

## 품질보증 반환 데이터의 여러 가지 분석방법

백재욱<sup>1</sup> · 조진남<sup>2</sup>

<sup>1</sup>한국방송통신대학교 정보통계학과 · <sup>2</sup>동덕여자대학교 정보통계학과

접수 2014년 10월 27일, 수정 2014년 11월 25일, 게재확정 2014년 12월 2일

### 요약

기업에서는 매달 제품이 일정 수량만큼 수송되어 판매되고, 이들 중 일부는 반환 또는 클레임이 제기된다. 이전의 연구에서는 품질보증 반환 데이터의 반환율을 그래프상에 어떻게 타점할 것인지 먼저 살펴보고, 이어서 이런 데이터를 좌측 및 우측 중도중단 데이터의 결합으로 생각하여 와이블 분포 등을 적합시켜 신뢰성분석을 실시해보았다. 본 연구에서는 우선 수송된 제품이 곧 판매된 것으로 보고, 이들 제품의 품질보증 반환 데이터에 대해 비모수적인 방법을 적용하여 시간에 따른 신뢰도의 추이를 살펴보고 모수적 방법에 의한 결과와 비교해본다. 한편, 제품이 생산되어 판매지로 수송된 경우 소비자의 손에 들어가기까지는 시차가 있기 마련인데, 이와 같이 수송된 제품이 시차를 두고 판매되는 경우 이런 제품의 품질보증 반환 데이터에 대해서는 어떻게 분석을 실시해야 하는지 살펴본다.

주요용어: 비모수적 방법, 시간차, 와이블분포, 품질보증 반환 데이터.

### 1. 서론

근래 기업은 품질보증 반환 (warranty returns) 데이터의 올바른 분석을 통해 해당 제품이 추후 고장이 얼마나 날 것인지 예측할 수 있으며, 나아가 품질보증 비용의 예측은 물론 수리에 필요한 여유 부품의 수를 파악할 수 있으므로 품질보증 반환 데이터의 체계적 관리를 위해 노력을 기울이고 있다.

하지만 품질보증 반환 데이터는 실험데이터 (experimental data)가 아니고 관측데이터 (observational data)에 속하므로 제대로 관리가 되지 않은 경우가 많아 특정 목적을 위해 데이터를 분석할 때에는 관측데이터로는 충분치 않아 실험을 하게 되는 경우를 보게 된다.

품질보증 반환 데이터에 대한 연구는 Suzuki (1985a, 1985b)를 시작으로 진행했으며, Karim 등 (2005)과 Wu (2012)는 품질보증 반환 데이터에 대한 전반적인 분석방법을 소개했다. Wu (2013)는 품질보증 반환 데이터를 여러 가지 형식의 표로 집계하여 분석하기도 했다.

Kalbfleisch 등 (1991)은 클레임 건수를 집계하는 구간이 매번 일정한 경우 NHPP (non-homogeneous Poisson process) 모형을 이용하여 기대 클레임 건수를 예측했으며, Kalbfleisch와 Lawless (1996)는 집계하는 구간이 매번 일정하지 않은 경우에도 기대 클레임 건수를 예측하는 방법을 제안했다. Suzuki 등 (2000, 2001)과 Karim 등 (2001)은 수리계 아이템의 경우에는 NHPP 모형을 적용하고, 비수리계 아이템의 경우에는 다항모형과 이의 근사치인 포아송 모형을 적용하고 EM (expectation-maximization) 알고리즘을 활용하여 품질보증 반환 데이터를 분석하였다.

<sup>1</sup> 교신저자: (110-791) 서울 중로구 동승동 169번지, 한국방송통신대학교 정보통계학과, 교수.  
E-mail: jbaik@knou.ac.kr

<sup>2</sup> (136-714) 서울시 성북구 월곡동 23-1, 동덕여자대학교 정보과학대학 정보통계학과, 교수.

