

의료보험 지식경영시스템 구축 방안에 관한 연구 - 데이터마이닝 기법을 활용한 부과체계 모형 개발 -

용왕식¹⁾, 채영문²⁾, 엄영진³⁾, 김석일⁴⁾, 강성홍⁵⁾, 호승희⁶⁾
국민건강보험공단¹⁾, 연세대학교 보건대학원²⁾⁶⁾, 보건복지부³⁾, 카톨릭의대 예방의학교실⁴⁾
인제대학교 보건관리학과⁵⁾

1. 서 론

1963년 의료보험법이 최초로 제정된 이후 1977년 직장의료보험 실시, 1989년 전국민 의료보험의 실시로 우리 나라의 의료보험제도는 사회보험으로서의 모습을 갖추게 되었다(한동관 등, 1998). 직장의료보험, 공무원 및 사립학교 교직원의료보험, 지역의료보험의 세 범주로 구분되어 독립채산형태로 운영되던 지역방식은 조합간 재정력 격차, 보험료부과의 불형평성, 관리운영의 비효율성 등을 이유로 통합방식으로서의 전환이 지속적으로 논의되어왔다(강복수 등, 1998). 이에 따라 1998년 10월 이후 일차적으로 지역과 공무원 및 사립학교교직원의료보험관리공단이 통합되었으며, 2000년 7월부터는 직장의료보험조합까지 통합되었다. 그러나 통합에 따른 문제들에 대한 우려와 함께, 의료보험의 본질이라고 할 수 있는 사회보장적 취지를 보장하는 형평한 의료보험료부과기준의 마련이 핵심적인 사안으로 대두되고 있다.

현재 우리 나라에서 의료보험료 부과액의 형평성은 실제소득과의 상관관계에서 찾고 있다. 직장가입자의 경우 근로소득에 의해, 지역가입자의 경우 소득, 재산, 자동차, 성, 연령 등의 제 요소를 고려하여 간접적으로 소득능력을 추정하여 보험료를 부과하고 있다(국민의료보험관리공단, 2000). 지역가입자의 부과요소별 배분비의 적정성과 등급표의 적정성 여부의 확인을 위해서는 실제소득과 부과요소간 상관관계를 분석하

여야 하나, 실제소득의 파악이 매우 어렵다. 1997년 12월 현재 지역조합가입자 대상 소득과세자료 보유현황은 약 28%로 매우 낮으며, 과세소득자료의 정확성도 의문시되고 있다.

한편 최근 대두되고 있는 지식경영은 21세기를 이해하기 위한 새로운 패러다임이라고 할 수 있다. 유준수(2000)는 지식경영이란 기업을 포함해서 조직이 보유하고 있는 다양한 정보로부터 지식을 창출하고, 이렇게 창출된 지식자산을 획득, 활용, 도입, 전이하는 등의 모든 과정을 정보통신기술을 시스템화하여, 정보통신기술을 중심으로 활용 가능한 형태의 시스템을 만들며, 이를 토대로 기업 및 조직의 부가가치를 창출하고 경쟁력을 확보하는 경영절차라고 정의하였다. 현재 기업에서의 지식경영은 생존 및 성장을 위한 가장 기본적인 조건이 되었으며, 정부에서도 21세기 국가경영 지표로서 지식을 표명하고, 전국가적인 지식화 사업을 추진하고 있다.

보건의료분야에서 지식경영 도입은 크게 의료보험분야와 병원분야의 두 분야로 나누어 생각할 수 있다. 특히 의료보험에 관련된 정보는 의료전달체계, 보건의료 자원의 활용, 의료이용 및 의료보험체계와 같은 중요한 보건의료정책 결정에 중요한 영향을 미친다. 그러나 이러한 전국적인 규모의 방대한 자료가 단지 의료보험 심사, 급여, 연보발행과 같은 부분에만 제한적으로 사용되고 있는 것이 사실이다(호승희 등, 1999).

의료보험의 통합을 앞두고 보험료 형평부과에 대한 논의가 지속되고 있는 가운데, 보험료 부과요소에 대하여 신정보기술에 의한 정보분석 접근이 필요하다. 따라서 이 연구에서는 보건 의료분야에의 지식경영 도입의 일환으로 데이터 마이닝 기법을 이용하여 지역가입자의 보험료 부과체계와 부과요소별 상관관계를 분석하고, 합리적인 의료보험료부과기준의 마련을 위한 최적 모형을 제시하였다는데 그 의의가 있다.

II. 연구내용 및 방법

1. 연구대상

이 연구는 도시와 농촌 복합형 도시인 고양시에 거주하는 공무원, 교직원가입자를 제외한 직장의료보험가입자 중, 2000년 4월 현재 국민 의료보험관리공단 전산자료의 보험료부과 화일에 등록되어 있는 재산 등의 보험료 부과자료를 보유한 피보험대상자로 하였다. 이들을 연구대상으로 선정한 이유는 의료보험조직의 완전 통합을 앞두고 의료보험료 부과의 형평성과 공평성을 보장할 수 있는 통합된 의료보험료 부과체계를 개발하기 위해서는, 직장가입자에게 지역 보험료 부과체계를 적용함으로써, 직장가입자와 지역가입자에 대한 보험료부과 방안을 도출해 내기 위함이다.

따라서 전국 직장가입자 1,652만명 중 재산 등 지역의료보험료 부과요소를 보유한 고양시 지역 거주자 234,224명 전수를 대상으로 하였으며, 이는 전국 직장가입자의 1.42%에 해당한다.

