Analysis of Reliability Data:

Life Distribution or Recurrence Analysis?

David Trindade, Ph.D.
David.Trindade@bloomenergy.com
QPRC 2016 Tempe, AZ June 2016

Abstract

In analyzing reliability data from ***repairable systems***, engineers may incorrectly model by fitting a **lifetime distribution** to the **times between failures**. However, such an analysis, which **disregards the time order of the observations**, can lead to invalid conclusions. **Recurrence analysis** methodology provides appropriate techniques for modeling reliability data from repairable systems.

This talk will include examples, illustrated using JMP software, of both the improper application of the lifetime distribution approach to repairable system data and the discovery analysis resulting when the correct recurrence analysis is applied. This discussion will also highlight some **limitations of the "mean time between failures"** (**MTBF**) as a reliability summary statistic and describe how MTBF values reported in the literature can be artificially inflated.

Objectives

- Describe the differences in the analysis of reliability data from ***non-repairable*** and ***repairable*** systems.
- Illustrate several applications using JMP's **Reliability Analysis** platform.
- Discuss some limitations on the use of the summary statistic **MTBF** as a measure of reliability.

Definition of a Repairable System

A system is ***repairable*** if, following a failure at some time ***t***, it can be **restored to satisfactory operation** by any action, excluding replacing the entire system.

Examples of Repairable Systems

Repairable Systems include:

- Automated Manufacturing Equipment
- Personal Computers
- Airplanes, Trains, Trucks
- Automobiles
- TV’s
- Network Servers
- Robots
- Software programs

Examples of Repair Actions

Restoring methods include:

- Replacing failed components, e.g. circuit boards
- Rebooting a computer
- Correcting a software bug
- Adjusting settings
- Swapping parts
- Automatic switchover to a redundant component
- Reestablishing electrical power
- Sharp blow with an object (Fonzie approach!)

Quantitative Measures for the Reliability of Repairable Systems

Age of System at Repair

The total accumulated running hours, days, cycles, miles, etc. on a system.

We'll use **time** in hours or days for our analyses.

Times Between Repairs

Called **interarrival** times.

We'll assume actual **repair times** are **negligible** compared to **operating hours**. Hence, downtime and availability will not be considerations in this discussion.

Reliability of Repairable Systems

Function of many factors

- Basic system design
- Operating conditions
- Environment
- Software robustness
- Type of repairs
- Quality of repairs
- Materials used
- Suppliers

Many factors can **vary** during system operation.

Key Property of Repairable Systems

FAILURES OCCUR SEQUENTIALLY IN TIME.

Non-Repairable Vs. Repairable Systems

For non-repairable components, the **usual** assumption made for analysis is that the **times to failure** are ***independent*** and ***identically distributed*** **(*iid*)** random samples from a ***single population of lifetimes*****.**

For repairable systems, we will show that it is a poor practice to **group together** all times between failures for analysis, disregarding any **occurrence order** of the data.

Are *iid* assumptions justified for repairable systems?

Case Study Example

Production Equipment:

A new design for a manufacturing system is based on a **single**, **replaceable *component***, a circuit board. Upon system failure, repairs are made by replacing the failed board with a **new**, **identical** board from a stockpile.

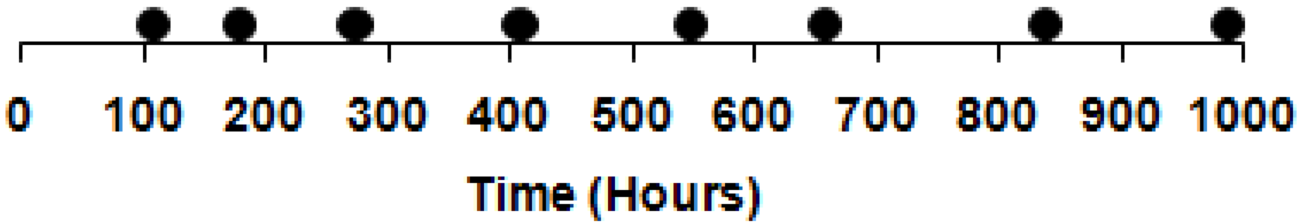
Objective of Analysis:

Engineers wanted to model the reliability of the system based on failure data obtained during the first 1,000 hours (i.e., 42 days) of operation under **accelerated** conditions.

System Repair History

Repairs (board replacements) were done at system ages 108, 178, 273, 408, 548, 658, 838, and 988 (hours).

A dot plot of repair times is shown below.



Case Study Weibull Analysis

The engineers thought it was reasonable to analyze the data using the **time to failure for each replaced board**, that is, the operating **hours between repairs,** and to treat those times as a group of *independent observations* from a ***single*** *lifetime distribution.*

In doing so, the actual **order** in which the times between failures occurred (i.e., **age of the system at the time of board replacement**) was ignored.

Weibull analysis methods were used for modeling. This common analysis of times between failures is often incorrect and misleading.

Interarrival Times

The board times to failure, based on the times ***between*** repairs, the **interarrival times,** are calculated below:

Repair Time (System Age)	Time Between Repairs (Interarrival Times)
108	108
178	178-108 = 70
273	273-178 = 95
408	408-273 = 135
548	548-408 = 140
658	658-548 = 110
838	838-658 = 180
988	988-838 = 150

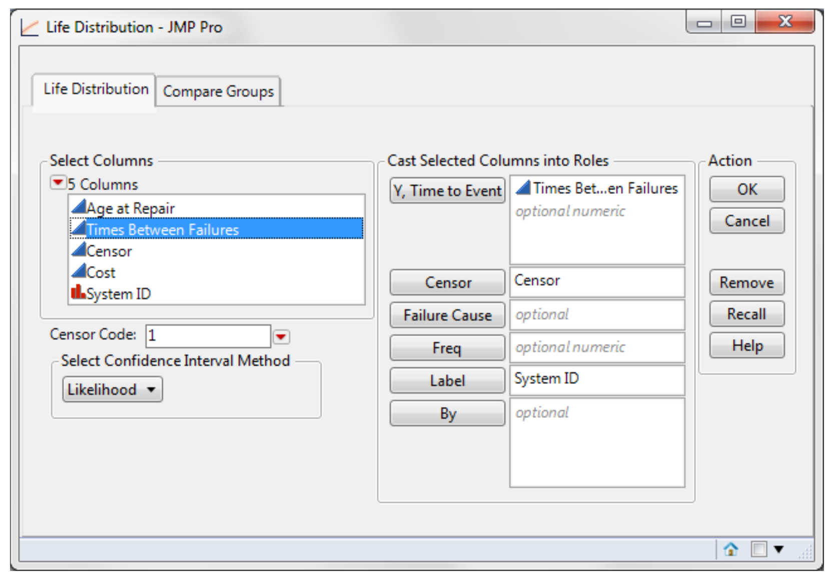
JMP Data Table and Life Distribution Analysis

Production Equipment Data Table

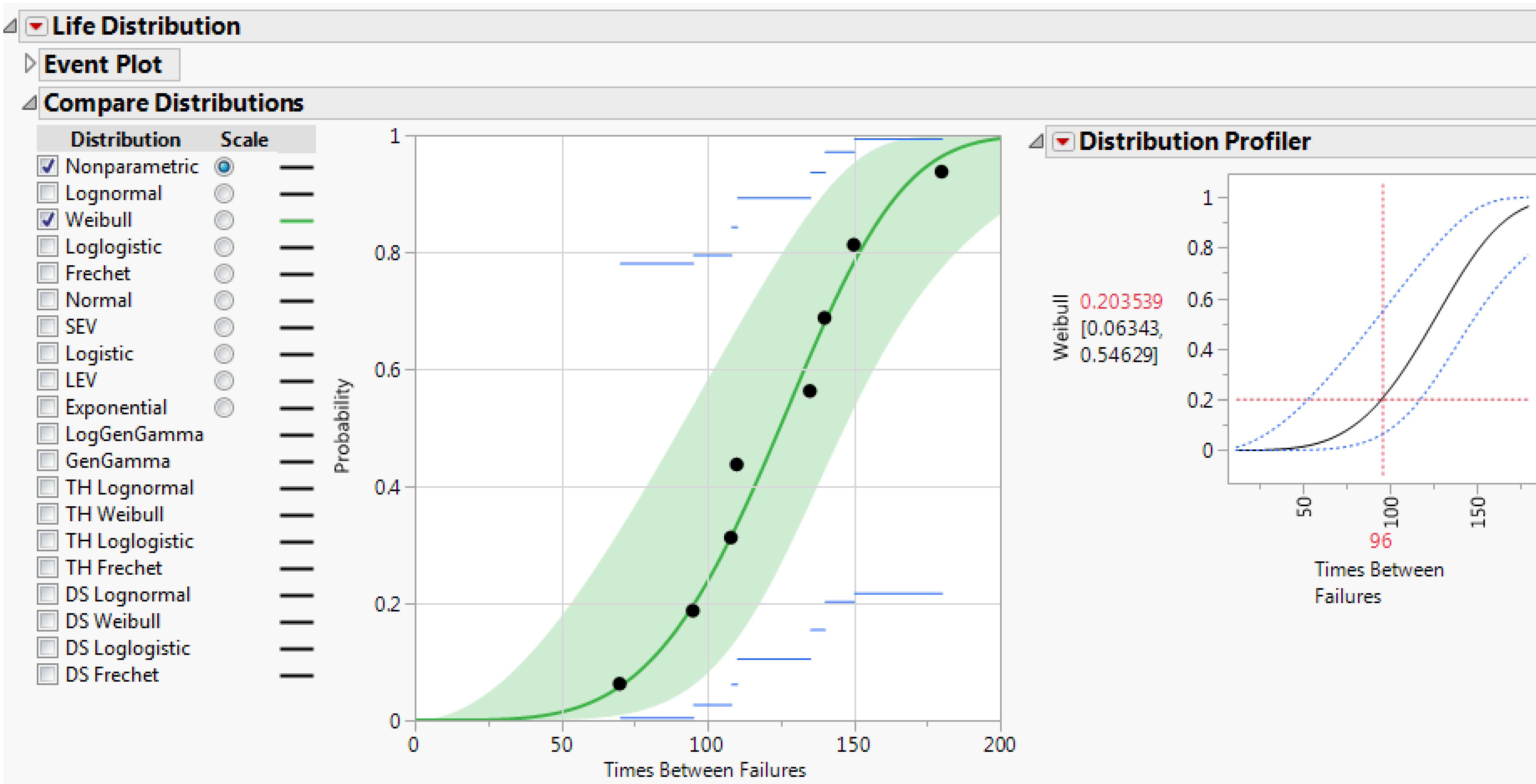
	Age at Repair	Times Between Failures	Censor
1	108	108	0
2	178	70	0
3	273	95	0
4	408	135	0
5	548	140	0
6	658	110	0
7	838	180	0
8	988	150	0
9	1000	12	1

The **Life Distribution** platform in JMP [1] is run using Times Between Failures for **Y, Time to Event**

Censor = 0 represents a **failure**
Censor = 1 is a ***suspension*** (unit still operating at test end)

