

碩士學位論文

기하구조를 고려한 교차로 교통사고예측모형

2002 年 2 月 日

西南大學校 大學院

土木工學科

李 周 眩

목 차

표목차.....	IV
그림목차.....	VI
영문초록.....	VII
제 1장	
서론.....	1
제 1절 연구 배경 및 목적.....	1
제 2절 연구의 범위.....	2
제 3절 국내외 연구동향.....	3
제 4절 연구 방법.....	4
제 2장 교통사고 기본분석 이론.....6	
제 1절 교통사고요인 특성과 정의.....	6
제 2절 교통사고의 유형과 발생원인.....	8
제 3절 교통사고 분석방법.....	9
제 4절 다중선형회귀이론.....13	
1) 다중회귀의 원리.....	13
2) 다중회귀분석의 목적 및 기본가정.....	14
3) 다중회귀모형의 적합도 검증.....	15
제 5절 다중비선형회귀분석.....17	
1) 비선형 회귀 모형의 일반.....	17
2) 비선형 회귀모형의 추정.....	19
제 6절 수량화이론 제 I 류.....22	
1) 데이터 표현.....	23
2) 수량화이론 제 I 류의 해석방법.....	24
3) 수량을 구하는 계산처리.....	29
4) 예측에 좋은 척도.....	32
(1) 중상관계수와 편상관계수.....	32
(2) 편상관계수와 범위.....	37
제 3장 교통사고데이터 특성 분석.....39	
제 1절 단순빈도분석.....	39

1) 교차로 사고위치별 교통사고발생분포.....	39
2) 교차로 사고위치에 따른 교통사고유형 분포.....	40
3) 교차로 사고유형별 교통사고 발생분포.....	41
4) 월별 교차로 교통사고 발생분포.....	42
5) 요일별 교차로 교통사고 발생분포.....	43
6) 보행자 사고원인별 교차로 교통사고 발생분포.....	44
7) 운전자 사고원인별 교차로 교통사고 발생분포.....	45
제 2절 상관관계 분석(Correlate Analysis).....	46
제 4장 교차로 교통사고 예측모형.....	49
제 1절 교차로 교통사고유형별 교통사고예측모형.....	49
1) 후미충돌에 대한 사고예측모형.....	49
2) 정면충돌에 대한 사고예측모형.....	52
3) 직각충돌에 대한 사고예측모형.....	55
4) 측면접촉에 대한 사고예측모형.....	58
제 2절 다중선형회귀모형.....	61
제 3절 다중비선형회귀모형.....	65
제 4절 수량화 I 류 이론을 이용한 모형.....	66
제 5절 구축된 교통사고 예측모형간의 비교.....	70
제 5장	
결론.....	74
참고문헌.....	77
국문초록.....	80
부록1. 전라북도 도로교통안전관리공단 사고조사 요인표.....	83
부록2. 도로 교통량 통계연보(건설 교통부).....	84
부록3. 교통사고지점의 기하구조 설계도.....	85

표 목차

- 표 1. 구간별 교통량 현황
- 표 2. 회귀직선 판정을 위한 분산분석표
- 표 3. 해당패턴(✓ 표로 표시)
- 표 4. 수량화 I 류를 위한 데이터 표시기호
- 표 5. 행렬 A와 벡터 B
- 표 6. 수량화의 Cross 집계
- 표 7. 외적기준의 Category 화
- 표 8. 전체 데이터의 교차로 교통사고건수와 기하구조요인과의 상관분석
- 표 9. 97년 데이터의 교차로 교통사고건수와 기하구조요인과의 상관분석
- 표 10. 98년 데이터의 교차로 교통사고건수와 기하구조요인과의 상관분석
- 표 11. 99년 데이터의 교차로 교통사고건수와 기하구조요인과의 상관분석
- 표 12. 후미충돌에 대한 ANOVA 분석표
- 표 13. 후미충돌에 대한 다중선형회귀모형결과
- 표 14. 정면충돌에 대한 ANOVA 분석표
- 표 15. 정면충돌에 대한 다중선형회귀모형결과
- 표 16. 직각충돌에 대한 ANOVA 분석표
- 표 17. 직각충돌에 대한 다중선형회귀모형결과
- 표 18. 측면접촉에 대한 ANOVA 분석표
- 표 19. 측면접촉에 대한 다중선형회귀모형결과
- 표 20. 97년도 교차로 ANOVA 분석표
- 표 21. 97년도 교차로 다중선형회귀분석결과
- 표 22. 98년도 교차로 ANOVA 분석표
- 표 23. 98년도 교차로 다중선형회귀분석결과
- 표 24. 99년도 교차로 ANOVA 분석표
- 표 25. 99년도 교차로 다중선형회귀분석결과
- 표 26. 전체 교차로 교통사고 ANOVA 분석표

- 표 27. 전체 교차로 교통사고 다중선형회귀분석결과
- 표 28. 국도 17호의 교차로 교통사고 다중선형회귀모형
- 표 29. 각 연도별 다중비선형 회귀모형
- 표 30. 각 연도별 교차로 수량화 I 류 모형
- 표 31. 구축된 모형간의 결정계수 및 추정오차 비교

그림 목차

- 그림 1. 연구대상 범위
- 그림 2. 교통사고건수예측모델 개발과정
- 그림 3. 교통사고 발생원인
- 그림 4. 교차로 사고위치별 교통사고발생분포
- 그림 5. 교차로 사고위치에 따른 교통사고 유형별 분포
- 그림 6. 교차로 교통사고유형별 교통사고발생분포
- 그림 7. 월별 교차로 교통사고발생분포
- 그림 8. 요일별 교차로 교통사고발생분포
- 그림 9. 보행자 사고원인별 교차로 교통사고 발생분포
- 그림 10. 운전자 사고원인별 교차로 교통사고 발생분포
- 그림 11. 후미충돌에 대한 수량화 I 류 모형
- 그림 12. 정면충돌에 대한 수량화 I 류 모형
- 그림 13. 직각충돌에 대한 수량화 I 류 모형
- 그림 14. 측면접촉에 대한 수량화 I 류 모형
- 그림 15. 97년 교차로 교통사고 수량화 I 류 모형
- 그림 16. 98년 교차로 교통사고 수량화 I 류 모형
- 그림 17. 99년 교차로 교통사고 수량화 I 류 모형
- 그림 18. 전체 교차로 교통사고 수량화 I 류 모형
- 그림 19. 97년 교차로 다중회귀모형과 수량화 I 류 모형의 예측값 비교
- 그림 20. 98년 교차로 다중회귀모형과 수량화 I 류 모형의 예측값 비교
- 그림 21. 99년 교차로 다중회귀모형과 수량화 I 류 모형의 예측값 비교
- 그림 22. 전체 데이터를 이용한 교차로 교통사고 예측모형의 예측값과 실측값 비교

ABSTRACT

Forecasting Model of Traffic Accident at the Intersection Considering Road Geometrical Construction

Lee Ju-Hyun

Department of Civil Engineering major
Graduate School of Seonam University

Directed by Professor Sung, Soo-Lyeon

Generally, the traffic accident on roads produces multiply human factors, vehicle factors and environmental factors.

By establishing the forecasting model of traffic accident through the study using the multiple linear regression model, the multiple non-linear regression model and the multidimensional quantification I. model, this research prevents the traffic accident from happening caused by the increase of vehicles and analyzes the accident more systematically and logically.

First, the research shows the traffic accident distribution at the intersection on National Road No. 17(Chonju~Namwon) in 1997, 1998, and 1999. To know the properties of the annual traffic accident, it practices the simple analysis on the traffic accident types. The result is that the accident on the intersection is mainly happened in the approaching part and inside the intersection. Most type of accident is the clash of the rear.

The elements causing the traffic accident at the intersection are the volume, the cross gradient and the width of road. There are the multiple linear regression model,

the multiple non-linear regression model and the multidimensional quantification I. model in the traffic accident types. From the models, the cross gradient has the greatest influence on the accident at the intersection.

To study the effect of the volume and the geometric of highways on the accident, this research makes the patterns of the traffic accidents using the multiple linear regression theory, the multiple non-linear regression theory and the multidimensional quantification I. theory. As the result of comparison of three models, the R-Square (R^2) using the multidimensional quantification I. theory is highest. In the comparison between the estimate value and the actual survey value and the result of the plot, the estimate value of the model using the multidimensional quantification I. theory is proximate to the actual survey value. Due to that, the pattern using the multidimensional quantification I. theory is more effective for the analysis.

제 1장 서 론

제 1절 연구배경 및 목적

교통은 우리의 생활에 밀접한 영향을 미치고 있으며, 각종 사회 환경의 급격한 발전으로 인하여 자동차의 중요성 및 필요성이 더욱 절실해짐과 아울러 자동차의 대중화로 교통사고의 심각성은 중요한 사회문제로 대두되고 있는 실정이다. 도로교통은 인간과 자동차, 도로, 시설 등 복합적으로 구성되고 있는데, 교통사고는 이들 요소 중 어느 하나만이 불완전하여 발생하는 경우도 있지만 대부분은 몇 가지 요소가 상호 작용하여 발생하고 있다. 전북의 교통사고 발생건수 증가추세를 보면 1991년도에 9,669건에 불과하던 것이 1999년대에 들어서 13,611건으로 연평균 4.37%의 증가율을 보이고 있다. 반면 교통사고의 사망자수는 1991년도에 0.083명/건이었는데 1999년도에는 0.047명/건으로 연평균 -2.91% 감소의 경향을 보이고 있지만, 전국과 비교했을 때 전북의 교통사고 1건당 사망자 발생은 전북이 전국보다 1.38배정도 더 높게 나타났다. 이는 전북에서 발생하고 있는 교통사고가 전국과 비교하여 더 치명적임을 알 수 있다.^[22]

특히, 본 연구대상은 전국에서 사고율이 가장 높은 것으로 알려져 있고, 실제로 과업구간의 사고발생현황을 전국의 통계와 비교하여 보면, km당 사고건수는 6.83건/km로 전국 평균(3.16건/km) 대비 2.16배이며, 일반국도 평균(5.70건/km)의 1.20배인 것으로 나타난 구간이다. 그리고, 더 심각한 것은 교통사고율이나 치사율로서 연구대상구간의 km당 사망자수는 0.97인/km로 전국평균(0.14인/km)에 비해서는 6.78배가 높으며, 일반국도(0.37인/km) 대비는 2.61배, 전국대비는 3.18배로 사고발생시 치사율이 매우 높은 것으로 나타났다[22]. 또한, 전체적으로 볼 때 남원~임실 구간이 임실~전주 구간보다 교통량이 적게 나타나고 있으며, 실제 사고 발생건수도 임실~전주 구간이 남원~임실 구간보다 많은 것으로 나타났다.

표 1. 구간별 교통량 현황

(단위 : 대/일)

구간	1997	1998	1999
남원~임실	20,970	17,712	18,977
임실~전주	31,191	25,561	27,590

이에 본 연구는 전북권의 교통사고가 많은 도로에 있어서 치사율이 가장 높은 국도 17호(전주~남원)를 대상으로 경찰청에서 제공하는 교통사고자료와 도로교통안전관리공단에서 제공하는 교통사고 잦은 곳 기본개선계획 '97년, '98년, '99년 보고서의 기하구조 자료를 바탕으로 데이터를 구축하여 교통사고를 분석하였다. 기존에 연구된 거시적 분석의 틀을 벗어나 본 연구에서는 지역간 도로상의 교차로에 있어서 교통사고 원인의 분석과 교통사고 발생 유형화를 통하여 교통사고 원인의 상호관련성을 미시적 접근방법으로 교통사고 유형별 교통사고의 영향요인을 명확하게 하고자 하였다. 또한 교차로 교통사고 예측모형을 선형 및 비선형 다중회귀모형과 수량화 제 I 류에 의한 모형을 구축하여 교통사고의 감소를 위한 교통안전대책의 기본자료로 활용하고자 하는데 본 연구의 목적이 있다.

제 2절 연구의 범위

본 연구 대상구간은 국도 17호선(전주~남원)구간이다. 이 구간은 직선구간, 곡선구간, 굴곡구간 등이 복합적으로 존재하고 있으며, 특히 전주~임실 구간은 도로 양측이 시가화 되어 있어 도로의 확장이나 접근 통제가 어려운 실정이고, 반면, 임실~남원 구간은 비교적 도로양측에 시가화가 덜 되었지만, 직선구간이 많아 과속의 가능성이 상존하는 구간이다. 또한 남원측에서는 도로의 고저차가 심하여 내리막에서의 고속주행이 자주 발생하는 구간이다. 본 연구의 대상구간을 그림1에 나타내었으며, 이 연구대상구간에서 발생하는 교차로 교통사고에 대하여 사고 다발지점이라고 고려되는 총71개 교차로의 교통사고에 대해 분석하였다. 시간적 범위로는 97년~99년까지 3년 동안의 교통사고자료를 경찰청 및 도로교통안전관리공단으로부터 제공받아 본 연구에 활용하였다. 본 논문의 대

상범위를 아래 그림으로 나타내었다.

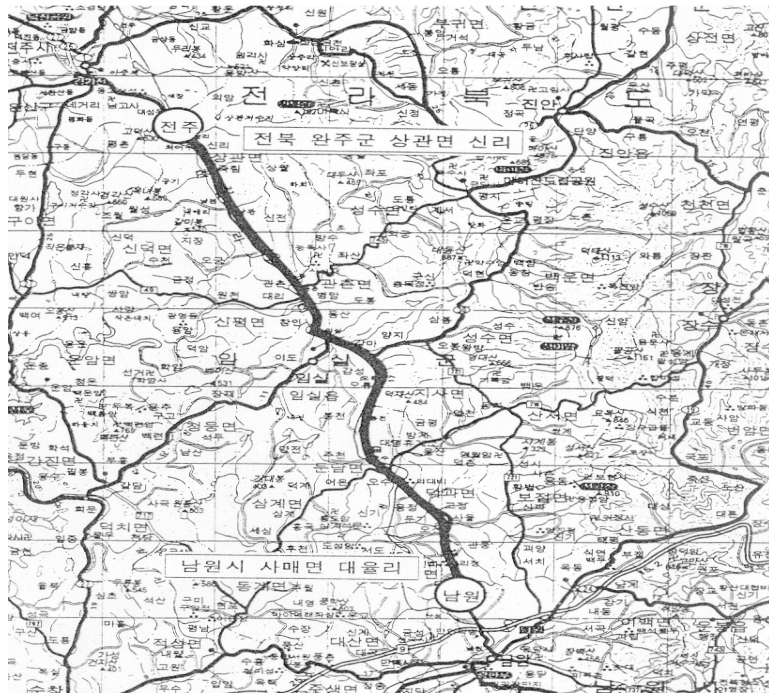


그림 1. 연구대상 범위

제 3절 국내외 연구동향

다중회귀이론을 이용하여 김홍상(김홍상 ; 1987, 서독 칼루스 공과대학 박사학위논문)은 가장 최근에 개발된 횡단분석형 상관관계 모형으로 공업국가 중심의 19개국 1980년 교통사고 통계를 이용하여 다중회귀분석을 실시한 바 있다. 그는 독립변수로 인구수, 자동차 보유대수, 도로 총 연장, 전 국민 총생산액, 연간 주행거리, 국토 면적, 이륜차를 제외한 차량보유 수, 정부예산중 도로 교통부분이 차지하는 비율 등을 이용하고 있다. 또한 그는 동일 자료를 가지고 2개의 독립변수(인구수, 자동차 보유대수)로 59개국을 모델에 포함시켜 다음과 같은 선형 및 지수 모델을 제시하였다. 김홍상의 모형을 자세히 살펴보면 교통사고 사망자수는 인구수 및 자동차수가 많아짐에 따라 사망자수가 증가함을 알

수 있으며 횡단모형의 결정계수(0.932)가 지수모형의 결정계수(0.873)보다 높음을 알 수 있다.^[6]

$$\widehat{Y} = 770.5 + 0.0446(P) + 0.23898 \cdot V : R^2 = 0.932$$

$$\widehat{Y} = 0.00048266(P)^{0.66646} \cdot (V)^{0.29192} : R^2 = 0.873$$

- \widehat{Y} : 사망자수 (인)
- P : 인구수 (1,000인당)
- V : 자동차수 (1,000대당)
- R^2 : 결정계수

강경우는 「공간자기회귀모형을 이용한 고속도로 교통사고 분석에 관한 연구」에서 공간적 상관관계를 고려한 교통사고분석을 위하여 교통사고에 영향을 미치는 변수로 단위 구간별 교통사고건수를, 그에 대한 설명변수로 단위구간별 교통량, I.C. 유무 및 화물차량 비율을 이용하여 공간자기회귀분석을 하였다.(1997).^[1]

三田晴利(일본, 1994)는 수량화 I 류를 이용한 全國觀光交通實態調査를 이용한 交通發生量 模型의 提案에 관한 논문을 제시한바 있다.^[34]

中川茂는 수량화 I 류 이론을 이용하여 도로주변 경관의 조화에 대해 주성분분석을 실시하고 이 결과를 바탕으로 수량화 I 류 모형을 구축하여 도로주변경관평가를 실시하였다.^[35]

제 4절 연구방법

교통사고건수예측모형개발을 위해서 먼저 경찰청에서 제공하는 사고데이터와 도로교통

안전관리공단에서 제공하는 교통사고 잦은 곳 기본개선계획서를 통해 교통량과 교차로 기하구조를 바탕으로 연구하였다. 또한 각각의 사고유형에 따라서도 분석을 하였다.

본 논문은 지역간 도로상의 교차로 교통사고의 유형에 따른 사고요인과 특성을 분석하여 교차로 교통사고예측모형을 구축하기 위하여 다음의 3단계로 분석을 실시하였다.

- 1) 기존 교통사고의 관련이론 연구
- 2) 교차로 교통사고에 대한 사고자료의 수집 및 기초통계 분석
- 3) 지역간 도로상의 교차로 교통사고 요인에 의한 교통사고예측모형 설정

본 연구에서는 교통사고시에 각 경찰서에서 조사한 교통사고 요인조사표를 기본자료로서 이용하였고, 교차로 기하구조는 건설교통부의 국도 17호선(전주~남원)의 교통안전개선방안 설계도를 사용하였으며, 또한 현황조사를 통해 보완하였다. 그리고 교통량자료는 건설교통부에서 간행된 도로교통량 통계연보를 사용하였다.

경찰청에서 조사된 사고요인 자료를 사용하여 종속변수로는 사고건수를 이용하였고, 설명변수로는 교통량과 교차로 기하구조 요인을 사용하여 선형 및 비선형 다중회귀모형과 수량화 I 류에 의한 지역간 도로상의 교차로 교통사고 예측모형을 설정하여 실측값과 각 구축된 모형의 예측값의 비교를 통해 유효한 사고예측모형을 제안하였다.

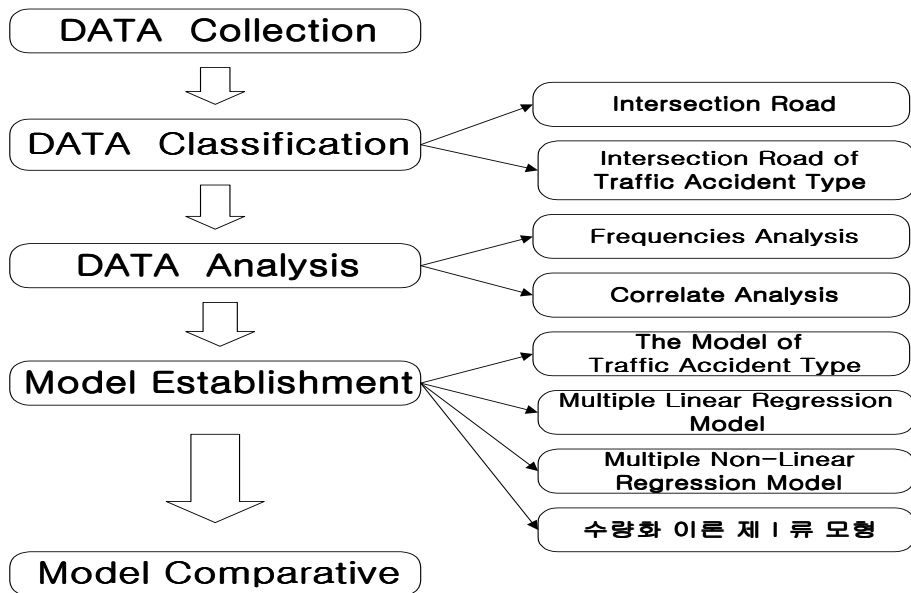


그림 2. 교통사고예측모형 개발과정

제 2장 교통사고 기본 분석 이론

도로교통사고는 일반적으로 도로에 있어서 차량의 교통으로 인하여 사람을 사상하거나 물건을 손괴한 사고를 말한다. 즉, 교통사고는 차량을 수단으로 공간적으로는 도로에 한정하여 사람과 사물에 피해를 입히는 행위라고 할 수 있다. 광의로는 차마, 궤도차 또는 기차, 노면전차는 물론 항공기, 선박 등의 모든 교통기관의 통행중 사람을 사상하거나 물건을 손괴한 경우를 말한다.

제 1절 교통사고 요인 특성과 정의

본 연구에서는 교통사고 분석에 대해 미시적인 접근을 시도하였다. 따라서, 교통사고의

요인으로써 교통량, 종단경사, 횡단경사, 곡선반경, 도로폭 등을 고려하여 분석을 하였으며, 이러한 사고요인들을 선정하기 위해 교차로교통사고에 영향을 많이 주는 요인 순으로 구분한 뒤 요인간의 상관관계를 고려하여 상관성이 큰 변수들에 대해서는 상호간의 상관계수값을 비교하여 큰 변수를 요인으로 채택하였다. 또한, 다른 방법으로써 기존 문헌과 자료를 통하여 연구대상지역에서 고려될 수 있는 교통흐름에 영향을 주는 요인들도 함께 고려하였다.

1) 교통량

교통량은 도시의 발전경향, 시가화의 상황, 토지 이용과 교통량 증감에 따른 교통밀도, 주행속도에 따른 운전자 심리에 영향을 주는 요소이다. 또한, 교통량이 증가하여 교통상황이 복잡하게 되면 운전자 상호간의 의사소통에 무리가 따르게 되고, 운전자의 감정을 자극하여 공격적 감정표출 및 공격적 행동으로 인한 사고를 유도하는 유발요인이 되기도 한다.(이순철, 2000).^[16] 따라서, 본 연구에서는 사고요인인 교통량에 대해 연평균 일 교통량(AADT : Annual Average Daily Traffic)을 이용하였다.