2. 분석항목의 선정

본 연구는 고양시에 거주하는 직장의료보험 가입자가 1999년 국세청에 신고한 비과세소득을 제외한 근로소득과 직장인이 보유하고 있는 재산, 자동차 등의 자료를 이용하여, 지역의료보험료 부과요소간의 상관관계를 파악하여 보험료 부과규칙을 도출한 후 통합보험료 부과모형을

제시하고자 하였다.

따라서 종속변수는 지역의료보험료 부과요소에서 추정근로소득 결정요인을 추론할 수 있는 근로소득, 즉 가산정보수월액으로 설정하였다. 독립변수로는 근로소득과 직접적 관련이 있다고 할 수 있는 재산과표액, 재산보험료, 자동차보험료, 성·연령, 재산, 자동차에 의해 산정된 평가소득 점수, 나이, 직장보험료, 직장보험료 보수월액, 평가점수 합계, 평가등급, 평가소득 보험료, 지역보험료, 보험료의 차이, 직장보험료를, 지역보험료를, 보험료를 차이의 18개 변수로 설정하였다

3. 데이터마이닝 기법의 적용

데이터 수집 단계에서는 호스트에 저장된 데이터를 텍스트 형태로 다운 받은 후 이들 자료를 이용하여 데이터베이스를 구축한 후 분석 목적에 적합하도록 각 데이터를 가공한 후 통계 작업과 데이터마이닝 작업을 위하여 SAS의 data set 형태로 데이터를 변환하였다.

이 연구의 분석도구로서 데이터베이스 구축 과정에서 MS Access를 사용하였으며, 통계 및 데이터마이닝 작업을 위해서는 SAS(Statistical Analysis System)와 데이터마이닝 패키지인 Enterprise Miner 3.0을 사용하였으며, 분석대상에 따라 그 특성에 적합한 분산분석, 상관분석, 의사결정기법, Paired T-검정을 이용하여 자료를 분석하였다.(Table1)

Table 1. 분석대상별 분석방법

분 석 대 상	분석방법
○ 소득과 보험료부과 요소간의 상관관계	상관분석
- 평가소득점수(성연령,재산자동차),재산자동차	
- 평가소득보험료	
- 직장·지역보험료, 보험료를	분산분석
○ 평가등급과 보험료, 보험료를 차이	
○ 보험료와 평가점수 합과의 차이, 보험료와 평가소득 보험료와의 차이	상관분석
○ 직장과 지역의료보험료 부과 비교분석	상관분석
○ 소득결정요인 분석	의사결정분석
○ 추정근로소득의 타당성검증	Paired T-검정

III. 연구결과

1. 의료보험료 부과시스템 평가 및 지식경영시스템 전략

의료보험의 보험료는 사회보험의 원칙에 따라 기여금으로 납부되는 단기보험적 성격이다. 국민건강보험법의 개정에 따라 지역가입자에 대해 부과표준소득으로 보험료를 부과 산정하는 부과체계 개발이 시급하나 소득과악율이 28%에 이르지 못하고 있고, 지역보험료 부과요소의 배분율은 실질소득과 거리가 있어 이에대한 논란이 그치지 않고 있다. 현 보험료 부과시스템은 직장과 지역간의 보험료 부과의 형평성을 기하고 보험료 부과요소와 소득간의 상관관계를 도출, 분석하여 의료보험료 부과체계의 새로운 지식을 추출.적용한 공평한 보험료 부과체계 개발이 시급한 실정이다. 현 의료보험료 부과체계의 문제점으로는 지역가입자의 보험료 부과체계상 부담능력을 반영하는 부과요소의 적정성에 대한 문제, 직장가입자와 지역가입자간 보험료 부담의 형평성문제, 재산과 자동차로 직장과 지역가입자군간의 보험료 부담비율에 대한 결정문제, 직장근로자의 근로소득과 지역가입자의 부과요소간 상관관계문제등을 들 수 있다. 이는 직장.지역간또는 지역의료보험료 부과요소의 객관성과 상관(유의)성에 대한 지식을 추출하여 패턴을 추론하고, 관계규칙을 설정하여 지식경영시스템 전략차원에서 합리적인 보험료 부과모델을 제시하였다.

2. 데이터마이닝 기법을 활용한 의료보험료 부과소득 추정 방안.

가. 보수일액에 대한 결정요인

소득결정요인은 현재의 지역의료보험료 부과요소로 근로소득을 어느 정도 설명하는 지를 알아보기 위해 의사결정기법과 회귀분석기법

중 보다 더 안정적으로 소득결정요인을 확인할 수 있는 방법으로 선택하고자 독립변수는 상관분석결과를 고려하여 가장 의미 있는 변수라고 판단된 재산보험료, 자동차보험료, 평가소득점수(성, 연령), 평가소득점수(재산), 평가소득점수(자동차)로 하였다. 5개 항목으로 제한한 이유는 기타변수는 독립변수로 선정된 변수와 다중공선성이 있기 때문에 적절하지 않았기 때문이다.

소득결정요인을 분석한 결과 평균제곱오차(ASE)가 의사결정나무분석(decision tree)은 1,135,512이고, 회귀분석(regression)은 1,176,560으로 나타나 의사결정나무분석이 보다 안정적인 기법임이 확인되었다. 즉, 의사결정나무분석이 소득의 결정요인을 보다 잘 알 수 있게 하였다.

따라서, 지역보험료 부과체계에 근거하여 직장인의 소득을 파악하는 모델은 의사결정기법에 근거하였다.