제품의 수명에 영향을 미치는 요인인 사용기간과 사용량을 동시에 고려한 2차원 품질보증 데이터에 대한 모델링 방법은 Baik과 Jo (2010)에 의해 제안되었으며, 품질보증기간이 끝난 후의 비용을 최소화 하는 방법에 대해서는 Jung (2006, 2008)이 살펴보았고, Baik (2010)은 기업에서 통상 월별 또는 분기별 클레임건수를 1000제품 당 나타낸 WCR (warranty call rate)을 어떻게 분석하는지 살펴보았다. Baik과 Jo (2011)에서는 필드 고장 요약 데이터를 활용하여 미래에 고장이 얼마나 날 것인지 예측하였다. 이어 Baik과 Jo (2014)에서는 표로 요약된 품질보증 반환 데이터에 대해 적절한 그래프에 의해 반환율을 표현하고, 좀 더 구체적인 반환시기가 주어진 경우 모수적 방법에 의해 분석을 어떻게 하는지 살펴보았다. 하지만 이런 품질보증 반환 데이터에 대해서도 비모수적 방법에 의해 시간에 따라 신뢰도가 어떻게 떨어지는지 살펴볼 필요가 있다. 한편, 기업에서 제품이 수송되어 판매처로 보내지는 경우 제품이 실제로 고객에게 판매되기까지는 시간이 걸릴 수 있다. 따라서 이런 시차 (time-lag)가 있는 경우 품질보증 반환 데이터가 어떻게 변하고 분석결과가 또한 어떻게 달라지는지 살펴볼 필요가 있다.

이에 본 연구의 2절에서는 우선 우리가 가지고 있는 품질보증 반환 데이터가 어떤 형태인지 먼저 소개하고, 비모수적인 방법을 적용하여 데이터를 분석할 경우 그 결과가 어떻게 나오는지 살펴보고 모수적인 방법에 의한 결과와 비교해본다. 이어 3절에서는 제품이 판매처로 수송되어 실제 판매될 때까지 시차가 있을 경우 분석해야 될 데이터가 어떻게 달라지는지 살펴보고, 분석결과 또한 어떻게 달라지는지 살펴보고자 한다. 마지막으로 4절에서는 본 연구에서 살펴본 것들을 정리하고 추후 연구과제를 제시한다.

## 2. 비모수적인 방법

어느 회사의 제품이 2012년 6월부터 매달 일정 수량 판매되었고, 그 중 일부가 반환 (클레임 제기)되었다 (Table 2.1 참조). 이들 데이터는 2013년 10월까지 조사되었고, 현재 시점은 2013년 11월이라고 하자. 따라서 2012년 6월에 판매된 354개의 제품 중 9개가 현재 시점까지 반환되었다. 이 제품에 대한 품질보증기간은 1년이므로 제조사는 이 제품이 품질보증기간인 1년 내에 반환될 가능성은 얼마나 되는지 알고 싶으며, 좀 더 구체적으로는 시간이 지나면서 신뢰도가 어떻게 떨어지는지 알고 싶어 한다. Table 2.1에서  $n(t)$ 는 시점  $t$ 에서 제조사가 판매한 수량이고  $r(t)$ 는  $n(t)$  중 현재 시점까지 반환된 수량을 나타낸다.

Table 2.1 Warranty returns data

t	Month	$n(t)$	$r(t)$
1	2013년 10월	358	0
2	2013년 9월	286	0
3	2013년 8월	205	1
4	2013년 7월	274	1
5	2013년 6월	293	0
6	2013년 5월	204	2
7	2013년 4월	225	0
8	2013년 3월	224	0
9	2013년 2월	201	0
10	2013년 1월	288	0
11	2012년 12월	186	2
12	2012년 11월	163	3
13	2012년 10월	208	1
14	2012년 9월	218	5
15	2012년 8월	128	0
16	2012년 7월	154	2
17	2012년 6월	354	9

이런 데이터는 Baik과 Jo (2014)에서 설명한 바와 같이 제품이 필드에 들어오는 시기가 다르므로 인해 중도중단 시점이 여럿이며, 각 시점에서 현재까지 몇 개가 고장 났는지 그리고 이후에 몇 개가 고장날 것인지 보여주므로 좌측 및 우측 중도중단이 함께 나타나는 데이터라는 것을 알 수 있다. 이런 좌측 및 우측 중도중단 데이터에 대해서도 비모수적인 방법으로 시간에 따른 신뢰도의 추이를 살펴볼 수 있다. Figure 2.1은 Minitab 16에 의한 결과를 보여준다.

