자동차엔진의 품질보증데이터 분석

2004년 7월 1일

한국방송통신대학교
정보통계학과

백재욱 (jbaik@knou.ac.kr)

Outline

1. Why the Analysis

2. Contents

2.1 Phase 1 : Data Cleaning
(1) Data Filtering(including the identification of important components)
(2) Population Formulation

2.2 Phase 2 : Analysis at the Component Level
(1) Reliability Analysis
(2) Warranty Cost Analysis

2.3 Phase 3 : Analysis at the Subassembly Level

2.4 Phase 4 : Analysis at the Assembly(Engine) Level

3. Summary and further considerations
3.1 Summary
3.2 Further considerations

1. Why the Analysis

Warranty on some components of engine for 10 years or 100K miles whichever comes first.

2. Contents

2.1 Phase 1 : Data Cleaning

(1) Identification of important components

Table 1. Data filtering process

Raw data

↓

Deleted if the cell for the exchanged component(A) is null (in order to use only exact information provided)

↓

Deleted if the data are of inappropriate vehicle types (vehicle types X3 and LC saved, and vehicle type RC in 1996 and 1997 deleted)

↓

Deleted if the numbers for the exchanged components(A) don' t start with 0, 1, 2, 3 and 9.

↓

Arrange the remaining data in terms of the exchanged components in an ascending order

↓

Deleted if the component is not listed in the LPB

Table 2. Pareto analysis on 1996 data

Table 3. Pareto analysis on 1996-2000 data combined

Table 4. Number of claims for important components in each subassembly(1996 to 2000 data)

(2) formulation of population

1) # of cars which have not claimed(# of censored cars)

** If the cell for the exchanged component(A) is null it is replaced by the component causing the problem. But the component causing the problem is not always the same as the exchanged component. **

① predicted monthly number of cars sold

Table 5. monthly number of cars sold(supplied by the company)

Table 6. Annual data acquisition rates (also considering production to sale period)

Table 7. predicted monthly number of cars sold based on the acquisition rates

② monthly number of cars claimed
** Had to arrange the data in terms of the month sold **
②-1. monthly number of claims based on the month sold [claims data with sales dates 1900.01.01 and 2000.01.01(manufactured in the middle of the year but sold on Jan 1^{st}, 2000) ignored]

Table 8. monthly number of claims

②-2. monthly number of claims based on the month sold (the sales dates of claims data with sales dates 1900.01.01 and 2000.01.01 estimated and put into the whole data set)

Table 9. Regression analysis results

Table 10. Monthly number of claims based on the month sold (after the sales dates of claims data with sales dates 1900.01.01 and 2000.01.01 are estimated)

②-3. Monthly number of cars claimed

Table 11. Monthly number of cars claimed

③ Monthly number of censored cars(before adjustment)

Table 12. Monthly number of censored cars(before adjustment) [Table 7 □ Table 11]

④ Monthly number of censored cars(after adjustment)

Figure 1. Production to sale in months (1996 data)

Table 13. How to adjust the predicted monthly number of cars sold

Table 14. Adjusted monthly number of censored cars (after adjustment for 1996 and for May and June of 1998)

Table 15. Censored life for for each month

2) life for each component from the claims data
Note : Replacing the null exchanged component(A) with the component causing the problem we had so many same claims data. In this case they have been treated as one claim.
The following are the lives that can be identified from the claims data.
① For each claim data calculate 'date replaced □ date sold'

② If a component is replaced more than once in the same car calculate 'date replaced recently □ date replaced previously' (Mileages could be calculated as in ① and ②.
③ For any claim data calculate 'April 15 2001 □ date replaced' since April 15 2001 was the date the claims data was collected. This is censored life.
⑤ In some cars some components had been replaced. But there are other components that had not been replaced at all. In this case calculate 'April 15 2001 □ date sold' for other components. This is censored life.

In summary, population for any component in the engine is formed by combining data from 1) and 2).