나. 근로소득에 대한 의사결정분석

근로소득에 대한 상관분석결과 유의한 상관관계를 보이면서 상관관계가 높은 변수는 재산보험료, 자동차보험료, 평가소득점수(성, 연령), 평가소득점수(재산), 평가소득점수(자동차)로 나타나서 이들 5개 변수를 이용하여 의사결정분석을 하였다.

본 연구에서 포함된 재산과표액, 자동차세액 등은 의사결정분석에 사용된 변수와 중복된 변수이므로 분석대상 변수에서 제외하였다. 근로소득에 영향을 미치는 독립변수가 모두 연속변수이므로 의사결정 분석에 있어 이를 적절히 그룹화할 필요가 없었다. 의사결정 분석을 실시한 결과 그룹의 수를 3으로 하는 것이 가장적절하다고 판단하였다. 이에 따라 의사결정 나무분석을 한 결과는 Table 2, Fig 1 과 같다.

Table 2. 의사결정분석을 위한 적정 그룹수

그룹수	평균제곱오차 (AVE SQ ERR)
2	1.26
3	1.24
4	1.24
5	1.24
10	1.23

자동차보험료가 2,050보다 작고, 재산보험료가 6,650보다 작으면서 평가소득점수(성, 연령)가 6보다 작은 집단이라는 것을 알 수 있다. 즉, 근로소득이 낮은 계층은 자동차나 재산이 거의 없으면서 성, 연령의 평가점수는 낮다.

2번 계층의 월 평균추정근로소득은 1,580,413원인데, 2번 계층을 결정하는 규칙으로는 자동차보험료, 재산보험료는 1번과 같으면서, 평가소득점수(성, 연령)는 6이상 14.5 이하임을 알 수 있다.

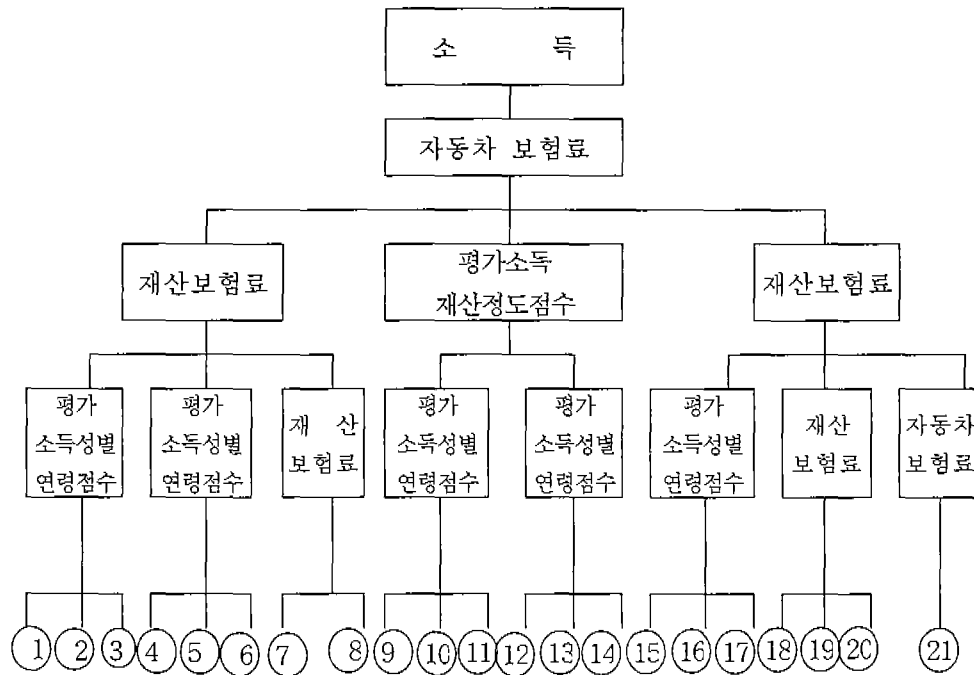


Fig1. 월 추정근로소득에 관한 의사결정도(소득에 대한 변수의 추이)

그룹의 수를 3으로 하여 의사결정분석을 한 결과 소득계층은 23단계로 분류되었으나 신뢰도를 높이기 위하여 극단치인 22, 23 구간은 제외하였다. 최소 소득계층인 1번 집단의 평균소득은 1,254,843원이었으며, 최대 소득계층인 19번은 3,262,211원으로 현재의 변수를 가지고 소득계층에 대한 분류를 할 수 있음을 알 수 있었다.

각 소득계층을 결정하는 규칙은 다음과 같다.

1). 1~3번 계층[①1,254,843 ②1,580,413 ③1,734,998]

1번 계층의 월 평균추정근로소득은 1,254,843원으로 나타났다. 1번 계층을 결정하는 규칙은

3번 계층의 월 평균추정근로소득은 1,734,998원으로, 3번 계층을 결정하는 규칙은 자동차보험료, 재산보험료는 1번과 같으면서, 평가소득점수(성, 연령)가 14.5이상이다.

2). 4~6번 계층[④1,534,427 ⑤1,879,011 ⑥1,546,208]

4~6번 계층의 월 평균추정근로소득은 1,534,427원부터 1,879,011원 사이이다. 재산보험료는 6,650원 이상 16,150원 이하이며, 이들은 평가소득점수에 따라 월 추정근로소득이 달라진다는 것을 알 수 있었다.

