

다변량 모형을 이용한 보증데이터 분석 방법 연구

김종걸* · 성기우*

*성균관대학교 산업공학과

A Study on Analysis Method of Warranty Data Using Multivariate Model

Jong-Gurl Kim* · Ki-Woo Sung*

*Dept. of Industrial Engineering, Sungkyunkwan University

Abstract

The purpose of the warranty data analysis can be classified into two categories. Two goals is a failure cause analysis and life prediction analysis. In this paper first, we applied multivariate analysis method that can be estimated in consideration of various factors on the failure cause warranty data. In particular, we apply the Tree model and Cox model. The advantage of the Tree is easy to interpret this result as compared to other models. In addition Cox model can quantitatively express the risk.

Second, this paper proposed a multivariate life prediction model (AFT) considering a variety of factors. By applying the actual warranty data confirmed the usability.

Keywords : Warranty Data, Accelerated failure time, AFT model, Tree model, Cox model

1. 서론

제품 개발에 있어서 우선적으로 선행되어야 하는 것은 해당 부품 또는 제품의 목표 수명을 정하는 것이다. 부품의 목표수명에 따라서 제품 설계를 하고 통제된 환경에서 짧은 시간 동안 제품의 수명을 파악할 수 있는 가속수명시험을 통해 그 부품 또는 제품의 수명을 평가한다. 그러나 현실적으로 단품 단계에서의 시험은 실제 사용 환경을 잘 반영하지 못하며, 필드 고장 모드의 재현율이 떨어진다. 이를 극복하고 필드에서의 제품(단품)의 수명을 파악하기 위해 부단히 노력하고 있다.

필드 수명을 파악 할 수 있는 방법은 크게 두 가지로 볼 수 있다. 첫 번째 방법은 특정 부품의 주행거리별 또는 사용기간별 부품의 특성치를 측정하여 잔존수명을 파악하는 방법과 다른 한가지 방법은 보증데이터를 활용하는 방법이다. 보증데이터는 보증기간 동안 서비스 센터에 접수된 클레임으로부터 얻어진 필드 고객

데이터로 고객의 실제 사용 환경에서 제품 고장을 파악하고, 수명을 확인하기에 효과적인 데이터이다.

보증이란 제품이 현재, 혹은 미래에 의도한 대로 올바르게 작동함을 제조업자가 소비자에게 보장하는 것을 의미한다. 좁은 의미로 보증은 소비자에 대하여 결함이 있는 제품을 비용 없이 혹은 할인된 비용으로 수리, 교환해 준다는 보장을 의미한다. 어느 품목이 보증기간 내에 고장이 났을 경우, 소비자는 제조업자 혹은 판매자에게 접촉하여 보증 이익에 대하여 결함이 있는 제품의 수리 혹은 교환에 대한 클레임을 제출할 수 있다. 보증 클레임은 제품의 실제 현장에서의 작동과 신뢰성 특성들에 대한 정보를 제공한다.

보증데이터 분석 목적은 크게 세 가지로 분류할 수 있다. 첫째 현상 파악이다. 현상파악은 각 부품별 보증 클레임데이터를 이용하여 각 부품의 현 수준을 분석하는 방법으로 다변량 분석 방법이다.

† Corresponding Author : Jong Gurl Kim, Sungkyunkwan University, Industrial Engineering, Suwon, Gyeonggi-do, E-mail: jgkim@skku.ed

Received April 19, 2015; Revision Received June 18, 2015; Accepted June 20, 2015.

