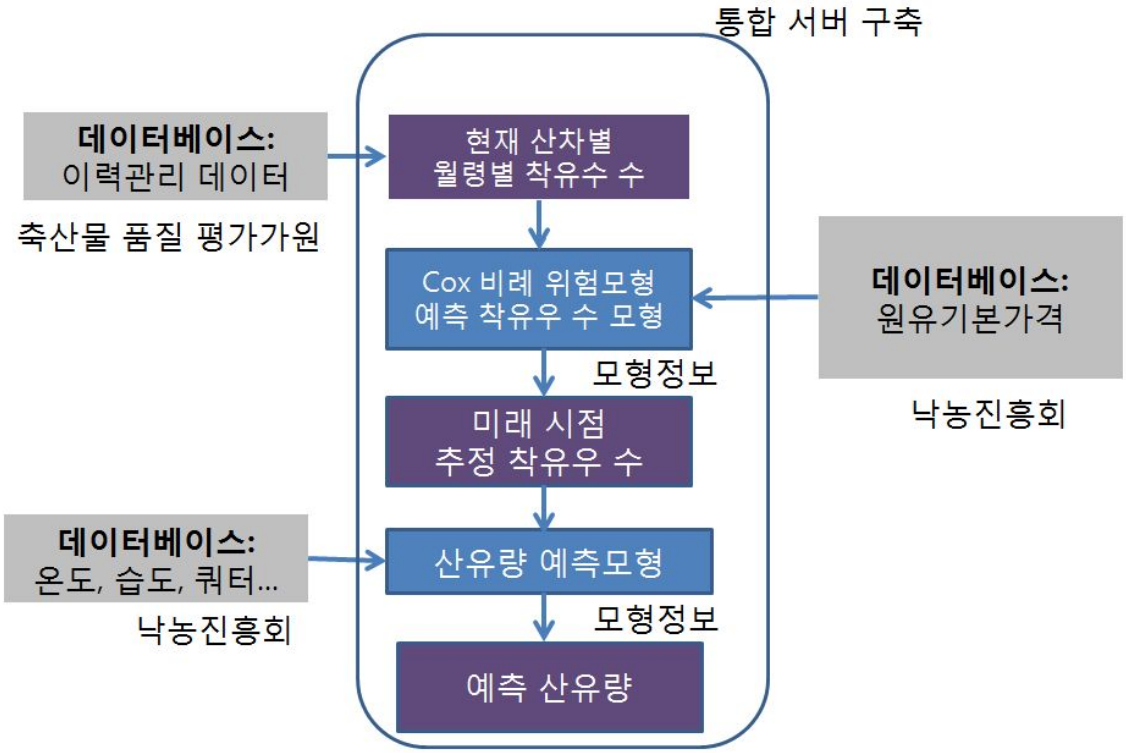
④ 소스 코드 exe.r을 실행한다.

※ 주의: 주어진 모형은 전국 단위 혹은 특정 지역단위에 따른 생존 모형 및 산유량 회귀모형이므로, 특정 농가군에 대한 산유량 예측모형이 필요한 경우 두 모형을 새로 적합해야 한다.

5.7 데이터 베이스 시스템의 구축안

본 연구는 R프로그램에 기반하여 데이터를 정리하고 모형을 생성하였다. 향후 데이터 베이스를 구축하고자 하는 경우 업무 흐름도를 제안한다.

먼저 사용해야 하는 원본 데이터는 축산물품질평가원의 이력관리 데이터, 낙농진흥회의 산유량/쿼터 데이터이다. 이력관리 데이터는 특정 시점에서 소의 월령, 산차, 상태를 파악할 있게 변환되어야 한다. 변환 프로그램은 5.1, 5.2에 설명되어 있다. 그리고 본 연구에서 제안한 생존모형의 적합결과와 미래시점의 공변량 정보를 이용하여 예측 착유우 두수를 생성한다. 다음으로 본 연구에 제안된 산유량 회귀정보와 공변량 데이터를 이용해 예측 산유량을 생성한다. 이 과정이 [그림 15] 에 나와 있다.