3). 7~8번 계층[⑦2,142,230 ⑧2,471,462]

7~8번 계층의 월 추정근로소득은 2,142,230원과 2,471,462원으로 나타났다. 이 계층을 결정하는 규칙은 자동차보험료가 2,050원보다 작으면서 재산보험료가 16,150원보다 큰 집단이라는 것을 알 수 있다. 보다 구체적으로는 재산보험료가 16,150원 이상이고 21,850원 보다 작을 때 월 추정근로소득은 2,142,230원, 재산보험료가 21,850원 이상일 때에는 2,471,462원이라는 것을 알 수 있었다.

4). 9~14번 계층 [㉑1,544,087 ㉒1,872,552 ㉓

2,143,476 ㉔ 2,215,594 ㉕2,616,812 ㉖ 2,312,653]

9~14번 계층을 결정하는 주요규칙은 자동차보험료, 평가소득점수(재산), 평가소득점수(성, 연령)으로 나타났다. 13번 계층의 평가소득성별연령점수 13.1~16.05구간이 16.05이상인 구간 보다 추정소득이 높은 이유는 가족구성원의 수가 적으면서 경제활동능력이 있는 세대원으로 구성된 가족의 소득이 높기 때문이다.

5). 15~17번 계층 [㉗2,220,574 ㉘2,745,497

㉙2,332,351]

Table 3. 월 추정근로소득의 결정규칙

소득 계층	월 추정근로소득 결정규칙(변수)					소득 계층	월 추정근로소득 결정규칙(변수)				
	월평균근로소득 (원) 표본수(세대)	평가소득 재산정도 점수(점)	평가소득 성별연령 점수(점)	재산 보험료 (원)	자동차 보험료 (원)		월평균근로소득 (원) 표본수(세대)	평가소득 재산정도 점수(점)	평가소득 성별연령 점수(점)	재산 보험료 (원)	자동차 보험료 (원)
1	1,254,843 (8,316)	-	<6	-	-	12	2,215,594 (1,338)	-	<3.1	-	-
2	1,580,413 (6,362)	-	6≤평점<14.5-	-	-	13	2,616,812 (5,307)	-	13.1≤평점<16.05	-	-
3	1,734,998 (2,823)	-	≥14.5	-	-	14	2,312,653 (1,902)	-	≥16.05	-	-
4	1,534,427 (1,655)	<11.6	-	-	-	15	2,220,574 (4,493)	-	<14.5	-	-
5	1,879,011 (2,422)	11.6≤평점<14.65-	-	-	-	16	2,745,497 (4,968)	-	14.5≤평점<16.05	-	-
6	1,546,208 (1,089)	≥14.65	-	-	-	17	2,332,351 (1,331)	-	≥16.05	-	-
7	2,142,230 (1,453)	-	-	<21,850	-	18	2,881,564 (3,601)	-	-	<23,750	-
8	2,471,462 (2,320)	-	-	>21,850	-	19	3,262,211 (4,160)	-	-	23,750≤재산<40,850	-
9	1,544,087 (1,654)	-	<6.15	-	-	20	2,789,087 (389)	-	-	≥40,850	-
10	1,872,552 (6,225)	-	6.15≤평점<13.1	-	-	21	2,477,656 (1,018)	-	-	-	<26,350
11	2,143,476 (11,964)	-	≥13.1	-	-						

15~17번 계층을 결정하는 주요규칙은 자동차보험료, 재산보험료, 평가소득점수(성, 연령)으로 나타났다. 16번 계층의 평가소득성별연령점수 14.5~16.05구간이 16.05이상인 구간보다 월 추정근로소득이 높은 이유는 13번 계층과 유사하다.

6). 18~20번 계층[㉔2,881,564 ㉕3,262,211 ㉖2,789,087]

18~20번 계층을 결정하는 주요규칙은 자동차보험료, 재산보험료로 나타났다. 19번 계층의 소득이 3,262,211원인 경우에 재산보험료 23,750~40,850원 구간이 20번 계층의 재산보험료가 40,850원 이상인 구간보다 월 추정근로소득이 높은 이유는 재산보험료를 반드시 많이 낸다고 추정근로소득이 높을 수 없으며 반드시 정비례 관계가 성립되지 않고 재산보험료 계층별 차이가 있음을 의미한다.

7). 21번 계층[㉗ 2,477,656]

21번 계층을 결정하는 주요규칙은 자동차보험료, 재산보험료로 나타났으며 재산보험료 보다는 자동차보험료가 소득결정요인에 더 큰 영향을 미쳤다.

3. 데이터마이닝 기법을 활용한 통합의료보험 부과체계 모형

종래의 지역의료보험에서는 수작업 중심의 개념으로 소득추정(파악)이나 보험료 부과체계를 개발하는데 있어 해당되는 요인들을 각기 분리하여 독립적으로 적용하였으나 본 연구에서는 각 계층의 특성에 따라 여러 가지 요인을 동시에 고려하여 소득을 파악할 수 있는 방안을 의사결정기법(decision tree)을 이용하여 제시하였다. 이와 같이 소득을 추정한 후 소득에 대해 적정 보험요율을 곱하여 보험료를 부과할 수 있는 통합보험료 부과체계모형을 개발하였다(Table 4)

예를 들어 1구간의 경우 Table 3의 소득결정

규칙에서 나타내 보이듯이 지역보험료 부과요소중 자동차 보험료가 2,050원 보다 작고, 재산보험료가 6,650원 보다 작으며, 평가소득 성별연령점수가 6 보다 작은 8,316세대에 대하여는 월평균 추정근로소득은 1,254,843원으로 나타낼 수 있다.

