

# 회귀분석 및 검정을 통한 친환경자동차검사결과 분석

김상화\*, 박지양\*\*, 박성윤\*\*\*, 김종훈\*\*\*, 황무성\*\*\*\*, 신성동\*\*

\*경희대학교 컴퓨터공학과

\*\*한국교통안전공단 첨단연구개발처

\*\*\*충남대학교 전기공학과

\*\*\*\*동아대학교 산업경영공학과

e-mail:shkim18@khu.ac.kr

## Analysis of Green Vehicle Inspection Results Through Regression and Test

Sang-Hwa Kim\*, Ji-Yang Park\*\*, Seong-Yun Park\*\*\*, Jong-Hoon Kim\*\*\*, Mu-Seong Hwang\*\*\*\*, Seong-Dong Shin\*\*

\*Dept. of Computer Engineering, Kyunghee University

\*\*Advanced R&D Department, Korea Transportation Safety Authority.

\*\*\*Dept. of Electrical Engineering, Chungnam University

\*\*\*\*Dept. of Industrial Management Engineering, Donga University

### 요약

국제적으로 온실가스·미세먼지 감축 및 글로벌 미래차 시장을 선점하기 위해 친환경 자동차의 연구개발 및 보급이 활발하게 진행되고 있다. 이러한 친환경자동차의 보급에 따른 운행자동차의 안전을 위해 한국교통안전공단에서는 자동차 검사를 주기적으로 시행하고 있다. 자동차 검사에서는 실제로 자동차 전반적인 육안검사, 전조등검사, 제동력검사, 배출가스 검사 등을 대표적으로 진행하고 있고 이에 대한 결과를 통계적으로만 공표를 진행하고 있다.

점차적으로 증가하는 친환경자동차에 대해서 검사전반적인 데이터 통계데이터를 넘어서 제작사의 제작기술 및 성능 향상 유도, 고객의 사전정비 및 정기적인 관리를 유도하기 위해 조금 더 심층적인 분석이 필요한 상황이다.

본 연구에서는 이러한 검사데이터에 대해서 기초적인 통계가 아닌 다양한 통계 분석을 통해 검사데이터 항목 간 회귀 분석 및 검정 분석을 통해 항목 간 관계성 및 통계적 유의미한 인자들을 도출하고자 한다. 통계적 분석을 위해서는 데이터의 특성 분석을 수행해야 하고, 결측치 및 이상치에 대한 데이터 전처리 방법 등을 고려하여 분석을 적용하여 자동차 검사의 최종결과 시 중요 변수로서 어떤 것 들이 있는지에 대해서 분석 하고자 한다.

### 1. 서론

국제적으로 탄소중립을 위해 저탄소 무공해 차량개발에 노력을 다하고 있고 전기차는 친환경 미래 모빌리티의 근간으로 온실가스·미세먼지 감축 및 글로벌 미래차 시장 선점을 위해 보급이 가속화가 되어 지고 있다.<sup>[1]</sup> 또한, 하이브리드 자동차 또한 배출가스 규제를 극복하기 위해 개발이 되어 있고 미래의 자동차 시장은 전기 및 하이브리드 차량이 선도할 것으로 보인다.

운행자동차 안전을 위해 한국교통안전공단에서는 자동차 검사를 주기적으로 시행하고 있다.<sup>[2]</sup> 실제로 작년기준 1200만 대를 검사하였으나, 작년 자동차 검사결과 기준 약 20%가 부적합을 받음으로서 국민의 안전에 대한 노력이 절실히 지고

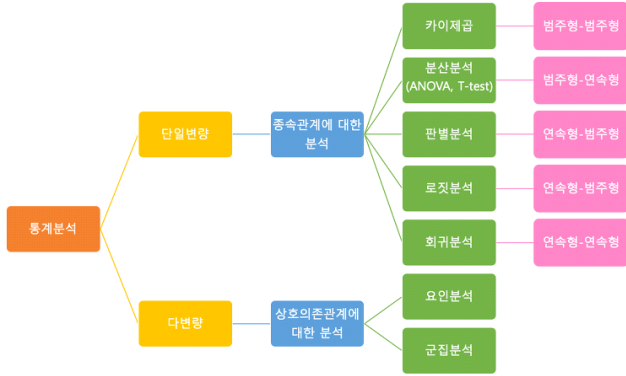
있다.<sup>[3]</sup> 매년 자동차검사결과에 대해서 자동차 용도별, 차종별, 규모별, 시도별, 차령별, 연료별, 누적주행거리별 결과가 통계적으로 공표가 되어 지고 있으나, 이는 1차원적인 통계결과 이다.<sup>[4]</sup>