2.2 Phase 2 : Analysis at the component level

(1) Reliability analysis

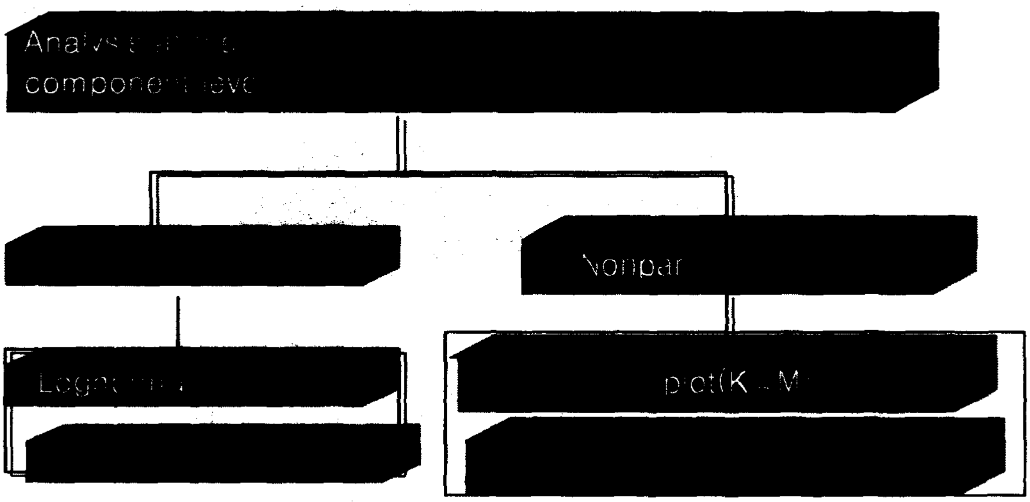

Figure 2. Analysis at the component level

Figure 3. Weibull and Lognormal plotting for component 25100(pump assy-water)

Figure 4. Reliability analysis in case Weibull distribution is valid

Figure 5. Nonparametric analysis for important components in cooling subsystem

Figure 6. Nonparametric analysis for important components in lubrication subsystem

Figure 7. Nonparametric analysis for important components in timing subsystem

(2) Warranty cost analysis

Nelson, W. (1998), "An Application of Graphical Analysis of Repair Data," Quality and Reliability Engineering International, 14, 49-52.

Figure 8. Warranty cost analysis for component 25100(pump assy-water) in cooling subsystem

Figure 9. Warranty cost analysis for component 26300(filter ass-oil)) in cooling subsystem

Figure 10. Warranty cost analysis for component 24321(belt-timing) in timing subsystem

2.3 Phase 3 : Analysis at the subsystem level

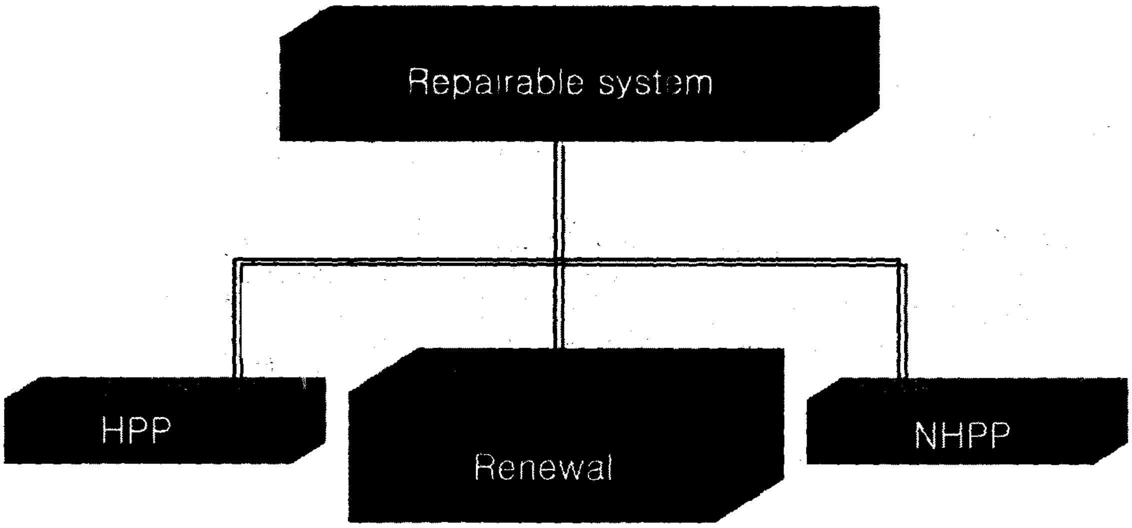

Figure 11. Different ways of analysis

Assumption : independence [-> $h(t)=h1(t)+h2(t)$]

Figure 12. Hazrd plot for the cooling subsystem if Weibull distributions for component 25100 and 25500 were valid

Figure 13. Nonparametric hazrd plot for the cooling subsystem with components 25100 and 25500

Figure 14. Nonparametric hazrd plot for the lubrication subsystem
with only one component 26300

Figure 15. Nonparametric hazrd plot for the timing subsystem
with components 21350, 24312, 24321, 24410

Warranty cost for a subsystem?