동세대에 대하여 보험요율을 최저 1.7% 적용시 21,340원을, 2.8% 적용(직장보험료율, 2000. 7. 1부 적용)시 35,140원을, 3.4% 적용(공무원보험료율, 2000. 7. 1부 적용)시 42,660원을 부담하는 것으로 나타났다.

이중 직장이나 공무원·교직원의 경우 50%는 본인인, 나머지 50%는 사용자가 부담하고 있으나 지역가입자의 경우는 전액을 세대별로 부담(국고 부담 미포함시)하여야 한다.

4. 데이터마이닝 기법에서 도출된 규칙을 적용한 보험료와 현적용 보험료의 비교, 분석

데이터마이닝기법에서 도출된 규칙을 적용한 보험료와 현적용 지역보험료에 대한 비교, 분석을 실시하였다. 데이터마이닝기법에서 도출된 규칙을 적용한 보험료는 의사결정나무분석에서 나타난 월평균 추정근로소득에 표본집단 지역평균보험료율(1.7%)을 적용하였으며 현적용 지역보험료는 근로소득결정규칙(Table 5)에 의하여 산출된 값을 적용하였다. 이 두 값에 대한 paired T-test 결과, 95%신뢰구간에서 양측검정 유의확률이 0.72로 통계적으로 유의하지 않았다.

종합해 볼 때 Table 6과 같이 마이닝결과 구간별 평균보험료는 36,879원이었고 현적용 지역보험료 구간별 평균보험료는 36,739원으로 통계적으로 유의한 차이는 나타나지 않았지만 소득구간별 보험료 중 지금내는 보험료보다 덜 부담해야 하는 구간이 21개 구간 중 62%에 해당하는 13개 구간이었으며 반면에 더 부담해야 하는 구간은 38%에 해당하는 8개 구간이었다. 변동폭이 큰 구간을 보면 (Table 5)에서 나타난 바와 같이 하락 1,2구간,

Table 4. 데이터마이닝 결과 통합보험료 부과체계 모형 (단위: 원, %, 세대)

구간 구분	월평균추정 근로소득	보험료율 (17-34)	월보험료			세대수
			계	본인부담	사용자부담	
1	1,254,843	1.7	21,340	10,670	10,670	8,316
		2.8	35,140	17,570	17,570	
		3.4	42,660	21,330	21,330	
2	1,580,413	1.7	26,860	13,430	13,430	6,362
		2.8	44,260	22,130	22,130	
		3.4	53,740	26,870	26,870	
3	1,734,998	1.7	29,500	14,750	14,750	2,823
		2.8	48,580	24,290	24,290	
		3.4	58,980	29,490	29,490	
4	1,534,427	1.7	26,080	13,040	13,040	1,655
		2.8	42,960	21,480	21,480	
		3.4	52,180	26,090	26,090	
5	1,879,011	1.7	31,940	15,970	15,970	2,422
		2.8	52,620	26,310	26,310	
		3.4	63,880	31,940	31,940	
6	1,546,208	1.7	26,280	13,140	13,140	1,089
		2.8	43,300	21,650	21,650	
		3.4	52,580	26,290	26,290	
7	2,142,230	1.7	36,420	18,210	18,210	1,453
		2.8	59,980	29,990	29,990	
		3.4	72,840	36,420	36,420	
8	2,471,462	1.7	42,020	21,010	21,010	2,320
		2.8	69,200	34,600	34,600	
		3.4	84,020	42,010	42,010	
9	1,544,087	1.7	26,240	13,120	13,120	1,654
		2.8	43,240	21,620	21,620	
		3.4	52,500	26,250	26,250	
10	1,872,552	1.7	31,840	15,920	15,920	6,225
		2.8	52,440	26,220	26,220	
		3.4	63,660	31,830	31,830	
11	2,143,476	1.7	36,440	18,220	18,220	11,964
		2.8	60,020	30,010	30,010	
		3.4	72,880	36,440	36,440	
12	2,215,594	1.7	37,660	18,830	18,830	1,338
		2.8	62,040	31,020	31,020	
		3.4	75,340	37,670	37,670	
13	2,616,812	1.7	44,480	22,240	22,240	5,307
		2.8	73,280	36,640	36,640	
		3.4	88,980	44,490	44,490	
14	2,312,653	1.7	39,320	19,660	19,660	1,902
		2.8	64,760	32,380	32,380	
		3.4	78,640	39,320	39,320	
15	2,220,574	1.7	37,740	18,870	18,870	4,493
		2.8	62,220	31,110	31,110	
		3.4	75,540	37,770	37,770	
16	2,745,497	1.7	46,680	23,340	23,340	4,968
		2.8	76,880	38,440	38,440	
		3.4	93,340	46,670	46,670	
17	2,332,351	1.7	39,640	19,800	19,800	1,331
		2.8	65,300	32,650	32,650	
		3.4	79,300	39,650	39,650	
18	2,881,564	1.7	48,980	24,490	24,490	3,601
		2.8	80,680	40,340	40,340	
		3.4	97,980	48,990	48,990	
19	3,262,211	1.7	55,460	27,730	27,730	4,160
		2.8	91,340	45,670	45,670	
		3.4	110,920	55,460	55,460	
20	2,789,087	1.7	47,420	23,710	23,710	389
		2.8	78,100	39,050	39,050	
		3.4	94,820	47,410	47,410	
21	2,477,656	1.7	42,120	21,060	21,060	1,018
		2.8	69,380	34,690	34,690	
		3.4	84,240	42,120	42,120	

