

공 학 석 사 학 위 논 문

대형 교통사고 발생지점의 유형화와 영향요인
판별 및 분류모델 구축에 관한 연구

2005년 8월

부산대학교 환경대학원

도 시 계 획 학 과

김 봉 기

공 학 석 사 학 위 논 문

대형 교통사고 발생지점의 유형화와 영향요인
판별 및 분류모델 구축에 관한 연구

지 도 교 수 정 현 영

2005년 8월

부산대학교 환경대학원

도 시 계 획 학 과

김 봉 기

김봉기의 공학석사 학위논문을 인준함

2005년 6월 일

주 심 손 태 민 (인)

위 원 이 성 호 (인)

위 원 정 현 영 (인)

<제목 차례>

제1장 서론	1
제1절 연구 목적	1
제2절 연구 방법 및 범위	4
 제2장 교통사고의 일반적 현황	6
제1절 교통사고의 발생 추이	6
제2절 대형 교통사고의 발생 추이	8
제3절 분석 대상 대형 교통사고의 발생 특성	10
1. 빈도분석에 따른 발생 특성	10
2. 교차분석에 따른 발생 특성	17
 제3장 발생지점 유형화와 영향요인 판별, 분류모델의 이론적 배경	23
제1절 군집분석의 이론적 배경	23
제2절 판별모델의 이론적 배경	25
제3절 분류모델의 이론적 배경	29
1. 이산형 목표변수에 대한 분리 기준	30
2. 연속형 목표변수에 대한 분리 기준	32
3. 가지치기	33
4. 결손 값에 대한 대체규칙	33
5. 의사결정나무의 장점 및 단점	34
 제4장 발생지점 유형화와 영향요인 판별, 분류모델 구축	36
제1절 대형 교통사고 충돌 형태별 영향요인 판별, 분류모델	36
1. 판별모델	36
2. 분류모델	41

제2절 대형 교통사고 발생지점 유형화	46
제3절 집단별 충돌 형태별 영향요인 판별, 분류모델	53
1. 제1집단	53
2. 제2집단	60
3. 제3집단	67
4. 제4집단	74
제4절 발생지점 유형별 특성과 교통사고 예방 대책	82
1. 판별모델 결과	82
2. 분류모델 결과	88
3. 판별, 분류모델의 종합적 결과	92
제5장 문제점 및 개선방안	95
제1절 문제점	95
제2절 개선 방안	95
제6장 결론	97
<참고문헌>	99

<표 차례>

<표 2-1> 교통사고 발생 추이	7
<표 2-2> 전체 교통사고와 대형 교통사고 발생 추이	9
<표 4-1> 충돌 형태별 현황(전체)	36
<표 4-2> 충돌 형태별 외적 변수의 판별 범위(전체)	36
<표 4-3> 충돌 형태별 판별모델의 판별 결과(전체)	38
<표 4-4> 위험 도표(전체)	42
<표 4-5> 이익 도표(전체)	43
<표 4-6> 최적 유형화 수준의 지표	46
<표 4-7> 집단별 사례 지점	50
<표 4-8> 집단별 특성	51
<표 4-9> 충돌 형태별 현황(제1집단)	53
<표 4-10> 충돌 형태별 외적 변수의 판별 범위(제1집단)	53
<표 4-11> 충돌 형태별 판별모델의 판별 결과(제1집단)	54
<표 4-12> 위험 도표(1집단)	58
<표 4-13> 이익 도표(1집단)	58
<표 4-14> 충돌 형태별 현황(제2집단)	60
<표 4-15> 충돌 형태별 외적 변수의 집단 중심값(제2집단)	60
<표 4-16> 충돌 형태별 판별모델의 판별 결과(제2집단)	61
<표 4-17> 위험 도표(2집단)	65
<표 4-18> 이익 도표(2집단)	65
<표 4-19> 충돌 형태별 현황(제3집단)	67
<표 4-20> 충돌 형태별 외적 변수의 판별 범위(제3집단)	67
<표 4-21> 충돌 형태별 판별모델의 판별 결과(제3집단)	68
<표 4-22> 위험 도표(3집단)	72
<표 4-23> 이익 도표(3집단)	72
<표 4-24> 충돌 형태별 현황(제4집단)	74

<표 4-25> 충돌 형태별 외적 변수의 판별 범위(제4집단)	74
<표 4-26> 충돌 형태별 판별모델의 판별 결과(제4집단)	76
<표 4-27> 위험 도표(4집단)	80
<표 4-28> 이익 도표(4집단)	80
<표 4-29> 발생지점 유형별 충돌 형태별 판별 특성	85
<표 4-30> 판별모델에 따른 발생지점 유형별 교통사고 예방 대책	87
<표 4-31> 발생지점 유형별 충돌 형태별 분류 특성	89
<표 4-32> 분류모델에 따른 발생지점 유형별 교통사고 예방 대책	91
<표 4-33> 발생지점 유형별 교통사고 예방 대책	92
<표 5-1> 표준 원 자료 시트(작성 예시)	96

<그림 차례>

<그림 2-1> 충돌 형태	10
<그림 2-2> 법규위반 유형	11
<그림 2-3> 발생 시간대	12
<그림 2-4> 총 차로수	13
<그림 2-5> 노면 포장상태	14
<그림 2-6> 도로 중앙부 형태	15
<그림 2-7> 타 차량 제공 요인	16
<그림 2-8> 충돌 형태와 법규위반 유형	17
<그림 2-9> 충돌 형태와 발생 시간대별	18
<그림 2-10> 충돌 형태와 총 차로수	19
<그림 2-11> 충돌 형태와 노면 포장상태	20
<그림 2-12> 충돌 형태와 도로 중앙부 형태	21
<그림 2-13> 충돌 형태와 타 차량 제공 요인	22
<그림 4-1> 충돌 형태별 외적 변수의 Data Score Histogram(전체)	37
<그림 4-2> 수량화 이론 II류에 의한 충돌 형태별 판별모델(전체)	40
<그림 4-3> 충돌 형태별 영향요인 분류모델(전체)	41
<그림 4-4> Ward법에 의한 수상도(Dendrogram)	49
<그림 4-5> 충돌 형태별 외적 변수의 Data Score Histogram(제1집단)	54
<그림 4-6> 수량화 이론 II류에 의한 충돌 형태별 판별모델(제1집단)	56
<그림 4-7> 충돌 형태별 영향요인 분류모델(1집단)	57
<그림 4-8> 충돌 형태별 외적 변수의 Data Score Histogram(제2집단)	61
<그림 4-9> 수량화 이론 II류에 의한 충돌 형태별 판별모델(제2집단)	63
<그림 4-10> 충돌 형태별 영향요인 분류모델(2집단)	64
<그림 4-11> 충돌 형태별 외적 변수의 Data Score Histogram(제3집단)	68
<그림 4-12> 수량화 이론 II류에 의한 충돌 형태별 판별모델(제3집단)	70
<그림 4-13> 충돌 형태별 영향요인 분류모델(3집단)	71

<그림 4-14> 충돌 형태별 외적 변수의 Data Score Histogram(제4집단)	75
<그림 4-15> 수량화 이론 II류에 의한 충돌 형태별 판별모델(제4집단)	78
<그림 4-16> 충돌 형태별 영향요인 분류모델(4집단)	79

제1장 서론

제1절 연구 목적

자동차화(Motorization)의 진행으로, 일상생활에서 필수품화된 자동차는 그 수요가 지속적으로 증가하고 있는데 반해, 교통여건은 이에 부응하지 못하고 있는 현실 여건 하에서 교통사고는 이미 사회문제화 되고 있는 실정에 있다.

특히 교통사고의 주된 원인을 교통사고 요인의 5M이라는 규정¹⁾에 근거하여 살펴보면, 인간(Man = 운전자 등), 기계(Machine = 자동차 등) 그리고 환경(Medium = 도로 등)의 세 가지 요인들이 그 목적(Mission)을 달성하기 위해서는 상호 균형(Balance)을 잘 유지해야 함은 물론이고, 이를 위해 관리(Management)를 잘 해야 한다.

그러나 이 중에서 인간 - 기계 - 환경계의 균형이 붕괴되거나, 어느 한 요인이라도 잘못되면 곧 교통사고가 발생할 수 있으며, 더욱이 목적의 불분명과 관리의 잘못 역시도 교통사고의 발생 원인이 될 수 있으므로, 이러한 교통사고의 방지를 위해서는 이들 5가지 요인 모두가 중요한 역할을 담당해야 한다.

이처럼 다양한 요인들에 의해 발생하는 교통사고 중에서 대형 교통사고²⁾의 경우는 발생 건수 면에서 1990년부터 2003년에 걸쳐 10년 평균 증가율이 6.1%의 비율로, 사망자수는 8.4%의 비율로 매년 감소되고 있는 실정에 있다³⁾. 하지만 사망자와 부상자의 발생으로 인한 생산성 저하, 물질 피해로 인하여 재산상의 피해를 크게 유발시킨다는 점에서 그 어떤 분야보다도 국가적 차원에서의 연구가 절실히 필요하다.

그러나 지금까지의 교통사고에 대한 선행적 연구 결과들⁴⁾의 대부분은 주

1). 全日本交通安全協會, 人と車, 第6巻 第2號 通巻 47號, 平成 2年 2月號, pp.13~14.

2). 경찰청, 교통사고 처리 지침 제2조, 제3호, '대형 교통사고라 함은 사망 3명 또는 부상 20명 이상의 사고와 기타 사회 물의를 야기한 사고를 말한다.'라고 규정하고 있음

3). 도로교통안전관리공단, 2004년판 대형 교통사고 사례 분석, 2004, p.16.

로 전체 교통사고에 대한 예측모델 위주로 연구되어 왔다. 이에 반해, 대형 교통사고의 경우는 전체 교통사고에서 매우 중요한 위치를 점하고 있음에도 불구하고 단지 교차분석(Cross-tabs Analysis)⁵⁾에만 의존한 연구⁶⁾가 몇몇 있을 뿐, 대형 교통사고 발생지점의 특성별 유형화와 충돌 형태에 미치는 영향요인에 대해서는 심층적인 분석 및 연구가 수행된 적이 전무하다.

또한 대형 교통사고에 있어 충돌 형태에 미치는 영향요인에 대한 분석은 차후 대형 교통사고 감소를 위한 대책 수립 시 가장 중요하게 고려되어야 할 주된 요인으로 그 무엇보다도 우선적으로 취급되어야 할 만큼 대단히 중요한 의미를 갖고 있다.

따라서 본 연구에서는 이러한 대형 교통사고에 대한 연구의 일환으로서 대형 교통사고에 있어 발생지점을 특성별로 유형화하기 위하여 군집분석(Cluster Analysis)⁷⁾을 행하고, 아울러 충돌 형태에 미치는 영향요인에 대한 영향 정도를 판별할 수 있도록 하기 위하여 수량화 이론 II류(Quantification II Method)⁸⁾⁹⁾¹⁰⁾를 이용하여 판별모델을 구축하였다.

4). ① 이일병, 임현연, 한국의 교통사고 모형 개발에 관한 연구, 대한교통학회지 제8권 제1호, 1990, pp.73~88.

② 이일병, 임현연, 교통사고 예측모형 개발과 사고의 단기 예측, 대한국토도시계획학회지 제25권 제3호 통권 58호, 1990.11., pp.139~160.

5). 최종성, SPSS Ver 10 을 이용한 현대통계분석, 복두출판사, 2000, pp.181~188.

6). ① 이주형, 손동혁, 윤문교, 교통사고의 발생특성과 그에 따른 사고 요인분석에 관한 연구, 대한국토계획학회지 국토계획, 제25권 제1호 통권 56호, 1990. 3., pp.135~154.

② 김대웅, 임채문, 도로교통에 있어서 운전자 시각특성 분석, 대한교통학회지 제8권 제2호, 1990, pp.7~25.

7). 최종성, 전개서, pp.341~354.

8). 田中 豊, 垂水共之 編, Windows版, 統計解析ハンドブック-多變量解析-, 共立出版株式會社, 1996, pp.160~169.

9). 허명희, 수량화 방법론의 이해, 자유아카데미, 1992, pp.2.1~2.18.

10). 노형진, SPSS/Amos에 의한 사회조사분석 -범주형 데이터 분석 및 공분산

아울러 데이터 마이닝 기법의 일종인 의사결정나무분석법 중의 한가지인 C&RT(Classification and Regression Trees)¹¹⁾¹²⁾¹³⁾를 이용하여 분류모델을 구축하고 이에 대한 적합성을 평가함으로써, 대형 교통사고 감소를 위한 정량적 척도의 간략화를 도모하였으며, 이를 근간으로, 대형 교통사고와 관련한 자료의 구축과정에서 대형 교통사고 조사와 관련한 개선 방안들을 제시하고자 하는데도 연구 목적을 두고자 하였다.

구조분석 - 형설출판사, 2003, pp.89~101.

11). 최종후, 한상태, 강현철, 김은석 공저, Answer Tree를 이용한 데이터 마이닝 의사결정나무분석, SPSS 아카데미, 1998, pp.42~48.

12). Answer Tree 2.0 User's Guide, SPSS Inc., 1998, pp.56~57, pp.143~152, pp.189~190, pp.192~202.

13). Claude Seidman 저, 권오주, 이정무 역, Microsoft SQL Server 2000 데이터 마이닝, Technical Reference, 정보문화사, 2001, pp.115~201.

제2절 연구 방법 및 범위

일반적으로 교통사고는 단독 인자에 의해서 일어날 수도 있지만, 대부분의 경우 여러 가지 인자들이 복합적으로 작용하여 발생하는 경우가 많아 그 원인과 결과를 정확히 규명한다는 것이 매우 어려운 경우가 많다.

따라서 본 연구에서는 대형 교통사고에 있어 여러 인자들의 복합적 작용을 알아보기 위한 미시적 분석의 일환으로 발생지점을 특성별로 유형화하고, 아울러 충돌 형태에 미치는 영향요인의 영향 정도를 판별 및 분류할 수 있는 모델을 구축하고자 하였다.

이를 위해, 먼저 1989년, 1991년, 1992년, 1997년, 2000년, 2001년, 2002년, 2003년, 2004년까지 총 9차례 발간된 “대형 교통사고 사례 분석” 보고서¹⁴⁾ 상에 나타난 대형 교통사고 사례 300건에 대하여 일차적으로 30개 항목의 자료를 구축하였다.

이후 자료의 동질성을 분석해 본 결과, 자료가 비교적 충실하게 구축된 충돌 형태를 포함한 7개 항목에 대해 51개 지점을 제외한 249개 지점을 유형화하기 위하여 군집분석을 행하였다.

또한 충돌 형태별에 대해 법규위반 유형, 사고발생 시간대별, 총 차로수, 노면 포장상태, 도로 중앙부 형태, 타 차량 제공요인 등의 6가지의 영향요인을 중심으로 영향정도를 분석하기 위하여, 수량화 이론 II류 분석법을 이용하여 인자득점 및 편상관계수(Partial Correlation)¹⁵⁾ 등을 산출하였으며, 경계점(Cutting Score)¹⁶⁾과 적중률(Hit Ratio)¹⁷⁾ 등으로 신뢰성을 검증한 후 최적의 판별모델을 구축하고자 하였다.

14). ‘대형 교통사고 사례 분석’ 보고서는 1989년, 1991년, 1992년, 1997년, 2000년,

2001년, 2002년, 2003년, 2004년까지 9차례 발간되었음

15). 최중성, 전게서, p.198

다중 상관분석에서 다른 변수들과의 관계는 고정되고 두 변수만의 관계에 대한 강도를 나타낸 것을 의미함

16). 채서일, 김종범, SPSS/PC+를 이용한 통계분석, 법문사, 1988, p.112

아울러 지니지수(Gini Index)¹⁸⁾에 의한 이진분리(binary split)¹⁹⁾를 수행하는 C&RT 분석법을 이용하여 최적의 분류모델도 구축하고자 하였다.

$$\text{Cutting Score(경계점)} = \frac{N_2 C_1 + N_1 C_2}{N_1 + N_2}$$

N_i : 집단의 표본수, C_i : 집단의 중심값(*Centroid*)

17). 상계서, p.115

$$\text{Hit Ratio(적중율)} = \frac{\sum_{i=1}^n P_{ii}}{T} \times 100(\%)$$

P_{ii} : i 집단을 i 집단으로 올바르게 판별한 사례수, T : 전체 사례수

18). 최종후, 한상태, 강현철, 김은석 공저, 전계서, p.42

19). 상계서, p.42

제2장 교통사고의 일반적 현황

제1절 교통사고의 발생 추이

교통사고에 대한 정의는 여러 학자들에 따라 약간의 차이가 있으나, 대부분 차량 등에 의하여 사람을 사상케 하거나 물건을 손상하는 것이라고 규정하고 있는 경우가 일반적이다.

이에 대한 정의를 도로교통법 제50조 제1항에 의거하여 보면, 자동차의 교통으로 인하여 사람을 사상케 하거나 물건을 손상하는 것으로 정의하고 있으며, 이에선 사망, 중상, 경상, 부상, 충돌, 추돌, 접촉, 전도, 전복, 추락 등의 10가지 유형으로 구분하고 있다.

이에 따라 1994년을 기점으로 과거 10년간 우리나라 교통사고 발생 추이를 <표 2-1>에서 살펴보면, 교통사고 발생건수는 1994년 266,107건에서 2003년 240,832건으로 10년 연평균 0.8%씩 감소하였고, 사망자 역시도 10,087명에서 7,212명으로 3.6%씩 감소하였으나, 부상자의 경우는 350,892명에서 376,503명으로 1.1%씩 증가한 것으로 나타나고 있다.

특히 자동차 등록대수가 1,000만대를 돌파한 1997년도에는 교통사고가 총 246,452건이 발생하여, 사망자 11,605명, 부상자 343,159명의 피해가 야기 되었으며, 이를 1일 평균으로 환산하여 보면, 615건의 사고가 발생하여, 32명이 사망하고, 786명이 부상을 입었던 것으로 나타나고 있다.

또한 인구 10만 명 당 사망자수와 차량 1만대 당 사망자수에 있어서는 각각 4.2%와 12.0%씩 감소하는 것으로 나타났다.

<표 2-1> 교통사고 발생 추이

구 분	교 통 사 고			사 망 자	
	발생건수(건)	사망(명)	부상(명)	인구 10만명 당 (명/10만명)	차량 1만대 당 (대/1만대)
1994	266,107	10,087	350,892	23	14
1995	248,865	10,323	331,747	23	12
1996	265,052	12,653	355,962	27	13
1997	246,452	11,603	343,159	25	11
1998	239,721	9,057	340,564	19	9
1999	275,938	9,353	402,967	20	8
2000	290,481	10,236	426,984	21	8
2001	260,579	8,097	386,539	17	6
2002	231,026	7,222	348,149	15	5
2003	240,832	7,212	376,503	15	4
연평균 증가율 ²⁰⁾ (%)	-0.8	-3.6	1.1	-4.2	-12.0

자료 : 도로교통안전관리공단, 2004년판 교통사고 통계 분석, 2004, p.17.

20). 도로교통안전관리공단, 2004년판 교통사고 통계 분석, 2004, p.4

$$10\text{년 연평균 증가율} = \left\{ \left(\frac{b}{a} \right)^{\frac{1}{n}} - 1 \right\} \times 100$$

b : 당해 년도 값, a : 기준 년도 값, n : 당해 년도 - 기준 년도

제2절 대형 교통사고의 발생 추이

다음에 제시된 <표 2-2>는 1994년부터 2003년까지의 10년간 발생한 전체 교통사고와 대형 교통사고의 발생 추이를 나타낸 것이다.

이를 보면, 대형 교통사고는 매년 전체 교통사고의 0.07%씩을 점하는 비율로 발생하였으며, 그 건수 면에서는 1995년과 1996년에 259건으로 가장 많이 발생하였다가, 2002년에는 최저치인 107건으로 10년 연평균 6.1%씩의 비율로 감소한 것으로 나타났다.

아울러 사망자의 경우도 1995년의 경우 772명으로 가장 높게 나타났다가, 2003년에는 247명으로 최저치를 나타내는 등 연평균 8.4%씩 감소한 것으로 나타났다.

<표 2-2> 전체 교통사고와 대형 교통사고 발생 추이 (단위 : 건, %)

구분 년도별	전체 교통사고 ^{자료 1}	대형 교통사고 ^{자료 2}	
	발생 건수	발생 건수	사 망 자
1994	266,107	221 (0.08)	627
1995	248,865	259 (0.10)	772
1996	265,052	256 (0.10)	732
1997	256,452	215 (0.09)	595
1998	239,721	182 (0.08)	477
1999	275,938	159 (0.06)	399
2000	290,481	215 (0.07)	624
2001	260,579	142 (0.06)	301
2002	231,026	108 (0.05)	256
2003	240,832	119 (0.04)	247
연평균 증가율(%)	-0.8	-6.1	-8.4

주 : ()내는 전체 교통사고에 대한 대형 교통사고의 비율(%)을 나타낸 것임

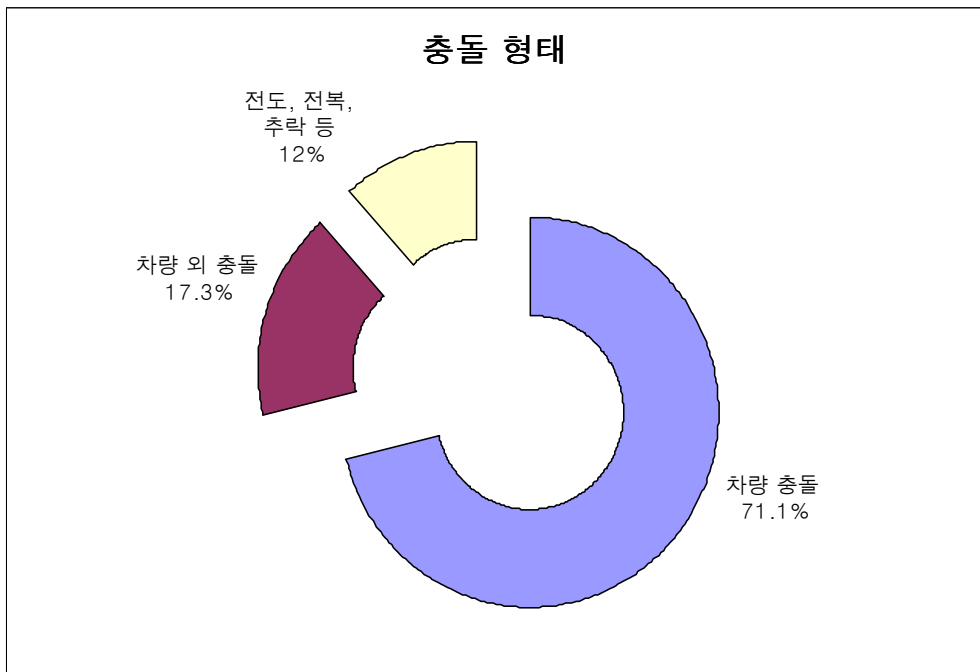
자료 1 : 도로교통안전관리공단, 2004년판 교통사고 통계 분석, 2004, p.4.

자료 2 : 도로교통안전관리공단, 2004년판 대형 교통사고 사례 분석, 2004, pp.16.

제3절 분석 대상 대형 교통사고의 발생 특성

1. 빈도분석(Frequency Analysis)²¹⁾에 따른 발생 특성

금번 분석 대상 대형 교통사고 자료는 249건으로, 이와 관련한 제 특성들을 살펴보고자 한다. 먼저 대형 교통사고의 충돌 형태는 <그림 2-1>에서 보는 바와 같이, 차량 충돌이 전체의 71.1%인 177건으로 가장 많은 것으로 나타났으며, 다음으로는 차량 외 충돌이 17.3%인 43건, 전도, 전복, 추락 등이 11.6%인 29건의 순으로 나타났는바, 이를 볼 때, 대형 교통사고는 속도의 제곱에 비례하여 충격력이 커짐으로 인하여 그 피해가 커지는 특성 상 차량 충돌에 의한 발생건수가 가장 많이 발생하는 것으로 나타났다.



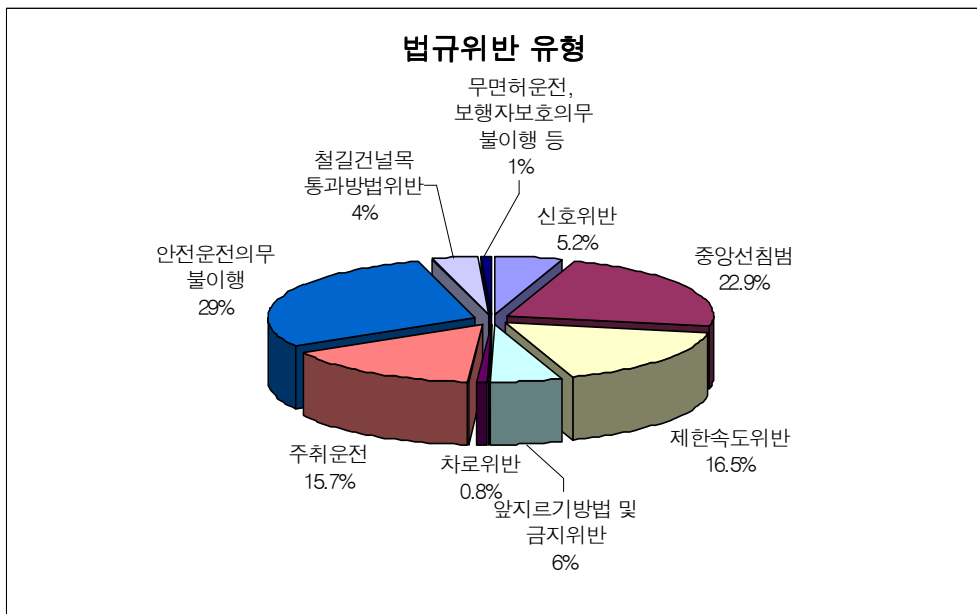
<그림 2-1> 충돌 형태

21). 최종성, 전게서, pp.145~156.

다음으로 대형 교통사고의 법규위반 유형을 <그림 2-2>에서 살펴보면, 안전운전의무불이행이 28.9%인 72건으로 가장 많은 것으로 나타났으나, 이는 교통사고에 있어 교통사고의 원인을 정확하게 알 수 없을 경우, 안전운전의무불이행으로 대다수가 처리되는 점을 감안한다면, 중앙선 침범 22.9%인 57건이 사실 상 가장 많은 것으로 볼 수 있을 것이며, 다음으로는 주취운전이 15.7%인 39건으로 높게 나타나고 있어 중앙선 침범과 주취운전이 대형 교통사고의 주된 원인인 것으로 판단된다.

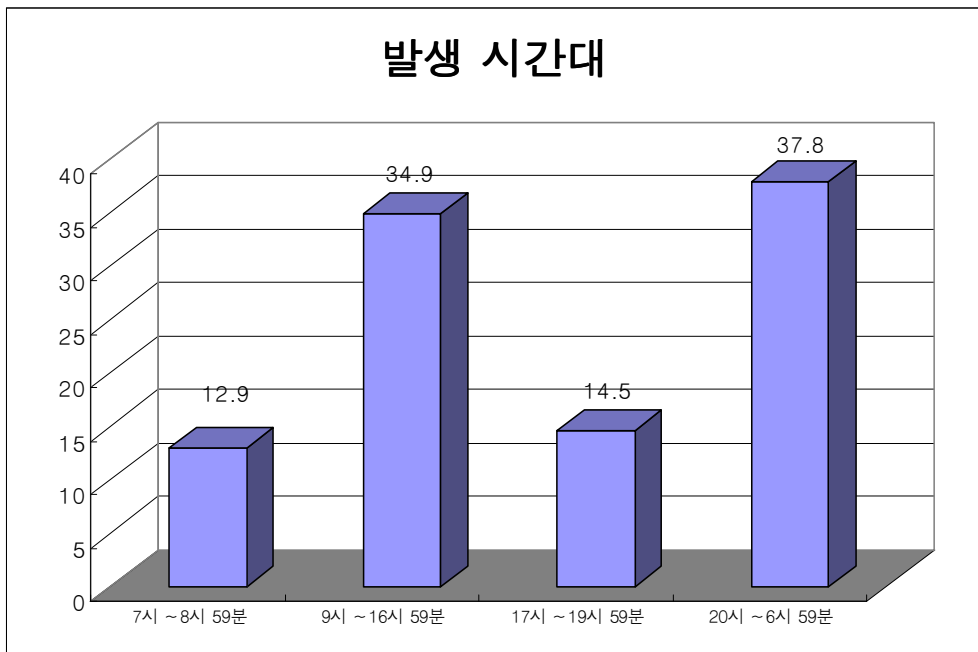
특히 이 중에서 중앙선 침범은 대향 차량과의 정면충돌 등을 야기하게 될 가능성이 높게 되고, 이로 인하여 운전자가 사망할 확률을 높이는 교통사고의 중대한 원인인 점을 감안한다면, 중앙선 침범은 다른 그 어떤 위반보다도 중요하게 다루어져야 할 대형 교통사고의 원인이라 볼 수 있을 것이다.