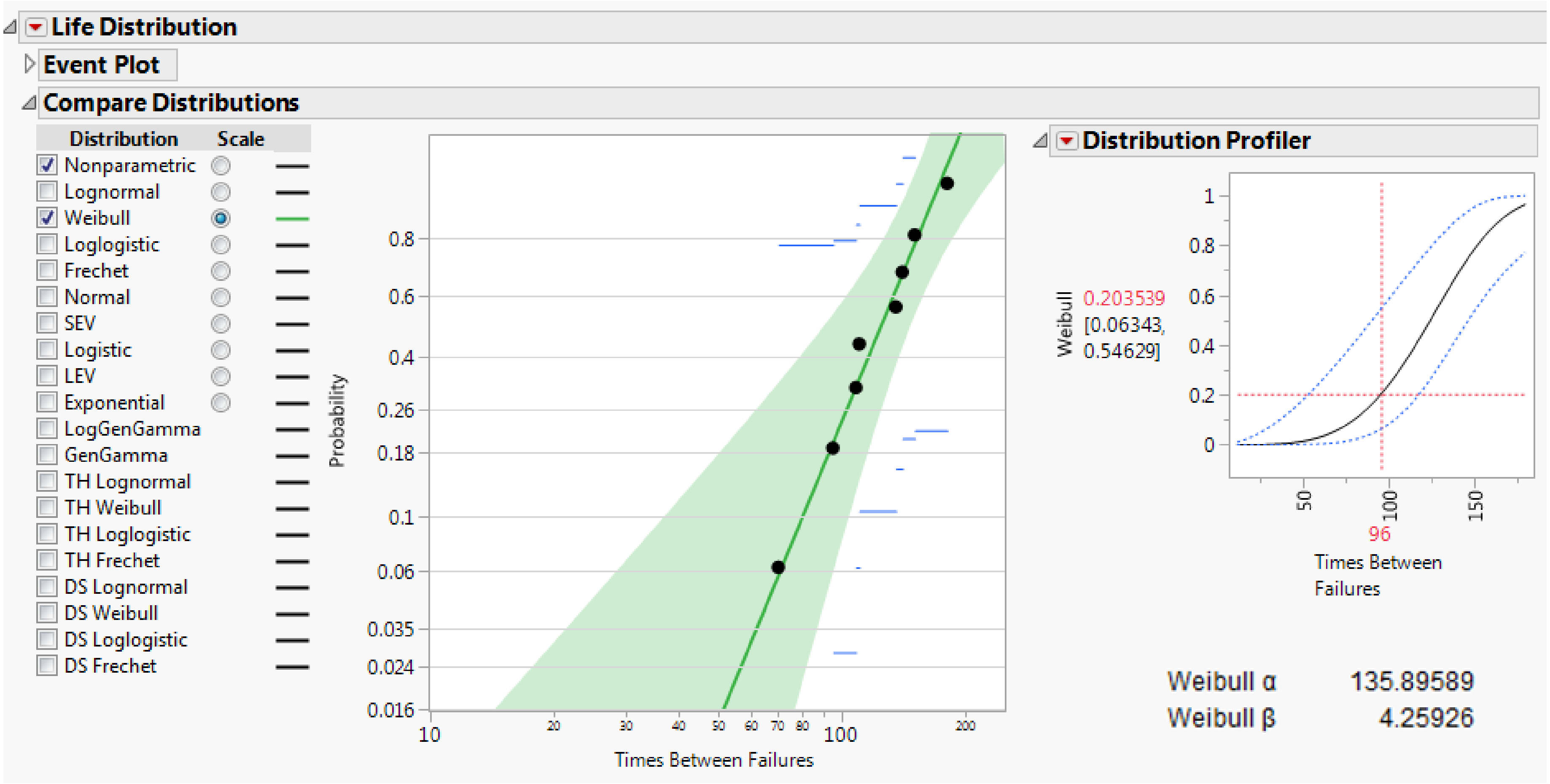


JMP Life Distribution Analysis: CDF Plot



The Weibull distribution appears to be a good fit to the data.

JMP Life Distribution Analysis: Probability Plot



Weibull probability plot shows data points falling close to a straight line.

Estimated Weibull Parameters

The Weibull parameter estimates show a **characteristic life α** ≈ 136 hours and a **shape parameter β** ≈ 4.3.

For the Weibull distribution, $\beta > 1.0$ indicates an **increasing hazard rate (instantaneous failure rate)**.

Statistics

Model Comparisons

Distribution	AICc	-2Loglikelihood	BIC
Weibull	84.311856	78.311856	82.706305

Summary of Data

Nonparametric Estimate

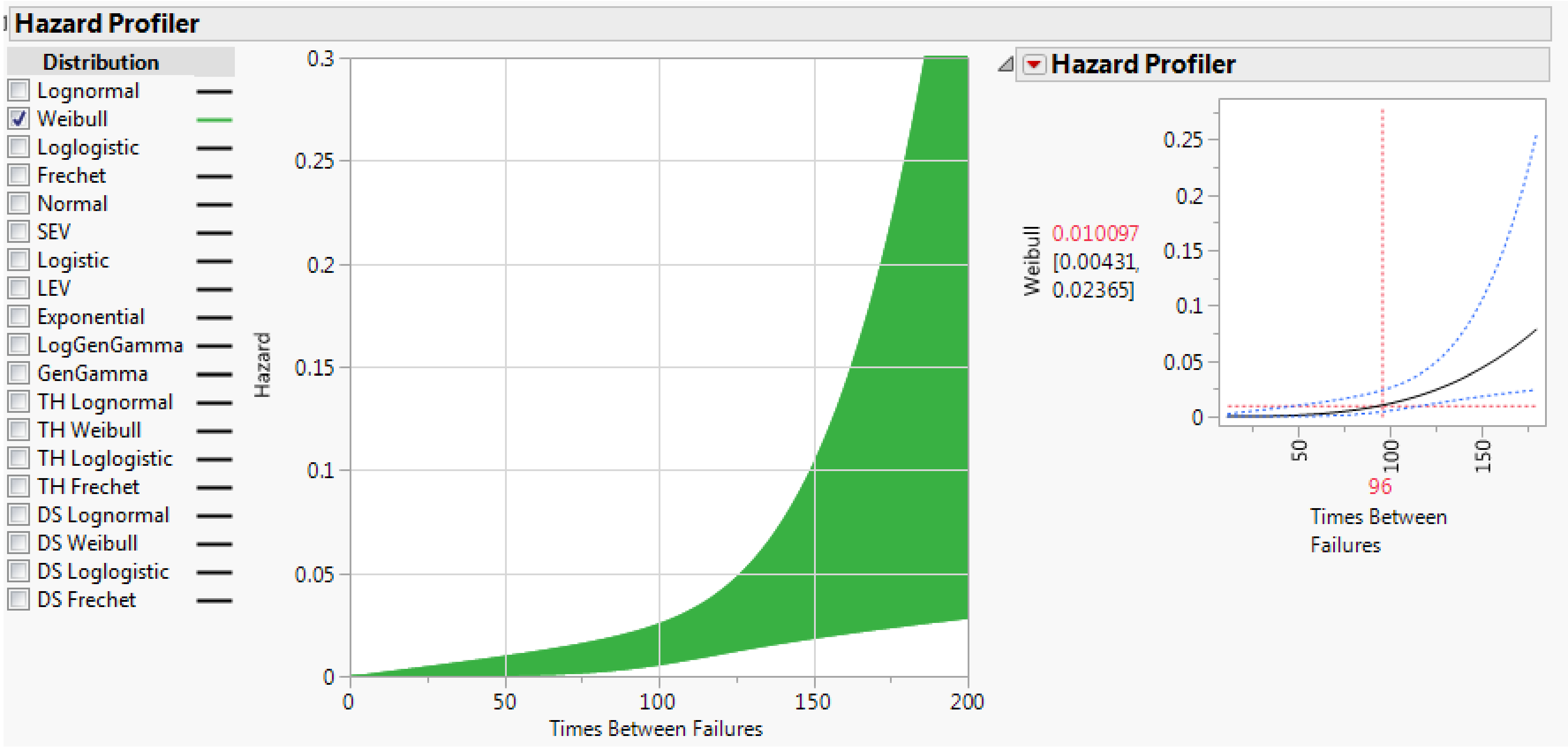
Parametric Estimate - Weibull

Parameter	Estimate	Std Error	Lower 95%	Upper 95%
location	4.91189	0.087661	4.70942	5.10195
scale	0.23478	0.064657	0.14529	0.43808
Weibull α	135.89589	11.912835	110.98812	164.34286
Weibull β	4.25926	1.172967	2.28267	6.88273
Mean (Wald CI)	123.62069	11.618392	102.82333	148.62458

Criterion	
-2*LogLikelihood	78.311856
AICc	84.311856
BIC	82.706305

Weibull Hazard Rate Profiler

JMP's Hazard Profiler shows an increasing hazard rate.



Engineers' Interpretation of Analysis

Engineers concluded times between repairs were well modeled by a Weibull distribution.

The estimated Weibull shape parameter, $\beta > 1$, indicated an ***increasing*** "failure rate."

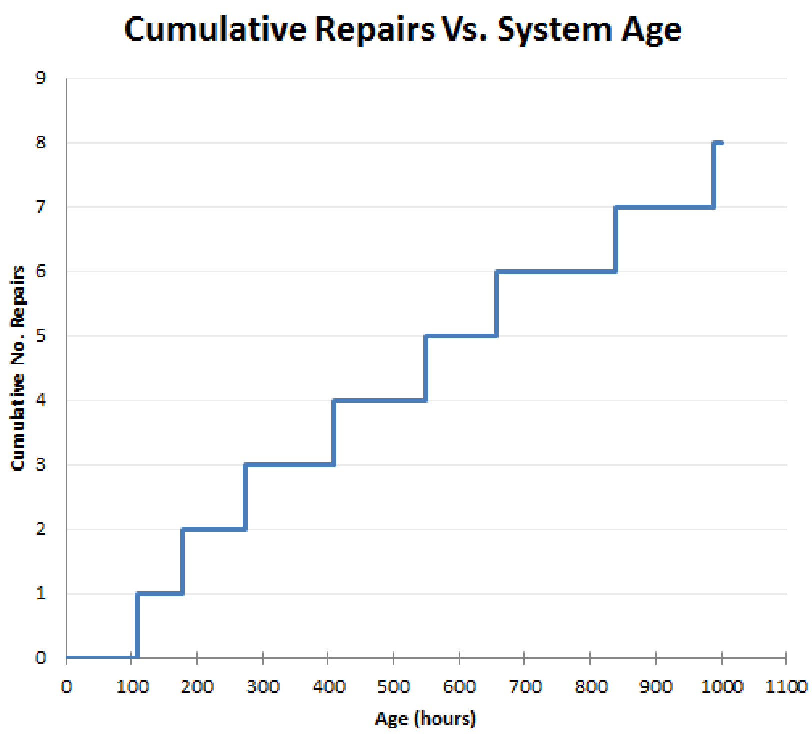
The inference was that the equipment had worsening reliability and should be considered for additional redesign, repair, or maintenance.

Were these conclusions justified or misleading based on the analysis methods applied?

We shall show that this analysis was not appropriate for such data and yielded erroneous conclusions.

Cumulative Plot for Analysis of Repairable Systems

- A very useful graph for the reliability analysis of repairable data is the **cumulative plot**: the **cumulative number** of repairs is plotted against the system **age**. [2,7]
- For this example, the system experienced board replacements at the following **ages**: 108, 178, 273, 408, 548, 658, 838, and 988 hours.