2) 종단경사

종단노선은 도로의 경사나 vertical curve를 포함하는데, 특히 Roy Jorgensen and Associates(1978)과 Organization for Economic Cooperation and Development(1976)에 의하면 upgrade나 downgrade를 포함한 경사지에서 사고율이나 심각도가 증가한다고 보고하였다. 또한, Transportation Research Board(1987a)의 보고에 의하면 종단경사(crest, sag, grade)에 의한 시거확보의 부족으로써 발생한 사고가 경사가 고려되지 않는 지역에서는보다 52%정도 높게 발생했다고 보고하였다. 이렇듯 사고에 많은 영향을 미치는 요인으로써 도로의 종단경사를 교통사고분석에 고려하기 위해서는 도로의 경사도나 curve를 동시에 고려하는 방안이 강구되어야 할 것이다.(Pak Poy and Kneebon, 1988)

3) 횡단경사

차도에서 배수를 위하여 노면의 중심을 정점으로 하고 양쪽으로 향하여 경사진 횡단경사(crossfall)를 붙인다. 노면에 내린 우수는 횡단경사로 측구 또는 가거로 유도된다. 횡단

경사가 크면 클수록 배수상 유리하지만 운전상의 안전과 쾌적면으로 보면 수평에 가까울수록 좋다. 따라서 횡단경사는 배수에 지장이 없는 범위에서 될 수 있는대로 완만하게 하는 것이 좋다. 자동차의 핸들 조작에 영향이 없는 횡단경사의 한도는 2% 정도이다.

즉, 보도, 자전거도 등에는 측구 또는 가거의 배수를 위하여 차도쪽을 향하여 직선 내리막경사로 2%를 표준으로 하는 횡단경사를 붙인다.

4) 곡선반경

자동차가 도로의 곡선부에서도 직선부와 마찬가지로 안정되고 쾌적하게 주행할 수 있도록 곡선부의 곡선반경은 어느 값 이상으로 유지되어야 한다. 자동차가 곡선부를 주행할 때 생기는 위험은 원심력에 의해 곡선부 외측으로 미끄러지거나 뒤집어 지는 것이다. 그 위험성은 자동차의 주행속도와 도로의 곡선반경, 편경사 및 노면의 마찰계수에 따라 달라진다.

5) 도로폭

도로의 폭은 자동차 주행에 주행 안전성을 확보할 수 있는 중요한 요소이다. 따라서 도로의 최소 폭은 안전한 주행을 보장할 수 있는 최소의 폭으로 설치되어야 하며 이 값은 설계속도가 커짐에 따라 증가한다. 도로의 폭이 규정보다 작게 설치되어 있을 경우 옆의 진행차량, 혹은 반대편의 진행차량에 의해서 운전자는 불안감을 느낄 수 있고 이에 따라 운전상태에 영향을 미칠 수 있다. 설계속도에 따른 차로의 폭은 설계속도 80 km/h 일 경우 3.50 m , $60\sim 80\text{ km/h}$ 일 경우는 3.25 m , 60 km/h 미만일 경우는 3.00 m 를 최소 차로폭으로 규정하고 있다.

제 2절 교통사고의 유형과 발생원인

교통사고의 유형은 많은 형태로 구분할 수 있는데 일반적으로 교통사고를 다음과 같이 4가지로 구분한다.

1) 차량간 사고(multiple vehicle)

- 2) 차량단독 사고(single vehicle)
- 3) 차와 사람간 사고(vehicle-pedestrian)
- 4) 차와 고정물체간 사고(vehicle-fixed object)

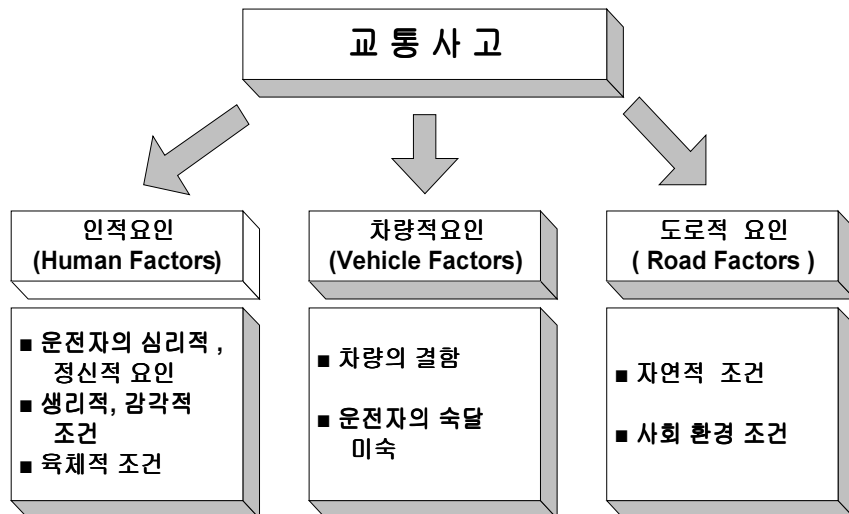


그림 3. 교통사고의 발생 원인

일반적으로 도로교통사고의 요인으로는 인적요인(Human Factors), 차량적요인(Vehicle Factors), 도로적요인(Road Factors)으로 크게 세 요인으로 분류할 수 있다. 인적요인으로는 운전자의 심리적, 정신적인 것과 생리적, 감각적 조건, 육체적 조건 등이 있고, 차량적 요인으로는 차량의 결함이나 운전자의 숙달 미숙으로 인하여 교통사고가 발생하며, 도로적 요인으로는 자연적인 조건과 사회환경적인 조건으로 인해서 도로교통사고의 요인이 된다.

제 3절 교통사고 분석방법

- 1) 교통사고 많은 지점 표시도에 의한 방법

대상지역이 넓지 않고 교통사고건수가 적을 때 벽에 설치해서 살펴보는 방법.

2) 교통사고건수에 의한 방법(Frequency Method)

교통사고를 건수만으로 나타내는 방법으로 적용할 수 있는 지역은 비교적 교통량이 적은 도로에 적당하다. 또한 도로 등급이나 지역에 따라서 그 기준을 달리 적용하거나 인명 피해나 물질 피해 중 하나만을 가지고 표현할 때 사용된다. 도로 구간에 대해서는 2,000mile이하에서 적용되고 기간은 1년 단위를 사용한다.

3) 교통사고율에 의한 방법

일반적으로 정해진 도로 구간에서 교통량과 주행거리를 고려한 교통사고 노출 정도의 변수인 1억대-km당 교통사고수를 통해 교통 안전성을 분석하는 방법인데, 식(2.1)과 식 (2.2)으로 타나낸다.

(i)도로 구간이 길 경우

$$RSEC = \frac{(100,000,000)(A)}{(365) \cdot (T) \cdot (V) \cdot (L)} \quad (2.1)$$

여기서, RSEC : 해당 도로구간의 교통사고율

A : 분석기간 동안 교통사고 발생건수

T : 분석기간(년)

V : 연평균 일 교통량(대/일)

L : 구간 길이(km)

(ii) 도로 구간이 짧을 경우

$$RSP = \frac{(100,000,000)(A)}{(365)(T)(V)} \quad (2.2)$$

여기서, RSP : 도로지점의 교통사고율

V : 연평균 일 교통량(대/일)로서 해당 접근로의 교통량 합계

4) 교통사고의 치상 정도를 감안한 접근방법

모든 형태의 교통사고를 재산피해만이 발생한 교통사고(EPDO : Equivalent Property Damage Only)로 환산하는 방법이다. 부서에 따라 EPDO를 산출하기 위한 가중치가 다소간 차이를 보일 수 있다.

5) 전통적인 통계적 분석방법

교통사고 건수나 교통사고율에 의한 방법을 통해 일단 후보지점들을 선정하고 나면, 이들을 대상으로 보다 상세한 분석을 수행할 수 있게 된다. 또 한가지 선정방법은 발생 특성이 대체로 정규분포를 따를 것이라는 가정을 전제로 하고 있으며, 만약 이 가정이 성립한다면 우리는 다음과 같은 수식을 통해 교통사고 많은 지점을 선정할 수 있다.

$$OB_i > XA + (K)(S) \quad (2.3)$$

여기서, OB_i = i 지점에서의 교통사고 건수 혹은 교통사고율

XA = 수집된 모든 조사지점의 평균교통사고 건수 혹은 평균교통사고율

K = 신뢰도에 따른 계수로서 신뢰도는 90, 95, 99%를 사용함

S = 수집된 모든 조사지점의 표본 표준편차

6) 임계 교통사고율이 적용에 의한 방법

임계 교통사고율방법은 다소 변형된 값을 이용하는 방법이다. 이 방법의 다른 방법들과의 차이점은 다음과 같다

- 교통사고율에 대해서만 적용할 수 있다.
- 이 방법에서는 일단의 대상 지역들에 대한 교통사고 발생건수가 포아송분포를 따르는 것으로 가정한다.

- 해당 지점의 교통사고율을 해당 지점과 유사한 지점의 평균교통사고율과 비교한다.

이 방법은 한 지점이나 한 구간에 적용할 수 있으며, 지점인 경우 백만 통과차량당 교통사고율을 적용하며, 구간인 경우 1억대-km당 교통사고율을 적용한다.

만약 해당 지점의 교통사고율이 다음 조건을 만족시키게 되면 이 지점은 교통사고가 유난히 많이 발생하는 것이다.

$$OBR_i > XS + K\sqrt{\frac{XS}{V_i}} + \frac{1}{2V_i} \quad (2.4)$$

여기서,

OBR_i = 지점 i 의 조사결과를 나타낸 교통사고율

XS = 지점 i 와 유사한 특성을 갖는 도로지점들에서 조사된 평균 교통사고율

V_i = 지점 i 의 교통량

K = 신뢰도에 따른 계수

문제는 해당 지점과 유사한 특성을 갖는 도로지점으로 과연 어느 지점들을 포함시킬 것인가의 문제이다. 일반적으로 넓은 범위의 정의를 내려서 유사성을 규정하고 있다.

7) 확률적 분석방법

종전의 방법은 신뢰도의 선정이 임의적이라는 단점을 지니고 있다. 또 한 가지 지적되는 문제점은 교통사고의 분석에 있어 현재의 상황과 과거의 상황이 연결되지 않음으로써 현재의 상황이 크게 달라진 경우 교통사고 분석이 큰 의미가 없어진다는 점이다.

이러한 문제점을 감안하여 새로운 접근방법이 정립되었는데 이 방법은 조건부 확률에 기초한 통계이론을 근간으로 하고 있다. 즉 이 방법은 해당 조사지점이 갖고 있는 교통사고이력과 교통량 및 도로조건하에서, 이 지점이 위험한지 아닌지의 여부를 확률을 통해 분석하여 알고자 하는 것이다.

제 4절 다중선형회귀이론(Multiple Linear Regression Theory)

1) 다중회귀의 원리

회귀분석의 궁극적인 목적은 종속변수를 예측할 수 있는 하나의 선형방정식을 도출하는 데 있다. 이 방정식을 회귀방정식이라고 하는데 이 회귀방정식은 종속변수의 실제값과 회귀방정식에 의해서 예측될 수 있는 값의 차이를 최소화하는 선형함수이다. 회귀방정식의 도출은 오차의 제곱의 합을 최소화하는 최소제곱법을 이용하게 되는데 회귀방정식의 일반적인 모형은 다음과 같다.

(1) 독립변수가 한 개인 경우(單純回歸分析 : Simple Regression)

$$Y = \alpha + \beta_1 X_1 + \varepsilon \quad (2.5)$$

(2) 독립변수가 두 개 이상인 경우(多重回歸分析 : Multiple Regression)

$$Y = \alpha + \beta_1 X_1 + \dots + \beta_i X_i + \varepsilon \quad (2.6)$$

다중회귀방정식의 경우 $\alpha, \beta_1, \dots, \beta_i$ 를 추정해야 한다. α, β_i 를 모수(parameter)라고 하는데 이는 최소제곱법(least square method)에 의해서 추정한다. 최소제곱이란 다음의 식에서처럼 실제값과 회귀방정식으로 예측되는 값과의 차이를 최소화하는 $\hat{\alpha}, \hat{\beta}_i$ 를 추정해내는 방법이다.

$$\varepsilon_1^2 = \sum_{i=1}^n [Y_i - (\hat{\alpha} + \hat{\beta} X)]^2 : \text{최소제곱법(O.L.S)} \quad (2.7)$$

$$\hat{Y} = \hat{\alpha} + \hat{\beta} X : \text{추정된 회귀식} \quad (2.8)$$

여기서, $\hat{\alpha}$: 상수

$\hat{\beta}$: 기울기($\hat{\beta}$ 가 X의 변화에 따른 Y의 변화량을 결정해 주는 계수)

다시 말해서, 회귀분석이란 한 변수를 이용하여 다른 변수의 값을 설명하거나 예측할 수 있는 모형으로 자료를 분석하는 것이다. 이 때 설명하는 변수를 설명변수 또는 독립변수라 하고, 설명이 되거나 예측이 되는 변수를 종속변수 또는 반응변수라 한다. 즉, 다중회귀분석은 쌍으로 관찰된 연속형 변수들 사이의 관계에 있어서 한 변수를 원인으로 하고 다른 변수들을 결과로 하여 종속변수와 설명변수들 사이의 식을 구하고 그 식을 이용하여 종속변수의 변수값을 예측하는 통계적 분석방법이다.

설명변수가 하나인 경우를 단순회귀분석이라 하고 여기서 구한 모형을 단순회귀모형(Linear Regression Model)이라 하며, 설명변수가 두 개 이상인 경우를 다중회귀분석이라 하는데 여기서 구한 모형을 다중회귀모형(Multiple Regression Model)이라 한다.

단순회귀모형에 의한 방정식과 다중회귀모형에 의한 방정식은 식 (2.5)와, 식(2.6)으로 나타낸다.

2) 다중회귀분석의 목적 및 기본가정

다중회귀분석의 목적은 다음과 같다.

- ① 여러 개의 독립변수로 종속변수를 설명하는 모형을 만든다.
- ② 모형의 적합성을 검정한다.
- ③ 각 회귀계수의 유의성을 검정한다.
- ④ 결정계수의 크기를 검토한다.
- ⑤ 회귀식을 이용해서 종속변수의 값을 예측/추정 한다.
- ⑥ 여러 개의 독립변수 중에서 유의한 변수만 선정하여 유의한 모형을 만들 수 있다.
- ⑦ 독립변수들만이 종속변수에 미치는 영향의 정도를 비교할 수 있다.

회귀분석을 실행하기 위해서는 자료에 대하여 다음과 같은 가정을 해야 한다.

① 정규분포성(normality)

: 같은 X값을 가지는 개체들의 Y값은 정규분포를 따른다.

② 독립성(Independence)

: 종속변수 Y는 통계적으로 서로 독립이어야 한다. 즉, 한 관측값이 다른 관측값들에 영향을 받아서는 안 된다.

③ 선형성(Linearity)

: 주어진 X값을 가지는 Y의 기대값(평균)은 모두 일직선 위에 있다.

④ 동일 분산성(equal variance, homoscedasticity)

: 모든 X에서 Y의 분산이 모두 같다.

3) 다중회귀모형의 적합도 검증

어떠한 통계분석에서도 중요하게 취급되는 부분이 데이터로부터 설정한 모형이 잘 적합하는가에 대한 검토를 하는 과정이다. 이를 적합도(Goodness of fit)검토라 하는데 이는 분석하고자 하는 데이터가 주어진 가정사항을 잘 만족하는지를 검토하는 것도 포함하고 있다.

(1) 결정계수(R^2)

종속변수의 전체 변동 중에서 회귀모형에 의하여 설명된 변동의 비율을 말한다.

$$R^2 = \frac{\sum(\widehat{Y}_i - \overline{Y})^2}{\sum(Y_i - \overline{Y})^2} \quad (2.9)$$

결정계수가 클수록(1에 가까울수록) 좋은 회귀모형이라고 할 수 있으나, 절대적으로 얼마 이상이어야 된다는 규칙은 없다.

(2) 수정결정계수(Adjusted R-Square)

회귀방정식에서 포함되는 독립변수의 수가 증가할수록 결정계수 R^2 의 값은 상승된다.

이러한 문제를 해결하기 위해서 독립변수의 수가 증가될수록 자유도의 수는 감소된다는 사실에 착안하여 자유도를 고려한 수정결정계수를 사용하게 된다.

$$\overline{R^2} \text{ (or } AdjR^2) = 1 - \frac{\sum (Y_i - \widehat{Y}_i)^2 / (n - k)}{\sum (Y_i - \overline{Y})^2 / n - 1}$$

(2.10)

여기서, n : 표본의 수

k : 추정치들의 수

(3) 분산비 F 값

회귀모형을 이용해서 독립변수로 종속변수를 설명을 할 수 있게 되는 경우 “모형이 유의하다”라고 하는데, 모형이 유의하다는 것은 독립변수의 계수가 0과 유의하게 다르다는 것이다. 모형의 유의성 검정은 분산분석표를 이용해서 할 수 있다.

SST : 총 제곱 합(Sum of Square of Total)

SSE : 오차제곱합(Sum of Squares of Error)

SSR : 회귀제곱합(Sum of Squares of Regression)

표 2. 회귀직선 판정을 위한 분산분석표

분산요인	제곱합	자유도	제곱평균	F	Sig F
회귀 (Regression)	SSR	P^*	$MSR = \frac{SSR}{P}$	$\frac{MSR}{MSE}$	$F(p, n - p - 1; \alpha)$
잔차 (residual)	SSE	$n - p - 1$	$MSE = \frac{SSE}{n - p - 1}$		
계(total)	SST	$n - 1$			

P^* : 독립변수의 수

(4) 표준화(회귀)계수

종속변수에 영향을 미치는 독립변수들의 어느 변수가 더 중요한지 우열을 가릴 수 없는 경우 회귀계수를 단위에 상관없는 계수로 만드는 것을 회귀계수의 표준화라고 하고, 이렇게 만들어진 회귀계수를 표준화(회귀)계수 또는 베타(beta)계수라고 한다.

검정결과 회귀계수가 유의하다는 결론에 도달하면 ‘독립변수가 종속변수에 영향을 미치며, 회귀계수가 예측에 필요한 변수다’라고 해석한다. 또한 유의하지 않은 결과가 나오면 ‘독립변수는 종속변수를 설명하는데 별 도움이 되지 않는다’거나 또는 ‘변수 X와 변수 Y는 선형의 관계가 없다’라고 결론을 짓는다.

제 5절 다중비선형회귀이론

1) 다중비선형회귀의 일반

주어진 모형이 비선형 모형(Nonlinear Model)이라도 적당한 변환을 하면 선형으로 바뀌어지는 것들이 있는데, 이와 같은 성질을 갖는 비선형 모형을 내부적으로 선형(Intrinsically Linear)이라 한다. 다음과 같은 모형이 내부적으로 비선형 모형이다.

$$Y = \exp(\alpha + \beta X + \varepsilon) \quad (2.11)$$

왜냐하면, 이 모형의 양변에 \log 함수를 취하면, 원래 모형을 아래와 같이 쓸 수 있고,

$$\log Y = \alpha + \beta X + \varepsilon \quad \Leftrightarrow \quad Y' = \alpha + \beta X + \varepsilon \quad (2.12)$$

결과되는 모형은 반응 변수가 Y' 이고 회귀 변수가 X 인 단순 회귀 모형으로 귀착되기 때문이다. 앞의 모형과 형태는 비슷하지만, 다음 모형은 내부적으로 선형이 아니다.

$$Y = \exp(\alpha + \beta X) + \varepsilon \quad (2.13)$$

비선형 모형에는 어떠한 정형도 있을 수 없으므로, 여기서는 다음과 같이 비선형 모형을 표기한다.

$$Y = \mathcal{A}(X_1, X_2, \dots, X_k; \beta_1, \beta_2, \dots, \beta_\rho) + \varepsilon \quad (2.14)$$

이 모형에서 회귀 변수의 개수 k 와 추정해야 할 모수의 개수 ρ 는 오히려 다른 경우가 더 많은데 유의해야 한다. 그리고, 선형 회귀분석에서와 같이 절편의 개념도 없으므로 β_0 와 같은 표기를 피한다. 선형 모형과 마찬가지로, 오차항 ε 에 대하여는 다음과 같은 가정을 한다.

$$E(\varepsilon) = 0 \quad \text{Var}(\varepsilon) = \sigma^2$$

여기서, 각 관측에 대한 오차들은 서로 독립으로, 오차들 간에 상관관계는 없다. 그리고 통산 ε 의 분포로 정규분포를 가정한다.

비선형 모형을 수집된 n 개의 관측에 대하여 다시 쓰면 다음과 같다.

$$Y_i = \mathcal{A}(X_{1i}, X_{2i}, \dots, X_{ki}; \beta_1, \beta_2, \dots, \beta_\rho) + \varepsilon_i, \quad i = 1, 2, \dots, n \quad (2.15)$$

또, 만일 다음과 같은 벡터, 혹은 행렬 표기를 약속하면,

$$X = [X_1, X_2, \dots, X_k]'$$

$$\beta = [\beta_1, \beta_2, \dots, \beta_\rho]'$$

위에서 제시한 모형은 간단히 하면

$$Y = f(X; \beta) + \varepsilon \quad (2.16)$$

으로 쓸 수 있으며, 위의 모형에 대한 기대값은

$$E(Y) = f(X; \beta)$$

로서, 비선형 모형의 적합에서 우리가 바라는 목표는 최대한 오차를 줄이면서 자료에 잘 들어맞는 모형을 결정하는 것이다. 모형 $E(Y) = f(X; \beta)$ 는 현재 알지 못하는 모수로 결정되므로, 우리의 일차적인 목표는 당연히 모수 β 에 대한 좋은 추정이라 하겠다. 덧붙여, i 번째 관측의 회귀 변수들의 값을 다음과 같이 벡터로 표기하면

$$X = [X_{1i}, X_{2i}, \dots, X_{ki}]'$$

$$\text{모형} \quad Y_i = f(X_{1i}, X_{2i}, \dots, X_{ki}; \beta_1, \beta_2, \dots, \beta_p) + \varepsilon_i, \quad i = 1, 2, \dots, n$$

$$Y_i = f(X_i; \beta_p) + \varepsilon_i \quad (2.17)$$

로 쓸 수 있다.

실제 자료 분석에서 비선형 모형을 적합하는 대표적인 경우와 모형의 예로 다음과 같은 것들이 있다. 첫째로, 화학 제품 등에서 경과된 반응 시간(T)과 농축도(Y)에 대한 모형은 다음 형태로 주어진다.

$$Y = \beta_1 [1 - \exp(-\beta_2 T)] \quad (2.18)$$

또, 식물이나 조직의 성장도(Y)와 자란 시간(T) 간의 관계는 다음 모형으로 주어진다.

$$Y = \beta_1 / [1 + \beta_2 \exp(-\beta_3 T)] \quad (2.19)$$

2) 비선형 회귀모형의 추정

비선형 모형의 적합에도 최소제곱법(Least Squares Method)을 사용한다. 즉, 오차들의 제곱합

$$Q(\beta) = \sum_{i=1}^n \varepsilon^2 = \sum_{i=1}^n [Y_i - f(X_i; \beta)]^2 \quad (2.20)$$

를 최소화하는 β 들의 추정치 $\widehat{\beta}$ 을 찾는다.

