Overview of R Package SMRD
Statistical Methods for Reliability Data in R

Maj Jason K. Freels, USAF\textsuperscript{1}    William Q. Meeker\textsuperscript{2}

\textsuperscript{1}Air Force Institute of Technology
\textsuperscript{2}Iowa State University

April 4, 2017
1 Background

2 SMRD Package Features

3 Example

4 Remaining Work & Spin-off Projects

5 Summary
Background
Statistical Methods for Reliability Data is a foundational text for analyzing failure time and survival data. Along with the text, the authors developed an S-Plus software package to utilize the methods for industry data. Today, R is the most popular statistical computing language in the world. SMRD is a software package to implement methods from Statistical Methods for Reliability Data in R. This presentation introduces SMRD and details several features of the package.
The R Project for Statistical Computing

- A statistical programming environment for data analysis and graphics
- Developed by Ross Ihaka and Robert Gentleman at the University of Auckland
- Open-source implementation of the ‘S’ language created by Becker et al. at Bell Labs
- A pre-eminent tool for statistics and data science
- One of the fastest growing technical computing languages in the world
  - Used for data processing and visualization, computational statistics, and natural language processing etc.
  - Heavily used by Google, Facebook, Twitter, Microsoft, etc.
In R, the fundamental unit of shareable code is called a package. Packages bundle together code, data, documentation, and tests to easily make analysis methods with others. Currently 225 packages are available on the Comprehensive R Archive Network (CRAN). Many more available from the Bioconductor and GitHub repositories. The huge variety of packages is a key reason why R is so successful. Chances are that someone has already solved a problem that you’re working on. You can benefit from their work by downloading their package.
SMRD - Development History

- Meeker developed a large collection of FORTRAN subroutines as part of contracted efforts at Bell Labs and Iowa State
- Meeker & Escobar wrapped the FORTRAN code into an S-Plus package called SPLIDA (S-Plus Life Data Analysis)
- SPLIDA serves as the companion software for Statistical Methods for Reliability Data 1st ed.
- Meeker attempted to translate SPLIDA into R under the name RSplida
  - Not user-friendly - couldn’t be installed as a traditional R package
  - Couldn’t be used with modern IDE’s (i.e. RStudio, Visual Studio, Eclipse, etc.)
- Freels & Meeker sign MOU to share FORTRAN code for purpose of developing an R package
- Aim is to publish SMRD to the CRAN for use with the 2nd edition of Statistical Methods for Reliability Data
SMRD Package Features
Importing Data From Multiple Sources

- The SMRD package includes over 120 documented datasets
- For importing external data, SMRD leverages several other R packages
- Excel files
  - XLConnect
  - readxl
  - xlsx
- CSV/TSV files
  - base
  - utils
  - readr
- Info, Minitab, S, SAS, SPSS, Stata, Systat and Weka files
  - foreign
  - HMISC
Flexible Event Definitions - Utilize Data As It Exists

- Organizations can use different terms to describe similar events
  - ‘Failure’ = ‘Failed’ = ‘Fail’ = ‘dead’ = ‘died’
  - ‘right’ = ‘rcensored’ = ‘suspended’ = ‘alive’
  - ‘left’ = ‘doa’ = ‘Icensored’
  - ‘interval’ = ‘int’ = ‘icensored’ = ‘grouped’

- SMRD minimizes data pre-processing by accepting many event definitions

- Event definitions can also be mixed

- For users familiar with the R package `survival`
  - SMRD includes functions to map event definitions to the correct numeric values
  - Produce `Surv`-class objects directly
## SMRD Default Event Definitions

<table>
<thead>
<tr>
<th>Failed</th>
<th>Left-censored</th>
<th>Right-censored</th>
<th>Interval-censored</th>
</tr>
</thead>
<tbody>
<tr>
<td>event</td>
<td>1</td>
<td>a</td>
<td>b</td>
</tr>
<tr>
<td>exact</td>
<td>l-censored</td>
<td>alive</td>
<td>bin</td>
</tr>
<tr>
<td>d</td>
<td>left-censored</td>
<td>c</td>
<td>i</td>
</tr>
<tr>
<td>dead</td>
<td>left</td>
<td>censor</td>
<td>interval</td>
</tr>
<tr>
<td>died</td>
<td>leftcensored</td>
<td>censored</td>
<td>i-censored</td>
</tr>
<tr>
<td>f</td>
<td>start</td>
<td>end</td>
<td>intervalcensored</td>
</tr>
<tr>
<td>fail</td>
<td>mstart</td>
<td>mend</td>
<td>intervalcensored</td>
</tr>
<tr>
<td>failed</td>
<td>2</td>
<td>noreport</td>
<td>3</td>
</tr>
<tr>
<td>failure</td>
<td></td>
<td>r</td>
<td>( r )-censored</td>
</tr>
<tr>
<td>report</td>
<td></td>
<td>right-censored</td>
<td></td>
</tr>
<tr>
<td>repair</td>
<td></td>
<td>removed</td>
<td></td>
</tr>
<tr>
<td>repaired</td>
<td></td>
<td>right</td>
<td></td>
</tr>
<tr>
<td>replaced</td>
<td></td>
<td>rightcensored</td>
<td></td>
</tr>
<tr>
<td>replacement</td>
<td></td>
<td>s</td>
<td>survived</td>
</tr>
<tr>
<td>s</td>
<td></td>
<td>survive</td>
<td>suspend</td>
</tr>
<tr>
<td>suspended</td>
<td></td>
<td>suspended</td>
<td>0</td>
</tr>
<tr>
<td>0</td>
<td></td>
<td>( b )-censored</td>
<td></td>
</tr>
</tbody>
</table>
SMRD can estimate reliability measures for many types of failure data