아울러 주취운전 역시 대향 차량의 운전자의 입장에서 보면, 자신의 의지와 전혀 관계없이 불의의 교통사고를 당하게 되고, 이로 인하여 본인의 목숨을 잃을 수도 있으므로, 이 역시 사회적 문제로서 엄벌을 처해야 할 중범죄로 다루어져야 할 대형 교통사고의 원인이라 생각된다.



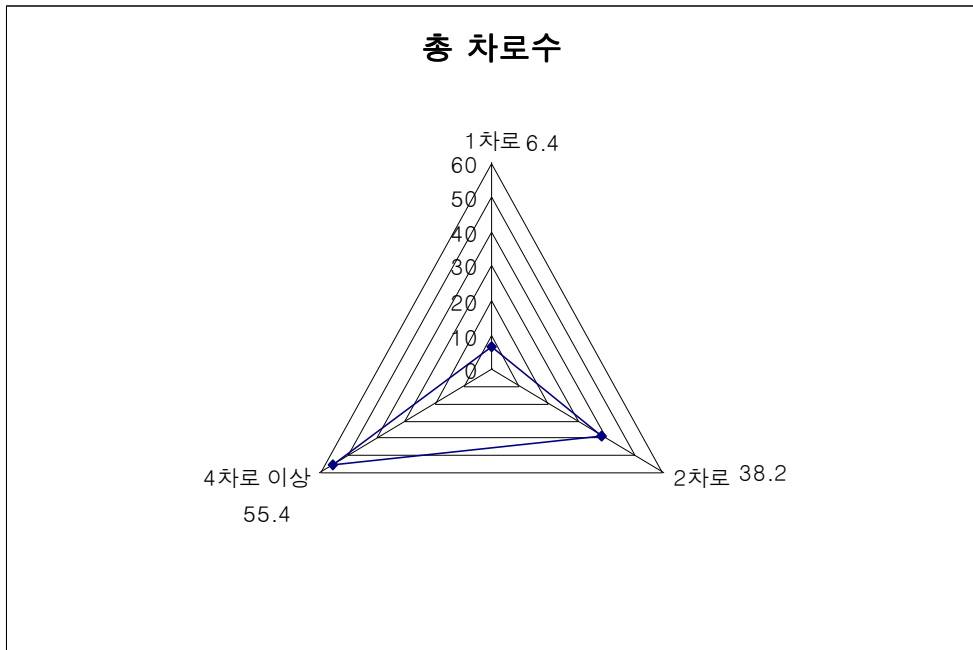
<그림 2-2> 법규위반 유형

다음으로 대형 교통사고의 발생 시간대별 특성을 <그림 2-3>에서 살펴보면, 운전 중 장애요인이 많아, 운전자의 입장에서 보면, 운전을 하기가 매우 어려운 시간대인 20시부터 익일 06시 59분까지의 야간 시간대가 37.8%인 94건으로 가장 많으며, 다음으로는 일상생활의 활동으로 인하여 자동차 사용이 빈번해지는 시간대인 09시부터 16시 59분까지가 34.9%인 87건으로 많이 나타나고 있는 등 대형 교통사고는 야간과 주간의 양 시간대에 걸쳐 골고루 많이 발생하는 것으로 나타났다.



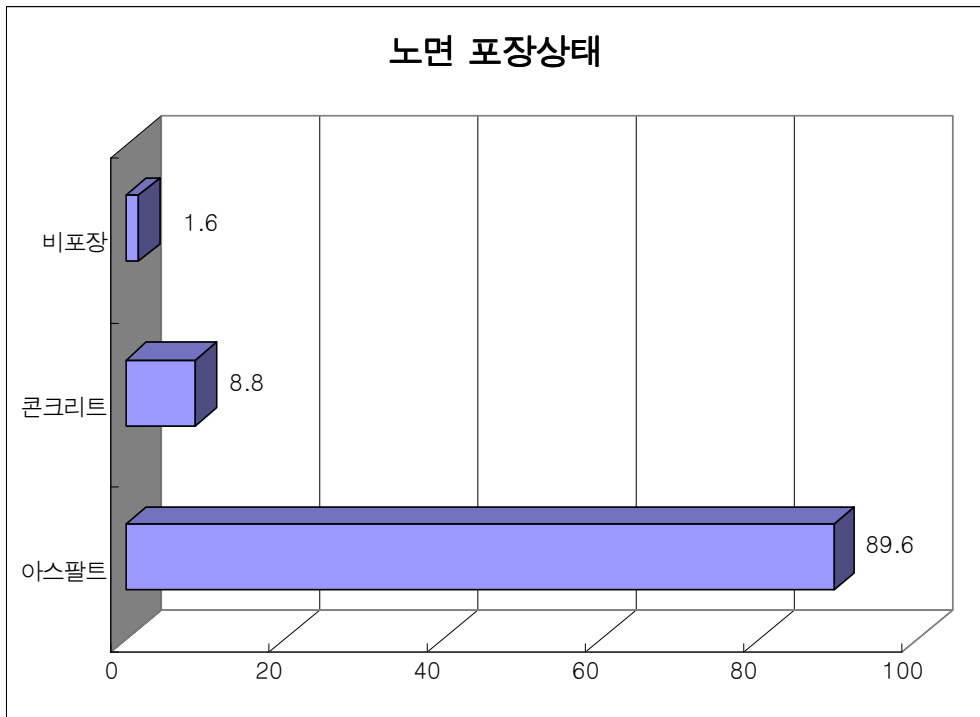
<그림 2-3> 발생 시간대

대형 교통사고가 발생한 도로의 총 차로수는 <그림 2-4>에서 보는 바와 같이, 도로의 조건이 가장 좋은 4차로 이상인 도로가 55.4%인 138건으로 가장 많으며, 다음으로는 2차로가 38.2%인 95건으로 나타나 도로의 조건이 좋을수록 주의력 집중력이 상대적으로 약화되기 쉬워 대형 교통사고가 많이 발생하는 것으로 나타났다.



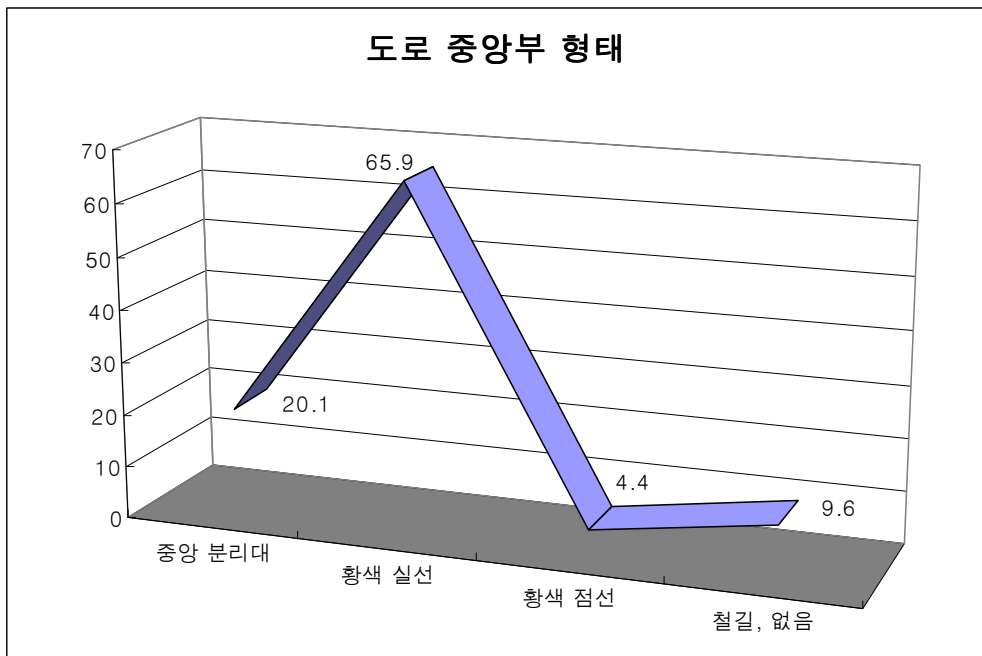
<그림 2-4> 총 차로수

대형 교통사고가 많이 발생하는 도로의 노면 포장상태는 <그림 2-5>에서 보는 바와 같이, 아스팔트 포장인 도로가 89.6%인 223건으로 대다수를 차지하고 있으며, 다음으로는 콘크리트 포장인 도로가 8.8%인 24건으로 나타나, 이 역시 조건이 좋은 도로에서 대형 교통사고가 많이 발생하는 것으로 나타났다.



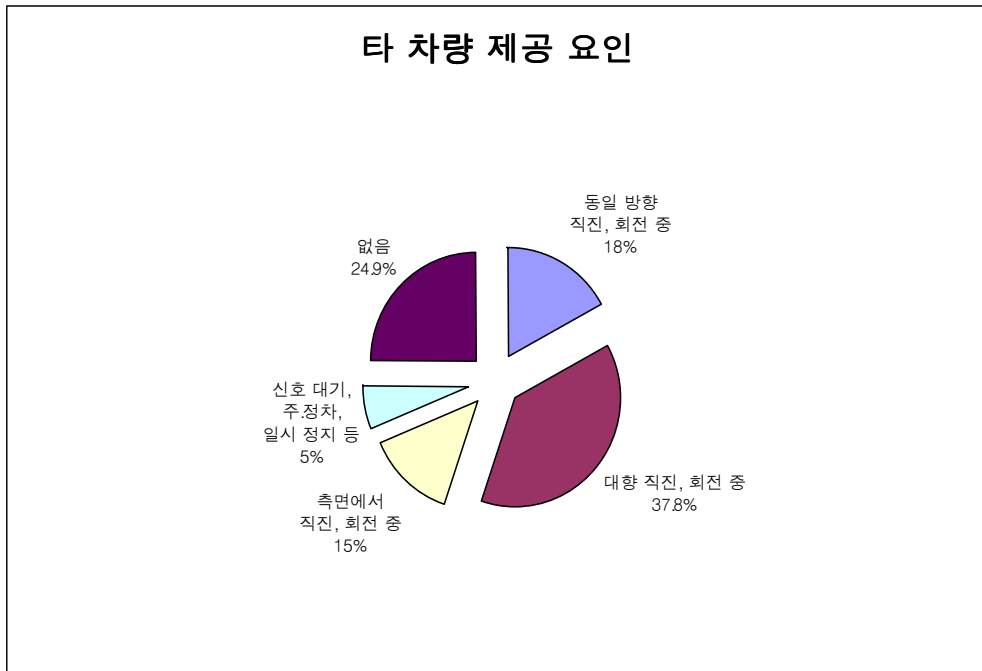
<그림 2-5> 노면 포장상태

대형 교통사고가 많이 발생하는 도로의 중앙부 형태는 <그림 2-6>에서 보는 바와 같이, 황색실선이 설치되어 있어 앞지르기가 금지되는 구간에서 대형 교통사고가 65.9%인 164건으로 가장 많이 발생하는 것으로 나타났으며, 다음으로는 중앙분리대가 설치된 도로에서 20.1%인 50건이 발생하는 등 비교적 대향 교통간의 흐름이 명확하게 분리가 잘 되는 도로 조건이 좋은 구간에서 중대한 법규 위반 행위를 하게 됨으로 인하여 대형 교통사고가 많이 발생하는 것으로 나타났다.



<그림 2-6> 도로 중앙부 형태

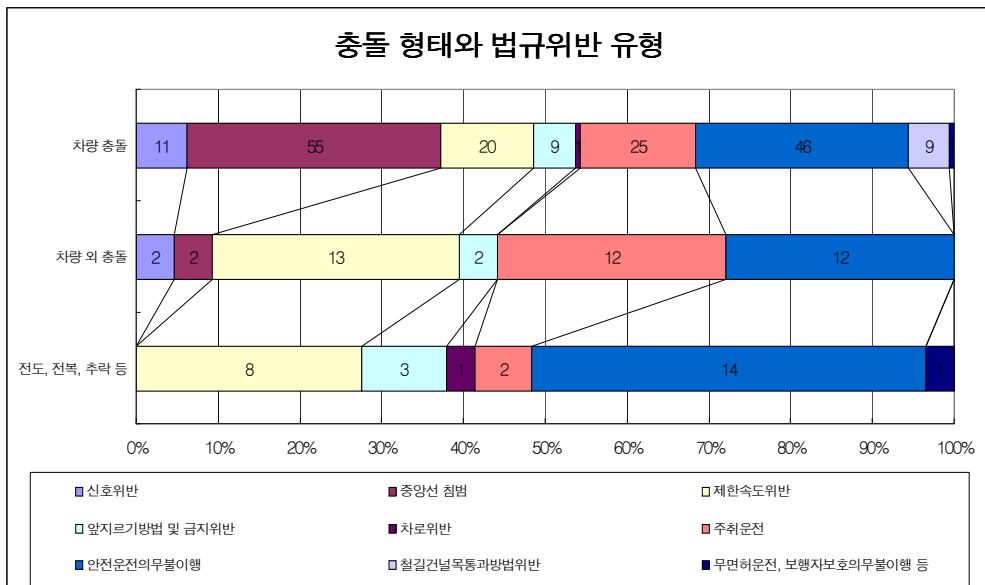
대형 교통사고 발생 당시 타 차량의 제공 요인으로는 <그림 2-7>에서 보는 바와 같이, 대향으로부터 직진 내지 회전 중인 경우가 37.8%인 94건으로 가장 많이 나타나, 이들 대향 차량과의 충돌이 피해를 크게 하여 대형 교통사고화 하는 주된 요인으로 작용하는 것으로 생각된다,



<그림 2-7> 타 차량 제공 요인

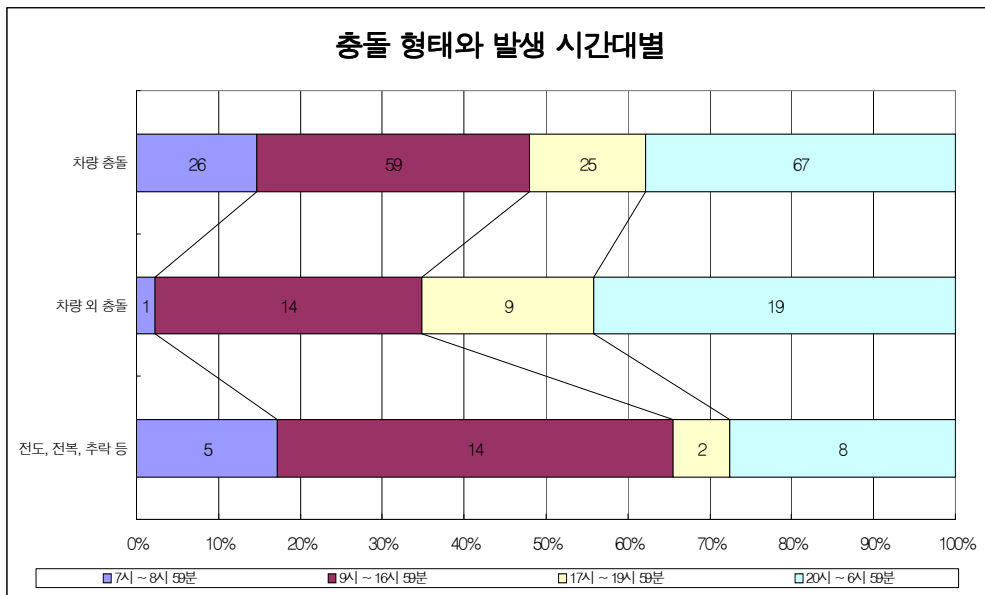
2. 교차분석에 따른 발생 특성

다음으로 대형 교통사고에 있어 충돌 형태와 법규위반 유형 간의 상호 관련성을 교차분석을 통해 살펴보면, <그림 2-8>에서 보는 바와 같다. 충돌 형태에 있어 가장 많이 발생한 차량 충돌의 경우는 대향 차량과의 충돌로 인하여 교통사고 결과가 대형화하기 쉬운 중앙선 침범이 55건으로 가장 많은 것으로 나타났으며, 차량 외 충돌의 경우는 속도의 제속에 비례하여 충격력이 커짐으로 인하여 충돌 시 피해가 큰 제한속도 위반이 13건으로 가장 많이 나타났으며, 전도, 전복, 추락 등의 경우는 안전운전의무불이행이 14건으로 가장 많이 나타났다.



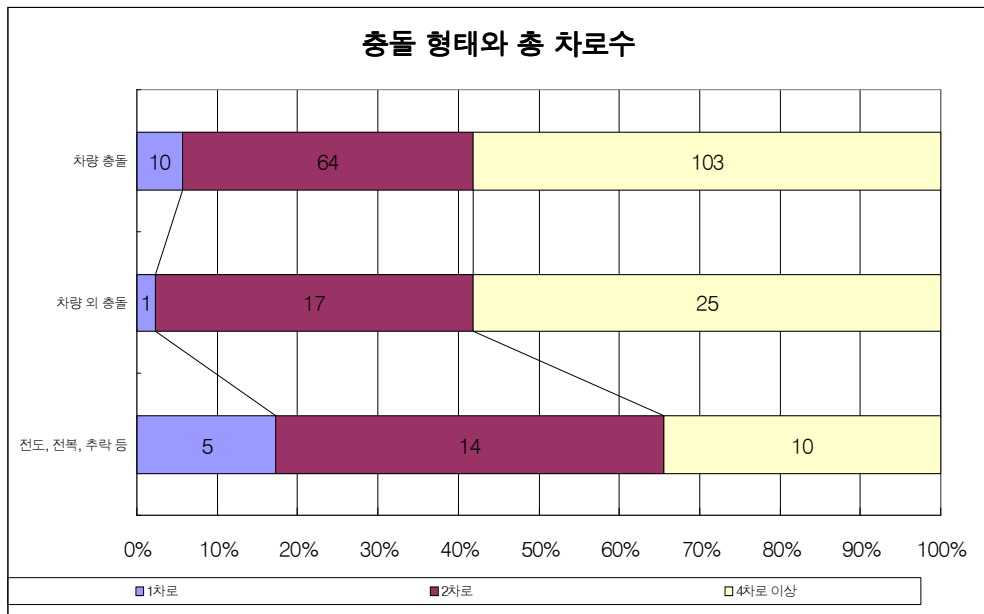
<그림 2-8> 충돌 형태와 법규위반 유형

다음으로 대형 교통사고에 있어 충돌 형태와 시간대별 간의 상호 관련성을 <그림 2-9>에서 보면, 운전 중 장애요인이 많아 운전자의 입장에서 보면 운전을 하기가 매우 어려운 시간대인 20시부터 익일 6시 59분까지의 야간 시간대가 67건과 차량 외 충돌의 경우 19건으로 가장 많은 것으로 나타났다. 반면에 일상 생활의 활동으로 인하여 자동차 사용이 빈번해지는 시간대인 9시부터 16시 59분 사이에는 장거리운전으로 인한 졸음운전 내지는 점심 식사시간 이후의 식곤증과 어울려져서 단독으로 발생할 가능성이 높은 전도, 전복, 추락 등의 교통사고가 14건으로 가장 많은 것으로 나타나는 등 명확한 대조를 이루고 있는 것으로 나타났다.



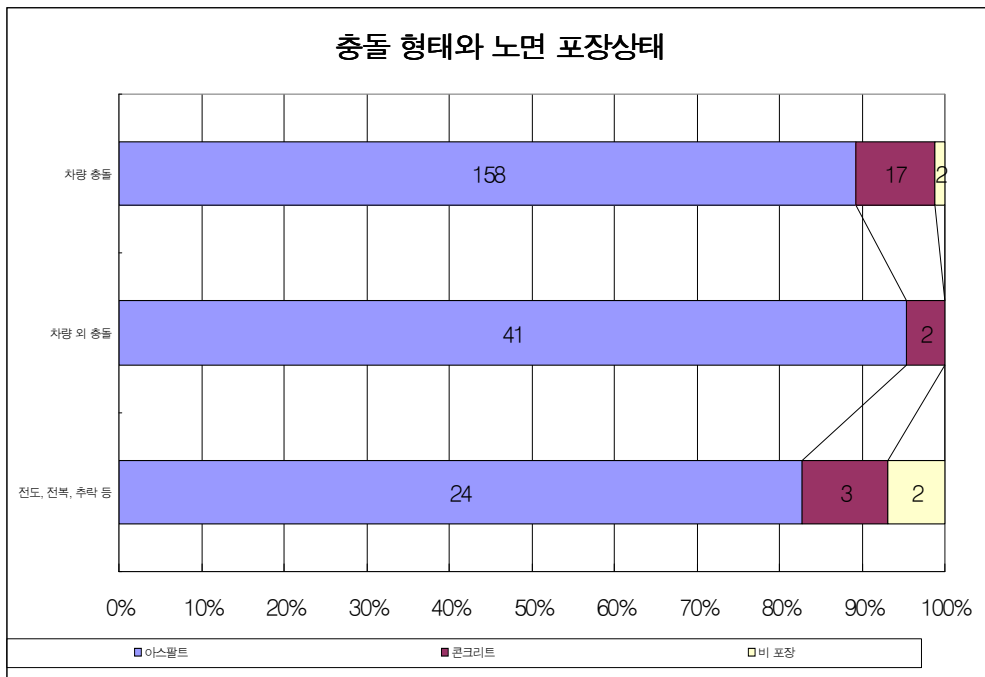
<그림 2-9> 충돌 형태와 발생 시간대별

그리고 대형 교통사고에 있어 충돌 형태와 총 차로수 간의 상호 관련성을 <그림 2-10>에서 보면, 차량 충돌의 경우와 차량 외 충돌의 경우는 도로의 조건이 매우 좋은 4차로 이상의 구간에서 103건과 25건으로 가장 많이 발생한 것으로 나타났으나, 전도, 전복, 추락 등의 경우는 상대적으로 도로 조건이 비교적 좋지 않은 2차로 구간에서 가장 많은 14건이 발생한 것으로 나타났다.



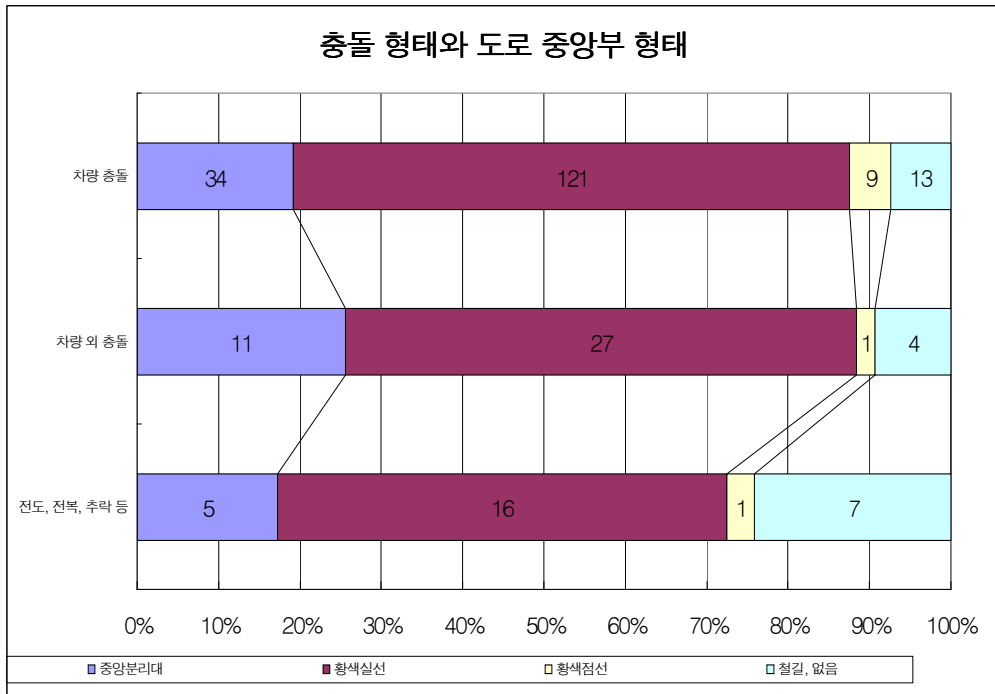
<그림 2-10> 충돌 형태와 총 차로수

아울러 대형 교통사고에 있어 충돌 형태와 노면 포장상태 간의 상호 관련성을 <그림 2-11>에서 보면, 차량 충돌의 경우와 차량 외 충돌의 경우 그리고 전도, 전복, 추락 등의 경우 모두 도로의 주행조건이 아주 좋은 아스팔트 구간 위에서 대형 교통사고가 각각 158건, 41건, 24건으로 가장 많이 발생한 것으로 나타났다.



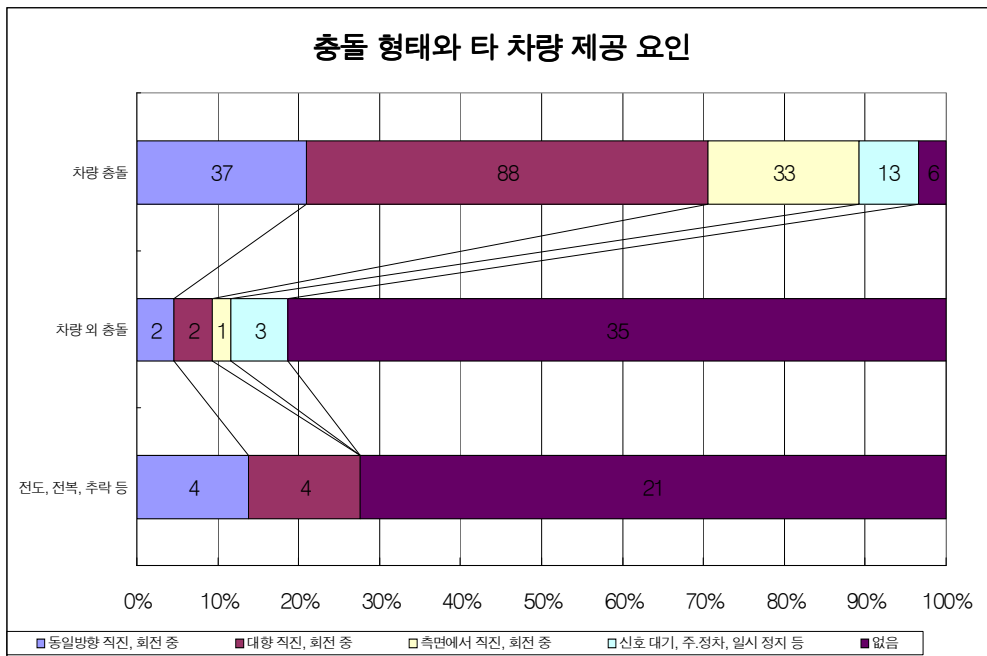
<그림 2-11> 충돌 형태와 노면 포장상태

또한 대형 교통사고에 있어 충돌 형태와 도로 중앙부 형태 간의 상호 관련성을 <그림 2-12>에서 보면, 차량 충돌의 경우와 차량 외 충돌의 경우 그리고 전도, 전복, 추락 등의 경우 모두 앞지르기가 금지되는 황색실선이 설치된 구간에서 대형 교통사고가 각각 113건, 24건, 14건으로 가장 많이 발생한 것으로 나타났다.



<그림 2-12> 충돌 형태와 도로 중앙부 형태

하지만 대형 교통사고에 있어 충돌 형태와 타 차량 제공 요인 간의 상호 관련성을 <그림 2-13>에서 보면, 차량 충돌의 경우는 대향으로부터의 직진 내지는 회전 중인 경우가 88건으로 가장 많이 나타났으나, 차량 외 충돌의 경우와 전도, 전복, 추락 등의 경우는 교통사고의 발생 특성 상 타 차량 제공 요인이 없는 경우가 각각 36건과 21건으로 가장 많이 발생한 것으로 나타났다.