찾는 방법 자체는 선형 모형인 경우와 다를 바가 없다. 모수 벡터 β 의 함수이자 오차 제곱합인 Q 를 $\beta_j (j = 1, 2, \dots, p)$ 에 대하여 편미분하고 0으로 놓은 다음, 결과되는 p 개의 연립 방정식, 즉, 정규방정식을 풀면 된다. 정규방정식은 다음과 같이 표현할 수 있다.

$$\sum_{i=1}^n [Y_i - f(X_i; \widehat{\beta})] \left[\frac{\partial f(X_i; \beta)}{\partial \beta_j} \right]_{\beta = \widehat{\beta}} = 0, \quad j = 1, 2, \dots, p \quad (2.21)$$

그러나, 비선형 모형의 함수 형태는 문제마다 제각각이고, 어떠한 정형이 있을 수 없으므로, 이와 같은 정규 방정식의 일반해는 수식으로 표현하여 풀기 힘들다. 따라서 컴퓨터로 가능한 해를 찾을 수밖에 없다. 컴퓨터를 이용하여 해를 구한다할 때, 가장 직관적이고 원시적인 방법은 적당한 모수값들을 정규 방정식에 대입해보아 p 개의 모든 방정식이 모두 0이 되는 값을 추정값로 잡으면 된다. 그러나, 이 방법은 극히 비효율적이며 성공할 확률은 거의 없다. 왜냐하면, 복잡한 p 개의 연립 방정식을 만족하는 모수의 추정값들을 구하려는 것보다는 차라리 우리가 최소화하고 싶은 오차 제곱합 Q 를 여러 모수값들에서 산출하여 그 중 최소값을 주는 모수값을 결정하고, 또 그 모수값들 근방에 위치한 모수값들을 더 잘 쪼개어 또 Q 값을 산출해보고 하는 식이 더 효율적이기 때문이다.

이런 식으로 최적의 모수값을 찾는 체계적인 방법으로 격자탐색(Grid Search)이 있다. 이 방법에서는 먼저 각 모수 β_j 마다 적당한 초기값(Initial Value)을 주고, 그 초기값 근방으로 동일 폭의 격자를 치고, 각 교차점에서 오차 제곱합 Q 의 값을 산출한 후 그 중 최소값을 찾는다. 그 다음 단계로, 최소값이 나온 격자점 근처에 더 작은 폭의 격자를 치고 또 Q 값들을 산출하는 식으로 반복해 나가다가 더 이상 Q 값이 향상, 즉, 더 작아지지 않으면 탐색을 끝낸다. 이 격자 탐색을 하다보면 자연적으로 산출된 Q 값들을 그래프로 그려서, 오차 제곱합들의 값의 추이를 눈으로 보는 편이 더 효과적인 때가 많다. 이런 Q 값들의 그래프를 흔히 제곱합 표면(Sum of Squares Surface)이라 한다.

제곱합 표면은 결국 p 개의 모수들에 대한 함수의 그래프이므로, 이 그래프는 결국 $p+1$ 차원에서 정의된다. 모수가 하나일 때 제곱합 표면의 그래프를 그리는 것은 쉽다. 두 개일 때도 등고선도(Contour Plot)를 그리면 되므로 큰 문제가 안 된다. 그러나, 모수의 개수가 세 개 이상일 때는 현실적으로 그래프를 그릴 방법이 없다. 이런 경우에는 한번에 두 개의 모수를 선택하여 이 두 개의 모수에 대하여 등고선도를 그려볼 수 있다. 그러나, 이런 방법은 그래프 해석이 간단하지 않다.

이상에서 언급한 격자 탐색법은 NLIN 절차에서 채용하는 탐색법에 비하여 초보적이라 하겠으나, 탐색시 초기값을 결정하기 위하여 많이 이용된다. 비선형 적합 기법들을 이용하려면 필히 모수들의 초기값을 잘 지정해야만 한다. 초기값 지정이 잘못되면 아무리 오래 계산을 하더라도 모수값 추정에 실패하기 때문이다.

그리고, 주어진 초기값을 기초로 Q 를 최소로 하는 모수를 추정하는데 성공하였다면, 이 해가 정말 우리가 목표로 하던 답인지는 알 수 없다. 왜냐하면, 답이 나왔다는 말은 정규방정식을 만족시킨다는 말과 동일한데, 하나의 제곱합 그래프에는 미분해서 0이 되는 점이 둘 이상일 수 있기 때문이다. 다른 말로, 곡면이 휘어지면서 여기저기에 국부최소(Local Minimum)점들이 존재할 수 있기 때문이다. 컴퓨터 알고리즘에서는 하나의 초기값 근방에서 최소점을 발견하면, 그 것이 국부최소든 전체최소든 가리지 않고 프로그램을 끝내버린다. 이런 까닭에 컴퓨터로 결정된 해를 흔히 최적해(Optimum Solution)라 부른다.

비선형 최소제곱법(Nonlinear Least Squares Method)은 OR 분야에서 많이 다루는 비선형 최적화문제(Nonlinear Optimization Problem)의 특별한 경우로 간주할 수 있다. 다

음 방법들은 비선형 최적화문제의 해를 구하는 방법들 중 그것들에 대해 확실히 알려진 것들이다.

- 급 강하법(Steepest Descent Method)
- 수정 경사법(Modified Gradient Method)
- 뉴턴의 방법(Newton Method)

여기서 급 강하법과 수정 경사법은 개념상 거의 유사한 방법이다. 그리고, 뉴턴의 방법은 뉴턴-랩슨방법(Newton-Raphson Method)이라고도 부른다. 그러나, 비선형 회귀분석을 위하여는, 이들 방법보다는 다음과 같이 전문적으로 비선형 최소제곱법을 푸는 방법들을 이용하는 편이 더 효율적이다.

- 가우스-뉴턴의 방법(Gauss-Newton Method)
- 마아퀘트의 방법(Marquardt Method)
- DUD(Doesn't Use Derivatives : 도함수를 사용않함)

앞에서도 언급하였지만, 자신의 문제에 알맞은 좋은 탐색 방법을 선택하는 것 못지않게 좋은 초기값을 선택하는 것도 중요하다. 위 모든 탐색 기법들은 공통적으로 정해진 알고리즘을 반복(Iteration) 처리하여 최적해를 구한다. 초기값을 잘 주면 빨리 최적해를 찾을 것이고, 잘못 주면 궁극적으로 아무런 해도 구하지 못할 수 있다. 그렇다면, 초기값을 어떻게 선택해야할까? 사실, 여기에는 특별한 왕도가 없다. 단지, 과거의 경험, 모형 자체에서 주어지는 모수값의 범위 등을 고려하여야 하며, 전혀 정보가 없다면 격자탐색을 미리 해서 최소값이 나올만한 몇 군데 점을 중심으로 초기값을 바꿔가면서 탐색할 수 밖에 없다.

제 6절 수량화이론 제 I 류

數量化理論이란 정도, 상태, 유무 또는 ‘예·아니오’와 같은 질적 데이터에 수량을 주고, 중회귀분석, 주성분분석, 판별분석과 같은 다차원해석을 행하는 수법이다.

수량화 제 I 류 방법은 수량화 분석을 함에 있어서, Y 의 발생을 예측하기 위해, 그 현상에 관련된 m 개의 항목(Item)중 구분(Category)의 반응으로써 설명특성 (X_1, X_2, \dots, X_m)이 측정될 때 그것들의 질적설명특성 X_j 에 기초하여 목적의 특성수량 Y 를 예측하거나 Item과 Category의 요인분석을 하는 질적데이터의 분석 방법론이다.

1) 데이터 표현

수량화 이론 제 I 류의 질적데이터 표현법으로써 δ 변수를 사용한다. δ 변수는 1이거나 0으로 표현되는 것으로 수학적 표현에 자주 이용된다. \checkmark 는 Item과 Category로 개체(Sample)가 해당할 때의 표시이다.

\checkmark 표시를 δ 변수의 내용으로써 표현하면 \checkmark 의 위치는 1, 기타는 0으로 표현하면 된다.

기본적인 질적데이터는 다음과 같이 2가지 방법의 표시로 사용된다.

(1) Item · Category - $C_{(jk)}$ 에 해당할 경우 표현

$$\delta_{i(jk)} = \begin{cases} 1 : \text{item } j\text{의 Category} - k\text{인 } C_{(jk)}\text{에 개수 } i\text{가 해당할 때} \\ 0 : \text{그 외} \end{cases}$$

(2) 일렬의 Category - $C_{(j)}$ 의 반응으로 표시할 경우

$$\delta_{i(j)} = \begin{cases} 1 : \text{Category} - C_{(j)}\text{에 개체 } i\text{가 반응할 경우} \\ 0 : \text{비해당} \end{cases}$$

표3에 수량화 데이터의 해당패턴을 나타내었다.

표 3. 해당 패턴(✓표로 표시)

A		X_1				X_2				...	X_m		
B		1	2	...	l_1	$l_1 + 1$	$l_1 + 2$...	$l_1 + l_2$ L		
D	<div>C E</div>	$C_{(11)}$	$C_{(12)}$...	$C_{(1l_1)}$	$C_{(21)}$	$C_{(22)}$...	$C_{(2l_2)}$...	$C_{(m1)}$...	$C_{(ml_m)}$
k_1	1	✓						✓		...			✓
k_2	2		✓				✓			...	✓		
k_3	3				✓				✓	...			✓
⋮	⋮			⋮			⋮			⋮	✓		
k_n	n			✓		✓				...		✓	

2) 수량화 이론 제 I 류의 해석방법

임의의 수량을 Y 로 표시하고, 이때 데이터로서 수량 Y 와 몇 개의 수량 Y 에 관계하는 질적인 량 (X_1, X_2, \dots, X_m)이 얻어지게 된다. 이 수량 Y (외측기준의 특성)를 몇 개의 질적인 것(설명변수)으로 예측한다고 하자. 그것은 질적인 것을 수량화하고, 수량 $\{x_{(jk)}\}$ 를 표현된 수량 예측량 \widehat{Y} 의 선형식으로 표현되어진다.

$$\widehat{Y} = X_1 + X_2 + \dots + X_m = \sum_{j=1}^m X_j \quad (2.22)$$

$$\text{단, } X_j = \sum_{k=1}^{l_j} \delta_{(jk)} x_{(jk)}$$

$$\delta_{(jk)} = \begin{cases} 1 : C_{(jk)} \text{에 해당} \\ 0 : C_{(jk)} \text{에 비해당} \end{cases}$$

l_j : Item X_j 의 Category 수

수량 $\{x_{(j,k)}\}$ 은 실측량 Y 와 예측량 \widehat{Y} 차의 제곱평균이 최소가 되도록 구한다.

이 방법을 최소제곱법이라 하고, $(\max_{x_{(j,k)}} E(Y - \widehat{Y})^2)$ 이다. 이 방법에 의하여 각

Category $C_{(j,k)}$ 에 가장 좋은 양 $(x_{(j,k)}^*)$ 를 결정한다. 이 예측은 Y 와 \widehat{Y} 의 상관 계수를 최대로 되는 예측에 대응하고 선형관계를 이용한 수량화의 예측 중에서 가장 오차가 적은 예측이다. 표4는 데이터의 표시기호를 나타낸 것이다.

표 4. 수량화 I 류를 위한 데이터 표시 기호

	설 명 특 성 (질적 데이터)										목적특성	
Item	X_1				\cdots	X_f	\cdots	X_m				(수량데이터)
Category	$C_{(11)}$	$C_{(12)}$	\cdots	$C_{(1l_1)}$	\cdots	$C_{(j,k)}$	\cdots	$C_{(m1)}$	$C_{(m2)}$	\cdots	$C_{(ml_m)}$	Y
수량 개체	$x_{(11)}$	$x_{(12)}$	\cdots	$x_{(1l_1)}$	\cdots	$x_{(j,k)}$	\cdots	$x_{(m1)}$	$x_{(m2)}$	\cdots	$x_{(ml_m)}$	
1	$\delta_{1(11)}$	$\delta_{1(12)}$	\cdots	$\delta_{1(1l_1)}$	\cdots	$\delta_{1(j,k)}$	\cdots	$\delta_{1(m1)}$	$\delta_{1(m2)}$	\cdots		
2	$\delta_{2(11)}$	$\delta_{2(12)}$	\cdots	$\delta_{2(1l_1)}$	\cdots	$\delta_{2(j,k)}$	\cdots	$\delta_{2(m1)}$	$\delta_{2(m2)}$	\cdots		
3	$\delta_{3(11)}$	$\delta_{3(12)}$	\cdots	$\delta_{3(1l_1)}$	\cdots	$\delta_{3(j,k)}$	\cdots	$\delta_{3(m1)}$	$\delta_{3(m2)}$	\cdots		
\vdots	\vdots	\vdots	\vdots	\vdots	\vdots	\vdots	\vdots	\vdots	\vdots	\vdots		
i	$\delta_{i(11)}$	$\delta_{i(12)}$	\cdots	$\delta_{i(1l_1)}$	\cdots	$\delta_{i(j,k)}$	\cdots	$\delta_{i(m1)}$	$\delta_{i(m2)}$	\cdots		
\vdots	\vdots	\vdots	\vdots	\vdots	\vdots	\vdots	\vdots	\vdots	\vdots	\vdots		
n	$\delta_{n(11)}$	$\delta_{n(12)}$	\cdots	$\delta_{n(1l_1)}$	\cdots	$\delta_{n(j,k)}$	\cdots	$\delta_{n(m1)}$	$\delta_{n(m2)}$	\cdots		

Item · Category $C_{(j,k)}$, $j = 1, 2, \cdots, m$; $k = 1, 2, \cdots, l_j$ 의 수량 $\{x_{(j,k)}\}$ 가 결정된 것으로써 $E(Y - \widehat{Y})^2$ 로 표현한다.

$$\begin{aligned}
 E(Y - \widehat{Y})^2 &= E(Y^2) - 2E(Y\widehat{Y}) + E(\widehat{Y}^2) \\
 &= E(Y^2) - 2 \sum_{j=1}^m E(YX_j) + \sum_{j=1}^m \sum_{u=1}^m E(X_j X_u)
 \end{aligned}
 \tag{2.23}$$

$X_j = \sum_{k=1}^{l_j} \delta_{(j,k)} x_{(j,k)}$ 를 식 (2.22)에 대입하면,

$$\begin{aligned}
E(Y - \widehat{Y})^2 &= E(Y^2) - 2 \sum_{j=1}^m \sum_{k=1}^{l_j} E(Y \delta_{(jk)}) \cdot x_{(jk)} \\
&+ \sum_{j=1}^m \sum_{u=1}^m \sum_{k=1}^{l_j} \sum_{v=1}^{l_u} E(\delta_{(jk)} \cdot \delta_{(uv)}) x_{(jk)} \cdot x_{(uv)}
\end{aligned}
\tag{2.24}$$

여기서, $E(Y - \widehat{Y})^2$ 를 최소로 하는 수량 $\{x_{(jk)}\}$ 구하려면,

$$\frac{\partial E(Y - \widehat{Y})^2}{\partial x_{(jk)}} = 0, (j = 0, 1, \dots, m; k = 1, 2, \dots, l_j) \tag{2.25}$$

되는 조건에서 도입된 식 (2.25)의 연립일차방정식을 $\{x_{(jk)}\}$ 에 관하여 풀면 된다.

$$\sum_{u=1}^m \sum_{v=1}^{l_u} E(\delta_{(jk)} \cdot \delta_{(uv)}) x_{(uv)} = E(Y \delta_{(jk)}) \tag{2.26}$$

$$(j = 1, 2, \dots, m; k = 1, 2, \dots, l_j)$$

그리고 최적해 $\{x_{(jk)}^*\}$ 을 이용한 예측량 \widehat{Y} 의 기대값 $E(\widehat{Y})$, 평균평방오차 $E(Y - \widehat{Y})^2$ 의 관계는 다음 식과 같다.

$$E(Y) = E(\widehat{Y})$$

$$\min E(Y - \widehat{Y})^2 = \sigma_Y^2 (1 - \rho_{Y\widehat{Y}}^2)$$

$$\rho_{Y\widehat{Y}} = \frac{\sqrt{\sum_{j=1}^m \sum_{k=1}^{l_j} E(Y \delta_{(jk)} x_{(jk)}^*) - \{E(Y)\}^2}}{\sigma_Y} \tag{2.27}$$

여기서, $\rho_{\widehat{Y}} : \text{중상관계수.}$

$\sigma_Y^2 : \text{실측값 } Y \text{의 분산}$

이것으로부터 $\sigma_{\widehat{Y}}$ 의 값이 1에 가까운 경우에 \widehat{Y} 에 의해 Y 의 예측이 된다. 이상은 모집단에 관하여 생각한 것이지만, 이것은 실제의 데이터 해석에서 Sample로 치환하여 계산 처리하면 된다. 그리고 $E(Y, \delta_{(jk)})$, $E(\delta_{(jk)} \delta_{(uv)})$ 를 표본값으로 변환한다. 또 Sample의 각 값을 정의한다.

$n : \text{Sample 수}$

$n_{(jk)} : \text{Item} \cdot \text{Category } C_{(jk)} \text{에 해당 Sample 수}$

$n_{(jk)(uv)} : \text{Item} \cdot \text{Category } C_{(jk)} \text{의 해당 Sample}$

$(j, u = 1, 2, \dots, l_j; v = 1, 2, \dots, l_u)$

y_i 의 개체 i (제 i 번째 Sample 수)의 실측량 Y 로 하면,

$$E(\delta_{(jk)} \delta_{(uv)}) = \frac{1}{n} \sum_{i=1}^n \delta_{i(jk)} \delta_{i(uv)} = \frac{n_{(jk)(uv)}}{n}$$

$$E(Y \delta_{(jk)}) = \frac{1}{n} \sum_{i=1}^n y_i \delta_{i(jk)} = \frac{1}{n} \sum_{i=1}^{n_{(jk)}} y_i$$

단, $n_{(jk)(jk)} = n_{(jk)}$, $n_{(jk)(jv)} = 0$, ($k \neq v$)

기호 $\sum_{i=1}^{n_{(jk)}}$ 는 $C_{(jk)}$ 에 해당하는 개체 i 만의 합이다.

식 (2.26)에 Sample 값을 대입하면,

$$\sum_{u=1}^m \sum_{v=1}^{l_u} \frac{n_{(j,k)(u,v)}}{n} x_{(u,v)} = \frac{1}{n} \sum_{i=1}^{n_{(j,k)}} y_i \quad (2.28)$$

($j = 1, 2, \dots, m; k = 1, 2, \dots, l_j$)

식 (2.28)의 양변에 n 을 곱하면 식 (2.29)이 Sample에 대한 식이 된다.

$$\sum_{u=1}^m \sum_{v=1}^{l_u} n_{(j,k)(u,v)} x_{(u,v)} = \sum_{i=1}^{n_{(j,k)}} y_i \quad (2.29)$$

($j = 1, 2, \dots, m; k = 1, 2, \dots, l_j$)

외적기준이 수량 벡터로서 얻어진 경우에 대해서도 같은 방법으로 분석할 수 있다. 여러 가지 구성별 외적기준을 Y_1, Y_2, \dots, Y_T 로 얻을 수 있으며, 이런 경우는 설명 Item X_1, X_2, \dots, X_m 과 외적기준 Y_1, Y_2, \dots, Y_T 는 대응한다. 즉, 동일 설명 Item으로부터 Y_1, Y_2, \dots, Y_T 을 정도 높게 예측할 수 있다면 좋다. 이를 위해 설명 Item 예측과 다차원적인 수량을 주는 것이 된다. 그것에는 설명 Item · Category · 벡터를 고려하여 그 양 벡터간의 벡터 중상관을 고려하여 최대가 되도록 설명 Item을 수량화한다. 이것은 각각 외적기준 Y_T 를 각별히 고려하여 수량화를 하는 것과 같다. 즉, 벡터형태로 외적기준이 얻어지는 경우에는 각각 1차원으로 고려하여 그것에 수량화를 행하는 것이 좋고, 또 동일 설명 Item에 대해서 얼마의 차이가 있고 외적기준이 있는 경우에는 다음과 같은 수량을 구한다.

다음으로, Sample이 설명 Item중 Category에 1개만 해당하는 것이 아니고, 몇 개의 Category에 해당하는지를 설명하는 경우를 다룬다 이것은 식 (2.22)을 다음과 같이 전개함으로써 Item · Category 수량을 구할 수 있다.

$$(\text{예측량}) \quad \widehat{\mathcal{Y}} = X_1 + X_2 + \cdots + X_m = \sum_{j=1}^m X_j$$

$$\text{단, } X_j = \sum_{k=1}^{l_j} p_{(j,k)} x_{(j,k)}$$

$p_{(j,k)}$ 는 $C_{(j,k)}$ 에 해당하는 비율로서 $\sum_{k=1}^{l_j} p_{(j,k)} = 1$ ($j = 1, 2, \dots, m$) 이다.

예측량 $\widehat{\mathcal{Y}}$ 를 아래와 같이 정의하면 해 $\{x_{(j,k)}\}$ 는 다음과 같은 연립 일차방정식의 해도 구해진다.

$$\sum_{u=1}^m \sum_{v=1}^u E(p_{(j,k)} \cdot p_{(u,v)}) x_{(u,v)} = E(Y p_{(j,k)}) \quad (2.30)$$

$$(j = 1, 2, \dots, m; k = 1, 2, \dots, l_j)$$

$$\text{단, } E(p_{(j,k)} \cdot p_{(u,v)}) = \frac{1}{n} \sum_{i=1}^n p_{i(j,k)} p_{i(u,v)}$$

$$E(Y \cdot p_{(j,k)}) = \frac{1}{n} \sum_{i=1}^n p_{i(j,k)} y_i$$

$p_{i(j,k)}$ 은 $C_{(j,k)}$ 에 sample i 가 강하게 해당하는 경우, y_i 는 Sample i 의 외적기준 수량이다.

3) 수량을 구하는 계산처리

수량 $\{x_{(j,k)}\}$ 를 구하기 위해 해당하는 식(식 (2.29)을 행렬로 표현한 식)의 행렬 A 는 $(m-1)$ 개의 링크 누락이 있기 때문에 기계적으로 Item X_1 을 제외하고 Item X_2, X_3, \dots, X_m 의 제1 Category $C_{(21)}, C_{(31)}, \dots, C_{(m1)}$ 의 개소의 행렬 A 의 행과 열의 각각 요소, 벡터 B 의 그것들의 Category에 대응할 요소를 삭제하고 해석한다.

