**Weibull Distribution**
12/15/2025

In experimental data analysis, often one finds that the data distribution does not follow perfect normal distribution as suggested by the central limit theorem. Rather the data are heavily skewed, indicative of a possible non-normal distribution. Indeed, Weibull distribution is broadly used for data that measure “survival” rates of a subject, especially in material research, medicine, and life sciences.

Distribution Functions

A two-parameter ($\alpha > 0, \beta > 0$) Weibull cumulative distribution function (cdf) is defined as below:

$$cdf(x) = \begin{cases} 1 - exp\left[-\left(\frac{x}{\alpha}\right)^{\beta}\right] & x \geq 0 \\ 0 & x < 0 \end{cases}$$

where $\alpha$ is “scale” and $\beta$ “shape” of the function. The Weibull probability density function (pdf) is the derivative of cdf(x):

$$pdf(x) = \frac{\beta}{\alpha}\left(\frac{x}{\alpha}\right)^{\beta-1} exp\left[-\left(\frac{x}{\alpha}\right)^{\beta}\right]$$

The $m^{th}$ moment of the Weibull distribution can be computed using the following formula:

$$E\left(X^m\right) = \alpha^m \Gamma\left(1 + \frac{1}{\beta}\right)$$

Therefore,

$$\mu = E(X) = \alpha\Gamma\left(1 + \frac{1}{\beta}\right)$$

$$\sigma^2 = \alpha^2\left[\Gamma\left(1 + \frac{2}{\beta}\right) - \Gamma^2\left(1 + \frac{1}{\beta}\right)\right]$$

Parameter Estimation

One of the challenges in applying Weibull distribution to experimental data is the initial guess of the parameters. It is because the maximum likelihood estimation procedures are non-linear therefore extremely sensitive to the initial values to begin the non-linear search. Often, the procedure fails to start when the initial values are not sufficiently close to the objective function minimum. The method of moment (MoM) as mentioned above unfortunately does not provide adequate estimates for the initial values. Rather, the “double logarithmic regression” is still the tried and true practical approach. The Weibull cumulative distribution function, $cdf(x)$, is transformed as below:

$$ln\left[-ln\left(1 - cdf(x)\right)\right] = \beta ln(x) - \beta ln(\alpha)$$

The parameters, $\alpha$ and $\beta$, are solved using linear regression with $ln(x)$ as the explanatory variable and the double logarithmic term as the response variable.

Application Examples

Sandia National Lab published a report (*SAND81-8262*) in 1981 in which the wedge bond strength of a 1-mil gold wire was measured. The breaking strength was found to follow the Weibull distribution. The data and the Weibull model are shown in Figure 1. The scale parameter is found as 6.46 grams.

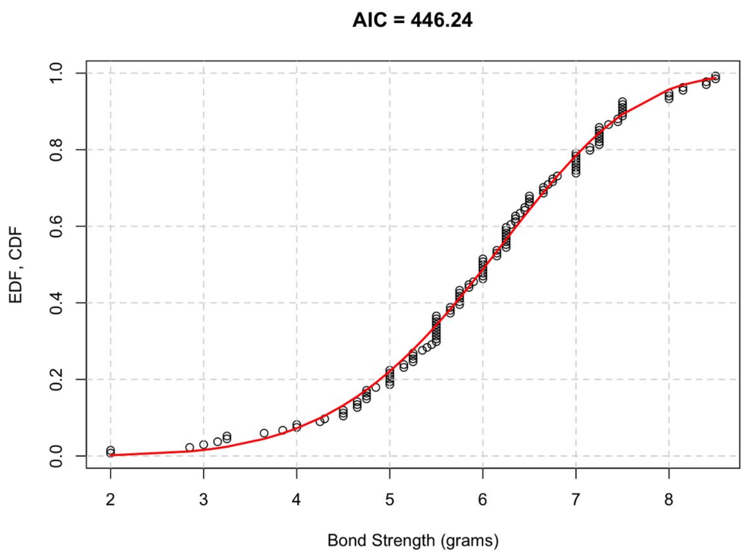


*Figure 1, 1-mil Gold Wedge Bond Strength*

| | Estimate | Std. Error |
|---|---|---|
| shape | *5.395364* | *0.3683935* |
| scale | *6.463973* | *0.1091287* |

The bladder cancer patient remission time was also found to follow the Weibull distribution (*Statistical Methods for Survival Data Analysis*, **vol 476**, *Wiley*, 2003). A single Weibull model is used first as shown in Figure 2. The model does not fully describe the data on the top left side of the graph.