WC(cooling subsystem) = WC(component 25100) + WC(component 25500)
If the replacement cost of each component is 1$ the warranty cost for the cooling subsystem is as shown below.

Figure 16. Warranty cost for the cooling subsystem

2.4 Phase 4 : Analysis at the assembly(engine) level

h(t)=h1(t)+h2(t)+ …. + hn(t) where n is the number of components in the assembly.

Figure 17. Nonparametric hazard plot for engine

Can also get the warranty cost for the engine as in the subassembly.

3. Summary and Further Works

3.1 Summary

Found important components in terms of frequency in the assembly (but not in terms of money).
Component 39*** was most important.
The failure mode was N69(no light on warning signal)
The cause of the failure was C15(bad connection).

Formed a population for each component.

Performed reliability and warranty cost analyses
- At the component level
- At the subsystem level
- At the system level.
- ** They don't trust the warranty cost analysis. **

Reliability improvement
- Among all the subsystems front ? subsystem is most vulnerable (among other things due to the large number of components in it), especially components 39*** and 28***.

3.2 Further considerations

1.Problems with the raw data (Stumbling block to accurate data analysis)

Too much data missing in the cell for the exchanged component(A) [more than 50%]

Whai if I delete them all?

Too much work to sort out for engine data.

Unexpected data, for example

More than one claim for the same car but with different sales dates.

Sales dates of 1900.01.01

If there are more than one claim with sales dates of 1900.01.01 for the same car we only need to estimate the sales date once.

Need to check whether some sales dates are different for the same car (some sales dates of 1900.01.01).

** Any thing wrong can happen to the raw data (eg. Sales date > repair date).

2. Need to communicate with the same language in terms of failed components, failure modes, the causes of the failures among A/S centers and data analysts.

3. Need to know how long the first owner holds the car.

4. Some complaints are not reported. How do I incorporate it into the reliability and warranty cost analyses?

5. In this case study nonparametric methods have been applied (parametric methods are hard to apply with long period of use).

5.1 Parametric methods

Are there any parametric methods that can be applied to this type of data? For instance a mixture of Weibull distributions or exponentiated Weibull? Nothing like that will likely to work when the hazard plot shows several humps(Sure I can model it by decomposing whole time zone into several disjoint intervals. But I will lose sense of the engineering interpretation).

Or I can sort the whole data into several groups where each group represents a particular failure mode and each failure mode can be can be explained by a Weibull distribution, a mixture of Weibull distributions, or an exponentiated Weibull distribution.

5.2 Nonparametric methods(Only valid for the period when the data were collected?)

Based on 5 years of data can I say something about the reliability and warranty cost, say at year 6 or year 10 (Sure it will be more difficult to predict them at year 10).

6. After reliability improvement action is performed we need to check Reliability growth.

6.1 Need to check reliability growth with same components but with different component numbers.

6.2 Need to check how much warranty cost is saved.

6.3 Can perhaps check the reliability growth by comparing the number of repairs and their failure modes(failure mode as a weight)

7. I considered only time for reliability and warranty cost analyses.

Need to consider the usage (mileage) as well. -> Can use usage as an acceleration factor(?) or can use two dim'al analysis.

8. Only considered one dependent variable (or consider two dependent variables in 7 above).

Why not consider some independent variables? For instance, environmental, usage, manufacturing conditions may need to be used so that they can be linked to the dependent variables. In this case study dealer can show usage condition (Not many independent variables in this dataset). Proportional hazard model can be used to explore the effect of independent variables.

9. Management

9.1 The DBMS that can emcompass all of the above is needed for management.

9.2 DBMS that can encompass the lifecycle [Design -> Manufacturing -> A/S (->another loop)]

10. Follow-up analysis on a small number of cars for year 11 performance.

11. Aggregation of components analyses = subsystem analysis ?

Aggregation of subsystem analyses = system analysis ?

(Analysis in terms of reliability and warranty cost)