# An Application of Weibull and KM Analysis to Determine Failure and Survival Rates in Automobile Industry using R

Nirbhay Pherwani | BE | Vivekanand Education Society's Institute of Technology |
Mumbai University | Mumbai | India

## 1. Abstract

This paper focuses on the automobile vehicle, automotive component failure detection based on failure modes using Weibull analysis and other survival analysis techniques. Detailed attention is paid to three areas: 1) overall failure rates are described statistically first, and data cleaning and definitions of 'failed' data within spell period are made; 2) Kaplan-Meier life curves and Nelson Allen Curve are used to compare the failure mode reliability over time and explore factor related to the vehicle failure; 3) Weibull regressions, with three parameters, are applied to fit real-world reliability data from different test conditions to analyze the vehicle failure.

## 2. Introduction

Automotive components fail over time, especially beyond the warranty time period. It is important to study the reasons why automotive components fail, and to understand the component reliability, associated with various manufacturing, environmental and testing conditions.

Most reliability test data are closely time-related, and failures happened within the warranty time window or beyond, a technique of 'time-to-event' or survival analysis is very suitable for such reliability data, especially Weibull model is a well-established tool to fit the test data and to predict the future failure trend beyond the available test duration. The main goals of this project are to apply a Weibull model to the automotive component reliability analysis, and then explore the failure rate over service time or mileage, and to provide some statistically-based insights into the component failures by the following process:

- Provide descriptive summary of available failure data of a component 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.