즉 부품의 현 수명(신뢰도)를 분석할 수 있는 생명표법, 카렌마이어 방법이 대표적이라고 할 수 있다. 둘째, 고장 원인 분석이다. 부품의 고장에 생산부터 고객의 사용조건까지 다양한 인자가 영향을 줄 것이다. 원인분석은 현상분석처럼 단순히 사용 시간을 가지고 분석하는 것이 아닌 다양한 원인변수를 통해서 원인을 파악하는 다변량 방법이다. 즉 단변량 분석이 아닌 좀 다양한 인자를 고려하여 분석하는 것이 차이점 이다. 본 연구에서는 원인 분석에 다변량 방법 중 Tree기법과 Cox모형을 적용하고자 한다. 두 방법을 제시하는 이유는 이 두 방법의 장단점을 이용하여 최적을 결과를 얻기 위함이다. Tree의 장점은 결과 해석이 다른 모형에 비해 쉽다. 단점은 유의한 변수가 무엇인지는 알 수 있으나 정량적으로 표현하기 어렵다. 반면에 Cox 모형의 경우는 위험도를 정량적으로 표현할 수 있다. 즉 다양한 인자의 기여도를 비교적 쉽게 찾을 수 있다. 보증데이터는 다양한 고객 정보와 고장에 대한 정보를 가지고 있으며 이를 바탕으로 원인분석이 가능하다. 셋째, 단변량이 아닌 다변량 수명예측 방법으로 그 모형의 구조에 따라 비례적 위험함수 모형 (proportional hazards)의 가정을 적용하는 Cox 모형과 가속화 시간 (accelerated failure time)을 적용하는 AFT 모형으로 분류할 수 있다. 수명예측에서는 예측에 적합한 AFT모형을 통해서 생존 시간 자체에 대한 설명변수의 효과를 모형화하고 각 부품의 수명을 예측하고자 한다. 이 방법을 제시하는 이유는 Cox모형의 경우 준모수적 방법으로 위험함수의 모수적 유형을 지정하지 않아도 된다는 장점을 가지고 있지만, 위험 함수 보다는 설명변수의 효과를 추정에 그 주목적이 있기 때문에 모형의 결과를 수명예측에 사용하기에는 부적합하다.

본 논문에서는 다양한 인자를 고려하여 고장 원인을 추정할 수 있는 다변량 분석방법을 보증데이터에 적용해 보았고, 실제 유의함을 확인하였다. 특히 Tree 모형 및 Cox모형을 통해서 서로의 장단점을 보완하였다. 또한 부품의 수명 및 수명에 대한 설명변수의 효과를 모형화 하여 보증데이터를 분석하는 AFT 모형을 제시하였고 실제 보증데이터를 통해서 AFT 모형의 활용성을 확인하였다.

2. 본 론

2.1 보증데이터란

보증데이터란 부품의 수리 보증비용이나 현장에서의 신뢰도를 산출하기 위한 목적의 필드 서비스 데이터로

서 차량 판매 후 보증기간 내에 결함이 발생하면 해당 부품을 무상으로 수리 또는 교환을 하고, 이 결함이 발견된 차량으로부터 얻어진 고장관련 데이터를 말한다. 특징은 사용기간, 주행거리뿐만 아니라 고장에 대한 정보를 알 수 있으며, 고객의 정보도 알 수 있다.

2.2 보증데이터 분석방법 연구 동향

보증데이터 분석 시 통계적 분석 방법에 따라 모수적 분석방법과 비모수적 분석방법으로 나눌 수 있으며, 더 세부적으로는 주행거리 또는 사용 기간을 이용한 분석방법으로 구분 할 수 있다. 또한 단변량 방법과 다변량 방법으로도 구분할 수 있다. 단변량 분석은 시간에 대한 생존을 분석으로 생명표법과 Kaplan-Meier방법이 대표적이다. 다변량 방법으로는 Semi parametric survival analysis으로는 Cox propotional hazards 모형, Parametric survival analysis 으로는 Exponential regression survival 모형 , Weibull distribution 모형, normal linear regression 모형, log-normal linear regression, AFT model 등이 있다.

<Table1> Classification of the analysis method

Parametric models	Nonparametric model	Semi-Parametric models
Exponential distribution Weibull distribution Log-normal distribution Gamma and generalized gamma Gompertz distribution Log-logistic Linear exponential AFT model	Life table method (=Actual method) Kaplan-Meier method (=Product limit method) Nelson-Aalen method (=Nelson and Fleming-Harrington)	Cox proportion al hazard model

원인분석을 위한 다변량 분석방법 동향은 다음과 같다. 다변량 인자의 원인분석 방법은 일반적으로 다양한 요인을 나타내는 회귀 모델을 사용한다. 이런 경우 로그선형모형을 많이 사용하며 편리하다. Kalbfleisch와 Lawless[1]는 현장 데이터 수집을 위한 절차를 제안하였고 공변량에 관련한 추가적인 정보와 함께 필드 고장 데이터로 부터 수명 분포를 추정하기 위한 회귀 모델을 사용했다. Hu와 Lawless[2]는 공변량에 대한 추가적인 정보 와 관측중단을 통한 추정 절차를 개발하였다. 또한 불완전한 정보로써의 보증 모델링을 위한 기술을 제안했다. 이때 클레임을 Poisson 프로세스를 따른다고 가정했다. Hu와 Lawless[3]는 불완전한 응답 변수와 공변량을 포함하는 상황을 고려했다. 그들은

