Style Guide for Loss Data Analytics

An open text authored by the Actuarial Community

2020-08-13

Contents

Preface

It is important for chapters in the *Open Actuarial Textbooks* project to have a consistent look and feel. The *Style Guide for Loss Data Analytics* aims to assist contributors in developing chapters consistently throughout the *Loss Data Analytics* book.

This guide is set up as sample chapters containing

- information regarding suitable content for a chapter,
- methods to implement chapter elements and
- conventions for consistent notation.

Chapter 1

Chapter Structure

In this chapter, you learn how to:

- Determine what and what not to include in a chapter
- Include technical supplements as needed
- Assess types of exercises and book resources that are appropriate for a chapter

---

1.1 Chapter Preview and Learning Objectives

- *Chapter Preview.* Begin with a chapter preview to set the stage of the chapter.

- *Learning Objectives.* At the beginning of each chapter and section, describe in a few bullet points what the reader can expect to learn.

1.2 Main Body

Split the chapter into 4-7 sections; within each section, introduce 0-5 subsections. Do not develop a deeper hierarchy (e.g., a "sub-subsection"). Use nonlinear aspects of the web. For example, detailed mathematical developments can go into a technical appendix or are simply hidden (using javascript "hide/show" tools) unless the viewer really wants to see the details. Case studies and historical references can be included in "side-bars," a supporting webpage. For the main body of the chapter, think about "25 pages" in length (whatever that means....).

1.2.1 What to Include

- Within the chapter, use boxed and numbered lists of procedures for easy reference.
- It is certainly okay (and expected) to use mathematical notation although please adhere to the conventions described in Section 5. Each chapter should have examples interwoven within theory, allowing readers to see the development of the theory along with the importance of the applications.
- Distinguish between an *"Example"* and a *"Special Case"*. The former shows how to relate the mathematics to a practical situation likely to be encountered by a practicing actuary. The latter looks at a subset of a general (usually) mathematical result. A few special cases are certainly acceptable but we want to focus on developing examples.
- Think of graphical ways to visualize/summarize relationships that you want to emphasize.
- Begin each section with a short bullet list describing the learning objectives of that section. Finish each section with a short quiz on these learning objectives. As of this writing (July 2018), quizzes are multiple-choice.
- Include short exercises/examples/special cases that can be readily solved by the viewer (with solutions using "hide/show" features) within the main body. These serve to reinforce concepts and provide benchmarks for understanding.

1.2.2 What Not to Include

- Do not include development of equations/formulas in the main body of the text. The main body of the text will be devoted to presenting results, providing context and intuition as to the importance of the results.
- Do not include references to the literature. This will appear in the last section on "Further Reading and References."
- Do not include graphs whose information could easily be summarized by a table.

1.3 Technical Supplements

We want our viewers to understand the underpinnings of the theory (the old analogy of "what is going on under the hood to see how the engine works" - no black boxes.) So, there will be occasions when you feel like a short development or "proof"; is reasonable. Put this in an appendix. Technical supplements should develop the theory in a step-by-step fashion, building on each concept in a crisp, mathematical fashion.

1.4 Contributors and Further Resources

1.4.1 Contributors

Make sure that contributors are listed at the end of the chapter. The following provides an example.

```
####Contributors {-}

- Edward W. (Jed) Frees, University of Wisconsin-Madison,
is the principal author of the initital version of this chapter.
Email: jfrees@bus.wisc.edu for chapter comments and
suggested improvements. Helpful improvements provided by
Alyaa Nuval Binti Othman and Aisha Nuval Binti Othman.
```

1.4.2 Further Reading and References

Do not finish with a "preview of upcoming chapter"; finish instead with a "Further Reading and References." This consists of a series of references with one or two lines of annotation for each reference that the interested reader could follow up on (self-citations are okay!). Historical developments are particularly nice in this section.

Chapter 2

Interactive Book Features

An advantage of publishing on the web is that we can produce an online version that contains many interactive objects such as quizzes, computer demonstrations, interactive graphs, video, and the like, to promote **deeper learning**.

We want viewers to interact with the book; it is **not** written to accommodate the "armchair reader," that is, one who passively reads and does not get involved by attempting the exercises in the text. Consider an analogy to sports; there is a great deal that you can learn about the game just by watching. However, if you want to sharpen your skills, then you have to go out and play the game.

##Examples and Exercises

Even for traditional offline (pdf) version of the book, we include problems that allow readers to develop their learning.