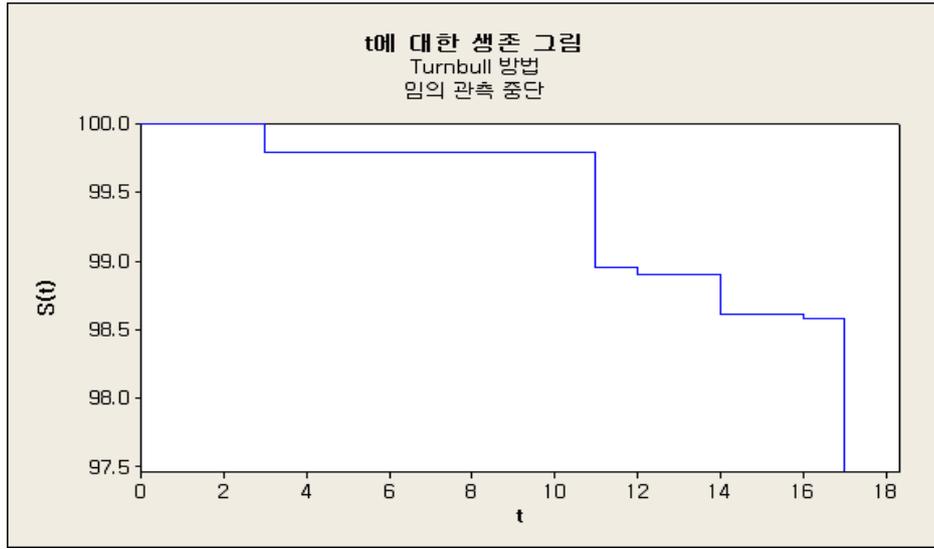


Figure 2.1 Reliability transition for left and right censored data (nonparametric method)

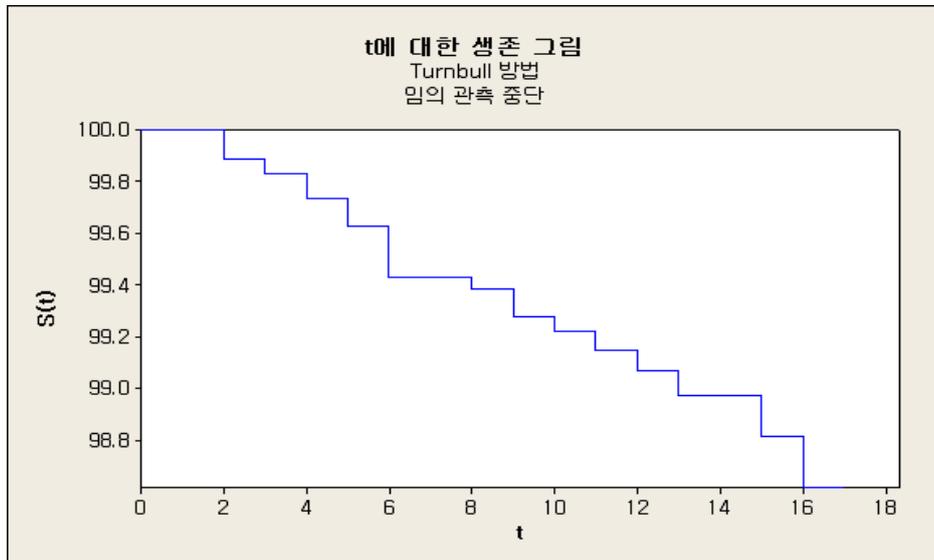
이로부터 제품이 12개월 이상 및 16개월 이상 작동할 확률 즉, 신뢰도  $R(12)$ 는 99% 정도,  $R(16)$ 은 98.5% 이상이 된다는 것을 알 수 있다. 하지만 비모수적 방법으로는 제품의 수명이 관측된 시간인 17개월까지만 생존확률의 추이를 살펴볼 수 있다는 단점이 있다. 한편, Baik과 Jo (2014)에서는 Table 2.1의 데이터에 대해 와이블분포를 적용하는 경우  $R(12)$ 와  $R(16)$ 이 각각 0.9903과 0.9836으로 나와 앞의 비모수적 방법에 의한 결과와 차이가 없음을 알 수 있다. 하지만 모수적 방법을 적용했으므로 척도모수  $\eta$ 와 형상모수  $\beta$ 가 각각 151.3과 1.826으로 추정되었으며, 형상모수가 1보다 크므로 시간이 흐르면서 고장률이 증가하는 경향을 보인다는 것을 알 수 있었다. 또한 Baik과 Jo (2014)에서는 확률밀도 함수가 우측으로 기울어진 모양을 취하므로 대표치로 중위수를 구하는 것이 의미가 있으며, 이는 123개월 (약 10년) 정도가 되며, 품질보증기간이 끝나는 12개월 시점까지 제품이 살아있을 가능성은 99% 정도가 됨을 볼 수 있었다.

