

## 도로위의 기상요인이 교통사고에 미치는 영향 - 부산지역을 중심으로 -

이경준<sup>1</sup> · 정임국<sup>2</sup> · 노윤환<sup>3</sup> · 윤상경<sup>4</sup> · 조영석<sup>5</sup>

<sup>12345</sup>부산대학교 통계학과

접수 2015년 3월 26일, 수정 2015년 4월 6일, 게재확정 2015년 4월 27일

### 요 약

교통사고는 인구의 증가와 그에 따른 자동차의 증가로 인하여 매년 증가하고 있다. 그러한 교통사고의 원인은 운전자의 부주의뿐만 아니라 도로상의 기상상황에 의해 영향을 받는다. 특히, 강수량, 시계, 습도, 흐림 정도, 기온 등에 의해 많은 교통사고들이 영향을 받는다. 따라서 본 연구는 다양한 기상요인의 영향 정도에 따른 교통사고 발생 유무의 분석을 목적으로 하였다. 부산 해운대구의 센텀남대로 및 해운대로의 2013년도 교통사고 발생 자료와 지역별 상세 기상 관측 자료인 AWS 기상자료(시간당 강수량, 강수유무, 기온, 풍속), 시간대, 요일을 활용하여 로지스틱 회귀모형 및 의사결정나무모형을 이용하여 분석하였다. 그 결과 기상요인 중 강수유무와 기온이 교통사고 발생에 영향을 미치는 요인으로 나타났다. 이러한 결과는 도로위의 기상상태에 따른 교통사고의 발생을 예측하는데 유용하게 사용할 수 있을 것이다.

주요용어: 교통사고, 기상요인, 로지스틱 회귀분석, 의사결정나무분석, 지역별 상세 관측 자료, 해운대구.

### 1. 서론

현대 산업의 발달에 따라 인간의 삶은 윤택해지고 이에 따라 점차 편안함을 추구하기 시작하였다. 따라서 자연스럽게 자동차는 기하급수적으로 증가하여 90년대 후반에는 천만대 이상의 차량이 등록되었고 오늘날에도 점점 증가하고 있다. 이러한 차량의 증가와 맞물려 자동차에 의한 사고가 빈번해지고 있고, 그 피해도 증가하고 있는 실정이다. 자동차 사고는 인간에게 물질 재산뿐만 아니라 생명에도 영향을 미치고 있어 이에 대한 관심은 매우 중요하다. 따라서 우리나라 경찰청에서는 매년 증가되고 있는 교통사고의 발생을 줄이기 위해 교통사고통계 책자를 발간하여 여러 요소별 통계자료를 제공하고 있다.

자동차 사고는 졸음운전, 과속, 음주운전과 같은 운전자의 운전습관 또는 부주의에 의해서 발생하기도 하고, 최근 발생하고 있는 포트홀과 같은 도로의 열악한 사정으로 인해서도 발생하기도 한다. 그리고 운전자의 습관, 도로의 사정뿐만 아니라 그날그날의 도로상의 날씨 상태, 즉 기상 상태에 따라서 교통사고는 많은 차이를 나타내고 있다.

<sup>1</sup> (609-735) 부산광역시 금정구 부산대학교로63번길 2, 부산대학교 통계학과, 박사후연구원.

<sup>2</sup> (609-735) 부산광역시 금정구 부산대학교로63번길 2, 부산대학교 통계학과, 석사과정.

<sup>3</sup> (609-735) 부산광역시 금정구 부산대학교로63번길 2, 부산대학교 통계학과, 박사과정.

<sup>4</sup> (609-735) 부산광역시 금정구 부산대학교로63번길 2, 부산대학교 통계학과, 박사과정.

<sup>5</sup> 교신저자: (609-735) 부산광역시 금정구 부산대학교로63번길 2, 부산대학교 통계학과, 교수.