JMP Recurrence Analysis Platform

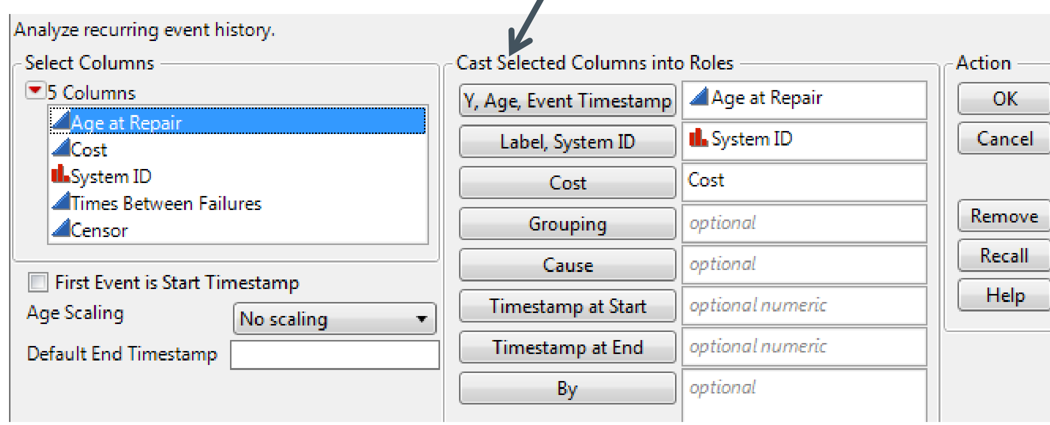
The **Recurrence Analysis** platform in JMP is run using Age at Repair for **Y, Age at Event**

Data Table

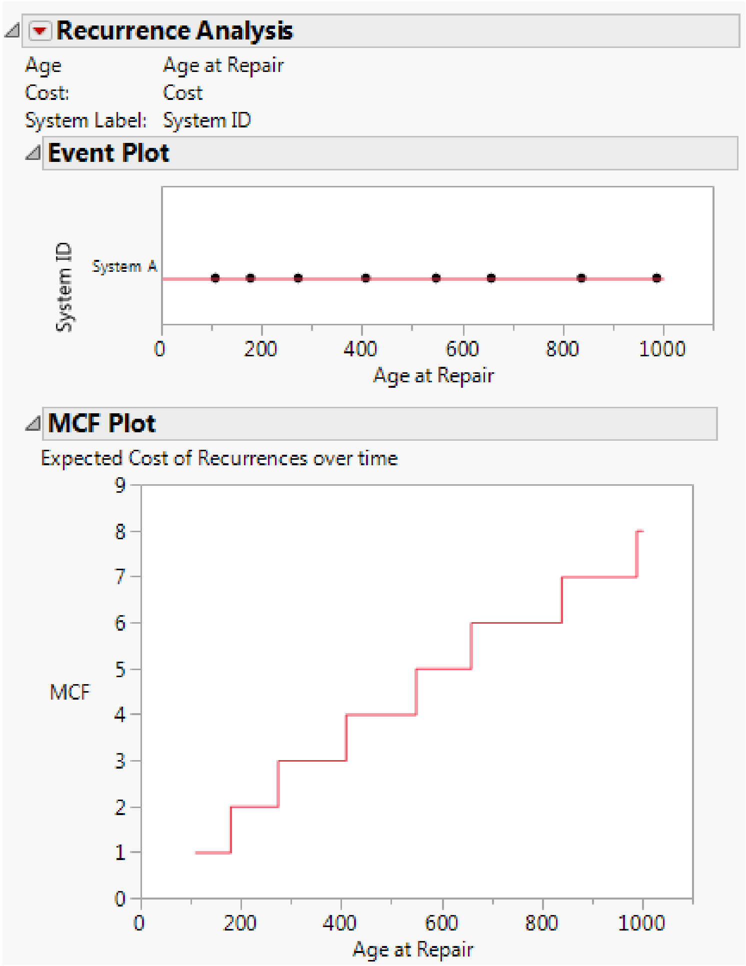
	Age at Repair	Cost	System ID
1	108	1	System A
2	178	1	System A
3	273	1	System A
4	408	1	System A
5	548	1	System A
6	658	1	System A
7	838	1	System A
8	988	1	System A
9	1000	0	System A

Cost = 1 represents a repair,
Cost = 0 is a **censored** observation.

Recurrence Analysis Input Box



Cumulative Plot of Number of Repairs Vs. Age



JMP shows both an **Event Plot** and a ***Mean Cumulative Function*** or **MCF** Plot. [2,3,4,7]

MCF plot for a single system is the **cumulative plot.**

The plot shows **no evidence of system reliability getting worse** with system age.

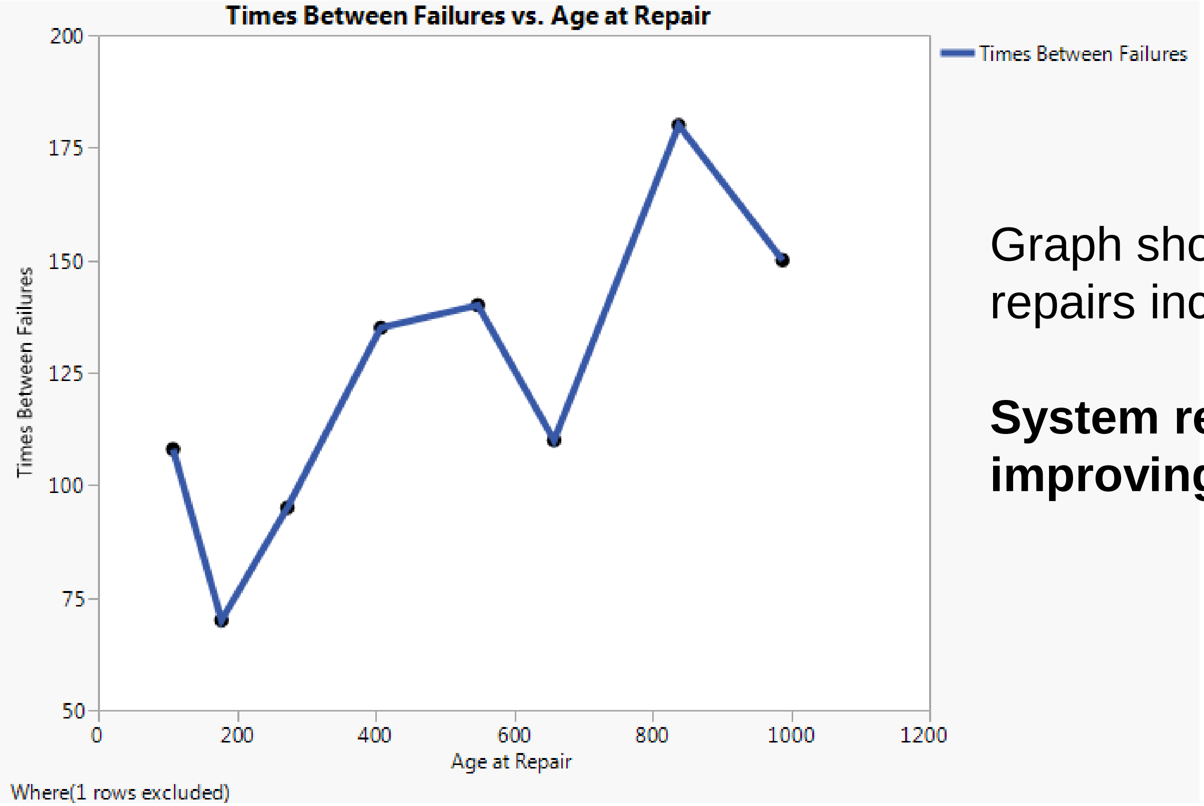
Let's consider the ***sequential* times between repairs**.

Sequence of Failure Times in Repairable Systems

- If the times between successive failures are getting **longer**, then the system reliability is **improving**.
- Conversely, if the times between failures are becoming **shorter**, the reliability of the system is **degrading**.
- Thus, the ***sequence*** of system failure times is very important.
- If the times show no trend (relatively **stable** behavior), the system may be neither improving or degrading, which may be suggestive of a **renewal process** with **constant mean time between failures** (**MTBF**). Following failures, system is restored to good as new.

Plot of Times Between Repairs Vs. Age

The plot of the sequential **times between repairs** versus the **system age** at repair shows an **increasing trend**.



Graph shows times between repairs increasing.

System reliability is actually improving!

Results/Implementation

- Analysis of the **repairable** system data using **non-repairable** Weibull analysis methods produced a **false conclusion**.
- Wrong interpretation was caused by the neglect of the occurrence order of failures in Weibull analysis.
- The correct **recurrence analysis** showed an **improving** trend in the repairable system history.
- With the correct analysis, engineers avoided unneeded and expensive corrective actions and instead focused on finding the reasons for the improvement.

Analysis of Multiple Repairable Systems

Data may come from **many similar systems** possibly subjected to **multi-censoring**.

Examples:

- Servers installed in the field at **different dates** throughout the year will have **different ages** at a specified calendar date.
- For autos, vehicles sold on the **same date** can have **different mileages** at the time of analysis.

Reliability Issues for Multiple Systems

Warranty analysis seeks answers to:

- What's the **mean number** of repairs by age *t*?
- What's the **repair rate** for all systems at age t?
- What's the **variation** in the mean number of repairs at a given age?
- What's the expected **age to first repair**? To ***k*th** repair?
- What is the mean repair cost?
- Are costs of repairs increasing or decreasing?
- Are **spare parts** adequate?
- Are there any **location dependent** issues?

Methods for Analysis of Multi-System Data

Davis in a 1952 paper [5] analyzed the number of miles between successive major failures of bus engines by comparing **distributions of interarrival miles** to **first** failure, **between first and second** failures, and so on.

He found that the average inter-repair miles to be decreasing. Early interarrival miles were nearly normal, but later miles were more exponentially distributed.

Davis 1952 Bus Engine Data

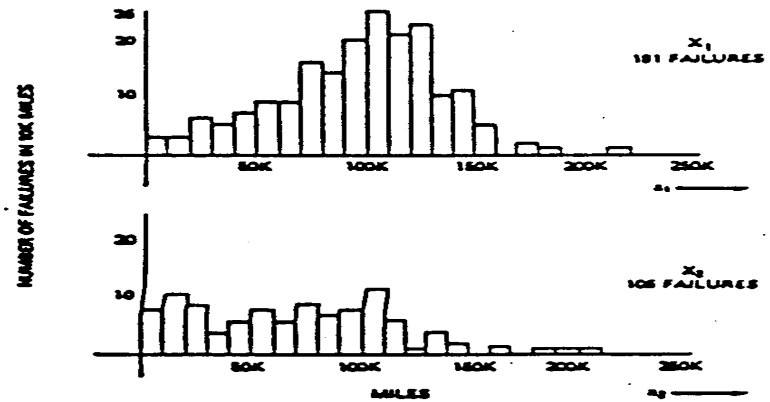


Figure 5-4 Interarrival miles of first and second failures of bus engines (191 engines originally put into service).

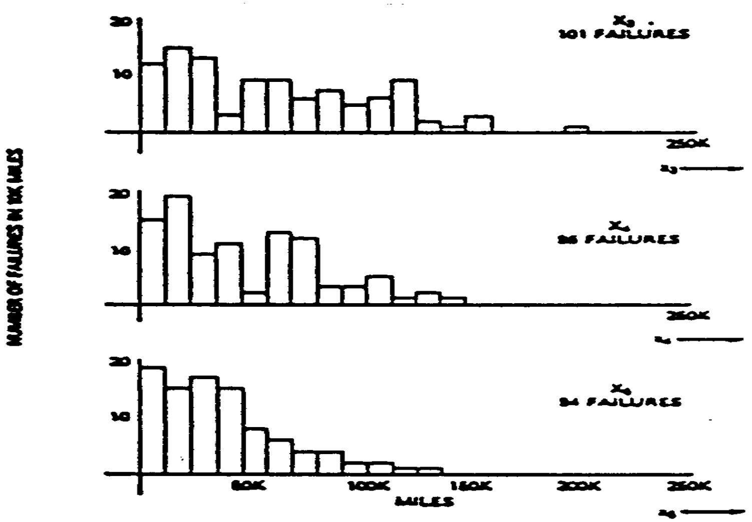


Figure 5-5 Interarrival miles of third, fourth and fifth failures of bus engines (191 engines originally put into service).