즉, 처음부터 $x_{(21)} = x_{(31)} = costs = x_{(m1)} = 0$ 으로서 연립방정식 $AX = B$ 를 해석한다.

실제에는 표5의 행렬과 벡터의 요소를 삭제한 것을 행렬 A^* 와 벡터 B^* 로 했을 때의 연립 일차방정식 $A^*X^* = B^*$ 을 해석한다. 또한 $A^*X^* = B^*$ 의 해 $\{x^*_{(j,k)}\}$ 를 각 Item내의 수량 평균을 0으로 기준화하여 분석한다.

Item 평균을 0에 기준화 할 때에는 기준화 수량을 $\{x_{(j,k)}'\}$ 하면,

$$x_{(j,k)}' = x^*_{(j,k)} - \overline{X_j} \quad (2.31)$$

$$\text{단, } \overline{X} = \frac{1}{n} \sum_{k=1}^{l_j} n_{(j,k)} x^*_{jk} \quad (j = 1, 2, \dots, m; k = 1, 2, \dots, l_j)$$

구한다. 이들의 수량에 따른 Category의 총수 ($L = \sum_{j=1}^m l_j$)의 차원으로 분석을 한다.

4) 예측에 좋은 척도

각 Item · Category 의 수량 $\{x_{(jk)}^*\}$ 이 구해지고 그 수량을 이용해서 예측량 $\widehat{Y} = \sum_{j=1}^m \sum_{k=1}^{I_j} \delta_{(jk)} x_{(jk)}^*$ 를 구했을 때, 이 \widehat{Y} 에 어느 정도 예측이 잘 되고 있을까, 또한 설명특성으로써 이용한 개개의 Item이 예측에 대해서 어느 정도 기여하고 있을까 등의 척도로써 전자에서는 중상관계수, 후자에서는 편상관계수 또는 범위를 이용한다.

(1) 중상관계수와 편상관계수

외적기준 Y (실측량 : 목적특성)와 총합량 \widehat{Y} (예측량 : 몇 가지의 설명특성을 선형식으로 합성한 량)의 상관계수의 중상관계수가 좋으려면, 이 계수의 값이 1에 가까울수록 예측이 좋다고 말한다. 수량화에서 중상관계수 $\gamma_{Y\widehat{Y}}$ 는 식 (2.27)에서 구할 수 있다. Sample에서의 식 (2.27)은 다음과 같다.

$$\gamma_{Y\widehat{Y}} = \frac{\sqrt{\frac{1}{n} (X^*, B^*) - \overline{Y}^2}}{S_Y} \quad (2.32)$$

단, n : Sample 수,

$$\overline{Y} : \text{실측값의 평균값} \quad \overline{Y} = \frac{1}{n} \sum_i^n y_i$$

$$S_Y : \text{실측량의 표준편차} \quad S_Y = \sqrt{\frac{1}{n} \sum_{i=1}^n (y_i - \overline{Y})^2}$$

(X^*, B^*) : 수량 $\{x_{(jk)}^*\}$ 를 구할 때의 $A^* B^* = B^*$ 의 해에 벡터 X^* 와

$$\text{정수 } B^* \text{의 벡터 내 } (X^*, B^*) = \sum_j \sum_k x_{(jk)}^* \left(\sum_i^{(jk)} y_i \right)$$

또한 Item · Category의 수량 $\{x_{(j,k)}^*\}$ 에서 각각의 개체예측량은 다음 식

$$\widehat{y}_j = \sum_j \sum_k \delta_{i(j,k)} x_{(j,k)}^*$$

로 구하고, 가장 일반적인 중상관계수를 구하는 식으로 계산한다.

$$r_{Y\widehat{Y}} = \frac{C_{Y\widehat{Y}}}{S_Y S_{\widehat{Y}}} \quad (2.33)$$

$$\text{단, } S_Y : \text{실측량의 표준편차 } S_Y = \sqrt{\frac{1}{n} \sum_{i=1}^n (y_i - \overline{Y})^2}$$

$$\overline{Y} : \text{실측량의 평균치 } \overline{Y} = \frac{1}{n} \sum_{i=1}^n y_i$$

$$S_{\widehat{Y}} : \text{예측량의 표준편차 } S_{\widehat{Y}} = \sqrt{\frac{1}{n} \sum_{i=1}^n (\widehat{y}_i - \overline{\widehat{Y}})^2}$$

$$\overline{\widehat{Y}} : \text{예측량의 평균값 } \overline{\widehat{Y}} = \frac{1}{n} \sum_{i=1}^n \widehat{y}_i$$

$$C_{Y\widehat{Y}} : \text{실측량과 예측량의 공분산}$$

$$C_{Y\widehat{Y}} = \sqrt{\frac{1}{n} \sum_{i=1}^n (y_i - \overline{Y})(\widehat{y}_i - \overline{\widehat{Y}})}$$

일반적인 중상관계수를 구하는 방법은 상관행렬을 구하여 그 역행렬의 요소로부터 계산한다. 상관행렬의 요소는 각 특성간의 상관계수로 구성되어 있다. 상관행렬 R 은 다음과 같이 표시한다.

$$R = \begin{bmatrix} \mathbb{Y}_{X_1 X_1} & \mathbb{Y}_{X_1 X_2} & \cdots & \mathbb{Y}_{X_1 X_m} & \mathbb{Y}_{X_1 Y} \\ \mathbb{Y}_{X_2 X_1} & \mathbb{Y}_{X_2 X_2} & \cdots & \mathbb{Y}_{X_2 X_m} & \mathbb{Y}_{X_2 Y} \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ \mathbb{Y}_{X_m X_1} & \mathbb{Y}_{X_m X_2} & \cdots & \mathbb{Y}_{X_m X_m} & \mathbb{Y}_{X_m Y} \\ \mathbb{Y}_{Y X_1} & \mathbb{Y}_{Y X_2} & \cdots & \mathbb{Y}_{Y X_m} & \mathbb{Y}_{Y Y} \end{bmatrix}$$

단, $\mathbb{Y}_{X_j} \mathbb{Y}_{X_j X_j} = 1 \quad (j, u = 1, 2, \dots, m)$

$$\mathbb{Y}_{Y Y} = 1$$

$$\mathbb{Y}_{X_j X_u} = \mathbb{Y}_{X_u X_j}$$

$$\mathbb{Y}_{X_j Y} = \mathbb{Y}_{Y X_j}$$

$\mathbb{Y}_{X_j X_u}$: 설명특성의 Item X_j 와 Item X_k 의 상관계수

$\mathbb{Y}_{X_j Y}$: 설명특성의 Item X_j 와 외측기준 Y 의 상관계수

이 상관행렬 R 의 역행렬 R^{-1} 를 다음과 같이 표현하면,

$$R^{-1} = \begin{bmatrix} \mathbb{Y}_{X_1 X_1}^{(-1)} & \mathbb{Y}_{X_1 X_2}^{(-1)} & \cdots & \mathbb{Y}_{X_1 X_m}^{(-1)} & \mathbb{Y}_{X_1 Y}^{(-1)} \\ \mathbb{Y}_{X_2 X_1}^{(-1)} & \mathbb{Y}_{X_2 X_2}^{(-1)} & \cdots & \mathbb{Y}_{X_2 X_m}^{(-1)} & \mathbb{Y}_{X_2 Y}^{(-1)} \\ \mathbb{Y}_{X_3 X_1}^{(-1)} & \mathbb{Y}_{X_3 X_2}^{(-1)} & \cdots & \mathbb{Y}_{X_3 X_m}^{(-1)} & \mathbb{Y}_{X_3 Y}^{(-1)} \\ \mathbb{Y}_{X_m X_1}^{(-1)} & \mathbb{Y}_{X_m X_2}^{(-1)} & \cdots & \mathbb{Y}_{X_m X_m}^{(-1)} & \mathbb{Y}_{X_m Y}^{(-1)} \\ \mathbb{Y}_{Y X_1}^{(-1)} & \mathbb{Y}_{Y X_2}^{(-1)} & \cdots & \mathbb{Y}_{Y X_m}^{(-1)} & \mathbb{Y}_{Y Y}^{(-1)} \end{bmatrix}$$

$$\text{단, } \mathbb{Y}_{X_j X_k}^{(-1)} = \mathbb{Y}_{X_j Y_j}^{(-1)}, \quad \mathbb{Y}_{X_j Y}^{(-1)} = \mathbb{Y}_{Y X_j}^{(-1)}$$

상관행렬 R 과 그의 역행렬 R^{-1} 는 대칭행렬이다. 상관행렬 R 의 역행렬 R^{-1} 을 구하면 그 요소를 이용하여 중상관계수를 구할 수 있다. 또한 후술하는 편상관계수와 그 역행렬의 요소에서 계산할 수 있다.

<상관계수>

$$\gamma_{I, X_1 X_2, \dots, X_m} = \sqrt{1 - \frac{1}{\gamma_{I I}^{(-1)}}} \quad (2.34)$$

<편상관계수>

$$\gamma_{I X_j, X_1 X_2, \dots, X_m} = - \frac{\gamma_{X_j I}^{(-1)}}{\sqrt{\gamma_{X_j X_j}^{(-1)} \gamma_{I I}^{(-1)}}} \quad (j = 1, 2, \dots, m) \quad (2.35)$$

Item X_j 와 X_u 간의 상관계수 $\gamma_{X_j X_u}$ 와 Item X_j 와 외적기준 I 간의 상관계수 $\gamma_{X_j I}$ 는 Item · Category의 수량 $\{x_{(j k)}^*\}$ 이 구해지면 식에서 계산할 수 있다. 표6에는 수량화의 Cross집계를 나타내었다.

(a) Item X_j 와 X_u 간의 상관계수 $\gamma_{X_j X_u}$ 를 구하는 방법

Item · Category의 수량 $\{x_{(j k)}^*\}$ 도 수량계산을 이용하여 행렬 A 요소 $\{n_{(j k)(u v)}\}$ 로부터 계산된다.

$$\gamma_{X_j X_u} = \frac{C_{X_j X_u}}{S_{X_j} S_{X_u}} \quad (2.36)$$

$$\text{단, } \overline{X_j} = \frac{1}{n} \sum_{k=1}^{l_j} n_{(j k)} x_{(j k)}^* \quad \overline{X_u} = \frac{1}{n} \sum_{v=1}^{l_u} n_{(u v)} x_{(u v)}^*$$

$$S_{X_j} = \sqrt{\frac{1}{n} \sum_{k=1}^{l_j} n_{(j k)} (x_{(j k)}^*)^2 - \overline{X_j}^2}$$

$$S_{X_u} = \sqrt{\frac{1}{n} \sum_{v=1}^{l_u} n_{(u v)} (x_{(u v)}^*)^2 - \overline{X_u}^2}$$

$$C_{X_j X_u} = \frac{1}{n} \sum_{k=1}^{l_j} \sum_{v=1}^{l_u} n_{(j k)(u v)} x_{(j k)}^* x_{(u v)}^* - \overline{X_j} \overline{X_u}$$

$\overline{X_j}, \overline{X_u}$ 는 평균치

S_{X_j}, S_{X_u} 는 표준편차

$C_{X_j X_u}$ 는 공분산

표 6. 수량화 Cross 집계

	Item X_u							
	Category		$C_{(u1)}$	$C_{(u2)}$...	$C_{(uv)}$...	$C_{(ul_u)}$
	Category	수 량	$x^*_{(u1)}$	$x^*_{(u2)}$...	$x^*_{(uv)}$...	$x^*_{(ul_u)}$
Item X_j	$C_{(j1)}$	$x^*_{(j1)}$	$n_{(j1)(u1)}$	$n_{(j1)(u2)}$...	$n_{(j1)(uv)}$...	$n_{(j1)(ul_u)}$
	$C_{(j2)}$	$x^*_{(j2)}$	$n_{(j2)(u1)}$	$n_{(j2)(u2)}$...	$n_{(j2)(uv)}$...	$n_{(j2)(ul_u)}$
	\vdots	\vdots	\vdots	\vdots	\vdots	\vdots	\vdots	\vdots
	$C_{(jk)}$	$x^*_{(jk)}$	$n_{(jk)(u1)}$	$n_{(jk)(u2)}$...	$n_{(jk)(uv)}$...	$n_{(jk)(ul_u)}$
	\vdots	\vdots	\vdots	\vdots	\vdots	\vdots	\vdots	\vdots
	$C_{(jl)}$	$x^*_{(jl)}$	$n_{(jl)(u1)}$	$n_{(jl)(u2)}$...	$n_{(jl)(uv)}$...	$n_{(jl)(ul_u)}$
	합	계	$n_{(u1)}$	$n_{(u2)}$...	$n_{(uv)}$...	$n_{(ul_u)}$
								n

(b) Item X_j 와 외측기준 Y 간의 상관계수를 구하는 방법

Item · Category의 수량 $\{x^*_{(jk)}\}$ 와 수량계산에 사용된 벡터 B 의 요소 $\{\sum_{i=1}^{m_{(jk)}} y_i\}$ 로부터 계산한다.

$$r_{X_j Y} = \frac{C_{X_j Y}}{S_{X_j} S_Y} \quad (2.37)$$

$$\overline{Y} = \frac{1}{n} \sum_{i=1}^n y_i$$

$$S_Y = \sqrt{\frac{1}{n} \sum_{i=1}^n y_i^2 - \bar{Y}^2}$$

$$C_{X_j Y} = \frac{1}{n} \sum_{k=1}^{I_j} x_{jk}^* \left(\sum_{i=1}^{n_{(jk)}} y_i \right) - \bar{Y} X_j, \quad (j = 1, 2, \dots, m)$$

(X_j, S_x 는 식 (2.37)에 나타낸 것과 같다.)

표7에는 외적기준의 Category화를 나타낸 것이다.

표 7. 외적기준의 Category화

	Category	수 량	외적기준의 합
Item X_j	$C_{(j1)}$	$x_{(j1)}^*$	$\sum_i^{(j1)} y_i$
	$C_{(j2)}$	$x_{(j2)}^*$	$\sum_i^{(j2)} y_i$
	\vdots	\vdots	\vdots
	$C_{(jk)}$	$x_{(jk)}^*$	$\sum_i^{(jk)} y_i$
	\vdots	\vdots	\vdots
	$C_{(jl)}$	$x_{(jl)}^*$	$\sum_i^{(jl)} y_i$

(2) 편상관계수와 범위

수량화이론에는 설명 Item X_j , $j = 1, 2, \dots, m$ 의 외적기준 Y 의 기여도를 각 Item내의 Category의 수량 $\{x_{(jk)}'\}$ 의 범위를 R_j 를 이용한다.

$$R_j = \max x_{(jk)}' - \min x_{(jk)}'; j = 1, 2, \dots, m.$$

이때 설명 Item의 기여 순서는 범위 R_j 이 큰 순위로 표현한다. 그 이유로서는 예측량을 만드는 방법이 $\widehat{Y} = \sum_{j=1}^m X_j$; $X_j = \sum_{k=1}^{I_j} \delta(jk) x_{(jk)}'$ 로 있기 때문에 Item 중 Category 수량의 범위(폭)가 클수록 예측량 \widehat{Y} 에 대한 영향력이 커지기 때문이다.

반대로 범위가 작은 Item은 그것을 더하여도 각각 $\widehat{\mathcal{P}} = \{\widehat{\mathcal{P}}_i\}$ 에 거의 영향을 미치지 않는다. 범위라는 통계량이 나타나 있지 않는 경우에는 중회귀분석 양식에 편상관계수를 이용한다. 편상관계수로는 설명특성 X_1, X_2, \dots, X_m 과 목적특성 Y 를 포함한 $m+1$ 차원의 공간상에 특성 X_j 와 Y 의 상관관계를 조사할 때 단순한 X_j 와 Y 의 상관계수를 취급하는 것이 아니고 X_j 와 Y 에 포함되어 있는 다른 특성 X_1, X_2, \dots, X_m 의 성분의 영향을 제거한 후 X_j^* 와 Y^* 간의 상관계수를 나타낸 것이다.

이 상관계수 $\mathbb{Y}_{X_j, Y}^*$ ($= \mathbb{Y}_{X_j, Y, X_1, X_2, \dots, X_m}$)을 갖고 설명 Item X_j 의 예측에 대한 설명력의 척도로 한다. 편상관계수를 구하는 방법은 상관행렬의 역행렬의 요소를 이용하여 식 (2.36)으로 구한다. 또한 편상관계수와 중상관계수 간에는 다음과 같은 중요한 관계가 있다.

$$\begin{aligned}\mathbb{Y}_{Y, X_q, X_1, X_2, \dots, X_{q-1}}^2 &= \frac{\mathbb{Y}_{Y, X_1, X_2, \dots, X_q}^2 - \mathbb{Y}_{Y, X_1, X_2, \dots, X_{q-1}}^2}{1 - \mathbb{Y}_{X_q, X_1, X_2, \dots, X_{q-1}}^2} \\ &= 1 - \frac{1 - \mathbb{Y}_{Y, X_1, X_2, \dots, X_q}^2}{1 - \mathbb{Y}_{X_q, X_1, X_2, \dots, X_{q-1}}^2}\end{aligned}$$

단, $\mathbb{Y}_{Y, X_q, X_1, X_2, \dots, X_{q-1}}$ 는 설명변수(X_1, X_2, \dots, X_q)를 목적특성 Y 로의 $q+1$ 차원에 의하여 목적변량 Y 를 설명변량 X_q 와의 편상관계수, $\mathbb{Y}_{Y, X_q, X_1, X_2, \dots, X_{q-1}}$ 는 설명특성(X_1, X_2, \dots, X_{q-1})과 목적특성 Y 의 q 차원에 의한 중상관계수이다.

그리고 $\mathbb{Y}_{Y, X_q, X_1, X_2, \dots, X_q}$ 는 설명특성 X_q 를 1개 더한 $q+1$ 차원에 의한 목적특성 Y 와의 중상관계가 된다. 편상관계수의 2승 $\mathbb{Y}_{Y, X_q, X_1, X_q, X_1, X_2, \dots, X_{q-1}}^2$ 은 설명특성 X_q 를 더하기 전의 중상관계수 $\mathbb{Y}_{Y, X_q, X_1, X_2, \dots, X_{q-1}}$ 에 더해서 중상관계수 $\mathbb{Y}_{Y, X_q, X_1, X_2, \dots, X_q}$ 가 어느 정도 상승할까가 목표의 척도가 된다. 즉, 설명특성 X_q 는 어느 정도 예측에 기여할까 하는 것이 목표가 된다.

제 3장 교통사고데이터 특성 분석

본 장에서는 국도17호선(전주~남원)에 있어서 교차로를 대상으로 교통량(연평균 일교통량:AADT), 기하구조(종단경사, 곡선반경, 횡단경사, 도로폭)를 고려하여 먼저 사고유형(후미충돌, 정면충돌, 직각충돌, 측면충돌)에 대해서 교통사고발생분포를 분석하였다.

그리고, 각 지점 특성별 사고유형(후미충돌, 정면충돌, 직각충돌, 측면접촉), 사고위치(접근부, 유출부, 횡단보도, 교차로내, 기타 등)로 구분해 교차분석 및 빈도분석을 하였으며, 국도 17호선(전주~남원)내에서의 교통 사고의 특성을 알아보기 위하여 월별 교통사고분포, 요일별 교통사고분포, 운전자의 사고원인 분포(신호위반, 중앙선침범, 과속, 음주운전, 안전운전 불이행, 기타), 보행자 사고원인 분포(횡단, 무단횡단, 승하차, 차로내 사고, 차로외 사고, 기타)등으로 구분하여 빈도분석을 하였다.

또한, 교통사고 예측모형을 구축하기 위해 교통사고에 영향을 주는 요인을 분석하고자 사고건수와 교통량 및 교차로 기하구조 요인과의 상관관계를 분석하였다.

제 1절 단순빈도분석

1) 교차로 사고위치별 교통사고발생분포

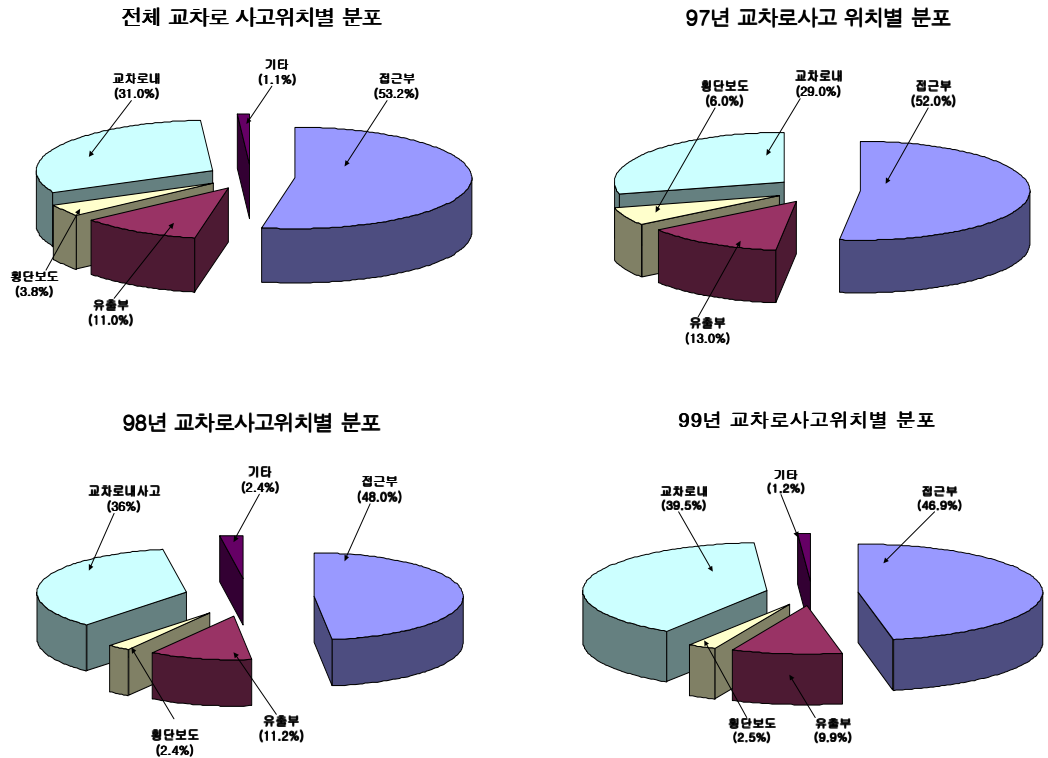


그림 4. 교차로 사고위치별 교통사고발생 분포

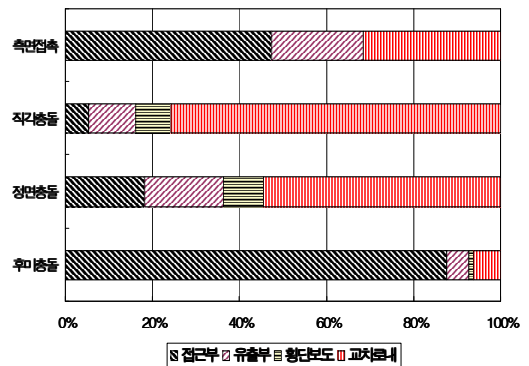
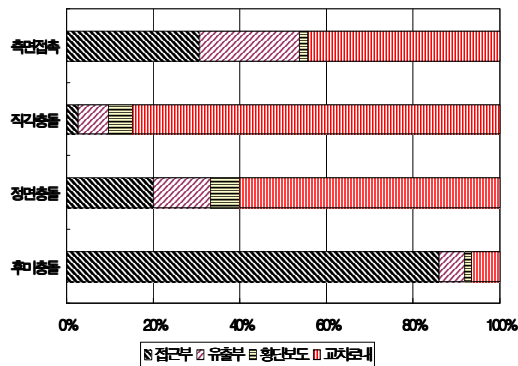
각 연도에 따른 교차로 사고위치별 분포와 전체적인 사고위치별 교통사고발생 분포를 그림 4에서 나타내었다. 교차로 사고위치별 교통사고분포를 나타낸 결과 접근부의 사고가 가장 많이 일어났고(전체:53.2%, '97년:52.0%, '98년:48.0%, '99년:46.9%), 그 다음으로 교차로 내에서 사고(전체:31.0%, '97년:29.0%, '98년:36.0%, '99년:39.5%)가 발생한 것으로 분석되었다.