E-mail: choys@pusan.ac.kr

Lynn과 Barbara (1978)는 미국 세인트루이스에서 교통사고를 시간별, 요일별, 그리고 강수량을 크기별로 분류하여 교통사고 발생률 증가와의 상관성을 조사하였다. 그 결과, 피해의 심각성과 요소별 크기와는 뚜렷한 상관성이 없었으나, 16~21시, 금, 토요일에, 강수가 많을 때 사고 발생률도 증가한 것으로 나타나 강수변화와 교통사고율과는 높은 상관성을 가지고 있다고 보고하고 있다. 그리고 Ibrahim과 Hall (1994) 역시 강수가 적을 (light rain)때는 교통흐름과 속도에는 미미한 감소가 있으나, 강수가 많을 (heavy rain)때에는 상당한 감소가 있다고 분석하고 있다.

Oh 등 (2002)은 비가 오는 경우 주간과 야간의 속도 감소 차이가 크게 나타나고, 주간이 빗길 사고와 관련성이 높다고 분석하고 있다. 또한 Jeong (2011)은 강우 시 고속도로의 교통사고 특성에 관한 연구를 하였다. Jeong (2011)은 강우 상황을 분류하였고, 이슬비가 개별적인 빗길사고와 관련성이 높다고 판단하였으나 사고빈도가 높은 구간은 계절적 요인인 폭우가 위험요인으로 작용한다고 보고하고 있다.

하지만, 이들은 기상요소 중 하나만의 요소를 다루고 있고, 여러 가지 기상요소를 종합적으로 사용한 연구는 거의 없는 실정이다. 그리고 교통사고 발생률에 대한 연구만 진행되고 있을 뿐, 직접적인 교통사고 발생 예측을 하는 연구는 많이 진행되고 있지 않다. 또한 부산지역은 해안지역에 위치하고 있기 때문에 기상변화가 잦을 뿐 아니라 교통량이 많은 지역이라 교통사고 예측이 매우 필요한 지역이라고 할 수 있다.

따라서 본 연구에서는 부산지역, 특히 해안과 인접해 있는 해운대구의 교통사고 발생자료와 지역별 상세 기상 관측 자료 (AWS 기상자료) 및 시간대 및 요일을 이용하여 부산지역 교통사고가 기상상태 및 시간대에 따라 어떻게 변화하는지 알아보고, 교통사고와 기상요소의 통계적인 관계를 이용하여 기상 예보를 이용하여 교통사고 발생을 예측해 볼 수 있는 모형을 산출하고자 한다.

## 2. 자료 및 방법

### 2.1. 조사기간 및 조사자료

본 연구에서는 기상 상태와 교통사고 발생과의 관계를 도출하기 위하여 기상 상태는 기상청에서 제공하는 AWS 기상자료를 이용하였다. AWS 기상자료는 매 분별로 기상자료를 제공하기 때문에 그 양이 방대하고 그에 대응되는 교통사고 발생 건수는 매우 작기 때문에, AWS 기상자료의 1시간 평균 자료를 이용하였다. 그리고 분석대상구간은 2013년도 부산 해운대 지역을 대상으로 하였고 시간대와 요일을 고려하였고, 부산지방경찰청에서 제공한 시간당 교통사고 일부자료를 이용하였다. 교통사고 유무별 요일, 시간대, 기상자료의 기초통계량은 다음 Table 2.1과 Table 2.2와 같다.

Table 2.1 Summary of day of the week and times

		Accident				Total	
		No		Yes		N	%
		N	%	N	%		
Day of the week	Mon	1083	14.2	165	14.5	1248	14.2
	Tue	1109	14.6	163	14.3	1272	14.5
	Wed	1110	14.6	138	12.1	1248	14.2
	Thu	1078	14.1	170	14.9	1248	14.2
	Fri	1086	14.3	162	14.2	1248	14.2
	Sat	1045	13.7	203	17.8	1248	14.2
	Sun	1109	14.6	139	12.2	1248	14.2
Times	0~3	1000	13.1	95	8.3	1095	12.5
	3~6	1029	13.5	66	5.8	1095	12.5
	6~9	975	12.8	120	10.5	1095	12.5
	9~12	963	12.6	132	11.6	1095	12.5
	12~15	936	12.3	159	13.9	1095	12.5
	15~18	926	12.2	169	14.8	1095	12.5
	18~21	885	11.6	210	18.4	1095	12.5
	21~24	906	11.9	189	16.6	1095	12.5
Existence of rainfall	No	7320	96.1	1075	94.3	8395	95.8
	Yes	300	3.9	65	5.7	365	4.2
Total		7620	100.0	1140	100.0	8760	100.0