<그림 2-13> 충돌 형태와 타 차량 제공 요인

제3장 발생지점 유형화와 영향요인 판별, 분류모델의 이론적 배경

제1절 군집분석²²⁾²³⁾²⁴⁾의 이론적 배경

군집분석은 서로 섞여 있는 이질적인 대상(개체 또는 변수)을 유사도(Measure of Similarity)에 의해 몇 개의 균질적인 집단으로 분류하여 데이터 전체의 구조에 대한 이해를 돕기 위한 분석방법으로, 개체들을 유사도가 높은 개체들끼리 또는 거리가 가까운 개체들끼리 하나의 군집으로 병합해 나가는 방법이다.

군집의 유형으로는 계통적 군집법(Hierarchical Clustering Method)과 상호배타적 군집법(Disjoint Clustering Method)으로 구분되며, 본 연구에서는 전자인 계통적 군집법을 이용하였다.

그리고 군집화 방법으로는 최단연결법(Single Linkage Method), 최장연결법(Complete Linkage Method), 군집 내 평균 연결법(Within Average Linkage Method), 군집 간 평균 연결법(Between Average Linkage Method), 중심연결법(Centroid Linkage Method), 와드법(Ward's Method) 등이 사용되고 있으나, 본 연구에서는 교통 분야에서 실용적으로 많이 쓰이고 있는 와드법을 이용하였다.

특히 와드법은 다음의 식(1)과 같이 N_i 개의 개체들로 구성된 군집 C_i 의 중심점을 군집의 평균값으로 사용하여 중심점과 각 개체들의 차이를 제곱함으로써 산출하여 군집 간 유사도 또는 거리로 이용한다.

$$E = \sum_{i=1}^g \sum_{j=1}^p \sum_{k=1}^{n_i} (X_{ijk} - \bar{X}_{ij})^2 \text{-----} (1)$$

초기에는 N 개의 개체들이 각각의 군집을 이루므로 $E = 0$ 이 된다. 단계

22). 상계서, pp.341~354.

23). 김종섭, SAS v.8.2를 이용한 통계분석의 이해, 한울출판사, 2003, pp.497~509.

24). Claude Seidman 저, 권오주, 이정무 역, 전계서, pp.203~235.

별로 군집을 만들어 가는 과정에서 모든 가능한 병합(Agglomeration) 중 두 군집의 병합으로 인한 E 값의 증가가 최소가 되도록 군집을 병합한다.

이 때 크기가 N_1 과 N_2 인 두 군집 C_1 과 C_2 를 묶을 때 생기는 E 값의 증분은 다음의 식(2)와 같다.

$$E(C_1, C_2) = \frac{n_1 n_2}{n_1 + n_2} d^2(\overline{X_1}, \overline{X_2}) \text{-----} (2)$$

여기서 $E(C_1, C_2)$ 를 두 군집 C_1 과 C_2 의 거리로 정하여 집단 내 거리의 제곱의 합이 최소화하는 방식으로 군집화를 실시하게 된다.

제2절 판별모델의 이론적 배경

수량화 모델은 Data의 성격이 주로 정성적이어서, 정량적 형태의 Data를 적용하는 일반적인 통계 분석방법을 적용토록 함으로서, 수치로 나타낼 수 없는 질적인 것에 수량을 부여하여 목적변수에의 정성적 영향요인을 판별하기 위한 방법으로, 이러한 정성적 Data들에 대해서는 요인별 분석을 가능케 한다.

본 연구에서는 수량화 이론 I ~ IV류²⁵⁾²⁶⁾ 중 질적 요인에 의한 “질적 외적 기준”을 예측 혹은 판별하기 위한 방법으로서, 정성적 속성의 각 Category에 적당한 수치를 부여하여 정량적인 변수처럼 다변량 분석을 적용하는 독창적인 이론으로서, Dummy 변수법²⁷⁾과 유사한 수량화 이론 II류를 적용하여 분석하고자 하였다.

이러한 수량화 이론 II류는 R 개의 정성적 속성에 관한 지식을 이용하여, 각각의 개체가 T 군의 어디에 속하는 것인가를 판별하는 판별문제를 해석하는 모델이다.

즉, 정성적 설명변수인 Item의 Category로부터 외적 기준의 분류를 실시하는 수법으로, 일반적으로 어떤 무한 모집단에서 추출한 Data에 대하여 외적기준을 가장 잘 설명할 수 있도록 Category에 득점을 부여한다. 이때의 설명변수는 상대비이고, 그 득점은 Category Score이며, 편상관계수의 범위(Range)가 클수록, 외적 기준에 미치는 영향이 크다고 판단할 수 있다.

그리고 표본수가 적은 경우에도 Item의 Category에 반응하면 분석이 가능해지고, 외적 기준과 설명변수의 인과관계를 정량화 할 수 있다는 점에서 대단히 편리하다.

각 개체가 R 개 Category의 kj 개 선택치 중에서 어느 하나에 반응한다고 하는 상황을 가정할 때, i 번째의 개체가 j 번째의 속성에 관하여 k 번째의

25). 田中 豊, 垂水共之 編, 전게서, pp.152~186.

26). 허명희, 전게서. pp.1.1~4.16.

27). 채서일, 김범중, 전게서, p.42, 85.

Category에 반응(회답)했을 때만을 1, 기타 $kj - 1$ 개의 Category에 반응을 했을 때에는 0을 취하는 $\delta_i(jk)$ 가 되는 양을 도입하면,

$$1 : \text{Category } k \text{에 반응(회답)했을 때} \\ \delta_i(jk) = 1 \text{ -----} (3)$$

0 : 이 외의 Category에 반응(회답)했을 때

이 $\delta_i(jk)$ 에 관해서는 다음과 같은 관계식이 성립한다.

$$\sum_{k=1}^{kj} \delta_i(jk) = 1 \text{ -----} (4)$$

$$\sum_{i=1}^n \delta_i(jk) = n_{jk} \text{ -----} (5)$$

$$\sum_{k=1}^{kj} \sum_{i=1}^n \delta_i(jk) = n \text{ -----} (6)$$

또한 개체 i 가 j 속성의 k Category에 반응했다고 하면,

$$0 \quad (k \neq k') \\ \delta_i(jk)\delta_i(jk') = \text{-----} (7) \\ 1 \quad (k = k')$$

지금 R 개의 각각의 속성을 kj 개의 각각의 Category에 대하여 $X_{jk} = (j = 1, \dots, k, k = 1, \dots, kj)$ 가 되는 수치를 부여 했을 때, 개체 i 에 대한 새로운 합성변수를 다음과 같이 정의한다.

$$a_i = \sum_{j=1}^R \sum_{k=1}^{kj} \delta_i(jk) X_{jk} \text{ -----} (8)$$

여기에서 a_i : 합성변량

$\delta_i(jk)$: Dummy변수

X_{jk} : 고유Vector

즉, R 개의 정성적 속성에 관한 지식을 사용하여 각각의 개체가 T 개 군의 어딘가에 속하고 있는가를 판별하는 판별문제를 해석하는 Model이다.

만약 (8)식의 X_{jk} 치의 부여방법이 완벽에 가깝게 되면, T 개 군의 구분을 가로축에, a 를 세로축에 두면, T 와 a 의 상관비 η 혹은 η^2 은 1에 가까운 수치를 취한다. 따라서 이때에는 급간 분산을 전 분산으로 나눈 값인 $\eta^2 = \delta_b^2 / \delta^2$ 이 되는 η^2 이 최대가 되도록 X_{jk} 의 수치를 결정하게 된다.

따라서 식(8)의 경우,

$$\delta^2 = \frac{1}{n} \sum_{i=1}^n a_i^2 \bar{a}^2 \text{-----} (9)$$

$$\delta_b^2 = \sum_{t=1}^T \frac{n_t}{n} (\bar{a}_t - \bar{a})^2 \text{-----} (10)$$

$$\bar{a} = \frac{1}{n} \sum_{i=1}^n a_i \text{-----} (11)$$

$$\bar{a}_t = \frac{1}{n} \sum_{i(t)=1}^{n_t} a_i(t) \text{-----} (12)$$

로 된다.

여기에서 η^2 을 최대로 하는 X_{jk} 를 구하기 위해서는 X_{uv} 로 편미분하여 각각을 0으로 두면 된다.

즉,

$$\frac{\sigma \eta^2}{\sigma X_{uv}} = 0 \text{ (} U = 1, \dots, R, V = 1, \dots, ku \text{) -----} (13)$$

$$\frac{\delta^2 \frac{\sigma \delta_b^2}{\sigma X_{uv}} - \delta_b^2 \frac{\sigma \delta^2}{\sigma X_{uv}}}{(\delta^2)^2} = 0 \text{-----} (14)$$

$$\frac{\sigma\delta_b^2}{\sigma X_{uv}} = \eta^2 \frac{\sigma\delta^2}{\sigma X_{uv}} \quad (U = 1, \dots, R, \quad V = 1, \dots, ku) \text{ ----- (15)}$$

여기서

$$\frac{\sigma\delta_b^2}{\sigma X_{uv}} = \frac{2}{n} \sum_j \sum_k \left[\sum_t \frac{g^t(jk)g^t(uv)}{n_t} - \frac{1}{n} n_{jk} n_{uv} \right] X_{jk} \text{ -- (16)}$$

$$\frac{\sigma\delta_b^2}{\sigma X_{uv}} = \frac{2}{n} \sum_j \sum_k \left[\sum_t f_{jk}(uv) - \frac{1}{n} n_{jk} n_{uv} \right] X_{jk} \text{ ----- (17)}$$

따라서 식(15)은

$$\sum_{j=1}^R \sum_{k=1}^{kj} h_{uv}(jk) X_{jk} = \eta^2 \sum_{j=1}^R \sum_{k=1}^{kj} \left[f_{uv}(lm) - \frac{1}{n} n_{lm} n_{uv} \right] X_{lm}$$

단,

$$h_{uv}(jk) = \sum_{t=1}^T \frac{g^t(jk)g^t(uv)}{n_t} - \frac{1}{\eta} n_{jk} n_{uv} \text{ ----- (18)}$$

식(18)을 행렬식으로 나타내면

$$HX = \eta^2 FX \text{ ----- (19)}$$

여기서 $H = [k_{uv}(jk)]$: 정방행렬

$$F = \left[f_{uv}(lm) - \frac{1}{n} \eta_{jk} \eta_{uv} \right] : \text{정방행렬}$$

식(17)을 변형하면

$$F^{-1}HX = \eta^2 X \text{ ----- (20)}$$

$$|F^{-1}H - \eta^2 I| = 0 \text{ ----- (21)}$$

여기서 η^2 을 최대로 하는 고유벡터를 구하면 된다.

제3절 분류모델의 이론적 배경

분류모델의 대표적인 분석방법인 의사결정나무분석법²⁸⁾은 의사결정규칙(decision rule)을 도표화하여, 관심대상이 되는 집단을 몇 개의 소집단으로 분류(classification)하거나 예측(prediction)을 수행하는 분석방법이다. 특히 분석과정이 나무구조에 의해서 표현되기 때문에, 분류 또는 예측을 목적으로 하는 다른 방법들인 신경망(Neural Networks), 판별분석(Discriminant Analysis), 회귀분석(Regression Analysis) 등에 비해, 분석 과정을 쉽게 이해하고 설명할 수 있다는 장점을 가지고 있다.

그리고 의사결정나무분석은 탐색과 모형화라는 두 가지 특성을 모두 가지고 있는 바, 탐색단계에서는 차원의 축소 및 변수 선택, 교호효과의 파악, 범주의 병합 또는 연속형 변수의 이산화 등이 포함되고, 모형화 단계에서는 세분화, 분류, 예측 등이 포함되어 있다. 즉, 의사결정나무는 판별분석 또는 회귀분석 등과 같은 모수적(parametric) 모형을 분석하기 위해서 사전에 이상치(outlier)를 검색하거나, 분석에 필요한 변수를 찾아내고, 모형에 포함되어야 할 교호효과를 찾아내는 데 사용될 수도 있고, 그 자체가 분류 또는 예측 모형으로 활용될 수 있다.

의사결정나무는 하나의 나무구조를 이루고 있고, 그 기능에 따라, 뿌리마디(root node), 자식마디(child node), 부모마디(parent node), 끝마디(terminal node) 중간마디(internal node) 그리고 가지(branch) 등으로 분류되는, 마디(node)라고 불리는 구성요소들로 이루어져 있다.

그리고 의사결정나무는 뿌리마디로 시작하여, 각 가지가 끝마디에 이를 때까지 자식마디를 계속적으로 형성해 나감으로써 완성되며, 이를 형성하는 알고리즘으로는 CHAID, C&RT, QUEST 등이 있다.

이들은, 하나의 부모마디로부터 자식마디들이 형성될 때, 어떤 예측변수를 이용하여, 어떻게 분리하는 것이 목표변수의 분포를 가장 잘 구별해 주는지에 대한 정도를 나타내는 분리기준(splitting criterion)에서 서로 다른 것을 이용한다. 또한 더 이상 분리가 일어나지 않고 현재의 마디가 끝마디가 되도록 하는 정지규칙(stopping rule) 그리고 지나치게 많은 마디를 가지는 의사결정나무가 새로운

28). 최종후, 한상태, 강현철, 김은석 공저, 전개서, pp.17~26.

자료에 적용할 때, 예측오차(prediction error)가 매우 크지는 것을 방지하기 위하여, 형성된 의사결정나무에서 적절하지 않은 마디를 제거하여, 적당한 크기의 부나무(subtree) 구조를 가지도록 하는 가지치기(pruning) 등에서도 서로 다른 형성과정을 가지고 있다.

이 중에서 본 연구에 이용된 C&RT 분석법²⁹⁾은 이산형 목표변수인 경우 적용하는 지니 지수(Gini Index), 연속형 목표변수인 경우 분산의 감소량을 이용하여 부모 마디로부터 자식 마디가 2개만 형성되게 하는 이진분리(binary split)를 수행하는 알고리즘이다.

1. 이산형 목표변수에 대한 분리 기준

(1). 지니 지수

지니 지수는 불순도(impurity)를 측정하는 지수이다.

먼저 각 마디에 속하는 개체를 그 마디에서 도수가 가장 많은 목표변수의 한 범주에만 모두 할당하는 분류 규칙을 고려해 보고자 한다.

임의의 한 개체가 목표변수의 j 번째 범주로부터 추출되었고, 그 개체를 목표변수의 j 번째 범주에 속한다고 오분류(misclassification)할 확률은 $R(j)R(j)$ 가 된다. 여기에서 $R(j)$ 는 각 마디에서 한 개체가 목표변수의 j 번째 범주에 속할 확률이다. 이러한 오분류할 확률을 모두 더하면

$$G = \sum_{j=1}^c \sum_{\#j} R(j)R(j) \text{-----} (22)$$

를 얻을 수 있고, 이는 이와 같은 분류규칙 하에서 오분류 확률의 추정치(estimate)가 된다. 여기서 c 는 목표변수의 범주수를 말한다.

일반적으로 C&RT는 이산형 목표변수에 대해서는 지니 지수를 분리 기준으로 사용한다. 지니 지수는 각 마디에서 불순도(impurity) 또는 다양도(diversity)를 재는 측도 중의 하나로써

$$G = \sum_{j=1}^c R(j)(1-R(j)) = 1 - \sum_{j=1}^c R(j)^2 = 1 - \sum_{j=1}^c (n_j/n)^2 \text{-----} (23)$$

29). 상계서, pp.42~48.

와 같이 표현될 수 있다. 여기에서 n 은 그 마디에 포함되어 있는 관찰치 수를 말하고, n_i 는 목표변수의 i 번째 범주에 속하는 관찰치 수를 말한다. 지니 지수는 n 개의 원소 중에서 임의로 2개를 추출하였을 때, 추출된 2개가 서로 다른 그룹에 속해 있을 확률을 의미하며 Simpson의 다양도 지수(diversity index)로도 알려져 있다. 목표변수의 범주가 2개인 경우에는 지니 지수는 다음과 같이 표현될 수 있다.

$$G=2P(1)P(2)=2\left(\frac{n_1}{n}\right)\left(\frac{n_2}{n}\right) \text{-----}(24)$$

이는 카이제곱 통계량을 사용하는 것과 같은 결과를 갖는다.

C&RT는 이 지니 지수를 가장 많이 감소시켜주는 예측변수와 그 변수의 최적 분리를 자식마디로 선택하는데, 지니 계수의 감소량은 다음과 같이 계산된다.

$$\Delta G=G-\frac{n_L}{n} G_L-\frac{n_R}{n} G_R \text{-----}(25)$$

여기서 n 은 부모 마디의 관측치 수를 말하고, n_R 과 n_L 은 각각 자식 마디의 관측치 수를 의미한다. 즉, 자식 마디로 분리되었을 때의 불순도가 가장 작도록, 다시 말하자면 순수도가 가장 크도록 자식 마디를 형성하는 것이며, 이는 자식 마디에서의 불순도의 가중 합을 최소화하는 것과 동일하다.

(2). Twoing

자주 사용되지는 않으나 C&RT의 분류 기준으로 사용되는 것으로 Twoing 기준이 있다. 일반적으로 지니 지수와 Twoing 기준은 유사한 나무구조를 형성하는 경우가 많으나, 목표변수의 범주 수가 큰 경우에는 Twoing 기준을 사용하면 계산 시간이 많이 걸리기 때문에 지니 지수가 더 선호되는 경향이 있다.

여기서 Twoing과 Ordered Twoing의 계산 과정을 보면, 먼저 각 마디에서 목표변수의 범주 집합을 $C=\{1, 2, \dots, c\}$ 이라 할 때, 이를 임의의 두 집합 C_1 과 $C_2=C-C_1$ 로 그룹화 하는 경우를 생각해 보자.

이러한 임의의 그룹화에 대해서 목표변수의 범주 수를 2라고 가정한 후

$R(1)R(2)$ 의 감소량을 계산하여 이를 $\Delta Q(C_1)$ 이라고 표현해 보자.

이 때 모든 가능한 C_1 에 대해서 $\Delta Q(C_1)$ 을 최대화하는 것은 결국 다음과 같은 Twoing 지수 T 를 최대화하는 것과 같게 된다.

$$T = \frac{P_L P_R}{4} \left[\sum_{j=1}^c |P_L(j) - P_R(j)| \right]^2 = \frac{1}{4} \left(\frac{n_L}{n} \right) \left(\frac{n_R}{n} \right) \left[\sum_{j=1}^c \frac{n_{jL}}{n_L} - \frac{n_{jR}}{n_R} \right]^2 \quad (26)$$

목표변수가 순서형일 때는 C_1 을 선택할 때 순서적으로 인접한 목표변수의 범주만이 C_1 으로 그룹화 되도록 제한을 둘 수 있는데, 이렇게 하여 계산된 Twoing 기준의 값을 Ordered Twoing 기준이라고 한다.

(3). 다른 분리 기준들

지니 지수와 유사한 것으로써 여러 소프트웨어나 알고리즘에서 많이 사용되고 있는 분리 기준으로

$$Mis(R(1), R(2)) = \sum_{j=1}^c R(j) \log_2 R(j) \quad (27)$$

와 같은 것들이 있다. 특히 $-\sum R(j) \log_2 R(j)$ 는 엔트로피(entropy) 지수라고 불리우며 다항분포(multinomial distribution)에서의 우도비 검정 통계량(likelihood ratio test statistics)을 사용하는 것과 같은 것으로 알려져 있다. 비교적 최근에 널리 알려진 알고리즘인 C4.5는 엔탈로피 지수를 분리 기준으로 사용하고 있다.

2. 연속형 목표변수에 대한 분리 기준

목표변수가 연속형일 때는 각 마디의 다양도의 측도로써 다음과 같은 분산(variance)을 고려할 수 있다.

$$V = \frac{1}{n} \sum_{i=1}^n (y_i - \bar{y})^2 \quad (28)$$

이는 마디 목표변수의 평균을 그 마디에 속하는 모든 개체의 예측값으로 사용하는 것을 고려한다면, 예측 오차를 최소화하는 것과 동일하다 할 수 있다.

명목형 목표변수인 경우와 마찬가지로 분산의 감소량을 최대화하는 것은 다음

과 같은 기준

$$\Delta V = V - \frac{n_L}{n} V_L - \frac{n_R}{n} V_R \text{-----} (29)$$

를 최대화하는 것과 같다. 여기서 n 은 부모 마디의 관측치 수를 의미하며, n_R 과 n_L 은 각각 자식 마디의 관측치 수를 의미한다. 이는 자식 마디에서의 집단 내 분산(within variance) $R(L) V_L + R(R) V_R$ 을 최소화하는 것과 동일하다.

나무구조 하에서의 위험(risk)과 위험의 표준오차의 경우 위험은 자식 마디에서의 집단 내 분산의 가중 합이 된다. 이와 같이 불순도를 분산으로 측정하는 경우에 있어서 예측 위험은 나무구조의 모든 끝 마디에서 집단 내 분산들의 가중 합을 계산한 것과 같다.

3. 가지치기(Pruning)

의사결정나무를 얻은 후에 사용자가 직접 가지치기를 수행할 수도 있지만, C&RT 알고리즘에서는 가지치기 기준을 사전에 지정을 할 수 있다.

(1). 표준오차(Standard Error)

표준오차 규칙은 T 번째 마디에서의 위험을 $R(T)$ 라 하고, $R(T)$ 의 표준오차를 $SE(R(T))$ 라 할 때, T_k 번째까지의 나무구조가 T_0 번째까지의 마디를 모두 포함하고 있다고 가정하면,

$$R(T_k) \leq R(T_0) + \alpha SE(R(T_0)) \text{-----} (30)$$

인 경우에 있어 T_k 번째까지의 나무구조를 선택하는 것을 의미한다. 즉, $\alpha SE(R(T_0))$ 정도의 위험 감소량은 유의하지 않은 것으로 생각하여 더 이상 나무구조를 형성하지 않게 된다는 것을 의미한다.

(2). 최소 위험(Minimum Risk)

최소 위험을 가지는 나무구조를 의사결정나무로 선택한다.

4. 결손 값에 대한 대체규칙(Surrogate Rule)

C&RT에서는 각 마디에서 선택된 예측변수에 결손 값이 존재

하는 경우에 다른 예측변수를 이용하여 결손 값을 대체할 수 있다. 결손 값에 대한 대체규칙을 지정해 놓으면 의사결정나무의 형성과정 또는 새로운 자료의 분류과정에서 이 규칙에 따라서 각 예측변수의 결손 값을 대체할 수 있다.

대체규칙을 선택할 때, 선택된 변수 대신에 다른 변수를 선택하였을 경우 분류 또는 예측이 일치하는 정도를 의미하는 연관성(association)이 큰 순서대로 다른 예측변수로 대체하면 그 변수가 가지는 값에 따라 각 개체를 자식 마디에 할당하였을 때의 분류가 어느 정도가 일치하는지를 나타내어 줄 수 있다.

5. 의사결정나무의 장점 및 단점³⁰⁾

(1). 의사결정나무의 장점

1). 해석의 용이성

나무구조에 의해 모형이 표현되기 때문에 모형을 사용자가 쉽게 이해 할 수 있다. 또한 새로운 개체에 대한 판별 또는 예측을 하기 위해서 뿌리마디로부터 끝까지 단순하게 따라가면 되기 때문에, 새로운 자료에 모형을 적합 시키기 매우 쉽다. 그리고 나무구조로부터 어떤 예측변수가 목표 변수를 설명하기 위해서 더 중요한지를 쉽게 파악할 수 있는 장점이 있다.

2). 교호효과(interaction)의 해석

의사결정나무는 두 개 이상의 변수가 결합하여 목표변수에 어떻게 영향을 주는지 쉽게 알 수 있다. 특히 회귀분석이나 로지스틱 판별분석(logistic discriminant analysis)과 같은 모수적 모형에서는 모든 가능한 교호효과를 모형에 포함시켜 분석한다는 것은 거의 불가능하지만, 의사결정나무는 이것이 가능하므로, 유용한 예측변수나 비선형성(nonlinearity)을 자동적으로 찾아준다.

3). 비모수적 모형(nonparametric model)

의사결정나무는 선형성(linearity)이나 정규성(normality) 또는 등분산성(equal variance) 등의 가정을 필요로 하지 않는 비모수적 방법이다. 또한 의사결정나무에서는 변수가 순서형 또는 연속형일지라도 단지 순위(rank)만 분석에 영향을 주기 때문에, 이상치에 민감하지 않는다는 장

30). 상계서, pp.26~28.

점을 가지고 있다.

(2). 의사결정나무의 단점

1). 비연속성

의사결정나무에서는 변수가 연속형이라도 비연속적인 값으로 취급하기 때문에, 분리의 경계점 근방에서는 예측오류가 클 가능성이 있다. 따라서 최근에는 이러한 단점을 극복하기 위하여, 모수적 모형이나 신경망 등을 의사결정나무와 결합하는 경향이 강하게 대두되고 있다.

2). 선형성 또는 주 효과의 결여

통상의 경우 선형모형(linear model)에서는 주 효과(main effect)가 다른 예측변수와 관련시키지 않고서도 각 변수의 영향력을 해석할 수 있다라는 장점을 가지고 있지만, 의사결정나무에서는 선형(linear) 또는 주 효과 모형에서와 같은 결과를 얻을 수 없다라는 한계점이 있다.

3). 불안정성

분석용 자료(training data)에만 의존하는 의사결정나무는 새로운 자료의 예측에서는 불안정(unstable)할 가능성이 높다. 이와 같은 현상은 분석용 자료의 크기가 너무 작은 경우와 너무 많은 가치를 가지는 의사결정나무를 얻는 경우에 빈번히 발생된다. 따라서 검증용 자료(testing data)에 의한 교차 타당성(cross validation) 평가나 가지치기에 의해서 안정성 있는 의사결정나무를 얻는 것이 바람직하다고 여겨진다.

제4장 발생지점 유형화와 영향요인 판별, 분류모델 구축

제1절 대형 교통사고 충돌 형태별 영향요인 판별, 분류모델

1. 판별모델

먼저 분석 대상 대형 교통사고 자료 249건 전체에 대하여 충돌 형태별 영향요인 판별모델 구축을 위한 수량화 이론 II류 분석을 행하였다. 이를 위하여, 외적 변수(Outside Variable)로 충돌 형태를 <표 4-1>과 같이, 차량 충돌, 차량 외 충돌, 전도, 전복, 추락 등의 3가지 유형으로 구분하여 채택하였다. 충돌 형태 유형의 빈도분석 결과 순위를 보면, 차량 충돌, 차량 외 충돌, 전도, 전복, 추락 등의 순으로 나타났다.

<표 4-1> 충돌 형태별 현황(전체)

충돌 형태	사고 건수(건)	비율(%)
차량 충돌	177	71.1
차량 외 충돌	43	17.3
전도, 전복, 추락 등	29	11.6
계	249	100.0

그리고 설명변수 항목(Item)으로는 범규위반 유형, 사고 발생 시간대별, 총 차로수, 노면 포장 상태, 도로 중앙부 형태, 타 차량 제동 요인 등 6가지가 선택되었으며, 이들 항목과 이에 따른 유형(Category)들을 설정하여 충돌 형태별 유형에 대해 수량화 이론 II류 분석을 행하였다. 그 결과 구축된 판별모델은 판별식에 의해서만 외적 변수를 판별할 경우에는 각각의 범위가 중복되어 사례들을 명확한 판별을 할 수가 없어, 외적 변수의 점수(Score)를 이용하여 두 집단을 구분 짓는 경계를 산정하는 경계점으로 외적 변수의 범위를 설정하여 사례들이 정확하게 분류되는지를 살펴보았다. 그 결과 설정된 판별 범위는 <표 4-2>와 <그림 4-1>에서와 같이 나타났다.