앞의 Table 2.1의 데이터에서는 제품이 반환 또는 클레임 제기된 경우 구체적으로 언제 반환 또는 클레임 제기되었는지 알 수가 없다. 하지만 데이터 분석자의 끈질긴 추적으로 구체적인 반환 또는 클레임 제기 시기가 Table 2.2의 'time at failure'에서와 같이 알려져 있다고 하자. Baik과 Jo (2014)의 Table 4.1과 다른 것은 본 데이터의 경우 2013년 8월에서 time at failure가 2라는 점이다. 그 값을 달리 한 이유는 3절에서 시차를 반영해야 하는데, 이에 맞는 데이터가 필요하기 때문이다. 이런 경우에는 Table 2.1에서 좌측 중도중단 데이터가 구체적인 고장시간으로 나타난 것이므로 우측 중도중단 데이터만 있는 경우가 된다.

**Table 2.2** Warranty returns data (specific returns periods included)

Month	$n(t)$	$r(t)$	$t$	time at failure	right cens
2013년 10월	358	0	1	-	358개
2013년 9월	286	0	2	-	286개
2013년 8월	205	1	3	2	204개
2013년 7월	274	1	4	2	273개
2013년 6월	293	0	5	-	293개
2013년 5월	204	2	6	2, 5	202개
2013년 4월	225	0	7	-	225개
2013년 3월	224	0	8	-	224개
2013년 2월	201	0	9	-	201개
2013년 1월	288	0	10	-	288개
2012년 12월	186	2	11	3, 6	184개
2012년 11월	163	3	12	4, 5, 8	160개
2012년 10월	208	1	13	5	207개
2012년 9월	218	5	14	3, 6, 6, 9, 13	213개
2012년 8월	128	0	15	-	128개
2012년 7월	154	2	16	4, 11	152개
2012년 6월	354	9	17	2,4,6,6,9,10,12,15,16	345개

이런 우측 중도중단 데이터에 대해서도 비모수적인 방법으로 시간에 따른 신뢰도의 추이를 살펴볼 수 있다. Figure 2.2는 Minitab 16에 의한 결과를 보여준다.

**Figure 2.2** Reliability transition for right censored data (nonparametric method)

이로부터 제품이 12개월 이상 및 16개월 이상 작동할 확률 즉, 신뢰도  $R(12)$ 는 99.0% 이상, 신뢰도  $R(16)$ 은 98.6% 이상이 된다는 것을 알 수 있다. 하지만 Figure 2.1에서와 같이 비모수적 방법으로는 제품의 수명이 관측된 시간까지만 생존확률의 추이를 살펴볼 수 있다는 단점이 있다.