상승 20,21구간에서 변동율이 5~20%로 나타났다. 이에 대한 구간별 타당성을 검증한 결과 1,2구간에서 현적용 지역보험료가 마이닝결과 보험료보다 작게 나타나는데 영향을 미친 변수는 자동차보험료, 재산보험료, 평가소득 성별, 연령점수로 나타났다. 그 요인은 연령층이 낮고 가구원수(1~2명)가 적은 근로자일수록 연령층이 높은 근로자보다 추정근로소득에 비해 재산보유정도가 평균재산보험료보다 훨씬 낮게 나타났기 때문이다. 이는 지역보험료 부과요소인 재산보험료 등급구간에서 불형평성이 있음을 알 수 있었다. 한편 현적용보험료가 마이닝결과 보험료보다 크게 나타난 20,21구간의 변수는 평가소득 성별, 연령별 점수, 재산보험료, 자동차보험료로서 유의한 차이는 없었지만 주영향 변수 자동차보험료 등급구간별 보험료에서 일부 불형평성이 있었다.

이는 Table 6에서도 입증해 주듯이 현적용보험료 증가지수를 보면 변동구간에 해당하는 증가지수는 근로소득이 증가함에 따라 함께 증가하지 않는 것으로 나타났다. 그러나 월평균 추정근로소득과 마이닝결과 보험료 평균증가지수는 1.73으로 동일하게 나타났으며 현적용 지역보험료 평균증가지수는 2.15로 4개구간을 제외하고는 소득증가에 따라 비례적으로 증가하고 있는 것으로 나타났다. 또한 마이닝결과 보험료와 현적용 지역보험료의 구간별 차이정도를 Fig2에서 비교해 본 결과 1, 2, 20, 21구간을 제외하고는 유사하게 나타났다.

위 상황을 종합해 볼 때 월평균 추정근로소득에 의해 산출된 마이닝결과 보험료의 타당성은 입증되었으나 현실적으로 소득과약을 저조, 표본집단의 제한성 등 한계상황을 극복하기에는 너무나 제약점이 많았다.

Table 5. 데이터마이닝기법에서 도출된 규칙은 적용한 보험료와 현직용 지역보험료에 대한 대응표본 검정

구간 구분	월평균추정 근로소득	결과 비교			t값	유의 확률
		마이닝결과 (A)	현직용 지역보험료(B)	금액차이 (A-B)		
1	1,254,843	21,340	17,070	4,270	0.359	0.723
2	1,580,413	26,860	23,310	3,550		
3	1,734,998	29,500	27,550	1,950		
4	1,534,427	26,080	26,210	-130		
5	1,879,011	31,940	32,100	-160		
6	1,546,208	26,280	27,020	-740		
7	2,142,230	36,420	36,750	-330		
8	2,471,462	42,020	43,050	-1,030		
9	1,544,087	26,240	24,460	1,780		
10	1,972,552	31,840	30,250	1,590		
11	2,143,476	36,440	35,740	700		
12	2,215,594	37,660	38,440	-780		
13	2,616,812	44,480	45,000	-520		
14	2,312,653	39,320	40,690	-1,370		
15	2,220,574	37,740	37,190	550		
16	2,745,497	46,680	45,780	900		
17	2,332,351	39,640	39,960	-320		
18	2,881,564	48,980	49,520	-540		
19	3,262,211	55,460	56,310	-850		
20	2,789,087	47,420	49,630	-2,210		
21	2,477,656	42,120	45,600	-3,480		

Table6. 근로소득구간별 데이터마이닝기법에서 도출된 규칙을 적용한 보험료와 현직용 보험료의 증가지수 변화 추이

구간 구분	월평균추정 근로소득	증가지수 (1999년)	결과 비교			증가 지수
			마이닝결과 보험료(A)	증가 지수 (B)	현직용지역 보험료(B)	
1	1,254,843		21,340		17,070	
4	1,534,427	1.22	26,080	1.22	26,210	1.54
9	1,544,087	1.23	26,240	1.23	24,460	1.43
6	1,546,208	1.23	26,280	1.23	27,020	1.58
2	1,580,413	1.26	26,860	1.26	23,310	1.37
3	1,734,998	1.38	29,500	1.38	27,550	1.61
10	1,872,552	1.57	31,840	1.57	30,250	1.77
5	1,879,011	1.50	31,940	1.50	32,100	1.88
7	2,142,230	1.71	36,420	1.71	36,750	2.15
11	2,143,476	1.71	36,440	1.71	35,740	2.09
12	2,215,594	1.77	37,660	1.77	38,440	2.25
15	2,220,574	1.77	37,740	1.77	37,190	2.18
14	2,312,653	1.84	39,320	1.84	40,690	2.38
17	2,332,351	1.86	39,640	1.86	39,960	2.34
8	2,471,462	1.97	42,020	1.97	43,050	2.52
21	2,477,656	1.97	42,120	1.97	45,600	2.67
13	2,616,812	2.09	44,480	2.09	45,000	2.64
16	2,745,497	2.19	46,680	2.19	45,780	2.68
20	2,789,087	2.22	47,420	2.22	49,630	2.91
18	2,881,564	2.30	48,980	2.30	49,520	2.90
19	3,262,211	2.60	55,460	2.60	56,310	3.30
평균	2,170,878	1.73	36,879	1.73	36,739	2.15

IV. 고찰

본 연구에서는 의료보험분야의 지식경영의 일환으로 데이터마이닝 기법을 활용한 지역의료 보험료 소득결정요인을 분석하여 통합의료보험 부과체계 모형제시로 의사결정지원시스템 개발 방안을 모색하였다. 따라서 현재의 지역보험료 부과체계를 가지고서 직장인에게 보험료를 부과할 수 있는 방안을 마련했다는 데 큰 의의가 있으며 향후 통합의료보험료 부과시스템에 마이닝 기법이 도입된다면 추정소득을 결정할 수 있는 다양한 신규 부과요소를 추가로 개발·도입하여 데이터의 활용가치를 향상시킴으로써 보험료 부과정보 인프라구조의 구축과 함께 형평성 있는 보험료부과에 기여할 것으로 기대된다.

