DOI: http://dx.doi.org/10.26483/ijarcs.v8i9.4858

ISSN No. 0976-5697

Volume 8, No. 9, November-December 2017

International Journal of Advanced Research in Computer Science

TECHNICAL NOTE

Available Online at www.ijarcs.info

AN APPLICATION FOR COMPUTING FAILURE AND SURVIVAL RATES OF AUTO COMPONENTS

Nirbhay Pherwani
VESIT, University of Mumbai
Mumbai, India

Kunal Tolani
VESIT, University of Mumbai
Mumbai, India

Abstract: An application for computing failure of automobile components using Weibull and other survival analysis methods is presented through this paper. The following areas are touched upon: I) Failure rates are described first, along with data cleaning and 'failed' data is defined within spell periods; II) Kaplan-Meier life curves are used to compare the failure mode reliability over time; III) Vehicle failure is analyzed by applying Weibull regression to fit on real-world data from different conditions.

Keywords: Weibull model, Kaplan-Meier life curve, Survival Estimation

I. INTRODUCTION

It's important to take into account the reason for the failure of automobile components after the warranty period and to understand the reliability of components associated with various manufacturing, environmental and testing conditions.

The Weibull model is an established tool to compute failure trends beyond the available test duration. Also, survival analysis is a suitable technique when it comes to close time-related reliability data.

We have used the Weibull model here for the primary purpose of exploring failure rate over service time and to also provide insights when it comes to component failure using the three key sub-points mentioned below:

- Provide a descriptive summary of failure data of components or the vehicle model itself.
- Explore the component failure probability over test time, and compare the failure rates of the same component from a few different data sets;
- Improve the data fitting by using a three-parameter Weibull model.

II. DESCRIPTIVE SUMMARY OF DATA SAMPLE:

The Descriptive Summary contains information about the parts that eventually failed over time that is the number of parts of each type that failed, etc.

```
> # Descriptive statistics
>
> summary(time)
   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
  1.000   3.000   5.000   4.996   7.000  12.000
>
> summary(event)
   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
      1       1       1       1       1       1
>
> summary(z)
  Failure.Code     Model.Family.Desc   Part.Desc
 Min.   : 1.000   Min.   :1.000     Min.   :1.000
 1st Qu.: 3.000   1st Qu.:5.000     1st Qu.:4.000
 Median : 3.000   Median :5.000     Median :4.000
 Mean   : 3.348   Mean   :4.846     Mean   :4.572
 3rd Qu.: 3.000   3rd Qu.:5.000     3rd Qu.:6.000
 Max.   :12.000   Max.   :5.000     Max.   :6.000
>
> summary(group)
           VACUUM MODULATOR   Vacuum Modulator-Padmini VACUUM MODULATOR - PADMINI
                        352                          5                        192
VACUUM MODULATOR (EGR) R&R VACUUM MODULATOR (VGT) R&R       VACUUM MODULATOR EGR
                       1932                        329                       1777
>
```

Figure. 1. Descriptive Summary of Data

III. METHODS OF MODELING SURVIVAL AND FAILURE RATES

It is interesting when we look at the failure rate, say F(t) from the survival point of view. The survival probability, say S(t) has a straightforward relationship for a fixed sample. S(t) = 1 -F(t). i.e. 30% failure means 70% survival. Kaplan Meier is one of the known and useful tools when it comes to comparing survival probability over time. The below formula illustrates the KM survival curve.

$$\hat{S}(t) = \prod_{t_i < = t} \left(1 - \frac{d_i}{n_i}\right) = \prod_{t_i < = t} \left(\frac{s_i}{n_i}\right) \quad (1)$$

The same expression can be further represented by the below formula. Here 'di' denotes the failed automobile component, 'si' denotes the survivor component, and 'n' the total of the two. (i.e., failed and survived subjects).

$$F(t) = 1 - e^{-(t/\eta)^{\beta}} \qquad (2)$$

Or the same can be illustrated by the following linear transformation.

$$\log(-\log(S(t)) = \beta\log(t) - \beta\log(\eta) \qquad (3)$$

In the above equation (3), S(t) denotes the survival function, which is estimated from the KM curve mentioned earlier. F(t) in equation (2) indicates the failure probability with increasing time, 'β' is the slope of the linear plot or the shape, 'η' is the Scale parameter. When test data cannot be visualized as a linear plot, a Weibull with three parameters is used to get better data fitting, where a time shift t0 is included. Refer equation (4). [1]

$$F(t) = 1 - e^{-((t-t_0)/\eta)^{\beta}} \qquad (4)$$

R programming is used for computations.