의사우도추정의 두 가지 유형을 논의했고 수명분포의 비모수적 추정을 위한 방법을 제공했다. Attardi[4] 등은 자동차의 보증 자료들의 분석을 위해 혼합 와이블 회귀모델을 이용했다. 엔진타입과 자동차 모델은 회귀 모형에서 공변량으로 사용된다. Karim 과 Suzuki는 [5] 신뢰성 관련 요인과 연관된 공변량을 고려했고 공변량의 함수로써 구성 요소의 수명을 위한 와이블 회귀분석 모델을 제시했다. 제품이 운영되는 지역은 Vinta[6]와 Hrycej, Grabert[6,7]에 의해 영향 요인으로 고려되었다. Thomas 와 William [8]은 필드보증 데이터를 비례위험모형으로 분석하였고 에어컨 컴프레서를 사례로 들었다. Sang Hyun Lee[9]는 신경망모형을 통해서 보증데이터의 조기 경보 및 발견할 수 있는 방법을 소개하였다. 본 논문에서 제시하는 AFT모형의 경우는 자동차 보증데이터 클레임에 적용한 사례는 없었다.

2.3 보증데이터 원인분석 방법

2.3.1 의사결정나무

의사결정나무(decision tree)는 의사결정규칙(decision rule)을 나무구조(tree structure)로 도표화하여 분류(classification)와 예측(prediction)을 수행하는 분석 방법이다. 의사결정나무의 장점은 첫째 해석의 용이성이다. 모형의 이해가 쉽고, 새로운 자료의 모형에 적합하며, 어떤 입력변수가 목표변수를 설명하기에 좋은지 쉽게 파악할 수 있다. 둘째 교호효과 해석이 가능하다. 두 개 이상의 변수가 결합하여 목표변수에 어떻게 영향을 주는지 쉽게 알 수 있다. 셋째, 비모수적 모형으로 선형성, 정규성, 등분산성 등의 가정이 필요 없다. 의사결정나무의 단점으로는 분석용 자료에만 의존하기 때문에 새로운 자료의 예측에서는 불안정할 가능성이 높다. 의사결정 나무의 형성 과정은 다음과 같다. 첫 번째, 의사결정 나무의 형성 두 번째, 가지치기 세 번째, 타당성 평가 네 번째, 해석 및 예측 이다. 순수도와 분리 기준은 목표변수의 분포를 가장 잘 구별해주는 입력변수를 찾기 위해, 목표변수의 분포를 구별하는 정도를 순수도(purity) 또는 불순도(impurity)에 의해서 측정하고, 자식 마디의 순수도를 가장 높이는 변수를 분리 기준으로 한다.

분리 기준은 하나의 부모 마디로부터 자식 마디들이 형성될 때, 입력 변수의 선택과 범주의 병합이 이루어질 기준을 의미한다. 순수도는 목표 변수의 특정 범주에 객체들이 포함되어 있는 정도를 의미한다.

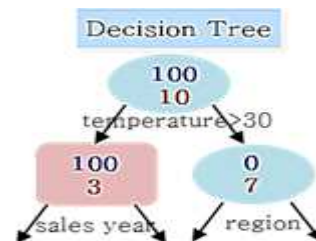
이산형 목표변수에 사용되는 분리 기준은 다음과 같

다. 카이제곱 통계량은 p-값이 가장 작은 예측변수와 그 때의 최적 분리에 의해서 자식 마디를 형성한다. 지니 지수는 불순도를 측정하는 하나의 지수로서 지니 지수를 가장 감소시켜 주는 예측변수와 그때의 최적 분리에 의해서 자식마디를 선택한다. 엔트로피 지수는 다항 분포에서의 우도비 검정통계량을 사용하는 것과 같은 것으로 알려져 있다. 이 지수가 가장 작은 예측 변수와 그 때의 최적분리에 의해서 자식마디를 형성한다.