Each chapter contains several "Examples" or "Exercises" - these are focused problems that generally ask the reader to do a calculation or provide an interpretation of a statistical issue. We call them "examples" when they appear in the main body of the text and "exercises" when they appear at the end of the chapter. The need for hand calculations has been advocated by Khamis (1991)[1]. Many teachers have found it useful to practice the mechanics in order to provide a solid foundation. With this foundation, we can get on to the real business; interpreting data using statistical principles.

We anticipate that substantial exercise banks will be built over time by users, professional associations, and those with commercial interests. When developing the chapter foundations, think about the following types of problems.

- **Hand Calculation.** These are problems that can be solved without the use of a computer. They typically reinforce a statistical or actuarial con-

[1] "Manual Computations—A Tool for Reinforcing Concepts and Techniques," H. J. Khamis, Pages 294-299, *The American Statistician*, Volume 45, 1991 - Issue 4

cept as well as highlight an issue that might be encountered in practice. For this book, these problems often involve algebraic and calculus manipulations. When writing these types of problems, readers are more motivated if they understand *why* the problem is important. What actuarial issue is being addressed? What statistical concept is being practiced? Often, a simple line or two prefacing a problem can help substantially in this regard.
- **Software.** These are problems that ask the reader to work with `R` software, such as calculating a function or reproducing a graph.
- **Data.** These are problems that ask the reader to work with data. The need for working with real data is well documented; for example, see Hogg (1972), Moore and Roberts (1989) or Singer and Willett (1990). By providing detailed guided tutorials that work with theory and data, we teach our students the essence of *Loss Data Analytics*. Of course, there are some important disadvantages to working with real data. Data sets can quickly become outdated. Further, the ideal data set to illustrate a specific statistical issue is difficult to find. Data exercises are complex and can span several chapter sections as well as chapters.

Examples and exercises are designed so that they illustrate a general analytics/statistical concept; they have the additional advantage in that they are often based on actuarial professional examinations and so provide readers with training for these assessment frameworks. Notably, *in the online version of the book, the solutions to the examples and exercises are hidden to encourage readers to actively solve (or at least consider) prior to revealing a solution.*

##Statistical Code

However, the limitation of exercise and examples done by hand is that they give **little insight** as to how the general analytics/statistical concepts can be used in applications or why more extensive treatments beyond the foundations would be needed. To address these limitations, throughout we include illustrate statistical code with sample (but real) datasets. The statistical language used throughout is `R`. For two reasons, the `R` code is hidden in the online version but can be interactively revealed but clicking a button.

- First, we want to focus on the analytics/statistical concepts and do not wish for readers to be distracted by software code that emphasizes implementation, not concepts.
- Second, not all readers will use `R` (there are many good alternative software programs available) and we want the book to be available to a broad readership.

Although we do not focus on developing `R` tutorials, we will provide guides and links to people who wish to learn `R`. Our focus is on teaching statistical methods and actuarial issues, not software. Over time, the project may also provide support for users of other software environments, such as Microsoft's Excel or Python.

##Other Interactive Features

A wonderful aspect of an online text is that we can readily incorporate other interactive features. In coming versions of the book, you can expect to see

- a glossary (hover the cursor over a technical word or phrase to reveal a definition),
- links to relevant applications of the basic concept, as well as
- end of the section quizzes that provide "low-level formative assessments" so that they reader can gauge his or her understanding of the material.

Some of the support material associated with the book also emphasizes interactive aspects. For example, detailed *R code* allows readers to learn complex statistical routines and provides sample code so that users can develop their own libraries of useful routines. Another site will feature `R-Shiny`, an interactive graphic tool that provides dynamic graphing features. The concept of active learning promotes one of the deeper learning goals set forth by some educational leaders: The ability to learn how to learn independently.

2.1 Additional Book Resources Supporting Each Chapter

We are hopeful that there will be several resources support the book that will appear outside of the chapter structure. Although not necessarily interactive, they will help develop users' learning. These include:

- **Case Studies and Historical Vignettes.** Similar to those appearing within the chapter, include short exercises/examples/special cases that can be readily solved by the viewer. These serve to reinforce concepts and provide benchmarks for understanding. Case studies can be used to emphasize different practices in different countries. Historical vignettes can be interesting in their own right and remind us all of the foundations of our discipline.
- **Data.** We anticipate developing a library of data sets that can be used by instructors who wish to emphasize different areas of practice.
- **Technical Supplements, Lists, and Tables.** The roles of technical supplements has already been described and there could be many. As is common in textbooks, we will also provide a place for lists or tables of organized facts for learners.

