

1. 신환자 이탈 분석



원자력병원 의무기록과

발표자:함승우




-목 차-

1. 활동목적

2. 자료수집 및 정제

3. 분석결과

- 1) 일반적 특성, 내원상태, 진료 상태에 따라 이탈환자 분석
 - 2) 로지스틱 분석을 통한 이탈에 영향을 주는 설명변수분석
 - 3) 판별분석을 통한 이탈에 영향을 주는 설명변수 분석
 - 4) 나무모형을 통한 이탈 예측 모형 분석
- 

1. 활동목적

- 병원의료 정보를 이용하여 이탈환자의 일반적 특성을 파악
- 이탈환자의 이탈 원인 변수를 모형화
- 가장 활용 가능성이 높은 모형을 찾아낸다
- 선택된 모형에 대해 예측도를 검증하여 모형의 활용 가능성을 제시
- 보건정보에 대한 세부적 분석을 통하여 데이터 속에 숨어 있는 정보를 활용
- 이탈방지 및 보다 향상된 의료정보제공 구축하여 표준화된 기초데이터 제공
- 고객에 대한 충분한 정보를 공유

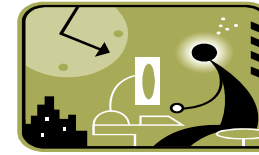
데이터 마이닝 (Data Mining)

- 대량의 가공하지 않은 데이터로부터 소량의 귀중한 정보(지식)을 찾아내는 과정
- 데이터 분석 및 예측모형 적합이라 할 수 있으므로 기존 통계학의 틀에 벗어난 것은 없으며
예측모형은 군집분석 모형, 연관규칙 모형, 나무 모형, 신경망 모형, 로지스틱 회귀모형등이다
- 활용은 고객관계관리, 신용평가, 부정행위 적발, 이미지 분석, 생명 정보학
- 데이터 마이닝을 통한 고객세분화(Customer Segmentation) 전략
 - 고객확보전략 : 신규고객 확보, 타사고객 흡수
 - 이탈방지전략 : 기존고객 이탈방지(포인트제, 마일리지)
 - 되찾기 전략 : 타사이동 자사유도(번호이동 서비스)
 - 교차판매 전략 : 자사고객 대한 새로운 서비스 제공(Sky Chanel open)



2. 자료수집 및 정제

자료수집
'06년1월1일~'07년 10월31일
외래환자 464,759건



신환 누적데이터(99,753)

신환중복제외(35,326)

예약

진료환자

이탈환자

- 2006년 환자가 예약부도가 발생하지 않고 2007년 1월 1일 기준 계속 진료 하는 환자
- 2007년 신환자가 동기간 예약부도 발생하지 않고 계속 진료 하는 환자

- 2006년 환자가 동기간 동안 진료 후 예약부도 발생 2007년 1월 1일 기준 계속 진료 하지 않은 환자
- 2007년 환자가 동기간 예약부도 발생하고 진료 받지 않은 환자

3. 분석결과

1) 일반적 특성, 내원상태, 진료 상태에 따라 이탈환자 분석

일반적특성, 내원상태, 진료상태

[투입변수]

- ◆성별(남.여)
- ◆연령별
- ◆지역별 (서울.지방)
- ◆보험구분 (일반,보호,보험,자보)
- ◆진료과(내과계,외과계)
- ◆특진유무(특진,일반)
- ◆초.재진
- ◆오전.오후진료
- ◆요일(월~일)
- ◆분기(1~4분기)
- ◆암환자

[종속변수]

- ◆ 이탈환자(예, 아니요)

H0: 투입변수와 이탈간에는 상호 관련성 없다(독립적)($p>0.005$)

H1: 투입변수와 이탈환자간에는 상호 관련성 있다 ($p<0.005$)

카이제곱 검정
(범주형데이터)