연속형 목표 변수에 사용되는 분리 기준은 다음과 같다. 분산분석에서의 F 통계량은 p-값이 가장 작은 예측 변수와 그 때의 최적분리에 의해서 자식마디가 형성, 가장 작은 p-값을 갖는 변수를 선택한다는 점에서 카이제곱과 F통계량의 p-값을 이용하는 방법은 같다. 분산의 감소량은 예측 오차를 최소화하는 것과 동일한 기준으로 분산의 감소량을 최대화하는 기준의 최적분리에 의해서 자식마디가 형성된다.

정지 규칙은 정지 규칙을 더 이상 분리가 일어나지 않고 현재의 마디가 끝마디가 되도록 하는 여러 가지 규칙을 의미한다.

가지치기는 지나치게 많은 마디를 가지는 의사결정나무는 새로운 자료에 적용할 때 예측 오차가 매우 클 가능성이 있다. 따라서 형성된 의사결정 나무에서 적절하지 않은 마디를 제거하여, 적당한 크기를 갖는 부나무(subtree)구조의 의사결정나무로 최종적인 예측 모형을 선택하는 것이 바람직하다.



[Figure 1] Tree model example

2.3.2 Cox 모형

1992년 Cox가 제안한 모형으로 특별한 생존 시간의 분포에 가정이 없고 상대위험도를 계산 하여준다. 실제로 생존시간에 대한 분포를 알 수 없으므로 적당한 모수적 모형을 찾기가 어려워서 모수적 모형으로는 생존시간(신뢰도)에 영향을 미치는 인자를 찾아내기가 어렵다. 결과 변수의 relative hazards에 자연로그를 취한 값 (natural log of relative hazard)을 가지고 모형을 세운다.

$$\lambda(t) = \exp(\beta\chi) \times \lambda_0(t) \quad (1)$$

$\lambda(t)$: 예후 인자를 가진 사람의 관찰 시간에서의 사망력(hazard)

$\lambda_0(t)$: 예후 인자를 모두 가지지 않은 사람의 관찰 시간에서의 사망력(baseline hazard)

$\beta\chi$: 예후 인자의 회귀계수

Cox 모형의 특성은 다음과 같다. 첫째, 생존에 영향을 주는 요인들에 대한 회귀분석이며, 둘째, 생존함수의 분포에 대한 가정 없이 추정하여 편리하다. 또한 추정된 계수는 상대적 위험도(relative hazard)를 나타낸다. 셋째, 공변수의 효과가 시간에 관계없이 일정하다는 가정 아래 수행되며 의학 연구에서 가장 널리 쓰이는 방법이다. Cox 모형에 투입할 변수의 개수, 종류, 변수 선택방법은 범주형, 연속형 변수 모두 가능하며 독립변수의 성질이 연속형이라도 비연속적인 범주형의 형태로 바꾸어서 입력한다. 또한 분류 후 골고루 분류되었는지 확인 필요 하다. 투입될 변수의 개수는 사건 발생한 수의 네 제곱근보다 많은 수의 변수는 안 된다. 예를 들면 $402(1/4) = 4.47$ 즉, 4개변수를 넘으면 안 된다. 변수 투입 방법은 Forward selection method 과 Backward selection method 방법 2가지가 있다. Cox 모형의 가정은 다음과 같다. 첫째, 비례위험 모형의 가정 둘째, Hazard function과 covariate 사이에는 서로 log-linear 관계가 있다는 가정 셋째, 모형의 적합성(goodness of fit) 검정을 통해서 만약 가정 위반 시 비례위험모형은 사용할 수 없으며 non-proportional hazard 모형 사용해야 한다.

2.3.3 AFT(Accelerated Failure Time)모형

모수적 생존분석 기법인 AFT 모형은 공변량과 생존시간 간의 비례 관계를 다음과 같이 가정한다.

$$T = \exp(\beta\chi) \cdot T_0 \quad (2)$$

식(2)에서 독립변수인 공변량 값이 0인 경우의 생존시간 확률변수를 T_0 라 하고, 사전적으로 어떤 기저분포(baseline distribution)를 갖는 것으로 가정한다.