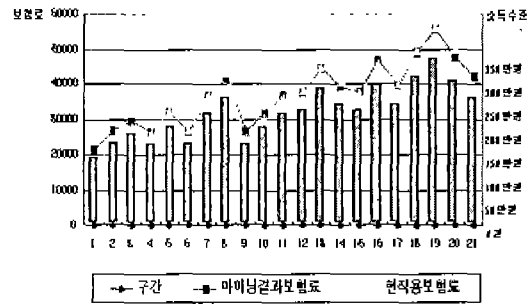


Fig 2. 데이터마이닝기법에서 도출된 규칙은 적용한 보험료와 현직용 지역보험료 변동 추이

이 연구는 지역의료보험료 소득결정요인을 분석하기 위하여 도시와 농촌의 복합도시인 고양시에 거주하는 직장의료보험가입자 중 1999년도 현재 국민의료보험관리공단 전산자료의 보험료 부과화일에 등록되어 있는 재산 등의 지역보

혐료 부과자료를 보유한 234,224명(직장가입피보험자 74,816명 포함)에 대한 보험료 부과요소와 근로소득(가산정보수월액)간의 상관관계를 파악하였다. 따라서 표본수가 총직장가입자 1,652만명 중 1.42%로 전 피보험자를 대표하는 데는 한계가 있었다. 연구결과 현 의료보험료 부과방식은 소득과약율이 저조하고 지역의료보험료 부과요소의 배분율은 실질소득과 거리가 있으며 지역·직장간 의료보험료 부담의 형평성문제는 집단간 갈등의 소지가 내재되어 있다. 데이터웨어하우스 및 데이터마이닝기법을 활용한 보험료 부과가능요소의 점진적 확대와 그 구성요소간의 상관분석을 통하여 패턴을 기초로 한 예측분석으로 지식경영시스템체계의 대전환이 긴요한 실정이다. 직장피보험자가 가지고 있는 재산, 자동차 등을 지역의료보험료 기준으로 환산하였을 때 세대당 월평균 36,006원을 부담하는 것으로 나타났다. 순수한 개인부담 보험료만 비교시 약 10,000원 정도를 지역가입자가 더 부담하는 것으로 나타났다. 추가부담부분에서는 국고보조금지원이나 보험료 부과요소의 조정등을 통해서 형평성을 유지시켜야 한다. 의사결정분석결과 산출된 월평균 추정근로소득은 2,178,878원으로 마이닝결과와 현적용 지역보험료 월평균 보험료액은 각각 36,879원과 36,739원으로 나타났다. 소득구간별 보험료 중 지금 내는 보험료보다 덜 부담해야 하는 구간이 21구간 중 62%에 해당하는 13개 구간이었으며 보험료를 최고로 적게 부담하는 구간은 21구간으로 최고 8%정도 덜 부담하는 것으로 나타났다. 반면에 더 부담해야 하는 구간은 21개 구간 중 38%에 해당하는 8개 구간이었으며 보험료를 최고로 많이 부담하는 구간은 1구간으로 20%정도 더 부담하는 것으로 나타났다. 보험료 변동폭이 큰 4개 구간을 분석한 결과 저소득층인 1,2구간은 재산보험료 등급구간에서 20,21구간은 자동차보험료 등급구간에서 일부 불형평성이 나타났으나 총 21구간중 동 4개구간을 제외한 나머지 17개 구간에서는 월평균 추

정근로소득과 데이터마이닝기법에 의해 도출된 규칙을 적용한 보험료 및 현적용 지역보험료 증가지수가 비례적으로 증가하고 있어 타당성이 입증되었으며 연구의 제한상황을 극복해 나간다면 실무적용에도 큰 무리가 없을 것으로 판단된다. 연구의 제한점을 극복하는 차원에서 향후의 연구방향은 소득과약율을 높이는 방법의 지속적인 연구와 현재 파악된 소득범위 내에서 데이터마이닝 등 신정보 기술을 이용한 추정소득모델을 지속적으로 개발하여 형평성있는 보험료 부과시스템을 구현함이 가장 시급한 과제이다. 국민건강보험법이 1998년 12월 31일 국회를 통과하면서 소득단일기준에 의한 보험료 부과조문이 시행도 해보지 못한 상태에서 불과 1년만인 1999년 12월 31일 개정된 바 있다. 그만큼 소득에 의한 단일부과체계 모형은 보험료부과체계의 가장 이상형이지만 완전한 소득을 파악하여 적용함은 현실적으로 불가능하다. 그렇지만 소득수준에 영향을 미치는 간접적 소득과약방법요소로서 재산요소인 금융재산, 골프회원권등 무체재산권을 들 수 있고, 인적자원 요소로 경험, 직업, 경제활동 참가율을, 소비지출요소로 각종 공과금, 사업장크기, 재료비들, 지역요소로 거주지 등을 들 수 있다.