특히, 접근부에서 교통사고 건수의 절반이 발생하고 있어 국도에 있어서 교차로 접근부의 교통안전시설의 확충 및 보완이 필요하다고 생각된다.

2) 교차로 사고위치에 따른 교통사고유형별 발생분포

(1) 전체

(2) 97년



(3) 98년

(4) 99년

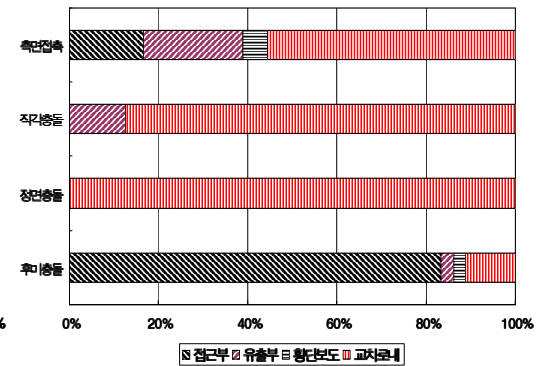
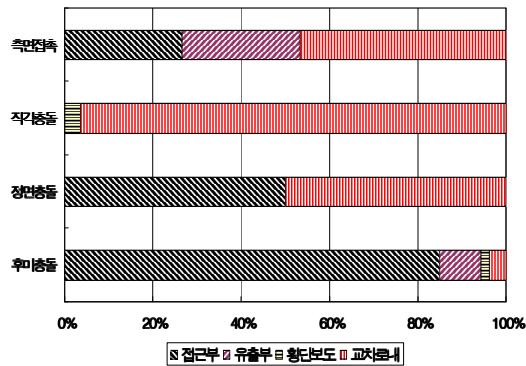


그림 5. 교차로 사고위치에 따른 교통사고유형별 분포

전체적으로 볼 때 후미충돌사고는 접근부에서 대체적으로 많이 일어났고, 직각충돌사고는 교차로내에서 많이 발생하는 것으로 나타났다. 또한 측면접속사고나 정면충돌사고는 교차로 내와 접근부에서 대체적으로 많이 발생하였다.

측면접속사고나 정면충돌사고와 같은 경우는 주방향으로 달리는 차량에 대해서 교차로에서 우회하려는 차량으로 인해 사고가 발생한 것으로 추정되며, 후미충돌사고는 교차로의 신호의 규제에 의해서 정지하려는 앞차와 뒷따라가는 차량이 앞차와의 거리를 확보하지 않은채 발생한 것이라고 판단된다.

3) 교차로 사고유형별 교통사고발생분포

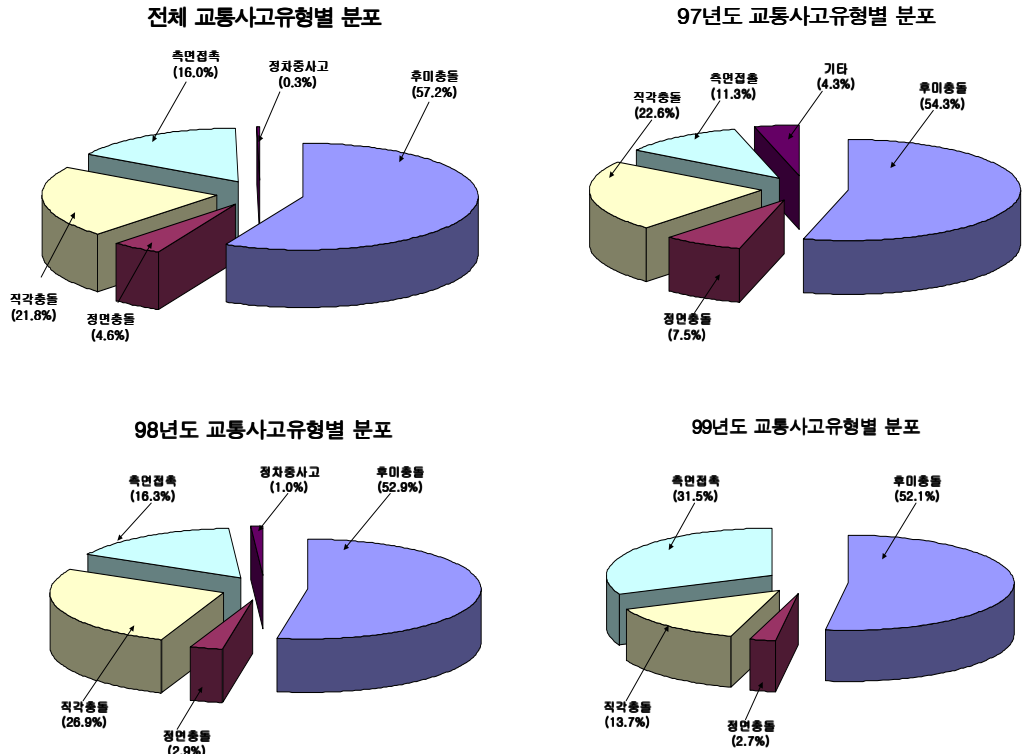


그림 6. 교차로 교통사고유형별 분포

각 연도에 있어서 교차로의 교통사고유형별 교통사고 발생분포를 분석한 결과 그림 6에서 볼 수 있듯이 전체적으로 볼 때 후미충돌(57.2%)이 교차로에 있어서 사고발생원인으로 가장 높고, 다음으로 직각충돌(21.8%), 측면접촉(16.0%)이 많이 발생하는 것으로 나타났다. 이는 과속하는 차량들이 신호의 규제를 받아 정지할 때 정지하는 앞차량과 뒤따라오는 뒷차와 이어지는 사고인 것으로 판단된다. 바꾸어 말하면, 앞차가 정지할 때 뒷차가 안전거리를 충분히 두지 않았기 때문인 것으로도 해석할 수 있다.

4) 월별 교차로 교통사고발생분포

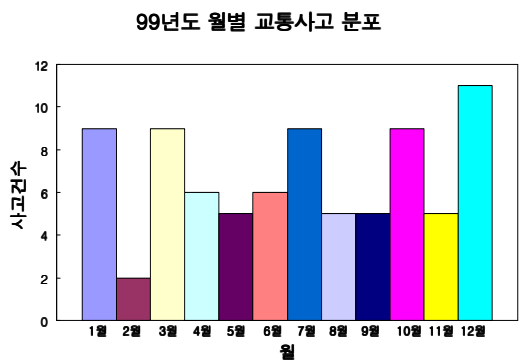
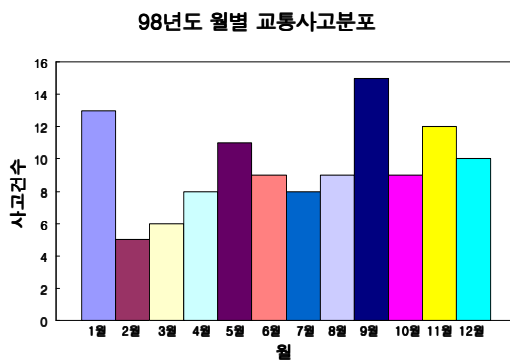
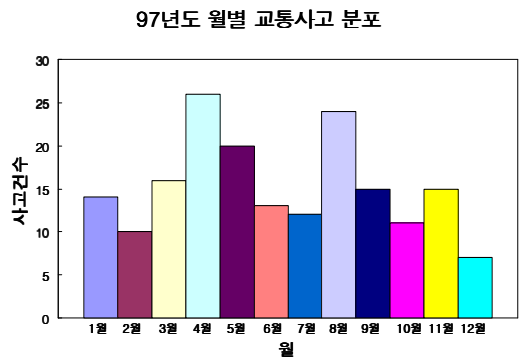
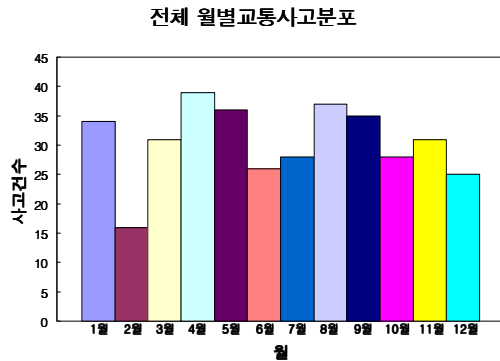


그림 7. 월별 교차로 교통사고발생 분포

교차로의 월별에 따른 교통사고발생 분포를 분석한 결과, 그림 7에서 볼 수 있는 바와 같이 전체적으로 4, 5월과 8, 9월에 사고가 많이 발생한 것으로 나타났는데 이는 날씨가 어느정도 춥지 않고, 바깥 외출하기에 적당한 달이기 때문인 것으로 판단된다. 또한, '97년의 경우 다소 차이는 많이 나지는 않지만 그 중에서 4월, 5월, 8월에 교통사고가 많이 발생했고, '98년에는 1월, 9월, 11월, 12월에 사고가 많이 발생했으며, '99년에는 1월과 3월, 7월, 10월, 12월에 교통사고가 많이 발생한 것으로 보아 특정한 날이 있거나 행사가 있는 달에 전주와 남원 사이에 위치한 나들이 장소나 남원 근교에 있는 지리산을 찾기 위한 차량들의 증가 때문이라고 판단된다.

5) 요일별 교차로 교통사고발생분포

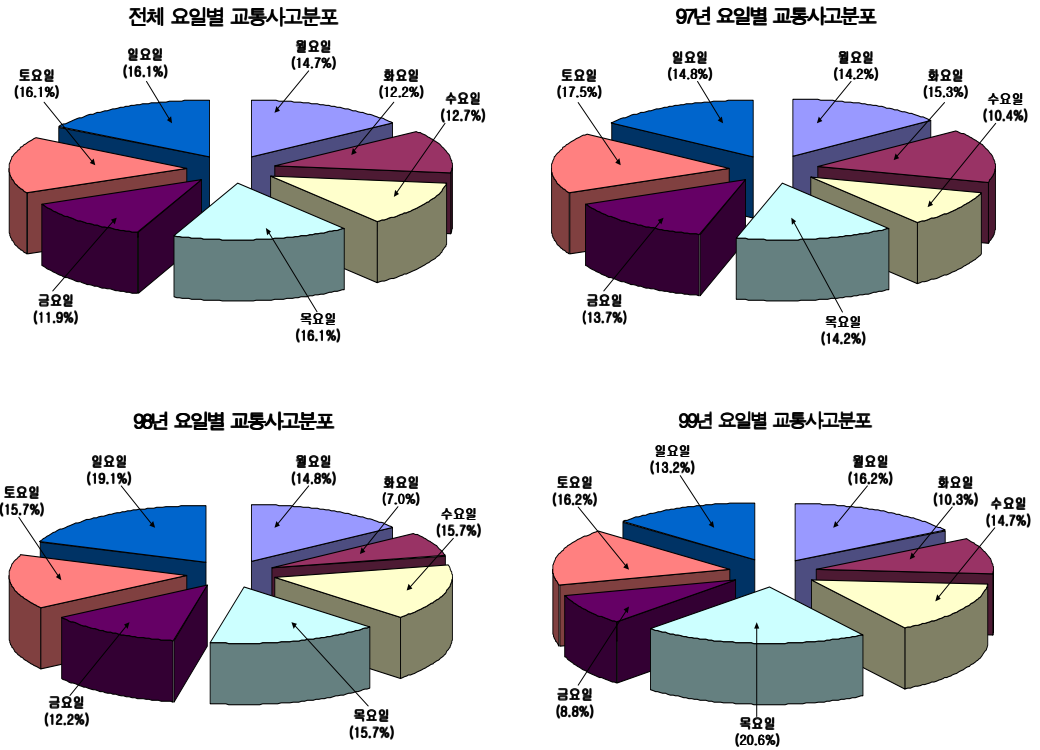


그림 8. 요일별 교차로 교통사고분포

요일별 교통사고 발생건수 추세를 볼 때 전체적으로는 그림 8에서와 같이 토요일과 일요일에 16.1%가 발생하여 가장 많았으며, 금요일에 11.9%로 가장 적었던 것으로 나타났다. 그 외 기타 평일의 경우에는 월요일과 목요일이 대체로 많은 분포를 나타냈다.

각 연도별로 볼 때도, 평일의 교통사고분포보다 휴일이나 주말의 교통사고분포가 대체로 많이 나타났다. 이처럼 평일에 비해 주말에 사고가 많은 것은 소득수준이 향상되고 또한 주변에 죽림 온천 등 휴양시설이나 남원의 근교인 지리산을 가기 위하여 남원을 지나야만 하는 원인과 남원의 관광요소 위치 등으로 인하여 주말 관광차량이 증가하고 있으며, 주말 휴무제가 광범위하게 실시되는 등 나들이 차량이 증가한데 원인이 있는 것으로 판단된다.

6) 보행자 사고원인별 교차로 교통사고발생분포

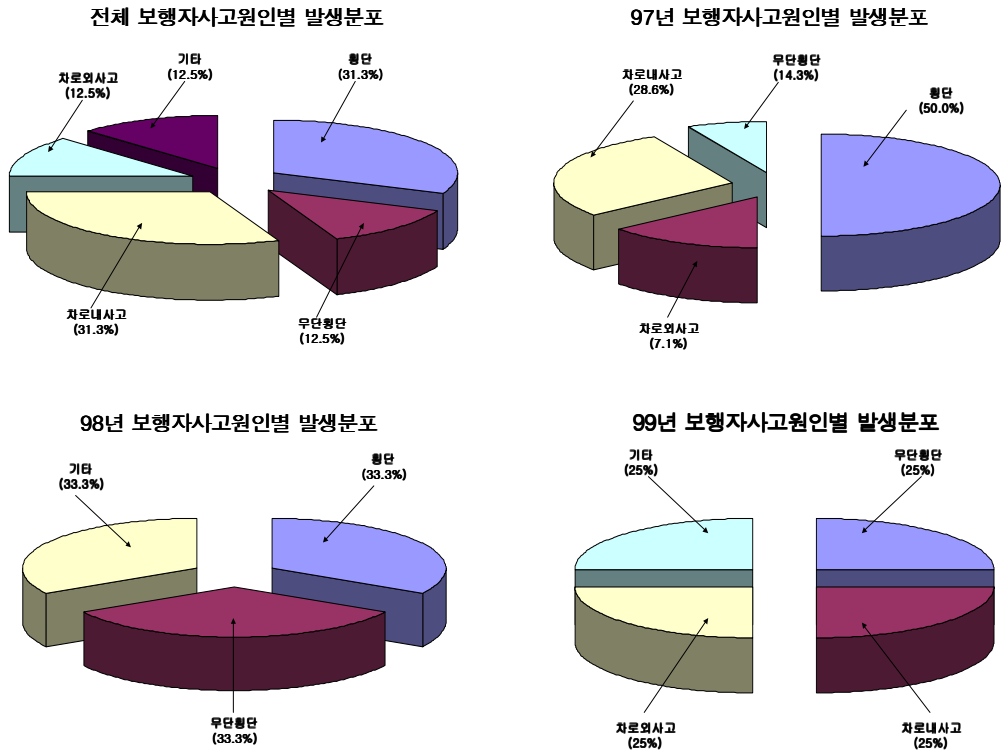


그림 9. 보행자사고원인별 교차로 교통사고발생분포

그림 9에서 볼 수 있듯이 교차로에 있어서 보행자에 관한 교통사고원인별로 분석한 결과, '97, '98, '99년 모두가 횡단보도 횡단시 부주의에 의한 사고가 가장 주된 원인으로 나타났다. 이는 본 연구 분석대상 교차로들은 국도상에 있어서 교차하는 지역이기 때문에 자동차들의 정지거리가 길어지고 이 때문에 사고의 위험도가 증가하는 경우와 분석 대상 지점을 통과하는 차량 속도가 높은 차량들이 신호에 의해 정지 규제를 받을 때 정지 시거가 부족하기 때문에 사고를 일으킬 위험도가 증가하는 것이라고 판단된다. 또한, 차량의 운행방향이 주방향에서 부방향으로 또는 부방향에서 주방향으로 우회전하는 경우, 운전자들이나 보행자의 부주의에서 오는 결과라 할 수 있겠다.

7) 운전자 사고원인별 교차로 교통사고발생분포

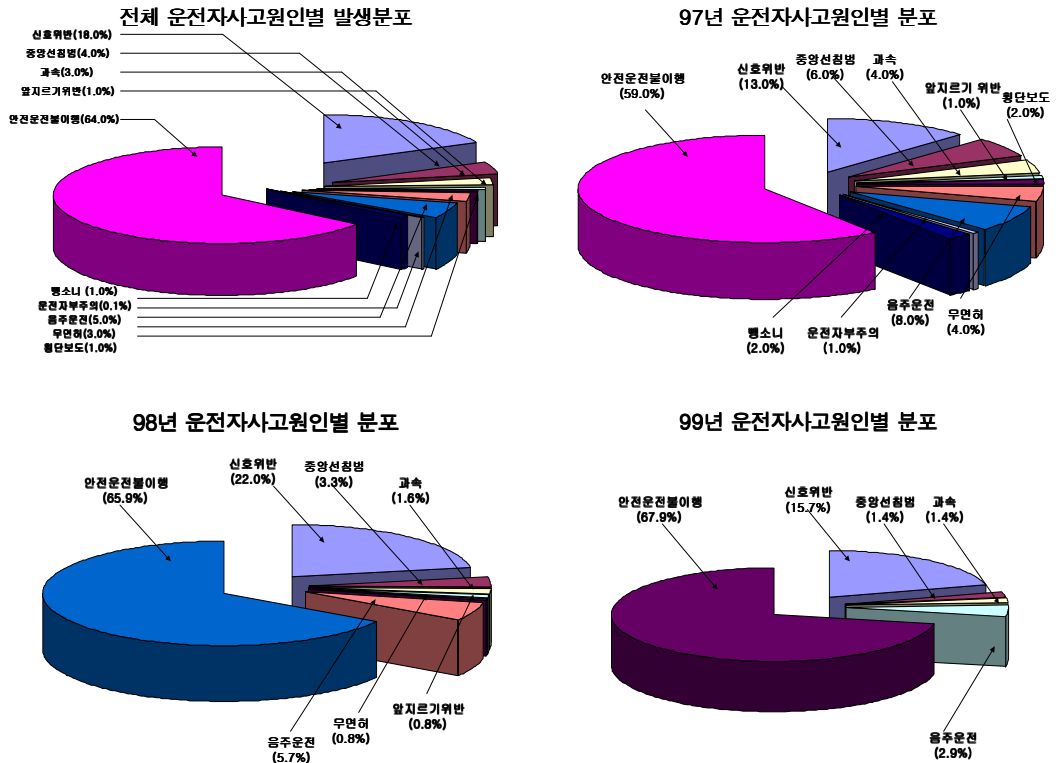


그림 10. 운전자 사고원인별 교차로 교통사고발생분포

각 연도에 따른 교차로에 있어서 운전자 사고원인에 대하여 분석한 결과 그림 10에서 볼 수 있듯이 지역과 지역을 연결하는 국도상의 교차로이기 때문에 '97, '98, '99년 모두가 안전운전불이행이 주된 원인으로 분석되었다. 또한 신호위반과 같이 운전자의 고의성이 부가된 교통사고의 요인은 운전자들이 교통안전에 대한 의식의 부재에 의해서 나타난 결과라 판단된다.

제 2절 상관관계분석(Correlate Analysis)

본 연구는 기존의 연구 자료를 감안하여 교통사고 서식에 나와있는 자료를 대상으로 국도상의 교차로 교통사고에 영향을 주는 요인의 상관관계를 분석하여 교통사고에 영향

이 있는 변수를 선정하였다.

전체 데이터와 '97년, '98년, '99년 교차로 교통사고 데이터의 사고건수와 교통량 및 교차로 기하구조 요인과의 상관관계를 분석한 결과 표8~표11와 같이 나타났다.