**Table 2.2** Descriptive statistics of AWS

Accident	AWS	min	Max	mean	S.D.
No (N=7620)	12-hourly precipitation	0.0	33.0	0.120	1.021
	temp	-10.4	33.2	14.572	9.030
	wind speed	0.0	10.0	2.357	1.609
Yes (N=1140)	12-hourly precipitation	0.0	18.0	0.190	1.160
	temp	-9.9	32.5	15.953	8.686
	wind speed	0.0	11.8	2.432	1.632
Total (N=8760)	12-hourly precipitation	0.0	33.0	0.130	1.041
	temp	-10.4	33.2	14.751	8.998
	wind speed	0.0	11.8	2.366	1.612

## 2.2. 분석방법

관측된 자료를 사용하는 관측연구는 각 그룹에 확률적 배분이 불가능하여 반응변수에 영향을 미치는 공변량이 불균형 자료일 때 반응변수의 차이 추정에 편향을 발생하는 선택편향 (selection bias) 의 문제가 발생하는 단점이 존재한다. 그래서 집단 간의 공변량의 균형을 시키고, 선택편향을 제거하기 위하여, 가장 비슷한 공변량의 특징을 갖고 있는 개체로 짝을 만들거나 비슷한 특성을 갖고 있는 집단으로 층화시킨 후 집단 간의 처리효과를 추정하는 방법을 고려하게 된다. 이를 위해 Rosenbaum과 Bubin (1983)은 관측연구에서 그룹 간의 확률적 배분을 위하여 성향점수 (propensity score)를 사용할 것을 제안하였다. 여기서 성향점수는 관측된 공변량에 대해 처리 그룹에 배치될 확률로 정의하는데, Rosenbaum과 Bubin (1983)은 성향점수에 의한 매칭 (matching)과 층화는 공변량의 불균형에 의해 발생하는 편향을 줄인다는 것을 보였다. 이러한 점을 볼 때, 성향점수를 이용한 방법은 실험이 불가능하거나 막대한 자본이 소요될 수 있는 분야에서 활용한다면 자료의 질적 향상에 도움이 되리라 생각할 수 있다.

성향점수를 추정하는 방법은 판별분석이나 로지스틱 회귀분석을 이용하여 추정할 수 있다. 두 방법 모두 공변량들이 주어진 조건 하에서 처리할당의 확률의 추정치를 구할 수 있다. 하지만 판별분석은 공변량들이 다변량 정규분포를 따른다는 것을 가정해야 하는 반면, 로지스틱 회귀분석은 그러한 가정에 비교적 자유롭고, 판별분석에 의한 성향점수를 이용한 매칭 방법보다 로지스틱의 성향점수를 이용한 방법이 편향을 더 많이 줄일 수 있으므로, 본 논문에서는 로지스틱 회귀에 의한 성향점수를 이용하였다. 그리고 교통사고 발생 유무에 영향을 미칠 수 있는 요인과 시간대를 공변인으로 하여, 교통사고 발생 유무 간에 공변인의 차이가 나지 않도록 매칭 방법은 일대일 매칭 (one-to-one matching) 방법을 고려하였다.

그리고 성향점수 매칭 방법을 이용하여 생성한 데이터를 이용하여, 기상상태에 따른 교통사고 발생 모형을 도출하기 위하여 3단계의 진행과정을 설정하였다. 1단계는 분석을 위한 자료구축단계로 기상자료와 교통사고 자료를 결합시켜 기초통계량을 구하고, 2단계는 교통사고 발생 유무별로 기상 상태의 차이를 알아보기 위하여 카이제곱 검정 (chi-squared test) 및 독립표본  $t$  검정 (independent sample  $t$  test)을 실시하고자 한다. 마지막으로 3단계는 기상상태가 교통사고 발생 유무를 예측하기 위한 모형을 도출하기 위하여 로지스틱 회귀분석 (logistic regression analysis) 및 의사결정나무분석 (decision tree analysis)을 실시하고자 한다.