흔히 사용되는 기저분포로는 와이블 분포, 지수분포, 극치 분포, 로그-정규분포 등이 있다. 식 (2)에서 생존시간에 영향을 미치는 공변량이 n개 존재하고, 척도모수(scale parameter)가 포함되면 다음과 같은 식 (3) 일반적인 AFT 모형이 구축된다.

$$T = \exp(\alpha + \beta_1\chi_1 + \beta_2\chi_2 + \dots + \beta_n\chi_n) \cdot T_0^\sigma \quad (3)$$

그리고 식 (3)의 양변에 로그를 취하면 식 (4)와 같은 최종 회귀식을 얻게 된다.

$$\log T = \alpha + \beta_1\chi_1 + \beta_2\chi_2 + \dots + \beta_p\chi_p + \sigma\epsilon \quad (4)$$

여기서, $\sigma\epsilon = \log(T_0)$ 로 정의한다.

AFT 모형은 식 (4)에서 보는 바와 같이 생존시간과 공변량 간의 다중회귀식이므로 분석과 해석이 용이하나, 앞서 설명한 바와 같이 절단된 자료를 포함하는 생존시간에 대해 모수적 분포 가정을 해야 하는 한계가 있다. 본 연구에서는 가속화 고장시간 모형의 이와 같은 한계점을 충분히 숙지하고 사전에 분포에 대한 적합도 검정과 추정된 형태모수(shape parameter)의 크기를 고려하여 기저분포를 와이블 분포로 가정하여 진행한다.

2.4 기존 분석방법과 차이점

기존 보증데이터 분석은 단변량(시간)을 이용하여 주행거리를 이용한 생명표법으로 분석하여 최종적으로 각 부품의 필드 수명을 예측하였다. 즉 고장 난 시점에 대한 정보만을 가지고 분석을 하였고, 다른 다양한 환경 및 영향을 주는 인자에 대해서 고려하지 않았다.

2.4.1 Tree 및 Cox 모형을 이용한 원인분석 방법

본 논문에서는 다변량 방법 중 Tree기법과 Cox모형을 적용하고자 한다. 두 방법을 제시하는 이유는 이 두 방법의 장단점을 이용하여 최적의 결과를 얻기 위함이다. Tree의 장점은 결과 해석이 다른 모형에 비해 쉽다. 단점은 유의한 변수가 무엇인지는 알 수 있으나 정량적으로 표현하기 어렵다. Cox 모형의 경우는 위험도를 정량적으로 표현할 수 있다. 즉 다양한 인자의 기여도를 비교적 쉽게 찾을 수 있다.

2.4.2 AFT(Accelerated Failure Time)모형을 이용한 예측 방법

AFT(Accelerated Failure Time)모형을 이용한 보증데이터 분석 방법은 단순히 사용 시간을 가지고 분석하는 것이 아닌 다양한 원인변수를 통해서 원인을 파악하는 다변량 방법이다. 즉 부품 수명에 영향을 주는 여러 가지 인자를 고려하여 부품의 필드수명을 예측하고 수명에 영향을 주는 인자의 영향도를 파악할 수 있다. Cox 모형과의 차이점은 Cox모형은 위험함수

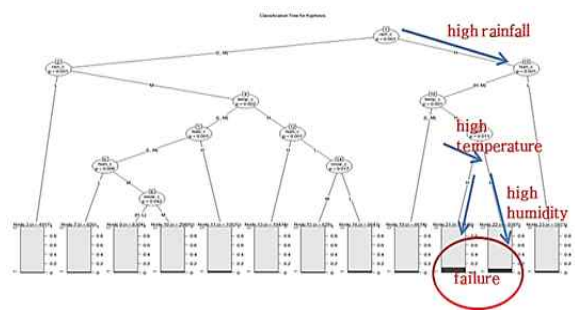
에 대한 설명변수의 효과를 모형화 하지만 AFT모형은 생존시간 자체에 대한 설명변수의 효과를 모형화 하기 때문에 AFT모형이 Cox모형보다 예측 목적에 적합하다. 따라서 본 연구에서는 필드 고장부품의 수명을 예측하고 설명변수의 수명 영향도를 알 수 있는 AFT모형을 통해 자동차 보증데이터를 분석하였다.