<표 4-2> 충돌 형태별 외적 변수의 판별범위(전체)

충돌 형태	판별 범위
차량 충돌	$-0.573 < (1) < -0.020$
차량 외 충돌	$-0.357 < (2) < 1.570$
전도, 전복, 추락 등	$0.651 < (3) < 1.171$



<그림 4-1> 층돌 형태별 외적 변수의 Data Score Histogram(전체)

또한 이러한 판별 범위 내에서 해당 집단을 해당 집단으로 올바르게 판별할 수 있는 확률이 어느 정도인지를 알아보기 위하여 적중률을 구하여 보았는바, <표 4-3>에서와 같이, 충돌 형태별에 있어서는 차량 외 충돌이 86.0%로 가장 높게 나타났고, 다음으로는 차량 충돌이 83.1% 그리고 전도, 전복, 추락 등이 34.5%의 순으로 나타났으며, 전체 적중률도 77.9%로 매우 높게 나타났다.

<표 4-3> 충돌 형태별 판별모델의 판별 결과(전체)

판별된 유형 미래의 유형	차량 충돌	차량 외 충돌	전도, 전복, 추락 등	계	적중율
차량 충돌	147 (83.1%)	15 (8.5%)	15 (8.5%)	177 (100.0%)	83.1%
차량 외 충돌	5 (11.6%)	37 (86.0%)	1 (2.3%)	43 (100.0%)	86.0%
전도, 전복, 추락 등	5 (17.2%)	14 (48.3%)	10 (34.5%)	29 (100.0%)	34.5%
계	157	66	26	249	77.9%

그리고 이러한 판별모델에서 외적 변수에 대한 항목들의 조합으로 구성되는 축이 외적 변수를 어느 정도까지 잘 나타내 주고 있는가를 보여주는 상관비를 구하여 보았는바, 제1축이 0.70998, 제2축이 0.14942로 나타났다. 이 중에서 제1축에 대한 <그림 4-2>의 판별모델이 상관비가 0.70998에서 대형 교통사고 발생 시의 충돌 형태인 외적 변수의 구분 요인을 가장 잘 설명하고 있는 판별모델로서 유효하다고 판단되었다.

다음으로 외적 변수에 대한 기여도를 나타내는 편상관계수는 계수가 큰 변수일수록 외적 변수에 대한 영향 정도의 평가에 있어서도 중요한 요인으로 작용한다고 볼 수 있다. 따라서 외적 변수인 대형 교통사고 발생 시의 충돌 형태에 가장 큰 영향을 미치는 영향요인을 편상관계수가 높은 순서로 보면, 타 차량 제공 요인, 법규위반 유형, 도로 중앙부 형태, 노면 포장상태, 총 차로수, 사고 발생 시간대별의 순으로 나타났는바, 인적 요인 > 도로·환경적 요인의 순으로 영향을 미치고 있음을 알 수 있었다.

특히 이 판별모델에서 외적 변수인 대형 교통사고 발생 시의 충돌 형태 유형 점수(Category Score)를 보면, 전도, 전복, 추락 등 > 차량 외 충돌 > 차량 충

들의 순으로 나타났는바, 이를 토대로 살펴 볼 때, 각 항목에서 유형 점수가 큰 유형은 외적 변수의 유형 점수가 큰 유형에 기여하고 있다고 해석할 수가 있을 것이다.

이를 종합하여 보다 구체적으로 살펴보면,

1). 인적 요인 중 타 차량 제공 요인에서는 동일 방향 직진, 회전 중인 경우 등 타 차량 제공 요인이 있는 모든 경우와 법규 위반 유형에서는 신호위반과 중앙선 침범 그리고 철길 건널목 통과방법 위반을 한 경우가 대형 교통사고 충돌 형태별에 있어 차량 충돌에 영향을 미치고 있는 것으로 나타났다.

반면에 타 차량 제공 요인에서 타 차량 제공 요인이 없는 경우와 법규 위반 유형에서 제한속도 위반, 앞지르기 방법 및 금지 위반, 차로 위반, 주취 운전, 안전운전 의무 불이행, 무면허 운전 및 보행자 보호 의무 불이행 등의 경우가 대형 교통사고 충돌 형태별에 있어 차량 외 충돌과 전도, 전복, 추락 등에 영향을 미치고 있는 것으로 나타났다.

2). 도로·환경적 요인 중에서는 도로 중앙부 형태에서는 중앙분리대와 황색실선 및 점선이 설치된 경우가, 노면 포장 상태에서는 아스팔트와 콘크리트 포장인 경우, 총 차로수에서는 도로 폭이 매우 좁은 1차로와 반대로 도로 폭이 매우 넓은 4차로 이상인 경우가, 사고 발생 시간대에서는 야간 시간대인 20시부터 익일 06시 59분 사이인 경우가 대형 교통사고 충돌 형태별에 있어 차량 충돌에 영향을 미치고 있는 것으로 나타났다.

반면에 도로 중앙부 형태에서는 철길이 지나가는 경우와 도로 중앙부에 별도의 시설이 없는 경우가, 노면 포장 상태가 비포장인 경우, 총 차로수에 있어서는 2차로인 경우 그리고 사고 발생 시간대별에 있어서는 출근 시간대인 07시부터 08시 59분 사이와 퇴근 시간대인 17시부터 19시 59분 사이의 모두를 포함한 시간대가 대형 교통사고 충돌 형태별에 있어 차량 외 충돌과 전도, 전복, 추락 등에 영향을 미치고 있는 것으로 나타났다.

항 목	범 주	빈 도	수 량 화 값	-2.0	-1.0	0	1.0	2.0	범 위	편상관계수
범규위반 유 형	1.신호위반	(13)	-0.22114						2.57304	0.44946 (2)
	2.중앙선 침범	(57)	-0.20709							
	3.제한속도위반	(41)	0.21058							
	4.앞지르기방법, 금지위반	(14)	0.56523							
	5.차로위반	(2)	1.28617							
	6.주취운전	(39)	0.08593							
	7.안전운전의무 불이행	(72)	0.04947							
	8.철길건널목 통과방법위반	(9)	-1.28687							
	9.무면허운전, 보행자보호 의무불이행 등	(2)	0.11416							
사고발생 시간대별	1.07:00~08:59	(32)	0.03503						0.16808	0.09076 (6)
	2.09:00~16:59	(87)	0.02105							
	3.17:00~19:59	(36)	0.09882							
	4.20:00~06:59	(94)	-0.06926							
총 차로수	1.1차로	(16)	-0.00223						0.21859	0.15672 (5)
	2.2차로	(95)	0.09882							
	3.4차로 이상	(138)	-0.06926							
노 면 포장상태	1.아스팔트	(223)	-0.01861						1.15875	0.21492 (4)
	2.콘크리트	(22)	-0.01862							
	3.비포장	(4)	1.14013							
도 로 중앙부 형 태	1.중앙분리대	(50)	-0.05019						0.91634	0.23541 (3)
	2.황색실선	(164)	-0.00750							
	3.황색점선	(11)	-0.52149							
	4.철길, 없음	(24)	0.39485							
타 차량 제공요인	1.동일방향 직진, 회전 중	(43)	-0.53371						2.14475	0.81671 (1)
	2.대향 직진, 회전 중	(94)	-0.56587							
	3.측면에서 직진, 회전 중	(34)	-0.56354							
	4.신호대기, 주·정차, 일시정차 등	(16)	-0.16185							
	5.없 음	(62)	1.57889							
Outside variable	1.차량 충돌	(177)	-0.53736	Eta-square (Correlation ratio) 0.70998						
	2.차량 외 충돌	(43)	1.30307							
	3.전도, 전복, 추락 등	(29)	1.34758							

<그림 4-2> 수량화 이론 II류에 의한 충돌 형태별 판별모델(전체)

목표변수의 오분률은 <표 4-4>와 같이, 위험 추정치(또는 예측 위험)가 0.136546으로, 이는 자식마디에서의 집단 내 분산을 의미하는 것으로 전체 분산에서 각 마디의 분산 증분(improvement)을 빼준 것이 0.136546인 것이다. 이러한 수치는 작을수록 더 선호되는데, 전체 분산이 0.289157(뿌리마디일 때의 전체 분산을 의미하는 것으로 작을수록 선호되는 예측 위험)임에 비하여 마디 내 분산은 0.136546인 것이다. 뿌리마디만 있을 경우와 회귀나무가 형성된 후의 분산 비율은 $0.136546 / 0.289157 = 0.47$ 로서, 뿌리마디만 있을 경우보다 53%(= 100% - 47%) 정도가 회귀나무에 의하여 추가적으로 설명되어지는 분산의 비율임을 뜻한다.

<표 4-4> 위험 도표(전체)

Misclassification Matrix					
		Actual Category			
		차량 충돌	차량 외 충돌	전도, 전복, 추락 등	Total
Predicted Category	차량 충돌	176	8	8	192
	차량 외 충돌	0	25	7	32
	전도, 전복, 추락 등	1	10	14	25
	Total	177	43	29	249
Resubstitution					
Risk Estimate		0.136546			
SE of Risk Estimate		0.02176			

그리고 각 마디에 대한 타당성을 평가하기 위하여, 해당 마디에서 자료가 얼마나 잘 분류되었는지를 살펴보기 위한 이익 도표(gain chart)는 <표 4-5>와 같다.

<표 4-5>이익 도표(전체)

Gain Summary												
Target variable : 충돌형태						Target category : 전도, 전복, 추락 등						
Node-by-Node							Cumulative					
Node	Node:n	Node:%	Resp:n	Resp:%	Gain(%)	Index(%)	Node:n	Node:%	Resp:n	Resp:%	Gain(%)	Index(%)
12	1	0.40	1	3.45	100.00000	858.62069	1	0.40	1	3.45	100.00000	858.62069
9	24	9.64	13	44.83	54.16667	465.08621	25	10.04	14	48.28	56.00000	480.82759
8	4	1.61	2	6.98	50.00000	429.31034	29	11.65	16	55.17	55.17241	473.72176
14	19	7.63	6	20.69	31.57895	271.14338	48	19.28	22	75.86	45.83333	393.53448
3	15	6.02	4	13.79	26.66667	228.96552	63	25.30	26	89.66	41.26984	354.35140
13	13	5.22	1	3.45	7.69231	66.04775	76	30.52	27	93.10	35.52632	305.03630
7	168	67.47	2	6.90	1.19048	10.22167	244	97.99	29	100.00	11.88525	102.04918
5	5	2.01	0	0.00	0.00000	0.00000	249	100.00	29	100.00	11.64659	100.00000

여기서 Node는 마디번호(node number), Node:n과 Node:%는 해당 마디번호에서의 응답 대상수와 백분율, Resp:n과 Resp:%는 해당 마디번호에서 목표변수의 원래 범주와 함께 올바르게 분류된 응답 대상수와 백분율을 의미한다. 그리고 Gain(%)는 (Resp:n)/(Node:n)의 비율로서, 해당 마디에서 목적범주가 제대로 분리된 응답 대상수의 비율을 의미하며, Index(%)는 Gain(%)이 전체 응답대상에서 목표범주 비율과 얼마나 비교되는지에 대한 측도를 의미한다.

<표 4-5>의 이익 도표에 의하면, 충돌 형태의 경우 첫 번째 마디(12번 마디)는 자료수(Node:n)가 1개소 중 응답 대상수(Resp:n)가 1개소로 응답률(Gain:%)은 100.00%를 나타내고 있다. 다음으로 이 마디에 대한 응답의 비율이 전체 응답의 비율과 얼마나 비교되는지를 나타내는 인덱스 점수(Index:%)를 보면, 858.62069%로, 이는 이 마디에 응답한 비율이 전체 자료에 대한 응답 비율보다 8배가량 높다는 것을 의미한다.

이와 같은 방식으로 이후의 마디에 대해서도 같은 방식으로 해석을 해 보면, 여섯 번째 마디(13번 마디)의 인덱스 점수는 66.05%로 100%보다 작은 바, 이에 해당하는 마디는 전체 자료보다 낮은 응답률을 갖는다는 것을 의미한다. 따라서 다섯 번째 마디와 여섯 번째 마디(3번 마디와 13번 마디)사이에 교차점(전체 자료보다 높은 응답률을 보이는 마디와 낮은 응답률을 보이는 마디의 교차점)이 있음을 알 수 있다.

또한 누적 통계량을 살펴보면, 다섯 번째 마디(3번 마디)까지를 취한다면, 전체 자료의 25%(Node:%, 63/249)에 해당하는 자료를 접촉하여 전체 응답 대상

의 89.65% (Resp:%, 26/29)의 응답률을 얻을 수 있게 된다. 여기서 추가적으로 여섯 번째 마디(13번 마디)와 일곱 번째 마디(7번 마디)까지를 포함한다면, 전체 자료의 97.99%(244/249)를 접촉하여 전체 응답 대상의 100.00%에 해당하는 응답률을 얻게 되는 것이다.

그러나 이 단계는 교차점을 지나면서부터 전체 자료의 접촉 비율과 포함 마디까지의 응답률과의 차이가 줄어들게 된다. 여기서 추가적으로 여섯 번째 마디(13번 마디)와 일곱 번째 마디(7번 마디)까지를 포함하면 응답률은 100.00%이지만, 전체 자료의 72.69%(181/249)를 별도로 더 접촉해야 함을 알 수 있다. 하지만 최소한 89.65%의 반응 비율을 예측하기를 원한다면, 다섯 번째 마디(3번 마디)까지를 목표(Target)로 삼으면 될 것이다.

이와 같은 이익 값을 이용하여 마디들이 어떻게 정의되는지를 나무지도 화면인 <그림 4-3>에서 보면, 충돌형태별에 있어 차량 충돌이 발생했던 177개소 중 171개소가 타 차량 제공 요인이 동일 방향 직진, 회전 중, 대향 직진, 회전 중, 측면에서 직진, 회전 중, 신호 대기, 주정차, 일시 정지 등인 경우에 속하는 것으로 나타났으며, 차량 외 충돌 사고가 발생했던 43개소 중에는 타 차량 제공 요인이 없는 경우가 35개소 포함되어 있는 것으로 나타났다. 이 중 차량 충돌이 발생했던 171개소가 속한 타 차량 제공 요인 중의 동일 방향 직진, 회전 중, 대향 직진, 회전 중, 측면에서 직진, 회전 중, 신호 대기, 주정차, 일시 정지 등인 경우는 다시금 법규위반 유형에 의해 나누어지며, 이 경우 앞지르기 방법 및 금지 위반, 차로 위반인 경우와 신호 위반, 중앙선 침범, 제한속도 위반, 주취운전, 안전운전 의무 불이행, 철길 건널목 통과방법 위반, 무면허 운전, 보행자 보호 의무 불이행 등인 경우 차량 충돌이 발생한 지점이 각각 10개소와 161개소가 속하는 것으로 나타났다. 이 중 신호 위반, 중앙선 침범, 제한속도 위반, 주취운전, 안전운전 의무 불이행, 철길 건널목 통과방법 위반, 무면허 운전, 보행자 보호 의무 불이행 등인 경우는 노면 포장상태에 의해 다시금 나누어지는데, 이 경우 아스팔트, 콘크리트인 경우와 비포장인 경우로 나누어지며, 이 중 아스팔트, 콘크리트인 경우가 159개소가 속하는 것으로 나타났다.

이에 반해 충돌 형태에 있어 차량 외 충돌 사고가 발생한 43개소 중 35개소가

속하는 타 차량 제공 요인이 없는 경우는 다시금 범규위반 유형에 의해 분리되지만, 이 중 신호위반, 제한속도 위반, 앞지르기 방법 및 금지 위반, 주취운전, 안전운전 의무 불이행, 무면허 운전, 보행자 보호 의무 불이행 등인 경우에 차량 외 충돌 사고가 발생한 지점 35개소 모두가 속하는 것으로 나타났다. 또한 이들 지점들은 다시금 사고발생 시간대별에 의해 분리되는데, 이 경우 17시부터 익일 06시 59분까지인 경우 차량 외 충돌사고가 발생한 지점이 25개소가 속하는 것으로 나타났다. 또한 사고발생 시간대별에 있어 17시부터 익일 06시 59분까지인 경우가 다시금 범규위반 유형에 의해 분리되는데, 이 중 신호위반, 제한속도 위반, 앞지르기 방법 및 금지 위반, 주취운전, 안전운전 의무 불이행인 경우에 25개소 모두가 속하는 것으로 나타났다. 이 중 범규위반 유형 중 신호위반, 제한속도 위반, 앞지르기 방법 및 금지 위반, 주취운전, 안전운전 의무 불이행의 경우는 다시금 범규위반 유형에 의해 분리되는데, 이 중 신호위반, 앞지르기 방법 및 금지위반, 주취운전인 경우와 제한속도 위반, 안전운전 의무 불이행인 경우가 차량 외 충돌사고가 발생한 지점이 각각 12개소와 13개소가 속하는 것으로 나타났다.

이를 종합하여 보면 차량 충돌의 경우는 사고 당시 타 차량 제공 요인이 동일 방향 직진, 회전 중, 대향 직진, 회전 중, 측면에서 직진, 회전 중, 신호 대기, 주·정차, 일시 정지 등을 하고 있는 경우, 범규위반 유형에서는 신호 위반, 중앙선 침범, 제한속도 위반, 주취운전, 안전운전 의무 불이행, 철길 건널목 통과방법 위반, 무면허 운전, 보행자 보호 의무 불이행 등인 경우, 노면 포장상태에서는 아스팔트, 콘크리트인 경우가 가장 많이 속하고 있었으며, 반면에 차량 외 충돌의 경우는 타 차량 제공 요인이 없는 경우 중 범규위반 유형 중 신호 위반, 제한속도 위반, 앞지르기 방법 및 금지 위반, 주취운전, 안전운전 의무 불이행, 무면허 운전, 보행자 보호 의무 불이행 등, 사고발생 시간대별에 있어서는 17시부터 익일 06시 59분까지인 경우, 또 다시 범규위반 유형에서는 신호 위반, 제한속도 위반, 앞지르기 방법 및 금지 위반, 주취운전, 안전운전 의무 불이행인 경우가 다시금 범규위반 유형에서 제한속도 위반, 안전운전 의무 불이행인 경우가 가장 많았다.

제2절 대형 교통사고 발생지점 유형화

대형 교통사고에 있어 발생지점을 특성별로 유형화하기 위하여 군집분석을 실시하였다. 먼저 지점 특성별에 따른 적정 집단 수의 결정³¹⁾은 <표 4-6>에서와 같이, 그 증가폭이 급격히 변화하는 단계에서 최적의 유형화 집단수를 결정하거나, 집단수를 판정하는데 사용되는 부편결정계수(Semi-partial R Square)값³²⁾과 초입방형 군집판정 기준(Cubic Clustering Criterion)의 값³³⁾을 동시에 고려하여 판단하였다. 그 결과 적정 집단수는 4개 집단인 것으로 나타났다.

<표 4-6> 최적 유형화 수준의 지표

군집수	부편결정계수	초입방형 군집판정 기준
10	0.0223	7.81
9	0.0234	6.90
8	0.0276	5.82
7	0.0284	5.13
6	0.0424	3.31
5	0.0567	0.84
4	0.0739	-1.50
3	0.0905	-3.00
2	0.1214	-2.20
1	0.1819	0.00

이 기준에 따라, 아래의 <그림 4-4> 수상도(Dendrogram)에서 본 연구대상 지점들을 최적의 집단수인 4개의 집단으로 구분하여 본 바, 해당 사례 지점들은 다음의 <표 4-7>과 같이 나타났다.

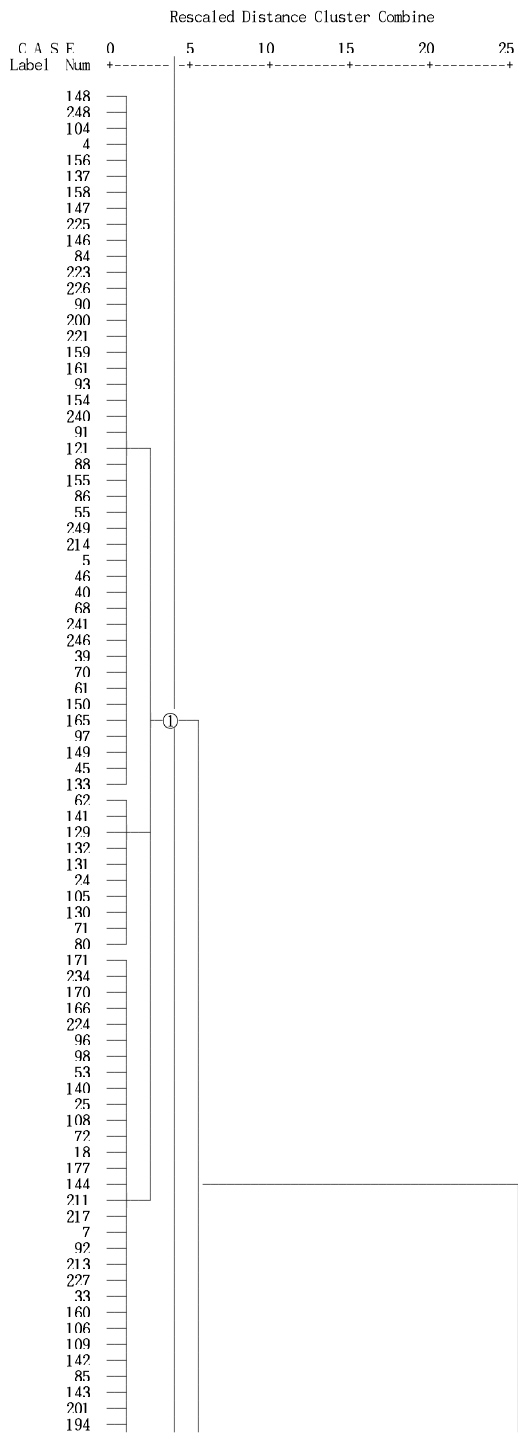
31). 김종섭, 전계서, p.500.

32). 허명희, 사회과학을 위한 다변량 자료 분석, 자유아카데미, 2000, pp.117~118.

33). 김기영, 전명식, SAS 군집분석, 자유아카데미, 1990, pp.55~60.

HIERARCHICAL CLUSTER ANALYSIS

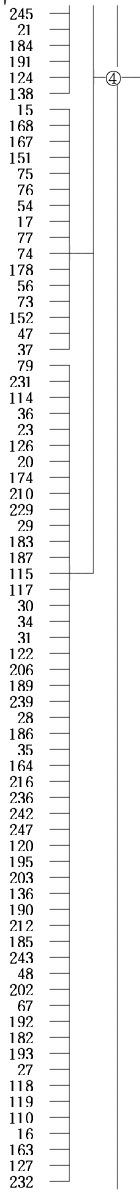
Dendrogram using Ward Method



<계속>

50		
87		
51		
176		
228		
172		
52		
173		
222		
38		
162		
157		
218		
153		
94	②	
198		
44		
244		
197		
219		
180		
181		
59		
95		
107		
235		
69		
175		
89		
233		
83		
139		
58		
60		
3		
6		
220		
1		
66		
179		
64		
101		
102		
10		
82		
8		
9		
125		
215	③	
196		
43		
12		
81		
78		
41		
199		
100		
65		
112		
14		
19		
99		
230		
145		
188		
237		
134		
135		
49		
205		
207		
11		
22		
13		
209		
42		
63		
208		
169		
238		
32		
111		
103		
128		
2		
57		
116		
123		
26		
113		
204		

<계속>



<그림 4-4> Ward법에 의한 수상도(Dendrogram)

<표 4-7> 집단별 사례 지점

구 분	대 상 지 점	비 고
1집단	4. 중앙고속도로(88), 5. 충남 아산군(88), 7. 강원 동해시(88), 18. 경기 남양주군(88), 24. 경기 안성군(88), 25. 경북 의성군(88), 33. 경북 금릉군(88), 39. 충북 증원군(88), 40. 광주 광산구(88), 45. 충북 증원군(90), 46. 인천시 북구(90), 50. 경부고속도로(90), 51. 경부고속도로(90), 53. 경기 남양주군(90), 55. 경북 경주시(90), 61. 경부고속도로(90), 62. 전북 김제군(90), 68. 남해고속도로(90), 70. 전남 목포시(90), 71. 강원 태백시(90), 72. 경북 경주시(90), 80. 경남 김해군(91), 84. 경부고속도로(91), 85. 광주 신관동(91), 86. 경기 광주군(91), 87. 중부고속도로(91), 88. 충남 당진군(91), 90. 창원시 팔용동(91), 91. 전북 완주군(91), 92. 원주시 문막면(91), 93. 광주시 북구(91), 96. 나주시 송월동(91), 97. 경기 용인군(91), 98. 충북 증원군(91), 104. 경기 양주군(91), 105. 경기 화성군(91), 106. 의정부시 호원동(91), 108. 호남고속도로(91), 109. 서울시 동작구(91), 121. 충북 제천시(96), 129. 경남 창원시(95), 130. 부산시 해운대구(96), 131. 경기 양평군(96), 132. 경북 문경군(95), 133. 충북 옥천군(95), 137. 호남고속도로(95), 140. 남해고속도로(96), 141. 충북 진천군(95), 142. 충북 단양군(96), 143. 서울시 노원구(96), 144. 경기 양평군(96), 146. 서울시 영등포구(96), 147. 경부고속도로(96), 148. 경북 구미시(95), 149. 전북 순창군(96), 150. 경기 시흥시(95), 154. 경기 수원시(96), 155. 서울시 종로구(95), 156. 부산시 동구(95), 158. 중부고속도로(96), 159. 경남 김해시(99), 160. 경남 거제시(99), 161. 경북 경주시(99), 165. 경남 통영시(99), 166. 대구시 달서구(99), 170. 전남 순천시(99), 171. 전북 정읍시(99), 177. 전남 해남군(99), 194. 울산시 남구(00), 200. 경기 양평군(01), 201. 경부고속도로(01), 211. 경기 양평군(02), 213. 경부고속도로(02), 214. 중앙고속도로(02), 217. 경기 남양주시(02), 221. 경기 안양시(02), 223. 인천 강화군(02), 224. 부산시 북구(02), 225. 남해고속도로(02), 226. 서울시 용산구(02), 227. 충남 연기군(02), 234. 강원 횡성군(03), 240. 충남 홍성군(03), 241. 경남 마산시(03), 246. 경기 평택시(03), 248. 충남 천안시(03), 249. 경부고속도로(03)	87개 지점
2집단	1. 경부고속도로(88), 3. 충북 음성군(88), 6. 강원 강릉시(88), 38. 경부고속도로(88), 44. 충북 청원군(90), 52. 경기 강화군(90), 58. 남해고속도로(90), 59. 인천시 북구(90), 60. 충남 예산군(90), 69. 전남 구례군(90), 83. 충북 증원군(91), 89. 강원 명주군(91), 94. 경기 파주시(91), 95. 충북 경주시(91), 107. 강원 양양군(91), 139. 충북 음성군(96), 153. 경기 부천시(96), 157. 영동고속도로(96), 162. 경기 시흥시(99), 172. 경남 통영시(99), 173. 경부고속도로(99), 175. 충남 서천군(99), 176. 경북 예천군(99), 180. 경기 오산시(00), 181. 부산시 기장군(00), 197. 경기 이천(01), 198. 경남 의령군(01), 218. 강원 강릉시(02), 219. 제주도 연동(02), 220. 전북 장수군(02), 222. 중앙고속도로(02), 228. 경북 봉화군(03), 233. 부산시 금정구(03), 235. 제주도 노형동(03), 244. 서울시 용산구(03)	35개 지점
3집단	8. 충북 경주시(88), 9. 경부고속도로(88), 10. 전북 군산시(88), 12. 서울시 용산구(88), 14. 경남 함안군(88), 19. 강원 홍천군(88), 41. 서울시 마포구(88), 43. 호남고속도로(90), 64. 충북 제천군(90), 65. 충남 음성군(90), 66. 서울시 성북구(90), 78. 서울시 노원구(91), 81. 경북 이성군(91), 82. 전남 영광군(91), 99. 전남 함평군(91), 100. 경기 고성군(91), 101. 경남 통영군(91), 102. 서울시 마포구(91), 112. 전북 완주군(91), 125. 충북 제천시(96), 145. 서해안고속도로(96), 179. 경남 김해군(99), 188. 경북 안동시(00), 196. 경남 산청군(01), 199. 경기 광주시(01), 215. 충남 태안군(02), 230. 전북 익산시(03), 237. 대전~진주간 고속도로(03)	28개 지점
4집단	2. 경북 경주시(88), 11. 경북 경산군(88), 13. 부산시 해운대구(88), 15. 경남 남해군(88), 16. 전남 승주군(88), 17. 충북 영동군(88), 20. 경기 화성군(88), 21. 전북 옥구군(88), 22. 경기 수원시(88), 23. 경남 김해시(88), 26. 경남 진양군(88), 27. 경기 남양주군(88), 28. 광주 광산구(88), 29. 경기 김포군(88), 30. 대전 동해시(88), 31. 경부고속도로(88), 32. 전북 남원시(88), 34. 경기 김포군(88), 35. 경북 월성군(88), 36. 서울시 영등포구(88), 37. 경기 남양주군(88), 42. 인천시 북구(88), 47. 경부고속도로(90), 48. 경기 여주군(90), 49. 서울시 강남구(90), 54. 안산시 일동(90), 56. 충북 증원군(90), 57. 호남고속도로(90), 63. 경부고속도로(90), 67. 경기 성남시(90), 73. 경기 남양주군(91), 74. 경남 함양군(91), 75. 경북 경주군(91), 76. 전남 나주군(91), 77. 충북 괴산군(91), 79. 대구시 달서구(91), 103. 전북 남원시(91), 110. 경남 산청군(91), 111. 전남 승주군(91), 113. 경남 고성군(91), 114. 강원 횡성군(91), 115. 충남 서산군(91), 116. 경기 화성군(91), 117. 충북 음성군(95), 118. 충북 영동군(96), 119. 대구시 북구(96), 120. 경남 함양군(96), 122. 전남 순천시(95), 123. 부산시 기장군(96), 124. 경기 김포군(96), 126. 충북 보은군(96), 127. 전북 전주시(96), 128. 경남 밀양시(96), 134. 서울시 동대문구(95), 135. 광주시 광산구(96), 136. 경기 성남시(96), 138. 남해고속도로(95), 151. 중앙고속도로(95), 152. 충북 진천군(95), 163. 전북 전주시(99), 164. 경남 창원시(99), 167. 경북 경주시(99), 168. 전남 완도군(99), 169. 중앙고속도로(99), 174. 충남 공주시(99), 178. 호남고속도로(99), 182. 전북 무주군(00), 183. 제주 북제주군(00), 184. 대전시 유성구(00), 185. 경기 광명시(00), 186. 강원 영월군(00), 187. 경기 포천군(00), 189. 전북 군산시(00), 190. 경기 남양주시(00), 191. 경남 거창군(00), 192. 경남 남해군(00), 193. 충남 부여군(00), 195. 경북 경주시(00), 202. 대구시 중구(01), 203. 경남 창원군(01), 204. 울산시 울주군(01), 205. 울산시 남구(01), 206. 88고속도로(01), 207. 인천시 중구(01), 208. 충남 논산시(01), 209. 전북 임실군(01), 210. 제주 북제주군(01), 212. 경기 가평군(02), 216. 경부고속도로(02), 229. 전남 진도군(03), 231. 경북 경주시(03), 232. 경기 고양시(03), 236. 인천시 중구(03), 238. 경남 거제시(03), 239. 충북 충주시(03), 242. 부산시 사상구(03), 243. 전남 순천시(03), 245. 인천시 연수구(03), 247. 경기 포천시(03)	99개 지점
	계	249개 지점

이들 각 집단에 속하는 지점들에 대해 해당 변수 항목에 대한 빈도분석 결과를 감안하여 공통적으로 나타난 특성들을 정리하여 보면 다음의 <표 4-8>과 같다.