### 3. 분석결과

#### 3.1. 카이제곱 검정 및 독립표본 $t$ 검정

Table 3.1과 Table 3.2는 사고유무에 따른 강수유무, 12시간 강수량, 온도, 풍속의 차이를 알아보기 위하여 카이제곱 검정 및 독립표본  $t$  검정을 실시한 결과이다. 먼저 Table 3.1의 사고유무에 따른 강수 유무를 살펴보면, 비가 오지 않은 날에 비하여 비가 온 날의 사고의 비율이 더 높은 것을 확인 할 수 있다 ( $\chi^2$  ( $df$ ) = 3.842 (1),  $p < 0.05$ ).

**Table 3.1** The result of chi-squared test

	Existence of accident		Total	$\chi^2$ ( $p$ )	
	No	Yes			
Existence of rainfall	No	1095 (96.1%)	1075 (94.3%)	2170 (95.2%)	3.842 (0.050)
	Yes	45 (3.9%)	65 (5.7%)	110 (4.8%)	
Total		1140 (100.0%)	1140 (100.0%)	2280 (100.0%)	

그리고 Table 3.2의 사고 유무에 따른 12시간 강수량, 온도, 풍속의 비교를 살펴보면, 온도의 경우 사고가 있는 날의 기온 ( $M = 15.953$ ,  $S.D = 8.6859$ )이 그렇지 않은 날의 기온 ( $M = 14.902$ ,  $S.D = 8.8067$ )에 비하여 더 높은 것을 확인할 수 있다 ( $t$  ( $df$ ) = -2.869 (2278),  $p < 0.01$ ). 하지만 사고 유무에 따라 12시간 강수량과 풍속은 차이가 존재하지 않은 것을 알 수 있다.

**Table 3.2** The result of independent samples  $t$ -test

	Existence of accident				$t$ ( $df$ )	$p$
	No		Yes			
	Mean	S.D	Mean	S.D		
12-hourly precipitation	1.99	10.211	1.84	9.088	0.360 (2278)	0.719
temp	14.902	8.8067	15.953	8.6859	-2.869 (2278)	0.004
wind speed	2.441	1.5866	2.432	1.6318	0.146 (2278)	0.884

#### 3.2. 로지스틱 회귀분석 (Logistic regression analysis)

로지스틱 회귀분석은 일반적 회귀분석과 유사한데, 종속변수가 일반적인 회귀분석과 달리 구간 비율 척도 변수가 아니라 0과 1로 나타낼 수 있는 이분형 범주형 변수라는 것이 다른 점이라 할 수 있다. 게다가 로지스틱 회귀분석은 회귀계수를 가지고 설명변수들의 오즈비 (odds ratio)를 쉽게 구할 수 있다는 이점이 있고 설명변수는 구간 비율척도와 범주형 변수를 모두 사용할 수 있다 (Park과 Kye, 2013; Park과 Choi, 2014).

AWS 기상자료에 대한 로지스틱 회귀분석을 통한 교통사고 발생 유무를 추정한 로지스틱 회귀모형의 계수는 Table 3.3과 같다.

**Table 3.3** The results of logistic regression analysis

	B	S.E.	$\chi^2$	$d.f.$	$p$	EXP(B)	95% C.I. for EXP(B)	
							LB	UB
existence of rainfall	.508	.224	5.163	1	.023	1.662	1.072	2.575
12-hourly precipitation	-.007	.005	2.145	1	.143	.993	.983	1.002
temp	.014	.005	8.431	1	.004	1.014	1.005	1.024
wind speed	-.016	.027	.345	1	.557	.984	.934	1.037
intercept	-.191	.100	3.637	1	.056	.826		

로지스틱 회귀모형의 계수를 살펴보면, 강수유무와 기온이 교통사고 발생유무에 통계적으로 유의미한 영향을 미치는 것을 확인 할 수 있다. 이는 앞의 결과와 같은 것으로, 먼저 강수유무의 경우 계수가 0.508 ( $p < .05$ )로 나타나 오즈비는  $\text{Exp}(0.508)=1.662$ 으로 비가 오지 않는 경우에 비하여 비가 오는 경우 교통사고가 발생할 확률이 1.662배 증가하는 것을 알 수 있다. 또한 기온의 경우, 계수가 0.014 ( $p < .01$ )로 나타나 오즈비는  $\text{Exp}(0.014)=1.014$ 로 기온이 1도 상승할수록 교통사고가 발생할 확률이 1.014배 증가하는 것을 알 수 있다. 즉, 비가 오고 기온이 높은 여름철의 경우 다른 경우에 비하여 더 높은 사고가 발생할 것으로 생각할 수 있다. 나머지 12시간동안 강수량, 풍속은 교통사고 발생유무에 통계적으로 유의미한 영향을 미치지 않는 것을 확인할 수 있다 ( $p > .05$ ).