전기 및 친환경자동차는 안전사고가 많이 발생하며 이를 방지하기 위해 관련 부처 및 제작사 등에서 많은 노력을 기울이고 있다. 이 중 한국교통안전공단에서는 자동차 검사를 진행하고 있으며, 신규 검사 방식 적용 등 다양한 부분에서 차량상태를 점검하여 데이터를 수집하고 있으나, 데이터에 대한 면밀한 분석은 진행되어 오고 있지는 않다.<sup>[5]</sup>

본 연구에서는 이러한 검사데이터에 대해서 기초적인 통계가 아닌 회귀 분석 및 검정을 통해 유효한 통계분석 데이터를 추출해보고 이에 따라 친환경자동차에 대한 사전정비 및 정기적인 관리를 유도하며, 제작사의 제작기술 및 성능향상을 유도하고자 한다.

## 2. 통계분석방법

통계적 분석을 위해서는 데이터의 특성 분석을 수행해야 하며, 결측치 및 이상치에 대한 데이터 전처리 방법 등을 고려하여 분석을 적용해야 한다. [그림 1]과 같이 각 변수의 유형(범주형/연속형)에 맞는 분석 방법을 적용해야 한다.<sup>[6]</sup>



[그림 1] 변수 유형에 따른 통계적 분석 방법

크기가 없고 'A', 'B', 'C'와 같이 분류할 수 있는 범주형 데이터로는 차종, 연료, 제작사, 차명 등이 있으며, 연속 또는 이산적으로 표현할 수 있는 수치데이터로서 주행거리, 전조등광량 등이 [표 1]과 같이 나타낼 수 있다.

[표 1] 자동차검사 데이터 유형 분류

구분	검사항목
범주형 데이터	검사종류
	차종/규모/연료
	제작사
	차명
	판정결과
	전조등 적합여부
	제동력 적합여부
	최종판정결과
...	
연속형 데이터	검사일자
	주행거리
	중량
	전조등광량
	제동력
	...

### 2.1 자동차검사결과 데이터 전처리

자동차검사결과와 통계적 분석의 용이성을 위하여 데이터전처리를 시행하였다. [표 2]와 같이 전처리를 진행하였으며, 자동차 검사결과 데이터 항목 중 데이터 수의 50% 이상이 결측치인 경우 분석 신뢰성이 낮

아지므로, 본 자동차 검사결과 데이터 분석에서는 사용하지 않는다.

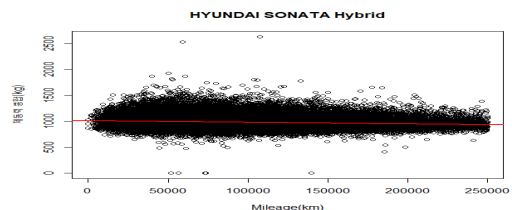
또한, 검사 대수가 500대 이하인 차량은 데이터 분석의 신뢰도가 떨어 질수 있어 분석에서 제외하였다.