Chapter 3

Samples of Writing in `R bookdown`

In this chapter, you learn how to:

- Reference other sections and equations
- Include in-text citation that links to the bibliography
- Include tables and figures not generated by `R` code
- Include a footnote
- Add tooltip descriptions for technical phrases
- Include chapter examples and R code

As we expand our contributor and reviewer base, it will be helpful to know more about the conventions used in the series regarding the details of `R markdown` and `R bookdown` used in the series. This chapter summarizes these conventions.

3.1 Section Labels and Learning Objectives

The following shows how to code Section titles and refer to them.

```
## Section Labels {#S:SectionLabels}
```

With that reference, one can readily refer to Section 3.1 in your text, as follows:

```
With that reference, one can readily refer to
Section \\ref{S:SectionLabels} in your text, as follows:
```

The following shows how to code learning objectives:

```
***
In this chapter, you learn how to:

- Reference other sections and equations
- Include in-text citation that links to the bibliography
- Include tables and figures
- Include a footnote

***
```

3.2 Equation References

Here is an example of a latex equation produced in `R markdown`, with reference number.

$$x + y = 1 \tag{3.1}$$

You can produce that equation using the following code.

```
\begin{equation}
  x + y = 1
\label{eq:ExampleEquation}
\end{equation}
```

With this, equation (3.1) can be referred to using the following code:

```
With this, equation \\eqref{eq:ExampleEquation} can be
referred to using the following code:
```

3.3 In-text Citations

Here is an example of an in-text citation made possible by `R bookdown` (Xie, 2015). This links to the bibliography where the full referece is displayed. As a convention we use the *APA* style citation.

```
Here is an example of an in-text citation made possible by
`R bookdown` [@xie2015]. This links to the bibliography
where the full reference is displayed.
As a convention we use the *APA* style citation.
```

3.4 Including Tables

We want to be able to include Latex tables (for mathematical type), as well as data-drive tables produced by `R`. In order to do that, we use html syntax in order to reference tables. This means that we have to number the tables by hand. Although a bit painful, it does gives us the flexibility needed.

Start with a Latex generated table.

Table 2.1. An Example of Including Tables using Latex in an `R markdown` Document

Policyholder	Number of claims
X	1
Y	2

`R markdown` does not have a convention for referencing non-R generated tables. For now, we reference them manually as in refer to Table 2.1. We do this by manually inserting an html anchor tag.

The following code produces this table.

```
<a id=tab:2.1></a>

[Table 2.1]: \#tab:2.1

##### Table 2.1. An Example of Including Tables using Latex in an `R markdown` Document {-}

$$
\begin{matrix}
    \begin{array}{c|c} \hline
    \text{Policyholder} & \text{Number of claims} \\\hline
    \textbf{X} & 1 \\\hline
    \textbf{Y} & 2 \\\hline
    \end{array}
\end{matrix}
$$
```

For reference, then use

```
`R markdown` does not have a convention for referencing
non-R generated tables. For now, we reference them manually
as in refer to [Table 2.1].
```

Now we give a data-driven table generated by `R` in Table 2.2.

Table 2.2. An Example of Including Tables using `R`

Minimum	First Quartile	Median	Mean	Third Quartile	Maximum
167	2226	4951	56332	11900	12922218

Here is the code to produce this table.

```
<a id=tab:2.2></a>

[Table 2.2]: \#tab:2.2
```

~~~~
```r
Insample <- read.csv("Insample.csv", header=T,
              na.strings=c("."), stringsAsFactors=FALSE)
Insample2010 <- subset(Insample, Year==2010)
InsamplePos2010 <- subset(Insample2010, yAvg>0)
summaryOutput <- t(as.matrix(summary(InsamplePos2010$yAvg)))
colnames(summaryOutput) <- c("Minimum","First Quartile",
                             "Median", "Mean", "Third Quartile", "Maximum")
kable_styling(knitr::kable(summaryOutput,digits=0,
                    align = "rrrrrr"), latex_options="scale_down")
```
~~~~

3.5 Including Figures

3.5.1 Figures Generated by R

Most figures are generated using R. Here is an illustrative figure.

