

## 주행거리별 운행차 배출가스 분포 추정 모델에 관한 연구

### A Study on the Inference Model of In-use Vehicles Emission Distribution according to the Vehicle Mileage

김 현 우\*  
Hyun woo Kim

#### ABSTRACT

To investigate the safety of the in-use vehicles emission against the tail-pipe emission regulation, in-use vehicles emission trend according to vehicle mileage should be known. But it is impossible to collect all vehicles emission data in order to know that. Therefore, it is necessary to establish a statistically meaningful inference method that can be used generally to estimate in-use vehicles emissions distribution according to the vehicle mileage with relatively less in-use vehicles emission data. To do this, a linear regression model that solved the problems of data normality and common variance of error was studied. As a way that can secure the data normality,  $\ln(\text{emission})$  instead of emission itself was used as a sampled data. And a reciprocal of mileage was suggested as a factor to secure common variance of error. As an example, 36 data of FTP-75 test were handled in this study. As a result, using average value and standard deviation at each mileage which were inferred from a linear regression model, probability density distribution and cumulative distribution of emissions according to the vehicle mileage were obtained and it was possible to predict the deterioration factor through full useful life mileage and also possible to decide whether those in-use vehicles will meet the tail-pipe emission regulations or not.

주요기술용어 : In-use vehicle emission(운행차 배출가스), Linear regression(선형 회귀), Normality(정규성), Common variance(등분산성), Error(오차), Probability density(확률 밀도)

#### 1. 서 론

자동차의 급격한 증가는 대기오염을 가중시키고 있고, 특히 자동차의 보급률이 높은 대도시를 중심으로 한 지역에서는 자동차가 대기오염의 주범으로 등장했다. 국내에서도 자동차 보급대수가 1000만대를 돌파한 1997년 이후에 자동

차에 의한 대기오염의 비율이 높아져서, 자동차 배출가스에 대한 관심이 높아진 상태다. 이에 따라 각국에서는 심화되고 있는 자동차에 의한 대기오염을 개선하고자 자동차의 배출가스 규제를 강화하고, 배출가스 보증기간도 연장하고 있으며, 또한 운행차의 배출가스에 대한 관리도 점차 강화하고 있다.

특히, 자동차의 주행성능과는 관계가 적은 배

\* 회원, 순천대학교 기계·자동차공학부

출가스 관련 부품에 결함이 있는 운행차는 유해한 배기가스의 배출을 증대 시킨다는 점과 운행차에 대한 적절한 관리는 대기오염을 줄일 수 있는 여타 수단과 비교해서 단기적으로 가장 효율적이라는 점에서, 운행차의 배출가스의 관리에 관한 규제가 새롭게 부각되고 있다.<sup>1)</sup> 이미 1990년대 중반부터 미국에서 시행하고 있는 OBD-II 규제는 운행차의 과도한 배기가스 배출을 자체 ECU를 통하여 자기진단(Self-Diagnosis) 하는 것을 의무화한 법규이며, 역시 미국에서 2000년부터 시행하고 있는 CAP2000(Compliance Assurance Program)규제도 배출가스 인증체계를 운행차 중심으로 변경한 제도이다.<sup>2,3)</sup>

운행차 배출가스의 관리에서 가장 중요한 것이 실제로 운행되고 있는 차량의 배출가스 양을 파악하는 것이다. 특히, 주행거리에 따른 운행차의 배출가스 양은 환경당국의 배출가스 규제, 개정에 중요한 정보가 될 뿐 아니라, 자동차 제작사의 배출가스규제 만족을 위한 차량 개발에도 매우 중요한 정보이다.<sup>4,5)</sup>

실제로 운행하고 있는 서로 다른 주행거리를 가지는 차량들의 배출가스 시험 결과를 가능한 많이 확보하는 것이 보다 정확한 운행차의 배출가스 양을 파악하는 데에 도움이 되지만, 여기에는 시간과 비용면에서 한계가 있다. 따라서, 적은 양의 배출가스 시험 결과로 운행차의 배출가스 경향을 파악할 수 있는 신뢰할 수 있는 통계적 모델이 필요하다.