### 3.3. 의사결정나무분석 (Decision tree analysis)

의사결정나무분석은 의사결정규칙 (decision rule)을 나무구조처럼 도표화하여 분류, 예측을 수행하는 분석 방법이며, 분류 또는 예측의 과정이 나무구조에 따른 추론규칙에 의해 표현되기에 다른 방법들에 비해 그 과정을 연구자가 쉽게 이해할 수 있고 또 설명할 수 있던 장점이 있다 (Cho, 2014; Lee와 Kim, 2014; Seok과 Lee, 2013). 본 논문에서는 Tree 노트에서 분리기준은 CHAID (chi-squared automatic interaction detection) 알고리즘을 이용하였고, 끝마디에 포함되는 최소 관측개체의 개수는 5개, 부모마디가 분리되기 위하여 요구되는 관측개체의 개수는 10개, 최대 가지분리의 개수는 2개로 지정하였다. 의사결정나무구조는 다음의 Figure 3.1과 같다.

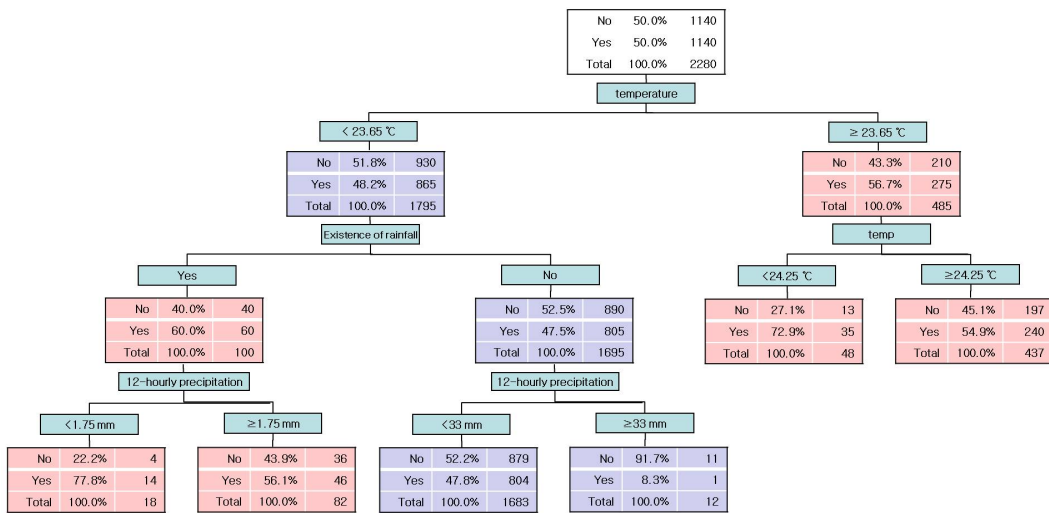


Figure 3.1 The construction of decision tree

이 그림을 통해 독립변수에 따라서 교통사고 발생여부가 어떻게 되는지를 알 수 있다. 먼저 전체 자료의 경우 교통사고 발생 1140건 (50.0%)과 교통사고 미발생 1140건 (50.0%)으로 총 2280건의 자료가 첫 번째 분류기준인 기온에 따라서 기온이 23.65도 이상인 경우 교통사고 발생 275건 (56.7%) 교통사고 미발생 210건 (43.3%)으로 분류되어 전체 자료에 비해서 교통사고 발생 정도의 순수도가 상승한 것을 확인할 수 있고, 기온이 23.65도 미만인 경우 교통사고 발생 865건 (48.2%) 교통사고 미발생 930건 (51.8%)으로 분류되어 전체 자료에 비해서 교통사고 미발생 정도의 순수도가 상승한 것을 확인할 수 있

