Ground Vehicle Reliability Analysis with the Mean Cumulative Function (MCF)

Presented at 2014 ICEAA Professional Development & Training Workshop June 2014

Caleb Fleming and Jennifer Scheel
Kalman & Company, Inc.
cfleming720@gmail.com
Jennifer.Scheel@Kalmancoinc.com

Outline

- Introduction
- Parametric vs nonparametric assumptions
- Recurrence data
- MCF overview
- Censoring
- Exact age vs interval data
- Single, group, and mixed data
- Generating defensible assumptions
- Ground vehicle analysis example
- Application
- Software
- References/Relevant Links

Introduction

- Statistical analysis is a critical component of cost estimating
- In order to develop accurate Cost Estimating Relationships (CER), tests for statistical significance are employable
- Two parent categories exist for tests of statistical significance
  - Nonparametric
  - Parametric

Introduction

- Parametric statistics
  - Widely understood and most recognizable
  - Always follow family of normal distributions
  - High levels of statistical power
  - High levels of precision
  - Generally sensitive to outliers
  - Require strict adherence to detailed test assumptions

Introduction

- Nonparametric statistics
  - Less commonly used and therefore less recognizable
  - Typically distribution-free
  - Results are generally robust to outliers
  - Require fewer and less strict, assumptions
  - Lower levels of statistical power
  - Helpful when used with behavioral research methods
  - Results generally reflect differences between groups of data

Parametric vs Nonparametric Assumptions

- Most notable difference is the emphasis on particular assumptions
- Parametric assumptions
  - Independent histories
  - Independent increments
  - Population follows parametric curve
  - Different types of recurrence are independent
  - Repair restores a unit to like-new or like-old condition

Parametric vs Nonparametric Assumptions

- Nonparametric assumptions
  - Target population is specified
  - Random sampling of the target population
  - Histories are independent of their censoring ages
  - Population history functions extend through the age range of the sample data
  - Population mean is finite over the range of data
  - All recurrence ages are distinct from each other and from the censoring ages

Parametric vs Nonparametric Assumptions

- Nonparametric models are valid even when parametric assumptions are met
- When the parametric assumptions are met, the parametric methodology will generally yield more accurate results