<표 4-8> 집단별 특성

집단(지점수)	변 수 유 형	특 성	동일 집단
1집단 (87개 지점)	충돌 형태별	차량 충돌	4집단
	법규위반 유형	안전운전 의무 불이행	2집단
	사고 발생 시간대별	20시~익일 06시 59분	2, 3집단
	총 차로수	4차로 이상	3, 4집단
	노면 포장상태	아스팔트	*
	도로 중앙부 형태	황색실선	*
	타 차량 제공 요인	대향 직진, 회전 중	4집단
2집단 (35개 지점)	충돌 형태별	차량 외 충돌	3집단
	법규위반 유형	안전운전 의무 불이행	1집단
	사고 발생 시간대별	20시~익일 06시 59분	1, 3집단
	총 차로수	2차로	
	노면 포장상태	아스팔트	*
	도로 중앙부 형태	황색실선	*
3집단 (28개 지점)	충돌 형태별	차량 외 충돌	2집단
	법규위반 유형	제한속도 위반	
	사고 발생 시간대별	20시~익일 06시 59분	1, 2집단
	총 차로수	4차로 이상	1, 4집단
	노면 포장상태	아스팔트	*
	도로 중앙부 형태	황색실선	*
	타 차량 제공 요인	타 차량 제공 요인 없음	
4집단 (99개 지점)	충돌 형태별	차량 충돌	1집단
	법규위반 유형	중앙선 침범	
	사고 발생 시간대별	09시~16시 59분	
	총 차로수	4차로 이상	1, 3집단
	노면 포장상태	아스팔트	*
	도로 중앙부 형태	황색실선	*
	타 차량 제공 요인	대향 직진, 회전 중	1집단

단, *는 모든 집단에 공통된 사항임

이를 보면, 먼저 1집단에 속하는 87개 지점들은 4집단과 같이 충돌 형태별에 있어서는 차량 충돌인 경우와 타 차량 제공 요인에 있어서는 대향 직진 내지 회전 중인 경우가 많았으며, 법규 위반 유형은 2집단과 같이 안전운전 의무 불이행이, 사고 발생 시간대별에 있어서는 2와 3집단과 같이 야간 시간대인 20시부터

익일 06시 59분 사이가 많았고, 총 차로수에서는 3과 4집단과 같이 도로 폭이 매우 넓은 4차로 이상인 경우가 많았던 것으로 나타났다.

다음으로 2집단에 속하는 35개 지점들은 3집단과 같이 충돌 형태별에 있어서는 차량 외 충돌인 경우가 많았으며, 법규 위반 유형은 1집단과 같이 안전운전 의무 불이행이, 사고 발생 시간대별에 있어서는 1과 3집단과 같이 야간 시간대인 20시부터 익일 06시 59분 사이가 많았고, 총 차로수에서는 2차로인 경우가 많았던 것으로 나타났다.

그리고 3집단에 속하는 26개 지점들에 있어서는 2집단과 같이 충돌 형태별에 있어서는 차량 외 충돌인 경우가 많았으며, 사고 발생 시간대별에 있어서는 1과 2집단과 같이 야간 시간대인 20시부터 익일 06시 59분 사이가 많았고, 총 차로수에서는 1과 4집단과 같이 도로 폭이 매우 넓은 4차로 이상인 경우가, 타 차량 제공 요인에 있어서는 타 차량 제공 요인이 없었던 경우가 많았던 것으로 나타났다.

끝으로 4집단에 속하는 89개 지점들에 있어서는 1집단과 같이 충돌 형태별에 있어서는 차량 충돌인 경우와 타 차량 제공 요인에 있어서는 대향 직진 내지 회전 중인 경우가 많았으며, 총 차로수에서는 1, 3집단과 같이 4차로 이상인 경우가 많았던 것으로 나타났다.

제3절 집단별 충돌 형태별 영향요인 판별, 분류모델

249개의 조사대상 지점을 이러한 집단 분류방법에 의거하여 4개의 집단으로 구분하여 수량화 이론 II류와 C&RT에 의하여 각 집단별 대형 교통사고 충돌 형태별 판별 및 분류모델을 구축하여 본 바, 각 집단별 특성은 다음과 같이 나타났다.

1. 제1집단

(1). 판별모델

먼저 87건이 속한 제1집단에는 외적 변수인 충돌 형태를 <표 4-9>와 같이 빈도분석을 한 결과, 그 순위를 보면, 차량 충돌이 93.1%로 가장 많으며, 다음으로는 차량 외 충돌, 전도, 전복, 추락 등이 같은 비율로 나타났다.

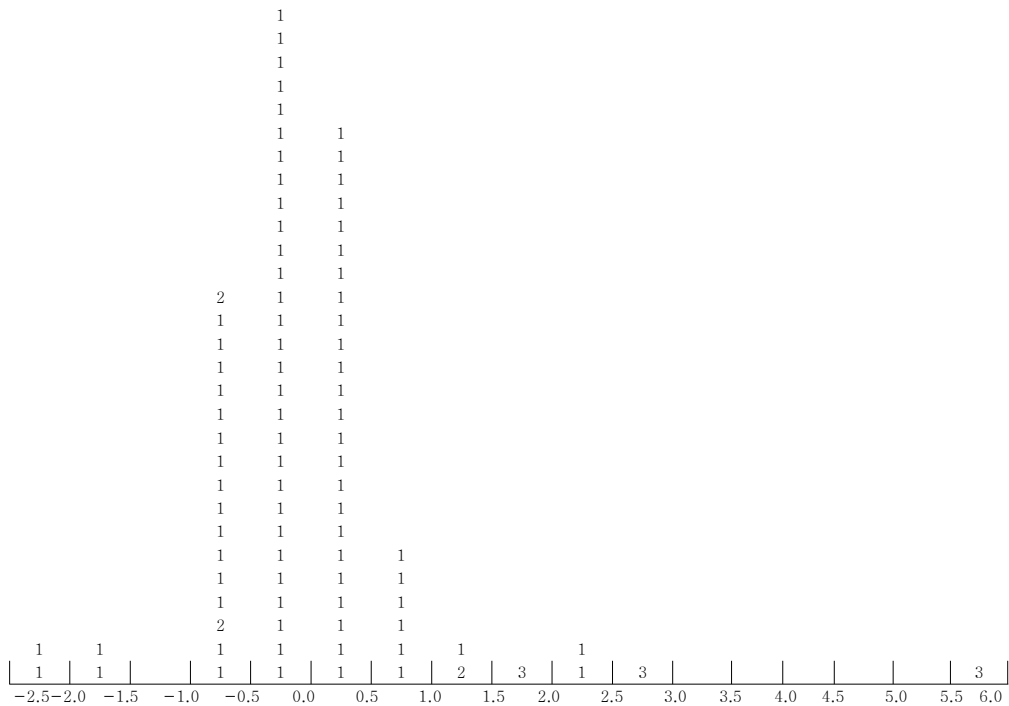
<표 4-9> 충돌 형태별 현황(제1집단)

충돌 형태	사고 건수(건)	비율(%)
차량 충돌	81	93.1
차량 외 충돌	3	3.4
전도, 전복, 추락 등	3	3.4
계	87	100.0

그리고 법규위반 유형에서만 신호위반, 중앙선 침범, 제한속도 위반, 앞지르기 방법 및 금지위반, 차로위반 등이 제외되었다. 따라서 이들 설명변수와 이에 따른 유형들을 설정하여 충돌 형태별 유형에 대해 수량화 이론 II류 분석을 행하였는바, 구축된 판별모델의 판별 범위는 <표 4-10>과 <그림 4-5>에서와 같이 나타났다.

<표 4-10> 충돌 형태별 외적 변수의 판별범위(제1집단)

충돌 형태	판별 범위
차량 충돌	$0.053 < (1) < 0.065$
차량 외 충돌	$-1.268 < (2) < 0.973$
전도, 전복, 추락 등	$-2.397 < (3) < -0.476$



<그림 4-5> 충돌 형태별 외적 변수의 Data Score Histogram(제1집단)

또한 적중률을 구하여 보았는바, <표 4-11>에서와 같이 충돌 형태별에 있어서는 차량 충돌이 82.7%로 가장 높게 나타났고, 다음으로는 차량 외 충돌과 전도, 전복, 추락 등이 66.7%의 순으로 나타났으며, 전체 적중률도 81.6%로 비교적 높게 나타났다.

<표 4-11> 충돌 형태별 판별모델의 판별 결과(제1집단)

판별된 유형 미래의 유형	차량 충돌	차량 외 충돌	전도, 전복, 추락 등	계	적중율
차량 충돌	67 (82.7%)	9 (11.1%)	5 (6.2%)	81 (100.0%)	82.7%
차량 외 충돌	1 (33.3%)	2 (66.7%)	—	3 (100.0%)	66.7%
전도, 전복, 추락 등	1 (33.3%)	—	2 (66.7%)	3 (100.0%)	66.7%
계	69	11	7	87	81.6%

그리고 상관비를 구하여 보았는바, <그림 4-6>의 판별모델에서 보는 바와

같이, 제1축이 0.43033, 제2축은 0.15117로 나타났다. 이 중 제1축에 대한 판별모델의 상관계수 0.43033에서 대형 교통사고 발생 시의 충돌 형태인 외적 변수의 구분 요인을 가장 잘 설명하고 있는 판별모델로서 유효하다고 판단되었다.

다음으로 편상관계수가 높은 순서로 볼 때, 총 차로수, 법규위반 유형, 타 차량 제공 요인, 도로 중앙부 형태, 노면 포장상태 그리고 사고 발생 시간대별의 순으로 나타나, 도로·환경적 요인 > 인적 요인의 순으로 영향을 미치고 있음을 알 수 있었다.

특히 이 판별모델에서 외적 변수인 대형 교통사고 발생 시의 충돌 형태 유형 점수(Category Score)를 보면, 전도, 전복, 추락 등 > 차량 외 충돌 > 차량 충돌의 순으로 나타났는바, 이를 토대로 하여, 보다 구체적으로 살펴보면 다음과 같은 특징이 나타났다.

1). 인적 요인 중 법규 위반 유형에서 주취운전, 철길건널목 통과방법위반을 한 경우와 타 차량 제공 요인에서는 측면에서 직진, 회전 중이거나, 신호대기, 주·정차, 일시정지 등인 경우가 대형 교통사고 충돌 형태별에 있어 차량 충돌과 차량 외 충돌에 영향을 미치고 있는 것으로 나타났다.

이에 반해 법규 위반 유형에서는 안전운전 의무 불이행, 무면허운전, 보행자 보호 의무 불이행 등의 경우와 타 차량 제공 요인에서는 동일 방향 직진, 회전 중이거나 대향 직진, 회전 중인 경우 그리고 타 차량 제공 요인이 없는 경우가 대형 교통사고 충돌 형태별에 있어 전도, 전복, 추락 등에 영향을 미치고 있는 것으로 나타났다.

2). 도로·환경적 요인 중에는 총 차로수는 도로 폭이 매우 넓은 4차로 이상인 경우, 도로 중앙부 형태에서는 중앙분리대와 황색점선인 경우, 노면 포장상태는 콘크리트 포장인 경우 그리고 사고 발생 시간대에서는 일상 시간대인 09시부터 16시 59분 사이와 야간 시간대인 20시부터 익일 06시 59분 사이인 경우가 대형 교통사고 충돌 형태별에 있어 차량 충돌과 차량 외 충돌에 영향을 미치고 있는 것으로 나타났다.

이에 반해 총 차로수에 있어서는 1차로와 2차로인 경우, 도로 중앙부 형태에서 황색실선인 경우와 철길이 지나가는 경우와 도로 중앙부에 별도의 시설이

없는 경우, 노면 포장상태는 아스팔트 포장인 경우와 비포장인 경우 그리고 사고 발생 시간대별에 있어서는 출근 시간대인 07시부터 16시 59분 사이, 퇴근 시간대인 17시부터 19시 59분 사이인 경우가 대형 교통사고 충돌 형태별에 있어 전도, 전복, 추락 등에 영향을 미치고 있는 것으로 나타났다.

항 목	범 주	빈 도	수 량 화 값	-2.0	-1.0	0	1.0	2.0	범 위	편상관계수
법규위반 유 형	1.주취운전	(26)	-0.55023						2.14945	0.46877 (2)
	2.안전운전의무 불이행	(51)	0.55668							
	3.철길건널목 통과방법위반	(9)	-1.59277							
	4.무면허운전, 보행자보호 의무불이행 등	(1)	0.25003							
사고발생 시간대별	1.07:00~08:59	(7)	0.33601						0.55288	0.15043 (6)
	2.09:00~16:59	(28)	-0.10218							
	3.17:00~19:59	(8)	0.45070							
	4.20:00~06:59	(44)	-0.07037							
총차로수	1.1차로	(9)	2.54386						2.99384	0.52788 (1)
	2.2차로	(23)	0.08061							
	3.4차로 이상	(55)	-0.44998							
노 면 포 장 상태	1.아스팔트	(73)	0.01633						2.01135	0.22134 (5)
	2.콘크리트	(12)	-0.37250							
	3.비포장	(2)	1.63885							
도 로 중 양 부 형 태	1.중앙분리대	(21)	-0.27815						1.10642	0.24188 (4)
	2.황색실선	(51)	0.21879							
	3.황색점선	(5)	-0.88764							
	4.철길, 없음	(10)	-0.08787							
타 차량 제공요인	1.동일방향 직진, 회전 중	(22)	0.00055						1.34660	0.38127 (3)
	2.대향 직진, 회전 중	(28)	0.58457							
	3.측면에서 직진, 회전 중	(22)	-0.76202							
	4.신호대기, 주·정차, 일시정지 등	(10)	-0.01420							
	5.없 음	(5)	0.10527							
Outside variable	1.차량 충돌	(81)	-0.12844	Eta-square (Correlation ratio) 0.43033						
	2.차량 외 충돌	(3)	-0.00130							
	3.전도, 전복, 추락 등	(3)	3.46905							

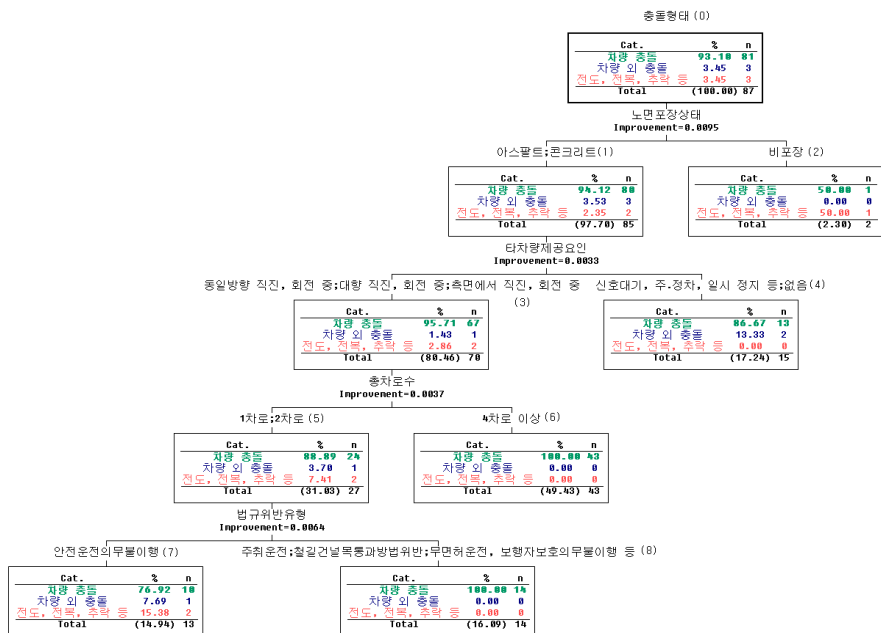
<그림 4-6> 수량화 이론 II류에 의한 충돌 형태별 판별모델(제1집단)

(2). 분류모델

다음으로는 전체 249개소 중 1집단에 속하는 87개소의 관찰대상과 1개의 목표변수 그리고 6개의 예측변수를 사용하여 분류모델을 구축하였는바, 그 모델은 <그림 4-7>과 같이 나타났다.

이를 보면 충돌 형태별에 대한 영향요인 분류모델에서는 예측변수로 노면 포장상태, 타 차량 제공 요인, 총 차로수, 법규위반 유형의 4 가지 변수가 채택되었다. 이 예측변수로 분리했을 때 분산의 감소 정도는 0.0095(Improvement = 0.0095)로 전체 분산을 거의 줄여주지 못하고 있다.

목표변수인 충돌 형태별의 빈도에서는 <그림 4-7>과 같이, “차량 충돌”이 가장 많이 나타났으며, 이후에는 “차량 외 충돌”과 “전도, 전복, 추락 등”이 같은 빈도인 것으로 나타났으며, 특히 “차량 충돌”인 경우는 “차량 외 충돌”과 “전도, 전복, 추락 등”의 경우에 비해 27배 정도 더 높게 나타났다.



<그림 4-7> 충돌 형태별 영향요인 분류모델(1집단)

주) ()안의 숫자는 마디 번호임

목표변수의 오분률은 <표 4-12>와 같이, 위험 추정치(또는 예측 위험)가 0.0689655로, 이는 자식마디에서의 집단 내 분산을 의미하는 것으로 전체 분산에서 각 마디의 분산 증분을 빼준 것이 0.0689655인 것이다. 이러한 수치는 작을수록 더 선호되는데, 전체 분산이 0.0689655이고 마디 내 분산은 0.0689655인 것이다. 따라서 뿌리마디만 있을 경우와 회귀나무가 형성된 후의 분산 비율이 $0.0689655 / 0.0689655 = 1$ 로서, 뿌리마디만 있을 경우와 회귀나무에 의하여 추가적으로 설명되어지는 분산 간에는 차이가 전혀 없는 것으로 나타났다.

<표 4-12> 위험 도표(1집단)

Misclassification Matrix					
		Actual Category			
		차량 충돌	차량 외 충돌	전도, 전복, 추락 등	Total
Predicted Category	차량 충돌	81	3	3	87
	차량 외 충돌	0	0	0	0
	전도, 전복, 추락 등	0	0	0	0
	Total	81	3	3	87
Resubstitution					
Risk Estimate		0.0689655			
SE of Risk Estimate		0.0271668			

그리고 각 마디에 대한 타당성을 평가하기 위하여, 해당 마디에서 자료가 얼마나 잘 분류되었는지를 살펴보기 위한 이익 도표는 <표 4-13>과 같다.

<표 4-13> 이익 도표(1집단)

Gain Summary												
Target variable : 충돌형태							Target category : 전도, 전복, 추락 등					
Node-by-Node							Cumulative					
Node	Node:n	Node:%	Resp:n	Resp:%	Gain(%)	Index(%)	Node:n	Node:%	Resp:n	Resp:%	Gain(%)	Index(%)
2	2	2.30	1	33.33	50.00000	1450.00000	2	2.30	1	33.33	50.00000	1450.00000
7	13	14.94	2	66.67	15.38462	446.15385	15	17.24	3	100.00	20.00000	580.00000
6	43	49.43	0	0.00	0.00000	0.00000	58	66.67	3	100.00	5.17241	150.00000
4	15	17.24	0	0.00	0.00000	0.00000	73	83.91	3	100.00	4.10959	119.17808
8	14	16.09	0	0.00	0.00000	0.00000	87	100.00	3	100.00	3.44828	100.00000

<표 4-13>의 이익 도표에 의하면, 충돌 형태의 경우 첫 번째 마디(2번 마

디)는 자료수가 2개소 중 응답 대상수가 1개소로 응답률은 33.33%를 나타내고 있다. 다음으로 이 마디에 대한 응답의 비율이 전체 응답의 비율과 얼마나 비교되는지를 나타내는 인덱스 점수를 보면, 1,450.00%로, 이는 이 마디에 응답한 비율이 전체 자료에 대한 응답 비율보다 14.5배가량 높다는 것을 의미한다.

이와 같은 방식으로 이후의 마디에 대해서도 같은 방식으로 해석을 해 보면, 세 번째 마디(6번 마디)의 인덱스 점수는 0.00%인 바, 이에 해당하는 마디는 전체 자료보다 낮은 응답률을 갖는다는 것을 의미한다. 따라서 두 번째 마디와 세 번째 마디(7번 마디와 6번 마디)사이에는 교차점이 있음을 알 수 있다.

또한 누적 통계량을 살펴보면, 두 번째 마디(7번 마디)까지를 취한다면, 전체 자료의 17.24%(Node:%, 15/87)에 해당하는 자료를 접촉하여 전체 응답 대상의 100.00% (Resp:%, 3/3)의 응답률을 얻을 수 있게 된다. 여기서 추가적으로 세 번째 마디(6번 마디)를 포함한다면, 전체 자료의 66.67%(58/87)를 접촉하여 전체 응답 대상의 100.00%에 해당하는 응답률을 얻게 되는 것이다.

그러나 이 단계는 교차점을 지나면서부터 전체 자료의 접촉 비율과 포함 마디까지의 응답률과의 차이가 줄어들게 된다. 여기서 추가적으로 세 번째 마디(6번 마디)를 포함하면 응답률은 100.00%이지만, 전체 자료의 49.42%(43/87)를 별도로 더 접촉해야 함을 알 수 있다. 하지만 100.00%의 반응 비율을 예측하기를 원한다면, 두 번째 마디(7번 마디)까지를 목표로 삼으면 될 것이다.

이와 같은 이익 값을 이용하여 마디들이 어떻게 정의되는지를 나무지도 화면인 <그림 4-7>에서 보면, 충돌형태별에 있어 차량 충돌이 발생했던 81개소 중 노면 포장상태 요인이 아스팔트, 콘크리트인 경우가 80개소, 타 차량 제공 요인은 동일 방향 직진, 회전 중, 대향 직진, 회전 중, 측면에서 직진, 회전 중인 경우가 67개소, 총 차로수에서는 4차로 이상인 경우가 43개소가 속하는 것으로 나타났다. 총 차로수에 있어 1, 2차로인 경우는 다시금 범규위반 유형에 있어 주취운전, 철길 건널목 통과방법 위반, 무면허 운전, 보행자 보호 의무 불이행 등인 경우가 14개소, 안전운전 의무 불이행인 경우 10개소가 속하는 것으로 나타났다.

2. 제2집단

(1). 판별모델

다음으로 제2집단에 대해 외적 변수인 충돌 형태를 <표 4-14>와 같이 빈도분석을 한 결과, 순위를 보면, 차량 외 충돌, 전도, 전복, 추락 등의 순으로 나타났다.

<표 4-14> 충돌 형태별 현황(제2집단)

충돌 형태	사고 건수(건)	비율(%)
차량 외 충돌	21	60.0
전도, 전복, 추락 등	14	40.0
계	35	100.0

그리고 범규위반 유형에서는 신호위반, 중앙선 침범, 제한속도 위반, 앞지르기 방법 및 금지위반, 차로위반, 철길 건널목 통과방법 위반이, 노면 포장상태에서는 비포장인 경우, 타 차량 제공 요인 전부가 제외되었다. 따라서 이들 설명변수와 이에 따른 유형들을 설정하여 충돌 형태별 유형에 대해 수량화 이론 II류 분석을 행하였는바, 구축된 판별모델은 외적 변수인 충돌 형태가 차량 외 충돌, 전도, 전복, 추락 등의 2가지로만 구분됨으로 인하여 <표 4-15>에서와 같이 외적 변수의 집단 중심값과 집단 표본수를 고려한 경계점을 산정하였는바 0.28³⁴⁾로 나타났다.

<표 4-15> 충돌 형태별 외적 변수의 집단 중심값(제2집단)

충돌 형태	집단 중심값
차량 외 충돌	-0.564
전도, 전복, 추락 등	0.845

이를 감안하여 볼 때, <그림 4-8>에서 보는 바와 같이, 사례별 판별 점수가 0.28보다 작으면 충돌 형태별에 있어 차량 외 충돌한 집단에 분류되며, 0.28보

34). 채서일, 김종범, 전계서, p.112

$$\text{Cutting Score(경계점)} = \frac{N_2 C_1 + N_1 C_2}{N_1 + N_2}$$

N_i : 집단의 표본수, C_i : 집단의 중심값(Centroid)

$$= \frac{(14 \times (-0.564)) + (21 \times 0.845)}{21 + 14} = \frac{-7.89 + 17.74}{35} = \frac{9.85}{35} = 0.28$$

특히 이 판별모델에서 외적 변수인 대형 교통사고 발생 시의 충돌 형태 유형 점수(Category Score)를 보면, 전도, 전복, 추락 등 > 차량 외 충돌의 순으로 나타났는바, 이를 토대로 하여, 보다 구체적으로 살펴보면 다음과 같은 특징이 나타났다.

1). 인적 요인 중 법규 위반 유형에서 주취운전을 한 경우가 대형 교통사고 충돌 형태별에 있어 차량 외 충돌에 영향을 미치고 있는 것으로 나타났다.

반면에 법규 위반 유형에서 안전운전 의무 불이행, 무면허운전, 보행자 보호 의무 불이행 등의 경우가 대형 교통사고 충돌 형태별에 있어 전도, 전복, 추락 등에 영향을 미치고 있는 것으로 나타났다.