IV. KAPLAN-MEIER LIFE CURVES

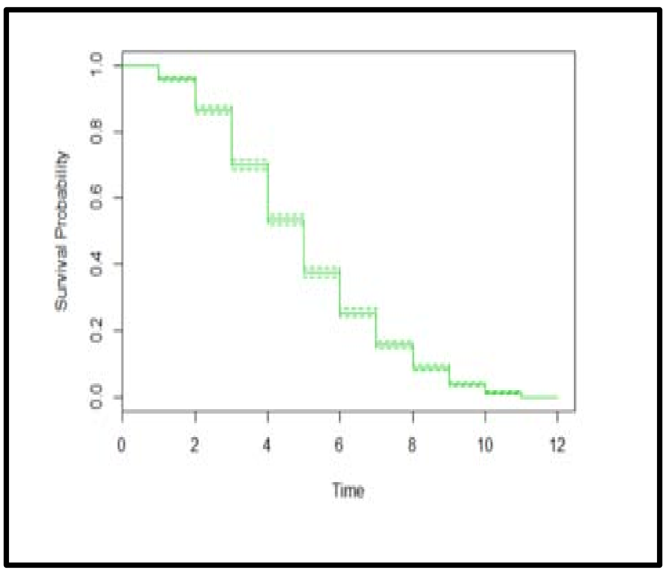

Figure 2. Kaplan Meier Non Parametric Analysis

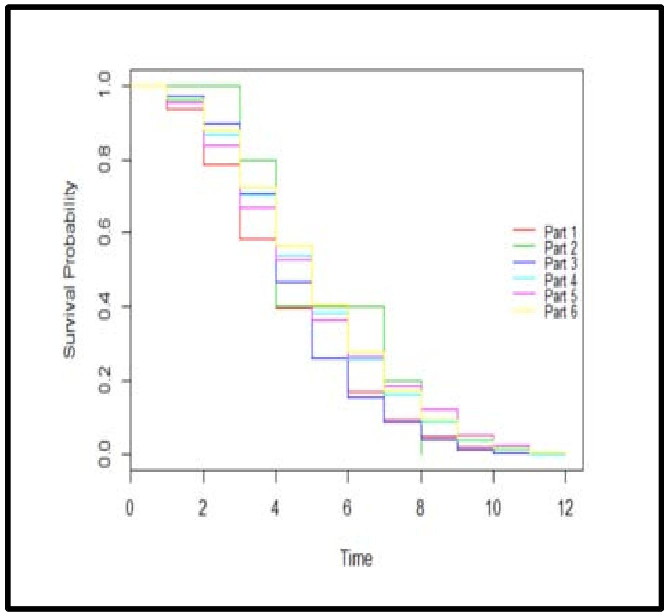

Figure 3. Kaplan Meier Non Parametric Analysis by group

The 'survival' package contains functions for carrying out survival analysis. Step 1 of any analysis is to create the survival object using the 'surv' function. Now we need to specify two arguments: first being the variable that records the subject at the event time and the second being the variable that identifies whether the event is a failure or censored. The 'survfit' function then uses the survival object to produce the Kaplan-Meier estimate of the survivor function. $surv denotes the KM estimate while $upper and $lower are the confidence limits (95%). On plotting the 'survfit' object, we get the estimated survivor function with a 95% confidence band. [2]

For example, for the part 'Brakes' we computed survival estimate over time.

```
> # Descriptive statistics
> summary(time)
   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
  1.000   5.000   8.000   7.093   9.000  12.000
> summary(event)
   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
      1       1       1       1       1       1
> library(survival)
> # Kaplan-Meier non-parametric analysis
> kmsurvival <- survfit(Surv(time,event) ~ 1 , data=mydata12)
> summary(kmsurvival)
Call: survfit(formula = Surv(time, event) ~ 1, data = mydata12)

 time n.risk n.event survival std.err lower 95% CI upper 95% CI
    1    927      17   0.9817 0.00441       0.9731       0.9903
    2    910      65   0.9115 0.00933       0.8934       0.9300
    3    845      35   0.8738 0.01091       0.8527       0.8954
    4    810      91   0.7756 0.01370       0.7492       0.8029
    5    719      73   0.6969 0.01510       0.6679       0.7271
    6    646      85   0.6052 0.01605       0.5745       0.6375
    7    561      96   0.5016 0.01642       0.4704       0.5349
    8    465     154   0.3355 0.01551       0.3064       0.3673
    9    311     101   0.2265 0.01375       0.2011       0.2552
   10    210      88   0.1316 0.01110       0.1115       0.1553
   11    122      73   0.0529 0.00735       0.0403       0.0694
   12     49      49   0.0000     NaN           NA           NA
```