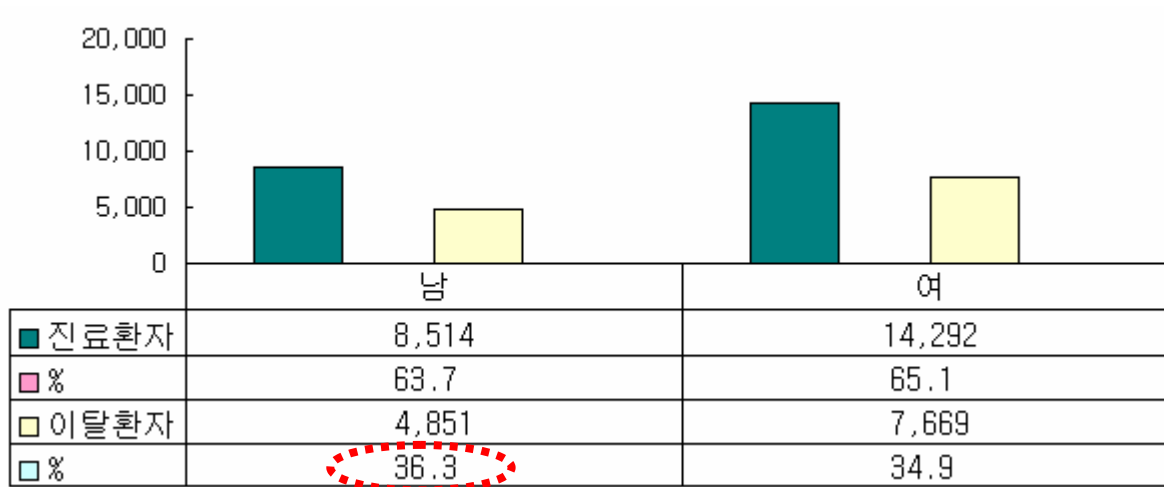


진료환자



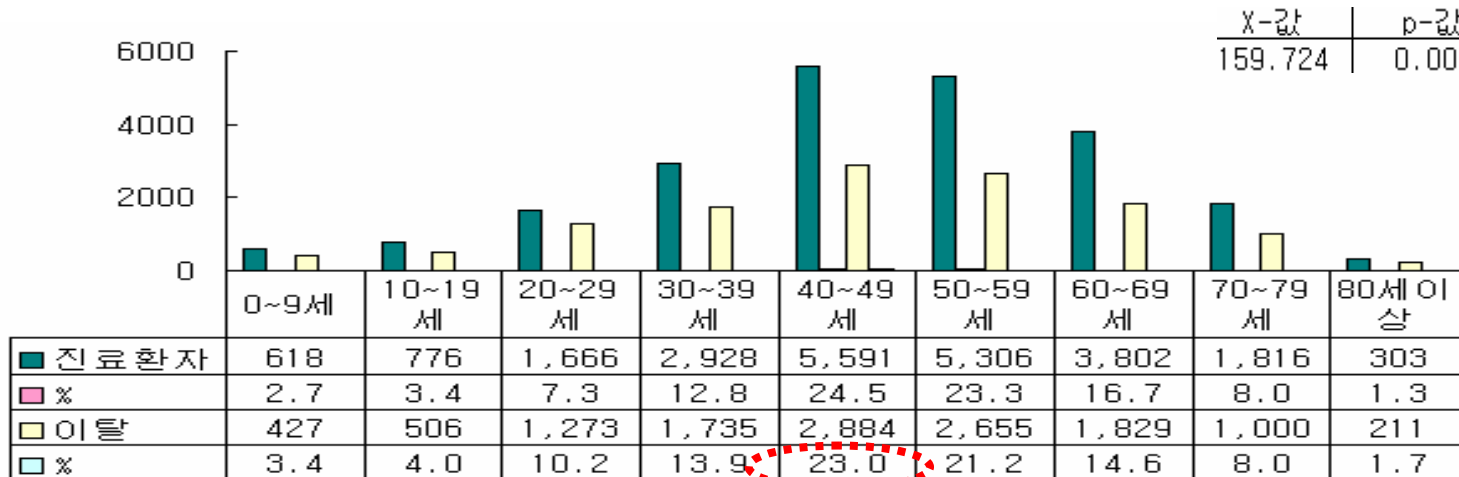
이탈환자

성별 : 남자(36.3%)가 여자(34.9%)보다 높았으며 유의한 차이가 있었다 (p<0.005)



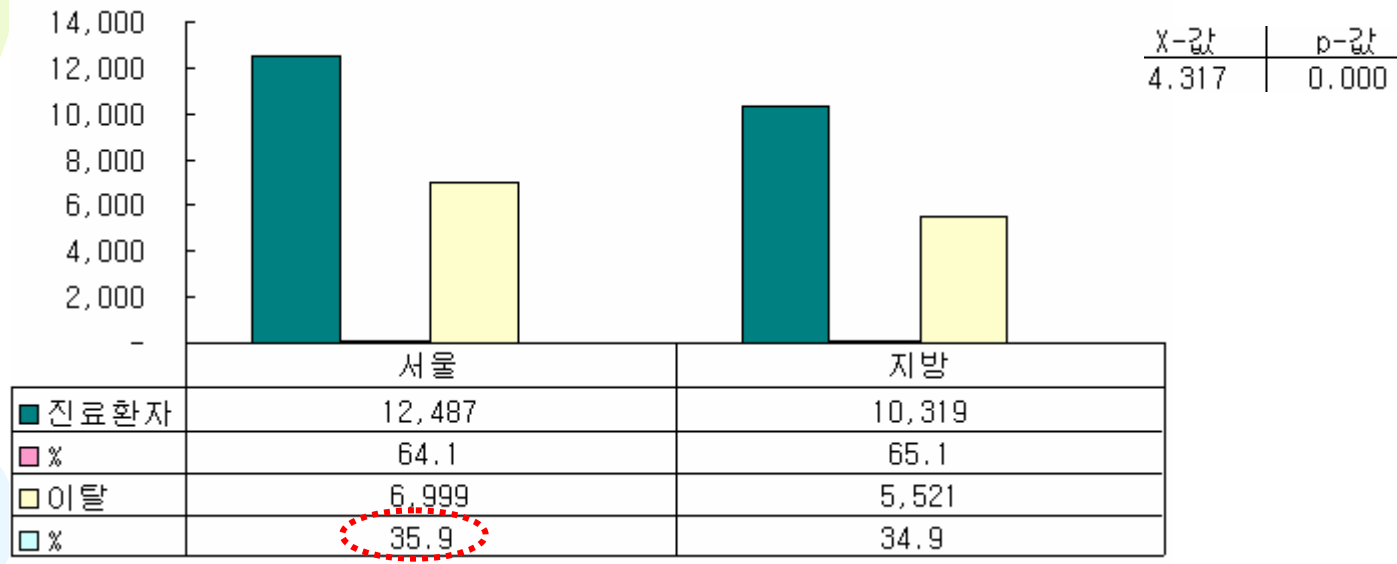
X-값	p-값
6.868	0.009

연령별 : 40~49세(23.0%)가 높았으며 유의한 차이가 있었다(p<0.005)