[표 2] 자동차검사결과 데이터 전처리 내용

구분	기존	수정
차명 및 제작사	<ul style="list-style-type: none"> <li>•(차명) Hyundai 소나타, 현대 소나타, 현대 쏘나타 등 동종이명 다수</li> <li>•(제작사) 기아 자동차(주), 기아 자동차 주식회사 등 동사이명 다수</li> </ul>	<ul style="list-style-type: none"> <li>•차명 및 제작사 이명 항목 통일 예) Hyundai소나타, 현대 소나타 → 현대소나타</li> </ul>
판정 결과	<ul style="list-style-type: none"> <li>•1부부적합, 1부불, 1부불+시정권고, 2부부적합, 2부불, 2부불+시정권고, 시정권고, 적합</li> </ul>	<ul style="list-style-type: none"> <li>•1부부적합, 2부부적합, 시정권고, 적합</li> </ul>
검사일자	<ul style="list-style-type: none"> <li>•20180101 ~ 20201231 연속형 변수</li> </ul>	<ul style="list-style-type: none"> <li>•2018, 2019, 2020으로 변경 ex) 20180101~20181231 → 2018</li> </ul>
사업용/비사업용	<ul style="list-style-type: none"> <li>•-</li> </ul>	<ul style="list-style-type: none"> <li>•차량 번호 중 "아, 바, 사, 자, 배, 하, 허, 호" 포함 차량은 사업용, 그 외 비사업용</li> </ul>
주행거리	<ul style="list-style-type: none"> <li>•주행 거리 90만 km 이상 차량 다수(10대)</li> </ul>	<ul style="list-style-type: none"> <li>•TS 확인 결과 주행거리 데이터 이상 확인 (1의 자리 수 추가 작성)</li> <li>•관련 내용 추후 협의하여 변경 진행</li> <li>•매 5만 km 차량 마다 분류</li> </ul>

## 3. 회귀분석

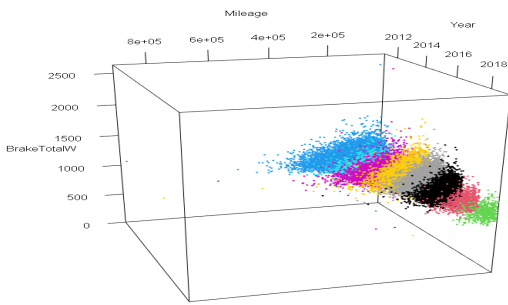
회귀분석은 독립변수(x)와 종속변수(y) 사이에 관계가 있는 경우 이를 수식적으로 나타내는 방법이다. 연속형-연속형 데이터의 관계를 분석하기 위해 주로 사용되므로 자동차검사결과 데이터 중 연속형 데이터인 제동력과 전조등 광도를 종속변수로 설정하여 각각 주행거리와 연식에 따른 회귀분석을 진행하였다.



[그림 2] 현대 소나타 하이브리드의 주행거리에 따른 제동력 종합(kg) 선형회귀분석

[그림 2]는 현대 소나타 하이브리드 자동차의 주행거리(x)에 따른 제동력 종합(kg)(y)관계를 선형회귀 분석을 통해 나타낸 그래프이다. [그림 2]로부터 1km 주행에 따라 제동력 종합(kg)은 약 0.0003kg 감소함을 알 수 있다.

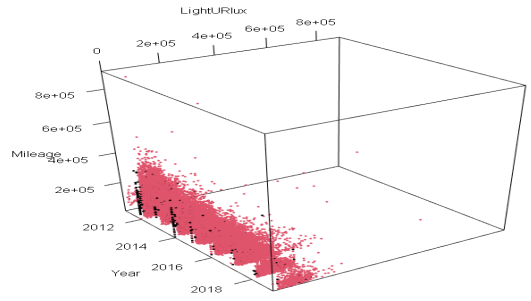
[그림 3]은 현대 소나타 하이브리드의 주행거리, 연식에 따른 제동력 종합(kg)을 산점도로 나타낸 그래프이다. 이에 따라 주행거리 1km에 따라 제동력이 0.000088 감소하는 것을 알 수 있고, 연식에 따라 12.27kg 증가 하는 것을 알 수 있으나, 두 그래프에서 확인 할 수 있듯이 단순히 직선으로는 정확히 추정하기는 힘든 것을 확인 할 수 있다.