다. 즉, 기온이 23.65도 이상인 경우 교통사고 발생의 확률이 높아지고, 23.65도 미만인 경우 교통사고 미발생의 확률이 높아진다는 것을 알 수 있다. 이와 같은 방법으로 다음 분류기준인 강수유무를 보면 비가 오는 경우 교통사고 발생의 순수도가 상승하는 것을 확인할 수 있고, 비가 오지 않는 경우 교통사고 미발생의 순수도가 상승하는 것을 확인할 수 있다. 거기에다 비가 오는 경우 12시간 동안의 강수량이 1.75mm미만인 경우 교통사고 발생의 순수도는 더욱더 높아지는 것을 확인할 수 있다. 그리고 비가 오지 않는 경우 12시간 동안의 강수량이 33mm이상인 경우 교통사고 발생의 순수도는 더욱더 높아지는 것을 확인할 수 있다. 즉, 기온이 23.65 이하이면서 12시간 동안 강수량이 1.75mm 미만인 지금 막 비가 오는 중인 그런 기상상태일 때 다른 경우에 비하여 더 높은 사고가 발생할 것으로 생각할 수 있다.

위의 분류기준을 이용하여 교통사고 발생 유무의 확률이 가장 높은 노드는 다음 Table 3.4와 같다.

**Table 3.4** The highest probability node

node	result	prob.	temp	Existence of rainfall	12-hourly precipitation
10	No	91.7%	< 23.65℃	No	≥ 33 mm
7	Yes	77.8%	< 23.65℃	Yes	<1.75 mm

#### 4. 결론

본 연구는 다양한 기상요인의 영향 정도에 따른 교통사고 발생유무를 분석하여 도로상의 교통사고 발생상황의 분류를 목적으로 하였다. 이를 위해 부산 해운대구의 센텀남대로 및 해운대로의 2013년도 교통사고 발생 자료와 지역별 상세 기상 관측 자료인 AWS 기상자료 (시간당 강수량, 강수유무, 기온, 풍속), 시간대, 요일을 활용하였다. 여기서 교통사고 발생 유무에 영향을 미칠 수 있는 요일과 시간대를 공변인으로 하여, 교통사고 발생 유무 간에 공변인의 차이가 나지 않도록 일대일 성향점수 매칭 방법을 사용하였다. 이를 통한 자료를 이용하여 로지스틱 회귀모형 및 의사결정나무모형을 이용하여 교통사고 발생유무에 대한 영향요인을 확인하고, 이 두 분석방법에서 제시하는 오분류표를 통한 예측정확성을 비교 비교함으로써 보다 예측 정확성이 높은 통계분석을 통해 교통사고를 유발할 확률이 높은 요인들을 확인하고자 하였으며, 그 결과를 요약하면 다음과 같다.

첫째, 카이제곱 검정과 독립표본  $t$  검정을 통해 교통사고 발생유무에 따라서 기온과 강수유무가 차이가 있는 것으로 나타났다. 즉, 교통사고가 발생한 경우 발생하지 않은 경우에 비하여 기온이 더 높고, 비가 오는 것을 확인할 수 있었다.

둘째, 로지스틱 회귀분석을 통해 기온과 강수유무가 교통사고 발생 확률에 영향을 주는 것으로 나타났다. 기온이 증가할수록, 비가 올수록 교통사고 발생확률은 더 높아지는 것을 확인할 수 있었다. 그리고 의사결정나무분석을 통해 기온이 23.65도 이상인 경우 그렇지 않은 경우에 비해 교통사고 발생이 더 많아지는 것을 알 수 있었다. 하지만 교통사고가 가장 많이 발생할 확률이 가장 높은 경우는 기온이 23.65도 이하이지만 비가 막 내리기 시작할 때라는 것을 확인할 수 있었다. 즉, 강수유무가 교통사고 발생유무에 많은 영향을 끼친다는 것을 알 수 있었다.