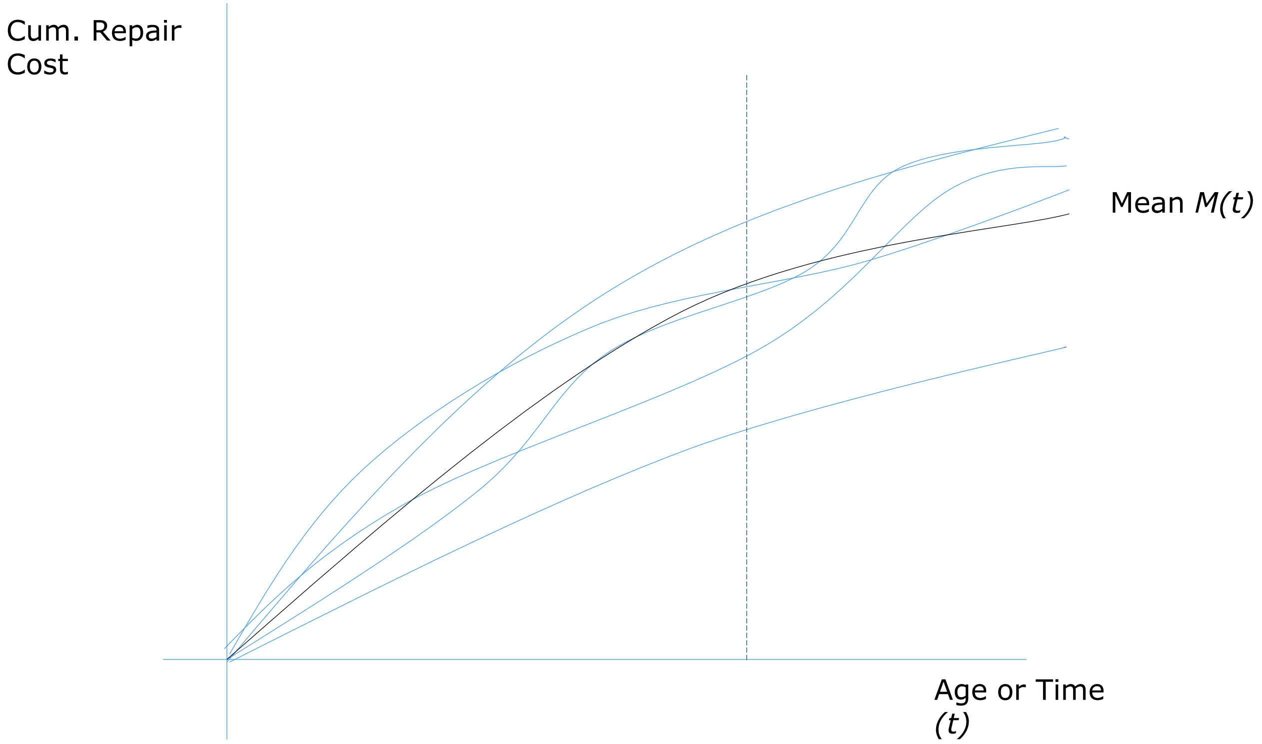
Recurrence Data

- Vehicle reliability is a common area of interest, specifically with regard to reparable subsystems and components
- Reliability analysis is derived from time-to-failure and time-between-failure data
  - Critical to life cycle cost analysis
- Points in these datasets are called “recurrence data”
  - Number of life cycle repairs to a transmission or fuel pump
- Recurrence data is oft modeled parametrically using the stochastic point process (Poisson)
  - Concern: Poisson process applies only to counts of recurrences, not the cost

Mean Cumulative Function Overview

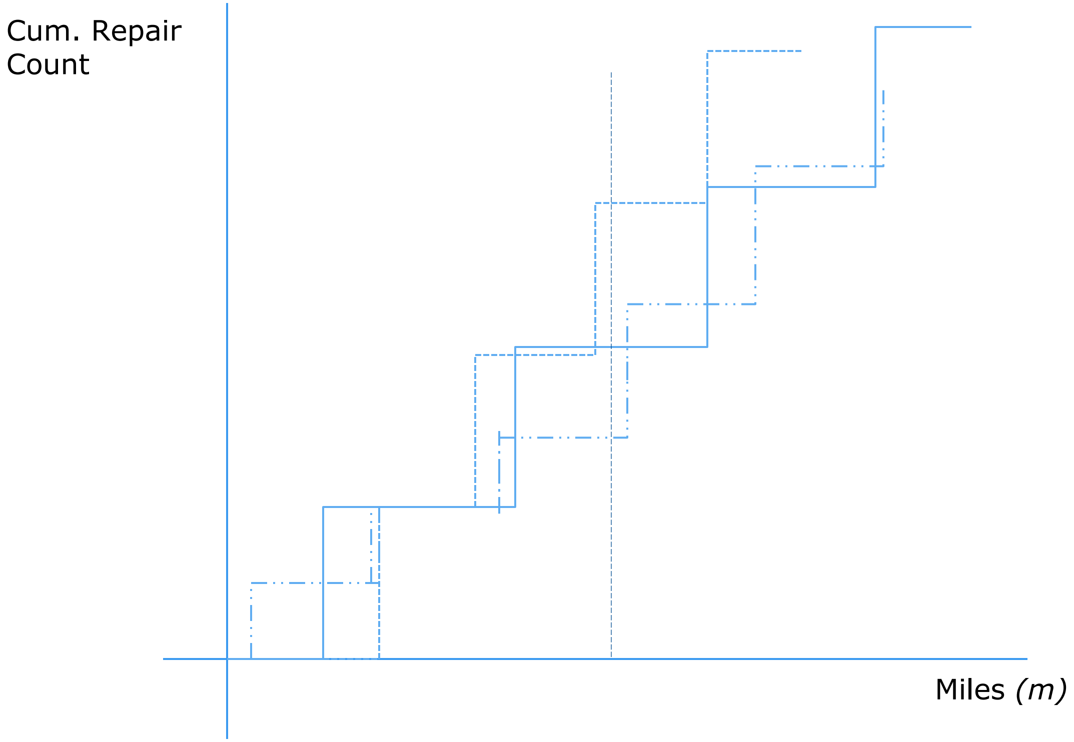
- The mean cumulative function (MCF) offers a nonparametric method that requires fewer assumptions, enables simplified methodologies, and yields more expansive outputs
  - The MCF could show event counts, costs, and maintenance down times (indicator of availability), among other values
- Population “value” for each function follows a staircase curve with unequal step rises
- Each model consists of a set of value curves
- At any age or time *t*, the corresponding distribution of the value curves has a mean *M(t)*
- This mean curve is the MCF