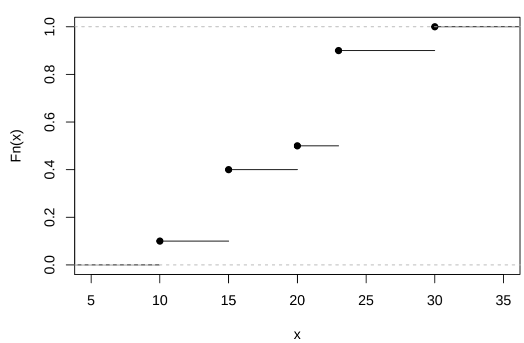

Figure 3.1: Empirical Distribution Function of a Toy Example

that we refer to as Figure 3.1. Here is the code for producing the figure:

~~~~
```{r EDFToy, echo = FALSE,
   fig.cap = 'Empirical Distribution Function of a Toy Example',
   out.width = '60%', fig.asp = 0.75, fig.align = 'center'}
~~~~

~~~
xExample <- c(10,rep(15,3),20,rep(23,4),30)
PercentilesxExample <- ecdf(xExample)
plot(PercentilesxExample, main = "", xlab = "x")
```
~~~

Here is is the code for referencing the Figure 3.1:

```
Here is is the code for referencing the Figure \\ref{fig:EDFToy}:
```

3.5.2 Figures Not Generated by R

For figures, we store the figures as png or jpeg files in a separate folder called "Figures". Then we use R code to call those figures for display so that we can reference them.

Here is such a figure:

Figure 3.2: An example of including figures in an R Markdown document

And here is the code that generates the figure:

```
"three backticks"{r, ExampleFigure, fig.cap = 'An example
of including figures in an R Markdown document',
out.width = '5%', fig.align = 'center', echo = FALSE}
knitr::include_graphics("Figures/RStudio-Ball.png")
"three backticks"
```

Here is is the code for referencing the Figure 3.2:

```
Here is is the code for referencing the Figure \\ref{fig:ExampleFigure}:
```

3.6 Including Footnotes

Try to minimize the use of footnotes. But, if you need them, here is how you can include a footnote [1].

```
Here is how you can include a footnote [^1].

[^1]: the footnote displays at the end of the chapter
```

[1] the footnote displays at the end of the chapter

3.7 Defining Glossary Terms

Use the `Gloss(term)` function to include a tooltip for technical phrases. `Gloss(term)` looks up the definition in `GlossFct.csv`. Accordingly, you can add new glossary terms for your chapter in the `csv` file.

Here is how you add a tooltip:

We can use the same `Gloss(term)` function to include tooltips for acronyms (e.g. iid) as listed in Section 5.2.

3.7.1 Glossary Conventions

- Include tooltip for a term only once in a chapter.
- Include tooltip when the term is defined. (e.g. "Frequency is …")
- If term is not explicitly defined, include tooltip at the term's first occurence in the chapter.
- Only include tooltip in the Chapter Preview or in examples if the term is not defined elsewhere in the chapter.

3.7.2 Glossary Terms with Multiple Definitions

A term may carry multiple definitions across different sections. In this case, use `Gloss(term, section)`:

3.8 Including Chapter Examples and R Code

As described in Section 2, chapter examples and R code allow readers to interact with the online version of the book.

Here is how to add a chapter example:

Similarly, you can add statistical code:

and also proof theory:

`HideExample`, `HideRCode`, and `HideProofTheory` are R functions that can hide these interactive features in the offline (pdf) version of the book. All you need to do is change `eval = TRUE` in `PdfOutput.Rmd`.

3.9 Useful Links

Naturally, you will want to learn more about coding in `R markdown`, `R bookdown` and so forth. The following provide some useful links for taking the next step.

- For an `R markdown` guide refer https://rmarkdown.rstudio.com/authoring_pandoc_markdown.html.
- For a `R bookdown` guide, see https://bookdown.org/yihui/bookdown/.

- For best practices in coding `R`, we suggest http://r-pkgs.had.co.nz/style.html.
- See also our online actuarial text resources at https://sites.google.com/a/wisc.edu/loss-data-analytics/online-actuarial-text-resources.

Chapter 4

Setting up a Glossary

In this chapter, you learn how to:

- Setup csv files for a glossary
- List terms and definitions by chapter
- List terms by the chapter in which they are first defined
- Use GitHub to store the project, make changes, seek feedback and facilitate collaboration

---

A glossary can serve as a quick reference when the reader needs to recall a term's definition or location within a book. This chapter can serve as a guide to setup a glossary in the `R Bookdown` environment. Please refer here for a refresher on `R Bookdown` if necessary.

4.1 csv File Setup

The following outlines how we currently store our glossary:

1. Setup an excel file with 3 columns: Term, Definition, Chapter first defined.
2. We suggest using one excel file for all chapters for easy reference and sharing. However, you will have to save each chapter separately. Alternatively, you may use one excel file per chapter instead of having one central file then separating it.
3. Save the Excel file as a csv file.
4. Please note that if you make changes, do so in the Excel file not the csv file.

4.2 Terms and Definitions by Chapter

This section describes how to create a list of important terms and their definitions listed according to the chapters in which they are found.