Miles to 1st failure
(closer to normal distribution).

Miles between 1st and 2nd failures.

Miles between 2nd and 3rd failures.

Miles between 3rd and 4th failures.

Miles between 4th and 5th failures.

(closer to exponential distribution)

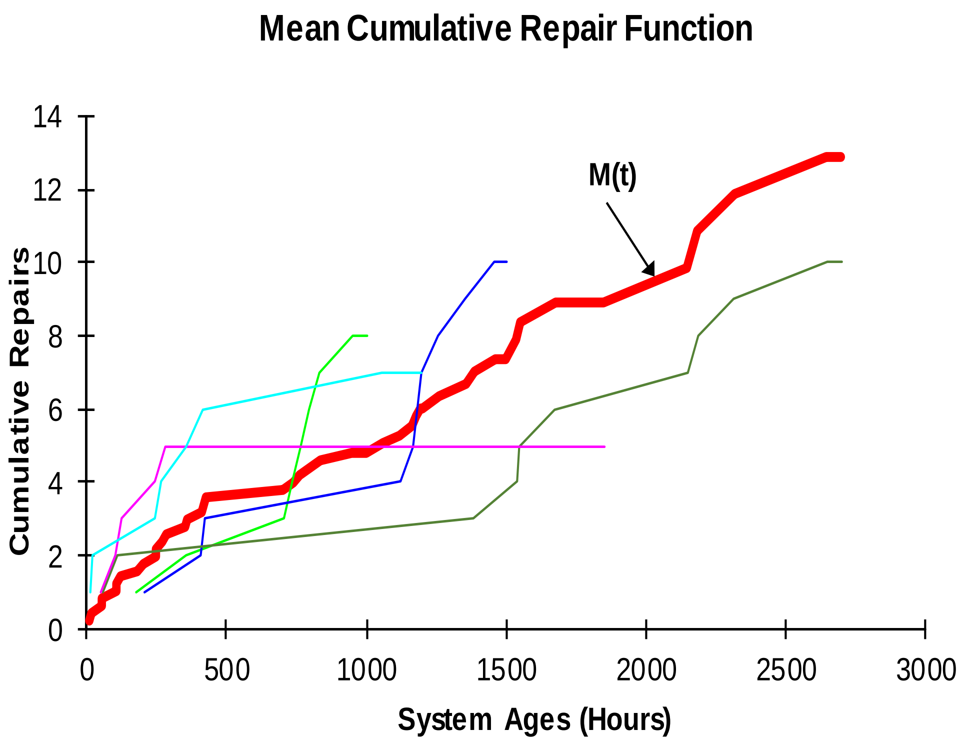
The Folly of Combining Davis Data

- It’s clear that the distributions of the miles between sequential repairs are different.
- Had all the data been combined together into one group and treated as a single population of lifetimes for analysis, the results would have been incorrect and misleading.
- **Neglecting the order of occurrence of the repairs can lead to invalid conclusions.**

Limitations of Davis Approach

- Method required considerable historical data on many systems.
- **No overall predictive models** were generated for the
  - **mean number of repairs** versus the bus **age** in miles
  - the **age** in miles to the ***k*th repair**

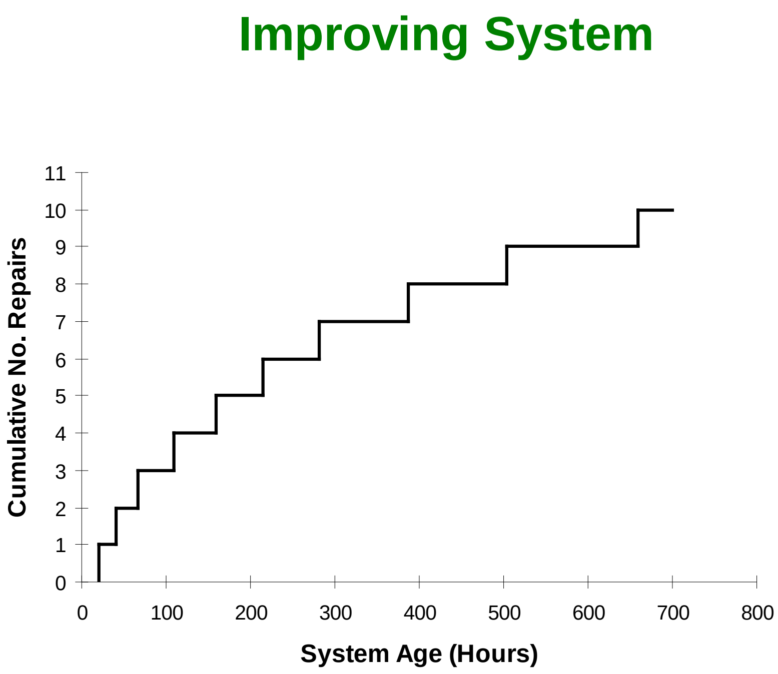
MCF Model for Multiple Systems

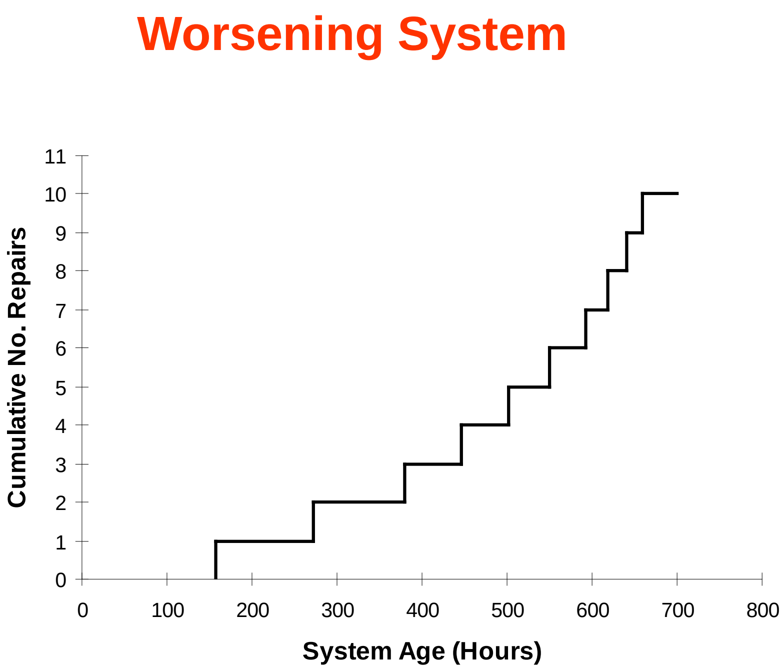


Consider a group of systems subject to repair actions. Represent **individual** repair histories *N*(*t*) using connecting lines between repairs, referencing all starting times back to zero. Example plot for **five** systems is shown.

MCF M(t) is **Mean Cumulative** (or Cost) **Function** = average number of repairs per system versus age. [2,3,4,7].

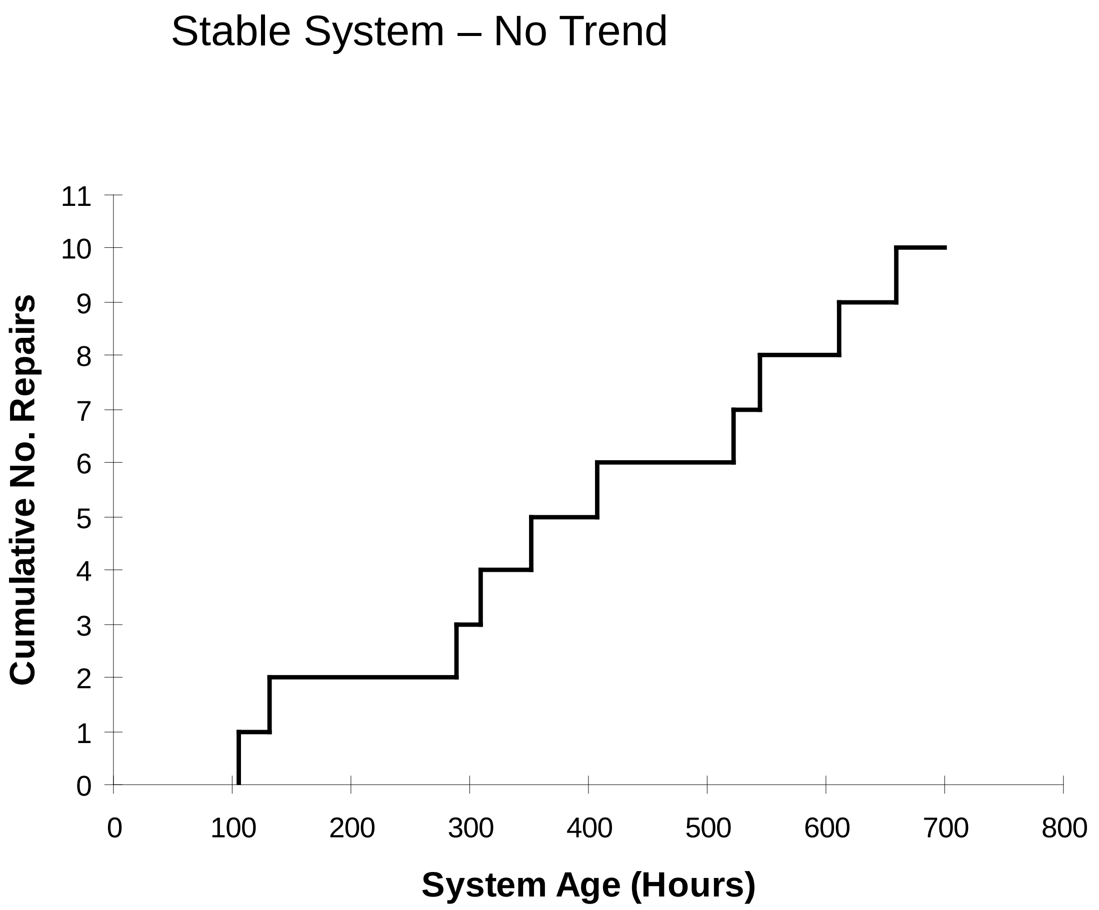
MCF Vs. Age Reveals Trends in Reliability





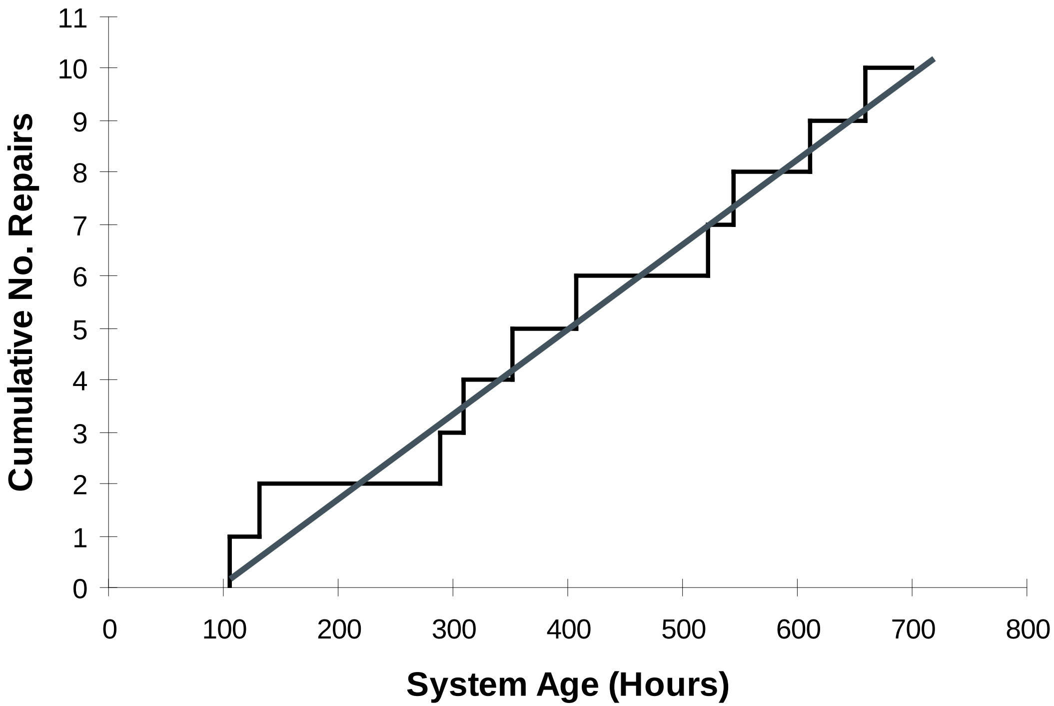
A single MTBF no longer applies as a valid measure.