Mean Cumulative Function Overview



Sample Continuous Cumulative History Function

Mean Cumulative Function Overview



Sample Discrete Cumulative History Function

Mean Cumulative Function Overview

- Purpose:
  - Determining recurrence rate behavior
    - Burn-in
    - Preventative replacement
    - Bath-tub effect
    - Retirement
  - Availability
  - Population comparison
- Calculation
  - For cost and count data, the “instantaneous” recurrence rate is found by calculating the derivative, or slope of the sample mean cumulative function, at a particular mileage or time

Censoring

- Occurs when an observation value is only partially known
  - Ex: A vehicle is removed from a study after 25,000 miles; We know the vehicle's transmission is reliable for at least 25,000 miles, though it may or may not be more than that
- Types of censoring
  - Right
  - Left
  - Interval
  - Type I
  - Type II
  - Random

Exact Age vs Interval Data

- Exact age with right censoring
  - Discrete events with precise ages of recurrence and right censoring times
    - Ex: Steering gear repairs on a mileage scale
  - Distinct values on the age scale with no ties
  - Numerous ties warrant conducting analyses using the alternative “interval method”
  - Most common form of recurrence data
  - Data presented in “time-event” plots

Kalman & Company, Inc.

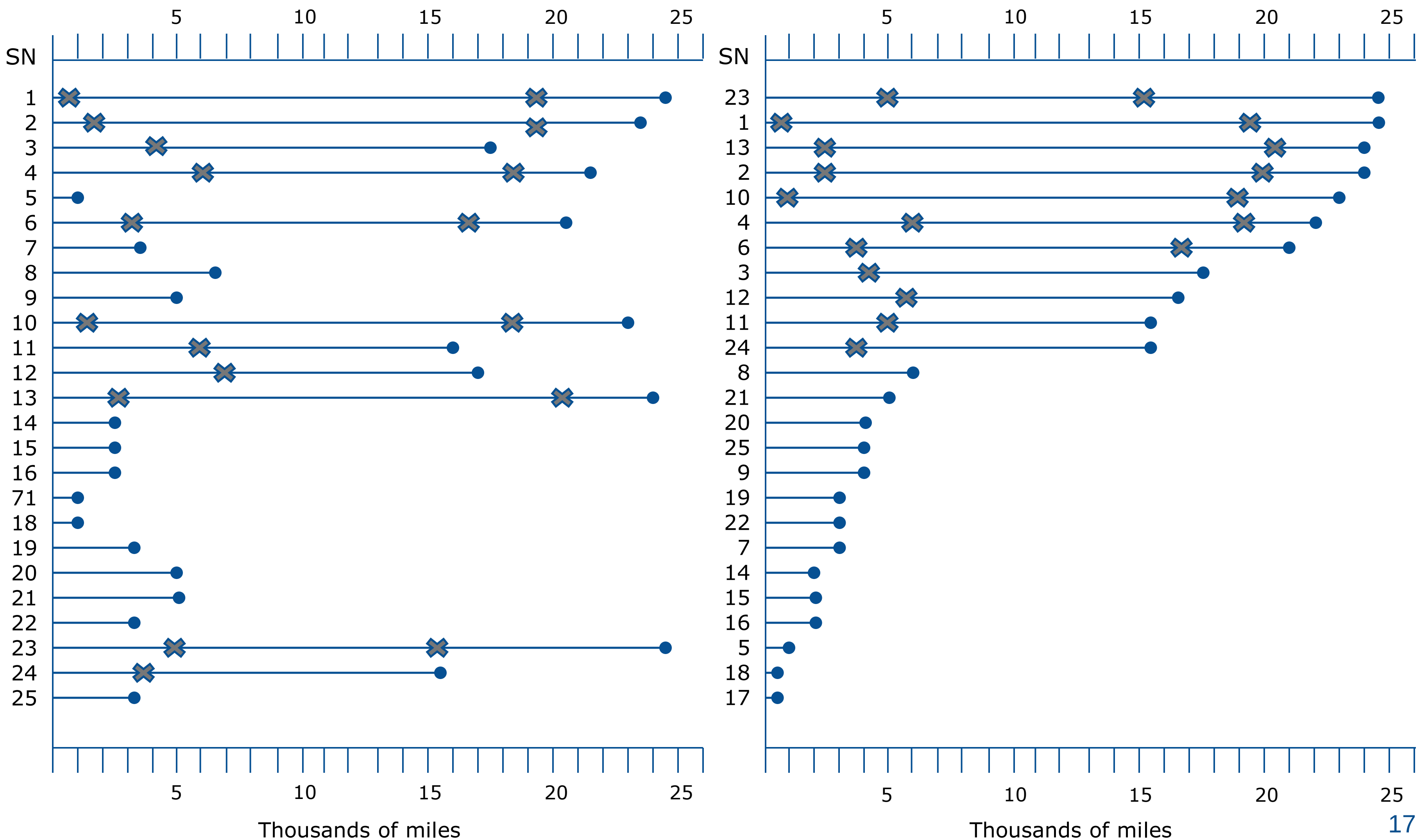
Exact Age vs Interval Data

Serial Number	Mileage		
0001	856	19323	24416+
0002	2877	19818	23676+
0003	4642	17233+	
0004	6609	18258	21137+
0005	1017+		
0006	3528	16963+	20407+
0007	3019+		
0008	6899+		
0009	4233+		
0010	1270	18736	22921+
0011	5656	15511+	
0012	6541	16332+	
0013	2536	20665	23931+
0014	2627+		
0015	2400+		

Serial Number	Mileage		
0016	2250+		
0017	864+		
0018	891+		
0019	3750+		
0020	4999+		
0021	5179+		
0022	3470+		
0023	5021	15205	24567+
0024	3280	15232+	
0025	4620+		

Sample exact age with right censored repair data

Exact Age vs Interval Data



Exact Age vs Interval Data

- Exact age with left censoring
  - Exact age characteristics apply
    - Discrete events with precise ages of recurrence and right censoring times
    - Distinct values on the age scale with no ties
    - Numerous ties warrant conducting analyses using the alternative “interval method”
  - Less common, as left censoring implies that a data gap exists from age zero to the first observation
    - Ex: The second owner of a vehicle is sometimes unaware of the maintenance plan in place for the miles accrued prior to their procurement

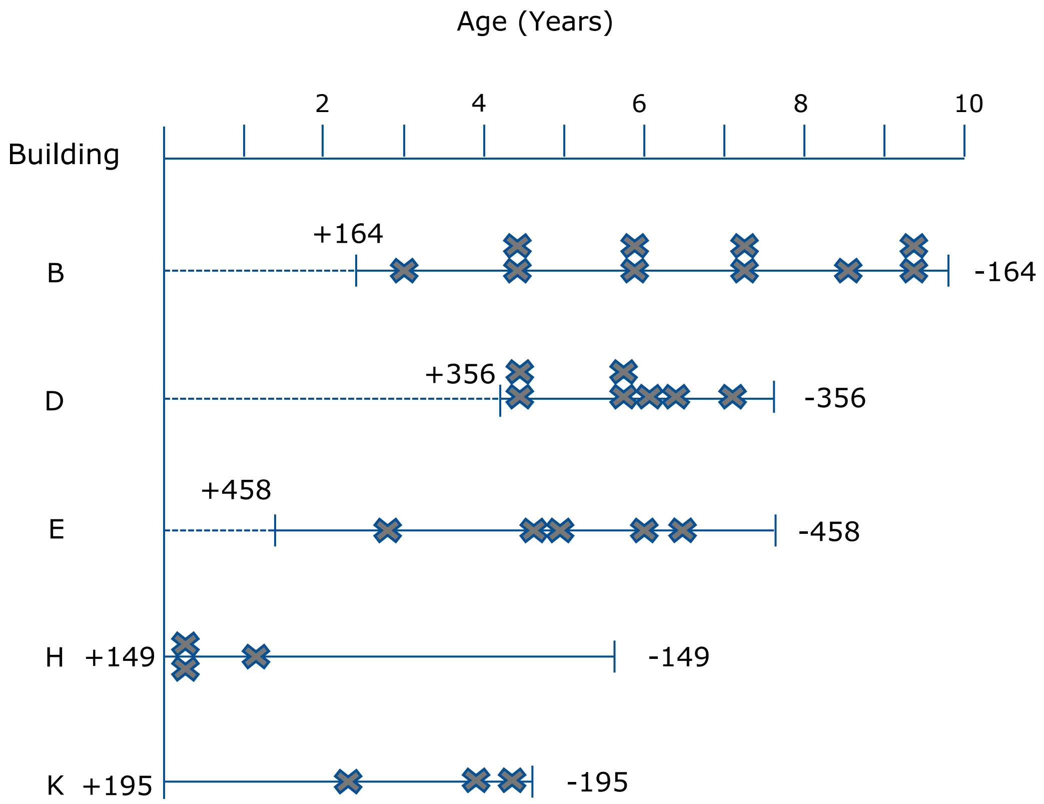
Kalman & Company, Inc.