한편, 일반적으로 운행차의 배출가스 양은 주행거리에 따라 증가하는 것으로 알려져 있어 선형 회귀 분석 방법을 통해 접근할 수 있으나, 배출가스 양의 분포 특성이 알려져 있지 않으므로 통계적인 자료처리의 용이함을 위해서는 배출가스 양의 분포에 대한 적절한 가정이 필요하고, 또한 차량이 일정 거리 이상을 주행하면 엔진 및 관련 부품들의 열화나 차량관리 상태, 운전자의 주행 습관 등에 의해 차량별 차이가 커지므로, 주행거리가 커짐에 따라 배출가스 양의 산포도가 커지는 경향이 있어 선형 회귀 분석을 위한 등분

산성(Common Variance) 조건에 부합되지 않는다는 문제가 있다. 이런 현상과 더불어 배출가스 양에 관련된 정보는 해당 차량의 Recall과 직접적인 관련이 있어, 발표된 연구 결과는 거의 없는 실정이다.<sup>6)</sup>

따라서, 본 연구에서는 선형 회귀 분석 방법을 사용하여 주행거리별 운행차 배출가스 양의 분포를 예측하는 일반적인 모델을 제안하고, 이를 통하여 운행차의 배출가스 규제 만족 여부를 판단하는 방법을 제시하고자 한다.

## 2. 배출가스 분포 추정 모델의 설정

본 연구에서 주행거리별 운행차 배출가스 양의 분포를 예측하는 선형 회귀 분석 방법을 제안하기 위하여 예로써 사용하는 운행차 배출가스 자료는 미국 EPA의 1999년 Recall Data Base에서 인용한 것으로, 주행거리가 3만~8만km인 차량들의 FTP-75시험결과이며, 대상차량들은 3.0 liter 전자제어엔진을 탑재한 1988~1990년식 자동변속기 차량으로 삼원 촉매를 장착하였으며 THC 0.25g/km @8만km의 규제를 적용 받는다.<sup>7)</sup>

Fig.1에 주행거리별 HC, CO, NOx의 분포를 보인다. 여기에서 Normalized Mileage(이하에서 N.M이라 표기)는 차량의 주행거리를 배출가스 보증거리인 80,000km로 나눈 값을 말한다.

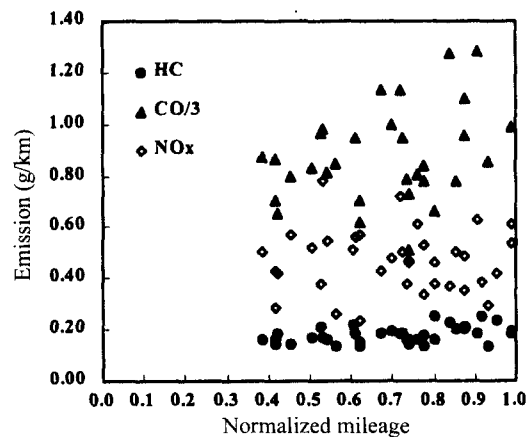


Fig. 1 Tailpipe emission distribution

### 2.1 회귀 직선의 요구 조건

많은 경우에 있어서 한 변수의 변화로부터 다른 변수의 변화를 예측하는 방법으로 선형 회귀 분석 방법을 사용하고 있다. 선형 회귀 분석은 측정된 변수들 사이에 존재하는 것으로 예상되는 직선적인 연관성을 규명하는 통계적 방법으로, 그 결과는 다음과 같은 형태를 가진다.

$$y = \beta_0 + \beta_1 x + \varepsilon \quad (1)$$

- $\beta_0$  : 회귀 직선의 절편
- $\beta_1$  : 회귀 직선의 기울기
- $\varepsilon$  : 오차

이때, 오차(Error)는 임의의 x에 대하여

- 오차가 정규분포를 이룰 것 (정규성)
- 오차가 등분산을 가질 것 (등분산성)

이라는 조건을 만족시켜야 한다. 이것은  $\hat{\varepsilon} \sim N(0, \sigma^2)$ 으로 표현됨을 말한다.

이들 조건들을 만족시키지 못하면, 얻어진 회귀 직선은 통계적 유의성을 갖지 못한다. 본 연구에서 인용한 자료에 있어서도, 자료 그대로 선형 회귀 분석을 하면, 얻어진 회귀 직선은 통계적 유의성을 가지지 못하여, 배출가스 양과 주행거리 사이의 연관성을 밝히는 데에 아무런 도움이 되지 못했다. 따라서, 주행거리와 배출가스 양 사이의 연관성을 밝히기 위해서는 위 조건들이 성립할 수 있는 모델을 설정할 필요가 있다.