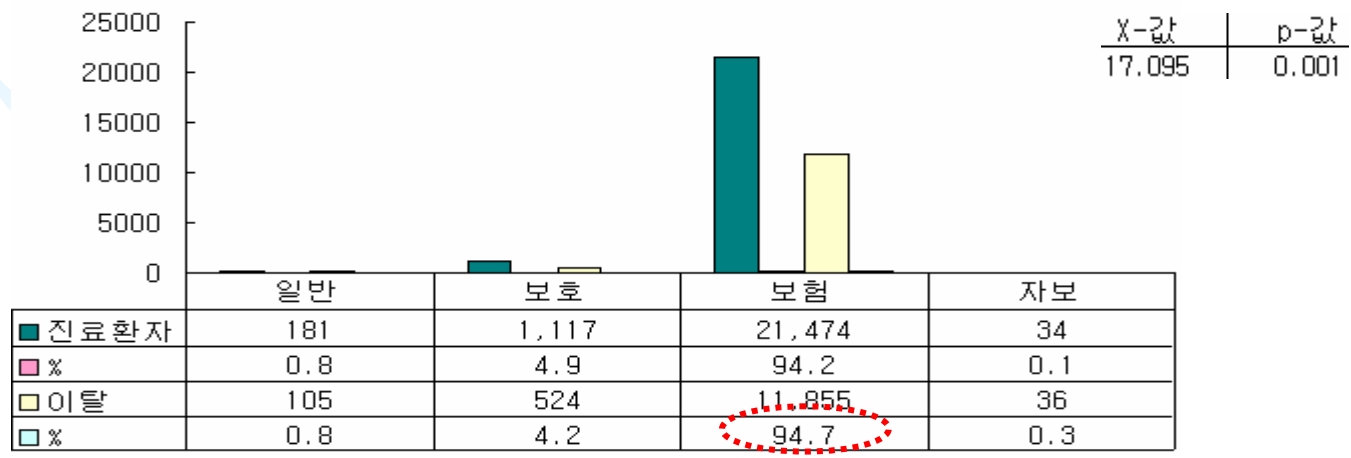


X-값	p-값
159.724	0.000

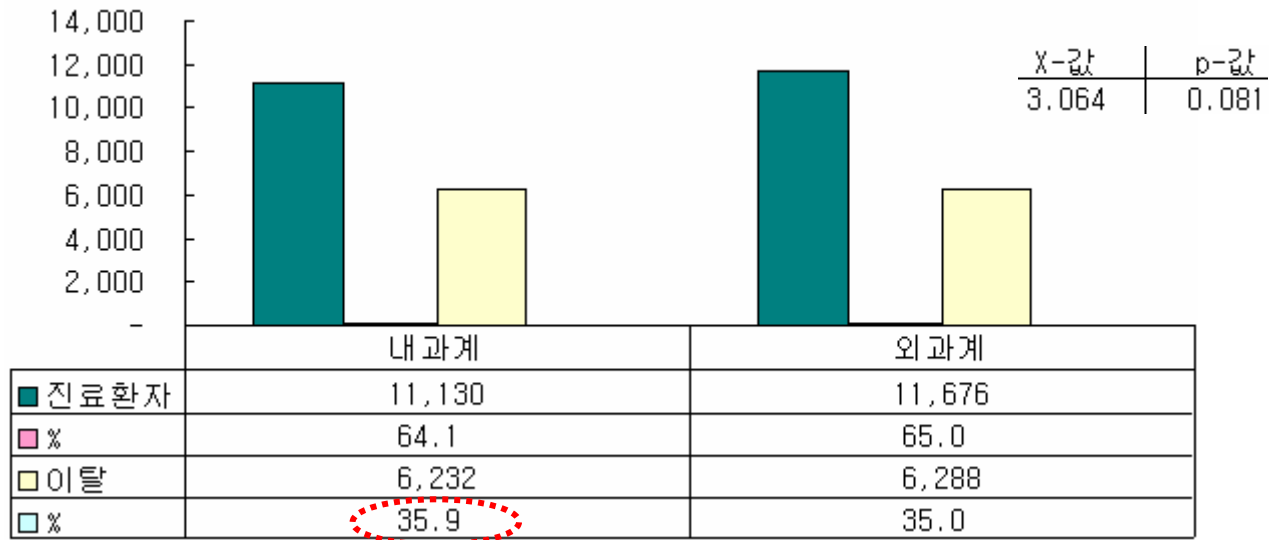
지역별 : 서울(35.9%)이 지방(34.9%)보다 높았으며 유의한 차이가 있었다($p < 0.005$)



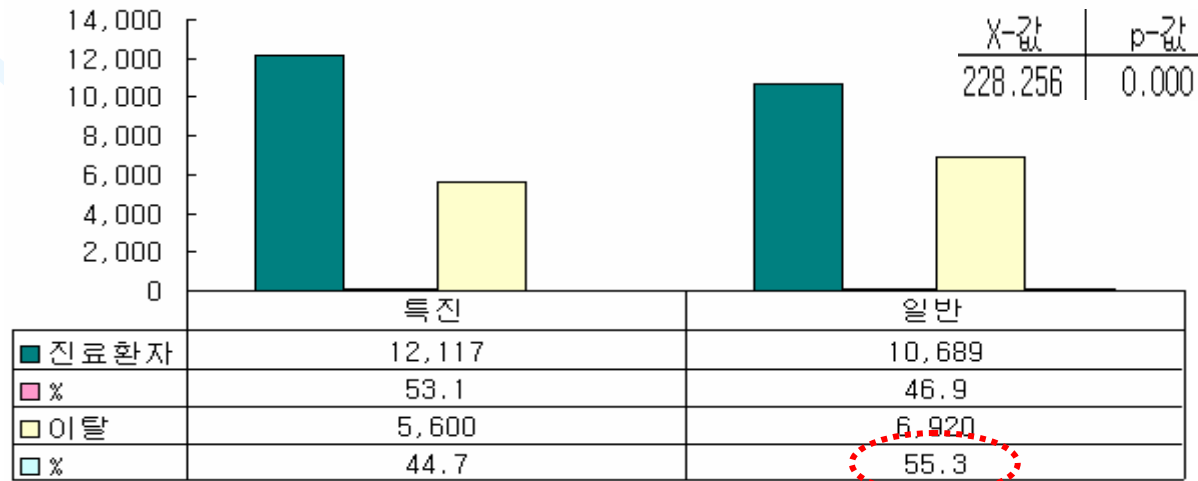
보험구분 : 보험(94.7%)이 높았으며 유의한 차이가 있었다($p < 0.005$)