3. 분석사례

분석데이터는 연료도어 부품의 클레임 데이터를 분석하였다. 분석 툴은 오픈소스 R을 이용하여 분석하였다. 특정 차종과 생산년도에 부품을 선정하였으며 작동 불량량의 고장모드 분석결과이다. 총 1만 여개의 Raw 데이터를 분석하였다. 독립변수는 변수간의 관계에서 다른 변수에 영향을 주는 변수를 의미하며, 종속변수는 독립변수에 의해 그 영향이 결정되는 변수를 의미한다. 독립변수(입력변수)는 우선 주 별 온도, 적설량, 습도, 강설량 이다. 또한 보증데이터에 고장난 시점의 주행거리, 사용일 등의 정보를 종속변수로 사용하였다. 현재는 환경 데이터만을 통해 환경원인을 추정하였지만, 향후 과제로 주 별 다양한 날씨, 도로, 인구통계학적 데이터를 이용하여 좀 더 자세히 수명영향인자를 분석할 수 있을 것으로 기대한다. 원인분석에 영향을 주는 인자는 시간에 대한 정보뿐만 아니라 제조 및 제품사양, 사용 강도와 작동 조건들도 있다. 예를 들어 사용자의 개인 특성, 기후 조건 등도 제품 수명에 영향을 주는 인자이다.

3.1 Tree 및 Cox 모형을 이용한

원인분석결과 보증 클레임 데이터 분석시 주요한 부분은 데이터의 정의이다. 클레임 데이터의 부품 및 기간에 대한 정보가 정확하게 있어야 한다. 의사 결정나무란 의사결정 규칙을 나무 구조로 도표화하여 관심대상이 되는 집단을 몇 개의 소집단으로 분류 하거나 예측(Prediction)을 수행하는 분석 방법이다. 특히 미국 주 별 온도, 습도, 강수량, 강설량 등 다양한 환경 데이터를 이용하여 부품의 수명에 영향을 주는 환경 인자를 찾고자 한다. 입력 변수는 주 별 온도, 적설량, 습도, 강설량 이다. 미국 날씨 통계를 이용하여 주 별 환경적 특성을 정리하였다. 분석결과 Tree모형을 통해 문제 발생원인 유추 할 수 있었다. 각 노드를 보면 고온 다습 강수량이 많은 지역에서 문제가 다발됨을 Tree모형을 통해 확인할 수 있다.



[Figure 2] Tree model results

현재는 환경 데이터만을 통해 환경원인을 추정하였지만, 향후 과제로 주 별 다양한 날씨, 도로, 인구통계학적 데이터를 이용하여 좀 더 자세히 원인을 분석할 수 있을 것으로 기대한다.

두 번째로 Cox모형을 통해서 고장원인에 따른 위험도를 수치화를 해보았다. 모형을 확인 해본 결과 고온지역보다 저온지역이 18% 위험도 감소하였고, 강수량이 많은 지역보다 적은 지역이 86% 위험도 감소하였다. 또한 다습한 곳보다 저습한 지역이 15% 위험도 감소를 알 수 있었다. 식5은 분석된 Cox모형이다. 결과적으로 COX 모형을 통해 고장부품의 고장원인 및 위험도 수치화 가능 하였다.

<Table 2> Tree model results

$$h(t) = [h_0(t)]e^{(-0.2 \cdot \text{temp}(L) - 0.4 \cdot \text{temp}(M) - 0.16 \cdot \text{hum}(L) + 0.28 \cdot \text{hum}(M) - 1.98 \cdot \text{rain}(L) - 0.6 \cdot \text{rain}(M))} \quad (5)$$

	coefficient	exp(coefficient)	z	Pr(> z)
low temperature	-0.20	0.82	-3.83	0.00013
high temperature	-0.40	0.67	-7.66	0.00000
low humidity	-0.16	0.85	-2.86	0.00419
middle humidity	0.28	1.32	5.44	0.00000
low rainfall	-1.98	0.14	-10.36	0.00000
middle rainfall	-0.60	0.55	-13.81	0.00000