We use an example from the glossary of Loss Data Analytics. First we display the list of terms and definitions for Chapter 1 of the book followed by the code used to generate it.

4.2.1 Chapter 1 Introduction to Loss Data Analytics

Term	Definition
loss adjustment expenses	Loss adjustment expenses are costs to the insurer that are directly attributable to settling a claims. For example, the cost of an adjuster is someone who assess the claim cost or a lawyer who becomes involve in settling an insurer's legal obligation on a claim
unallocated loss adjustment expenses	Unallocated loss adjustment expenses are costs that can only be indirectly attributed to claim settlement; for example, the cost of an office to support claims staff
allocated loss adjustment expenses	Allocated loss adjustment expenses, sometimes known by the acronym ALEA, are costs that can be directly attributed to settling a claim; for example, the cost of an adjuster
indemnification	Indemnification is the compensation provided by the insurer.
insurance claim	An insurance claim is the compensation provided by the insurer for incurred hurt, loss, or damage that is covered by the policy.
loss amount	The loss amount is the size of the loss incurred by the policyholder for incurred hurt, loss, or damage that is covered by the policy.
analytics	Analytics is the process of using data to make decisions.

Term	Definition
renters insurance	Renters insurance is an insurance policy that covers the contents of an apartment or house that you are renting.
homeowners insurance	Homeowners insurance is an insurance policy that covers the contents and property of a building that is owned by you or a friend.
automobile insurance	An insurance policy that covers damage to your vehicle, damage to other vehicles in the accident, as well as medical expenses of those injured in the accident.
property insurance	Property insurance is a policy that protects the insured against loss or damage to real or personal property. The cause of loss might be fire, lightening, business interruption, loss of rents, glass breakage, tornado, windstorm, hail, water damage, explosion, riot, civil commotion, rain, or damage from aircraft or vehicles.
nonlife insurance	Nonlife insurance is any type of insurance where payments are not based on the death (or survivorship) of a named insured. Examples include automobile, homeowners, and so on. Also known as property and casualty or general insurance.
casualty insurance	Causalty insurance is a form of liability insurance providing coverage for negligent acts and omissions. Examples include workers compensation, errors and omissions, fidelity, crime, glass, boiler, and various malpractice coverages.
valuation date	A valuation date is the date at which a company summarizes its financial position, typically quarterly or annually.

Term	Definition
underwriting	Underwriting is the process where the company makes a decision as to whether or not to take on a risk.
ratemaking	
reinsurer	A reinsurer is an insurance company that offers insurance to an insurer.
loss reserve	A loss reserve is an estimate of liability indicating the amount the insurer expects to pay for claims that have not yet been realized. This includes losses incurred but not yet reported (IBNR) and those claims that have been reported claims that haven't been paid (known by the acronym RBNS for reported but not settled).
technical provisions	Technical provisions is another name for loss reserves.
experience rating	
merit rating	
risk classification	Risk classification is the process of grouping policyholders into categories, or classes, where each insured in the class has a risk profile that is similar to others in the class.
cream skimming	
claims triage	
pure premium	Pure premium is the total severity divided by the number of claims. It does not include insurance company expenses, premium taxes, contingencies, nor an allowance for profits. Also called loss costs. Some definitions include allocated loss adjustment expenses (ALAE).

Term	Definition
loss cost	Loss cost is the total severity divided by the number of claims. It does not include insurance company expenses, premium taxes, contingencies, nor an allowance for profits. Also called pure premium. Some definitions include allocated loss adjustment expenses (ALAE).
rating variables	
coinsurance	Coinsurance is an arrangement whereby the insured and insurer share the covered losses. Typically, a coinsurance parameter specified means that both parties receive a proportional share, e.g., 50%, of the loss.
deductible	A deductible is a parameter specified in the contract. Typically, losses below the deductible are paid by the policyholder whereas losses in excess of the deductible are the insurer's responsibility (subject to policy limits and coninsurance).
policy limit	A policy limit is the maximum value covered by a policy.
personal lines	
dividend	A dividend is the refund of a portion of the premium paid by the insured from insurer surplus.
bonus	
retrospective premiums	The process of determining the cost of an insurance policy based on the actual loss experience determined as an adjustment to the initial premium payment.
prospective premiums	
claims adjustment	Claims adjustment is the process of determining coverage, legal liability, and settling claims.