2). 도로·환경적 요인 중에는 사고 발생 시간대에서는 퇴근 시간대인 17시부터 19시 59분 사이인 경우, 도로 중앙부 형태에서는 중앙분리대, 노면 포장 상태에서는 아스팔트 포장인 경우, 그리고 총 차로수에서는 도로 폭이 도로 폭이 매우 좁은 1차로인 경우와 2차로인 경우가 대형 교통사고 충돌 형태별에 있어 차량 외 충돌에 영향을 미치고 있는 것으로 나타났다.

반면에 사고 발생 시간대별에 있어 출근 시간대인 07시부터 08시 59분 사이, 일상 시간대인 09시부터 16시 59분 사이 그리고 야간 시간대인 20시부터 익일 06시 59분 사이인 경우, 도로 중앙부 형태에서 황색실선 및 점선인 경우, 철길이 지나가는 경우와 도로 중앙부가 없는 경우, 노면 포장 상태에서는 콘크리트 포장인 경우 그리고 총 차로수에 있어서는 매우 넓은 4차로 이상인 경우가 대형 교통사고 충돌 형태별에 있어 전도, 전복, 추락 등에 영향을 미치고 있는 것으로 나타났다.

항 목	범 주	빈 도	수 량 화 값	-2.0 -1.0 0 1.0 2.0 3.0						범 위	편상관계수
범규위반 유 형	1.주취운전	(13)	-0.95856							3.54585	0.56562 (1)
	2.안전운전의무 불이행	(21)	0.47019								
	3.무면허운전, 보행자보호 의무불이행 등	(1)	2.58729								
사고발생 시간대별	1.07:00~08:59	(1)	1.33930							2.46086	0.46604 (2)
	2.09:00~16:59	(12)	0.29851								
	3.17:00~19:59	(6)	-1.12156								
	4.20:00~06:59	(16)	0.11299								
충차로수	1.1차로	(2)	-0.07343							0.35628	0.13878 (5)
	2.2차로	(18)	-0.15749								
	3.4차로 이상	(15)	0.19878								
노면포장 상 태	1.아스팔트	(33)	-0.06774							1.18544	0.29264 (4)
	2.콘크리트	(2)	1.11770								
도 로 중앙부 형 태	1.중앙분리대	(7)	-1.08307							2.24360	0.36476 (3)
	2.황색실선	(23)	0.19569								
	3.황색점선	(1)	1.16053								
	4.철길, 없음	(4)	0.48000								
Outside variable	1.차량 외 충돌	(21)	-0.58084	Eta-square (Correlation ratio) 0.49648							
	2.전도, 전복, 추락 등	(14)	0.87125								

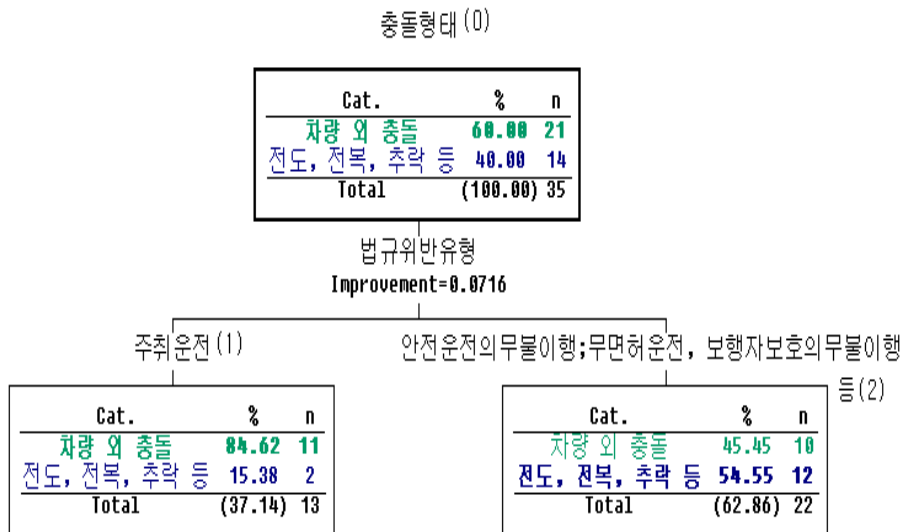
<그림 4-9> 수량화 이론 II류에 의한 충돌 형태별 판별모델(제2집단)

(2). 분류모델

다음으로는 전체 249개소 중 2집단에 속하는 35개소의 관찰대상과 1개의 목표변수 그리고 5개의 예측변수를 사용하여 분류모델을 구축하였는바, 그 모델은 <그림 4-10>과 같이 나타났다.

이를 보면 충돌 형태별에 대한 영향요인 분류모델에서는 예측변수로 법규위반 유형의 한 가지 변수만 채택되었다. 이 예측변수로 분리했을 때 분산의 감소 정도는 0.0716(Improvement = 0.0716)로 전체 분산을 7.2% 정도 줄여주고 있다.

목표변수인 충돌 형태별의 빈도에서는 <그림 4-10>과 같이, “차량 외 충돌”, “전도, 전복, 추락 등”의 순으로 나타났으며, 특히 “차량 외 충돌”인 경우는 “전도, 전복, 추락 등”의 경우에 비해 1.5배 정도 더 높게 나타났다.



<그림 4-10> 충돌 형태별 영향요인 분류모델(2집단)

주) ()안의 숫자는 마디 번호임

목표변수의 오분률은 <표 4-17>과 같이, 위험 추정치(또는 예측 위험)가 0.342857로, 이는 자식마디에서의 집단 내 분산을 의미하는 것으로 전체 분산에서 각 마디의 분산 증분을 뺀 것이 0.342857인 것이다. 이러한 수치는 작을수록 더 선호되는데, 전체 분산이 0.342857이고 마디 내 분산도 0.342857인 것이다. 따라서 뿌리마디만 있을 경우와 회귀나무가 형성된 후의 분산 비율은 $0.342857 / 0.342857 = 1$ 로서, 뿌리마디만 있을 경우와 회귀나무에 의하여 추가적으로 설명되어지는 분산 간에는 차이가 전혀 없는 것으로 나타났다.

<표 4-17> 위험 도표(2집단)

Misclassification Matrix				
		Actual Category		
		차량 외 충돌	전도, 전복, 추락 등	Total
Predicted Category	차량 외 충돌	11	2	13
	전도, 전복, 추락 등	10	12	22
	Total	21	14	35
Resubstitution				
Risk Estimate		0.342857		
SE of Risk Estimate		0.0802329		

그리고 각 마디에 대한 타당성을 평가하기 위하여, 해당 마디에서 자료가 얼마나 잘 분류되었는지를 살펴보기 위한 이익 도표는 <표 4-18>과 같다.

<표 4-18> 이익 도표(2집단)

Gain Summary												
Target variable : 충돌형태							Target category : 전도, 전복, 추락 등					
Node-by-Node							Cumulative					
Node	Node:n	Node:%	Resp:n	Resp:%	Gain(%)	Index(%)	Node:n	Node:%	Resp:n	Resp:%	Gain(%)	Index(%)
2	22	62.86	12	85.71	54.54545	136.36364	22	62.86	12	85.71	54.54545	136.36364
1	13	37.14	2	14.29	15.38462	38.46154	35	100.00	14	100.00	40.00000	100.00000

<표 4-18>의 이익 도표에 의하면, 충돌 형태의 경우 첫 번째 마디(2번 마디)는 자료수가 35개소 중 응답 대상수가 12개소로 응답률은 54.55%를 나타내고 있다. 다음으로 이 마디에 대한 응답의 비율이 전체 응답의 비율과 얼마나 비

교되는지를 나타내는 인덱스 점수를 보면, 136.36%로, 이는 이 마디에 응답한 비율이 전체 자료에 대한 응답 비율보다 약간 높다는 것을 의미한다.

이와 같은 방식으로 이후의 마디에 대해서도 같은 방식으로 해석을 해 보면, 두 번째 마디(1번 마디)까지의 인덱스 점수는 38.46%인 바, 이에 해당하는 마디는 전체 자료보다 낮은 응답률을 갖는다는 것을 의미한다. 따라서 첫 번째 마디와 두 번째 마디(2번 마디와 1번 마디)사이에 교차점이 있음을 알 수 있다.

또한 누적 통계량을 살펴보면, 첫 번째 마디(2번 마디)를 취한다면, 전체 자료의 62.86%(Node:%, 22/35)에 해당하는 자료를 접촉하여 전체 응답 대상의 85.71% (Resp:%, 12/14)의 응답률을 얻을 수 있게 된다. 여기서 추가적으로 두 번째 마디(1번 마디)를 포함한다면, 전체 자료의 100.00%(35/35)를 접촉하여 전체 응답 대상의 100.00%에 해당하는 응답률을 얻게 되는 것이다.

그러나 이 단계는 교차점을 지나면서부터 전체 자료의 접촉 비율과 포함 마디까지의 응답률과의 차이가 줄어들게 된다. 여기서 추가적으로 두 번째 마디(1번 마디)를 포함하면 응답률은 100.00%이지만, 전체 자료의 37%(13/35)를 별도로 더 접촉해야 함을 알 수 있다. 하지만 85.71%의 반응 비율을 예측하기를 원한다면, 첫 번째 마디(2번 마디)까지를 목표로 삼으면 될 것이다.

이와 같은 이익 값을 이용하여 마디들이 어떻게 정의되는지를 나무지도 화면인 <그림 4-10>에서 보면, 충돌형태별에 있어 차량 외 충돌이 발생했던 21개소 중 범규위반 유형 요인이 주취운전인 경우에 11개소, 안전운전 의무 불이행, 무면허 운전, 보행자 보호 의무 불이행인 경우는 10개소가 속하는 것으로 나타났다.

3. 제3집단

(1). 판별모델

다음으로 제3집단에 대해 외적 변수인 충돌 형태를 <표 4-19>와 같이 빈도분석을 한 결과 순위를 보면, 차량 외 충돌, 전도, 전복, 추락 등, 차량 충돌의 순으로 나타났다.

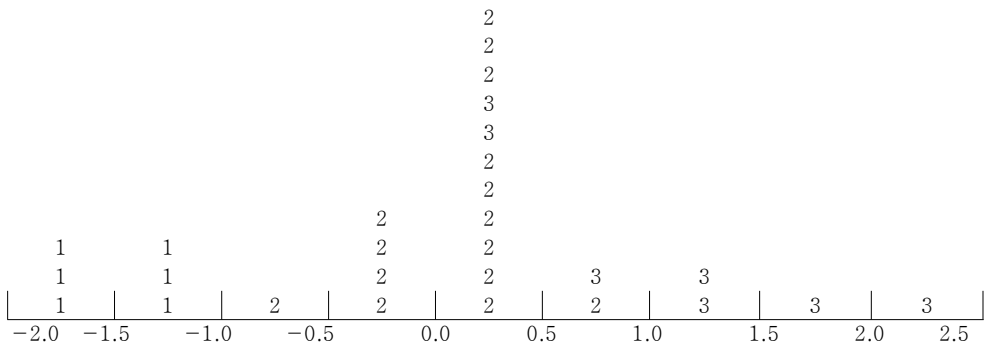
<표 4-19> 충돌 형태별 현황(제3집단)

충돌 형태	사고 건수(건)	비율(%)
차량 충돌	6	21.4
차량 외 충돌	15	53.6
전도, 전복, 추락 등	7	25.0
계	28	100.0

그리고 법규위반 유형에서는 차로위반, 주취운전, 안전운전 의무 불이행, 철길 건널목 통과방법 위반, 무면허 운전, 보행자 보호 의무 불이행 등이, 노면 포장 상태에서는 비포장인 경우, 타 차량 제공 요인에서는 동일 방향 직진, 회전 중, 대향 직진, 회전 중, 측면에서 직진, 회전 중이 제외되었다. 따라서 이들 설명변수와 이에 따른 유형들을 설정하여 충돌 형태별 유형에 대해 수량화 이론 II류 분석을 행하였는바, 구축된 판별모델의 판별 범위는 <표 4-20>과 <그림 4-11>에서와 같이 나타났다.

<표 4-20> 충돌 형태별 외적 변수의 판별범위(제3집단)

충돌 형태	집단의 중심값
차량 충돌	-2.048 < (1) < -0.328
차량 외 충돌	0.040 < (2) < 0.299
전도, 전복, 추락 등	-0.360 < (3) < 1.670



<그림 4-11> 충돌 형태별 외적 변수의 Data Score Histogram(제3집단)

또한 적중률을 구하여 보았는바, <표 4-21>에서와 같이, 충돌 형태별에 있어서는 전도, 전복, 추락 등이 85.7%로 가장 높게 나타났고, 다음으로는 차량 충돌이 83.3%, 차량 외 충돌이 60.0%의 순으로 나타났으며, 전체 적중률은 71.4%로 비교적 높게 나타났다.

<표 4-21> 충돌 형태별 판별모델의 판별 결과(제3집단)

판별된 유형 미래의 유형	차량 충돌	차량 외 충돌	전도, 전복, 추락 등	계	적중율
차량 충돌	5 (83.3%)	1 (16.7%)	—	6 (100.0%)	83.3%
차량 외 충돌	2 (13.3%)	9 (60.0%)	4 (26.7%)	15 (100.0%)	60.0%
전도, 전복, 추락 등	—	1 (14.3%)	6 (85.7%)	7 (100.0%)	85.7%
계	7	11	10	28	71.4%

그리고 상관비를 구하여 보았는바, <그림 4-12>의 판별모델에서 보는 바와 같이, 제1축이 0.80411, 제2축이 0.28817로 나타났다. 이 중 제1축에 대한 판별모델의 상관비 0.80411에서 대형 교통사고 발생 시의 충돌 형태인 외적 변수의 구분 요인을 가장 잘 설명하고 있는 판별모델로서 유효하다고 판단되었다.

다음으로 편상관계수가 높은 순서로 볼 때, 법규위반 유형, 사고 발생 시간대, 타 차량 제공 요인, 도로 중앙부 형태, 노면 포장상태, 총 차로수의 순으로 나타나 인적 요인 > 도로·환경적 요인의 순으로 영향을 미치고 있음을 알 수 있었다.

특히 이 판별모델에서 외적 변수인 대형 교통사고 발생 시의 충돌 형태 유형 점수(Category Score)를 보면, 전도, 전복, 추락 등 > 차량 외 충돌 > 차량 충돌로 나타났는바, 이를 토대로 하여, 보다 구체적으로 살펴보면 다음과 같은 특징이 나타났다.

1). 인적 요인 중 법규 위반 유형에서 신호 위반, 중앙선 침범을 한 경우와 타 차량 제공 요인에서는 신호 대기, 주정차, 일시 정지 등을 한 경우가 대형 교통사고 충돌 형태별에 있어 차량 충돌에 영향을 미치고 있는 것으로 나타났다.

이에 반해 법규 위반 유형에서는 제한속도 위반을 한 경우와 앞지르기 방법 및 금지위반인 경우, 타 차량 제공 요인에서는 타 차량 제공 요인이 없는 경우가 대형 교통사고 충돌 형태별에 있어 차량 외 충돌과 전도, 전복, 추락 등에 영향을 미치고 있는 것으로 나타났다.

2). 도로·환경적 요인 중에는 사고 발생 시간대에서는 야간 시간대인 20시부터 익일 06시 59분 사이의 경우와 도로 중앙부 형태에서는 중앙분리대가 설치되어 있는 경우와 황색실선인 경우, 노면 포장상태에서는 아스팔트인 경우 그리고 총 차로수에서는 도로 폭이 매우 좁은 1차로와 2차로인 경우가 대형 교통사고 충돌 형태별에 있어 차량 충돌에 영향을 미치고 있는 것으로 나타났다.

이에 반해 사고 발생 시간대별에 있어 출근 시간대인 07시부터 08시 59분 사이부터 퇴근 시간대인 17시부터 19시 59분 사이까지인 경우와 도로 중앙부 상태에서는 황색점선인 경우, 철길이 지나가는 경우와 도로 중앙부가 없는 경우가, 노면 포장상태에서는 콘크리트 포장인 경우 그리고 총 차로수에 있어서는 도로 폭이 매우 넓은 4차로 이상인 경우가 대형 교통사고 충돌 형태별에 있어 차량 외 충돌과 전도, 전복, 추락 등에 영향을 미치고 있는 것으로 나타났다.

항 목	법 주	빈 도	수 량 화 값	-2.0	-1.0	0	1.0	2.0	법 위	편상관계수
법규위반 유 형	1.신호위반	(2)	-0.69682						2.05687	0.77573 (1)
	2.중앙선 침범	(4)	-1.50292							
	3.제한속도위반	(21)	0.32626							
	4.앞지르기방법, 금지위반	(1)	0.55395							
사고발생 시간대별	1.07:00~08:59	(3)	1.62432						2.11484	0.76746 (2)
	2.09:00~16:59	(9)	0.10639							
	3.17:00~19:59	(4)	0.01396							
	4.20:00~06:59	(12)	-0.49052							
총차로수	1.1차로	(1)	-0.73396						0.88865	0.39643 (6)
	2.2차로	(8)	-0.27565							
	3.4차로 이상	(19)	0.15469							
노 면 포장상태	1.아스팔트	(26)	-0.08558						1.19816	0.45849 (5)
	2.콘크리트	(2)	1.11258							
도 로 중앙부 형 태	1.중앙분리대	(8)	-0.32892						1.31983	0.54428 (4)
	2.황색실선	(16)	-0.04512							
	3.황색점선	(1)	0.99092							
	4.철길, 없음	(3)	0.78743							
타 차량 제공요인	1.신호대기, 주·정차, 일시정차 등	(6)	-1.14093						1.45209	0.70771 (3)
	2.없 음	(22)	0.31116							
Outside variable	1.차량 충돌	(6)	-1.51925						Eta-square (Correlation ratio) 0.80411	
	2.차량 외 충돌	(15)	0.09230							
	3.전도, 전복, 추락 등	(7)	1.10444							

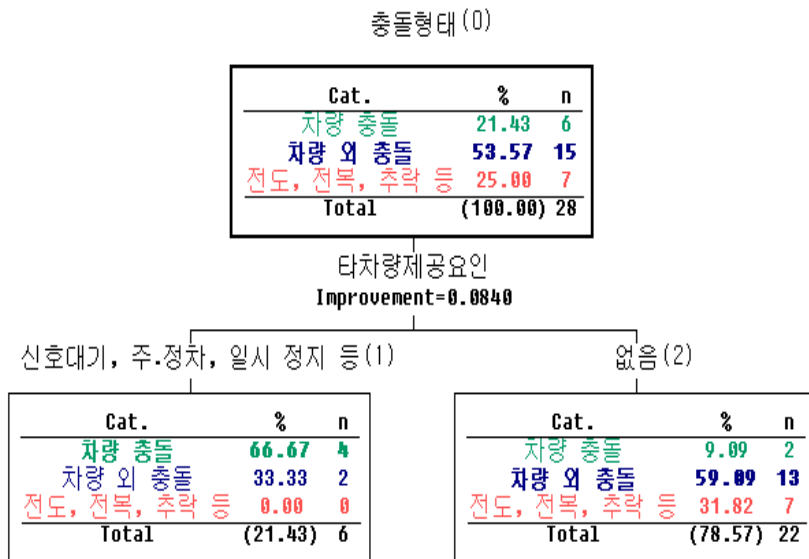
<그림 4-12> 수량화 이론 II류에 의한 충돌 형태별 판별모델(제3집단)

(2). 분류모델

다음으로는 전체 249개소 중 3집단에 속하는 28개소의 관찰대상과 1개의 목표변수 그리고 6개의 예측변수를 사용하여 분류모델을 구축하였는바, 그 모델은 <그림 4-13>과 같이 나타났다.

이를 보면 충돌 형태별에 대한 영향요인 분류모델에서는 예측변수로 사고발생 시간대별의 한 가지 변수만 채택되었다. 이 예측변수로 분리했을 때 분산의 감소 정도는 0.0840(Improvement = 0.0840)로 전체 분산을 8.4% 정도 줄여주고 있다.

목표변수인 충돌 형태별의 빈도에서는 <그림 4-13>과 같이, “차량 외 충돌”, “전도, 전복, 추락 등”, “차량 충돌”의 순으로 나타났으며, 특히 “차량 외 충돌”인 경우는 “차량 충돌”의 경우에 비해 2.5배 정도 더 높게 나타났다.



<그림 4-13> 충돌 형태별 영향요인 분류모델(3집단)

주) ()안의 숫자는 마디 번호임

목표변수의 오분률은 <표 4-22>와 같이, 위험 추정치(또는 예측 위험)가 0.392857로, 이는 자식마디에서의 집단 내 분산을 의미하는 것으로 전체 분산에서 각 마디의 분산 증분을 빼준 것이 0.392857인 것이다. 이러한 수치는 작을수록 더 선호되는데, 전체 분산이 0.464286임에 비하여 마디 내 분산은 0.392857인 것이다. 뿌리마디만 있을 경우와 회귀나무가 형성된 후의 분산 비율은 $0.392857 / 0.464286 = 0.84$ 로서, 뿌리마디만 있을 경우보다 16%(= 100% - 84%) 정도가 회귀나무에 의하여 추가적으로 설명되어지는 분산의 비율임을 뜻한다.

<표 4-22> 위험 도표(3집단)

Misclassification Matrix					
		Actual Category			
		차량 충돌	차량 외 충돌	전도, 전복, 추락 등	Total
Predicted Category	차량 충돌	4	2	0	6
	차량 외 충돌	2	13	7	22
	전도, 전복, 추락 등	0	0	0	0
	Total	6	15	7	28
Resubstitution					
Risk Estimate		0.392857			
SE of Risk Estimate		0.0922962			

그리고 각 마디에 대한 타당성을 평가하기 위하여, 해당 마디에서 자료가 얼마나 잘 분류되었는지를 살펴보기 위한 이익 도표는 <표 4-23>과 같다.

<표 4-23> 이익 도표(3집단)

Gain Summary												
Target variable : 충돌형태							Target category : 전도, 전복, 추락 등					
Node-by-Node							Cumulative					
Node	Node:n	Node:%	Resp:n	Resp:%	Gain(%)	Index(%)	Node:n	Node:%	Resp:n	Resp:%	Gain(%)	Index(%)
2	22	78.57	7	100.00	31.81818	127.27273	22	78.57	7	100.00	31.81818	127.27273
1	6	21.43	0	0.00	0.00000	0.00000	28	100.00	7	100.00	25.00000	100.00000

<표 4-23>의 이익 도표에 의하면, 충돌 형태의 경우 첫 번째 마디(2번 마디)는 자료수가 22개소 중 응답 대상수가 7개소로 응답률은 81.82%를 나타내고

있다. 다음으로 이 마디에 대한 응답의 비율이 전체 응답의 비율과 얼마나 비교되는지를 나타내는 인덱스 점수를 보면, 127.27%로, 이는 이 마디에 응답한 비율이 전체 자료에 대한 응답 비율보다 약간 높다는 것을 의미한다.

이와 같은 방식으로 이후의 마디에 대해서도 같은 방식으로 해석을 해 보면, 두 번째 마디(1번 마디)의 인덱스 점수는 0.00%인 바, 이에 해당하는 마디는 전체 자료보다 낮은 응답률을 갖는다는 것을 의미한다. 따라서 첫 번째 행과 두 번째 행(2번 마디와 1번 마디)사이에 교차점이 있음을 알 수 있다.

또한 누적 통계량을 살펴보면, 첫 번째 마디(2번 마디)를 취한다면, 전체 자료의 78.57%(Node:%, 22/28)에 해당하는 자료를 접촉하여 전체 응답 대상의 100.00% (Resp:%, 7/7)의 응답률을 얻을 수 있게 된다. 여기서 추가적으로 두 번째 마디(1번 마디)를 포함한다면, 전체 자료의 100.00%(28/28)를 접촉하여 전체 응답 대상의 100.00%에 해당하는 응답률을 얻게 되는 것이다.

그러나 이 단계는 교차점을 지나면서부터 전체 자료의 접촉 비율과 포함 마디까지의 응답률과의 차이가 줄어들게 된다. 여기서 추가적으로 두 번째 마디(1번 마디)를 포함하면 응답률은 100.00%이지만, 전체 자료의 21.42%(6/28)를 별도로 더 접촉해야 함을 알 수 있다. 하지만 100.00%의 반응 비율을 예측하기를 원한다면, 첫 번째 마디(2번 마디)까지를 목표로 삼으면 될 것이다.

이와 같은 이익 값을 이용하여 마디들이 어떻게 정의되는지를 나무지도 화면인 <그림 4-13>에서 보면, 충돌형태별에 있어 차량 외 충돌이 발생했던 15개소 중 타 차량 제공 요인이 없는 경우가 13개소 속하는 것으로 나타났다.

4. 제4집단

(1). 판별모델

끝으로 제4집단에 대해 외적 변수인 충돌 형태를 <표 4-24>과 같이 빈도분석을 한 결과 순위를 보면, 차량 충돌, 전도, 전복, 추락 등, 차량 외 충돌의 순으로 나타났다.

<표 4-24> 충돌 형태별 현황(제4집단)

충돌 형태	사고 건수(건)	비율(%)
차량 충돌	90	90.9
차량 외 충돌	4	4.0
전도, 전복, 추락 등	5	5.1
계	99	100.0

그리고 법규위반 유형의 경우 주취운전, 안전운전 의무 불이행, 철길 건널목 통과방법 위반, 무면허운전, 보행자 보호 의무 불이행 등이 제외되었으며, 타 차량 제공요인에서는 신호대기, 주·정차, 일시정지 등과 타 차량 제공 요인 없음의 경우가 제외되었다. 따라서 이들 6가지 설명변수와 이에 따른 유형들을 설정하여 충돌 형태별 유형에 대해 수량화 이론 II류 분석을 행하였는바, 구축된 판별모델의 판별 범위는 <표 4-25>와 <그림 4-14>에서와 같이 나타났다.

<표 4-25> 충돌 형태별 외적 변수의 판별범위(제4집단)

충돌 형태	집단의 중심값
차량 충돌	-0.170 < (1) < -0.033
차량 외 충돌	0.586 < (2) < 1.094
전도, 전복, 추락 등	-0.225 < (3) < 2.591



<그림 4-14> 충돌 형태별 외적 변수의 Data Score Histogram(제4집단)

또한 적중률을 구하여 보았는바, <표 4-26>에서와 같이, 충돌 형태별에 있어서는 전도, 전복, 추락 등이 100.0%로 가장 높게 나타났고, 다음으로는 차량 충돌이 75.6%, 차량 외 충돌이 25.0%의 순으로 나타났으며, 전체 적중률도 74.7%로 비교적 높게 나타났다.

<표 4-26> 충돌 형태별 판별모델의 판별 결과(제4집단)

판별된 유형 미래의 유형	차량 충돌	차량 외 충돌	전도, 전복, 추락 등	계	적중율
차량 충돌	68 (75.6%)	11 (12.2%)	11 (12.2%)	90 (100.0%)	75.6%
차량 외 충돌	1 (25.0%)	1 (25.0%)	2 (50.0%)	4 (100.0%)	25.0%
전도, 전복, 추락 등	-	-	5 (100.0%)	5 (100.0%)	100.0%
계	69	12	18	99	74.7%

그리고 상관비를 구하여 보았는바, <그림 4-15>의 판별모델에서 보는 바와 같이, 제1축이 0.35493, 제2축이 0.14926으로 나타났다. 이 중 제1축에 대한 판별모델의 상관비 0.35493에서 대형 교통사고 발생 시의 충돌 형태인 외적 변수의 구분 요인을 가장 잘 설명하고 있는 판별모델로서 유효하다고 판단되었다.

다음으로 편상관계수가 높은 순서로 볼 때, 범규위반 유형, 노면 포장 상태, 도로 중앙부 형태, 총 차로수, 사고 발생 시간대별, 타 차량 제공 요인의 순으로 나타나 인적 요인과 도로·환경적 요인이 순차적으로 영향을 미치고 있음을 알 수 있었다.

특히 이 판별모델에서 외적 변수인 대형 교통사고 발생 시의 충돌 형태 유형 점수(Category Score)를 보면, 전도, 전복, 추락 등 > 차량 외 충돌 > 차량 충돌로 나타났는바, 이를 토대로 하여, 보다 구체적으로 살펴보면 다음과 같은 특징이 나타났다.