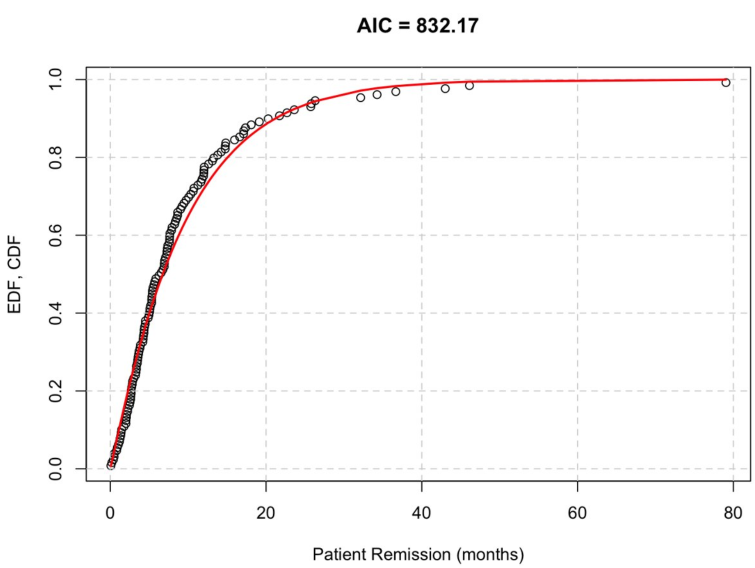


*Figure 2 Bladder Cancer Patient Remission Time.*

| | Estimate | Std. Error |
|---|---|---|
| shape | *1.047844* | *0.06757725* |
| scale | *9.558967* | *0.85265272* |

A bimodal Weibull model is used to improve the modeling accuracy as shown in Figure 3. The AIC reduced from 832 to 829. The original scale parameter of 9.55 month becomes one at 6.88 month and another at 14.2 month. The second mode comprises of 39% patient population.

In both models, the shape parameters are close to 1; therefore the data could also be described to follow the exponential distribution with almost constant hazard rates.

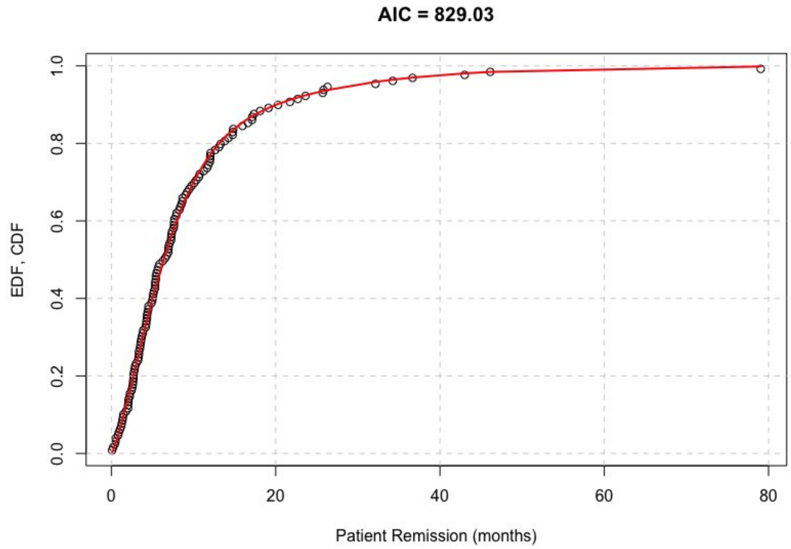


*Figure 3, Bladder Cancer Patient Remission Time.*

| | Estimate | Std. Error |
|---|---|---|
| sh1 | *1.4880357* | *0.4179047* |
| sh2 | *0.9967084* | *0.2066136* |
| sc1 | *6.8770633* | *1.1813058* |
| sc2 | *14.2438962* | *6.9812975* |
| p | *0.3915414* | *0.3317803* |

Laboratory guinea pig survival time after being infected by virulent tubercle bacilli (*Am. J. Hyg.* **vol 72**(1), 130-148 (*1960*)) is shown in Figure 4 with a Weibull model.