한편, 회귀 직선식의 질적 수준을 높이기 위해서는 사전에 어긋난 값(Outlier)을 제거하는 단계가 필요하다. 어긋난 값은 자료의 표준편차가  $\pm 2.5$  이상인 경우로 정의한다. 본 연구에서는 어긋난 값으로 3개의 자료를 제거해서 총 36개의 자료로 회귀분석을 하였다.

### 2.2 정규성 검토

보통 자동차의 배출가스 양은 엔진이나 제어 시스템, 그리고 부품의 편차로 인해, 각 차량마다 서로 차이를 보이는데, 이렇게 차이를 보이는 배출가스 양의 분포에 대한 확정된 모델은 아직 없

다. 또한, 주행거리별 배출가스 양의 분포를 예측하기 위하여 설정할 회귀 직선이 오차의 정규성 조건을 확보하기 위해서는 정규분포로 가정할 수 있는 종속변수의 설정이 필요하다.

자동차의 배출가스 양은 삼원촉매의 특성에 의해 직접적인 영향을 받는다. L. S. Socha, Jr. 등에 따르면 배출가스 양은 다음과 같이 삼원촉매의 활성표면적의 지수함수로 표현할 수 있다.<sup>8)</sup> 여기서 A, B, c는 상수이고, X는 활성표면적이다.

$$Emission = A + B \exp(-cX) \quad (2)$$

따라서,

$$X = f\{\ln(Emission)\} \quad (3)$$

활성표면적은 생산과정 중에 그 품질이 관리되고 있는 것으로 정규분포에 가깝다고 볼 수 있다. 또한, 대부분의 차량이 운전되는 패턴에는 일정한 경향이 있어,<sup>9)</sup> 일정한 주행거리를 경과하는 동안에 촉매에 가해지는 열충격은 대체로 일정할 것으로 예상되므로, 차량의 주행거리가 증가함에 따라 전반적으로 활성표면적이 감소할 지라도 활성표면적의 정규분포성은 유지된다고 가정할 수 있다. 따라서,  $\ln(Emission)$ 의 분포 또한 정규분포를 따른다고 할 수 있다.

$\ln(Emission)$ 의 분포가 정규분포에 더 근접함을 보이기 위해 좁은 범위의 주행거리 내에 비교적 많은 자료가 있는  $0.7 < N.M. < 0.8$ 의 자료를 선택하고, 이들의 잔차를 구한 다음, 잔차의 표준편차를 살펴보았다. Fig. 2는 HC와  $\ln(HC)$ 를 종속변수로 했을 때, 잔차들의 표준편차를 비교한 것이다. 여기서, 잔차들의 표준편차가 정규분포선에 가깝게 위치할수록 정규성이 양호하다는 것을 의미하는데,  $\ln(HC)$ 로 하는 경우가 정규분포에 더 가까운 것을 확인할 수 있다.

정규분포성은 임의의 주행거리에서 만족해야 하므로, 전체 자료를 주행거리에 따라 4개의 그룹으로 나누어, 각각의 그룹에서 자료의 분포를 정규분포로 가정할 수 있는가를 검토할 필요가 있다. 따라서, Kolmogorov-Smirnov Test<sup>10)</sup> 방법을

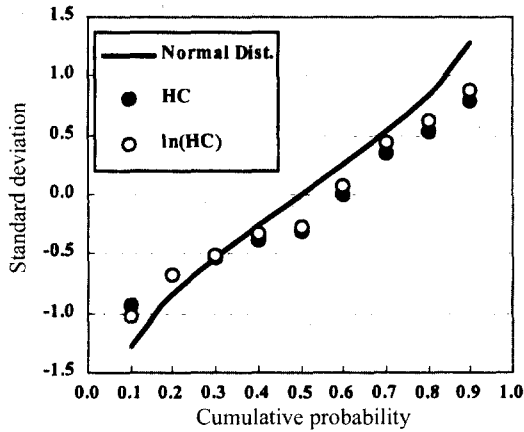


Fig. 2 Normality of residual

사용하여 각 그룹의 정규성을 확인하였다. 검토 결과 정규성에 대한 가정을 95% 신뢰수준으로 채택할 수 있었다.