표 8. 전체 데이터의 교차로 교통사고 건수와 기하구조요인과의 상관분석

	사고건수	교통량	종단구배	곡선반경	횡단구배	도로폭
사고건수	1.000					
교통량	0.213**	1.000				
종단구배	0.079	-0.102	1.000			
곡선반경	-0.063	0.145**	0.049	1.000		
횡단구배	-0.227**	0.119*	0.192*	-0.072	1.000	
도로폭	-0.138**	0.080	-0.212**	0.164**	-0.020	1.000

표 9. '97년 데이터의 교차로 교통사고 건수와 기하구조요인과의 상관분석

	사고건수	교통량	종단구배	곡선반경	횡단구배	도로폭
사고건수	1.000					
교통량	0.025	1.000				
종단구배	0.186*	0.065	1.000			
곡선반경	0.095	0.149*	0.070	1.000		
횡단구배	-0.330**	0.121	0.171*	-0.080	1.000	
도로폭	-0.280**	0.067	-0.113	0.225**	0.052	1.000

표 10. '98년 데이터의 교차로 교통사고 건수와 기하구조요인과의 상관분석

	사고건수	교통량	종단구배	곡선반경	횡단구배	도로폭
사고건수	1.000					
교통량	0.356**	1.000				
종단구배	-0.006	-0.246**	1.000			
곡선반경	-0.063	0.206*	0.051	1.000		
횡단구배	-0.230*	-0.052	0.268**	-0.161	1.000	
도로폭	0.035	0.152	-0.340**	-0.173	-0.110	1.000

표 11. '99년 데이터의 교차로 교통사고 건수와 기하구조요인과의 상관분석

	사고건수	교통량	종단구배	곡선반경	횡단구배	도로폭
사고건수	1.000					
교통량	0.391**	1.000				
종단구배	-0.100	-0.186	1.000			
곡선반경	-0.236	-0.075	0.046	1.000		
횡단구배	0.049	0.271*	0.191	-0.160	1.000	
도로폭	0.027	-0.344**	-0.280*	-0.156	-0.217	1.000

주 : 1) '97년 : 183표본, '98년 : 115표본, '99년 : 68표본

2) * : 위험율 5%에서 유의

** : 위험율 1%에서 유의

사고건수에 영향이 있는 변수들을 선정하기 위해 교통량과 기하구조요인 변수들에 대해 상관분석을 실시한 결과 '97년의 경우에는 종단경사, 횡단경사, 도로폭이 교통사고건수에 영향이 있는 유의한 변수로 나타났고, '98년의 경우 교통량, 횡단경사 이었으며, '99년의 경우에는 교통량만이 교통사고건수에 영향을 미치는 유의한 변수임을 알 수가 있었다.

제 4장 교차로 교통사고 예측모형

제 1절 교통사고유형별 교통사고예측모형

본 연구는 교차로의 사고유형(후미충돌, 정면충돌, 직각충돌, 측면접촉)에 대한 각각의 다중선형회귀모형, 다중비선형회귀모형, 수량화 I 류 모형을 구축하였다.

1) 후미충돌에 대한 교통사고예측모형

(1) 후미충돌에 대한 다중선형회귀모형(Multiple Linear Regression Model)

표 12. 후미충돌에 대한 ANOVA 분석표

	오차제곱합	자유도	평균제곱	분산비	유의도
선형회귀	2769.495	5	543.899	11.546	0.000
잔 차	8683.093	181	47.973		
합 계	11,452.588	186			

표 13. 후미충돌에 대한 다중선형회귀모형 결과

속성	계수(B)	표준오차	베타(β)	t 값	유의도
상수	62.325	12.224		5.099	0.000
교통량	0.293	0.115	0.167	2.550	0.012
횡단경사	-55.920	9.010	-0.406	-6.206	0.000
도로폭	-1.343	0.552	-0.159	-2.435	0.016
운전자성별	-4.257	2.053	-0.135	-2.074	0.040
왕복차로	-3.206	1.003	-0.209	-3.198	0.002
설명력	$R^2 = 0.242$ ($Adjusted R^2 = 0.221$)				

후미충돌에 대해 사고요인을 독립변수로 사용한 다중선형회귀모형에서 결정계수 (R^2)는 24.2%로 비교적 낮게 나타났으며, 수정결정계수의 값도 22.1%로 종속변수를 잘 설명하고 있지 못하는 것으로 나타났다. 그렇지만, 교통사고의 회귀분석시 보여주는 기존의 교통사고에 대한 회귀분석의 결정계수 값과 비교했을 때 어느정도 설명력을 보여주고 있는 것으로 판단된다. 그 이유는 교통사고의 주요 요인은 인적요인(Human Factor), 차량적 요인(Vehicle Factor), 도로환경적요인(Environmental Factor)등의 요인으로 구분할 수 있는데 이 중에서 인적요인(Human Factor)이 교통사고에 있어서 가장 큰 비중을 차지하고 있기 때문에 실질적으로 우리가 교통사고를 분석할 때 자료를 획득함에 있어서 기록에 의한 자료만 얻을 수 있기 때문이고, 또 그 자료만으로 분석할 때 모형에 대한 값은 그다지 높지 않을 것이라고 판단되기 때문에 본 논문의 선형회귀모형에서의 결정계수 (R^2)값은 그다지 낮지 않은 것이라고 생각된다. 분산분석의 검정통계량 F의 값은 유의수준 5% 안에 들어 양호한 것으로 검정되었으며, t 통계값도 모든 예측변수들에서 유의한 것으로 나타났다.

또한, 모형의 설명변수 중에서 횡단경사 요인이 교차로 후미충돌 교통사고에 가장 많은 영향을 미치고 있음을 알 수 있었다.

(2) 후미충돌에 대한 다중비선형회귀모형(Multiple Non-Linear Regression Model)

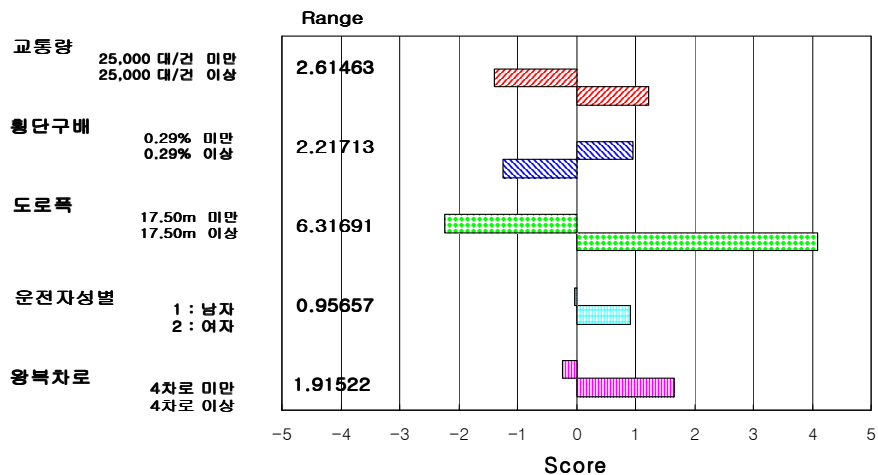
$$\hat{y} = e^{1.56695} \times (\text{교통량})^{0.69475} \times (\text{횡단경사})^{-3.65902} \\ \times (\text{도로폭})^{-1.46809} \times (\text{운전자성별})^{-1.33851} \times (\text{왕복차로})^{-1.13181}$$

$$R^2 = 0.278$$

다중선행회귀모형에서 사용한 설명변수를 대상으로 후미충돌에 대한 다중비선행회귀모형을 지수식으로 나타낸 결과 결정계수 (R^2)는 27.8%로 비교적 낮게 나타났으나 다중선행회귀모형에서 설명력을 나타낸 결정계수 (R^2)값보다는 높게 나타났다.

위 식을 보면 알 수 있듯이 교통량이 많을수록, 횡단경사가 작을수록, 도로폭이 작을수록, 운전자가 남성일수록, 왕복차로가 적을수록 교통사고가 많이 일어났던 것임을 알 수 있다.

(3) 후미충돌에 대한 수량화 I 류 모델



CORRELATION COEFFICIENT 0.61922

ESTIMATION ERROR 3.84800

그림 11. 후미충돌에 대한 수량화 I 류 모형

위 모형을 식으로 나타내어 보면 다음과 같다.

$$\begin{aligned} y = & -1.38422(x_{11}) + 1.23042(x_{12}) \\ & + 0.96361(x_{21}) - 1.25677(x_{22}) \\ & - 2.22950(x_{31}) + 4.08741(x_{32}) \\ & - 0.04092(x_{41}) + 0.91565(x_{42}) \\ & - 0.24580(x_{51}) + 1.66941(x_{52}) \end{aligned}$$

$$R^2 = 0.383$$

주 : x_{11} : 교통량 25,000 대/건 미만, x_{12} : 교통량 25,000 대/건 이상

x_{21} : 횡단경사 0.29% 미만, x_{22} : 횡단경사 0.29% 이상

x_{31} : 도로폭 17.50m 미만, x_{32} : 도로폭 17.50m 이상

x_{41} : 운전자 성별 (남자), x_{42} : 운전자 성별 (여자)

x_{51} : 왕복차로 4차로 미만, x_{52} : 왕복차로 4차로 이상

후미충돌에 대한 다중선형회귀모형과 다중비선형회귀모형에서 사용한 설명변수를 그대로 사용하여 수량화 I 류 이론을 이용한 모델을 구축한 결과 다중회귀모형(선형회귀모형, 비선형회귀모형)보다 모형의 설명력을 나타내는 결정계수 (R^2)값이 38.3%로 높게 나타났다.

2) 정면충돌에 대한 교통사고예측모형

(1) 정면충돌에 대한 다중선형회귀모형

표 14. 정면충돌에 대한 ANOVA 분석표

	오차제곱합	자유도	평균제곱	분산비	유의도
선형회귀	1.141	3	0.380	7.055	0.007
간 차	0.593	11	2.389E-02		
합 계	1.733	14			

표 15. 정면충돌에 대한 다중선형회귀모형 결과

속성	계수(B)	표준오차	베타(β)	t 값	유의도
상수	-0.774	0.488		-1.586	0.141
교통량	4.034E-02	0.014	0.534	2.965	0.013
횡단경사	3.321	0.896	0.668	3.706	0.003
기상상태	-0.225	0.108	-0.380	-2.090	0.061
설명력	$R^2 = 0.658$ ($Adjusted R^2 = 0.565$)				

정면충돌에 대한 다중선형회귀모형에서 결정계수 (R^2)는 65.8%로 대체적으로 높은 설명을 보이고 있으며, 수정결정계수값 또한 56.5%로 종속변수인 사고건수를 잘 설명하고 있는 것으로 분석되었다. 분산분석의 검정통계량 F의 값은 유의수준 5% 안에 들어 양호한 것으로 검정되었으며, t통계값도 모든 예측변수들에서 유의한 것으로 나타났다.

위 모형을 볼 때 후미충돌에서 교통사고건수에 횡단경사가 가장 많은 영향을 주었는데, 정면충돌에서도 마찬가지로 횡단경사가 정면충돌사고에 가장 많은 영향을 미치고 있음을 보여주고 있다. 또한, 기상상태의 요인도 정면충돌사고에 영향을 주고 있음을 알 수 있었다.

(2) 정면충돌에 대한 다중비선형회귀모형

$$\hat{y} = e^{-1.90249} \times (\text{교통량})^{0.97135} \\ \times (\text{횡단경사})^{1.02059} \times (\text{기상상태})^{-0.30466}$$

$$R^2 = 0.738$$

다중선형회귀모형에서 사용한 설명변수를 대상으로 정면충돌에 대한 다중비선형회귀모형을 지수식으로 나타낸 결과 결정계수 (R^2)는 73.8%로 매우 높은 설명력을 보여주었고, 다중선형회귀모형과 비교해볼 때 지수식으로 표현한 다중비선형회귀모형의 설명력이 더 높게 분석되었다. 여기서, 교통량이 많을수록 정면충돌사고가 많이 일어나고, 기상상

태가 맑은 날보다는 비나 눈이 오는 날에 교통사고가 많이 일어나는 것으로 해석할 수 있다. 이는 기상 조건이 악화될수록 도로 표면의 마찰계수가 적어져 제동거리가 짧아지고 시거가 나빠짐으로써 가시거리 확보가 불리해져 운전자의 주변 상황에 대한 대처 능력이 저하되어 교통사고발생이 많아지는 것으로 판단된다.

(3) 정면충돌에 대한 수량화 I 류 모형

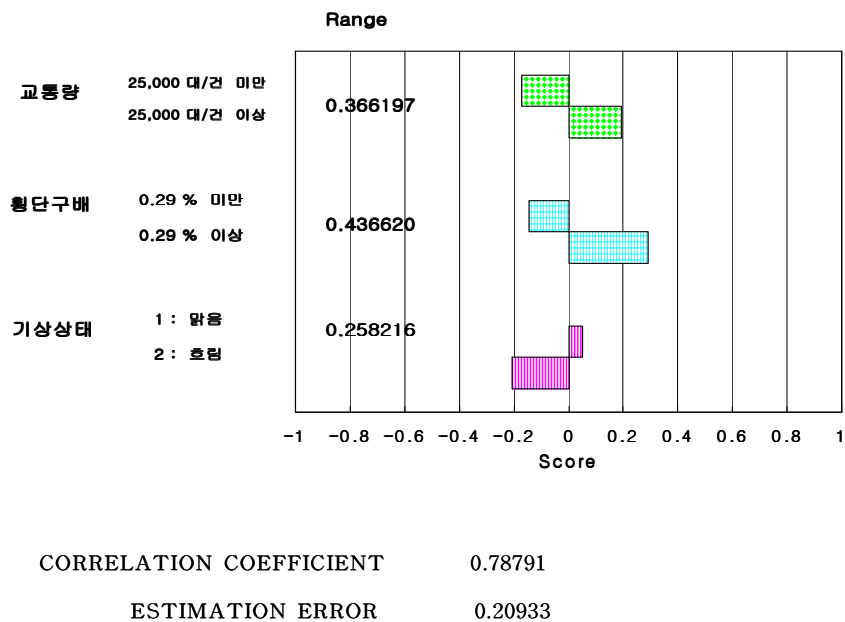


그림 12. 정면충돌에 대한 수량화 I 류 모형

위 모델을 식으로 나타내면 다음과 같다.

$$\begin{aligned}
y = & -0.17082(x_{11}) + 0.19531(x_{12}) \\
& -0.14554(x_{21}) + 0.29108(x_{22}) \\
& + 0.05164(x_{31}) - 0.20657(x_{32})
\end{aligned}$$

$$R^2 = 0.621$$

주 : $x_{11 \sim 12}$: 교통량, $x_{21 \sim 22}$: 횡단경사, $x_{31 \sim 32}$: 기상상태

다중선형회귀모형과 다중비선형회귀모형에서의 사고요인을 대상으로 수량화 I 류 이론을 이용한 모형을 구축한 결과 다중선형회귀모형 및 다중비선형회귀모형보다 그다지 높은 설명력을 보여주고 있지는 않지만 다른 사고유형에 비하여 어느정도 정면충돌 교통사고를 잘 설명하고 있는 것으로 분석되었고, 또한 모형의 설명력을 나타내는 결정계수 (R^2) 값이 62.1%로 높게 분석되었다.

3) 직각충돌에 대한 교통사고예측모형

(1) 직각충돌에 대한 다중선형회귀모형

표 16. 직각충돌에 대한 ANOVA 분석표

	오차제곱합	자유도	평균제곱	분산비	유의도
선형회귀	305.663	4	76.416	5.564	0.001
잔차	920.115	67	13.733		
합계	1,225.778	71			

표 17. 직각충돌에 대한 선형다중회귀모형 결과

속성	계수(B)	표준오차	베타(β)	t 값	유의도
상수	33.513	9.662		3.468	0.001
교통량	0.101	0.095	0.115	1.064	0.291
횡단경사	-33.652	9.317	- 0.386	- 3.612	0.001
도로폭	-0.901	0.501	- 0.194	- 1.800	0.076
운전자성별	-3.758	1.417	- 0.286	- 2.651	0.010
설명력	$R^2 = 0.249$ ($Adjusted R^2 = 0.205$)				

직각충돌에 대한 사고요인을 독립변수로 사용하여 다중선형회귀모형을 구축한 결과 결정계수 (R^2)값은 24.9%, 수정결정계수의 값은 20.5%로 나타났다. 분산분석의 검정통계량 F의 값은 유의수준 5% 안에 들어 양호한 것으로 검정되었고, 다른 사고유형에서와 마찬가지로 횡단경사가 직각충돌에 가장 큰 영향을 미치는 요인으로 나타났다.

(2) 직각충돌에 대한 다중비선형회귀모형

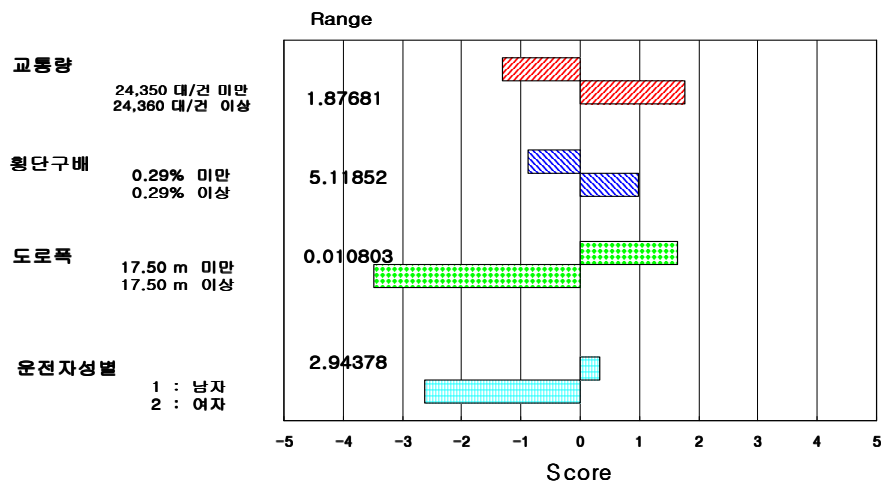
$$\hat{y} = e^{-40.80938 \times (\text{교통량})^{1.02910} \times (\text{횡단경사})^{-37.78891} \times (\text{도로폭})^{-2.52144} \times (\text{운전자성별})^{-1.51718}}$$

$$R^2 = 0.395$$

다중선형회귀모형에서 사용한 설명변수를 대상으로 직각충돌에 대한 다중비선형회귀모형을 지수식으로 나타낸 결과 결정계수 (R^2)는 39.5%로 그다지 높은 설명력을 보여주고 있지는 않았다. 하지만, 다중선형회귀모형과 비교해볼 때 지수식으로 표현한 다중비선형회귀모형의 설명력은 더 높게 분석되었다. 여기서, 교통량이 많을수록 교통사고가 많이 일어나고, 횡단경사와 도로폭이 작을수록 교통사고가 많이 일어나며, 운전자가 남성일수

록 사고가 많이 일어난 것으로 나타났다.

(3) 직각충돌에 대한 수량화 I 류 모형



CORRELATION COEFFICIENT 0.64847

ESTIMATION ERROR 3.14094

그림 13. 직각충돌에 대한 수량화 I 류 모형

위 수량화 I 류 모형을 식으로 나타내면 다음과 같다.

$$\begin{aligned}\hat{y} = & -0.886272(x_{11}) + 0.990540(x_{12}) \\ & + 1.63058(x_{21}) - 3.48344(x_{22}) \\ & - 0.36010E^+02(x_{31}) + 0.72020E^+02(x_{32}) \\ & + 0.327086(x_{41}) - 2.61669(x_{42})\end{aligned}$$

$$R^2 = 0.421$$

주 : $x_{11 \sim 12}$: 교통량, $x_{21 \sim 22}$: 횡단경사, $x_{31 \sim 32}$: 도로폭, $x_{41 \sim 42}$: 운전자성별

직각충돌에 대한 다중선형회귀모형과 다중비선형회귀모형에서 사용한 설명변수를 수량화 I 류 이론을 이용한 모형을 구축하는데 적용한 결과 다중회귀모형(선형회귀, 비선형회귀)의 결정계수 (R^2) 값보다 높게 분석되어서 수량화 I 류 이론을 이용한 모형의 설명력이 높게 나타났다. ($R^2 = 0.421$)

4) 측면접촉에 대한 교통사고예측모형

(1) 측면접촉에 대한 다중선형회귀모형

표 18. 측면접촉에 대한 ANOVA 분석표

	오차제곱합	자유도	평균제곱	분산비	유의도
선형회귀	262.589	4	65.647	3.647	0.011
간차	846.084	47	18.002		
합계	1,108.673	51			

표 19. 측면접촉에 대한 선형다중회귀모형 결과

속성	계수(B)	표준오차	베타(β)	t 값	유의도
상수	9.984	4.652		2.146	0.037
교통량	0.357	0.139	0.332	2.564	0.014
종단경사	0.793	0.485	0.223	1.634	0.109
횡단경사	-29.531	10.381	-0.391	-2.845	0.007
표면상태	-1.897	1.117	-0.218	-1.699	0.096
설명력	$R^2 = 0.237$ ($Adjusted R^2 = 0.172$)				

측면접촉에 대한 사고요인을 독립변수로 사용한 다중선형회귀모형에서 결정계수 (R^2)는 23.7%로, 수정결정계수의 값은 17.2%로 나타났다. 분산분석의 검정통계량 F의 값은 유의수준 5% 안에 들어 양호한 것으로 검정되었다. 또한, 모형의 각각의 설명변수에 대해서 교통량과 횡단경사의 요인이 측면접촉에 대한 교통사고에 가장 많은 영향을 미치고 있음을 알 수 있다.

(2) 측면접촉에 대한 다중비선형회귀모형

$$\hat{y} = e^{-3.25259} \times (\text{교통량})^{1.15093} \times (\text{종단경사})^{0.05437} \\ \times (\text{횡단경사})^{-1.42573} \times (\text{표면상태})^{-0.59284}$$

$$R^2 = 0.225$$

측면접촉에 대한 다중선형회귀모형에서 사용한 설명변수를 대상으로 다중비선형회귀모형을 지수식으로 나타낸 결과 결정계수 (R^2)는 22.5%로 낮은 설명력을 보여주고 있다. 또한 다중선형회귀모형보다도 더 낮은 설명력을 보여주고 있어서 모형으로써 좋은 값을 나타내주고 있지 않다. 여기서, 위 식을 살펴보면, 정면충돌의 다중비선형회귀식과 마찬가지로 교통량이 많을수록 교통사고가 많이 일어나고, 종단경사가 클수록, 횡단경사 작을수록 측면접촉 사고가 많이 일어나고 있음을 알 수 있다.

