

제 16회 SAS 분석 챔피언십 1차 제안서

팀 번호	SA1029		
팀 장	정유진	팀원 명	김유민, 이도진
연락처	010-2160-8467	이메일	dlx0021@naver.com

1. 교통사고에 영향을 미치는 원인과 징후에 대한 탐색 및 시각화 방안 제시 (1장 이내)

- ✓ 2016~2017년 서울시 교통사고 데이터 분석 및 방향성
 - 사고유형 중 '차대차'(70%)가 가장 큰 비중을 차지함. 차대차 사고 중에서도 '측면직각충돌'(22%)의 비중이 가장 큼. 따라서 교차로의 교차각과 교차거리, 신호 교차로인 경우 신호 등의 작동순서 및 시간 등 세부 요인을 구체적으로 분석할 필요가 있음.
 - 법규위반 중 '안전운전불이행'(56%)이 가장 큰 비중을 차지함. 따라서 가해자 정보, 주변 환경 등 사고원인의 1, 2차 요인을 세부적으로 규명할 필요가 있음.
 - 도로형태 중 '단일로-기타'(52.15%)가 가장 큰 비중을 차지함. 따라서 사고지점부근이 직선인지 곡선로인지의 여부, 도로폭이나 보도 등의 간격, 진행방향의 시야거리 등을 세부적으로 탐구할 필요가 있음.
 - 단, 가장 큰 비중을 보였던 위의 세 가지 요소를 원인으로 보고 분석할 예정이며, 이외의 나머지 요소들이 미치는 영향을 간과하지 않고 복합적으로 탐구할 것임.
- ✓ 교통 시설물 데이터 분석 및 방향성
 - 상관분석 결과 가장 유의한 3가지 요소를 교통사고 데이터와 함께 지도에 3개 구로 나누어 시각화 함. *[그림 1] 참조*
 - 교량, 자동차전용도로, 주거지역, CCTV, 횡단보도, 험프, 운행제한시설물, 버스정류장, 어린이 보호구역을 교통사고에 영향을 미치는 원인이라고 보고 각각의 구별 빈도를 구함. (여기서 발생하는 결측치는 각각 변수의 평균으로 대체함.) 이를 차후 회귀분석을 통한 모델링의 회귀변수로 지정할 것이며, 보다 세부적인 분석을 위해 구별 빈도가 아닌 도로구간별 빈도를 구할 것임.
- ✓ 차후 데이터 분석 방향
 - 도로 폭, 길이, 교차점, 교통량, 통행 속도 등의 도로 특성을 분석할 것임.
 - 연령대나 차종 등의 가해자 성향을 분석할 것임.
 - DTG 데이터와 위험행동기준을 비교 및 분석해 위험 구간을 예측하는 파생변수를 생성할 것임.
 - 구별 구분을 도로 구간별 구분으로 세분화할 것임.

2. 교통사고 위험구간 예측을 위한 파생변수 생성 및 모델링 방법론 제시 (1장 이내)