1). 인적 요인 중 범규 위반 유형에서 신호위반, 중앙선 침범, 제한속도 위반을 한 경우와 타 차량 제공 요인에서는 대향 내지 측면에서 직진, 회전 중인 경우가 대형 교통사고 충돌 형태별에 있어 차량 충돌과 차량 외 충돌에 영향을 미치고 있는 것으로 나타났다.

반면에 법규 위반 유형에서는 앞지르기 방법 및 금지위반을 한 경우와 차로 위반을 한 경우, 타 차량 제공 요인에서는 동일 방향 직진, 회전 중인 경우가 대형 교통사고 충돌 형태별에 있어 전도, 전복, 추락 등에 영향을 미치고 있는 것으로 나타났다.

2). 도로·환경적 요인 중에는 노면 포장 상태에서는 아스팔트 포장인 경우, 도로 중앙부 형태에서는 중앙분리대가 설치되어 있거나 황색점선인 경우, 총 차로 수에서는 도로 폭이 매우 넓은 4차로 이상인 경우 그리고 사고 발생 시간대에서는 일상 시간대인 09시부터 16시 59분 사이인 경우가 대형 교통사고 충돌 형태별에 있어 차량 충돌과 차량 외 충돌에 영향을 미치고 있는 것으로 나타났다.

반면에 노면 포장상태에서는 콘크리트 포장인 경우와 비포장인 경우, 도로 중앙부 상태에서는 황색실선인 경우, 철길이 지나가는 경우와 도로 중앙부가 없는 경우, 총 차로수에 있어서는 도로 폭이 매우 좁은 1차로인 경우와 2차로, 사고 발생 시간대별에 있어 출근 시간대인 07시부터 08시 59분 사이, 퇴근 시간대인 17시부터 19시 59분 사이인 경우와 야간 시간대인 20시부터 익일 06시 59분 사이가 대형 교통사고 충돌 형태별에 있어 전도, 전복, 추락 등에 영향을 미치고 있는 것으로 나타났다.

항 목	법 주	빈 도	수 량 화 값	-2.0	-1.0	0	1.0	2.0	3.0	4.0	법 위	편상관계수
법규위반 유 형	1.신호위반	(11)	-0.67711								4.40287	0.44578 (1)
	2.중앙선 침범	(53)	-0.35916									
	3.제한속도위반	(20)	-0.00855									
	4.앞지르기방법, 금지위반	(13)	1.47719									
	5.차로위반	(2)	3.72576									
사고발생 시간대별	1.07:00~08:59	(21)	0.01845								0.26088	0.07757 (5)
	2.09:00~16:59	(38)	-0.13394									
	3.17:00~19:59	(18)	0.10609									
	4.20:00~06:59	(22)	0.12694									
총차로수	1.1차로	(4)	0.31992								0.49735	0.12690 (4)
	2.2차로	(46)	0.16118									
	3.4차로 이상	(49)	-0.17743									
노 면 포장상태	1.아스팔트	(91)	-0.09927								3.65111	0.35560 (2)
	2.콘크리트	(6)	0.32158									
	3.비포장	(2)	3.55184									
도 로 중앙부 형 태	1.중앙분리대	(14)	-0.35695								1.39203	0.18673 (3)
	2.황색실선	(74)	0.03625									
	3.황색점선	(4)	-0.67542									
	4.철길, 없음	(7)	0.71661									
타 차량 제공요인	1.동일방향 직진, 회전 중	(21)	0.07430								0.11630	0.02452 (6)
	2.대향 직진, 회전 중	(66)	-0.01601									
	3.측면에서 직진, 회전 중	(12)	-0.04200									
Outside variable	1.차량 충돌	(90)	-0.15605	Eta-square (Correlation ratio) 0.35493								
	2.차량 외 충돌	(4)	-0.32290									
	3.전도, 전복, 추락 등	(5)	2.55066									

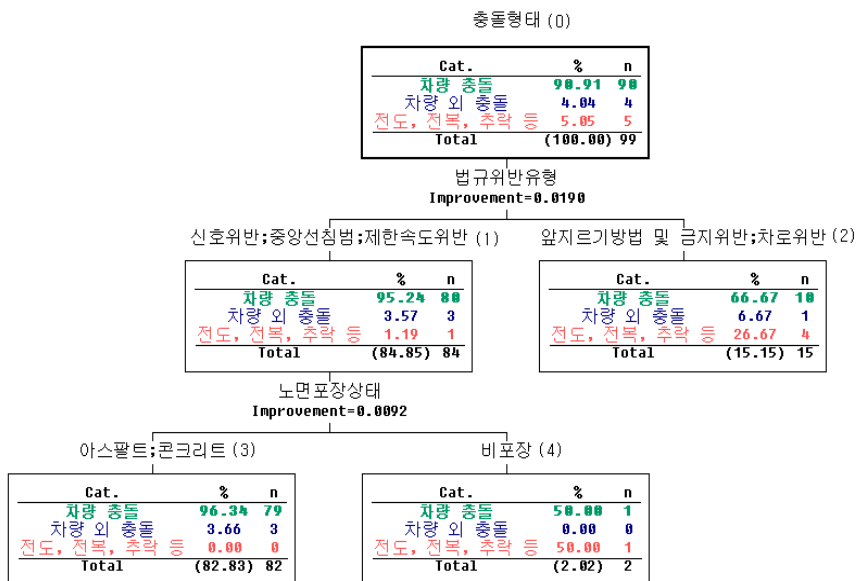
<그림 4-15> 수량화 이론 II류에 의한 충돌 형태별 판별모델(제4집단)

(2). 분류모델

끝으로 전체 249개소 중 4집단에 속하는 99개소의 관찰대상과 1개의 목표변수 그리고 6개의 예측변수를 사용하여 분류모델을 구축하였는바, 그 모델은 <그림 4-16>과 같이 나타났다.

이를 보면 충돌 형태별에 대한 영향요인 분류모델에서는 예측변수로 법규위반 유형, 도로 중앙부 형태의 두 가지 변수가 채택되었다. 이 예측변수로 분리했을 때 분산의 감소 정도는 0.0190(Improvement = 0.0190)로 전체 분산을 1.9% 정도 줄여주고 있다.

목표변수인 충돌 형태별의 빈도에서는 <그림 4-16>과 같이, “차량 충돌”, “전도, 전복, 추락 등”, “차량 외 충돌”의 순으로 나타났으며, 특히 “차량 충돌”인 경우는 “차량 외 충돌”의 경우에 비해 22.5배 정도 더 높게 나타났다.



<그림 4-16> 충돌 형태별 영향요인 분류모델(4집단)

주) ()안의 숫자는 마디 번호임

목표변수의 오분률은 <표 4-27>과 같이, 위험 추정치(또는 예측 위험)가 0.0909091로, 이는 자식마디에서의 집단 내 분산을 의미하는 것으로 전체 분산에서 각 마디의 분산 증분을 빼준 것이 0.0909091인 것이다. 이러한 수치는 작을수록 더 선호되는데, 전체 분산이 0.0909091이고 마디 내 분산도 0.0909091인 것이다. 따라서 뿌리마디만 있을 경우와 회귀나무가 형성된 후의 분산 비율은 $0.0909091 / 0.0909091 = 1$ 로서, 뿌리마디만 있을 경우와 회귀나무에 의하여 추가적으로 설명되어지는 분산 간에는 차이가 전혀 없는 것으로 나타났다.

<표 4-27> 위험 도표(4집단)

Misclassification Matrix					
		Actual Category			
		차량 충돌	차량 외 충돌	전도, 전복, 추락 등	Total
Predicted Category	차량 충돌	90	4	5	99
	차량 외 충돌	0	0	0	0
	전도, 전복, 추락 등	0	0	0	0
	Total	90	4	5	99
Resubstitution					
Risk Estimate		0.0909091			
SE of Risk Estimate		0.0288928			

그리고 각 마디에 대한 타당성을 평가하기 위하여, 해당 마디에서 자료가 얼마나 잘 분류되었는지를 살펴보기 위한 이익 도표는 <표 4-28>과 같다.

<표 4-28> 이익 도표(4집단)

Gain Summary												
Target variable : 충돌형태							Target category : 전도, 전복, 추락 등					
Node-by-Node							Cumulative					
Node	Node:n	Node:%	Resp:n	Resp:%	Gain(%)	Index(%)	Node:n	Node:%	Resp:n	Resp:%	Gain(%)	Index(%)
4	2	2.02	1	20.00	50.00000	990.00000	2	2.02	1	20.00	50.00000	990.00000
2	15	15.15	4	80.00	26.66667	528.00000	17	17.17	5	100.00	29.41176	582.35294
3	82	82.83	0	0.00	0.00000	0.00000	99	100.00	5	100.00	5.05051	100.00000

<표 4-28>의 이익 도표에 의하면, 충돌 형태의 경우 첫 번째 마디(4번 마디)는 자료수가 2개소 중 응답 대상수가 1개소로 응답률은 50%를 나타내고 있다. 다음으로 이 마디에 대한 응답의 비율이 전체 응답의 비율과 얼마나 비교되는지를 나타내는 인덱스 점수를 보면, 990.00%로, 이는 이 마디에 응답한 비율이 전체 자료에 대한 응답 비율보다 9.9배정도 높다는 것을 의미한다.

이와 같은 방식으로 이후의 마디에 대해서도 같은 방식으로 해석을 해 보면, 세 번째 마디(3번 마디)까지의 인덱스 점수는 0.00%인 바, 이에 해당하는 마디는 전체 자료보다 낮은 응답률을 갖는다는 것을 의미한다. 따라서 두 번째 마디와 세 번째 마디(2번 마디와 3번 마디)사이에 교차점이 있음을 알 수 있다.

또한 누적 통계량을 살펴보면, 첫 번째 마디(4번 마디)를 취한다면, 전체 자료의 2.02%(Node:%, 2/99)에 해당하는 자료를 접촉하여 전체 응답 대상의 20.00%(Resp:%, 1/5)의 응답률을 얻을 수 있게 된다. 여기서 추가적으로 두 번째 마디(2번 마디)를 포함한다면, 전체 자료의 17.17%(17/99)를 접촉하여 전체 응답 대상의 100.00%에 해당하는 응답률을 얻게 되는 것이다.

그러나 이 단계는 교차점을 지나면서부터 전체 자료의 접촉 비율과 포함 마디까지의 응답률과의 차이가 줄어들게 된다. 여기서 추가적으로 세 번째 마디(3번 마디)를 포함하면 응답률은 100.00%이지만, 전체 자료의 89.89%(82/99)를 별도로 더 접촉해야 함을 알 수 있다. 하지만 100.00%의 반응 비율을 예측하기를 원한다면, 두 번째 마디(2번 마디)까지를 목표로 삼으면 될 것이다.

이와 같은 이익 값을 이용하여 마디 들이 어떻게 정의되는지를 나무지도 화면인 <그림 4-16>에서 보면, 충돌형태별에 있어 차량 충돌이 발생했던 90개소 중 범규위반 유형 요인이 신호 위반, 중앙선 침범, 제한속도 위반의 경우는 80개소, 앞지르기 방법 및 금지 위반, 차로 위반의 경우는 10개소가 속하는 것으로 나타났으며, 이 중 신호 위반, 중앙선 침범, 제한속도 위반의 경우는 다시금 노면 포장상태에서 아스팔트, 콘크리트인 경우가 79개소가 속하는 것으로 나타났다.

제4절 발생지점 유형별 특성과 교통사고 예방 대책

1. 판별모델 결과

먼저 발생지점의 유형화에 따른 4개 집단의 판별모델에 있어 외적 변수인 충돌 형태별에 대한 영향요인들을 종합하여 보면, 다음의 <표 4-29>와 같다.

먼저 1집단에 속하는 87개 지점들은 주취운전 또는 철길 건널목 통과방법 위반을 한 상태에서, 진행 방향 전방에 타 차량이 동일 방향 혹은 측면에서 직진 또는 회전 중이었거나 신호 대기, 주·정차, 일시 정지 중이어서 이를 피하는 과정에서 정면충돌이나 추돌의 형태로 차 대 차 교통사고가 많이 발생한 것으로 나타났다.

아울러 교통정체가 시작되는 출근 시작 시간대부터 퇴근 시작 시간대 직전인 07시부터 16시 59분 사이 또는 시인성의 확보가 어려워 교통사고의 위험성이 높은 야간 시간대의 20시부터 익일 06시 59분 사이에 차 대 차 교통사고가 많이 발생하였으며, 도로 폭이 넓어 운전 조건이 매우 좋은 4차로 이상, 중앙분리대가 있는 콘크리트 포장 도로 상에서도 차 대 차 교통사고가 많이 발생하는 것으로 나타났다.

이에 반해, 무면허운전 또는 보행자 보호 의무 불이행 등의 위반 등을 한 상태에서, 타 차량이 대향 직진 또는 회전 중이어서 이들 차량을 피하고자 한 상황에서, 퇴근 시간대인 17시부터 19시 59분 사이에, 앞지르기가 금지가 되는 황색 실선이 설치된 아스팔트 포장의 2차로 도로 상에서 단독으로 다른 고정 물체와 충돌하는 등의 차량 외 충돌 내지는 전도, 전복, 추락한 교통사고가 많이 발생한 것으로 나타났다.

또한 2집단에 속하는 35개 지점들은 일상의 시간대에 해당하는 09시부터 16시 59분 사이 혹은 야간 시간대인 20시부터 익일 06시 59분 사이에, 철길 건널목 통과방법 위반을 한 상황에서, 진행 방향 전방에서 타 차량이 측면에서 직진 또는 회전 중이어서 이를 피하는 과정에서 차 대 차 교통사고가 많이 발생한 것으로 나타났으며, 도로 폭이 좁아 차량 교행이 불가능한 1차로나 반대로 도로 폭이 넓어 운전 조건이 매우 좋은 4차로 이상의 아스

팔트 포장 도로 상에서도 차 대 차 교통사고가 많이 발생한 것으로 나타났다.

이에 반해, 주취운전을 하였거나, 무면허운전 또는 보행자 보호 의무 불이행 등의 위반을 한 상태에서, 타 차량이 동일 방향 또는 정반대로 대향에서 직진 또는 회전 중이어서 이를 피하고자 하는 상황에서 단독으로 다른 고정물체와 충돌하는 등의 차량 외 충돌 내지는 전도, 전복, 추락한 교통사고가 많이 발생한 것으로 나타났다.

더불어 교통 정체가 심각한 출근 시간대인 07시부터 08시 59분 사이 또는 퇴근 시간대인 17시부터 19시 59분 사이에, 콘크리트 포장이 되어 있거나 비포장 상태의 2차로 도로 상에서 단독으로 다른 고정물체와 충돌하는 등의 차량 외 충돌 내지는 전도, 전복, 추락한 교통사고가 많이 발생한 것으로 나타났다.

그리고 3집단에 속하는 28개 지점들은 시인성의 확보가 어려운 야간 시간대인 20시부터 익일 06시 59분 사이에, 신호위반 또는 중앙선 침범을 한 상태에서, 타 차량이 진행 방향 전방에서 신호 대기, 주·정차 또는 일시 정지 중이어서 이를 피하는 과정에서 차 대 차 교통사고가 많이 발생한 것으로 나타났으며, 도로 폭이 좁아 차량 교행이 어려운 1차로 또는 2차로의 아스팔트 포장 도로 상에서도 차 대 차 교통사고가 많이 발생한 것으로 나타났다.

이에 반해, 제한속도 위반 또는 앞지르기 방법 및 금지위반을 한 상태에서, 교통 정체가 심각한 출근 시작 시간대인 07시부터 퇴근 시간대 끝 날 무렵인 19시 59분 사이에 단독으로 다른 고정물체와 충돌하는 등의 차량 외 충돌 내지는 전도, 전복, 추락한 교통사고가 많이 발생한 것으로 나타났으며, 도로 중앙부에 황색 실선 또는 점선이 설치되어 있는 콘크리트 포장의 도로 폭이 넓어 운전 조건이 매우 좋은 4차로 이상 도로 상에서도 단독으로 다른 고정물체와 충돌하는 등의 차량 외 충돌 내지는 전도, 전복, 추락한 교통사고가 많이 발생한 것으로 나타났다.

끝으로 4집단에 속하는 99개 지점들은 신호위반, 중앙선 침범, 제한속도

위반 등을 한 상태에서, 교통 정체가 심각한 출근 시작 시간대인 07시부터 퇴근 시간대 시작 직전인 16시 59분 사이에 중앙 분리대가 설치되어 있는 도로 폭이 넓어 운전 조건이 매우 좋은 4차로 이상의 아스팔트 포장 도로 상에서 차 대 차 교통사고가 많이 발생한 것으로 나타났다.

이에 반해, 교통 정체가 심각한 퇴근 시간대인 17시부터 19시 59분 사이 또는 시인성의 확보가 어려운 야간 시간대인 17부터 19시 59분 사이에, 앞지르기 방법 및 금지 위반을 하였거나 또는 차로위반을 한 상태에서, 타 차량이 동일 방향으로 직진 내지는 회전 중이어서 이들 차량을 피하고자 하는 상황에서 단독으로 다른 고정 물체와 충돌하는 등의 차량 외 충돌 내지는 전도, 전복, 추락한 교통사고가 많이 발생한 것으로 나타났으며, 도로 중앙 부에 앞지르기가 금지되는 황색실선이 설치되어 있는 콘크리트 포장 또는 비포장의 도로 폭이 좁아 차량 교행이 어려운 1차로 또는 2차로의 도로 상에서도 단독으로 다른 고정 물체와 충돌하는 등의 차량 외 충돌 내지는 전도, 전복, 추락한 교통사고가 많이 발생한 것으로 나타났다.

<표 4-29> 발생지점 유형별 충돌 형태별 판별 특성

집단	요 인	항 목	의식 변수에의 영향 요인 (충돌 형태별)			제 외 범 주
			차량 충돌	차량 외 충돌	전도, 전복, 추락 등	
1 (87)	인 적	법규 위반	· 주취운전 · 절길건널목 통과방법 위반	· 주취운전 · 절길건널목 통과방법 위반	· 안전운전 의무 불이행 · 무면허운전, 보행자 보호 의무 불이행 등	· 신호위반 · 중앙선 침범 · 제한속도 위반 · 앞지르기 방법 및 금지 위반 · 차로위반
		타 차량 제공 요인	· 측면에서 직진, 회전 중 · 신호 대기, 주정차, 일시 정지 등	· 측면에서 직진, 회전 중 · 신호 대기, 주정차, 일시 정지 등	· 동일 방향 직진, 회전 중 · 대향 직진, 회전 중 · 없음	
	도 로 · 환경적	사고 발생 시간대별	· 09:00~16:59 · 20:00~06:59	· 09:00~16:59 · 20:00~06:59	· 07:00~08:59 · 17:00~19:59	
		충 차로수	· 4차로 이상	· 4차로 이상	· 1차로 · 2차로	
		노면 포장상태	· 콘크리트	· 콘크리트	· 아스팔트 · 비포장	
2 (35)	인 적	법규 위반		· 주취운전	· 안전운전 의무 불이행 · 무면허운전, 보행자 보호 의무 불이행 등	· 신호위반 · 중앙선 침범 · 제한속도 위반 · 앞지르기 방법 및 금지 위반 · 차로위반 · 절길건널목 통과방법 위반
		타 차량 제공 요인			· 동일 방향 직진, 회전 중 · 대향 직진, 회전 중 · 측면에서 직진, 회전 중 · 신호 대기, 주정차, 일시정지 등 · 없음	
	도 로 · 환경적	사고 발생 시간대별		· 17:00~19:59	· 07:00~08:59 · 09:00~16:59 · 20:00~06:59	
		충 차로수		· 1차로 · 2차로	· 4차로 이상	
		노면 포장상태		· 아스팔트	· 콘크리트	· 비포장
3 (28)	인 적	법규 위반	· 신호위반 · 중앙선 침범	· 제한속도 위반 · 앞지르기 방법 및 금지 위 반	· 제한속도 위반 · 앞지르기 방법 및 금지 위반	· 차로위반 · 주취운전 · 안전운전 의무 불이행 · 절길건널목 통과방법 위반 · 무면허운전, 보행자 보호 의무 불이행 등
		타 차량 제공 요인	· 신호 대기, 주정차, 일시 정지 등	· 없음	· 없음	· 동일 방향 직진, 회전 중 · 대향 직진, 회전 중 · 측면에서 직진, 회전 중
	도 로 · 환경적	사고 발생 시간대별	· 20:00~06:59	· 07:00~08:59 · 09:00~16:59 · 17:00~19:59	· 07:00~08:59 · 09:00~16:59 · 17:00~19:59	
		충 차로수	· 1차로 · 2차로	· 4차로 이상	· 4차로 이상	
		노면 포장상태	· 아스팔트	· 콘크리트	· 콘크리트	· 비포장
4 (99)	인 적	법규 위반	· 신호위반 · 중앙선 침범 · 제한속도 위반	· 신호위반 · 중앙선 침범 · 제한속도 위반	· 앞지르기 방법 및 금지 위반 · 차로위반	· 주취운전 · 안전운전 의무 불이행 · 절길건널목 통과방법 위반 · 차로위반 · 보행자 보호 의무 불이행 등
		타 차량 제공 요인	· 대향 직진, 회전 중 · 측면에서 직진, 회전 중	· 대향 직진, 회전 중 · 측면에서 직진, 회전 중	· 동일 방향 직진, 회전 중	· 신호 대기, 주정차, 일시정지 등 · 없음
	도 로 · 환경적	사고 발생 시간대별	· 09:00~16:59	· 09:00~16:59	· 07:00~08:59 · 17:00~19:59 · 20:00~06:59	
		충 차로수	· 4차로 이상	· 4차로 이상	· 1차로 · 2차로	
		노면 포장상태	· 아스팔트	· 아스팔트	· 콘크리트 · 비포장	
4 (99)	도 로 · 환경적	도로 중앙부 형태	· 중앙분리대 · 황색점선	· 중앙분리대 · 황색점선	· 황색점선 · 절길, 없음	

주 : 집단 항목 내 ()안의 숫자는 그 집단에 속하는 지점수를 의미함

이에 따른 집단별 교통사고 예방 대책을 제시하여 보면, 다음의 <표 3-30>과 같다.

먼저 1집단의 경우에는 차량 충돌 및 차량 외 충돌사고를 예방하기 위해서는 인적 요인에 대해 주취운전에 대한 단속 강화와 철길건널목 안전시설 보강이 필요하고, 도로·환경적 요인에 대해서는 시인성의 향상을 위한 교통안전시설의 개선이 필요한 것으로 나타났다.

반면에 전도, 전복, 추락 등의 교통사고를 예방하기 위해서는 인적 요인에 대해 횡단보도 등의 보행시설 정비에 필요한 것으로 나타났다.

또한 2집단의 경우에는 차량 외 충돌사고를 예방하기 위해서는 인적 요인에 대해 주취운전에 대한 단속 강화가 필요하고, 도로·환경적 요인에 대해서는 중앙분리대의 정비가 필요한 것으로 나타났다.

이에 반해, 전도, 전복, 추락 등의 교통사고를 예방하기 위해서는 인적 요인에 대해 횡단보도 등의 보행시설 정비가 필요하며, 도로·환경적 요인에 대해서는 시인성의 향상을 위한 교통안전시설의 개선과 중앙선의 정비가 강구되어야 할 것으로 나타났다.

그리고 3집단의 경우에는 차량 충돌사고를 예방하기 위해서는 인적 요인에 대해 신호등과 중앙선의 정비가 필요하고, 도로·환경적 요인에 대해서는 1집단과 같이 시인성의 향상을 위한 교통안전시설의 개선이 필요한 것으로 나타났다.

이에 반해, 차량 외 충돌, 전도, 전복, 추락 등의 교통사고를 예방하기 위해서는 인적 요인에 대해 과속 방지턱의 보강 및 신규 설치와 중앙선의 정비가 필요한 것으로 나타났다.

끝으로 4집단의 경우에는 차량 충돌 및 차량 외 충돌사고를 예방하기 위해서는 인적 요인에 대해 신호등과 중앙선의 정비와 과속 방지턱의 보강 및 신규 설치가 필요한 것으로 나타났다.

이에 반해, 전도, 전복, 추락 등의 교통사고를 예방하기 위해서는 인적 요인에 대해 중앙선 및 차선의 정비가 필요한 것으로 나타났으며, 도로·환경적 요인에 대해서는 시인성의 향상을 위한 교통안전시설의 개선이 필요한 것으로 나타났다.

<표 4-30> 판별모델에 따른 발생지점 유형별 교통사고 예방 대책

집단	요 인	예방 대책 (충돌 형태별)		
		차량 충돌	차량 외 충돌	전도, 전복, 추락 등
1 (87)	인 적	· 주취운전에 대한 단속 강화 · 철길건널목 안전시설 보강	· 주취운전에 대한 단속 강화 · 철길건널목 안전시설 보강	· 횡단보도 등 보행시설의 정비
	도로·환경적	· 시인성 향상을 위한 교통안전시설의 개선	· 시인성 향상을 위한 교통안전시설의 개선	
2 (35)	인 적		· 주취운전에 대한 단속 강화	· 횡단보도 등 보행시설의 정비
	도로·환경적		· 중앙분리대의 정비	· 시인성 향상을 위한 교통안전시설의 개선 · 중앙선의 정비
3 (28)	인 적	· 신호등과 중앙선의 정비	· 과속 방지턱의 보강, 설치 · 중앙선의 정비	· 과속 방지턱의 보강, 설치 · 중앙선의 정비
	도로·환경적	· 시인성 향상을 위한 교통안전시설의 개선		
4 (99)	인 적	· 신호등과 중앙선의 정비 · 과속 방지턱의 보강, 설치	· 신호등과 중앙선의 정비 · 과속 방지턱의 보강, 설치	· 중앙선 및 차선의 정비
	도로·환경적			· 시인성 향상을 위한 교통안전시설의 개선

주 : 집단 항목 내 ()안의 숫자는 그 집단에 속하는 지점수를 의미함

2. 분류모델 결과

다음으로 발생지점의 유형화에 따른 4개 집단의 분류모델에 있어 외적 변수인 충돌 형태별에 대한 영향요인들을 종합하여 보면, 다음의 <표 4-31>과 같다.

먼저 1집단에 속하는 87개 지점들은 차량 충돌사고의 경우 주취운전, 철길건널목 통과방법 위반, 무면허 운전, 보행자 보호 의무 불이행 등을 한 상태에서, 진행 방향 전방에 타 차량이 동일 방향과 대향 혹은 측면에서 직진 또는 회전 중이어서 이를 피하는 과정에서 정면이나 측면 충돌의 형태로 많이 발생한 것으로 나타났다.

아울러 도로 폭이 넓어 운전 조건이 매우 좋은 4차로 이상, 아스팔트나 콘크리트 포장 도로 상에서도 많이 발생하는 것으로 나타났다.

또한 2집단에 속하는 35개 지점들은 차량 외 충돌사고의 경우 주취운전을 한 경우가 많이 발생하는 것으로 나타났다.

그리고 3집단에 속하는 28개 지점들은 차량 충돌사고의 경우는 타 차량이 진행 방향 전방에서 신호 대기, 주·정차 또는 일시 정지 중인 경우에 많이 발생하고 있는 것으로 나타났다.

끝으로 4집단에 속하는 99개 지점들은 차량 충돌사고의 경우 법규 위반 유형에 있어서는 신호위반, 중앙선 침범, 제한속도 위반 등을 한 경우 노면 포장상태에 있어서는 아스팔트와 콘크리트 포장인 경우에 많이 발생하는 것으로 나타났다.