(3) 측면접촉에 대한 수량화 I 류 모형

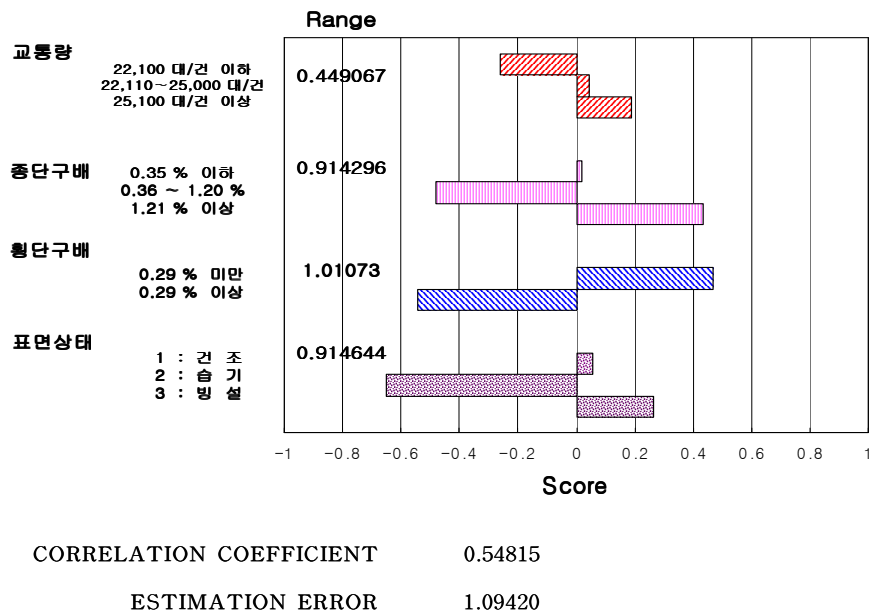


그림 14. 측면접촉에 대한 수량화 I 류 모형

위 모형을 식으로 나타내면 다음과 같다.

$$\begin{aligned}\hat{y} = & -0.259950(x_{11}) + 0.424540E01(x_{12}) + 0.189117(x_{13}) \\ & + 0.178250E01(x_{21}) - 0.478867(x_{22}) + 0.435429(x_{23}) \\ & + 0.466490(x_{31}) + 4.08741(x_{32}) \\ & + 0.559878E01(x_{41}) - 0.544238(x_{42})\end{aligned}$$

$$R^2 = 0.300$$

주 : $x_{11} \sim 13$: 교통량, $x_{21} \sim 23$: 종단경사, $x_{31} \sim 32$: 횡단경사, $x_{41} \sim 42$: 표면상태

측면접촉에 대한 수량화 I 류 이론을 이용한 모형을 구축한 결과 모형으로써 적합한지의 여부를 판단할 수 있는 계수인 결정계수 (R^2)값이 30.0%로 다중회귀모형(선형회귀, 비선형회귀)의 결정계수 (R^2) 값보다 높게 분석되어서 수량화 I 류 이론을 이용한 모형의 설명력이 높게 나타났다. 또한, 모형의 각각의 설명변수에 대해서 횡단구배가 종속변수에 가장 많은 영향을 미치고 있음을 알 수 있었다.

제 2절 다중선형회귀모형(Multiple Linear Regression Model)

교통량과 기하구조(종단경사, 곡선반경, 횡단경사, 도로폭)가 교통사고에 영향을 미치는 정도를 알아보기 위해 다중선형회귀이론을 이용해 모형을 구축하였다.

여기서, 모델의 적합도 검정(Goodness of Fit)으로 결정계수 (R^2)값과 분산분석의 검정통계량 F값을 사용하였다. 결정계수 값은 모형을 설명하는 설명변수가 많으면 많을수록 그 값이 증가하기 때문에 이러한 경향을 보완하기 위한 방법으로 수정결정계수(Adjusted R^2)값을 사용해 모형의 검정에 사용하였다.

또한 설명변수의 모형에 대한 설명력을 판단하기 위해 t 통계값을 이용하였다.

표 20. '97년 교차로 ANOVA 분석표

	오차제곱합	자유도	평균제곱	분산비	유의도
선형회귀	2357.551	5	471.510	11.103	0.000
잔차	7516.328	177	42.465		
합계	9873.880	182			

표 21. '97년 교차로 교통사고 다중선형회귀분석결과

속성	계수(B)	표준오차	베타(β)	t 값	유의도
상수	51.095	9.548		5.351	0.000
교통량	0.138	0.109	0.085	1.272	0.205
종단경사	1.353	0.402	0.228	3.367	0.001
곡선반경	-0.133	0.087	-0.105	-1.522	0.130
횡단경사	-48.931	80813	-0.376	-5.552	0.000
도로폭	-1.590	0.501	-0.217	-3.173	0.002
설명력	$R^2 = 0.239$ (<i>Adjusted</i> $R^2 = 0.217$)				

표 22. '98년 교차로 교통사고 ANOVA 분석표

	오차제곱합	자유도	평균제곱	분산비	유의도
선형회귀	524.920	5	104.984	6.624	0.000
잔차	1727.567	109	15.849		
합계	2252.487	114			

표 23. '98년 교차로 교통사고 다중선형회귀분석결과

속성	계수(B)	표준오차	베타(β)	t 값	유의도
상수	8.818	10.797		0.817	0.416
교통량	0.488	0.101	0.434	4.846	0.000
종단경사	0.656	0.349	0.177	1.877	0.063
곡선반경	-0.747	0.310	-0.216	-2.407	0.018
횡단경사	-24.929	7.541	-0.294	-3.306	0.001
도로폭	-0.253	0.577	-0.040	-0.437	0.663
설명력	$R^2 = 0.233$ ($Adjusted R^2 = 0.198$)				

표 24. '99년 교차로 교통사고 ANOVA 분석표

	오차제곱합	자유도	평균제곱	분산비	유의도
선형회귀	137.064	5	27.413	3.519	0.007
잔차	482.995	62	7.790		
합계	620.059	67			

표 25. '99년 교차로 교통사고 다중선형회귀분석결과

속성	계수(B)	표준오차	베타(β)	t 값	유의도
상수	-9.828	9.858		-0.997	0.323
교통량	0.369	0.105	0.462	3.518	0.001
종단경사	0.153	0.365	0.053	0.421	0.676
곡선반경	-0.780	0.469	-0.194	-1.664	0.101
횡단경사	-4.707	6.829	-0.084	-0.689	0.493
도로폭	0.520	0.452	0.151	-1.151	0.254
설명력	$R^2 = 0.221$ ($Adjusted R^2 = 0.158$)				

표 26. 전체 교차로 교통사고 ANOVA 분석표

	오차제곱합	자유도	평균제곱	분산비	유의도
선형회귀	2294.443	5	458.889	14.178	0.000
잔차	11651.713	360	32.366		
합계	13946.156	365			

표 27. 전체 교차로 교통사고 다중선행회귀분석결과

속성	계수(B)	표준오차	베타(β)	t 값	유의도
상수	23.996	6.492		3.696	0.000
교통량	0.393	0.067	0.289	5.826	0.000
종단경사	0.757	0.262	0.147	2.887	0.004
곡선반경	- 0.168	0.073	-0.114	-2.289	0.023
횡단경사	-33.779	5.614	-0.300	-6.017	0.000
도로폭	- 0.795	0.341	0.117	-2.332	0.020
설명력	$R^2 = 0.165$ ($Adjusted R^2 = 0.153$)				

표 20 ~ 표 27 의 분산분석과 선행회귀분석결과를 선행회귀모형식으로 나타내어 보면 표 28와 같다.

표 28. 국도 17호의 교차로 교통사고 다중선행회귀모형

구분	선행 회귀 모형식
97년 교차로	$y = 51.095 + (V) \times 0.138 + (G_r) \times 1.353$ $- (C_r) \times 0.133 - (C_g) \times 48.931 - (R_w) \times 1.590$
98년 교차로	$y = 8.818 + (V) \times 0.488 + (G_r) \times 0.656$ $- (C_r) \times 0.747 - (C_g) \times 24.929 - (R_w) \times 0.253$
99년 교차로	$y = -9.828 + (V) \times 0.369 + (G_r) \times 0.153$ $- (C_r) \times 0.780 - (C_g) \times 4.707 + (R_w) \times 0.520$
전체 교차로	$y = 23.996 + (V) \times 0.393 + (G_r) \times 0.757$ $- (C_r) \times 0.168 - (C_g) \times 33.779 - (R_w) \times 0.795$

주 : 1) V : 교통량, G_r : 종단경사, C_r : 곡선반경, C_g : 횡단경사, R_w : 도로폭

2) 단위 : 교통량(천대), 곡선반경(km), 종단경사 및 횡단경사(%), 도로폭(m)

교통사고요인에 대한 회귀분석의 결과 도출된 다중회귀식에서 설명변수로 연평균 일교통량, 종단경사, 곡선반경, 횡단경사, 도로폭이 사용되었으며, 분석하고자 하는 데이터가 주어진 가정사항을 잘 만족하고있는지를 검증하는 적합도 검정(Goodness of Fit)에는 결정계수(R^2)의 값과 분산분석의 검정통계량 F의 값을 사용하였다.

본 회귀모형의 결정계수 (R^2)는 '97년 교차로의 경우 0.239, '98년의 경우 0.233, '99년의 경우는 0.221로써 그다지 높은 설명력을 보여주고 있지는 않지만 교통사고의 회귀분석시 보여주는 기존의 교통사고에 대한 회귀분석의 결정계수 값과 비교했을 때 어느정도 설명력을 보여주고 있는 것으로 분석되었다. 그 이유는 교통사고의 주요 요인은 인적요인(Human Factor), 차량적요인(Vehicle Factor), 도로환경요인(Environmental Factor)등의 요인으로 구분할 수 있는데 이 중에서 인적요인(Human Factor)이 교통사고에 있어서 가장 큰 비중을 차지하고 있기 때문에 실질적으로 우리가 교통사고를 분석할 때 자료를 획득함에 있어서 기록에 의한 자료만 얻을 수 있기 때문이고, 또 그 자료만으로 분석할 때 모형에 대한 값은 그다지 높지 않을 것이라고 판단되기 때문에 본 논문의 선형회귀모델에서의 결정계수 (R^2)값은 그다지 낮지 않은 것이라고 생각된다.

각각의 연도에 따른 교차로 사고데이터 분산분석의 검정통계량 F의 값 또한 유의수준 5%안에 들어 양호한 것으로 검정되었다.

제 3절 다중비선형회귀모형(Multiple Non-Linear Regression Model)

다중비선형회귀모형은 다중선형회귀모형의 현실과의 괴리를 극복하고자 제한된 것이다. 교통사고는 일반적으로 비선형적인 발생특성을 나타내므로 이 접근법은 기존 방법에 비해 진일보한 방법으로 인식되고 있다. 이 모형의 범주에는 비선형 최소제곱법(Nonlinear Ordinary Least Square), 2단계 최소제곱법(Two Stage Least Square) 등의 방법이 있지만 이 방법은 선형 최소제곱추정의 구간분할을 통한 불연속 선형근사(piece-wise linear approximation)라는 한계를 지니고 있다.

본 논문은 자료의 특성에 따라 비선형적인 모형을 설정하기 위해 선형회귀모형에 사용

했던 설명변수에서 곡선반경의 변수를 빼 나머지 변수를 사용하여 비선형회귀모형을 설정하였다. 또한 각 연도에 따라 비선형회귀모형을 설정한 결과 선형회귀모형에서 구해진 결정계수(R^2)값과 그다지 차이가 나지 않았다. 각 연도에 따른 비선형회귀모형식은 다음과 같다.

표 29. 다중비선형회귀모형

구 분	결정계수(R^2)	다중비선형회귀모형
97년 데이터	0.219	$\hat{y} = e^{7.1309} \times (V)^{0.5558} \times (G_r)^{0.0907}$ $\times (C_g)^{-1.9247} \times (R_w)^{-3.0351}$
98년 데이터	0.205	$\hat{y} = e^{0.7948} \times (V)^{1.0227} \times (G_r)^{-0.1353}$ $\times (C_g)^{-1.2548} \times (R_w)^{-1.2123}$
99년 데이터	0.200	$\hat{y} = e^{-5.9914} \times (V)^{1.1765} \times (G_r)^{-0.0641}$ $\times (C_g)^{-0.1971} \times (R_w)^{1.3361}$
전체 데이터	0.153	$\hat{y} = e^{2.18141} \times (V)^{0.97447} \times (G_r)^{0.06960}$ $\times (C_g)^{-1.51066} \times (R_w)^{-1.67765}$

주) 단위 : 교통량(천대), 곡선반경(km), 종단경사 및 횡단경사(%), 도로폭(m)

V : 교통량, G_r : 종단경사, C_r : 곡선반경, C_g : 횡단경사, R_w : 도로폭

제 4절 수량화 I 류 이론을 이용한 모형

다중회귀이론을 이용하여 교통사고예측모형을 설정한 방법과 동일하게 사용했던 설명변수를 대상으로 수량화 I 류를 이용하여 교통사고예측모형을 구축하였다.

수량화 I 류를 이용하여 교통사고예측모형을 구축함에 있어서 각 요인들의 상대적 중요도, Item과 Category의 수량, Item의 범위, 중상관계수, 평균예측오차 등의 평가지수를 구할 수 있었다.

수량화 I 류 이론에 의한 데이터를 분석하기 위해 설명변수(사고요인)들이 사고 발생시 사고에 영향을 주는 상황을 동시적으로 고려하여 요인들에 대하여 각각의 경우를 조합하여 각 Item별로 Category화 한 후, 각 데이터 set을 만들었다

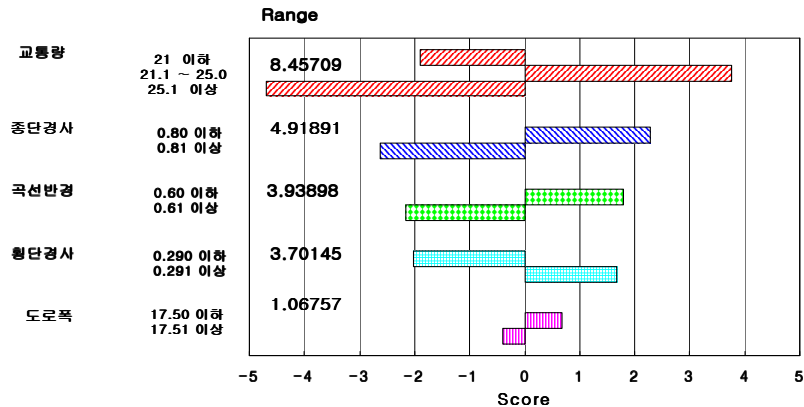


그림 15. '97년 교차로 교통사고 수량화 I 류 모형

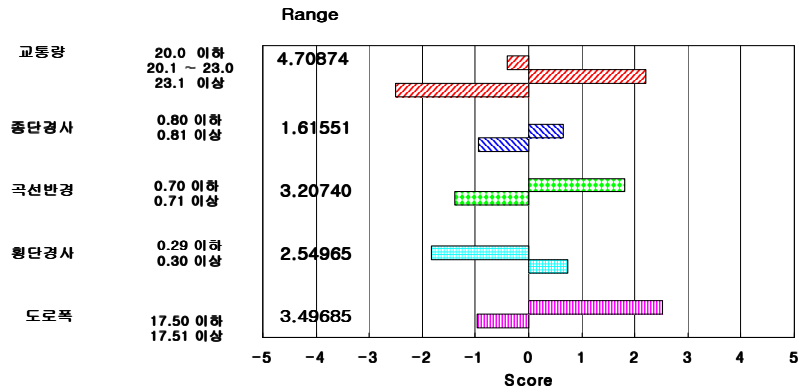


그림 16. '98년 교차로 교통사고 수량화 I 류 모형

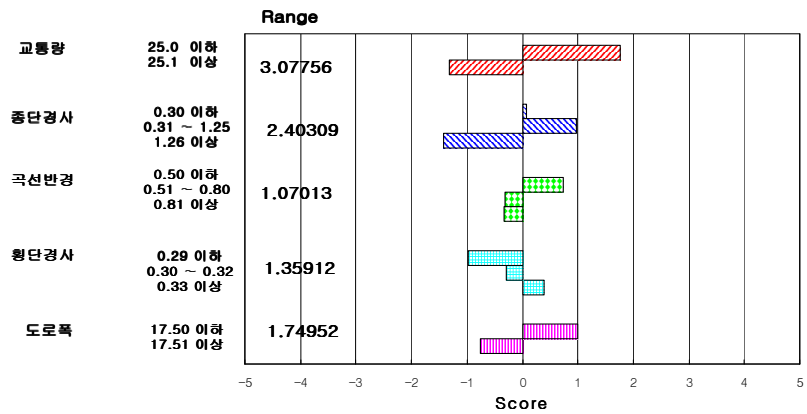


그림 17. '99년 교차로 교통사고 수량화 I 류 모형

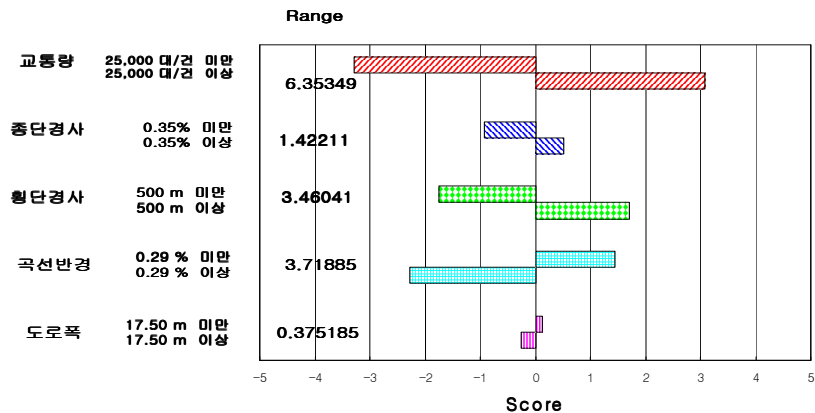


그림 18. 전체 교차로 교통사고 수량화 I 류 모형

수량화 I 류 이론에 의한 교통사고예측모형을 구축한 결과, 사고요인들이 분석된 모형을 얼마나 잘 설명하고 있는지의 여부를 판단할 수 있는 범위(Range)를 구할 수 있었다. 그런데, 주의할 점은 모형의 분석결과에서 Item의 범위(Range)의 값이 크면 클수록 요인의 중요도는 높아지게 되는 것이다.

수량화 I 류 이론에 의한 각 연도별 교차로 사고요인에 대한 Category 분석결과 교통량의 범위('97년 : 8.45709, '98년 : 4.70874, '99년 : 3.07756, 전체 : 6.35349)가 가장 큰 값을 보이고 있어 교통량이 사고요인으로 가장 큰 영향을 미치고 있음을 알 수 있다. 또한 사고건수를 외측기준으로 하여 교통사고예측모형을 구축한 결과, 모형 설명력을 나타내는 중상관계수 값은 '97년의 경우 79.90%이고, '98년의 경우 73.53%, '99년의 경우는 61.40%였으며, 전체의 경우에는 63.56%였다. 또한 추정오차(Estimation Error)은 '97년의 경우 4.41707, '98년의 경우 2.99961, '99년의 경우에는 2.38343이었고, 전체의 경우에는 4.76543으로 분석되었다. 그리고 Item의 범위는 교통량, 종단경사, 곡선반경, 횡단경사, 도로폭의 순으로 나열하였다.

다중회귀이론을 이용하여 교통사고예측모형을 분석했을 때에 사용했던 설명요인을 대상으로 수량화 I 류 이론을 이용한 모형을 구축하였는데 다중회귀이론을 이용하였을 때의 결정계수(R^2)값보다 높게 나타나고 있어 모형으로써 설명력이 높게 분석되었다.

수량화 I 류 이론을 이용하여 각각의 연도별로 교통사고를 분석해본 결과 '97년의 경우

에는 교통량이 21,000대/일~25,000대/일 때 교통사고가 가장 많이 일어나는 것으로 분석되었고, 종단경사가 0.81%이상일 때 사고가 가장 많이 일어났고, 도로폭은 17.5m 이상일 때 교통사고가 많이 일어나는 것으로 분석되었다. 또한, 99년의 경우에는 교통량이 25,000대/일 이상일 때와 종단경사가 0.30% 이상일 때 교통사고가 가장 많이 일어나는 것으로 분석되었으며 전체적으로는 교통량이 25,000대/건 이상일 때와 종단경사가 0.36% 이상일 때, 곡선반경이 0.51이상일 때, 횡단경사가 0.29% 미만일 때 그리고 도로폭이 17.5m 이하일 때 교통사고가 많이 일어나는 것으로 분석되었다. 따라서, 위 그림을 볼 때 교통량과 기하구조가 도로교통사고에 커다란 영향을 미치고 있음을 보여주고 있다.

수량화 I 류 이론을 이용하여 모형을 구축하였는데 이를 식으로 나타내어 보면 다음과 같다.