Exact Age vs Interval Data

Bldg. B	Bldg. D	Bldg. E	Bldg. H	Bldg. K
2.59 (+164)	4.45 (+356)	1.00 (+458)	0.00 (+149)	0.00 (+195)
3.30	4.47	2.58	0.17	2.17
4.62	4.47	4.65	0.17	3.65
4.62	5.56	4.79	1.34	4.14
5.75	5.57	5.85	5.09 (-149)	4.14 (-195)
5.75	5.80	6.73		
7.42	6.13	7.33 (-458)		
7.42	7.02			
8.77	7.05 (-356)			
9.27				
9.27				
9.33 (-164)				
10 replaced	7 replaced	5 replaced	3 replaced	3 replaced

Table 2: Sample exact age with left censored repair data

Exact Age vs Interval Data



Exact Age vs Interval Data

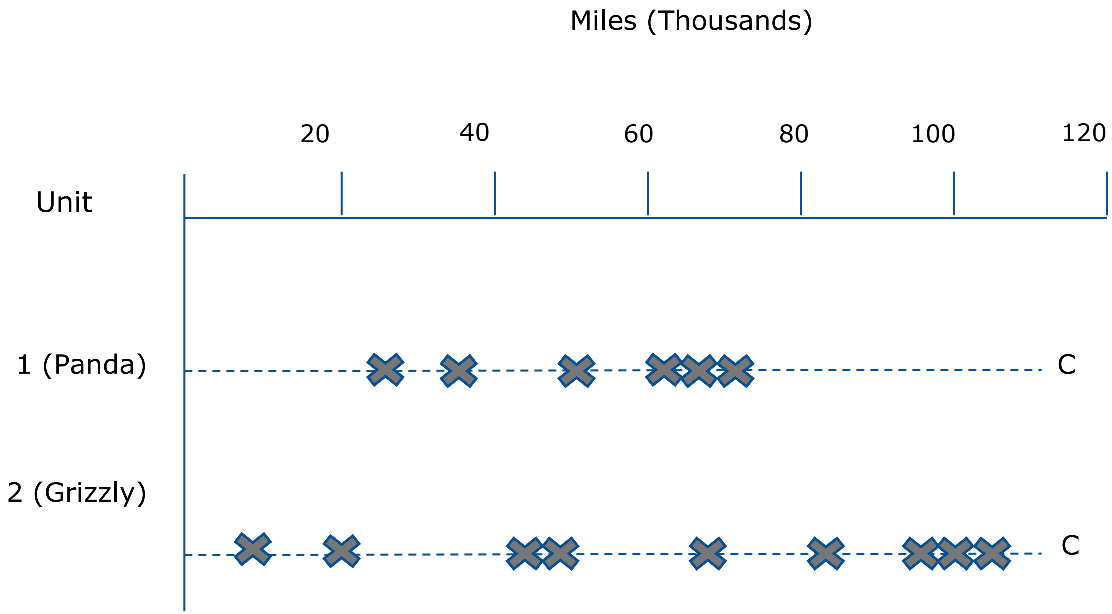
- Interval data
  - Exact event ages and censoring ages for a unit are unknown (not discrete), therefore the scale has been partitioned into intervals
  - Number of events within interval is known
  - Interval grouping sometimes occurs to accommodate large datasets where it’s OK to lose minor amounts of precision
    - Ex: Daily event data available, but reporting occurs at the yearly level

Exact Age vs Interval Data

	Mileage	Panda		Grizzly	
Interval	Range (K miles)	# of Engine Failures	# Censored	# of Engine Failures	# Censored
0	0-20	0	0	0	0
1	21-40	3	1	0	0
2	41-60	3	0	1	1
3	61-80	3	1	6	2
4	81-100	5	1	8	7
5	100-120	6	3	6	3
	Total:	20	6	21	13

Table 3: Sample interval repair data

Exact Age vs Interval Data



Exact Age vs Interval Data

- What's different?
  - Right censored data:
    - Estimate using the number of units "at risk" remaining
    - Calculate mean cost per unit for each recurrence, or number of units for each recurrence
  - Left censored data:
    - Estimate using the number entering the sample at a particular time
    - Calculate incremental mean number of recurrences per unit
    - Cost is calculated by dividing the cost by the number at risk at the specific recurrence