MCF Vs. Age Reveals Trends in Reliability



MCF Vs. Age Reveals Trends in Reliability

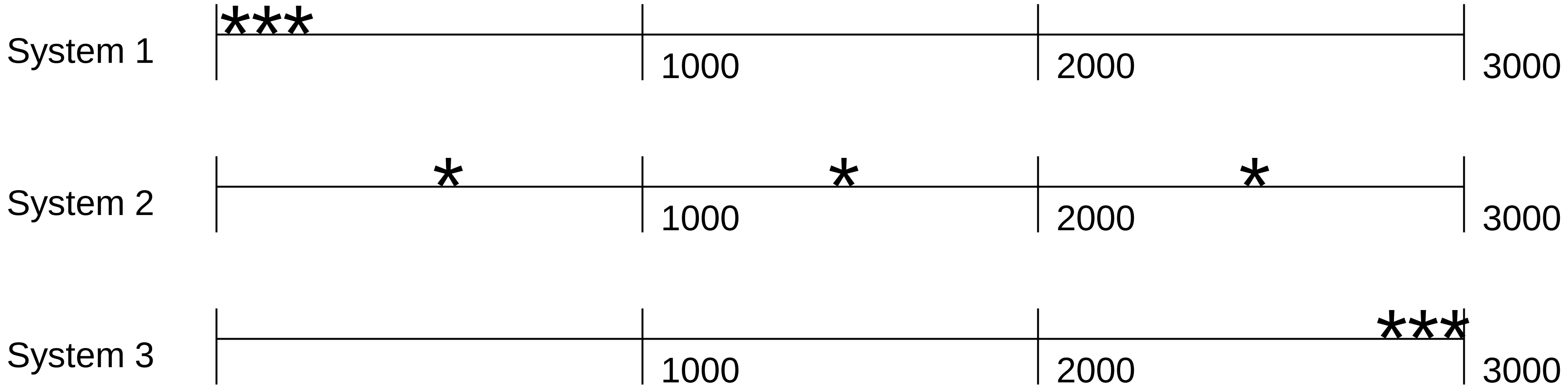
Stable System – No Trend



Here, MTBF = age/number of failures is a valid measure since a straight line can drawn through the data.

Caution MTBF – Hides Information

- Consider three systems operating for 3000 hours.
- System 1 had three failures at 30, 70, 120 hrs and no further failures.
- System 2 had three failures at 720, 1580, and 2550 hrs and no further failures.
- System 3 had three failures at 2780, 2850, and 2920 hrs and no further failures.



All systems have the same MTBF = 3000/3 = 1000 hours.

Is the reliability the same for all three systems?

Interpretation of MTBF as Typical Lifetime

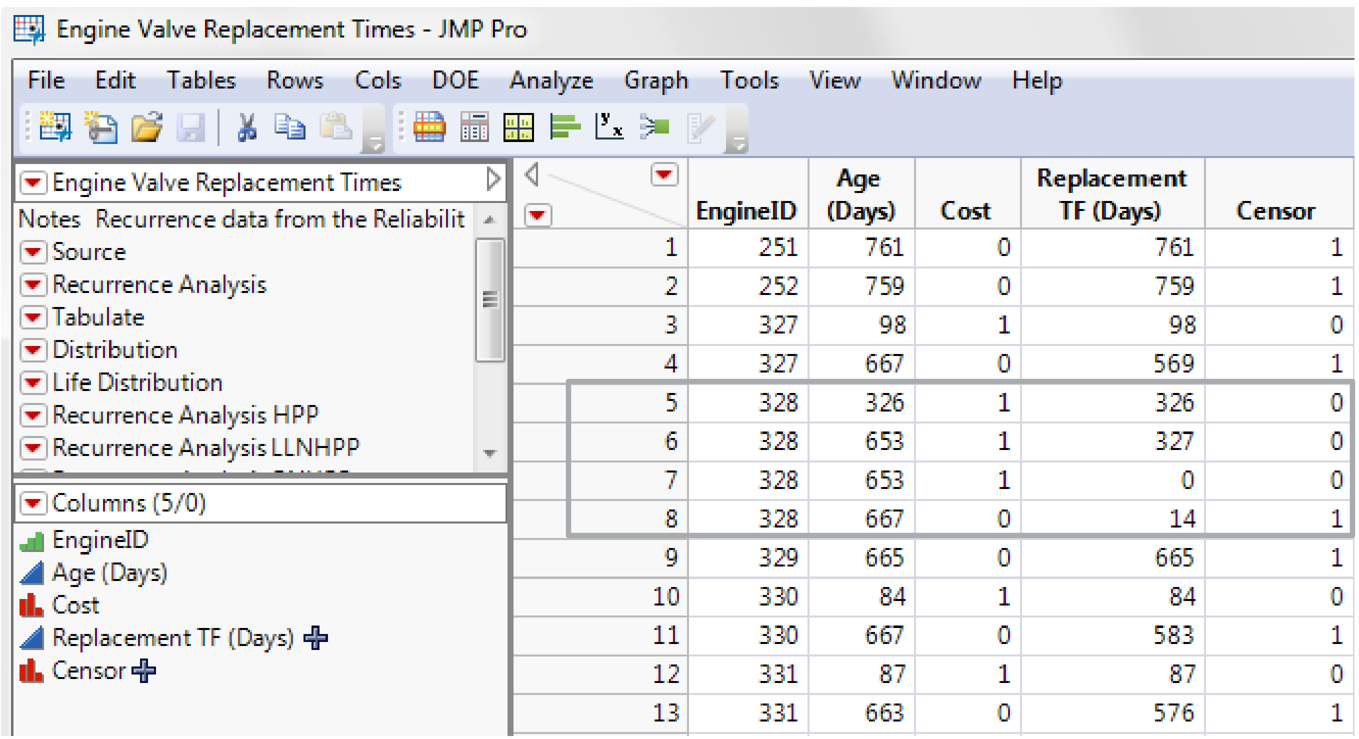
- MTBF is **not** the typical lifetime of a system
- Example:
  - During the years 1996-1998, the average annual death rate in the US for children ages 5-14 was 20.8 per 100,000 resident population.
  - The average failure rate is thus 0.02%/yr
  - The MTBF is the reciprocal of the average failure rate or 4,800 years!

Paper Clip Example of Inflated MTBF

- Bend three clips until each breaks: record 6, 5, and 7 breaks to failure
  - MTBF = 6
- Take a sample of 100 clips and bend each one three times. Only two break.
  - MTBF = 300/2 = 150

Example Analysis of Multiple Repairable Systems

We'll use the JMP sample data file **Engine Valve Seat.jmp** which records valve seat replacements in 41 locomotive engines. [6] Partial table is shown. Each engine has an ID.

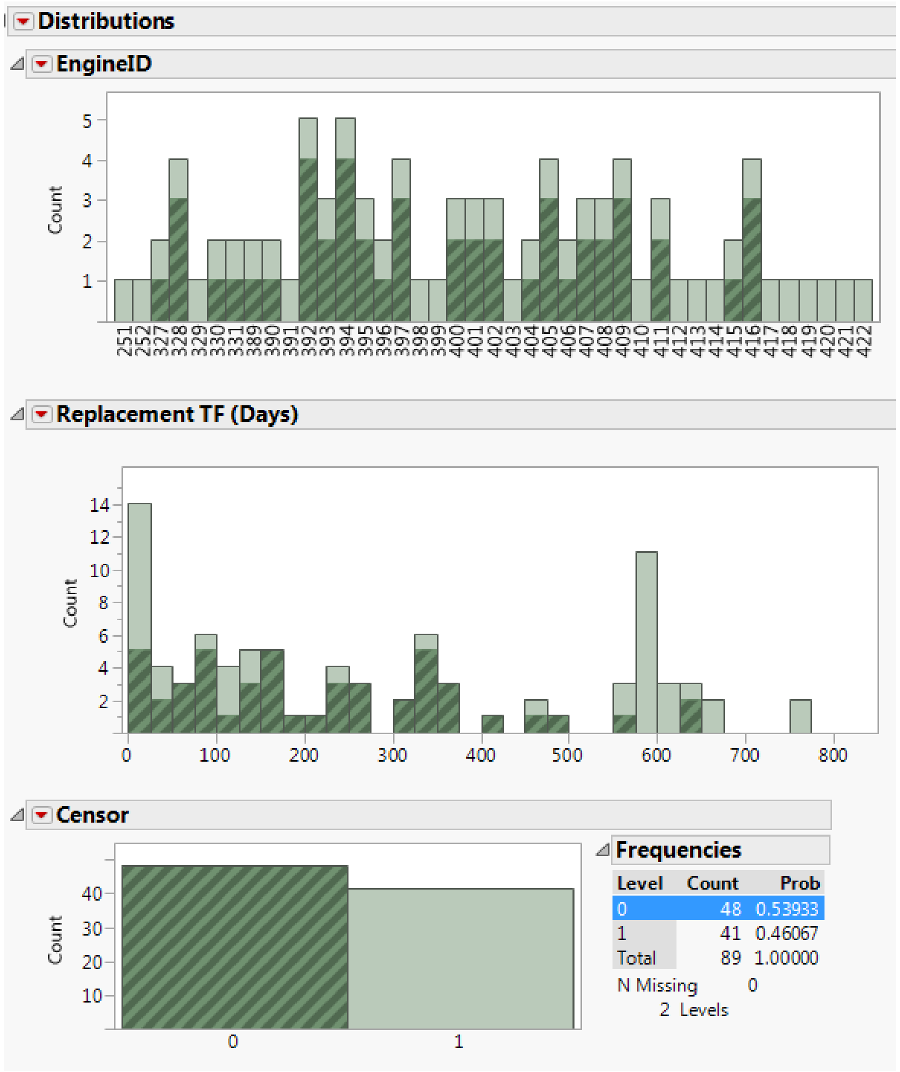


	EngineID	Age (Days)	Cost	Replacement TF (Days)	Censor
1	251	761	0	761	1
2	252	759	0	759	1
3	327	98	1	98	0
4	327	667	0	569	1
5	328	326	1	326	0
6	328	653	1	327	0
7	328	653	1	0	0
8	328	667	0	14	1
9	329	665	0	665	1
10	330	84	1	84	0
11	330	667	0	583	1
12	331	87	1	87	0
13	331	663	0	576	1

Engine 328
TF of Replacements (Cost = 1):
1st at 326
2nd at 327
3rd at 0
Censor at 14

Engine 328
Ages at Replacements:
326, 653, 653
Censor at 667

Non-Repairable Component Analysis: Distributions of TFs of Replaced Valves by Engines