표 30. 각 연도별 교차로의 수량화 I 류 모형

구분	R^2	수량화 I 류에 의한 Model
'97년	0.64	$y = -4.69075(x_{11}) + 3.76635(x_{12}) - 1.89060(x_{13})$ $- 2.63417(x_{21}) + 2.28474(x_{22})$ $- 2.15245(x_{31}) + 1.78653(x_{32})$ $+ 1.67880(x_{41}) - 2.02265(x_{42})$ $- 0.396691(x_{51}) + 0.670875(x_{52})$
'98년	0.54	$y = -2.50375(x_{11}) + 2.20498(x_{12}) - 0.406899(x_{13})$ $- 0.955260(x_{21}) + 0.660253(x_{22})$ $- 1.39452(x_{31}) + 1.81288(x_{32})$ $+ 0.731640(x_{41}) - 1.81801(x_{42})$ $- 0.973036(x_{51}) + 2.52381(x_{52})$
'99년	0.38	$y = -1.31249(x_{11}) + 1.76507(x_{12})$ $- 1.42099(x_{21}) + 0.982100(x_{22}) + 0.0755501(x_{23})$ $- 0.340696(x_{31}) - 0.310496(x_{32}) + 0.729433(x_{33})$ $+ 0.389466(x_{41}) - 0.288817(x_{42}) - 0.969652(x_{43})$ $- 0.746117(x_{51}) + 1.00340(x_{52})$
전체	0.40	$y = -3.28090(x_{11}) + 3.07259(x_{12})$ $- 0.913107(x_{21}) + 0.509008(x_{22})$ $- 1.74912(x_{31}) + 1.71130(x_{32})$ $+ 1.45299(x_{41}) - 2.26585(x_{42})$ $+ 0.132237(x_{51}) - 0.242947(x_{52})$

주 : $x_{11} \sim 13$: 교통량, $x_{21} \sim 23$: 종단구배, $x_{31} \sim 33$: 곡선반경, $x_{41} \sim 43$: 횡단구배 $x_{51} \sim 53$: 도로 폭

제 5절 구축된 교통사고 예측모형간의 비교

표 31. 구축된 모형간의 결정계수 및 추정오차 비교

모형	97년 데이터		98년 데이터		99년 데이터		전체 데이터	
	결정계수 (R^2)	추정 표준오차	결정계수 (R^2)	추정 표준오차	결정계수 (R^2)	추정 표준오차	결정계수 (R^2)	추정 표준오차
다중선형 회귀모형	0.24	6.517	0.23	3.981	0.22	2.791	0.165	5.689
다중비선형 회귀모형	0.22	11.209	0.21	6.215	0.20	3.996	0.153	5.691
수량화 I 류 모형	0.64	4.417	0.54	3.000	0.38	2.383	0.400	4.765

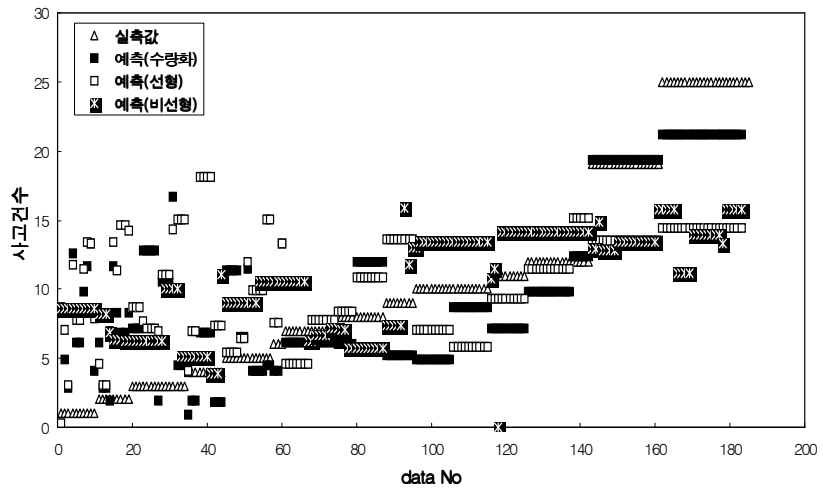


그림 19. '97년 교통사고 예측모형의 예측값과 실측값의 비교

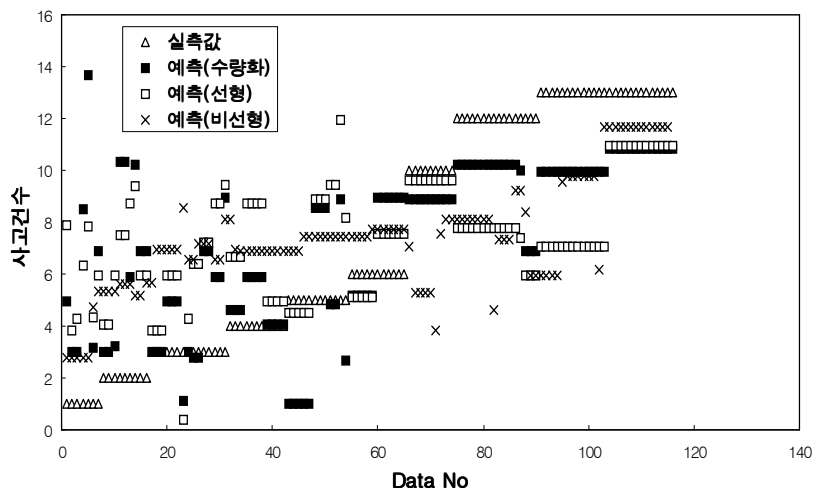


그림 20. '98년 교통사고 예측모형의 예측값과 실측값의 비교

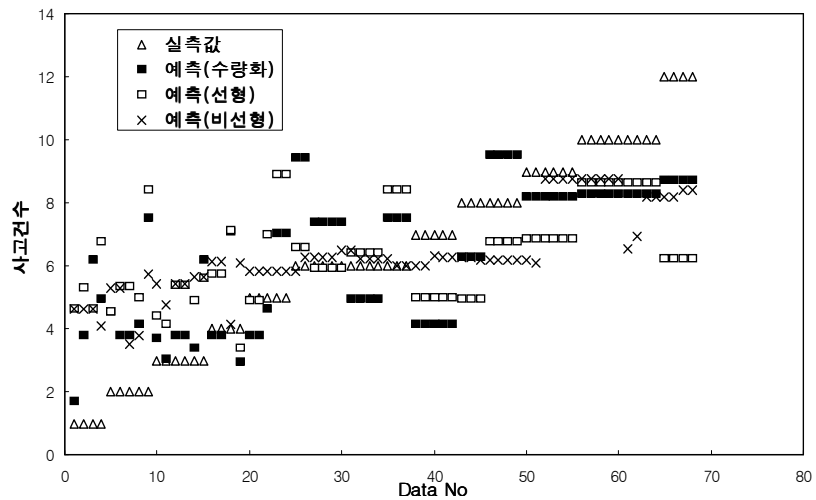


그림 21. '99년 교통사고 예측모형의 예측값과 실측값의 비교

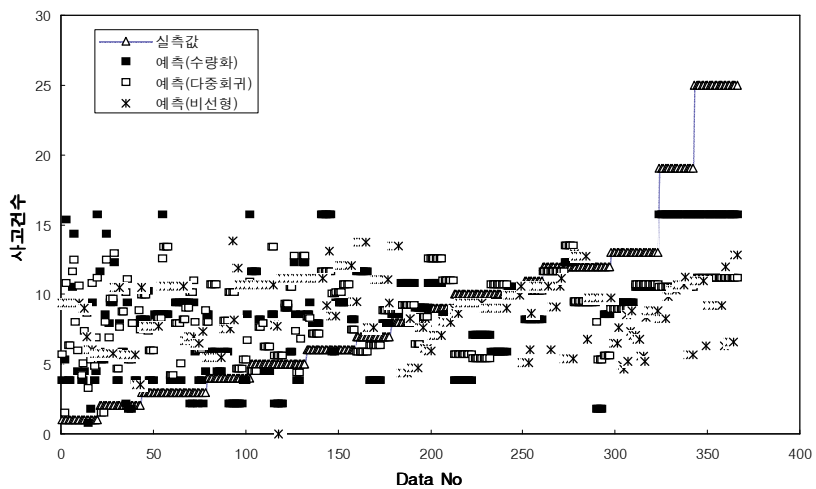


그림 22. 전체 데이터를 이용한 교통사고 예측모형의 예측값과 실측값의 비교

다중회귀이론을 이용한 모형과 수량화 I 류 이론을 적용한 모형의 설명력에 있어서 수량화 I 류 이론을 적용한 모형의 결정계수(R^2)값이 더 높게 나타났으며, 추정 표준오차도 수량화 I 류 모형이 더 작은 값을 보이고 있다. 이들 예측값을 실측값과 비교하여 plot하면 그림19~그림22과 같이 수량화 I 류 모형이 다중회귀모형보다 더 실측치에 근접하게 나타나고 있음을 알 수 있으므로 수량화 I 류 이론을 적용한 모형이 더 좋은 모형으로 분석되었다.

따라서, 여러 사고요인들을 고려한 교통사고 예측모형의 구축은 기존의 다중회귀모형보다는 수량화 I 류 이론을 이용한 방법이 더 유효한 방법이라고 생각된다.

제 5장 결 론

본 연구에서는 다중선형회귀, 다중비선형회귀 그리고 수량화 I 류를 이용한 분석을 통해 교통사고예측모형을 구축함으로써 지속적인 차량증가에 따른 교통사고를 미연에 방지하고, 보다 합리적이고 체계적으로 교통사고에 대한 분석을 시도하였다.

본 연구는, 국도 17호선(전주~남원)구간의 도로상에 있어서 교차로 교통사고의 '97, '98, '99년 각 연도별 교통사고와 이를 통합한 전체 데이터를 이용하여 교통사고를 분석하였다.

본 연구 대상구분은 각 지점 특성별로 사고유형(후미충돌, 정면충돌, 직각충돌, 측면접촉), 사고위치(접근부, 유출부, 횡단보도, 교차로내, 기타 등)로 구분해 교차분석 및 빈도 분석을 하였고, 국도 17호선(전주~남원)내에서의 교통 사고의 특성을 알아보기 위하여 월별 교통사고분포, 요일별 교통사고분포, 운전자의 사고원인 분포(신호위반, 중앙선침범, 과속, 음주운전, 안전운전 불이행, 기타), 보행자 사고원인 분포(횡단, 무단횡단, 승하차, 차로내 사고, 차로외 사고, 기타)등으로 구분하여 빈도분석을 하였다.

또한, 각각의 모형에 있어서 예측값을 구하고, 각각에 대한 예측값들을 실측값에 대하여 비교·plot함으로써 모형의 적합도를 비교할 수 있었다.

1. 단순 통계분석에 의한 교차로 교통사고의 특성

각 연도별로 교통사고의 특성을 알아보기 위해 교통사고 요인에 대한 단순분석을 실시한 결과 교차로에서의 사고는 대부분 접근부와 교차로내의 사고가 가장 많았고, 사고유형별로 분류해 볼 때 교차로에서의 사고는 후미충돌과 측면접촉사고가 대부분인데 후미충돌은 교차로의 접근부에서 대체적으로 사고가 많이 일어나고, 측면접촉이나 정면충돌 역시 교차로의 접근부와 교차로내에서 사고가 많이 발생하였다.

일반적으로 교차로에서의 후미충돌이나 측면접촉사고는 주방향으로 달리는 차량에 대해서 교차로에서 우회하려는 차량으로 인해 사고가 발생하고, 교차로 신호의 규제에 의해서 정지하려는 앞차와 뒷따라가는 차량이 앞차와의 거리를 충분히 확보하지 않은채 발생한 것이라고 사료된다.

월별 교통사고발생분포는 전체적으로 볼 때 남원과 전주 근교의 지역에 따라 특정한

행사로 인해 나들이하러는 차량의 증가로 교통사고가 많이 발생한 것이라고 생각된다. 특히 4, 5월이나 8, 9월은 바깥 외출하기에 적당한 달이기 때문에 지역을 찾는 차량들의 증가로 교통사고가 많이 발생한 것이라고 생각된다.

요일별 교통사고분포의 경우 평일의 교통사고분포 보다 휴일이나 주말의 교통사고분포가 대체로 많이 나타났다. 보행자 사고원인별 교통사고분포를 보면, 대체적으로 보행자가 횡단보도 횡단시 부주의에 의한 사고가 주된 원인으로 분석되었다. 이는 본 연구의 대상 지역이 국도이면서 교차하는 지역이기 때문에 자동차들의 정지거리가 길어짐에 따라 사고의 위험도가 증가하고, 또 대상지점을 통과하는 차량 속도가 높은 차량들이기 때문에 그만큼 신호에 의한 정지규제를 받을 때 정지시거가 부족하기 때문에 사고가 증가하는 것이라고 생각된다.

운전자 사고원인별 분포의 경우 안전운전불이행의 원인이 주된 원인으로 분석되었다.

2. 상관분석에 의한 교차로 교통사고 요인분석

본 연구는 기존의 연구 자료를 감안하여 교통사고 서식에 나와있는 자료를 대상으로 국도상의 교차로 교통사고에 영향을 주는 요인의 상관관계를 분석하여 교통사고에 영향이 있는 변수를 선정하였다.

각 연도에 따른 교차로 교통사고 요인들의 상관관계를 분석한 결과 '97년의 경우 종단구배, 횡단구배, 도로폭이 교통사고에 영향이 많게 분석되었고, '98년의 경우는 교통량, 곡선반경, 횡단구배 이었으며, '99년의 경우 교통량, 곡선반경으로 교통사고건수에 영향을 미치는 요인들을 알 수가 있었다. 또 전체데이터에 있어서 교차로 교통사고에 영향을 미치는 요인으로 교통량, 횡단구배, 도로폭 등으로 나타났다.

3. 교차로 사고유형에 따른 교통사고예측모형

교차로 사고유형을 후미충돌, 정면충돌, 직각충돌, 측면접촉으로 나누어 각 유형에 대해 다중선형 및 비선형회귀모형과 수량화 I 류 모형을 구축하여 적합한 모형들을 얻었으며, 모든 사고유형에 있어서 횡단경사의 요인이 교통사고에 가장 큰 영향을 미침을 알 수 있었다.

특히, 정면충돌의 경우 다른 사고유형과 달리 횡단경사가 클수록 교통사고가 많이 발

생하고 있는데 이는 정면충돌사고의 경우 곡선커브를 갖는 교차로에서 많이 발생하고 있음을 알 수 있어 본 연구대상 지역의 선형개량이 정면충돌 사고 발생억제 대안의 하나라고 생각된다.

4. 교차로 기하구조 요인을 고려한 교통사고예측모형

본 논문은 교통량과 기하구조(중단구배, 곡선반경, 횡단구배, 도로폭)가 교통사고에 영향을 미치는 정도를 알아보기 위해 다중선형회귀이론과 다중비선형회귀이론, 수량화 I 류 이론을 이용한 교통사고예측모형을 구축하였다.

구축된 세 모형을 비교해 본 결과 다중회귀이론을 이용한 모형보다 수량화 I 류 이론을 이용한 모형의 결정계수(R^2)값이 더 높게 나타났고, 이들의 실측값들에 대한 예측값들을 비교하여 Plot해본 결과 수량화 I 류 모형의 예측값이 실측값에 근접하게 나타나고 있음을 알 수 있다. 또한 추정표준오차도 수량화 I 류 이론을 이용한 모형이 더 작은 값을 보이고있어 수량화 I 류 이론을 이용한 모형이 더 유효한 분석방법이라고 생각된다.

위의 결과를 종합해 볼 때 교통사고의 원인으로서 교통량과 기하구조가 교통사고에 많은 영향을 미치고 있음을 알 수 있었고, 또 수량화 I 류 이론을 이용하여 모형을 구축함으로써 변수들의 Category화된 항목들이 외측기준에 얼마나 많은 영향을 미치는가를 구체적으로 파악할 수 있었다.

참고 문헌

- [1] 강경우(1997), 「공간자기회귀모형을 이용한 고속도로 교통사고 분석」, 『대한교통학회지』, 제15권, 제1호.
- [2] 건설교통부(1997, 1998, 1999), 『도로 교통량 통계연보』.
- [3] 건설교통부(1999), 『국도 17호선 전주~남원간 국도의 교통안전 개선방안 연구 최종 보고서』.
- [4] 김갑수, 신관식(1998), 「도시고속도로 기본구간의 교통특성분석(대구신천대로를 대상으로)」, 『대한교통학회지』, 제16권, 제4호. pp.99~112.
- [5] 김원철(2000), 「교차로 안전진단수법에 관한 연구」, 『원광대학교석사학위논문』.
- [6] 김홍상(1987), 「교통사고의 기술에 방법에 관한 연구」, 『서독 칼스루에 공과대학 박사학위논문』.
- [7] 김효종(1997) 「교통사고 유형에 미치는 영향요인 분석에 관한 연구」, 『전남대학교박사학위논문』.
- [8] 남궁현(2001), 「신호 교차로 교통사고 예측모형의 개발 및 적용(광주광역시 4-지 신호 교차로를 중심으로)」, 『전남대학교석사학위논문』.
- [9] 박병호(1995), 「충청권의 교통사고 예측모형 개발에 관한 연구」, 『대한교통학회지』, 제13권, 제1호, pp.63~82.
- [10] 박종철(1998), 「교차로 교통사고의 요인과 특성분석」, 『전남대학교석사학위논문』.
- [11] 서채연, 김효종(1995), 「교차 교통량 특성이 교통사고에 미치는 영향에 관한 연구」, 『대한 국토·도시계획학회지』, 제13권, 제2호(통권 76호), pp.255~266.
- [12] 오 철, 장재남, 장명순(1999), 「고속도로 시설물 구간의 교통혼잡도와 사고와의 관계분석(신갈~안산 고속도를 중심으로)」, 『대한교통학회지』, 제17권 제2호, pp.21~27.
- [13] 원제무(1997), 『도시교통론』, 박영사.
- [14] 원제무, 최재성(1997), 『교통공학』, 박영사.
- [15] 원제무(1993), 『알기쉬운 도시교통』, 세진사.

- [16] 이순철(2000), 『교통심리학』, 학지사.
- [17] 이용택, 김양지, 김대현, 임강원(2000), 「고속도로 교통사고의 계절성 검증과 요인분석」, 『대한교통학회지』, 제18권, 제5호, pp.7~16.
- [18] 이점호, 이동민, 최재성(2000), 「평면곡선부의 속도 및 교통사고 영향분석연구」, 『대한 교통학회지』, 제18권, 제1호, pp.35~43.
- [19] 이찬노, 최재성, 하태준(1998), 「신경망을 통한 도시간선도로의 교통사고 요인분석연구」, 『대한토목학회논문집』, 제18권, 제Ⅲ-6호, pp.803~810.
- [20] 임현연, 이일병(1992), 「부산시 교통사고예측모형의 개발」, 『대한교통학회지』, 제10권, 제3호. pp. 103~121.
- [21] 장수은, 김대현, 임강원(2000), 「신경망 이론을 이용한 통행발생 모형연구(선형/비선형회귀모델과의 비교)」, 『대한교통학회지』 제18권, 제4호, pp.95~105.
- [22] 전라북도 지방경찰청, 도로교통안전관리공단 전라북도지부(2000), 『전라북도 교통사고 잦은 곳 기본개선계획』.
- [23] 정충영, 최이규(1996), 『SPSS WIN을 이용한 통계분석』, 무역경영사.
- [24] 조인호(2001), 『SAS강좌와 통계컨설팅[Ⅰ][Ⅱ]』, 제일경제연구소.
- [25] Bonneson, J. A. and McCoy, P. T.(1993), *Estimation of safety at two-way stop-controlled intersections on rural highways*, Transportation Research Record, 1401 Transportation Research Board, National Research Council, Washington D.C., pp.83~89
- [26] Br de, U. and Larsson, J(1988), *The use of prediction models for eliminating effects due to regression-to-the mean in road accident data*, Accident Analysis and Prevention, Vol.20, No.4, pp.299-310.
- [27] K. M. Bauer and D. W. Harwood(1996), *Statistical Models of At-Grade Intersection Accidents*, Report No. FHWA-RD-96-125, Federal Highway Administration, Washington, D. C.
- [28] Miaou, S. and Lum. H, *Modeling vehicle accident and highway geometric design relationship*, Accidents Analysis and Prevention, Vol. 25, No 6, pp.689~709.

- [29] William R. Dillon, Matthew Goldstein(1984), *Multivariate Analysis Method And Applications*, Research Foundation The City University of New York.
- [30] G. A. F. Seber(1997), *Linear Regression Analysis*, John Willey & Sons, Inc.
- [31] 駒澤勉(1990), 『數量化理論とデータ處理』, 朝倉書店.
- [32] 有馬 哲, 石村貞夫(1987), 『多變量解析のはなし』, 東京図書株式会社.
- [33] 芳賀敏郎, 橋本茂司(1984), 『回歸分析と主成分分析』, 日科技連出版社
- [34] 山田晴利外 3人(1994), 「全國觀光交通實態調査を用いた交通發生量モデルの提案」, 『교통공학』, Vol.29, No.2, pp.18~26
- [35] 中川茂外 3人(1990), 「景觀評價システムに關する研究」, 『土木計劃學研究・講演 集』, No13, pp.53~56.

기하구조를 고려한 교차로 교통사고 예측모형

서남대학교 대학원
토목공학과 토목공학 전공

이 주 현

지도교수 : 성 수 련

< 국문초록 >

일반적으로 도로상에서 발생하는 교통사고는 인적요인(Human Factors), 차량적 요인(Vehicle Factors), 도로환경적요인(Environmental Factors)등이 복합적으로 발생한다.

본 연구는 다중선행회귀와 다중비선행회귀, 수량화 I 류를 이용한 분석을 통해 교통사고예측모형을 구축함으로써 지속적인 차량증가에 따른 교통사고를 미연에 방지하고, 보다 합리적이고 체계적으로 교통사고를 분석하였다.

먼저, 국도17호선(전주~남원)구간의 도로상에 있어서 97년, 98년, 99년 교차로 교통사고 분포를 나타내었고, 연도별 교통사고 특성을 알아보기 위해 교통사고 요인에 대한 단순분석을 실시하였다. 그 결과, 교차로에서의 교통사고는 대부분 접근부와 교차로내에서 많이 발생했고, 사고유형으로는 후미충돌사고가 가장 많았다.

본 연구에서는 대상구간의 교차로 교통사고에 영향을 주는 요인으로 교통량, 횡단경사, 도로폭 등으로 나타났고, 사고유형에 대하여 다중선행회귀모형, 다중비선행회귀모형, 수량화 I 류 모형을 각각 구축하였다. 모형을 구축한 결과 모든 사고유형에 있어서 횡단경사의 요인이 교차로 교통사고에 가장 큰 영향을 미침을 알 수 있었다.

본 논문은 교통량과 기하구조가 교차로 교통사고에 영향을 미치는 정도를 알아보기 위하여 다중선행회귀이론, 다중비선행회귀이론, 수량화이론 제 I 류를 이용한 교통사고예측

모형을 구축하였다. 구축된 세 모형을 비교해본 결과 수량화 I 류 이론을 이용한 모형의 결정계수 (R^2)값이 가장 높게 나타났고, 모형으로부터 구한 예측값과 실측값들을 비교·Plot 해본 결과 수량화 I 류 모형의 예측값이 실측값에 근접하게 나타나고 있음을 알 수 있었다. 때문에 수량화 I 류 이론을 이용한 모형이 더 유효한 분석방법임을 확인할 수 있었다.

감사의 글

이 논문을 완성하기까지 무엇보다도 지난 학부 4년과 대학원 2년 동안에 많은 격려와 뒤에서 뒷바라지를 해주신 부모님께 먼저 머리 숙여 감사의 말씀을 전합니다. 그리고 학교 생활을 열심히 할 수 있도록 여러모로 지켜봐 주시고 많은 지도와 격려를 해 주신 서남대학교 토목공학과 김 홍룡 교수님, 박 상우 교수님, 강 권수 교수님, 한 재익 교수님, 문 병석 교수님, 이 광준 교수님께도 진심으로 감사의 말씀을 올립니다.

이 논문을 쓰기까지 저에게 가장 관심을 가져주시고 아껴주시면서 정신적으로 많은 가르침을 주셨던 저의 지도교수님이신 성 수련 교수님께 머리숙여 감사의 말씀을 올립니다.

어떻게 보면 그렇게 긴 시간은 아니었지만 이것저것 대학원생활에 있어서 많은 도움을 주시고, 논문을 쓰기까지 물심양면으로 많은 가르침과 지도를 해주신 원광대학교 남궁문 교수님께도 감사의 말씀을 전합니다. 또한, 대학원 생활을 하면서 저의 많은 디딤돌이 되주신 원광대학교 이 병주 박사과정, 이 정택, 손 영국, 김 정일 학형, 김 장욱, 박 상명, 박 민수 학우들에게도 고마운 마음을 전합니다. 또 논문자료를 수집하는데 도움을 주신 도로교통안전관리공단 유 창남 선생님, 멀리서나마 논문을 완성하는데 많은 조언을 해준 김 원철, 서 승환님, 그리고, 대학원 생활을 같이 해온 서 영준, 정 부현 학우, 김 상한, 이 신재 학형, 김 영대 소장님께도 감사의 말씀을 전합니다.

끝으로, 사회에 나가서도 저에게 관심을 가져주신 모든 분들을 위해서 모든 일에 최선을 다하는 그런 사람이 되도록 열심히 노력하겠습니다.