<표 4-31> 발생지점 유형별 충돌 형태별 분류 특성

집단	요 인	항 목	C&RT에 의한 이차분리				제 외 범 주
			차량 충돌	차량 외 충돌	전도, 전복, 추락 등	비 분리	
1 (87)	인 적	법규 위반(4)	· 주위운전 · 절김 · 건널목 · 통과방법 위반 · 부면하 · 운전, · 보행자 · 보호 · 외부 · 불이행 등	· 안전운전 · 외부 · 불이행			· 신호위반 · 중앙선 침범 · 제한속도 위반 · 앞지르기 방법 및 · 금지 위반 · 차로위반
		타 차량 제공 요인 (2)	· 동일 · 방향 · 직진, 회 · 전 중 · 대향 · 직진, 회 · 전 중 · 측면에서 · 직진, 회 · 전 중	· 신호 · 대기, · 주정차, · 일시 · 정지 등 · 없음			
	도로·환경적	사고 발생 시간대별				· 07:00 · ~08:59 · 09:00 · ~16:59 · 17:00 · ~19:59 · 20:00 · ~06:59	
		충 차로수(3)	· 4차로 · 이상	· 1차로 · 2차로			
		노면 포장상태(1)	· 아스팔트 · 콘크리트	· 비포장	· 비포장		
		도로 중앙부 형태				· 중앙분리대 · 황색실선 · 황색점선 · 절김, 없음	
2 (35)	인 적	법규 위반(1)		· 주위운전	· 안전운전 · 외부 · 불이행 · 부면하 · 운전, · 보행자 · 보호 · 외부 · 불이행 등		· 신호위반 · 중앙선 침범 · 제한속도 위반 · 앞지르기 방법 및 · 금지 위반 · 차로위반 · 횡단 건널목 · 통과방법 위반
		사고 발생 시간대별				· 07:00~08:59 · 09:00~16:59 · 17:00~19:59 · 20:00~06:59	
	도로·환경적	충 차로수				· 1차로 · 2차로 · 4차로 이상	
		노면 포장상태				· 아스팔트 · 콘크리트 · 비포장	
		도로 중앙부 형태				· 중앙분리대 · 황색실선 · 황색점선 · 절김, 없음	
		도로 중앙부 형태				· 중앙분리대 · 황색실선 · 황색점선 · 절김, 없음	
3 (28)	인 적	법규 위반				· 신호위반 · 주위운전 · 안전운전 · 외부 · 불이행 · 부면하 · 운전, · 보행자 · 보호 · 외부 · 불이행 등	· 차로위반 · 주위운전 · 안전운전 · 외부 · 불이행 · 건널목 · 통과방법 위반 · 부면하 · 운전, · 보행자 · 보호 · 외부 · 불이행 등
		타 차량 제공 요인 (1)	· 신호 대기, 주정차, · 일시 정지 등	· 없음			· 동일 방향 직진, · 회전 중 · 대향 직진, 회전 중 · 측면에서 직진, · 회전 중
	도로·환경적	사고 발생 시간대별				· 07:00~08:59 · 09:00~16:59 · 17:00~19:59 · 20:00~06:59	
		충 차로수				· 1차로 · 2차로 · 4차로 이상	
		노면 포장상태				· 아스팔트 · 콘크리트	· 비포장
		도로 중앙부 형태				· 중앙분리대 · 황색실선 · 황색점선 · 절김, 없음	

<계속>

집단	요 인	항 목	C&RT에 의한 이점분리				제 외 범 주
			차량 충돌	차량 외 충돌	전도, 전복, 추락 등	비 분리	
4 (99)	인 적	법규 위반(1)	· 신호위반 · 중앙선 침범 · 제한속도 위반	· 앞지르기 방법 및 금지 위반 · 차로위반			· 주취운전 · 안전운전 의무 불이행 · 횡단보도 건널목 통과방법 위반 · 부면허운전, 보행자 보호 의무 불이행 등
		타 차량 제동 요인				· 동일 방향 직진, 회전 중 · 대향 직진, 회전 중 · 측면에서 직진, 회전 중	· 신호 대기, 주·정차, 임시 정지 등 없음
	도로·환경적	사고 발생 시간대별				· 07:00~08:59 · 09:00~16:59 · 17:00~19:59 · 20:00~06:59	
		중 차로수				· 1차로 · 2차로 · 4차로 이상	
		노면 포장상태(2)	· 아스팔트 · 콘크리트	· 비포장	비포장		
		도로 중앙부 형태				· 중앙분리대 · 황색실선 · 황색점선 · 절연, 없음	

주 : 집단 내 ()안의 숫자는 그 집단에 속하는 지점수를 의미하며,

항목 내 ()안의 숫자는 분리 순서 숫자임

이에 따른 집단별 교통사고 예방 대책을 제시하여 보면, 다음의 <표 4-32>와 같다.

먼저 1집단의 경우에는 차량 충돌사고를 예방하기 위해서 인적 요인에 대해서는 주취운전에 대한 단속 강화와 철길건널목 안전시설의 보강이, 도로·환경적 요인에 대해서는 도로 조건이 좋은데 따른 교통안전시설의 보강이 필요한 것으로 나타났다.

또한 2집단의 경우에는 차량 외 충돌사고를 예방하기 위해서 인적 요인에 대해 주취운전에 대한 단속 강화가 필요한 것으로 나타났다.

그리고 3집단의 경우에는 차량 충돌사고를 예방하기 위해서 인적 요인에 대해 신호등 시설의 정비와 주·정차 위반 단속의 강화도 필요한 것으로 나타났다.

끝으로 4집단의 경우에는 차량 충돌사고를 예방하기 위해서 인적 요인에 대해서는 신호등과 중앙선의 정비, 과속 방지턱의 보강, 설치가가 필요하며, 도로·환경적 요인에 대해서는 도로 조건이 좋은데 따른 교통안전시설의 보강이 필요한 것으로 나타났다.

<표 4-32> 분류모델에 따른 발생지점 유형별 교통사고 예방 대책

집단	요 인	예방 대책 (충돌 형태별)		
		차량 충돌	차량 외 충돌	전도, 전복, 추락 등
1 (87)	인 적	· 주취운전에 대한 단속 강화 · 철길 건널목 안전시설 보강		
	도로·환경적	· 도로 조건이 좋은데 따른 교통안전시설의 보강		· 도로 노면 포장
2 (35)	인 적		· 주취운전에 대한 단속 강화	
	도로·환경적			
3 (28)	인 적	· 신호등 시설의 정비 · 주·정차 위반 단속 강화		
	도로·환경적			
4 (99)	인 적	· 신호등과 중앙선의 정비 · 과속 방지턱의 보강, 설치		
	도로·환경적	· 도로 조건이 좋은데 따른 교통안전시설의 보강		· 도로 노면 포장

주 : 집단 내 ()안의 숫자는 그 집단에 속하는 지점수를 의미함

3. 판별, 분류모델의 종합적 결과

이상으로 판별모델과 분류모델의 구축을 통해서 분석된 결과를 종합하여 보면, 다음의 <표 4-33>에서와 같이, 발생지점 유형별 교통사고 예방 대책을 중점적으로 우선 시행되어야 할 대책과 이를 보완해 나가야 할 대책으로 구분하여 제시할 수 있었다.

<표 4-33> 발생지점 유형별 교통사고 예방 대책

집단	요 인	예방 대책 (충돌 형태별)					
		차량 충돌		차량 외 충돌		전도, 전복, 추락 등	
		중점 대책	보완 대책	중점 대책	보완 대책	중점 대책	보완 대책
1 (87)	인 적	· 주취운전에 대한 단속 강화 · 철길건널목 안전시설 보강			· 주취운전에 대한 단속 강화 · 철길건널목 안전시설 보강		· 횡단보도 등 보행시설의 정비
	도로·환경적	· 도로 조건이 좋은데 따른 교통안전시설의 보강	· 시인성 향상을 위한 교통안전시설의 개선		· 시인성 향상을 위한 교통안전시설의 개선	· 도로 노면 포장	
2 (35)	인 적			· 주취운전에 대한 단속 강화			· 횡단보도 등 보행시설의 정비
	도로·환경적				· 중앙분리대의 정비		· 시인성 향상을 위한 교통안전시설의 개선 · 중앙선의 정비
3 (28)	인 적	· 신호등 시설의 정비 · 주·정차 위반 단속 강화	· 중앙선의 정비		· 과속 방지턱의 보강, 설치 · 중앙선의 정비		· 과속 방지턱의 보강, 설치 · 중앙선의 정비
	도로·환경적		· 시인성 향상을 위한 교통안전시설의 개선				
4 (99)	인 적	· 신호등과 중앙선의 정비 · 과속 방지턱의 보강, 설치			· 신호등과 중앙선의 정비 · 과속 방지턱의 보강, 설치		· 중앙선 및 차선의 정비
	도로·환경적	· 도로 조건이 좋은데 따른 교통안전시설의 보강				· 도로 노면 포장	· 시인성 향상을 위한 교통안전시설의 개선

주 : 집단 항목 내 ()안의 숫자는 그 집단에 속하는 지점수를 의미함

먼저 1집단의 경우에는 차량 충돌사고를 예방하기 위한 중점적으로 시행해야 할 대책으로 인적 요인과 관련해서는 주취운전에 대한 단속 강화와 철길건널목 안전시설의 보강이, 도로·환경적 요인과 관련하여서는 도로 조건이 좋은데 따른

교통안전시설의 보강이 필요한 것으로 나타났으며, 추후 보완해야 할 대책으로는 도로·환경적 요인과 관련하여서는 시인성 향상을 위한 교통안전 시설의 개선이 필요한 것으로 나타났다.

또한 차량 외 충돌의 경우는 보완 대책으로서 인적 요인과 관련하여 주취운전에 대한 단속 강화와 철길건널목 안전시설의 보강이, 도로·환경적 요인과 관련하여서는 시인성 향상을 위한 교통안전시설의 개선이 필요한 것으로 나타났으며, 전도, 전복, 추락 등의 교통사고를 예방하기 위한 중점 시행 대책으로는 도로·환경적 요인과 관련하여 도로 노면 포장 상태가 불량한 도로의 포장이 필요한 것으로 나타났으나, 보완 대책에서는 인적 요인과 관련하여 횡단보도 등 보행시설의 정비에 필요한 것으로 나타났다.

2집단의 경우에는 차량 외 충돌의 경우 중점 대책으로 인적 요인과 관련하여 주취운전에 대한 단속 강화가 필요하며, 보완 대책으로는 도로·환경적 요인과 관련하여서 중앙분리대의 정비가 필요한 것으로 나타났으며, 전도, 전복, 추락 등의 교통사고를 예방하기 위해서는 보완 대책으로는 횡단보도 등 보행시설의 정비가 필요한 것으로 나타났으며, 도로·환경적 요인과 관련하여서는 시인성 향상을 위한 교통안전시설의 개선과 중앙선의 정비가 필요한 것으로 나타났다.

그리고 3집단의 경우에는 차량 충돌사고를 예방하기 위하여 인적 요인과 관련한 중점 대책으로는 신호등 시설의 정비가 필요하고, 주·정차 위반 단속의 강화도 필요한 것으로 나타났으며, 보완 대책으로는 중앙선의 정비가 필요한 것으로 나타났고, 차량 외 충돌 사고와 전도, 전복, 추락 등의 교통사고를 예방하기 위한 인적 요인과 관련한 보완 대책으로는 과속 방지턱의 보강, 설치와 중앙선의 정비가 필요한 것으로 나타났다.

끝으로 4집단의 경우에는 차량 충돌사고를 예방하기 위한 중점 대책으로는 인적 요인과 관련하여서는 신호등과 중앙선의 정비, 과속 방지턱의 보강, 설치가 필요하며, 도로·환경적 요인과 관련해서는 도로 조건이 좋은데 따른 교통안전시설의 보강이 필요한 것으로 나타났으며, 차량 외 충돌사고의 경우는 보완 대책으로 인적 요인과 관련하여 신호등과 중앙선의 정비, 과속 방지턱의 보강, 설치가 필요하며, 전도, 전복, 추락 등의 교통사고를 예방하기 위한 중점 대책으로는 도

로·환경적 요인과 관련해서 도로 노면 포장 상태가 불량한 도로의 포장이 필요한 것으로 나타났고, 보완 대책으로는 인적 요인과 관련해서는 시인성 향상을 위한 교통안전시설의 개선이 필요한 것으로 나타났다.

제5장 문제점 및 개선방안

제1절 문제점

본 연구의 결과 나타난 제반 문제점들 살펴보면, 다음과 같다.

대형 교통사고에 관한 자료에 대해 조사, 분석 측면에서 문제점을 살펴보면, 본 공단에서는 1989년, 1991년, 1992년, 1997년, 2000년, 2001년, 2002년, 2003년, 2004년까지 9차례에 걸쳐 “대형 교통사고 사례 분석” 보고서를 발간하였으며, 금번 본 연구에서 대형 교통사고 사례 300건에 대하여 일차적으로 30개 항목의 자료를 구축하였다.

하지만 자료의 동질성 측면에서 볼 때, 상당수의 변수들에 결측치(Missing Value)가 많아, 이 중에서 자료가 충실하게 구축된 충돌 형태를 포함한 7개 항목에 대해 51개 지점을 제외한 249개 지점만을 분석에 도입하게 되었다.

이로 인하여, 기 정리된 정보를 제대로 활용할 수 없게 되는 등 막대한 정보 손실이 초래되게 되어 보다 심층적인 분석은 하기 어렵다는 문제점이 나타났다.

제2절 개선 방안

정리된 상당수의 변수들에 결측치(Missing Value)가 많아, 기 정리된 정보를 제대로 활용할 수 없게 되는 등 막대한 정보 손실이 초래되게 되는 문제점을 해결하기 위해서는 대형 교통사고 조사, 분석 시 표준화된 원 자료 시트의 작성을 의무화할 필요가 있다.

다음의 <표 5-1>은 향후 대형 교통사고 조사, 분석 시 작성을 의무화할 필요성이 있는 표준 원 자료 시트의 작성 예시를 제시하고 있는바, 향후에는 대형 교통사고 1건당 이에 대한 작성을 의무화해야 할 것이며, 아울러 이들 항목뿐만 아니라 필요성이 입증된 새로운 항목의 도입도 계속되어야 할 것이며, 이후에는 이들 자료를 토대로 한 보다 심층적인 연구 결과가 도출될 수 있도록 해야 할 것으로 판단된다.

<표 5-1> 표준 원 자료 시트(작성 예시)

항	목	내	용
사고 원인		졸음운전, 과속	
법규 위반 유형		제한속도 위반	
사고 발생 지역		경기도 광주시	
사고 발생 시간대별		23시	
총 차로수		3차로	
도로 형태		우커브	
사고 야기 차량 차종		승합차	
충돌 형태		정면+우측전면	
사고 당시 기상 상태		비	
노면 포장상태		아스팔트	
도로 중앙부 형태		황색복선	
사고 야기 운전자 소지 면허종별		1종 대형	
사고 야기 운전자 면허 취득기간		5년 6개월	
사고 발생 지점 도로 폭원		25m	
사고 당시 추정 속도		105.6km/h	
타 차량 상태		대향 직진	
사고 야기 운전자 연령		37세	
사고 야기 운전자 사상 여부		사망	
사고 야기 운전자 성별		여자	
사고 발생 지점 도로 곡선 반경		797.6m	
사고 발생 지점 도로 종단 구배		3%	
사고 발생 지점 도로 횡단 구배		6%	
사고 발생 지점 도로 편구배		1%	
사고 당시 기온		16.2°	
사고 당시 습도		63.4%	
사고 당시 풍속		3.2m/s	
사고 당시 풍향		남남동	
사고 당시 강수량, 강설량		64m	
사고 발생 년도		2003년	

제6장 결론

본 연구는 대형 교통사고에 있어 발생지점을 특성별로 유형화하고, 아울러 충돌 형태에 미치는 영향요인에 대한 영향 정도를 판별하는 판별모델과 분류모델을 구축하였다.

그 결과를 요약하여 보면 다음과 같다.

1). 먼저 대형 교통사고 전체에 있어서의 충돌 형태별 영향 요인을 판별하는 모델에서는 전체 적중률이 77.9%로 매우 높게 나타났으며, 상관비는 제1축이 0.70998에서 판별모델이 유효하다고 판단되었으며, 제반 영향 요인들의 특성 역시 명확하게 구분되어 나타났다.

2). 본 연구대상 지점들의 유형화 결과 최적 집단수는 4개인 것으로 나타났으며, 집단별 전체 적중률은 1집단의 경우는 81.6%, 2집단은 74.3%, 3집단은 71.4%, 4집단은 74.7%로 각각은 매우 높게 나타났으며, 상관비는 제1축의 경우 1집단이 0.43033, 2집단이 0.49648, 3집단은 0.80411, 4집단은 0.35493에서 판별모델이 유효하다고 판단되었다.

3). 대형 교통사고 전체에 있어서의 충돌 형태별 영향 요인을 분류하는 모델에서는 차량 충돌의 경우 사고 당시 타 차량 제공 요인이 동일 방향 직진, 회전 중, 대향 직진, 회전 중, 측면에서 직진, 회전 중, 신호 대기, 주·정차, 일시 정지 등을 하고 있는 경우, 법규위반 유형에서는 신호 위반, 중앙선 침범, 제한속도 위반, 주취운전, 안전운전 의무 불이행, 철길 건널목 통과방법 위반, 무면허 운전, 보행자 보호 의무 불이행 등인 경우, 노면 포장상태에서는 아스팔트, 콘크리트인 경우가 가장 많이 속하는 것으로 나타났다.

반면에 차량 외 충돌의 경우는 타 차량 제공 요인이 없는 경우 중 법규위반 유형 중 신호 위반, 제한속도 위반, 앞지르기 방법 및 금지 위반, 주취운전, 안전운전 의무 불이행, 무면허 운전, 보행자 보호 의무 불이행 등, 사고발생 시간대별에 있어서는 17시부터 익일 06시 59분까지인 경우, 또 다시 법규위반 유형에서는 신호 위반, 제한속도 위반, 앞지르기 방법 및 금지 위반, 안전운전 의무 불이행인 경우가 가장 많이 속하는 것으로 나타났다.

4). 집단별 분류 특성으로 1집단은 차량 충돌사고의 경우 주취운전, 철길건널목 통과방법 위반, 무면허 운전, 보행자 보호 의무 불이행 등을 한 상태에서, 진행 방향 전방에 타 차량이 동일 방향과 대향 혹은 측면에서 직진 또는 회전

중이어서 이를 피하는 과정에서 정면이나 측면 충돌의 형태로 나타나거나, 도로 폭이 넓어 운전 조건이 매우 좋은 4차로 이상, 아스팔트나 콘크리트 포장 도로 상에서도 많이 발생하는 것으로 나타났다.

또한 2집단은 차량 외 충돌사고의 경우 주취운전을 한 경우가 많이 발생하는 것으로 나타났으나, 3집단은 차량 충돌사고의 경우 타 차량이 진행 방향 전방에서 신호 대기, 주·정차 또는 일시 정지 중인 경우에, 4집단 역시 차량 충돌사고의 경우 법규 위반 유형에 있어서는 신호위반, 중앙선 침범, 제한속도 위반 등을 한 경우 노면 포장상태에 있어서는 아스팔트와 콘크리트 포장인 경우에 많이 발생하는 것으로 나타났다.

5). 발생지점별 유형화에 따른 4개 집단의 판별 및 분류모델에 있어 충돌 형태별 제반 영향 요인들의 특성은 집단별로 명확한 차이를 보이는 것으로 나타나, 이에 대한 교통사고 예방 대책도 집단별로 달리 제시할 수 있었으며, 아울러 중점적으로 우선 시행되어야 할 대책과 이를 보완할 대책들에 대해서도 보다 체계적으로 제시할 수 있었다.

이상의 연구 결과 나타난 제반 문제점으로는 대형 교통사고를 조사, 분석하는 측면에서 볼 때 상당수의 변수들에 결측치가 많아, 분석 과정에서 막대한 정보 손실이 초래되게 되고 이로 인하여 보다 심층적인 분석을 하기 어려웠다는 점이다.

따라서 이러한 문제점들을 해결하기 위해서는 대형 교통사고 조사, 분석 시 표준화된 원 자료 시트의 작성을 의무화할 필요가 있는 것으로 판단되었다.

<참고문헌>

<국내서>

- 1). Claude Seidman 저, 권오주, 이정무 역, Microsoft SQL Server 2000 데이터 마이닝, 정보문화사, 2001.
- 2). 김종섭, SAS v.8.2를 이용한 통계분석의 이해, 한울출판사, 2003.
- 3). 최종후, 한상태, 강현철, 김은석 공저, Answer Tree를 이용한 데이터 마이닝 의사결정 나무분석, SPSS 아카데미, 1998.
- 4). 노형진, SPSS/Amos에 의한 사회조사분석 -범주형 데이터 분석 및 공분산구조분석 - 형설출판사, 2003.
- 5). 채서일, 김종범, SPSS/PC+를 이용한 통계분석, 법문사, 1988.
- 6). 최종성, SPSS Ver 10을 이용한 현대통계분석, 복두출판사, 2000.
- 7). 허명희, 사회과학을 위한 다변량 자료 분석, 자유아카데미, 2000.
- 8). 허명희, SAS 군집분석, 자유아카데미, 1990.
- 9). 허명희, 수량화 방법론의 이해, 자유아카데미, 1992.
- 10). 최종후, 한상태, 강현철, 김은석 공저, Answer Tree를 이용한 데이터 마이닝 의사결정 나무분석, SPSS 아카데미, 1998.

<국내 논문>

- 11). 김대웅, 임채문, 도로교통에 있어서 운전자 시각특성 분석, 대한교통학회지 제8권 제2호, 1990.
- 12). 이일병, 임현연, 한국의 교통사고 모형 개발에 관한 연구, 대한교통학회지 제8권 제1호, 1990.
- 13). 이일병, 임현연, 교통사고 예측모형 개발과 사고의 단기 예측, 대한국토·도시계획학회지, 제25권 제3호 통권 58호, 1990. 11.
- 14). 이주형, 손동혁, 윤문교, 교통사고의 발생 특성과 그에 따른 사고 요인분석에 관한 연구, 대한국토계획학회지 국토계획, 제25권 제1호 통권 56호, 1990. 3.

<국내 자료>

- 15). 경찰청, 교통사고 처리 지침

- 16). 도로교통안전관리공단, 대형 교통사고 사례 분석,
1989, 1991, 1992, 1997, 2000, 2001, 2002, 2003, 2004.

<동양서>

- 17). 田中 豊, 垂水共之 編, Windows 版 統計解析ハンドブック
－多變量解析－, 共立出版株式會社, 1996.
- 18). 澤駒 勉, 數量化理論とデータ處理, 朝倉書店, 1982.
- 19). 林知己夫, 數量化の方法, 東洋經濟新報, 1985.
- 20). 日本建築學會, 建築·都市計劃のための調査·分析方法, 井上書院,
1987.
- 21). 全日本交通安全協會, 人と車, 第6卷 第2號 通卷 47號, 平成 2年
2月號.

<서양서>

- 22). Markos Papageorgiou, Concise Encyclopedia of Traffic
and Transportation Systems, Pergamon Press, 1991.
- 23). Answer Tree 2.0 User's Guide, SPSS Inc., 1998.

The Study on the Typical Patterns of Big Traffic Accidents Happening Lot's and Establishment of Discriminant and Classification Models of Impact Factors

Kim, Bong Ki

Department of Urban Engineering
Graduate School, Busan National University

Abstract

Big traffic accidents caused by vehicle, are reduced with rate of 6.1% in occurrence cases, rate of 8.4% of deaths from 1990 to 2003 in average rate.

However, it gives rise to property loss because productivity decline, material loss are caused by deaths and the wounded and then it is necessary to study the governmental research more than any other fields.

The analysis of impact factors about the crash type among the big traffic accidents are very important significance in order to establish the strategy for big traffic accidents reduction after this.

Therefore, this study performed the Cluster Analysis in order to

classify the typical patterns of big traffic accident happening lots and enforced the Quantification II Method and Classification and Regression Trees Method to discriminate the effect degree to impact factor by crash type and then it executed the study to contrive the simplification of quantity index for big traffic accidents reduction.

The results are summarized as follows.

1). First, total hit ratio shows very high as 77.9% in model to discriminate the effect factor by each crash type in all big traffic accidents and correlation ratio is justified as discrimination model is effective since the first axis is 0.70998 and impact factors are also shown dividedly and clearly.

2). The typical patterns of big traffic accident happening lots are shown as 4 groups. Total hit ratio of each group are very high as in case of the first group is 81.6%, the second group is 74.3%, the third group is 71.4%, and the fourth group is 74.7% and correlation ratio is justified as the first group is 0.43033, the second group is 0.49648, and the third group is 0.80411, and the fourth group is 0.35493 in case of the first axis and then it is justified that discrimination models are effective.

3). The model of impact factor by crash type in all big traffic accidents, another vehicle providing factor is going straight ahead to same direction, during turning, going straight ahead to opposite direction, going straight ahead during turning, waiting for signal during turning, parking and stopping, temporary stopping and so forth when traffic accidents occur in case of vehicle crash and traffic signal contravention, invasion of central line, violation of speeding, violation of method of passing another car ahead and prevention, drunken driving, failure in safe drive duty, unlicensed driving, failure in pedestrian-protection-duty and so forth are very occurred in case of violation of regulations and asphalt and concrete are very come under in case of road-surfacing status.

While, if another vehicle factor is not provided, traffic signal contravention, violation of speeding, violation of method of passing

another car ahead and prevention, drunken driving, failure in safe drive duty, unlicensed driving, failure in pedestrian-protection-duty and so forth in cases of crash except vehicle are very occurred and in case of time of accident occurrence, traffic signal contravention, violation of speeding, violation of method of passing another car ahead, failure in safe drive duty are very occurred during from 17:00 to 06:59 of the next day.

4). The first group, as the feature of each group, is presented as front or side crash during drunken driving, contravention of railroad passing method, unlicensed driving, failure in pedestrian-protection-duty and so forth in case of vehicle crash accident since another vehicle is going straight ahead or is turning at same direction and opposite direction or side or is frequently presented at four roads or more and asphalt or concrete paved road that is good for driving condition since road width is wide.

And second group is shown that drunken drive are very many occurred as in case of another case of crash accident except vehicles but another vehicle is waiting for traffic signal, parking and stopping or temporary stopping at the front direction in case of vehicle crash accident in third group and

Third group is occurred when another vehicle is waiting for signal change or is parking and stopping or is temporary stopping in case of vehicle crash accident and the fourth group is occurred in case of traffic signal contravention, invasion of central line, violation of speeding, violation of speed limit and so forth in vehicle crash accidents and asphalt and is occurred in case of asphalt and concrete pave as road surfacing condition.

5). The features of impact factor by crash type are distinct difference among groups in discriminant and classification model type of four kinds of group by typical patterns of big traffic accident happening lots and strategies to decrease big traffic accidents on these features are presented with each group dissimilarly and preferential strategy that will

be executed preponderantly and supplementary strategy can be suggested systematically.

Significant data loss is caused in the process of analysis since collected variables have missing values and then further deep analysis was not executed.

Therefore, this study has concluded that standard data sheet drawing is necessary when research the big traffic accidents research, analysis in order to solve these problems.

감사의 글

항상 변함없는 모습으로 저에게 격려를 보내주신 모든 분들께 감사드립니다.

본 논문이 완성되기까지 끊임없는 학문적 지도와 격려를 해주신 정현영 지도교수님께 깊은 감사를 드리며, 부족한 본 논문을 위해 심사를 맡아 더 좋은 논문으로 거듭나게 해주신 손태민 교수님, 이성호 교수님과 보다 좋은 논문이 될 수 있도록 조언을 아끼지 않으신 최열 교수님께도 감사를 드립니다.

또한 본 논문의 시작부터 마무리까지 아낌없는 지도편달과 항상 관심을 가져주신 도로교통안전관리공단 고상선 박사님께 진심으로 깊은 감사의 마음을 전하며, 논문의 진행 과정에서 격려와 조언을 주신 이종달 부장님, 정우택 부장님 그리고 직장 동료에게도 감사의 마음을 전합니다.

이 외에도 논문이 완성되기까지 도움을 주셨던 모든 분들에게도 이 지면을 통해 다시 한번 감사의 마음을 전합니다.

그리고 항상 저를 지켜 봐 주시고, 격려해주시고, 물심양면으로 저에게 사랑과 희생으로 일관해 주신 존경하는 어머니, 장인어른, 항상 용기를 잃지 말라고 하시며 힘이 들 때 나에게 새로운 힘을 주셨던 누님과 자형들, 동생 봉근이 그리고 처남에게도 감사의 마음을 전합니다.

아울러 어려운 여건 속에서도 늘 따뜻한 내조와 사랑을 아끼지 않았던 사랑하는 아내 경미와 우리 두 공주 령희, 수원이와 이 작은 결실을 맺게 된 데 대해 기쁨을 함께 나누고자 합니다.

2005년 6월 김 봉 기 드림