Exact Age vs Interval Data

- What's different?
  - Interval Data
    - Define the interval, and calculate the number of recurrences and censor points within each
    - Calculate the average number of recurrences per sample unit within each interval, or average total cost of all recurrences in an interval
    - Calculate the mean cumulative function using the same methodology as with exact age data

Single, Group, and Mixed Data

- Single
  - Estimating the population MCF for a single type of event
    - Example: Gears, gear trains, driveshaft
- Group
  - Estimating the population MCF for a group of events
    - Example: Transmission
- Group with events eliminated
  - Estimating the population MCF for a group if particular failure modes were eliminated
    - Example: Upgrading all driveshafts and only looking at other transmission components
    - Note: Independence assumption necessary

Single, Group, and Mixed Data

- Mixed
  - Multiple distinct events or types of recurrences take place within the sample dataset
    - Example: Engine and transmission repairs included together
  - In order to estimate a mix of *K* events, the data must be able to be partitioned into *K* types
    - Ex: Repairs attributed specific components within a subsystem would need to be separable
  - Model consists of *N* units in a population of *N* vectors, each with *K* cumulative history functions and *K* MCFs
  - MCFs are summative, meaning that the analyst could group together the MCFs for common Work Breakdown Structure (WBS) elements

Generating Defensible Assumptions

- Define the repair "event" to be analyzed:
  - Preventative replacement
  - Scheduled maintenance
  - Component replacement vs repair (patching a tire vs replacing a tire)
  - Adjustments (inflating a tire vs replacing a tire)
- Define "age" units:
  - Appropriate measures of system usage
    - Days
    - Mileage
    - Usage cycles
    - Energy output
    - Including or excluding downtimes
  - Determining which age to use
    - Age at failure
    - Age at repair
    - Age at completion reporting

Generating Defensible Assumptions

- Determine what costs to include
  - Labor
  - Materials
  - Warranty repair
  - Preventative maintenance costs
  - Depreciation
- Identify potential risk
  - Sampling error
  - Reporting error
  - Measurement error
  - Model error

Ground Vehicle Analysis Example

- Panda is a fictional, lightweight, ground vehicle in preproduction testing
- Preproduction tests were conducted at Panda Proving Ground with vehicles accumulating 30,000 test miles – the equivalent of 180,000 mission miles
- Using historical test data, complete the following objectives:
  1. Evaluate the per Panda mean cumulative number of repairs at different life cycle intervals
  2. Evaluate the per Panda mean cumulative cost of repair at different life cycle intervals

Ground Vehicle Analysis Example

- Step one – Order mileages
  - Arrange the sample recurrence and censoring mileages from smallest to largest
  - Denote censoring mileages with a “+”
  - If ties are present (there are not any in our dataset), order randomly
  - If there are common recurrence and censoring mileages, note the recurrence event before the censoring
- Step two – Identify the number “at risk”
  - For each sample mileage, in the second column write the number of remaining units (r)
    - Example: If we were to start with three engines and have one replaced, the censoring would leave two engines remaining at risk of experiencing recurrent repair