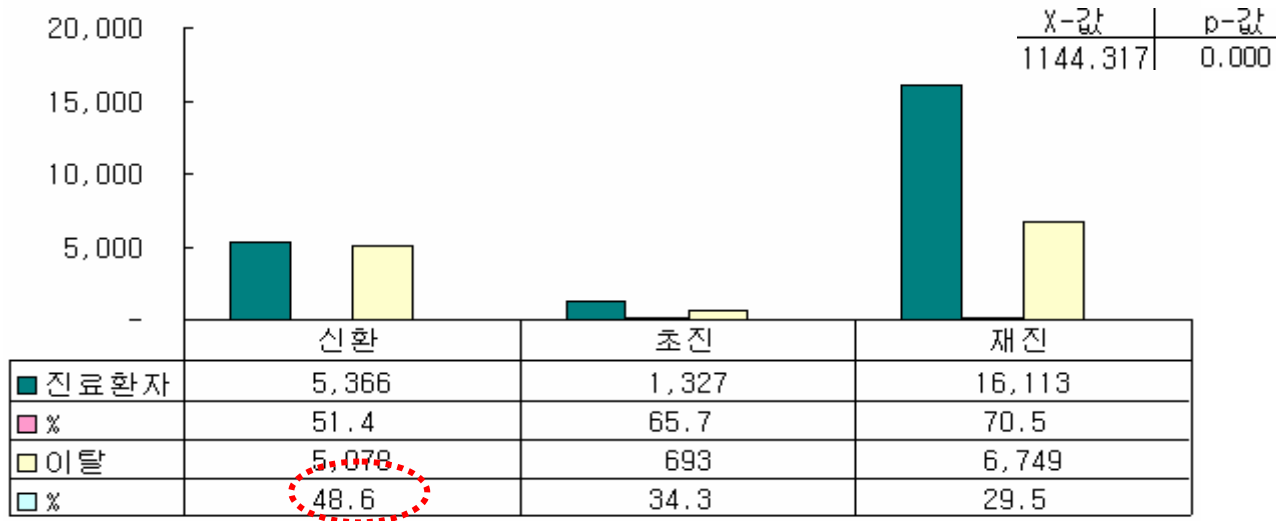
진료과(내.외과계) : 내과계(35.9)가 높았으나 유의한 차이는 없었다($p>0.005$)



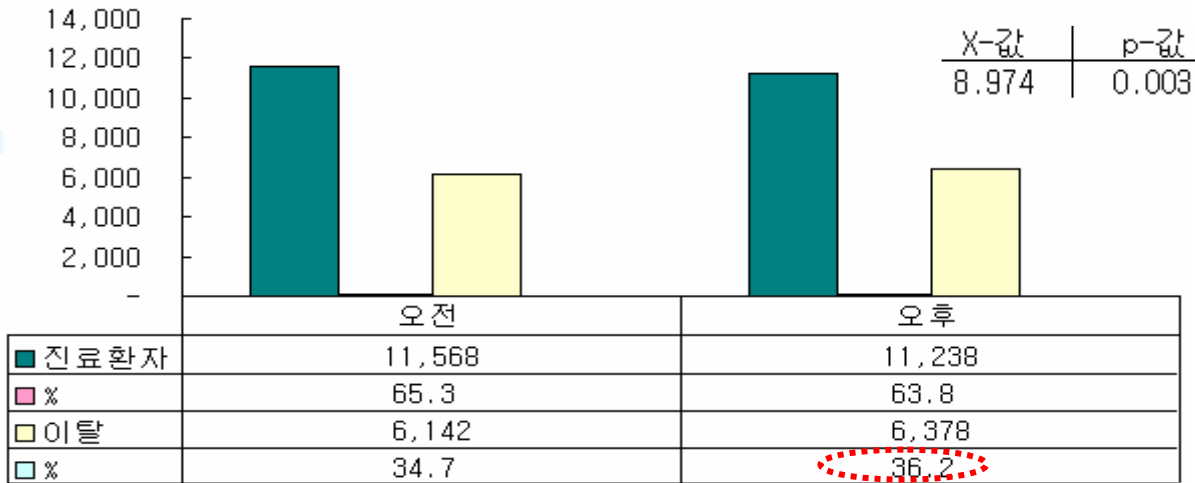
특진유무 : 일반(55.3)이 높았으며 유의한 차이를 나타냈다($p>0.005$)



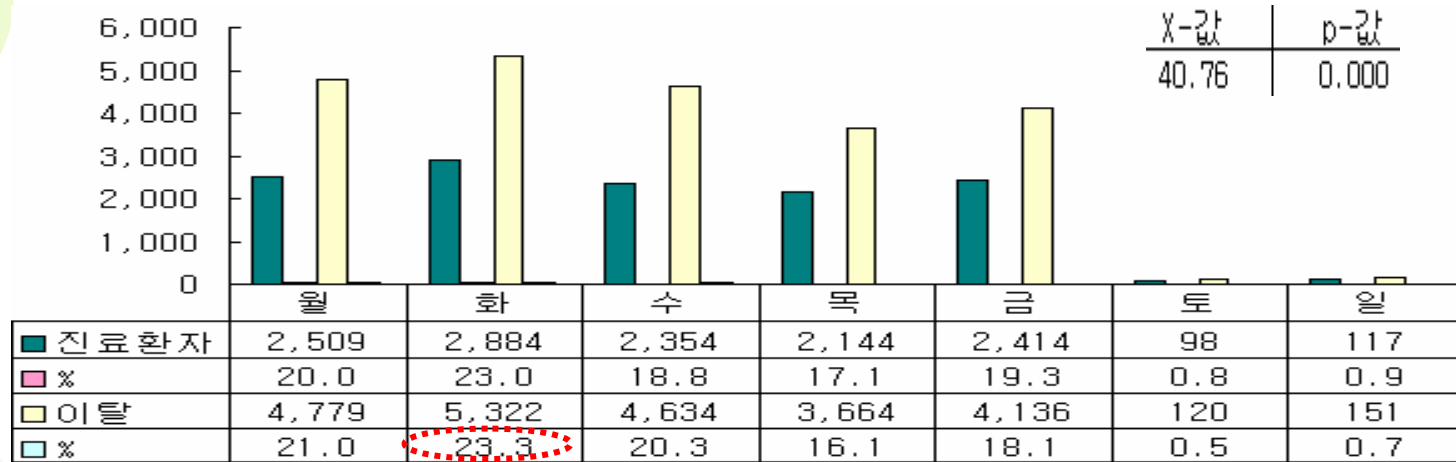
초.재진 : 신환(48.6%)이 높았으며 유의한 차이를 나타냈다(p<0.005)