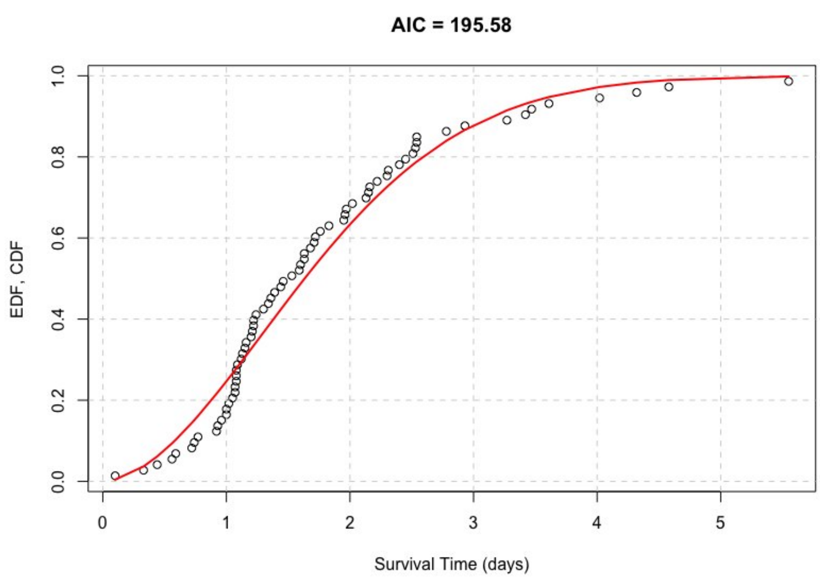


*Figure 4, Guinea Pig Survival Time*

| | Estimate | Std. Error |
|---|---|---|
| shape | *1.825357* | *0.1587044* |
| scale | *1.995885* | *0.1362873* |

Again, the data in Figure 4 are not fully described by the single Weibull model. A bimodal Weibull model is therefore introduced and shown in Figure 5. The AIC reduced slightly from 195.58 to 195.57.

The bimodal Weibull model suggests that approximately 19% of the guinea pigs showed a greater initial resistance to the virulent tubercle bacilli infection with a scale parameter of 3.6 days in comparison with the other guinea pigs with a scale parameter of 1.6 days..

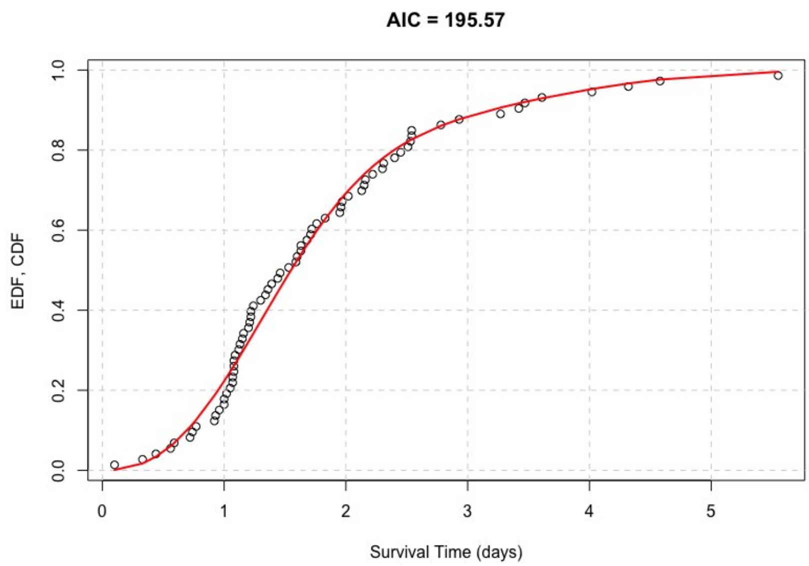


*Figure 5. Guinea Pig Survival Time*

| | Estimate | Std. Error |
|---|---|---|
| sh1 | *2.4386908* | *0.3344186* |
| *sh2* | *3.1270926* | *2.6147775* |
| sc1 | *1.6061072* | *0.2163924* |
| sc2 | *3.6299954* | *1.4587202* |
| p | *0.1872351* | *0.2212426* |

## Concluding Remarks

Weibull distribution is an important tool in studying the survival behavior of the research subject, in material research, medicine, and life sciences. Often, a single distribution may not completely explain the population; a multimodal model may be required before taking on more exotic models that are difficult to explain. Due to the non-linearity of the maximum likelihood estimator, the choice for the initial guess of the parameters becomes tricky for the success or failure of the model. The tried and true “double logarithmic regression” has proven effective. A R function $guess(x)$ is listed below and $x$ is the vector of experimental data.

```
cdf=function(x){cdf=x; for (i in 1:length(x)) cdf[i]=i/(len+1);
  return(cdf)}
guess=function(x){y=log(-log(1-cdf(x))); lx=log(x); m=lm(y~lx);
  list(shape=m$coefficients[2],
  scale=exp(-m$coefficients[1]/m$coefficients[2]))}
```