## 3. Descriptive Summary of Data Sample

The Descriptive Summary contains information about the parts that eventually failed over time i.e. 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
>
```

## 4. Methods of Modeling Survival and Failure Rates

It is of great interest to observe the component failure rate, F(t), or from an opposite point of view, the survival probability varying over a test time, S(t), while there is a simply relationship between the two: 'S(t)= 1 – F(t)', i.e., 20% failures means 80% survival rate among a fixed sample. One of the most useful tools to compare the survival probability over time is a method proposed by Kaplan and Meier. The Kaplan-Meier survival curve is described by the following formula:

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

Where 'di' is 'deceased' subject or failed automotive component, and 'si' is the 'survivor' or 'alive' component, and 'ni ' is the total (both failed and suspension components) in the study at any moment beyond time zero. Or, turning the problem around, the failure probability over test time, F(t), equal to '1-S(t)', it can be further expressed by following Eq. (2) in Weibull model 2 :

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

Or, equivalently it can be visualized by the following 'linear' transformation, as Eq. (3):

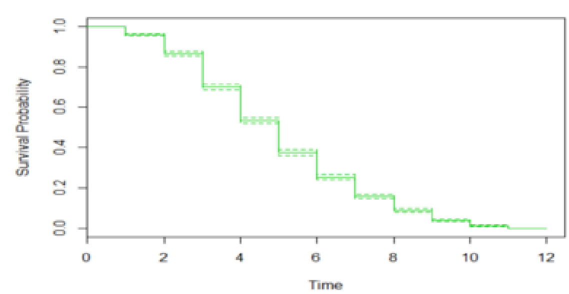
$$\log(-\log(S(t)) = \beta \log(t) - \beta \log(\eta) \qquad (3)$$

In the above Eq. (3), S(t) is survival function, which can be estimated from the Kaplan-Meier curve discussed earlier, and F(t) of Eq. (2) is the accumulation of failure probability as time increases, here 'β' is regarded as the 'Slope' of the 'linear' plot, or 'Shape' parameter, and 'η' is a 'Scale' parameter and is related to the intercept of the 'linear' plot. When a plot of test data is not visualized as a 'linear' plot as Eq. (3), especially at the earlier time stage, a Weibull with three parameters, as Eq. (4), provides a better data fitting, where a time shift, or threshold, t0, is included as Eq. (4) -

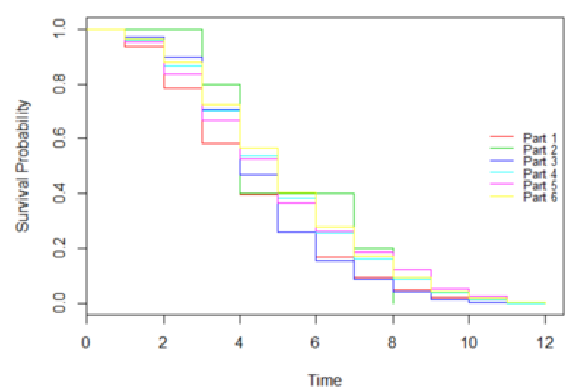
$$F(t) = 1 - e^{-((t - t_0)/\eta)^{\beta}} \qquad (4)$$

For the purpose of computations R programming is used.

## 5. Kaplan-Meier life curves (samples)




KAPLAN MEIER NON PARAMETRIC ANALYSIS




KAPLAN MEIER NON PARAMETRIC ANALYSIS BY GROUP

The package survival contains functions for carrying out survival analysis. The first step in any analysis is to use the Surv function to create a survival object. Now we need to specify two arguments: the variable that records the subject at the event time followed by the variable that identifies the nature of the event (failure or censored). The survfit function then uses the Surv object as the response variable in a formula expression to produce the Kaplan-Meier estimate of the survivor function. The standard approach is to create the Surv object and estimate the survivor function in a single expression. The component $surv of the survfit object is the Kaplan-Meier estimate while $upper and $lower components are the upper and lower 95% confidence limits. If we use the plot function on a survfit object, we get the plot of the estimated survivor function along with a 95% confidence band.

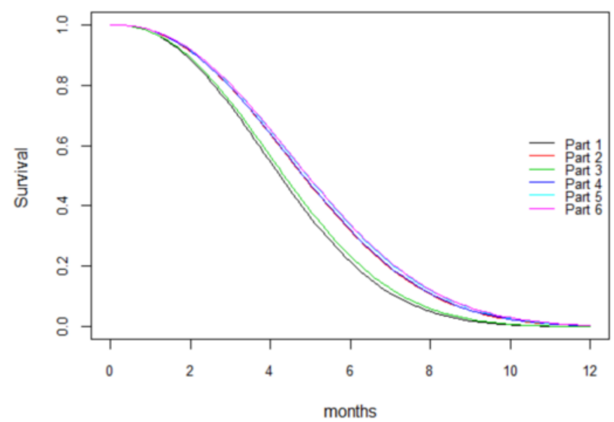
## 6. Weibull Modeling

**<u>Plotting the survivor function or and failure with the curve function</u>**

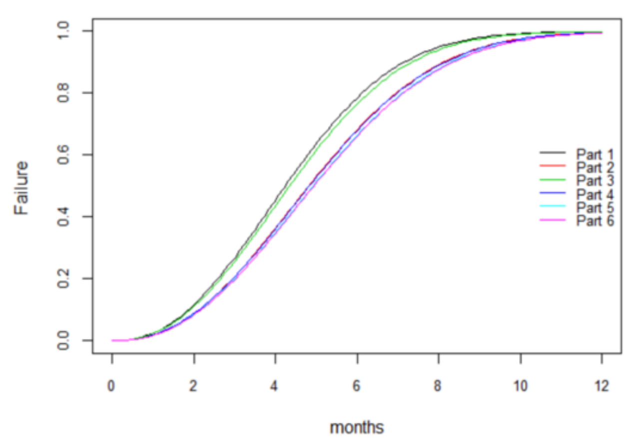
The labeling of the Weibull parameters in the survreg function even differs from that used by Weibull function in base R. The linear predictor is the estimate of what is usually called log scale, i.e. log λ, and what survreg calls the scale parameter is in fact the reciprocal of the Weibull shape parameter used in the base R Weibull functions. With those identifications understood, plotting the Weibull survivor function becomes straight-forward. We need to use the pweibull function with the argument lower.tail=FALSE, or equivalently, plot instead 1–pweibull.

In a similar fashion we can obtain estimates of the Weibull densities and the Weibull hazard functions in different situations.

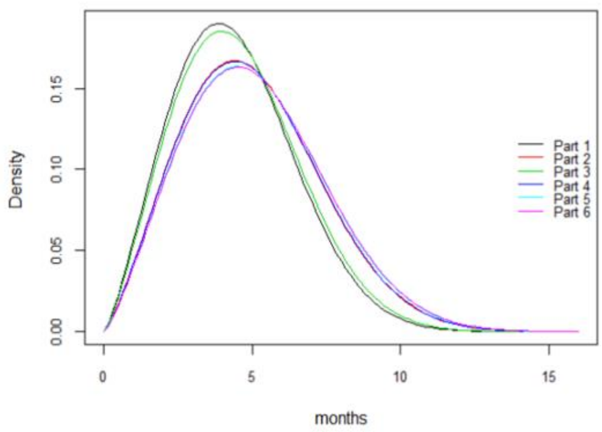
```
> survreg(Surv(time,event)~ group, dist='weibull' , data=mydatatry4) -> out.weib
>
> summary(out.weib)

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
```




**<u>WEIBULL SURVIVAL ANALYSIS</u>**




**<u>WEIBULL FAILURE ANALYSIS</u>**




WEIBULL DENSITY

## 7. Conclusion

- Modeling of automotive component and automobile vehicles' reliability from the simple statistical description, to 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.

## 8. References


1. https://sites.google.com/site/econometricsacademy/econometrics-models/
2. http://www-nrd.nhtsa.dot.gov/
3. http://www.unc.edu/courses/2010spring/ecol/562/001/docs/lectures/lecture24.htm
4. https://cran.r-project.org/manuals.html