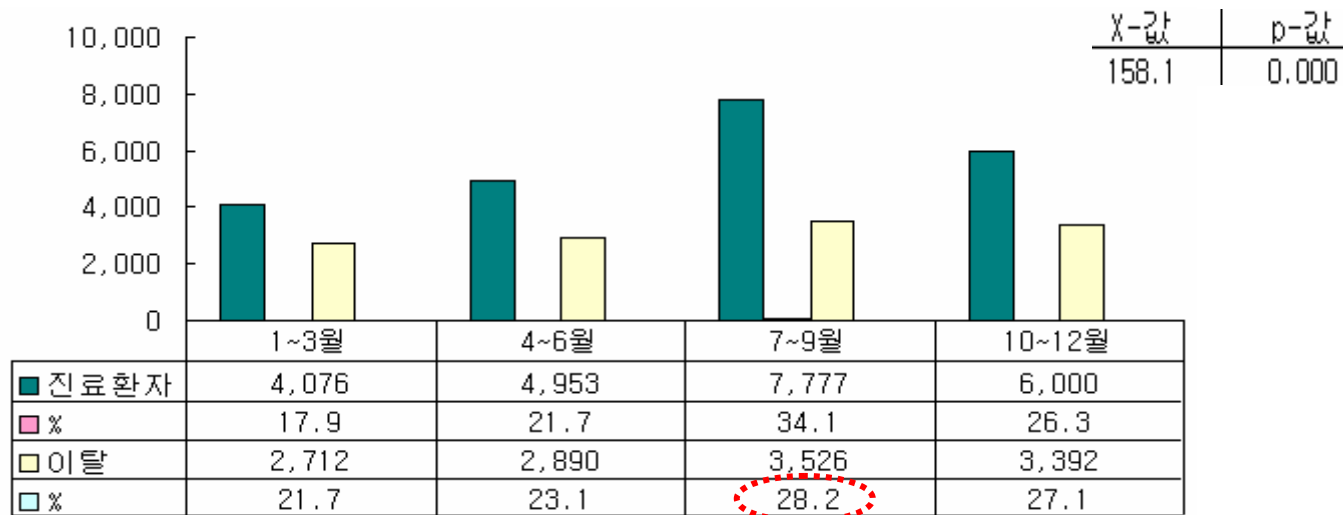
오전.오후 진료 : 오후(36.2%)가 높았으며 유의한 차이를 나타냈다(p<0.005)



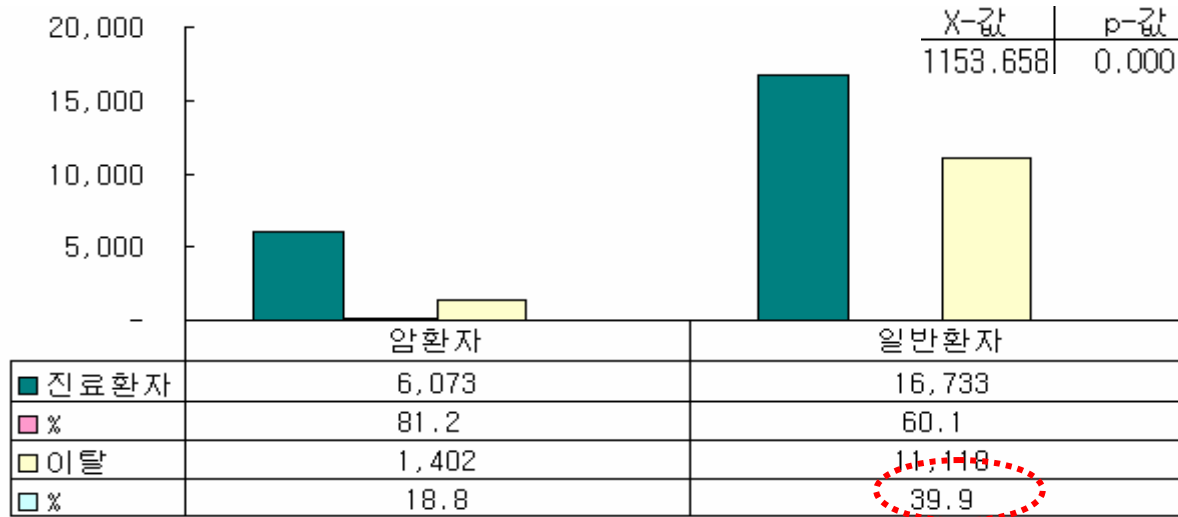
요일별 : 화요일(23.3%)이 높았으며 유의한 차이를 나타냈다($p < 0.005$)



분기별 : 이탈환자는 3분기(28.2%)가 높았으며 유의한 차이를 나타냈다($p < 0.005$)



암환자 : 일반환자(39.9%)가 높았으며 유의한 차이를 나타냈다($p < 0.005$)