Figure 4. Summary of Survival estimate for 'Brakes in Scorpio Hawk'

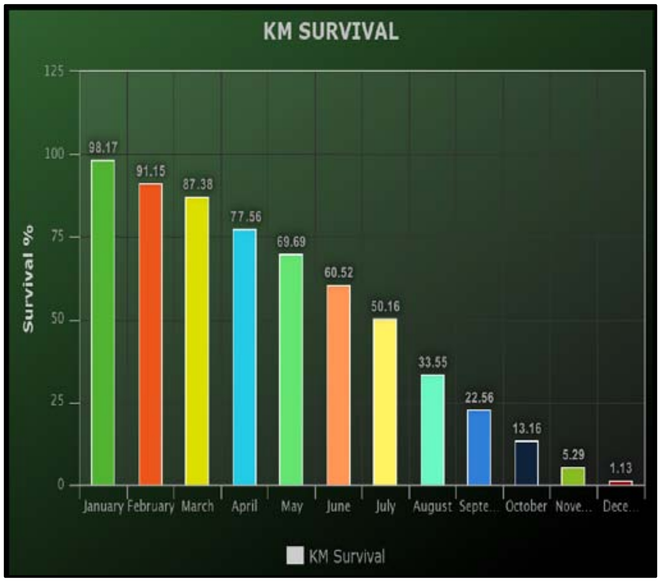

Figure 5. Survival plot across the year for 'Brakes in Scorpio Hawk'

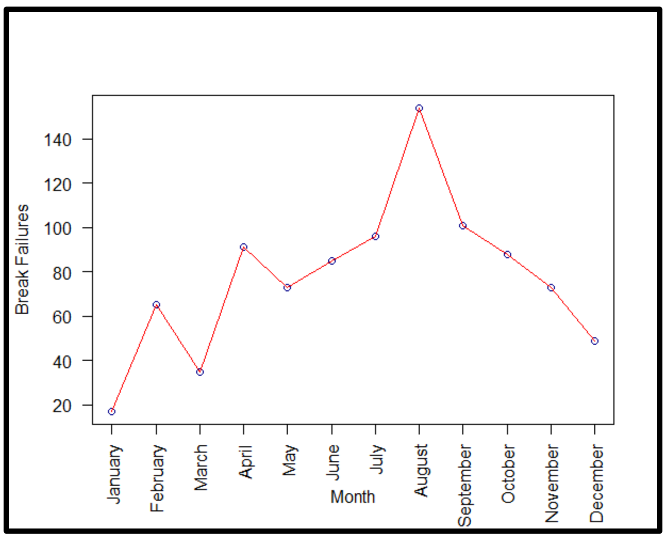

Figure 6. Brake Failure likelihood across the year for 'Brakes in Scorpio Hawk'

V. WEIBULL MODELING

We need to use the pweibull function with the argument lower.tail=FALSE, or equivalently, plot 1–pweibull because the Weibull parameters in the survreg function differ from the Weibull function in R. The scale parameter is basically the reciprocal of the Weibull shape parameter used in R. Also, the linear predictor is an estimate of log λ, i.e., the log scale.

Similarly, we can obtain estimates of the Weibull densities and the Weibull hazard functions in different situations. [3]

```
> survreg(Surv(time,event)~ group, dist='weibull' , data=mydatatry4) -> out.weib
>
> summary(out.weib)

Call:
survreg(formula = Surv(time, event) ~ group, data = mydatatry4,
    dist = "weibull")
                                  Value Std. Error      z        p
(Intercept)                      1.6038     0.0234  68.647 0.00e+00
groupVacuum Modulator-Padmini    0.1275     0.1962   0.650 5.16e-01
groupVACUUM MODULATOR - PADMINI  0.0251     0.0391   0.641 5.22e-01
groupVACUUM MODULATOR (EGR) R&R  0.1309     0.0253   5.183 2.18e-07
groupVACUUM MODULATOR (VGT) R&R  0.1498     0.0334   4.485 7.29e-06
groupVACUUM MODULATOR EGR        0.1531     0.0254   6.023 1.71e-09
Log(scale)                      -0.8308     0.0115 -72.143 0.00e+00

Scale= 0.436

Weibull distribution
Loglik(model)= -10190.2   Loglik(intercept only)= -10212.3
    Chisq= 44.31 on 5 degrees of freedom, p= 2e-08
Number of Newton-Raphson Iterations: 6
n= 4587
```