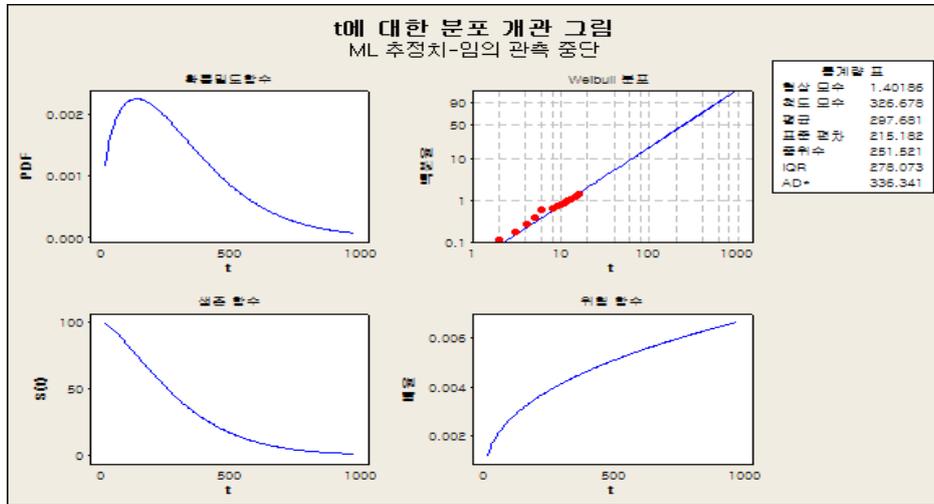


Figure 2.3 Fitting of Weibull distribution to Table 2.2 data

Table 2.2의 데이터에 대해 와이블분포를 적용하는 경우 척도모수  $\eta$ 와 형상모수  $\beta$ 는 각각 327과 1.402로 추정된다. 형상모수  $\beta$ 가 1보다 크므로 시간이 흐르면서 고장률이 증가한다는 것을 알 수 있다. 아울러 확률밀도함수가 우측으로 기울어진 모양을 취하므로 대표치로 중위수를 구하는 것이 의미가 있는데, 이는 252개월 정도가 된다. 또한 품질보증기간이 끝나는 12개월 시점까지 제품이 살아있을 가능성은 99%보다 훨씬 더 크게 나온다.

### 3. 시차가 있는 경우 데이터의 분석

Table 2.1이나 Table 2.2에서  $n(t)$ 는 시점  $t$ 에서 제조사가 판매한 수량이라고 했는데, 사실 제조사가 판매처로 제품을 수송한 시기인 경우가 많다. 자동차의 경우 실제 데이터로 그 시차를 살펴보면 짧으면 수송된 달에 팔리기도 하지만 길면 수송된 달에서부터 1년 후에나 팔리기도 한다. 이런 경우 시차 (time-lag)가 크므로 신뢰도 등의 분석은 판매된 시점을 기준으로 실시해야 한다. 본 논문에서는 분석의 편의상 제품이 수송된 달로부터 1달 정도 걸려 판매된다고 가정하자. 이런 경우 앞의 Table 2.2는 다음 Table 3.1과 같이 변한다 ( $n(t)$  열의 값이 1달씩 밀린다).