그러므로, 본 연구에서는 회귀 직선의 정규성 만족도를 높이기 위해서 우선 Emission값 대신에 ln(Emission)을 종속변수로 선정한다. 이하에서는 ln(Emission)을 ln(EM)으로 표현한다.

### 2.3 등분산성 검토

오차의 등분산성 조건은 각각의 자료들과 회귀 직선에 의한 추정치의 차이인 잔차(Residual)들의 분산을 검토함으로써 만족 여부를 판단할 수 있다. 등분산성이 모든 주행거리에서 만족해야 하므로, ln(EM)을 주행거리에 따라 4개의 그룹으로 나누고, 각 그룹에서 구한 표준잔차가 전체적으로 일정하다고 가정할 수 있는지를 Hartley F-max Test<sup>11)</sup> 방법으로 검토하였는데, 그 결과 ln(CO)의 경우에 95% 신뢰수준에서 등분산 가정이 기각되었다. 등분산성이 만족되지 못한 것은 운행차의 주행거리에 따른 자료의 산포도가 일정하지 않다는 것을 의미하는 것이다.

본 연구에서는 잔차가 주행거리에 비례하여 커지는 특성을 이용하여, 독립변수로 N.M. 대신에 1/N.M.를, 그리고 종속변수는 ln(EM) 대신에 ln(EM)/N.M.으로 하는 방법을 사용하여 등분산

성 조건을 만족시키고자 하였다. 이 경우, 회귀 직선은 다음과 같은 형태를 가진다.

$$\frac{\ln(EM)}{N.M.} = \frac{a}{N.M.} + b \quad (4)$$

Fig. 3은 종속변수를 각각 ln(EM)과 ln(EM)/N.M.으로 하였을 때의 표준잔차를 비교한 것이다. 여기에서 보면, ln(HC)/N.M.의 경우에, 주행거리가 작을 때는 ln(HC)에 비하여 표준잔차가 커지고, 반대로 주행거리가 커질 때는 표준잔차가 상대적으로 작아져서, 주행거리 전체에 걸쳐서 표준잔차가 일정한 범위 안에 있는 것을 확인할 수 있다.

이를 바탕으로, 종속변수를 ln(EM)/N.M.으로 하여 Hartley F-max Test방법으로 등분산성을 검토한 결과, 95% 신뢰수준에서 등분산 가정을 채택할 수 있었다.

이상의 방법으로 구한 각 배출가스 HC, CO, 그리고 NOx에 대한 회귀 직선의 특성치를 Table 1에 보인다. 여기에서 R<sup>2</sup>은 회귀 직선의 결정계수

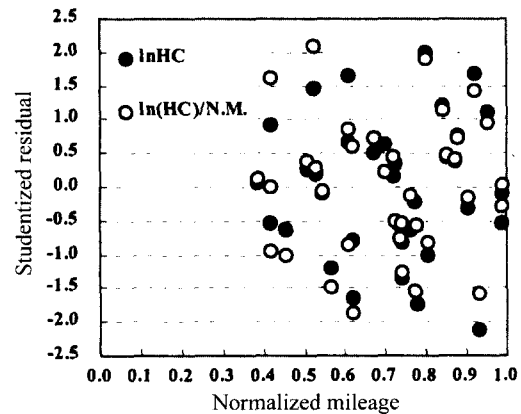


Fig. 3 Distribution of studentized residuals

Table 1 Characteristic values of linear regression

	a	b	R <sup>2</sup>	σ <sup>2</sup>
HC	-2.0078	0.3666	0.9445	0.2192
CO	0.6722	0.4293	0.4381	0.3429
NOx	-0.8661	0.1096	0.4292	0.4498

이고,  $\sigma$ 는 추정의 표준오차이다. CO나 NOx의 결정계수가 HC의 결정계수에 비하여 작은 값이나, 종속변수를  $\ln(EM)/N.M.$ 으로 변형하기 전에 비해서는 매우 큰 값이며, 또한 구해진 회귀 직선의 기울기는 유의수준 1%에서 매우 유의하다.