Figure 7. Summarized description for survreg

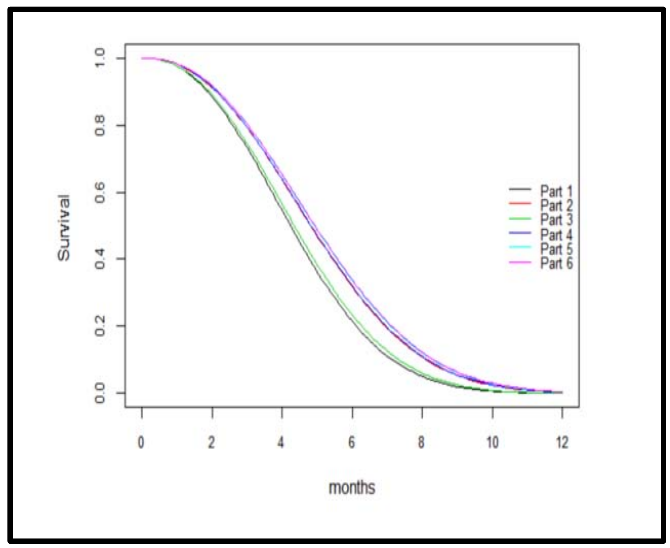

Figure 8. Weibull Survival Analysis

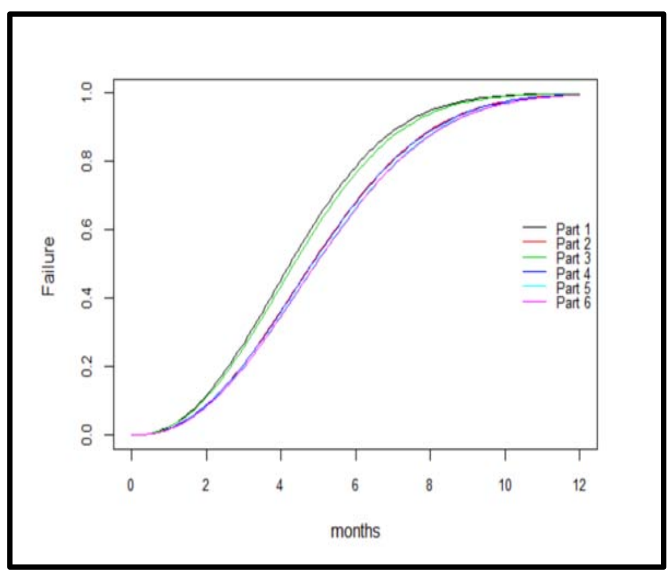

Figure 9. Weibull Failure Analysis

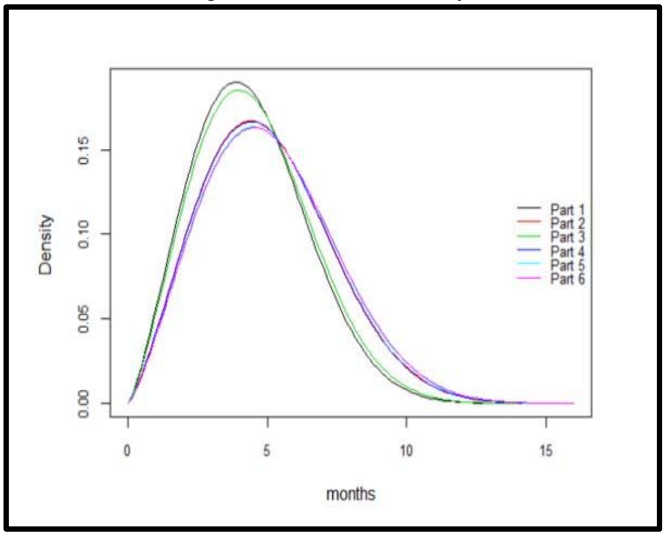

Fig. 10. Weibull Density

VI. CONCLUSION

- Modeling of automotive component and automobile vehicles' reliability from the simple statistical description, to the estimation of a reliability curve over time, to a proper mathematical model to fit the test data.
- Employing the Kaplan-Meier life curve permits us to compare the component reliability over time, and to evaluate the effect factors with statistical reliability.
- A Weibull model with two parameters (slope, β, and scale, η) can reasonably display the mean failure with a 'linear' model, while a Weibull model with three parameters can treat some 'nonlinearity' at earlier time stage much better.

VII. REFERENCES

[1] Kaplan, E. L. and Meier, P., Nonparametric Estimation from Incomplete Observations, J. of the American Statistical Association, 1958, p 261, 719-724.

[2] The R Core Team, R: A Language and Environment for Statistical Computing Reference Index, https://cran.r-project.org/doc/manuals/r-release/fullrefman.pdf

[3] Abernethy, Robert, The New Weibull Handbook, (3 rd Edition), 1999