- ✓ 구별 교통사고 빈도수 데이터 및 구별 교통시설물 개수 데이터를 7(train data) : 3(test data)로 분할함.
 - 교통 시설물 데이터의 교량, 자동차전용도로, 주거지역, 횡단보도, 험프, 운행제한시설물, 버스정류장, 어린이보호구역 변수의 빈도를 이용한 주성분 분석의 결과로 고유값(Eigenvalue)이 1보다 큰 3가지 주성분을 선택함. 이 3가지 주성분의 고유벡터(Eigenvector)를 통해 파생변수 prin1, prin2, prin3를 생성함. [그림 2] 참조
 - $\text{prin1} = 0.283595 * \text{교량빈도} + 0.218547 * \text{자동차전용도로} + 0.343001 * \text{주거지역} + 0.303224 * \text{cctv} + 0.463460 * \text{횡단보도} + 0.396733 * \text{험프} + 0.284312 * \text{운행제한시설물} + 0.335722 * \text{버스정류장} + 0.310615 * \text{어린이보호구역}$
 - $\text{prin2} = -0.129319 * \text{교량빈도} + 0.540566 * \text{자동차전용도로} - 0.248947 * \text{주거지역} + 0.132641 * \text{cctv} + 0.001012 * \text{횡단보도} + 0.290809 * \text{험프} - 0.592007 * \text{운행제한시설물} - 0.258303 * \text{버스정류장} + 0.331260 * \text{어린이보호구역}$
 - $\text{prin3} = 0.606380 * \text{교량빈도} - 0.183978 * \text{자동차전용도로} + 0.020881 * \text{주거지역} - 0.124828 * \text{cctv} + 0.045178 * \text{횡단보도} - 0.211247 * \text{험프} + 0.060473 * \text{운행제한시설물} - 0.577419 * \text{버스정류장} + 0.445759 * \text{어린이보호구역}$
- ✓ 3가지 주성분에 대한 설명
 - prin1- 횡단보도가 영향력이 있고 나머지 변수는 비슷한 정도의 영향력을 가짐.
 - prin2- 자동차전용도로와 운행제한시설물은 영향력이 있고, 횡단보도는 영향력이 미미함.
 - prin3- 교량, 버스정류장, 어린이보호구역은 영향력이 있고, 주거지역과 운행제한시설물은 영향력이 미미함.
 - (단, 고유벡터의 계수가 0.4 이상일 때 영향력이 있으며, 0.7 이상일 때 영향력이 크다는 기준을 적용함.) [그림 2] 참조
- ✓ 파생변수를 이용한 회귀분석(유의수준=6%)
 - Model: 사고발생빈도 예측값 = $-397.532 + 4.957 * \text{prin1} - 4.3148 * \text{prin2} + 7.28 * \text{prin3}$ [그림 3] 참조
 - 설명계수는 86.5%로 높은 설명력을 가짐.
 - 그러나, fit diagnostics의 표시된 그래프를 보면 등분산성 조건을 만족하지 못한다는 것을 발견함. 따라서 종속변수를 변환하는 방법, PROC GENMOD or PROC GLMMIX를 사용하는 방법, 등분산 가정을 하지 않는 모형을 선택하는 방법 등을 적용해 이 문제를 해결할 것임. [그림 4] 참조
 - 2차 제안시에는 더 다양한 파생변수들을 추가하고 데이터를 더욱 세분화하여 보다 정확한 분석을 진행할 것임.
 - 위의 모형을 통해 도출되는 예측값에 따라 위험지수등급을 분류할 것이며, 사고 위험 구간 및 사고 형태 예측은 이를 바탕으로 추후에 분석할 예정임.
- ✓ Train data를 통해 도출한 예측모형에 Test data를 대입하여 그 오차를 비교함. [그림 5]참조

3. 예상되는 분석결과와 그에 관한 활용방안 아이디어 제시 (1장 이내)

✓ 현재 모델링에 의해 예상되는 분석결과

- 3가지 주성분 중 첫번째 주요한 원인인 prin 1을 보면, 횡단보도와 험프가 각각 0.463, 0.396의 고유벡터계수로 가장 영향력이 큼. 횡단보도와 험프의 유무는 도로의 특성을 반영함.
- 두번째 주요한 원인인 prin 2를 보면 자동차 전용도로와 운행제한시설물이 각각 0.54, -0.59의 고유벡터계수로 가장 영향력이 큼. 자동차전용도로와 운행제한시설물은 차량과 도로와의 관계를 반영함.
- 마지막으로 주요한 원인인 prin 3를 보면 교량, 버스정류장, 어린이보호구역이 각각 0.6, -0.578, 0.44의 고유벡터계수로 가장 영향력이 큼. 교량, 버스정류장, 어린이보호구역은 모두 특수구역이라는 특성을 지님.
- 위의 주성분분석과 회귀분석을 시작으로 추가적인 여러 분석을 시행할 것이며, 위의 세 가지 주성분으로 도출된 각각의 요인들(횡단보도와 험프, 자동차전용도로와 운행제한시설물, 버스정류장과 교량, 어린이보호구역)은 교통사고 데이터 해석에 있어 서로 유의한 방향으로 분석될 것이라고 예상됨.