### 3. 주행거리별 배출가스 분포 추정

#### 3.1 확률 밀도

주어진 자료를 이용하여 회귀 직선을 구할 때, 독립변수  $x$ 에 관계없이 오차는  $N(0, \sigma^2)$ 인 확률 분포를 따른다는 가정이 채택되도록 자료를 설정, 변형하는 과정을 보였다. 따라서 모든  $x$ 의 값에 대하여 종속변수  $y$ 의 기대값은 회귀 직선식에서 얻어지는 값이며, 분산은  $\sigma^2$ 라고 생각할 수 있다. 그러므로, 임의의 주행거리에서 배출가스 양  $\ln\{(EM)/N.M.\}$ 은  $N\{(a/N.M.)+b, \sigma^2\}$ 인 확률 분포를 가진다고 할 수 있다.

주행거리가 50,000km, 65,000km, 그리고 배출가스 보증거리인 80,000km일 때의 HC, CO, NOx의 확률 밀도(Probability Density)를 Fig. 4 ~ Fig. 6에 보인다. 여기에서 보면 주행거리가 길어질수록 배출가스 양이 최빈값 이상인 범위에서 더 넓게 분포함을 확인할 수 있으며, Table 1에서 보이는  $\sigma^2$ 의 크기가 반영되어  $HC < CO < NOx$ 의 순서로 산포도가 큼을 알 수 있다. 이는 기존에 알려진 삼원촉매의 열화 특성에 부합되는 것이며,<sup>12)</sup> 이런 산포도의 순서를 보이는 이유는 크게 두 가지로 설명될 수 있다.

- 혼합비 제어의 정밀도가 떨어질 때, NOx나 CO의 촉매정확효율이 더 큰 폭으로 떨어진다.
- 고온에 의한 삼원촉매의 정확성능 저하는 NOx나 CO에서 더 크게 나타난다.

#### 3.2 누적분포

앞에서 구한 확률밀도를 누적하여 누적분포(Cumulative Distribution)를 구하였다. 주행거리 50,000km, 65,000km, 80,000km에서의 각 배출가스의 누적분포를 Fig. 7~ Fig. 9에 보인다. 여기에

