Estimator comparison

By: S. Civit

Contents

```
## S'està carregant el paquet requerit: printr

## Registered S3 method overwritten by 'printr':
##   method                from
##   knit_print.data.frame rmarkdown
```

Introduction

For the uninitiated, Weibull analysis is a method for modeling data sets containing values greater than zero, such as failure data. Weibull analysis can make predictions about a product's life, compare the reliability of competing product designs, statistically establish warranty policies or proactively manage spare parts inventories, to name just a few common industrial applications. In academia, Weibull analysis has modeled such diverse phenomena as the length of labor strikes, AIDS mortality and earthquake probabilities.

Learning by example: Weibull estimators method (MRR)

Example:

Imagine that you work for a toy company that wants to compare the reliability of two proposed designs (second part of study: out of scope now) for a jack-in-the-box spring housing (a children's toy consisting of a box with a model of a person inside it that jumps out and gives you a surprise when the top of the box is raised.

The desired **reliability at 400,000 cycles is 0.90**. In other words, the toy company would like 90 percent of the spring housings to survive at least 400,000 cycles. This reliability goal is expressed mathematically as **R(400,000) 0.90**. Ten units were assembled with each of the two housing designs (Design A and Design B). These 20 units were tested until their spring housings failed.

The data (see Excel file) don't clearly indicate whether either design meets the desired reliability goal. Both designs had at least one failure before 400,000 cycles, yet clearly the average number of cycles before failure exceeds 400,000 for both designs.

A comparison of sample averages using a Student's t test reveals no statistical difference between the average cycles for Design A and the average cycles for Design B (p-value = 0.965). But as a simple measure of central tendency, the sample average gives no information about the spread or shape of the distribution of failure times.

- Could the two designs averages be the same, but their reliability be quite different? (out of scope)
- How can you be more scientific about comparing the reliability of the two proposed designs? (out of scope)

MRR estimator method by Excel file

Modeling the data using Weibull analysis requires some preparation. Open **Excel file (Weibull.xls)** and focus on the data from Design A.

Estimating Weibull Parameters

Why can we expect the graph of the **ln(Cycles) vs. the transformed median ranks** to plot as a straight line?

With some effort, the Weibull cumulative distribution function can be transformed so that it appears in the familiar form of a straight line $Y = mX + b$ Here's how:

Comparing this equation with the simple equation for a line, we see that the left side of the equation corresponds to Y, β corresponds to m, and $\beta ln\alpha$ corresponds to b. Thus, when we perform the linear regression, the estimate for the Weibull Beta parameter comes directly from the slope of the line. The estimate for α parameter must be calculated as follows:

$$\hat{\alpha} = e^{-\left(\frac{b}{\beta}\right)}$$

Interpreting the results

The Weibull shape parameter, called β, indicates whether the failure rate is increasing, constant or decreasing. A β <1.0 indicates that the product has a decreasing failure rate. This scenario is typical of "infant mortality" and indicates that the product is failing during its "burn-in" period. A Beta=1.0 indicates a constant failure rate. Frequently, components that have survived burn-in will subsequently exhibit a constant failure rate. A $\beta > 1.0$ indicates an **increasing failure rate**. This is typical of products that are wearing out. Such is the case with the spring housings-both designs A and B have Beta values much higher than 1.0. The housings fail due to fatigue, i.e., they wear out.

The Weibull characteristic life, called, is a measure of the scale, or spread, in the distribution of data. For example, with Design A housings, about 37 percent of the housings should survive at least 693,380 cycles.

While this is interesting, it still doesn't reveal whether either jack-in-the-box design meets the reliability goal of R(400,000) 0.90. For this, you need to know the formula for reliability assuming a Weibull distribution:

$$R(t) = e^{-\left(\frac{x}{\hat{\alpha}}\right)^{\hat{\beta}}}$$

where x is the time (or number of cycles)until failure

$$R(400000) = e^{-\left(\frac{400000}{693380}\right)^{4.25}} = 0.908$$

Statement

We will look at how we can begin to **determine estimates of the parameters** for each lifetime distribution, based on test data. These estimates can then be used to construct reliability functions and plots, as well as other life data statistics, such as the MTBF.

The simplest and longest-used method for parameter estimation is that of probability plotting.

This methodology involves plotting the failure times on a specially-constructed plotting paper to determine the fit of the data to a given distribution and, if applicable, estimates of the distribution's parameters.

However the main purpose of this example, is **to compare the Weibull parameters estimation by MRR and MLE**

MRR Median Ranks Regression

The Median Ranks method is used to obtain an estimate of the unreliability for each failure. The median rank is the value that the true probability of failure, $Q(T_j)$, should have at the j^{th} failure out of a sample of N units at the 50% confidence level.

The rank can be found for any percentage point, P, greater than zero and less than one, by solving the cumulative binomial equation for Z. This represents the rank, or unreliability estimate, for the j^{th} failure in the following equation for the cumulative binomial:

$$P = \sum_{k=j}^{N} \binom{N}{k} Z^k (1-Z)^{N-k}$$

where N is the sample size and j the order number.

The median rank is obtained by solving this equation for Z at $P = 0.50$:

$$0.50 = \sum_{k=j}^{N} \binom{N}{k} Z^k (1-Z)^{N-k}$$

For example, if $N = 4$ and we have four failures, we would solve the median rank equation for the value of Z four times; once for each failure with $j = 1, 2, 3and4$. This result can then be used as the unreliability estimate for each failure

Benard's Approximation for Median Ranks

Another quick, and less accurate, approximation of the median ranks is also given by:

$$MR = \frac{j - 0.3}{N + 0.4}$$

This approximation of the median ranks is also known as Benard's approximation.

Simulation study: Pipeline

the goal of this simulation study will be to analize ** MRR Weibull estimator**.

We'll perform a simulation study under the following simplifying conditions:

- **Generate (pipeline)**:
  - Generating a iid sample size **n** from a **Weibull two parameter**
  - α (small, medium large)
  - β (small, medium large)
  - n (small, medium large)
- **Analyse (pipeline)**:
  - **MRR estimation** MLE estimation (optional)
- **Sumaryze (pipeline)**:
  - **Bias, MSE, precision,**
- **Results (pipeline)**:
  - Summarize in a table and in a new section (Results) the most relevant results
- **Conclusions (new section)**:
  - Using bullet points the most important conclusions.

Write a report according to ADMEP sections