2. 로지스틱 회귀분석을 통한 이탈에 영향을 주는 설명변수분석

로지스틱은 종속변수(질적변수)와 설명변수 사이의 인과관계를 설명하고 주어진 설명변수의 수준에서 종속변수를 예측

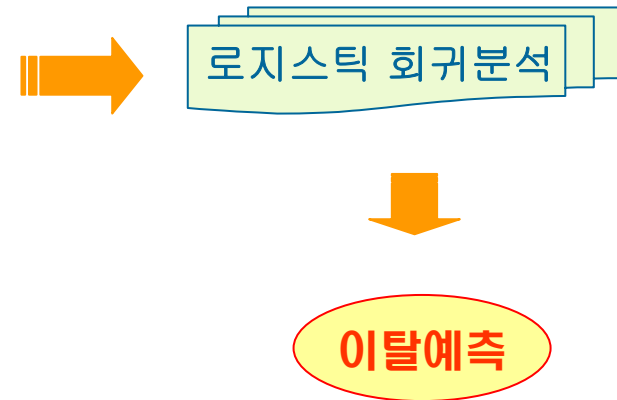
연구 : 설명변수는 이탈여부에 유의적인 영향을 미친다

[설명변수] z점수화

- ◆ 성별(1 남, 2 여)
- ◆ 보험구분(1일반, 2보호, 3보험, 4자보)
- ◆ 내외과계(1내과계, 2외과계)
- ◆ 특진유무(1특진, 2일반)
- ◆ 초재구분(1신환, 2초진, 3재진)
- ◆ 오전.오후진료(1오전, 2오후)
- ◆ 요일(1월, 2화, 3수, 4목, 5금, 6토, 7일)
- ◆ 분기(1분기, 2분기, 3분기, 4분기)
- ◆ 암구분(1암환자, 2일반환자)

[종속변수]

- ◆ 이탈여부 (1 예, 2 아니요)



- 로지스틱 분석으로 이탈여부에 예측 정확도는 전체적으로 65.7%
- 성별, 보험구분, 내외과계, 특진유무, 초.재구분, 오전.오후진료, 요일, 암진료여부, 분기별등이 Z점수(기술평균전환 점수)를 설명변수로 하였을때 이탈예측에 유의하였다(p<0.05)

분류표

	제곱합	예측치			정확도
		예	아니요	전체	
관찰치	예	2,875	9,456	23.0	
	아니요	2,473	20,333		
	전체				89.2
-2LL=43626.433					65.7
카이제곱(절편모형-이론모형)=2306.965(df 9, p=0.000), Nagelkerke R2 =0.087					

회귀분석

독립변수	회귀계수	표준오차	Ward	자유도	유의확률	Exp(B)
(상수)	.648	.012	2065.628	1	0.000	1.913
성별	.036	.012	9.469	1	0.002	1.037
보험구분	.053	.012	11.499	1	0.001	.962
내외과계	.134	.012	19.768	1	0.000	.949
특진유무	.331	.012	127.817	1	0.000	.875
초재구분	.042	.011	841.203	1	0.000	1.392
오전.오후진료	.041	.012	13.326	1	0.000	.959
요일	.393	.011	12.567	1	0.000	.960
암구분	.028	.014	848.419	1	0.000	.675
분기	.648	.012	5.596	1	0.018	.973

$$\text{Logit(이탈)} = 0.65 + 0.04(\text{성별}) + 0.5(\text{보험}) + 0.13(\text{내.외}) + 0.33(\text{특진}) + 0.04(\text{초.재}) + 0.04(\text{오전.후}) + 0.39(\text{요일}) + 0.03(\text{암}) + 0.65(\text{분기})^{14}$$

3. 판별분석을 통한 이탈에 영향을 주는 설명변수 분석

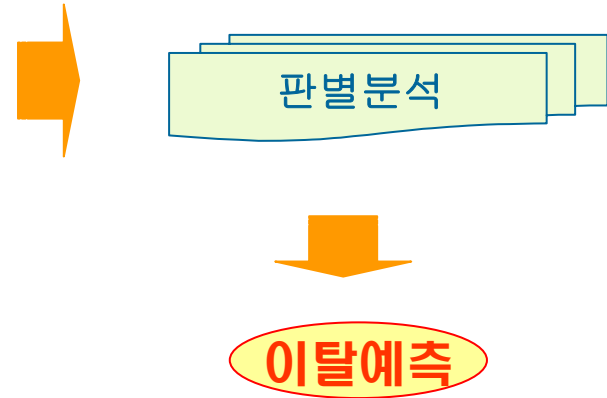
판별분석은 설명변수를 통하여 집단 구성원을 판별 혹은 예측하고 설명변수 중 영향력이 가장 큰 변수가 무엇인지 찾고자 할 때 사용한다.

[설명변수]

- ◆ 성별(1 남, 2 여)
- ◆ 보험구분(1일반, 2보호, 3보험, 4자보)
- ◆ 내외과계(1내과계, 2외과계)
- ◆ 특진유무(1특진, 2일반)
- ◆ 초재구분(1신환, 2초진, 3재진)
- ◆ 오전.오후진료(1오전, 2오후)
- ◆ 요일(1월, 2화 3수 4목 5금 6 토 7일)
- ◆ 분기(1분기, 2분기, 3분기, 4분기)
- ◆ 암구분(1암환자, 2일반환자)

[종속변수]

- ◆ 이탈여부(1 예, 2 아니요)



- 판별분석으로 이탈여부에 예측 정확도는 60.4%
- 분류함수 계수는 성별, 보험구분, 내외과계, 특진유무, 초.재구분, 오전.오후진료, 요일, 암진료여부, 분기별등이다

피셔의 선형판별함수(Fisher's Linear Discriminant Function)

이탈함수= $-0.75 - 0.03(\text{성별}) + 0.03(\text{보험}) + 0.04(\text{내.외}) + 0.09(\text{특진}) - 0.23(\text{초재}) + 0.03(\text{오전.후}) + 0.03(\text{요일}) + 0.23(\text{암}) + 0.2(\text{분기})$

진료함수= $-0.71 - 0.01(\text{성별}) + 0.01(\text{보험}) + 0.02(\text{내.외}) + 0.05(\text{특진}) - 0.13(\text{초재}) + 0.01(\text{오전.후}) + 0.02(\text{요일}) + 0.13(\text{암}) + 0.01(\text{분기})$

분류표

	이탈여부	예측치		
		예	아니요	전체
관찰치	예	7,036(56.2)	5,484(43.8)	12,520(100)
	아니요	8,506(37.3)	14,306(62.7)	22,806(100)