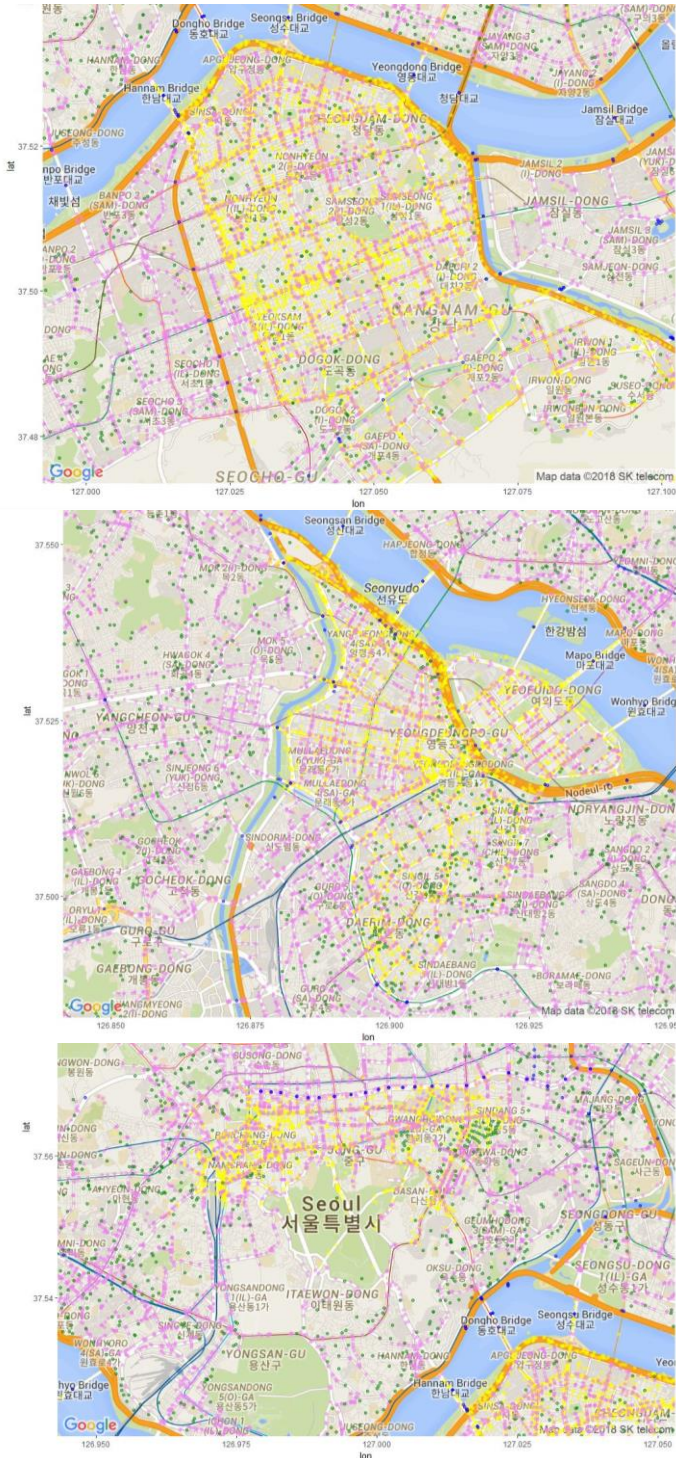
✓ 재구성한 모델링에 의해 예상되는 분석결과와 활용방안 아이디어

- 그러나, 현재 모델링한 결과는 각 변수를 늘렸을 때, 사고빈도가 늘어나는 것으로 나옴. 면밀한 분석을 통해 모델링을 재설정한다면, 사고빈도를 줄이는 변수들이 있을 것으로 예상됨.
- 재설정된 모델링을 통해 도로구간별로 각 변수가 부족한 곳에 해당 시설물을 추가적으로 설치하도록 제안할 것임. 또한 같은 주성분을 지닌 변수들의 특성을 모두 가진 시설물을 개발하는 방안을 제시할 것임.
- 예를 들어, 횡단보도와 과속방지턱이 같은 주성분을 가지는 것으로 나타난다면, 이 두 시설물을 융합한 과속방지턱 겸용 횡단보도 장치를 설치하도록 제안할 수 있음. 이는 횡단보도 자체의 높이를 늘리므로 어린이보호구역에 효과가 있을 것으로 예상됨.
- 만약 횡단보도와 어린이보호구역이 같은 주성분을 가지는 것으로 나타난다면, 어린이 보호구역에 위치한 횡단보도에 자동차와 아이들의 통행을 제한할 수 있는 안전 바(녹색어머니회 깃발과 같은 역할)를 설치하는 방안을 제안할 수 있음.