Term	Definition
Commercial line	Commercial line is insurance purchased by commercial ventures (businesses)
line of business	A line of business is a classification of business written by insurers.
claims leakage	Claims leakage respresents money lost through claims management inefficiencies.
fraud detection	
case reserve	A case reserve is an estimate of the insurer's future liability made by the claims adjuster.
adjuster	An adjuster is a person who investigates claims and recommends settlement options based on estimates of damage and insurance policies held.
life Insurance	Life insurance is a contract where the insurer promises to pay upon the death of an insured person. The person being paid is the beneficiary.
capital allocation	

Here is the code used for producing the list of terms and definitions:

```
### Chapter 1 Introduction to Loss Data Analytics

{r}
library(pander)
chapter1 <- read.csv("csv/Chapter1.csv", header=TRUE,
                     na.strings=c("."), stringsAsFactors=FALSE)
table1.1 <- cbind(chapter1[, 1], chapter1[, 2])
final.table1.1 <- as.data.frame(table1.1)
names(final.table1.1) <- c("Term", "Definition")
pander(final.table1.1)
```

4.3 Terms and Chapter First Defined

This section describes how to create a list of terms by the chapter in which they are first defined. Certain terms are defined multiple times throughout the book, so this list can help the reader refer to the chapter in which a term is first used and defined. The terms listed here are sorted in alphabetical order.

We use an example from the glossary of Loss Data Analytics. Here, we use Chapter 1 and Chapter 2 of the book. We display the list of terms by chapter first defined then show the code used to generate it.

```
Warning: package 'dplyr' was built under R version 3.6.3
```

Term	Chapter first defined
adjuster	1
aggregate claims	2
allocated loss adjustment expenses	1
analytics	1
automobile insurance	1
Bernoulli distribution	2
Binomial distribution	2
bonus	1
capital allocation	NA
case reserve	1
casualty insurance	1
claims adjustment	1
claims leakage	1
claims triage	1
coinsurance	1
Commercial line	1
cream skimming	1
deductible	1
Distribution function F(x)	2
dividend	1
experience rating	1
fraud detection	1
Frequency	2
Gamma Distribution	2
homeowners insurance	1
indemnification	1
insurance claim	1
life Insurance	NA
line of business	1
loss adjustment expenses	1
loss amount	NA
loss cost	1
loss reserve	1
Maximum Liklihood Estimator	2
merit rating	1
Mixture	2

Term	Chapter first defined
Moment generating function	2
Negative binomial	2
nonlife insurance	1
personal lines	1
Poisson	2
policy limit	1
Probability generating function	2
Probability mass function f(x)	2
property insurance	1
prospective premiums	1
pure premium	1
ratemaking	1
rating variables	1
reinsurer	1
renters insurance	1
retrospective premiums	1
risk classification	1
Severity	2
Survival function S(x)	2
technical provisions	1
unallocated loss adjustment expenses	1
underwriting	1
valuation date	1
Zero Modifided Distribution	2
Zero Truncated Distribution	2

Here is the code used for producing the list of terms by chapter first defined:

```
{r}
# Chapter 1
table2.1 <- cbind(chapter1[, 1], chapter1[, 3])

# Chapter 2
table2.2 <- cbind(chapter2[, 1], chapter2[, 3])

# Concatenate tables
table2 <- rbind(table2.1, table2.2)

# Sort alphabetically --> do not change
sort.table2 <- table2[order(table2[,1]), ]

# Remove duplicates --> do not change
library(dplyr)
```

```
final.table2 <- as.data.frame(sort.table2)
names(final.table2) <- c("Term", "Chapter first defined")

# Generate table --> do not change
pander(distinct(final.table2, Term, .keep_all= TRUE))
```

Note that some lines of code say "do not change". This is because these lines of code apply to the concatenated table which includes all the chapters. We need the concatenated table before we can sort all terms alphabetically and remove duplicates to generate the final table.

4.4 Glossary on GitHub

This section describes how to setup a glossary repository on GitHub. By doing so, users can store the project, make changes, seek feedback and facilitate collaboration. We include suggestions for these different types of users:

- someone who wants to do a book like ours
- the reviewer/reader who simply wants to suggest altering or adding a definition
- a contributing author who needs to compile a csv file of definitions and include them into the chapter using tooltip

4.4.1 Repository Creation to Supplement a Book

The following are suggestions on how to setup a repository on GitHub to store your glossary project. Here, we do not get into GitHub features in detail but we do suggest a place which you can refer to.

1. We suggest referring to Happy Git to get started on setting up GitHub and linking it to R Studio.
2. Once you have done so, you can store and update your glossary project on GitHub. Happy Git describes how you can make changes locally, commit and push the changes to GitHub.