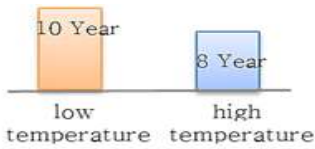
3.2 AFT(Accelerated Failure Time) 분석 결과

보증데이터에서 특정 차종의 부품을 선정하여 AFT모형을 통해 각 환경 인자가 부품수명에 어떤 영향을 주는 알아보았다. 우선 AFT모형을 통해서 다양한 모형을 만들 수 있다. 첫 번째로 연료도어 작동불량이 온도에 영향을 많이 받는 것으로 알려져 있으므로 온도만을 용한 모형을 만들어 분석해 보았다. 온도만을 고려한 AFT 모형을 만들어 B10 life 와 B50 life를 고온과 저온에 따라 수명에 영향을 주는지 보았다. <Table 3> 결과처럼 고온일 때 저온일 때 보다 수명을 낮았다. B10 life 기준은 고온은 8년, 저온은 10

년의 수명을 예측하였다. 따라서 문제 부품은 고온일 때 수명이 낮아짐을 확인하였다. 즉 온도에 따른 부품 수명이 고온보다 저온이 1.2배 더 높음을 확인하였다.

<Table 3> Comparison of life time

	low temperature	high temperature
B50 (50%tile)	15,393 (day)	12,086(day)
B10 (10%tile)	3,760(day)	2,953(day)



[Figure 3] lifetime compared according to temperature

둘째로 온도, 습도, 강수량, 강설량 등의 환경 조건을 이용한 최적의 모형을 찾아보았다. 이때 최적의 모형을 찾기 위해 단계별 변수 선택법(stepwise selection)을 적용하였다. 모형의 선정의 기준은 AIC 값을 사용하였다. Akaike(1974)가 모형의 복잡성(complexity)과 모델의 로그우도를 동시에 고려한 모형선정 방법을 제시하였다. 이 방법은 AIC(Akaike's Information Criterion)라 불리 우며 낮은 AIC 값을 형성하는 모형이 우수한 모형이다. 모형의 AIC는 다음과 같이 정의 되어진다.

$$AIC = -2LL + 2(k + c + 1) \quad (6)$$

여기서 LL은 모형의 로그 우도값이며, k는 변수의 수를 나타낸다. 또한 c는 설정된 분포의 파라메타를 나타낸다. <Table 4>은 각 모형별 AIC값을 나타낸 값으로 최적의 모형은 온도와 습도로 만들어진 모형이 가장 낮은 AIC값을 가지고 있다. 따라서 최적 모형은 온도와 습도가 포함된 AFT 모형이 선정 되었다.

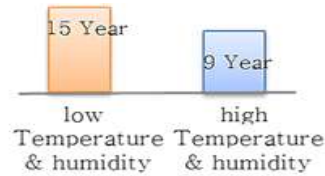
<Table 4> AIC comparison

NO	model	AIC
1	temperature + humidity + rainfall + snowfall	85631.1
2	temperature + rainfall + snowfall	85690.8
3	temperature + humidity + rainfall	85630.1
4	temperature + humidity	85605.6
5	temperature	86235.3

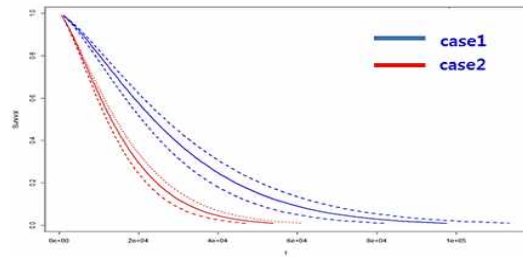
온도와 습도를 고려한 최종모형으로 B10 life 와 B50 life를 구해 비교해 보았다. <Table 5>를 보면 저온/저습일 때 B10 life 5,812일, 고온/다습 일 때 B10 life가 3,210일로 분석되었다. 분석결과 저온/저습이 고온/다습 보다 수명이 1.8배 더 높음을 확인 하였다.

<Table5> Comparison of life time

	Case1	Case2
	low temperature & humidity	high temperature & humidity
B50 (50%tile)	23,724 (day)	13,106 (day)
B10 (10%tile)	5,812 (day)	3,210 (day)



[Figure 4] lifetime compared according to temperature & humidity



[Figure 5] B10 life lifetime compared according to temperature & humidity

[Figure 5]은 조건1(저온 / 저습)과 조건2(고온 / 다습)의 B10 life를 비교한 그래프이며 AFT모형을 통해서 온도, 습도 조건 및 다양한 연관 조건들에 대한 생존율 및 수명 비교 가능하였고 특히 저습이 고온/다습 보다 수명이 1.8배 더 높음을 확인하였다. 이처럼 AFT 모형을 통해서 필드에서의 수명을 예측하였고, 특히 주요인자인 온도 및 습도에 따라서 필드수명에 달라짐을 확인하였다.