Ground Vehicle Analysis Example

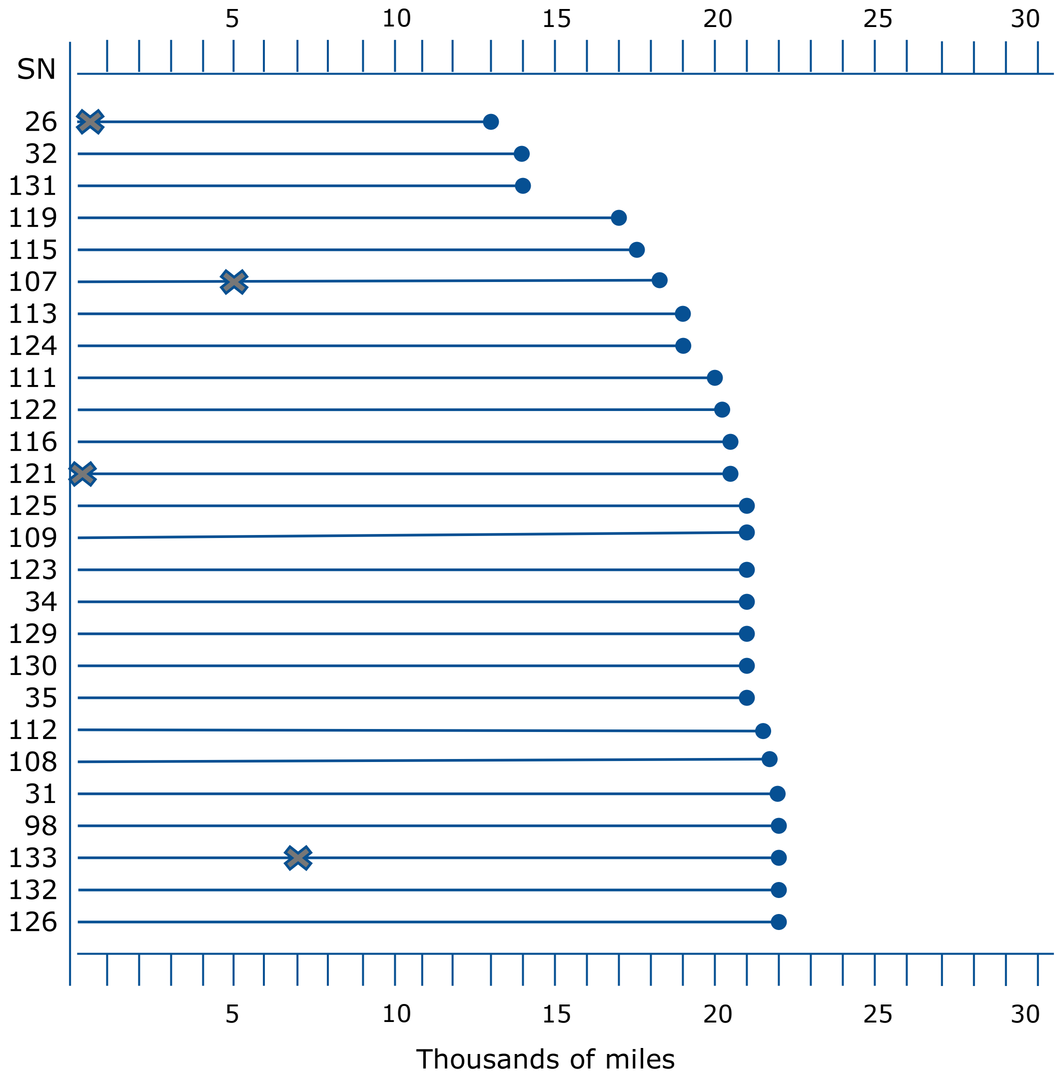
- Step three – Calculate the mean
  - Calculate the observed incremental “mean number of recurrences per unit” for each mileage in column 3 as 1/r
- Step four– Generate the sample MCF
  - Calculate the sample MCF at each recurrence by summing the preceding increments
  - No MCF is calculated at the censoring mileages, however the censoring affects the number at risk for each recurrence
- Step five – Graph
  - Plot each recurrence MCF value against age
  - Resulting plot is the nonparametric sample MCF