4.4.2 Feedback from Reviewers/ Readers

We assume that the author already has a glossary repository on GitHub in order to use the issue feature to receive feedback. The following is an excerpt out of the glossary for Loss Data Analytics on how readers can make suggestions:

When using the glossary, we encourage the reader to provide feedback regarding the terms and their definitions. For example, if the reader feels that there is a better definition for a particular term, the following instructions outline how the reader can suggest improvements.

- First, open up the issues tab on our repository on GitHub.
- Click on "create an issue".

- Indicate which chapters you want to make changes to in the title.
- Specify the terms and definitions you wish to change, add or remove.
- Click "Submit new issue".

4.4.3 Collaboration from Authors

4.4.3.1 Definitions Compilation

Aside from readers, collaborators can also contribute to the glossary. For example, professors can get authors to assist in compiling definitions.

Collaborators can setup their own GitHub accounts. They can fork the project, make changes locally and make a pull request. The project owner can then merge these changes to update the project. These processes are outlined on Happy Git as well.

As mentioned previously, we suggest compiling definitions in the Excel file instead of csv. This is because making changes directly to the csv file may result in space distortions in the R output of the glossary.

4.4.3.2 Definitions in-text using `Tooltip`

Further, in Loss Data Analytics, we use `tooltip` which allows readers to hover over a word in the text so that they may receive the definition as in the following example from the introduction chapter:

When introducing data methods, we will focus on losses that arise from obligations in insurance contracts. This could be the amount of damage to one's apartment under a renter's insurance agreement, the amount needed to compensate someone that you hurt in a driving accident, and the like. We call these obligations *insurance claims* An insurance claim is the compensation provided by the insurer for incurred hurt, loss, or damage that is covered by the policy.. With this focus, we will be able to introduce generally applicable statistical tools and techniques in real-life situations.

The following is the `tooltip` code associated with the above output:

```
<a href="#" class="tooltip" style="color:green">*insurance claims*<span style="font-si
```

Note that our version of `tooltip` is customized within our style.css file. If you prefer another style, you will have to modify the code or replace the style.css file to suit your needs.

Chapter 5

Conventions for Notation

Chapter Preview. Loss Data Analytics will serve as a bridge between actuarial problems and methods and widely accepted statistical concepts and tools. Thus, the notation should be consistent with standard usage employed in probability and mathematical statistics. See, for example, (Halperin et al., 1965) for a description of one standard.

5.1 General Conventions

- Random variables are denoted by upper-case italicized Roman letters, with X or Y denoting a claim size variable, N a claim count variable, and S an aggregate loss variable. Realizations of random variables are denoted by corresponding lower-case italicized Roman letters, with x or y for claim sizes, n for a claim count, and s for an aggregate loss.
- Probability events are denoted by upper-case Roman letters, such as Pr(A) for the probability that an outcome in the event ''A" occurs.
- Cumulative probability functions are denoted by $F(z)$ and probability density functions by the associated lower-case Roman letter: $f(z)$.
- For distributions, parameters are denoted by lower-case Greek letters. A caret or ''hat" indicates a sample estimate of the corresponding population parameter. For example, $\hat{\beta}$ is an estimate of β .
- The arithmetic mean of a set of numbers, say, $x_1, \ldots, x_n$, is usually denoted by $\bar{x}$; the use of x, of course, is optional.
- Use upper-case boldface Roman letters to denote a matrix other than a vector. Use lower-case boldface Roman letters to denote a (column) vector. Use a superscript prime "′" for transpose. For example, $\mathbf{x}'\mathbf{A}\mathbf{x}$ is a quadratic form.
- Acronyms are to be used sparingly, given the international focus of our audience. Introduce acronyms commonly used in statistical nomenclature

but limit the number of acronyms introduced. For example, *pdf* for probability density function is useful but *GS* for Gini statistic is not.

5.2 Abbreviations

Here is a list of abbreviations that we adopt. We italicize these acronyms. For example, we can discuss the goodness of fit in terms of the *AIC* criterion.

AIC	Akaike information criterion
BIC	(Schwarz) Bayesian information criterion
cdf	cumulative distribution function
df	degrees of freedom
iid	independent and identically distributed
glm	generalized linear model
mle	maximum likelihood estimate/estimator
ols	ordinary least squares
pdf	probability density function
pmf	probability mass function

5.3 Common Statistical Symbols and Operators

Here is a list of commonly used statistical symbols and operators, including the latex code that we use to generate them (in the parens).