[별첨] 추가가 필요할 경우에만 제출 (최대 3장 이내)

[그림1] 강남구, 영등포구, 중구

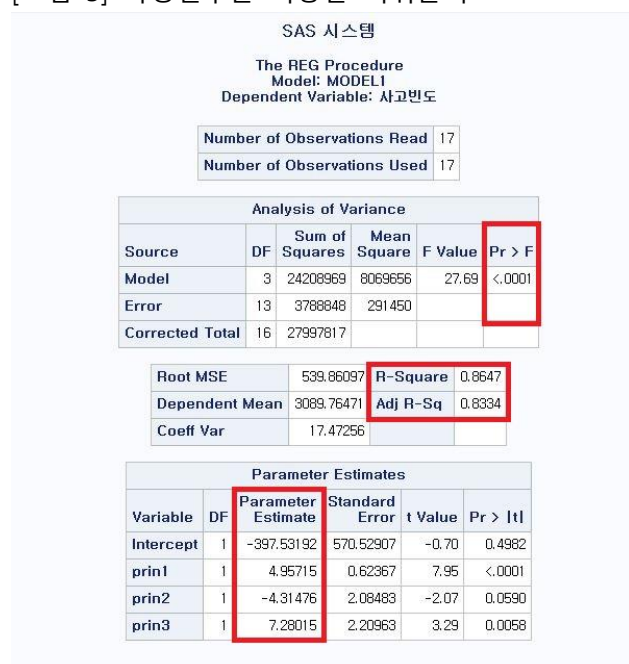
Yellow – 교통사고 지점, Purple – 횡단보도, Green – 거주지역, Blue – 교량



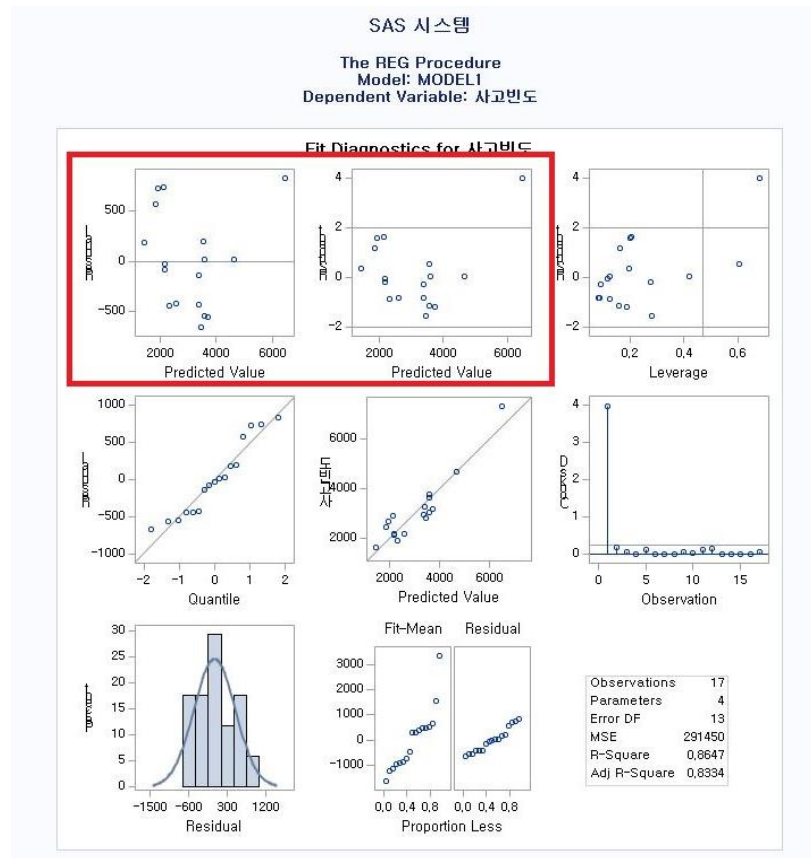
[그림 2] 교통시설물데이터 빈도 주성분분석 결과



[그림 3] 파생변수를 이용한 회귀분석



[그림 4]



[그림 5]