Table 3.1 Warranty returns data (1 month of time lag considered)

t	n(t)	# at risk	# failures	right cens
1	286	3611	4	286
2	205	3321	2	204
3	274	3115	3	273
4	293	2839	3	293
5	204	2543	5	202
6	225	2336	0	225
7	224	2111	1	224
8	201	1886	2	201
9	288	1683	1	288
10	186	1394	1	184
11	163	1209	1	160
12	208	1048	1	207
13	218	840	0	213
14	128	627	1	128
15	154	498	1	152
16	354	345	0	345

하지만 Table 2.2에서 Table 3.1로의 작은 변화가 신뢰성 등의 분석결과에 어느 정도 영향을 미치는 것을 볼 수 있다. 특히 제품의 수송으로부터 판매까지의 시차가 길면 길수록 그 영향이 커지는 것을 짐작할 수 있다. Figure 3.1은 Minitab 16에 의한 결과를 보여준다. 이로부터 제품이 12개월 이상 작동할 확률 즉,  $R(12)$ 는 99.0% 정도가 됨을 알 수 있다. 하지만 Figure 2.1이나 Figure 2.2에서와 같이 비모수적 방법으로는 제품의 수명이 관측된 시간까지만 생존확률의 추이를 살펴볼 수 있다.

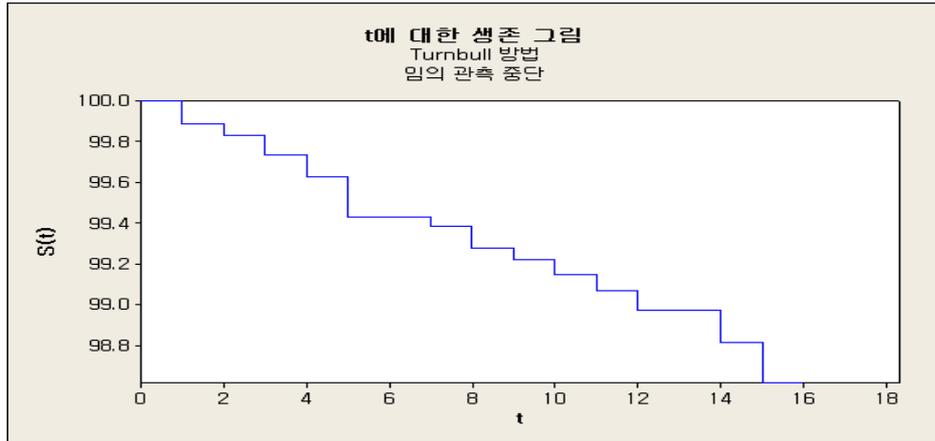


Figure 3.1 Fitting of Weibull distribution to Table 3.1 data

Table 3.1의 데이터에 대해 와이블분포를 적용하는 경우 척도모수  $\eta$ 와 형상모수  $\beta$ 는 Figure 3.2에서와 같이 각각 616과 1.15로 추정되어, Table 2.2의 데이터에 대해 와이블분포를 적합시켜 추정된 결과인 척도모수  $\eta$ 와 형상모수  $\beta$ 의 추정치인 327과 1.4와 많이 다름을 알 수 있다. 특히 시차를 고려하는 경우 형상모수의 값이 1.15로 추정되어 시간이 흐르면서 고장률이 약간만 증가하는 경향을 보이게 된다. 아울러 확률밀도함수가 우측으로 기울어진 모양을 취하므로 대표치로 중위수를 구하는 것이 의미가 있으며, 이는 447개월로 이전 추정치인 252개월 훨씬 크다는 것을 알 수 있다. 하지만 품질보증기간이 끝나는 12개월 시점까지 제품이 살아있을 가능성은 99%보다 훨씬 더 크게 나와 1개월의 시차를 고려하지 않은 경우와 비슷하게 나온다는 것을 볼 수 있다.