$I(\cdot)$	binary indicator function (I). For example, $I(A)$ is one if an outcome in event A occurs and is 0 otherwise.
$\Pr(\cdot)$	probability (`\Pr`)
$\mathrm{E}(\cdot)$	expectation operator (`\mathrm{E}`). For example, $\mathrm{E}(X) = \mathrm{E}\ X$ is the expected value of the random variable X, commonly denoted by μ.
$\mathrm{Var}(\cdot)$	variance operator (`\mathrm{Var}`). For example, $\mathrm{Var}(X) = \mathrm{Var}\ X$ is the variance of the random variable X, commonly denoted by σ^2.
$\mu_k = \mathrm{E}\ X^k$	kth moment of the random variable X. For k=1, use $\mu = \mu_1$.
$\mathrm{Cov}(\cdot,\cdot)$	covariance operator (`\mathrm{Cov}`). For example, $\mathrm{Cov}(X,Y) = \mathrm{E}\{(X - \mathrm{E}\ X)(Y - \mathrm{E}\ Y)\} = \mathrm{E}(XY) - (\mathrm{E}\ X)(\mathrm{E}\ Y)$ is the covariance between random variables X and Y.
$\mathrm{E}(X\|\cdot)$	conditional expectation operator. For example, $\mathrm{E}(X\|Y = y)$ is the conditional expected value of a random variable X given that the random variable Y equals y.
$\Phi(\cdot)$	standard normal cumulative distribution function (`\Phi`)
$\phi(\cdot)$	standard normal probability density function (`\phi`)
$\sim$	means is distributed as (`\sim`). For example, $X \sim F$ means that the random variable X has distribution function F.
$se(\hat{\beta})$	standard error of the parameter estimate $\hat{\beta}$ (`\hat{\beta}`), usually an estimate of the standard deviation of $\hat{\beta}$, which is $\sqrt{Var(\hat{\beta})}$.
H_0	null hypothesis
H_a or H_1	alternative hypothesis

5.4 Common Mathematical Symbols and Functions

Here is a list of commonly used mathematical symbols and functions, including the latex code that we use to generate them (in the parens).

$\equiv$	identity, equivalence (`\equiv`)
$a := b$	defines a in terms of b
$\implies$	implies (`\implies`)
$\iff$	if and only if (`\iff`)
$\to, \longrightarrow$	converges to (`\to`, `\longrightarrow`)
$\mathbb{N}$	natural numbers $1, 2, \ldots$ (`\mathbb{N}`)
$\mathbb{R}$	real numbers (`\mathbb{R}`)
$\in$	belongs to (`\in`)
$\notin$	does not belong to (`\notin`)
$\subseteq$	is a subset of (`\subseteq`)
$\subset$	is a proper subset of (`\subset`)
$\cup$	union (`\cup`)
$\cap$	intersection (`\cap`)
$\emptyset$	empty set (`\emptyset`)
A^c	complement of A
$g * f$	convolution $(g * f)(x) = \int_{-\infty}^{\infty} g(y) f(x-y) dy$
$\exp$	exponential (`\exp`)
$\log$	natural logarithm (`\log`)
$\log_a$	logarithm to the base a
$!$	factorial
$\mathrm{sgn}(x)$	sign of x(`sgn`)
$\lfloor x \rfloor$	integer part of x, that is, largest integer $\leq x$ (`\lfloor`, `\rfloor`)
$\lvert x \rvert$	absolute value of scalar x
$\varGamma(x)$	gamma (generalized factorial) function (`\varGamma`), satisfying $\varGamma(x+1) = x\varGamma(x) (\varGamma)$
$B(x, y)$	beta function, $\varGamma(x)\varGamma(y)/\varGamma(x+y)$

5.5 Further Readings

To make connections to other literatures, see (Abadir and Magnus, 2002) http://www.janmagnus.nl/misc/notation.zip for a summary of notation from the econometrics perspective. This reference has a terrific feature that many latex symbols are defined in the article. Further, there is a long history of discussion and debate surrounding actuarial notation; see (Boehm et al., 1975) for one contribution.

Bibliography

Abadir, K. and Magnus, J. (2002). Notation in econometrics: a proposal for a standard. *The Econometrics Journal*, 5(1):76–90.

Boehm, C., Engelfriet, J., Helbig, M., Kool, A. I., Leepin, P., Neuburger, E., and Wilkie, A. (1975). Thoughts on the harmonization of some proposals for a new international actuarial notation. *Blätter der DGVFM*, 12(2):99–129.

Halperin, M., Hartley, H. O., and Hoel, P. G. (1965). Recommended standards for statistical symbols and notation: Copss committee on symbols and notation. *The American Statistician*, 19(3):12–14.

Xie, Y. (2015). *Dynamic Documents with R and knitr*. Chapman and Hall/CRC, Boca Raton, Florida, 2nd edition. ISBN 978-1498716963.