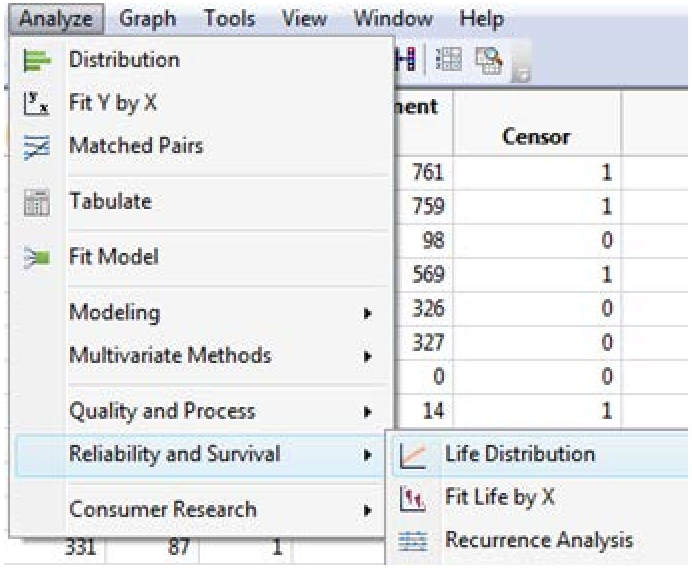


41 Engine IDs.
Shaded areas are replacements.
14 Engines had more than one replacement.

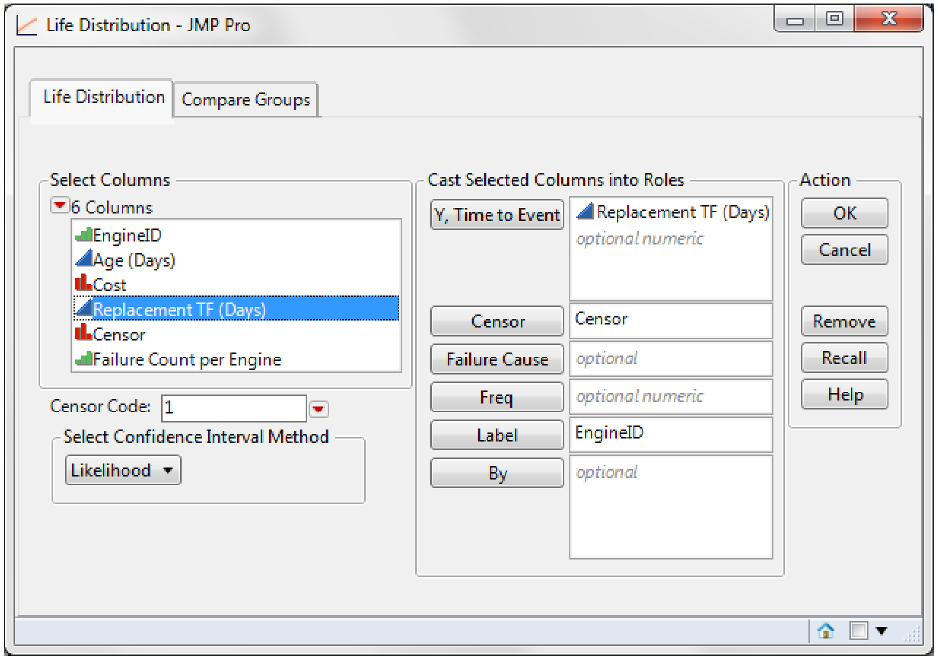
Interarrival TFs and Censoring Times of Valves

Censor Code:
Censor = 1
Replacement = 0.

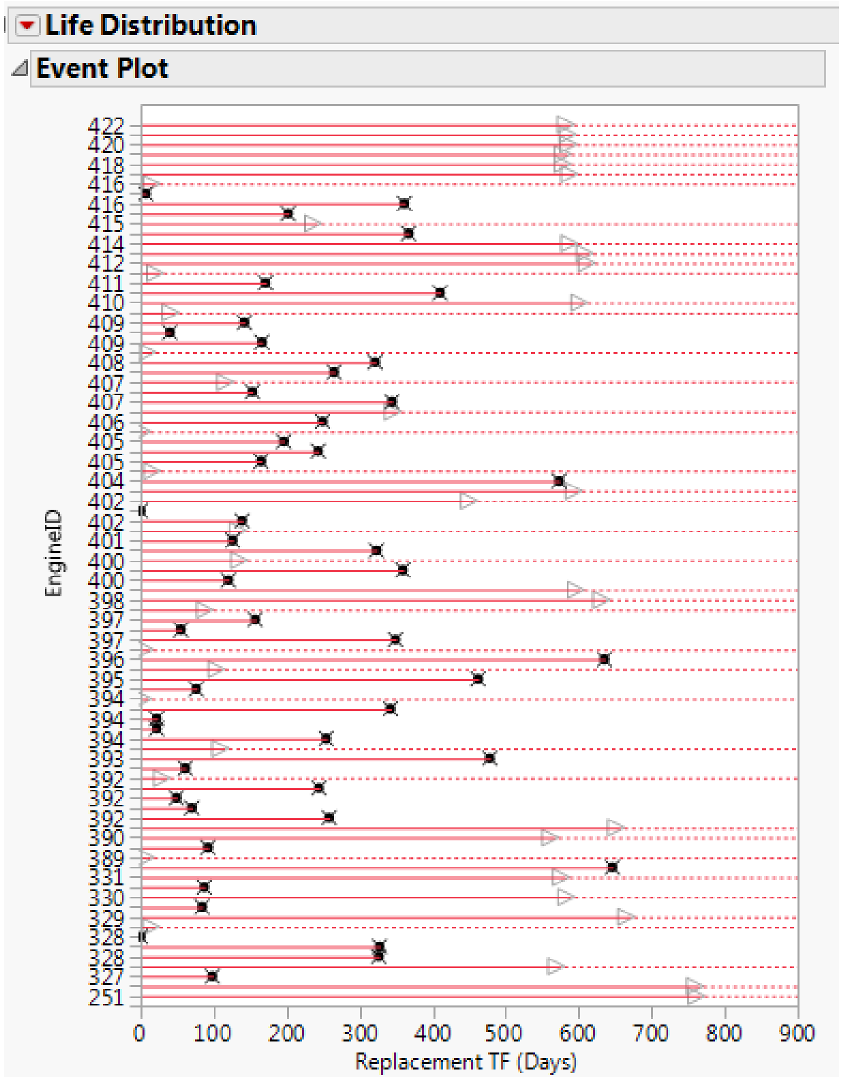
Invalid Analysis of Valve Seats as Non-Repairable Components



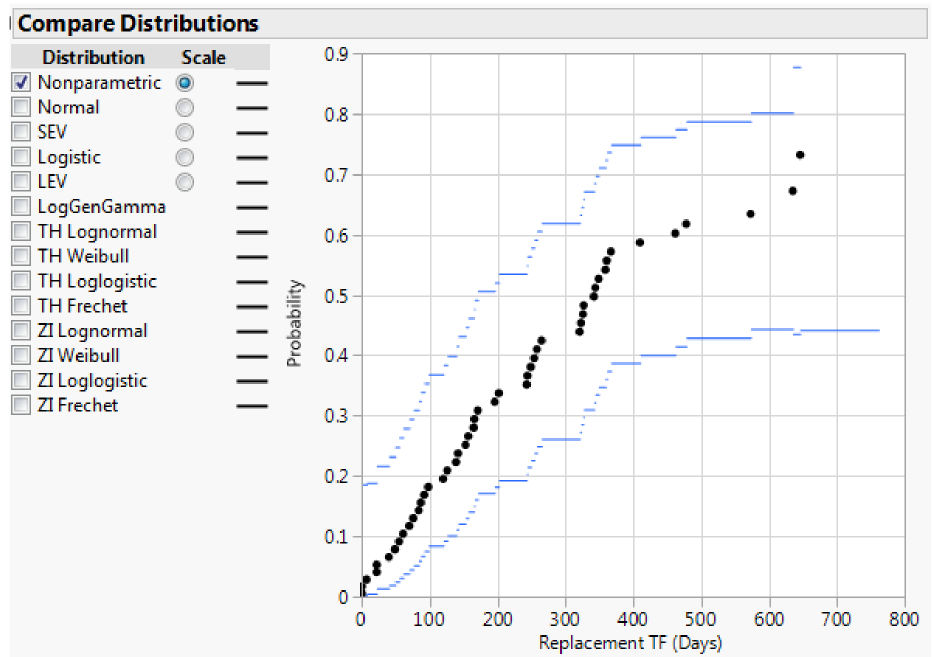
We will incorrectly analyze the data using the **time to failure** for each replaced component, as measured by the **time between successive replacements** for each locomotive.



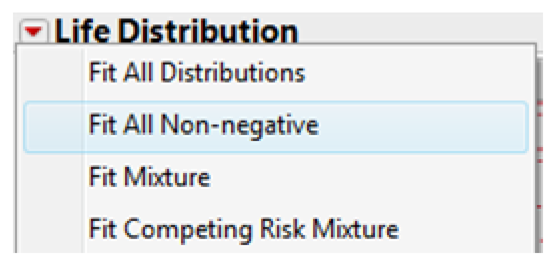
Event Plot and Nonparametric Plot for Times Between Replacements (Incorrect Analysis)



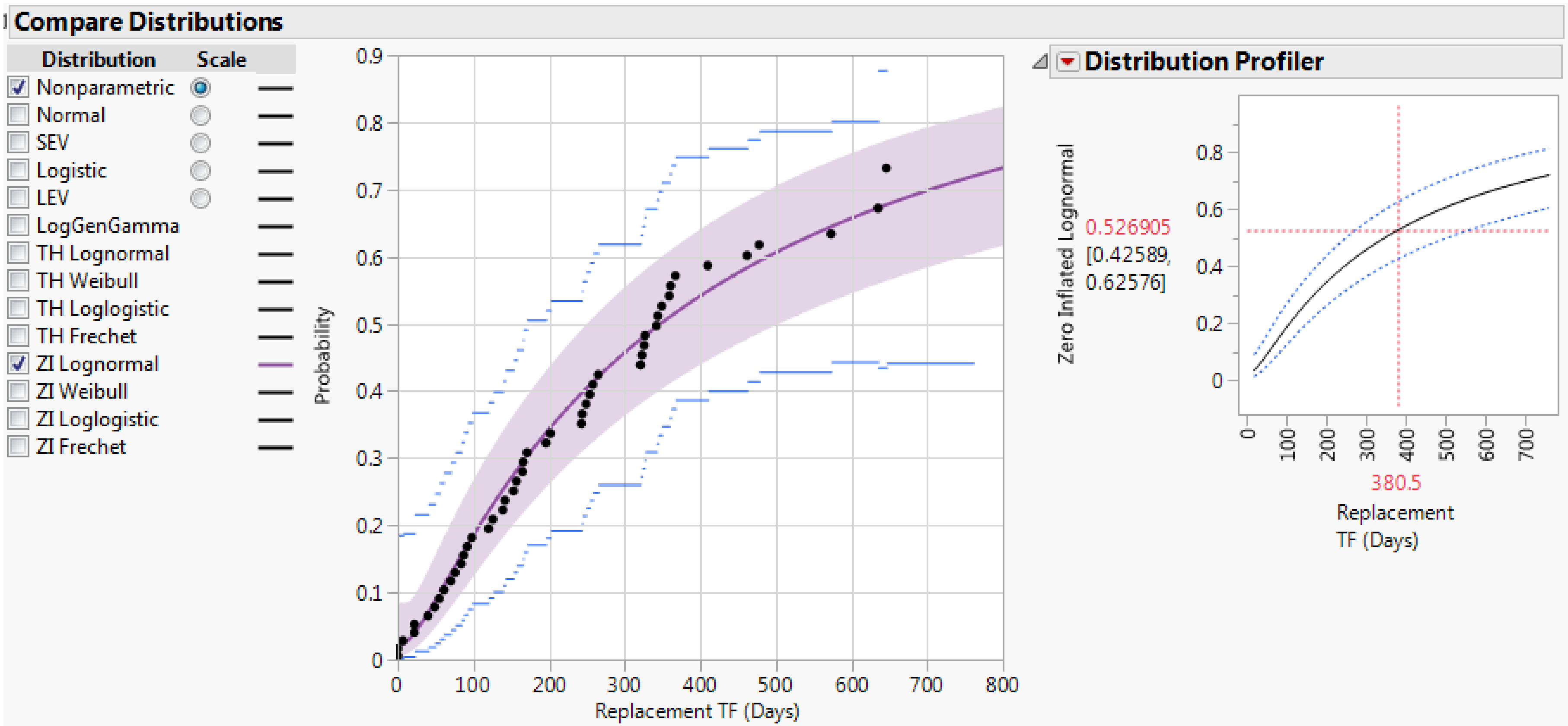
Arrow Indicates Censored Observation
X Indicates Repair.