판별분석

독립변수	고유값	Wilks 람다	정준상관	판별함수와 독립변수와의 상관
성별				.054
보험구분				.058
내외과계				-.036
특진유무	0.067	.937(0.000)	25.0%	.312
초재구분				.062
오전.오후진료				.003
요일				.711
암구분				.142
분기				

3. 나무모형을 통한 이탈 예측 모형 분석

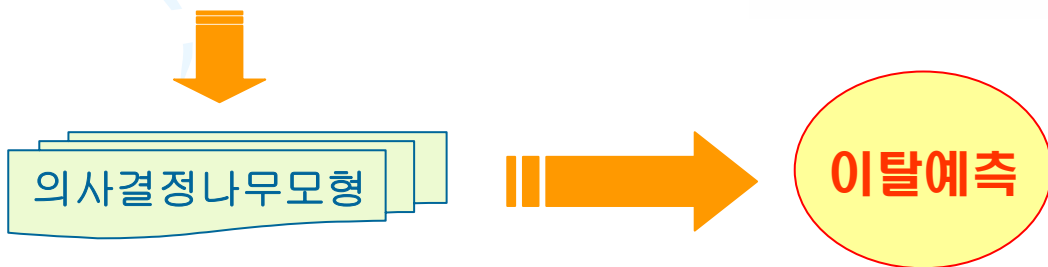
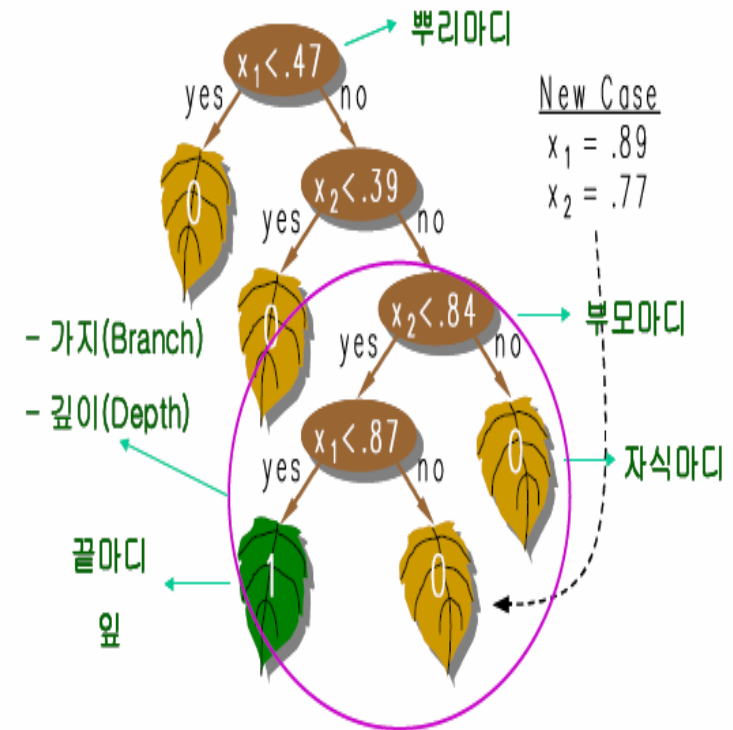
의사결정나무는 의사결정규칙(decision rule)을 도표화하여 관심대상이 되는 집단을 몇 개의 소 집단으로 분류(classification) 하거나 예측(prediction)을 수행하는 분석방법

[입력변수]

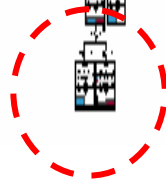
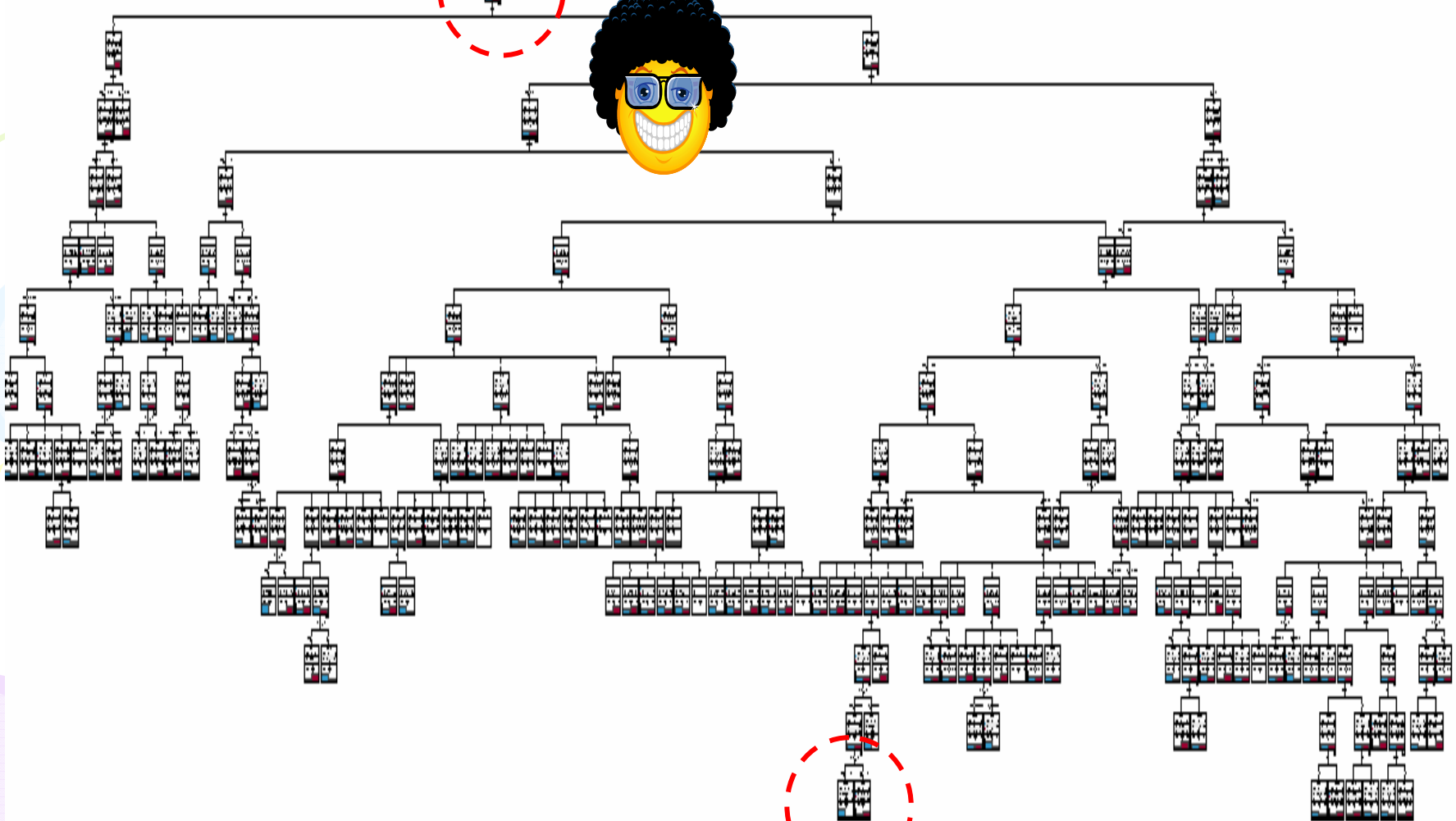
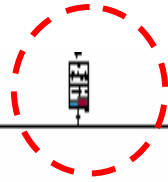
- ◆ 성별(1 남, 2 여)
- ◆ 보험구분(1일반, 2보호, 3보험, 4자보)
- ◆ 내외과계(1내과계, 2외과계)
- ◆ 특진유무(1특진, 2일반)
- ◆ 초재구분(1신환, 2초진, 3재진)
- ◆ 오전.오후진료(1오전, 2오후)
- ◆ 요일별(1월, 2화, 3수, 4목, 5금, 6토, 7일)
- ◆ 암구분(1암환자, 2일반환자)