[그림 3] 현대 소나타 하이브리드의 주행거리, 연식에 따른 제동력 종합(kg)변화

따라서, 결정계수인 R-squared를 계산하여 현재 회귀선이 실제값을 얼마나 잘 예측할 수 있는지 확인해 보았다. 현재 주행거리 및 연식에 따른 제동력 종합(kg)의 R-squared값은 0.04587로 예측에 대한 신뢰도가 낮음을 알 수 있다. R-squared를 높이기 위해 차량의 총 중량을 변수로 하여 다 변수 선형회귀를 수행하였다. 이에 따라 R-squared는 0.09222로 주행거리와 연식만으로 추정할 때보다 증가함을 확인 할 수 있었으며, 이후 자동차 검사데이터 중 연속형 데이터를 이용하여 제동력 예측을 수행 할 수 있는 독립변수 전체에 대해서 Backward elimination을 진행하였다. 이를 통해 만들어진 선형 회귀 분석 모델의 R-squared 값은 0.2185로 다소 모델의 설명력이 생성됨을 확인 할 수 있다.

마찬가지로, 주행거리에 따른 전조등(상향, 우) 광도는 [그림 4]와 같이 나타나며 1km 주행에 따라 0.04975 lux 감소하며, 연식이 오래 될수록 광도가 1,879lux씩 감소하는 것을 알 수 있다.



[그림 4] 현대 소나타 하이브리드의 주행거리, 연식에 따른 전조등(상향, 우) 광도

#### 4. 검정분석

통계적 분석에서 수집한 표본에 대해 수립한 가설이 옳고 그른지를 판정하는 것을 “가설 검정”이라고 한다. 통계적 검정 방법에는 분산분석법 및 카이제곱 검정 등이 있다.

##### 4.1 분산분석법(ANOVA)

분산분석법이란 두 집단 또는 그 이상의 집단 간의 평균의 차이를 검증하는 방법으로 주로 귀무가설을 비교하고자 하는 집단간의 평균이 같다고 설정하고, 대립가설은 평균이 다르다고 설정하여 검정한다.<sup>[7]</sup>

판정결과에 따른 제동력 주차(kg)를 일원분산분석을 1부부적합 - 2부부적합, 1부부적합 - 시정권고, 1부부적합 - 적합으로 수행하였다. 판정결과에 따른 제동력 주차(kg) 일원분산분석 결과는 [표 3]과 같이 나타났다. ANOVA 분석 결과를 확인해 보았을 때, 모든 집단간의 p-value가 유의수준( $\alpha$ )인 0.1보다 낮으므로 귀무가설을 기각한다. 이를 통해 모든 관계는 집단 간 평균이 차이나는 것을 확인 할 수 있고, 판정결과 별로 제동력 주차(kg)가 같지않다는 결론을 내릴 수 있다.

[표 3] 판정결과에 따른 제동력 주차(kg) 일원분산분석 결과

구분	F	p-value	평균(a)	평균(b)	기각여부
1부부적합 - 2부부적합	6.5917	0.01025	394.69	391.36	O
1부부적합 - 시정권고	84.415	<2.2e-16	394.69	402.25	O
1부부적합 - 적합	217.95	<2.2e-16	394.69	409.215	O
2부부적합 - 시정권고	58.07	2.561e-14	391.36	402.25	O
2부부적합 - 적합	101.94	<2.2e-16	391.36	409.215	O
시정권고 - 적합	76.14	<2.2e-16	402.25	409.215	O

## 4.2 카이제곱 검정

카이제곱 검정은 관찰된 빈도가 기대되는 빈도와 유의하게 다른지를 검증하는 방법이며, 범주형 자료로 구성된 데이터 분석에 이용한다. 일원 카이제곱 검정과 이원 카이제곱 검정으로 분류할 수 있으며, 일원카이제곱 검정은 한 개의 범주를 대상으로 하여 적합도를 검정하며, 이원 카이제곱 검정은 두 개 이상의 범주를 대상으로 독립성 및 동질성을 검정하는 방법이다.<sup>[8]</sup>

본 연구에서는 최종판정결과에 따라 어떤 변수가 가장 중요성을 갖는지 분석하였다.

본 통계적 분석의 목표는 최종판정결과를 유추하기 위해 관계성이 높은 인자를 추출하기 위한 방법이다. 카이제곱 검정 결과([표 4])를 통해 판정결과, 제조사, 연식, 주행거리, 차명이 최종결과를 유추하기 위한 중요변수로 선정 되었다.