Select

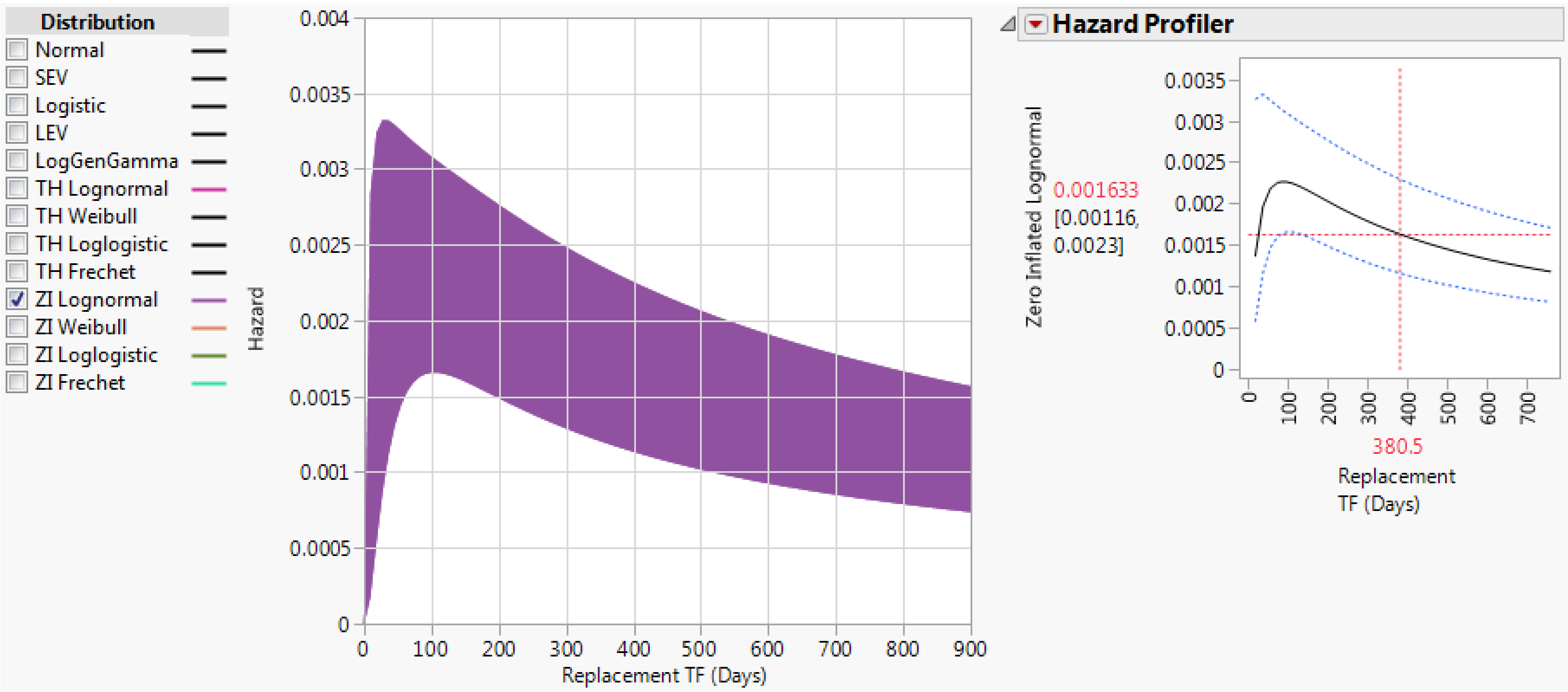


Lifetime Distribution Platform (Incorrect Analysis)



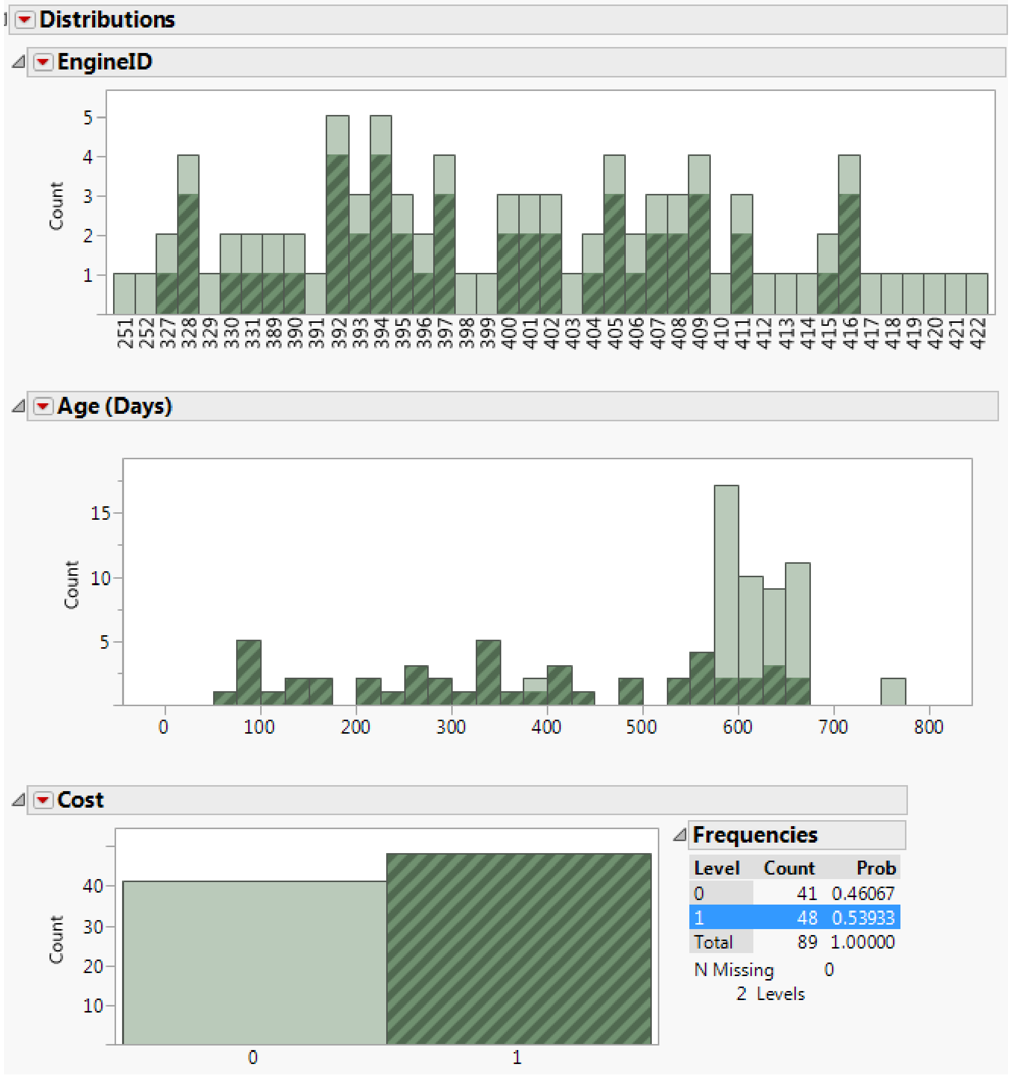
Zero Inflated Lognormal distribution appears to fit data very well.

Hazard Rate Profiler (Incorrect Analysis)



Profiler shows hazard rate increasing early in life, peaking around 100 days, and then continually decreasing thereafter.

Recurrence Analysis: Distributions of System Ages of Replaced Valves by Engines



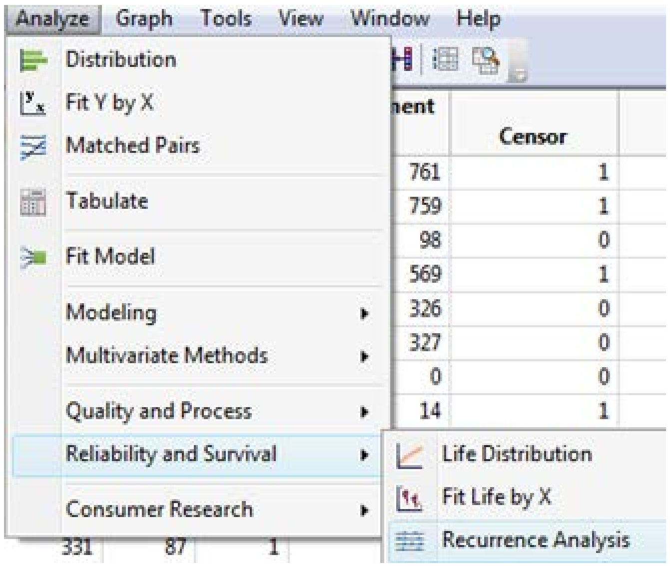
41 Engine IDs. Shaded areas are replacements.

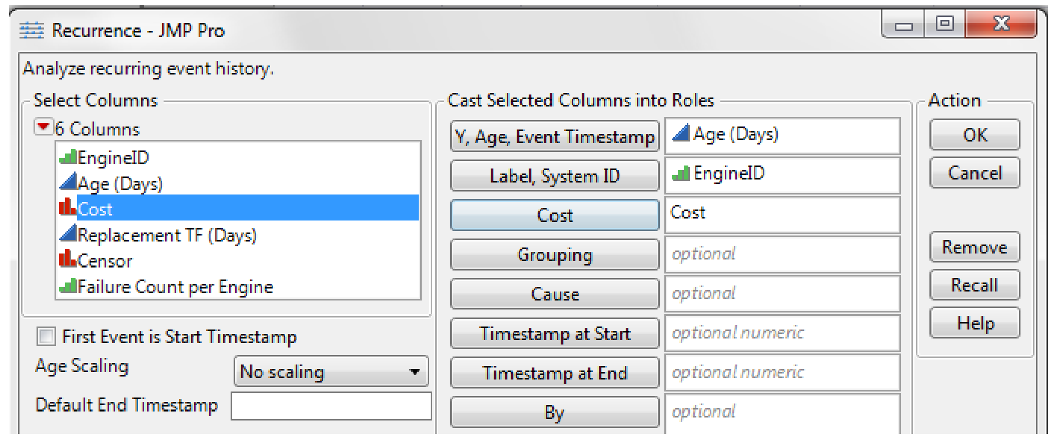
Age at Replacement and Censoring Times of Valves

Cost Code:
Censor = 0
Replacement = 1.

Proper Recurrence Analysis of Valve Seat Data

Y, Age, Event Timestamp:
Age column is entered.