Ground Vehicle Analysis Example

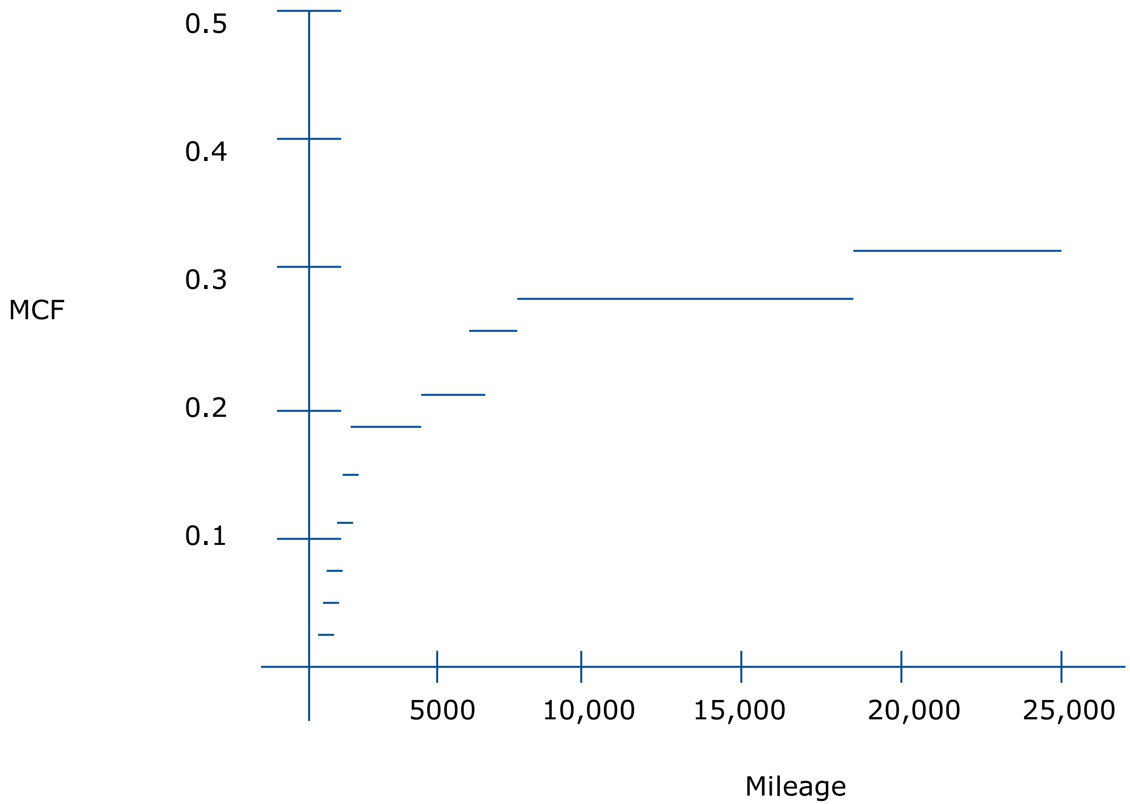
Mileage	r	Mean	MCF
28	34	0.03	0.03
48	34	0.03	0.06
375	34	0.03	0.09
530	34	0.03	0.12
1388	34	0.03	0.15
1440	34	0.03	0.18
5094	34	0.03	0.21
7068	34	0.03	0.24
8250	34	0.03	0.27
13809+	33		
14235+	32		
14281+	31		
17844+	30		
17955+	29		
18228+	28		
18676+	27		
19175+	26		
19250	26	0.04	0.31
19321+	25		
19403+	24		
19507+	23		
19607+	22		

Mileage	r	Mean	MCF
20425+	21		
20890+	20		
20997+	19		
21133+	18		
21144+	17		
21237+	16		
21401+	15		
21585+	14		
21691+	13		
21762+	12		
21876+	11		
21888+	10		
21974+	9		
22149+	8		
22486+	7		
22637+	6		
22854+	5		
23520+	4		
24177+	3		
25660+	2		
26744+	1		
29834+	0		

Ground Vehicle Analysis Example



Ground Vehicle Analysis Example



Ground Vehicle Analysis Example

- Results
  - Value achieved directly from the staircase estimate or plotted curve
  - At ~6,000 test miles (36,000 mission miles), there are 0.21 failures per car (21% failure rate)
  - At ~20,000 test miles (120,000 mission miles), there are 0.31 failures per car (31% failure rate)
  - Recurrence rate interpretation

Ground Vehicle Analysis Example

- Applying cost instead of events
  - Using the repair cost at each mileage, calculate the mean cost per “at risk” unit
  - Iteratively sum each MCF

Ground Vehicle Analysis Example

- Possible additional excursions with a larger scope of data
  1. Identify points on the MCF where repair costs sharply incline or decline
  2. Project mileage at which the most cost effective option becomes vehicle replacement vice repair
  3. Generate confidence intervals (more assumptions required)

Application

- Predictive modeling
  - Consumables and reparables costs
  - Secondary reparable costs
  - Service life extension program estimation
  - Cost growth
  - Failure rates
  - Population comparison
  - LCCE development

Application

- MCF provides an alternative to parametric estimation, providing precise point estimate values for component failure event quantities and costs
- Revealing statistically significant maintenance event tendencies
- Forecast repair part demand and curb the downtime associated with deadline component failures in a peacetime environment

Software

- A majority of the MCF calculations are done using simple calculators or spreadsheets
- Programs useful for more complex analysis, including determining confidence limits, include:
  - SAS by the SAS Institute
  - ReliaSoft by Weibull ++
  - SPLIDA features for S-PLUS of Insightful

References/Relevant Links

- This presentation, including sample data contained within, is derived from research, analysis, and documentation conducted and presented by Wayne Nelson in his book, “Recurrent Events Data Analysis for Product Repairs, Disease Recurrences, and Other Applications” (2002)
- Datasets presented in Mr. Nelson’s book can be downloaded at www.siam.org/books/sa10/