[그림 15] 산유량 예측을 위한 업무흐름도

5. 결론 및 제언

본 연구는 젖소 이력데이터를 이용하여 착유우 두수를 추정하는 방법과, 젖소의 생존률을 분석하는 모형을 통해 미래의 산유량을 예측하는 방법을 제안하였다. 산차별 착유우 두수에 기반한 월별 일평균 산유량 모형은 기후 요인에 의한 계절요인을 적절히 반영하고 있는 것으로 보이며, 젖소의 생애주기별 생존모형 역시 계절요인과 쿼터량 효과를 잘 반영하고 있는 것으로 나타났다. 산유량의 예측치는 1월부터 6월까지 증가하고 7월부터 9월까지 감소한 후 다시 증가하는 패턴을 보이고 있다.

본 연구에서 제안한 모형은 2010년 6월 이후부터 2013년 8월까지 젖소/육우 이력데이터, 생산량 쿼터 데이터를 사용하여 적합하였다. 현재 제안된 모형에서 2013년 9월부터 2013년 11월까지 원유 생산량 예측치는 50만 1천 톤으로 실측치가 51만 3천 톤임을 고려할 때 비교적 정확한 예측치를 주었다고 본다. 하지만 모형의 중기 예측시 착유우 두수의 변화가 전문가의 예측치보다 높게 나타나는 것으로 보아 원유 공급이 안정화 된 시기의 데이터가 더 축적 된 후 모형을 다시 적합할 필요가 있다. 본 모형에 의하면 2014년 예측 생산량은 214만 톤이며 월 총생산량이 후반기로 갈수록 증가하는 것으로 예상된다.

향후 이력데이터를 이용한 정확한 예측모형을 구축하기 위해서는 다음 작업이 반드시 이루어져야 한다.

1) 젖소/육우 출생시 모개체번호의 기록: 본문에서 언급한 것처럼 현재 이력데이터의 모개체번호가 많은 부분 결측되어 있어 착유시점을 파악하기가 곤란한 자료가 매우 많다. 모개체번호가 잘 기록된다면 착유시작 시점을 데이터로부터 정확하게 파악할 수 있을 것이다.
2) 농가번호의 매칭작업: 전국단위의 산유량 예측 뿐만 아니라 소규모 지역단위 혹은 특정 농가 포트폴리오에 대해 산유량을 예측하고자 할 경우, 축산물 품질평가원의 이력데이터로 파악할 수 있는 착유우 두수와 낙농진흥회에서 파악할 수 있는 농가별 생산량을 대응시킬 수 있어야 한다. 이를 위해서 농가대응에 필요한 농가코드의 일원화 작업이 이루어져야 한다.
3) 지속적인 데이터 및 모형관리: 데이터가 축적되면서 얻을 수 있는 정보도 많아진다. 본 연구에서 제안한 모형을 효과적으로 활용하기 위해서는 데이터와 모형을 지속적으로 갱신하는 것이 필요하다. 이를 통해 보다 정확한 중단기 예측이 가능할 것이라 생각된다.

부록 1 [생존함수 정보]

기저생존함수
- time: 월
- surv: 기저생존함수 ($Z(t)=0$인 경우 생존함수)
- dLambda: $dA(t)$

※ 주의: 시간에 의존하는 공변량을 가진 생존모형에서 생존확률을 구하기 위해서는 3장 생존함수의 계산을 참고하라.

출생 → 1산 착유

time	surv	dLambda
1	0.8913925	0.10860751
2	0.7647734	0.142046434
3	0.7022594	0.081741822
4	0.6582371	0.062686624
5	0.6334002	0.037732469
6	0.6130273	0.032164308
7	0.5963648	0.027180817
8	0.5884404	0.01328768
9	0.5825576	0.009997275
10	0.5782688	0.007362024
11	0.5756828	0.004472029
12	0.5738857	0.003121617
13	0.5730732	0.001415856
14	0.5714889	0.002764515
15	0.5703239	0.002038659
16	0.5695366	0.001380396
17	0.5687923	0.001306869
18	0.5683899	0.000707364
19	0.5677283	0.001164144
20	0.5672195	0.000896094
21	0.5667355	0.000853376
22	0.5661876	0.000966711
23	0.5651893	0.001763268
24	0.5642945	0.001583082
25	0.5628568	0.002547836
26	0.5612783	0.002804341
27	0.5596551	0.002892013
28	0.5576251	0.003627326
29	0.5550351	0.004644708