[표 4] 최종판정결과 유추를 위한 변수 중요성 평가 결과

구분	X-squared	구분	X-squared
판정결과	0.1152	전조등적합여부	0.0201
제조사	0.0361	제동력(주차)	0.0179
연식	0.0315	규모	0.015
주행거리	0.0314	사업용	0.01088
차명	0.0281	제동력적합여부	0.00623
차량중량	0.0273	연료	0.0056
총중량	0.0273	차종	0.00202
전조등(상향 좌광도)	0.0239	제동력(종합)	0
전조등(상향 우광도)	0.0203		

## 5. 한계점

아직 전자장치에 대한 검사 방법이 확정이 안되어 있고, 육안검사가 위주이기 때문에 본 연구에서는 전기 및 하이브리드 자동차에 대한 가장 큰 특성 중 하나인 자동차 전자장치에 대해서는 적용하지 않고 제동력과 전조등 결과에 대해 집중하여 분석을 하였다.

## 6. 결 론

본 연구는 전기 및 하이브리드 자동차에 대해 검사 결과를 바탕으로 데이터 전처리부터 검사 결과 데이터 중 제동력과 전조등 결과에 대해서 통계적 분석 방법인 회귀분석, ANOVA분석, 카이제곱 분석 등을 통해 분석하였다.

본 연구의 결과를 요약하면 다음과 같다.

- 1) 자동차검사결과 Raw데이터의 경우 데이터의 정제를 위해 차명 및 제작사, 판정결과, 검사일자, 사업용/비사업용, 주행거리로 표준화 작업이 필요하다.
- 2) 회귀분석을 통해 친환경자동차의 주행거리, 연식에 따라 제동력을 구할 수 있으며, 대표적으로 소나타 하이브리드는 1km 주행 시 제동력이 0.000088 감소, 연식에 따라 12.27kg 증가 하는 것을 알 수 있다.
- 3) 제동력 예측을 수행 할 수 있는 독립변수 전체에 대해서 Backward elimination을 진행해본 결과 선형회귀 분석 모델의 R-squared값은 0.2185로 다소 모델의 설명력이 생성됨을 확인 할 수 있다.
- 4) ANONA 분석 결과를 통해 1부 부적합, 2부 부적합, 시정권고, 적합 사이의 관계는 집단 간 차이가 나타나는 것을 확인 할 수 있다.
- 5) 카이제곱 검정을 통해 최종판정결과에 따라 어떤 변수가 가장 중요한지 알아본 결과 판정결과, 제조사, 연식, 주행거리, 차명이 최종결과를 유추하기 위한 중요 변수로 선정 할 수 있다.
- 6) 이를 통해 친환경자동차의 검사결과의 통계결과뿐만 아니라 분석결과를 적용하여 친환경자동차의 관리 및 연구를 할 때 주요한 인자 선정을 할 수 있는 기초 연구 자료로 활용 할 수 있다.

### 참고문헌

- [1] 대한민국 정부, “2050 탄소중립 전략”
- [2] 자동차관리법 제43조, 자동차종합검사의 시행 등에 관한 규칙 등
- [3] 한국교통안전공단, “통계 분석으로 본 2020년 자동차검사 - 한 해 동안 약 1,200만대 검사, 전년 대비 3.1%(37만대) 증가”, 2021.05.31.
- [4] [www.kosis.kr](http://www.kosis.kr), 국가통계포털
- [5] 박지양, “전기자동차 운행안전성 확보를 위한 검사 필요성 연구”, 2020 한국자동차안전학회 춘계학술대회, 2020.6
- [6] 남기성, “사회과학에서 통계분석 방법의 선택에 관한 연구”, 한국자료분석학회, 제6권 제5호, pp 1,255 - 1,265, 2004
- [7] 박병수, “ANOVA에 의한 해양사고의 통계분석”, 해양환경안전학회, 제13권 제3호, pp191 - 198, 2007
- [8] 홍선경, “카이제곱 검정 통계량과 유관계수의 안전한 다자간 계산”, 한국정보과학회, 2014 한국정보과학회 제 41회 정기총회 및 동계학술발표회, pp225 - 227, 2014.12