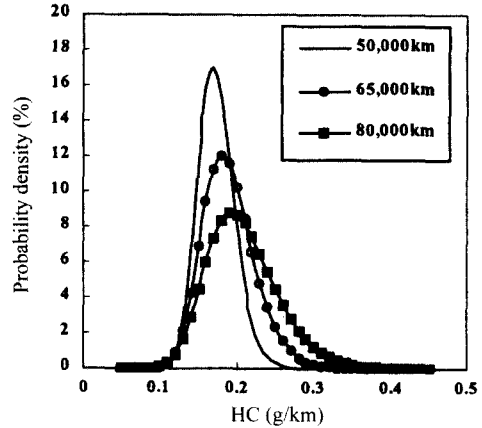


Fig. 4 Probability density of HC

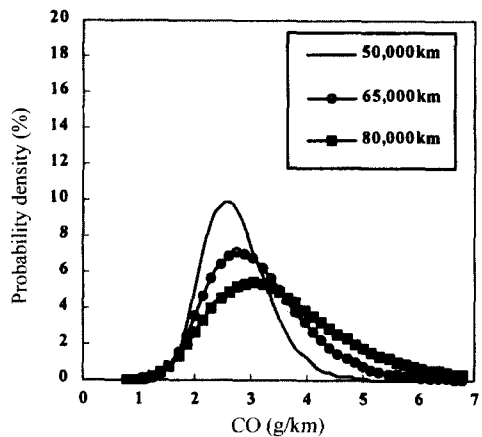


Fig. 5 Probability density of CO

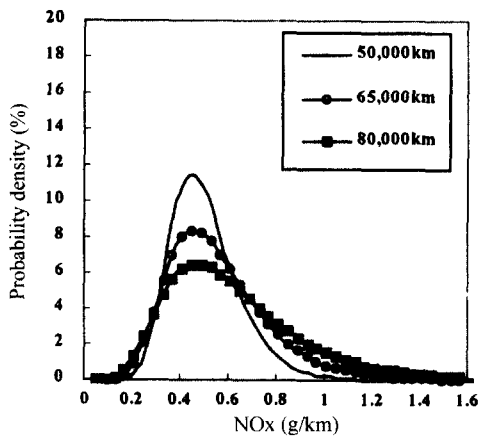


Fig. 6 Probability density of NOx

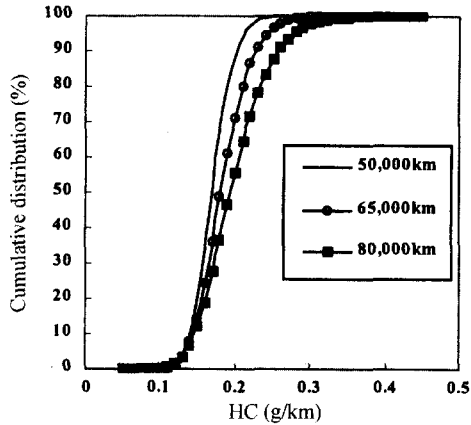


Fig. 7 Cumulative distribution of HC

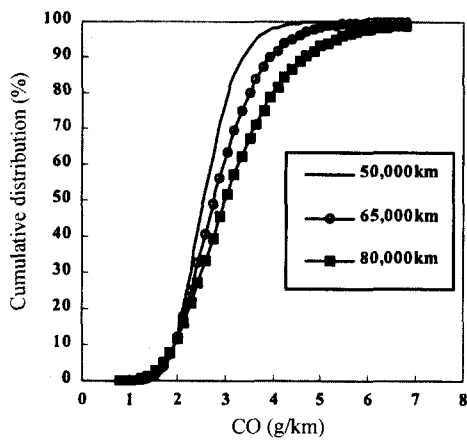


Fig. 8 Cumulative distribution of CO

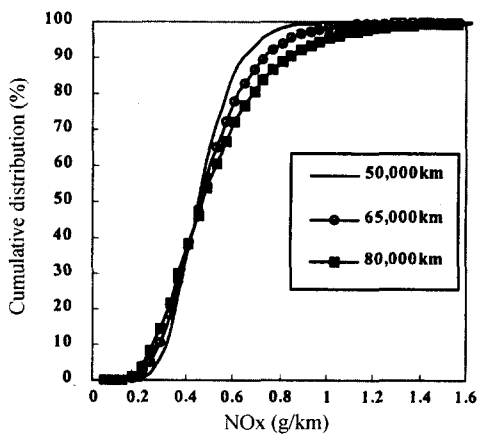


Fig. 9 Cumulative distribution of NOx

서 배출가스별 규제치에 해당하는 누적분포를 구하면, 각 배출가스별로 규제치를 만족할 것으로 예상되는 차량의 비율을 알 수 있는데, 80,000km에서 HC는 88%, CO는 15%, NOx는 73%를 보였다. CO가 HC와 NOx에 비해서 비정상적으로 낮은 값을 보이고 있으며, 또한, 주행거리가 약 15,000km일 때, CO의 평균값이 규제치에 근접하는 것으로 추정된다. 이것은 배출가스 보증거리 이내에서도 대부분의 차량이 규제치보다 더 많은 CO를 배출하고 있다는 것을 의미하며, 따라서 배출가스 Recall의 대상이 될 것으로 예상할 수 있다. 실제로 본 연구에서 인용한 차량은 1994~1995년에 걸쳐 Recall을 시행하였다.<sup>13,14)</sup>

### 3.3 열화계수와 배출가스 규제 만족 판정

구해진 회귀 직선에 의한 6,400km에서의 배출가스 추정값과 80,000km에서의 추정값 사이의 비율 열화계수(Deterioration Factor)라고 하는데, 이는 배출가스 보증거리를 주행하는 동안 배출가스 양이 초기에 비해서 얼마나 악화되었는가를 나타내는 지표이다. 본 연구에서 인용한 자료의 열화계수 계산결과를 Table 2에 보인다. 여기에서 보면, 운행차의 추정 열화계수는 인용 차량의 인증열화계수인 1.000(HC), 1.000(CO), 1.109(NOx)에 비해서 HC와 CO에서 높은 값을 보이고 있다.<sup>15)</sup> 이것은 실제의 차량 운행조건이 인증 조건에 비해서 HC와 CO에 있어서 더 가혹한 조건임을 말하는 것이다.