1산착유→1산건유

time	surv	dLambda
1	0.967719	0.032281032
2	0.9524341	0.015794717
3	0.9389908	0.014114707
4	0.9264283	0.01337871
5	0.9158342	0.011435467
6	0.9068646	0.009793878
7	0.8971221	0.010743074
8	0.8884001	0.009722168
9	0.8805793	0.008803292
10	0.8744652	0.00694318
11	0.8661264	0.009535914
12	0.8584854	0.00882202
13	0.8487781	0.011307549

1산건유→2산착유

time	surv	dLambda
1	0.9818356	0.01816437
2	0.9593844	0.02286657

2산착유→2산건유

time	surv	dLambda
1	0.9515833	0.04841667
2	0.923805	0.02919173
3	0.9049953	0.02036114
4	0.8885523	0.01816908
5	0.8737365	0.0166741
6	0.8604042	0.01525902
7	0.8455221	0.01729653
8	0.8322753	0.01566708
9	0.8211165	0.0134076
10	0.8104499	0.01299027
11	0.7988911	0.0142622
12	0.7865299	0.01547298
13	0.7734111	0.01667932

2산건유→3산착유

time	surv	dLambda
1	0.9700683	0.02993173
2	0.9334309	0.03776788

3산착유→3산건유

time	surv	dLambda
1	0.92062	0.07938002
2	0.886163	0.03742797
3	0.8599615	0.02956736
4	0.8384318	0.02503572
5	0.8184186	0.02386972
6	0.7980849	0.02484517
7	0.7811093	0.02127043
8	0.7657373	0.01967971
9	0.750089	0.0204356
10	0.7371853	0.0172028
11	0.7249667	0.01657467
12	0.7133951	0.01596165
13	0.699832	0.01901201

3산건유→4산착유

time	surv	dLambda
1	0.9637366	0.03626342
2	0.9166401	0.0488686

4산착유→4산건유

time	surv	dLambda
1	0.8915695	0.1084305
2	0.8444623	0.05283628
3	0.8124617	0.0378946
4	0.7856186	0.0330393
5	0.7612069	0.03107314
6	0.7378746	0.03065179
7	0.7204553	0.0236074
8	0.700231	0.02807148
9	0.6832258	0.02428509
10	0.6721713	0.01617994
11	0.6585391	0.02028085
12	0.6470095	0.01750782

13	0.6355977	0.01763773

4산건유→5산착유

time	surv	dLambda
1	0.9443567	0.05564334
2	0.8926797	0.05472188

5산착유→5산건유

time	surv	dLambda
1	0.8840168	0.11598325
2	0.8302104	0.0608658
3	0.79344	0.04429036
4	0.7535056	0.0503307
5	0.7166164	0.04895685
6	0.6946115	0.0307067
7	0.6797866	0.02134267
8	0.6569514	0.03359177
9	0.6364497	0.03120731
10	0.6131183	0.03665865
11	0.6015657	0.01884239
12	0.5897351	0.01966626
13	0.5759108	0.02344155

5산건유→6산착유

time	surv	dLambda
1	0.9690826	0.03091744
2	0.9491073	0.02061257

6산착유→6산건유

time	surv	dLambda
1	0.9050506	0.0949494
2	0.8672689	0.04174542
3	0.829089	0.04402315
4	0.7920162	0.04471506
5	0.7694357	0.0285102
6	0.7334471	0.04677268
7	0.7053227	0.03834548
8	0.6766678	0.04062668
9	0.6655102	0.01648902
10	0.642167	0.03507566
11	0.6234042	0.02921792
12	0.5965791	0.04302999
13	0.5728847	0.03971721