이러한 요소들을 실제 보험료부과에 활용하기 위해서는 관찰하기 쉽고, 수집하는데 비용이 적게 들고, 객관적 계량화가 용이하고, 객관적 타당성이 있어 모든 사람들의 수용가능성이 커야 한다. 동시에 부과요소간의 상관관계, 패턴, 추세 등을 지속적으로 연구·발전시켜 보편타당한 보험료 부과모델을 제시하여야 한다.

보험료 결정주기를 현행 월단위에서 연간 1회로 하여 변동조정에 따른 불필요한 행정력 낭비를 방지하는 방안도 전향적으로 고려해야 할 것이다. 이러한 과제를 해결하기 위해서는 접근도가 쉽고 효율적인 지식정보를 탐사·적용할 수 있는 신정보지식 기술인 데이터마이닝 기법활용을 실용화시킬 수 있다는 전략차원에서 단계적인 발전을 기대하면서 지식경영시스템의 구축·활용에 진력하여야 할 것이다.

참 고 문 헌

1. 강호규. 정보사회에서의 지식경쟁력강화방안. 한국전산원. 1998
2. 고재성. 데이터웨어하우징과 데이터마이닝 기법을 이용한 의사결정사례에 관한 연구. 성균관대학교 경영대학원 석사학위논문. 1997
3. 광용구. 데이터웨어하우스와 구축방법론(1). 경영과컴퓨터. 1996. p306~312
4. 광용구. 데이터웨어하우스와 구축방법론(2). 경영과컴퓨터. 1996. p308~312
5. 광용구. 데이터웨어하우스와 구축방법론(3). 경영과컴퓨터. 1996. p324~327
6. 광용구. 데이터웨어하우스와 구축방법론(4). 경영과컴퓨터. 1997. p296~299
7. 광용구. 데이터웨어하우스와 구축방법론(5). 경영과컴퓨터. 1997. p322~325
8. 국민의료보험관리공단. 외국의 보험료부담 분석 배부자료. 1999
9. 국민의료보험관리공단. 의료보험통합의 의의와 효과. 2000
10. 국민의료보험관리공단. 의료보험통합자료 모음집. 1999
11. 국민의료보험관리공단. 지역보험료 부과관리 업무처리요령. 2000
12. 김성훈. 공공기관지식관리의 전략과 성공요인에 관한 연구. 한국전산원. 1999
13. 서울경제신문. 선진국 보험요율 우리의 2~5배. 1998. 9. 19
14. 송석주. 한국사회보장정책의 발전적 전환에 관한 연구. 한남대학교 지역개발대학원 석사학위논문. 1997
15. 신중섭외 10인. 세계의 사회보장. 유몽출판사. 1996
16. 이두호, 차홍봉, 엄영진. 국민의료보장론. 나남. 1992
17. 이희석, 장재경. 데이터마이닝과 DLAP. 컴퓨터월드. 1997. p132~141
18. 장동인. 실무자를 위한 데이터웨어하우스. 대청. 1998
19. 정철홍. 데이터웨어하우스 시장현황 및 향후전망. 정보과학회지 1997; 15(5)
20. 조재희, 박성진. 데이터웨어하우징과 OLAP. 대청. 1996
21. 포스코경영연구소. 지식경영. 더난출판사. 1998
22. 한국능률협회 컨설팅/ERP/P1센터. 창조적 지식경영 구축을 위한 지식경영 및 지식관리시스템. KMAC와 Knowledgian. com이 공동개최한 뉴밀레니엄 특별강좌. 2000. 3. 7
23. 한국오라클. 데이터웨어하우스, 데이터마트 그리고 데이터베이스마케팅 - 오라클 세미나. 한국오라클. 1997. 4 ~ 5
24. 한국전산원. 전자정부 구현을 위한데이터웨어하우스 기술 및 표준화 연구. 1998
25. 호승희, 채영문, 조경원 등. 의료분야의 지식경영을 위한 데이터마이닝응용. 연세대학교 보건대학원. 1999
26. 홍의기. 데이터웨어하우스를 이용한 임원정보시스템 아키텍처 구현. 한국과학기술원 테크노경영대학원 석사학위논문. 1997
27. Adrains, P, Zantinge, D. Data Mining. Aoblisian-Wesley. 1996
28. Brooking A. Intellecture Capital. International Thomson Publishing Company. 1996
29. Bigus JP. Data Mining with Neural Network. McCraw-Hill. 1996
30. Fayyad UM, Piatetsky-Shapiro G, Smyth P. From Data Mining to Knowledge Discovery, In Advances in Knowledge Discovery and Data Mining. MIT Press. 1996. p1~34
31. Parsaye K. DLAP and Data Mining Bridging The Gap. <http://www.w.w.datamining.com/datamime/bridge.htm>
32. Garey MR, Johnson DS. Computers and Intractability: A Guide to the Theory of NP-Completeness, W.H. Freeman and Company. 1979
33. Nonaka. The Knowledge-Creating Company. Oxford University Press. 1995
34. Rijsbergen V. An algorithm for Information Structuring and Retrieval Computer Journal 1971; 14: 407~412
35. Voorhees E. The Effectiveness and Efficiency of Agglomerative Hierarchic Clustering in Document Retrieval. Ph.D. Dissertation, 1986
36. Zamir O, Etzioni O, Madani O, Karp R. Fast and Intuitive Clustering of WebDocuments. In Proceedings of the KDD 97, 1997