본 연구에서는 AWS 기상자료만을 사용하여 교통사고 발생유무에 영향을 미치는 요인을 분석하였지만, 교통사고 발생에는 AWS 기상자료와 같은 기초적인 기상자료 뿐만 아니라 태풍과 안개 같은 좀 더 특수한 상황의 기상상태들도 영향을 미칠 것이고, 도로위의 상황들 (과속방지카메라 설치유무, 신호등, 도로위의 함몰 등) 도 영향을 미칠 것이라 생각하므로 이들을 함께 고려한 연구가 필요한 것으로 생각한다. 또한 분석결과들을 통해 계절성이 교통사고 발생 확률에 영향을 주는 것으로 보이므로 계절에 따른 연구가 필요한 것으로 생각한다.

## References

- Cho, J. S. (2014). Analysis of employee's characteristic using data visualization. *Journal of the Korean Data & Information Science Society*, **25**, 727-736.
- Ibrahim, A. T. and Hall, F. L. (1994). *Effect of adverse weather conditions on speed-flow-occupancy relationships*, Transportation Research Board, Washington
- Jeong, S. J. (2011). *A study of the characteristics of traffic accidents in rainy conditions on freeways*, Master Thesis, University of Seoul, Seoul.
- Lee, J. Y. and Kim, H. J. (2014). Identification of major risk factors association with respiratory diseases by data mining. *Journal of the Korean Data & Information Science Society*, **25**, 373-384.
- Lynn A. S. and Barbara C. F. (1978). An analysis of the relationship between rainfall and the occurrence of traffic accidents. *Journal of Applied Meteorology*, **17**, 711-715.
- Oh, J. S., Shim, Y. U. and Cho, Y. H. (2002). Effect of weather conditions to traffic flow on freeway. *KSCE Journal of Civil Engineering*, **6**, 413-420.
- Park, C. and Choi, H. S. (2014). An educational tool for binary logistic regression model using Excel VBA. *Journal of the Korean Data & Information Science Society*, **25**, 403-410.
- Park, C. and Kye, M. J. (2013). Penalized logistic regression models for determining the discharge of dyspnea patients. *Journal of the Korean Data & Information Science Society*, **24**, 125-133.
- Rosenbaum, P. R. and Rubin, D. B. (1983). The central role of the propensity score in observational studies for causal effects. *Biometrika*, **70**, 41-55.
- Seok, K. H. and Lee, T. W. (2013). Comparison of data mining methods with daily lens data. *Journal of the Korean Data & Information Science Society*, **24**, 1341-1348.

## The effect of road weather factors on traffic accident - Focused on Busan area -

Kyeongjun Lee<sup>1</sup> · Imgook Jung<sup>2</sup> · Yunhwan Noh<sup>3</sup> · Sanggyeong Yoon<sup>4</sup> · Youngseuk Cho<sup>5</sup>

<sup>12345</sup>Department of Statistics, Pusan National University

Received 26 March 2015, revised 6 April 2015, accepted 27 April 2015

### Abstract

Them traffic accidents have been increased every year due to increasing of vehicles numbers as well as the gravitation of the population. The carelessness of drivers, many road weather factors have a great influence on the traffic accidents. Especially, the number of traffic accident is governed by precipitation, visibility, humidity, cloud amounts and temperature. The purpose of this paper is to analyse the effect of road weather factors on traffic accident. We use the data of traffic accident, AWS weather factors (precipitation, existence of rainfall, temperature, wind speed), time zone and day of the week in 2013. We did statistical analysis using logistic regression analysis and decision tree analysis. These prediction models may be used to predict the traffic accident according to the weather condition.

*Keywords:* AWS, decision tree analysis, Heaundae, logistic regression analysis, traffic accident, weather factor.

---

<sup>1</sup> Post doc. fellow, Department of Statistics, Pusan National University, Busan 609-735, Korea.

<sup>2</sup> Master's course, Department of Statistics, Pusan National University, Busan 609-735, Korea.

<sup>3</sup> Master's course, Department of Statistics, Pusan National University, Busan 609-735, Korea.

<sup>4</sup> Master's course, Department of Statistics, Pusan National University, Busan 609-735, Korea.

<sup>5</sup> Corresponding author: Professor, Department of Statistics, Pusan National University, Busan 609-735, Korea. E-mail: choys@pusan.ac.kr