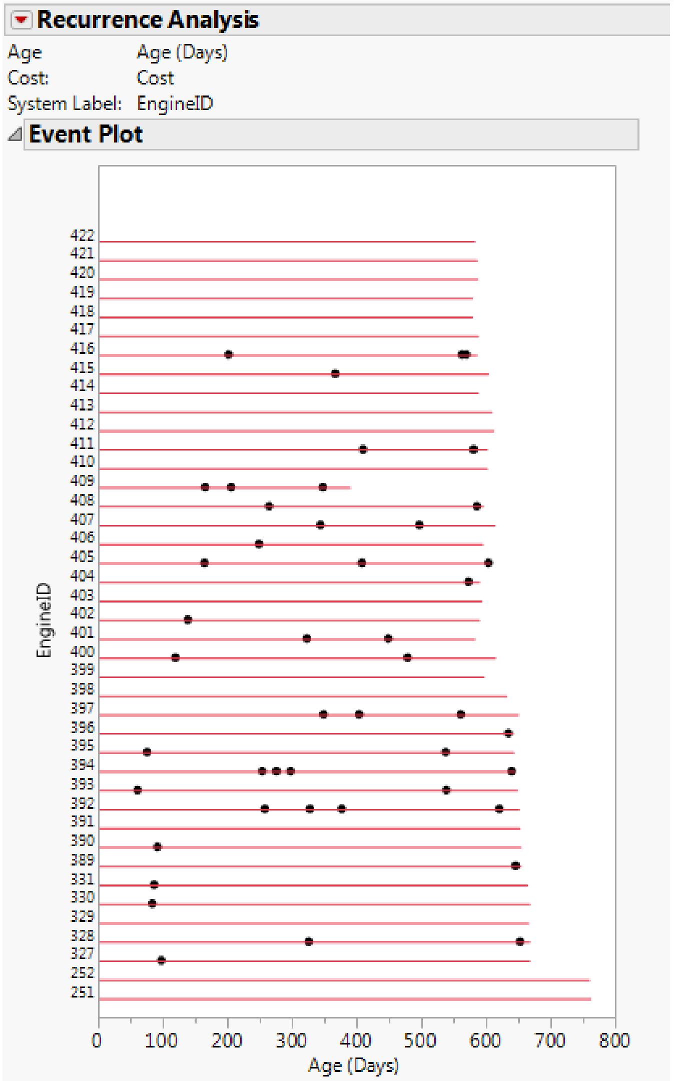


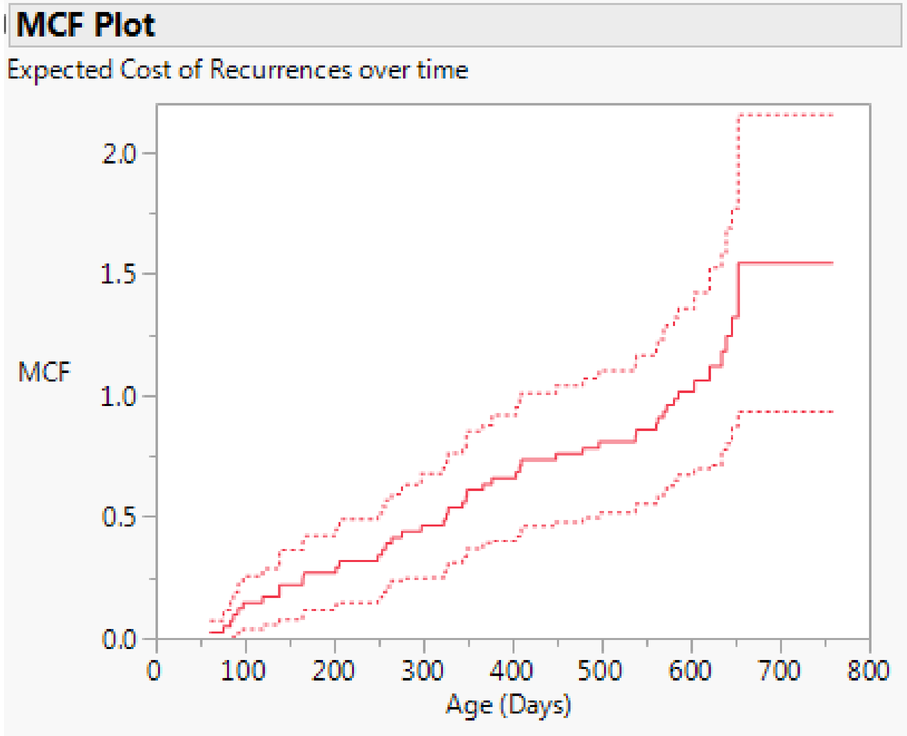


Cost:
1 = Replacement
0 = Censor

Label, System ID:
Required! Each and every system must have a censor time index of 0 in Cost column.

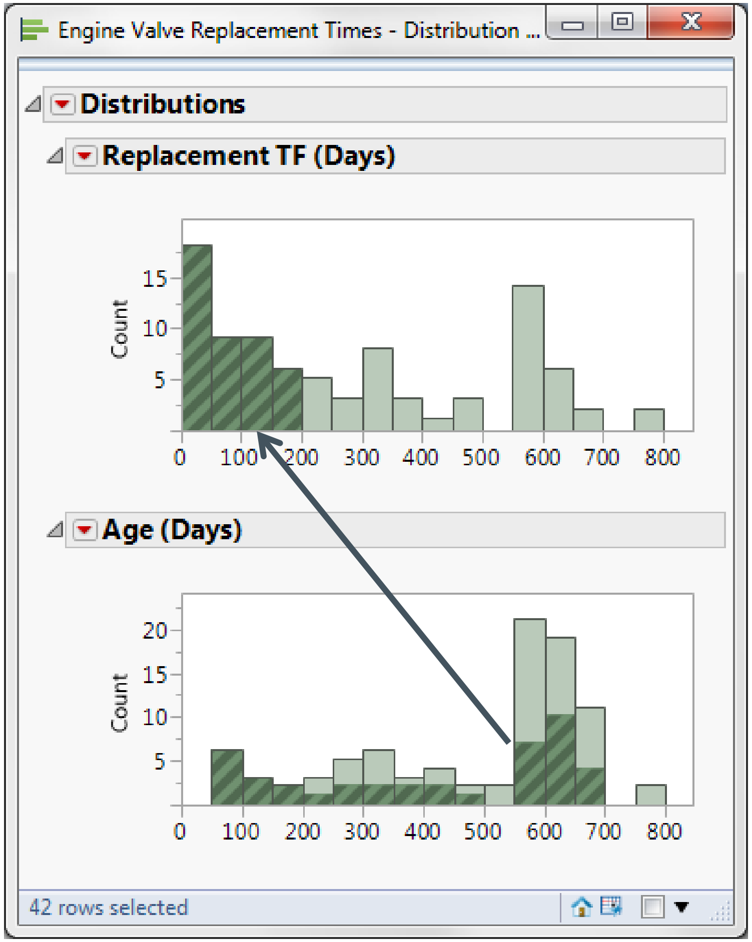
Analysis of Valve Seats on Locomotive as Repairable Systems





MCF plot shows repair rates increasing at a nearly constant rate until around 550 days, when the rate appears to increase. Is wearout occurring?

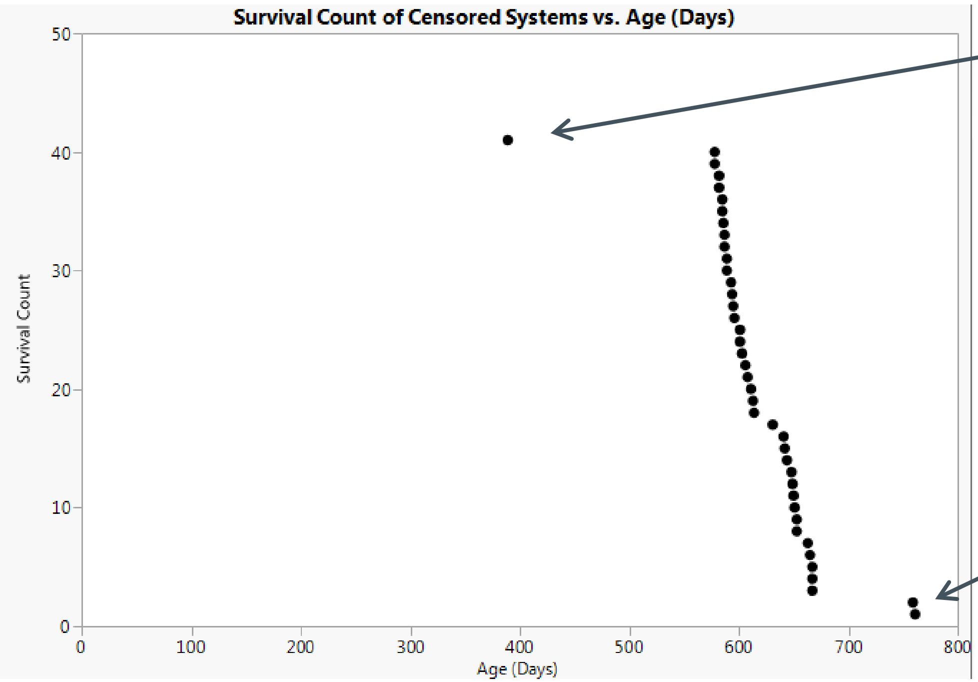
Comparison of Interarrival Times Vs. Ages



Note that several interarrival times of short duration occurring at oldest system ages show up as early failures when order is neglected, resulting in misleading interpretation of data.

Survival Distribution of Censoring Ages

A graph that is useful for repairable systems is a plot of the **number or percent of starting systems that exist as a function of the system ages**, that is, a plot of the number of censored observations versus the censoring ages of the systems.



One locomotive (#409) began operation over six months later than most systems. Nelson [3] reports that this system was “dropped into the water while being loaded on shipboard to go to China. Water removal and other cleanup delayed its start in service.” #409 had three replacements in this shortened time.

Two locomotives began operation over three months earlier than practically all other systems. Both systems had no replacement failures.

Calendar Date MCF Analysis (CMCF)

- Normally an MCF is plotted versus the systems **ages**. However, there may be applications where the MCF could be plotted versus the **calendar da**te to reveal issues that might be less evident by a typical MCF calculation. [7]
- For example, suppose a group of systems installed at various times in a facility during the year are **moved on the same date to a new location**. There may be multiple failures on or after the same date associated with the move and not related to system ages.
- As another example, a group of systems with different ages may all receive **a software upgrade** that causes issues. Again, the calendar date plot might be more revealing of a special cause variation on the common date.

Lessons Learned

Analysis of repairable system data by fitting lifetime distributions to times between repairs can produce misleading results.

For repairable systems, the **time order** in which failures occur is a very important factor for analysis.

For individual systems, a **cumulative plot** shows the repair history graphically. For multiple systems, the **MCF** plot can reveal trends in the collective behavior of a group of systems.

References

1. SAS Institute Inc. 2015. *JMP® 12 Reliability and Survival Methods*. Cary, NC: SAS Institute Inc.
2. Tobias, P.A. and Trindade, D.C. 2012 *Applied Reliability*, 3rd ed., Boca Raton, FL: CRC Press
3. Nelson, W.B. 2003. *Recurrent Event Data Analysis for Product Repairs, Disease Recurrence, and Other Applications*. Philadelphia, PA:SIAM
4. Meeker, W.Q. and Escobar, L.A. 1998, *Statistical Methods for Reliability Data*, New York: John Wiley & Sons
5. Davis, D.J. 1952, An analysis of some failure data. *J Am Stat Soc* 47:113-50
6. Nelson, W.B. 1995 Confidence limits for recurrence data – applied to cost or number of product repairs. *Technometrics* 37:147-157
7. Trindade, D.C. and Nathan, S., 2008 Field Data Analysis for Repairable Systems: Status and Industry Trends, *Handbook of Performability Engineering*, London, Springer-Verlag
8. Trindade, D.C. 1975 An APL program to numerically differentiate data. IBM TR Report (19.0361)