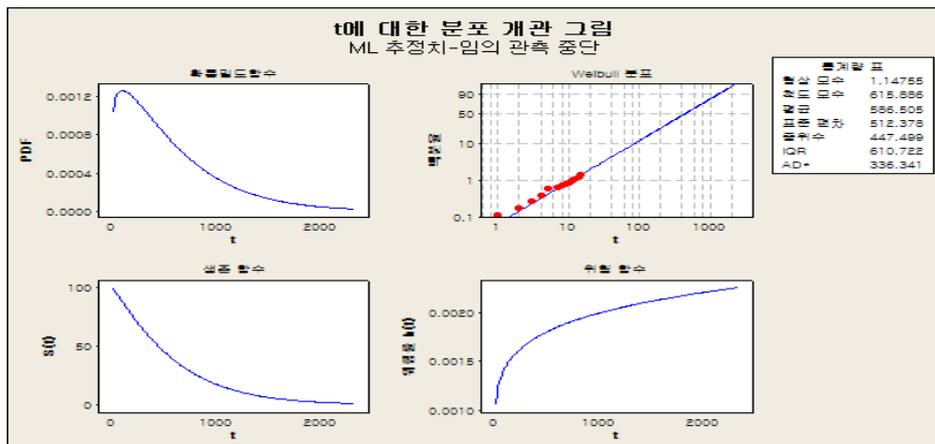


Figure 3.2 Fitting of Weibull distribution to Table 3.1 data

#### 4. 결론

기업에서는 매달 제품이 일정 수량만큼 수송되어 판매되고, 이들 중 일부는 반환 또는 클레임이 제기된다. Baik과 Jo (2014)에서는 품질보증 반환 데이터의 반환율을 그래프상에 어떻게 타점할 것인지 먼저 살펴보았고, 이어서 이런 데이터를 좌측 및 우측 중도중단 데이터의 결합으로 생각하여 와이블 분포 등을 적합시켜 신뢰성분석을 실시해보았다.

하지만 모수적 방법을 적용하여 신뢰성분석을 실시하기 전에 비모수적인 방법으로 신뢰성의 추이나 고장률의 추이를 살펴보는 것이 더욱 적절할 것이다. 따라서 본 연구에서는 품질보증 반환 데이터가 Table 2.1이나 Table 2.2의 형식으로 요약된 경우 시간이 지나면서 신뢰도가 어떻게 변하는지 비모수적인 방법으로 살펴보았고, 그 결과를 모수적인 방법으로 나온 결과와 비교하였다. 본 연구에서 살펴본 Table 2.1 데이터의 경우 비모수적인 방법으로 구한  $R(12)$ 와  $R(16)$ 이 각각 0.9903과 0.9836으로 나와 Baik과 Jo (2014)에서 제시한 모수적 방법에 의한 결과와 차이가 거의 없음을 알 수 있다.

한편, 제품이 생산되어 판매지로 수송된 경우 소비자의 손에 들어가기까지는 시차가 있기 마련인데, 이와 같이 수송된 제품이 시차를 두고 판매되는 경우 Table 2.2와 같은 데이터가 어떻게 변하며 분석결과는 어떻게 변하는지 살펴보았다. 구체적으로 1달의 시차를 가정할 수 있는 경우 Table 2.2의 데이터에서 첫 번째 행만 삭제되고 그 다음 행부터의 데이터만 고려하는 Table 2.3의 데이터가 되는데, 이런 아주 작은 변화가 시간  $t$ 에 따른 신뢰도의 추이에 적지 않은 영향을 미칠 수 있다. 앞의 데이터의 경우  $R(12)$ 는 0.99 정도로 추정치가 거의 변화가 없는 반면  $R(16)$ 은 아예 추정할 수 없게 되는 결과를 낳게 된다.

본 연구에서 분석의 편의상 제품이 수송되어 판매되기까지 1달이 걸린다고 가정했는데, 사실 그 시차는 확률변수로 표현할 수 있다. 따라서 추후 연구에서는 이런 시차를 확률변수로 표현하여 품질보증 반환 데이터를 분석하고자 한다.

다음으로 본 연구에서는 제품이 한번 고장 나면 그만인데, 실제 내구소비재의 경우 수명주기 동안 여러 번 고장 난다. 따라서 추후의 연구에서는 이런 수리계 아이템의 실제 고장 데이터에 맞는 모형을 제시하고자 한다.