[출력변수]

- ◆ 이탈여부(1이탈 2진료)



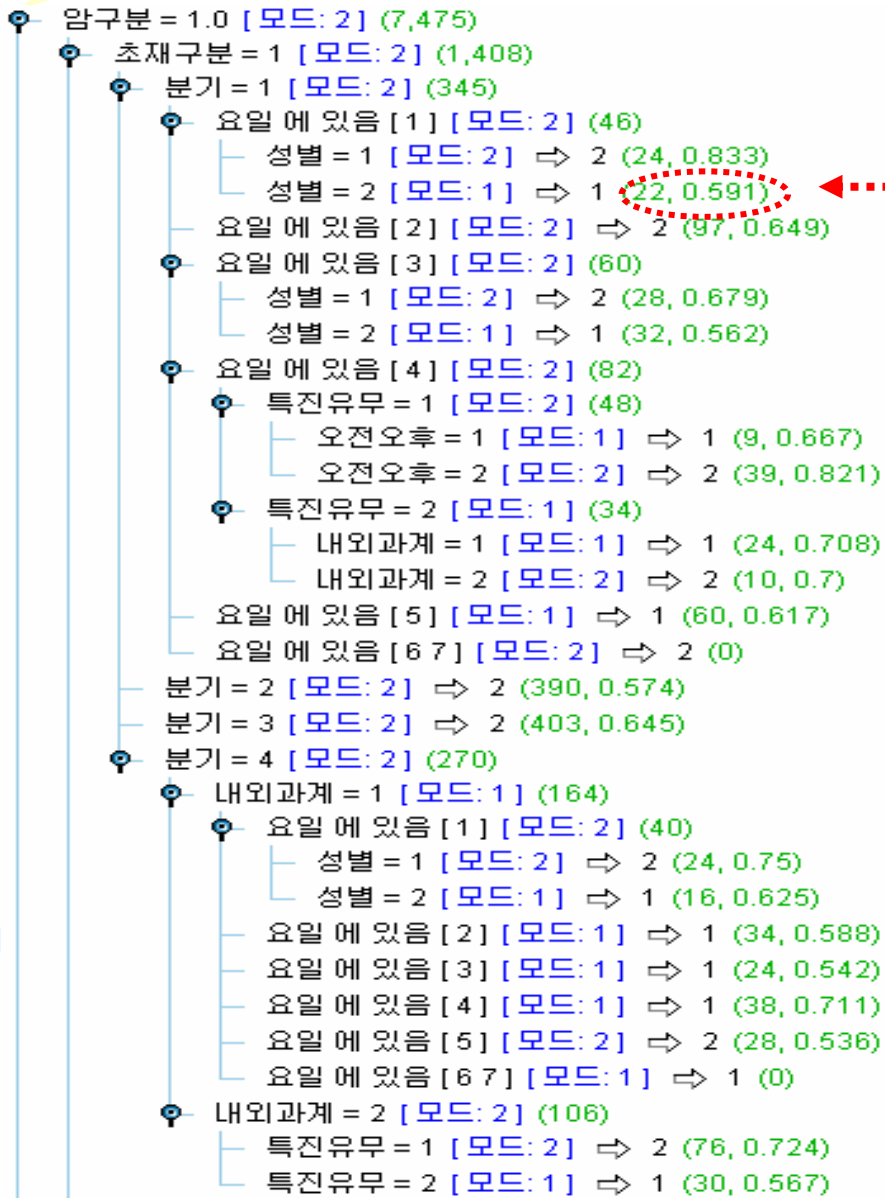
뿌리마디



뿌리마디

이탈모형

1	35.44	12520
2	64.56	22806
합계		100.00 35326



[분석 활용 예시]

이탈[모드:1]

진료[모드:2]

- 총이탈환자중 암구분
- 암환자중 진료환자(7,475)
- 초재진 환자중 신환자(1,408)
- 분기별로는 1분기(1~3월;345)
- 월요일 진료환자 중 여성환자 22명의

이탈율 59.1%

생략