4. 결론

본 논문은 다양한 인자를 고려하여 고장원인을 추정할 수 있는 다변량 분석방법을 보증데이터에 적용해 보았고, 실제 유의함을 확인하였다. 특히 Tree 모형

및 Cox 모형을 통해서 서로의 장단점을 보완하였고, 더욱 정확한 원인분석을 할 수 있었다. 예측 부문에서는 AFT 모형을 통해서 문제 부품의 필드 수명 및 수명 영향인자를 확인 할 수 있었다. 좀 더 많은 고장인자를 고려하려면 제품 고장에 영향을 주는 인자를 찾아 DB를 만들고 본 논문에서 제시한 모형을 적용해 본다면 제품 개발에 많은 도움이 될 것이다. 본 논문의 주요 결론은 다음과 같다.

1) 보증데이터의 다변량 분석방법에 대해서 살펴보고 연구 현황을 정리하였다. 보증데이터 예측 목적에 부합하는 분석방법에 대해서 조사 연구를 하였다.

2) 실 사례를 통해서 Tree 및 Cox모형을 적용하여 유용성을 확인하였다. 특히, 각 모형의 장단점을 서로 보완하여, 더욱 정확한 분석을 할 수 있었다.

3) 실 사례를 통해서 AFT 모형을 적용하여 유용성을 확인하였다. 특히, AIC 값을 통해 최적의 AFT 모형을 찾고 필드 수명 및 주요 인자를 찾아 수명 영향도를 확인하였다.

4) 보증데이터를 이용한 필드 문제부품 수명예측 및 원인분석에 대한 다변량 분석방법에 대한 연구방향을 제시하였다.

5. References

- [1] Kalbfleisch JD, Lawless JF. (1988), "Estimation of reliability infield-performance studies" *Technometrics*, 30:365-388.
- [2] Hu XJ, Lawless JF. (1996), "Estimation from truncated life time data with supplementary information on covariates and censoring times" *.Biometrika*, 83:747-761
- [3] Hu XJ, Lawless JF.(1997), "Pseudo-likelihood estimation in a class of problems with response-related missing covariates", *The Canadian Journal of Statistics*, 25:125-142
- [4] Attardi L, Guida M, Pulcini G.(2005), "A mixed-Weibull regression model for the analysis of automotive warranty data". *Reliability Engineering and System Safety*, 87:265-273.
- [5] Karim MR, Suzuki K.(2007), "Analysis of warranty data with covariates". *Proceedings*

of the Institution of Mechanical Engineers, Part O: *Journal of Risk and Reliability*, 221:249-255.

- [6] Vinta S. (2009), "Analysis of data to predict warranty cost for various regions " *Proceedings of Annual Reliability and Maintainability Symposium*, 78-82
- [7] Hrycej T, Grabert M.(2007), "Warranty cost forecast based on car failure data ". *IEEE*
- [8] *International Conference on Neural Networks - Conference Proceedings*, 108-113
- [9] Thomas L. and William J. Kolarik(1987),
- [10] "Proportional Hazards Analysis of Field
- [11] Warranty Data", *Reliability Engineering*, 18
- [12] Sang Hyun Lee(2007), "A Study on Warning / Detection Degree of Warranty Claims Data Using Neural Network Learning" , *IEEE International Conference*.
- [13] Jong-Gurl Kim, Ki-Woo Sung(2014), "Cause Analysis Method of Warranty Claims Data Using Cox and Tree model" , *the Korea Safety Management & Science*

저 자 소개

김 종 결



서울대학교 계산통계학 에서 석사, 한국과학기술원 산업공학과에서 박사학위를 취득 하였으며, 현재 한국 품질보증 /PL 연구회회장으로 활동하고 있으며, 성균관대학교 시스템경영공학과 교수로 재직 중이다.

성 기 우



고려대학교 통계학과에서 학사와 석사를 취득 후 현대자동차 남양연구소에서 근무하고 있다. 성균관대학교 산업공학과 박사학위 과정에 재학 중이며, 관심분야는 보증데이터 분석 및 신뢰성 공학, 가속열화시험 (ADT) 등이다.