#### References

- Baik J. W. (2010). The study on the analysis of quality assurance data. *Journal of the Korean Data & Information Science Society*, **21**, 621-628.
- Baik, J. W. and Jo, J. N. (2010). Various types of modelling for scale parameter in Weibull intensity function for two-dimensional warranty data. *Journal of the Korean Data & Information Science Society*, **21**, 555-560.
- Baik, J. W. and Jo, J. N. (2011). Predicting the future number of failures based on the field failures summary data. *Journal of the Korean Data & Information Science Society*, **22**, 755-764.
- Baik, J. W. and Jo, J. N. (2014) Reliability analysis of warranty returns data. *Journal of the Korean Data & Information Science Society*, **25**, 1-9.
- Jung, K. M. (2006). Optimal preventive maintenance policy for a repairable system. *Journal of the Korean Data & Information Science Society*, **17**, 367-377.
- Jung, K. M. (2008). Optimization of cost and downtime for periodic PM model following the expiration of warranty. *Journal of the Korean Data & Information Science Society*, **19**, 587-596.
- Kalbfleisch, J. D. and Lawless, J. F. (1996). Statistical analysis of warranty claims data. In *Product Warranty Handbook*, edited by Blischke, W. R. and Murthy, D. N. P., Marcel Dekker, 231-259.
- Kalbfleisch, J. D., Lawless, J. F. and Robinson, J. A. (1991). Methods for the analysis and prediction of warranty claims. *Technometrics*, **33**, 273-285.
- Karim, M. R. and Suzuki, K. (2005). Analysis of warranty claim data: A literature review. *International Journal of Quality and Reliability Management*, **22**, 667-686.

- Karim, M. R., Yamamoto, W. and Suzuki, K. (2001). Statistical analysis of marginal count failure data. *Lifetime Data Analysis*, **7**, 173-186.
- Suzuki, K. (1985a). Nonparametric estimation of lifetime distribution from a record of failures and follow-ups. *Journal of American Statistical Association*, **80**, 68-72.
- Suzuki, K. (1985b). Estimation of lifetime parameters from incomplete field data. *Technometrics*, **27**, 263-271.
- Suzuki, K., Karim, M. R. and Wang, L. (2001). Statistical analysis of reliability warranty data. In *Advances in Reliability*, edited by Rao, C. R. and Balakrishnana, N., Elsevier, 585-609.
- Suzuki, K., Yamamoto, W., Karim, M. R. and Wang, L. (2000). Data analysis based on warranty database. In *Recent Advances in Reliability Theory*, edited by Limnios, N. and Nikulin, M., Birkhauser, 213-227.
- Wu, S. (2012). Warranty data analysis: A review. *Quality and Reliability Engineering International*, **28**, 795-805.
- Wu, S. (2013). A review on coarse data and analysis. *Reliability Engineering and System Safety*, **114**, 1-11.

## Various types of analysis of warranty returns data

Jaiwook Baik<sup>1</sup> · Jinnam Jo<sup>2</sup>

<sup>1</sup>Department of Information Statistics, Korea National Open University

<sup>2</sup>Department of Statistics & Information, Dongduk Women's University

Received 27 October 2014, revised 25 November 2014, accepted 2 December 2014

### Abstract

A certain number of products are transported to be sold each month and some of them are returned for repair. In this study we first assume that the transported products are the ones that have been sold, Then nonparametric approach is applied to the warranty returns data to see how the reliability decreases over time. Parametric approach such as Weibull distribution is applied to the same data and the results for both nonparametric and parametric approaches are compared. Next we assume that there is a time lag between shipment and sale. Then both nonparametric and parametric approaches are applied to the time-lag data and the results are compared.

*Keywords:* Nonparametric method, time-lag, warranty returns data, Weibull distribution.

---

<sup>1</sup> Corresponding author: Professor, Department of Information Statistics, Korea National Open University, Seoul 110-791, Korea. E-mail: jbaik@knou.ac.kr

<sup>2</sup> Professor, Department of Statistics & Information, Dongduk Women's University, Seoul 136-714, Korea.