- Multiple failure modes
- Censored observations (right, left, and interval censoring)
- Truncated observations (right, left, and interval truncation)
- Failure times with explanatory variables (normal, Weibull, and logistic regression)
- Repeated measures degradation data
- Repairable system failures (recurring events)
- Physical degradation measures
- Failure times with prior information (Bayesian reliability)
- Reliability growth test data
- Reliability test simulations
Exporting & Visualizing Results

- With SPLIDA, the code was intended to remain underneath a GUI.
- Results of SPLIDA analyses were presented all together:
  - Plots
  - Tables
  - Single numeric values
  - Text summaries
- For GUI-based tools, presenting multiple results simultaneously is **GOOD**.
- For tools emphasizing reproducible research and literate programming, presenting multiple results simultaneously is **BAD**.
- A great deal of effort has gone into ensuring that specific results can be produced and called where desired.
Example
The Shockabsorber Dataset

This example demonstrates a few of the SMRD function to analyze the shockabsorber dataset used throughout the text.

<table>
<thead>
<tr>
<th>miles</th>
<th>mode</th>
<th>event</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>6700</td>
<td>Mode1 Failure</td>
</tr>
<tr>
<td>2</td>
<td>6950</td>
<td>Censored Censored</td>
</tr>
<tr>
<td>3</td>
<td>7820</td>
<td>Censored Censored</td>
</tr>
<tr>
<td>4</td>
<td>8790</td>
<td>Censored Censored</td>
</tr>
<tr>
<td>35</td>
<td>27410</td>
<td>Censored Censored</td>
</tr>
<tr>
<td>36</td>
<td>27490</td>
<td>Mode1 Failure</td>
</tr>
<tr>
<td>37</td>
<td>27890</td>
<td>Censored Censored</td>
</tr>
<tr>
<td>38</td>
<td>28100</td>
<td>Censored Censored</td>
</tr>
</tbody>
</table>
Creating life.data Objects

- The methods in the package require we create a life.data-class object

```r
shock.ld <- frame.to.ld(frame = shockabsorber,
                        response.column = 1,
                        failure.mode.column = 2,
                        censor.column = 3,
                        time.units = 'Miles')
```

- Since SPLIDA was written for a GUI, functions to produce results already existed
- Thus, once the life.data object has been created, many different plots and numeric results can be produced with a single line of code
# Producing Results From life.data Objects

<table>
<thead>
<tr>
<th>Plots</th>
<th>Numeric Results</th>
</tr>
</thead>
<tbody>
<tr>
<td>- Nonparametric CDF plots</td>
<td>- $F(t)$ at specified values of $t$</td>
</tr>
<tr>
<td>- Parametric CDF plots</td>
<td>- $h(t)$ at specified values of $t$</td>
</tr>
<tr>
<td>- ML CDF and hazard plots</td>
<td>- $t^{-1}(p)$ at specified values of $p$</td>
</tr>
<tr>
<td>- Explanatory variable plots</td>
<td>- ML parameter estimates and standard errors</td>
</tr>
<tr>
<td>- Multi-failure mode plots</td>
<td>- Logit and log transformed confidence intervals</td>
</tr>
<tr>
<td>- Likelihood surfaces</td>
<td>(pointwise and simultaneous)</td>
</tr>
<tr>
<td>- Relative likelihood curves</td>
<td></td>
</tr>
</tbody>
</table>
Nonparametric CDF plots

`plot(shock.ld)`
ML Plots ($F(t)$ & $h(t)$)

```r
mleprobplot(shock.ld, distribution = 'weibull')
mlehazplot(shock.ld, distribution = 'weibull')
```
Integration with Rmarkdown & Shiny

- By taking advantage of the existing SPLIDA code, SMRD is ideal for use with the shiny and rmarkdown packages
  - Allows for a seamless reliability workflow to create papers & presentations incredibly fast
  - The package can be used to create high-quality plots with a single line of code
  - Turn results into LaTeX tables instantly
  - Create interactive content
Remaining Work & Spin-off Projects
Remaining Work

Several tasks must be completed before the package can be published on the CRAN.

The following tasks are listed in order of time required to complete:

- Convert underlying FORTRAN code to C++
- Document functions
- Separate results for middle chapters
- Update plots
- Integrate shiny gadgets and rmarkdown templates to make the analysis workflow super-fast
Spin-Off Projects

- Teaching with the SMRD package throughout the development has spawned several projects
- SMRD.apps - R package of interactive visualizations and examples from the text
- A companion text to help SMRD readers learn R programming skills and along with the course content
- SMRD.resources R package for instructors to automatically create problem sets, homework solutions and in-class examples from the text
- teachingApps R package containing 150 apps to help students and instructors visualize statistical concepts and R programming
Summary
SMRD is an upcoming R implementing the methods presented in the Statistical Methods for Reliability Data in stuff goes here

Plan is to release in conjunction with the second edition on the book

Always looking for help testing the package
QUESTIONS?