한편, 운행차들의 배출가스 양은 어떤 산포도를 가지고 분포하고 있기 때문에, 80,000km에서의 배출가스 양이 규제치 이하인 차량의 비율로

Table 2 Predicted deterioration factor

		HC	CO	NOx
Predicted mean (g/km)	6,400km	0.14	2.03	0.42
	80,000km	0.19	3.01	0.47
Deterioration factor		1.401	1.484	1.106

써 배출가스 규제에 대한 만족 여부를 판정해야 한다. 본 연구에서는 80,000km에서의 배출가스 누적분포에서 특정 비율에 해당하는 배출가스 양을 판정기준(Threshold)으로 하여, 그것과 규제치를 비교함으로써 규제 만족 여부를 판정하였다. 판정기준을 결정짓는 누적분포 비율은 EPA의 Line-crossing 규정을 보완하여 60%로 하였다. EPA의 Line-crossing 규정은 회귀 직선에 의한 80,000km에서의 추정값(누적분포 50%에 해당)이 규제치 이하일 것을 요구하는 것이지만,<sup>16)</sup> 본 연구에서는 여기에 10% 더 엄격하게 적용하여 60%로 하였다. Table 3에 배출가스 규제 만족 여부를 정리한다.

Table 3 Decision results

	HC	CO	NOx
Emission standard [g/km]	0.25	2.1	0.62
Threshold [g/km]	0.21	3.3	0.53
Decision	Pass	Fail	Pass

#### 4. 결론

정확한 운행차의 배출가스 분포를 알기 위해서 많은 대수의 차량시험 결과를 수집하는 것은 현실적으로 불가능하다. 본 연구에서는 소수의 운행차 시험 결과로 임의의 주행거리에서의 배출가스 양의 분포를 추정하는 방법을 제안하였다. 이 결과를 종합하면 다음과 같다.

1) 독립변수를 1/N.M.로, 그리고 종속변수를  $\ln(EM)/N.M.$ 으로 설정함으로써 오차의 정규성과 등분산성을 확보하는 회귀 직선 분석 모델을 설정할 수 있었다.

2) 인용한 운행차의 배출가스별 산포도는  $HC < CO < NOx$ 의 순서로 나타났으며, 이는 삼원촉매의 열화특성에 부합되는 결과이다.

3) 구해진 회귀 직선에 의한 운행차의 열화계수는 HC, CO, NOx가 각각 1.401, 1.484, 1.106으로 계산되었다.

4) 80,000km에서의 60% 누적비율에 해당하는

배출가스 양과 규제치를 비교하는 방법으로 규제 만족 여부를 판정하였을 때, 인용 자료의 경우에 CO가 규제를 불만족하였다.

#### 후 기

이 논문은 2001년도 두뇌한국21사업에 의하여 지원되었음.

#### 참 고 문 헌

- 1) 환경부, 자동차 배출가스 종합대책, p.107, 1995.
- 2) EPA, Regulations Requiring On-Board Diagnosis: Federal Register, 1995.
- 3) EPA, Compliance Programs for New Light-Duty Vehicles and Light-Duty Trucks: Federal Register, 1999.
- 4) EPA, Determination of Running Emissions as a Function of Mileage for 1981-1983 Model Year Light-Duty Cars and Trucks: Draft, 1999.
- 5) EPA, Regulatory Impact Analysis Control of Air Pollution from New Motor Vehicles : Tier 2 Motor Vehicle Emissions Standards and Gasoline Sulfur Control Requirements, 1999.
- 6) H. W. Haskew, T. F. Liberty, In-Use Emissions with Today's Closed-Loop Systems, SAE 910339, 1991.
- 7) EPA, Recall Data Base (EPA420-B-99-003), 1999.
- 8) J. P. Day, L. S. Socha, Jr., Technique for the Analysis of FTP Emissions, SAE 920724, 1992.
- 9) S. Kishan, T. H. DeFries, C. G. Weyn, A Study of Light-Duty Vehicle Driving Behavior: Application to Real-World Emission Inventories, SAE 932659, 1993.
- 10) S. P. Gordon, F. S. Gordon, Contemporary Statistics, McGraw-Hill, pp.558-565, 1994.
- 11) R. J. Freund, W. J. Wilson, Statistical Methods, Academic Press, pp.231-232, 1997.
- 12) D. D. Beck, M. H. Krueger, D. R. Monroe, D. J. Upton, J. M. Lendway, D. R. Smith, Improvement of Laboratory Tests for Eval-

- uating Three-Way Catalyst Activity , SAE 920099, 1992.
- 13) EPA, Defeat Recall (92-09/DR-898), 1994.
- 14) EPA, Defeat Recall (92-39/DR-942), 1995.
- 15) EPA, 1988MY Certified Test Result Report, 1988.
- 16) EPA, Compliance with Emission Standards, 40 CFR 86.094-28.