## Chapter 1 - Revisions

Keep chapter structure as is but change title headers to emphasize the use of data in insurance activities.

**Current Version:**

1. Introduction to Loss Data Analytics

1.1 Relevance of Analytics to Insurance Activities

1.1.1 Nature and Relevance of Insurance

1.1.2 What is Analytics?

1.1.3 Insurance Processes

1.2. Insurance Company Operations

1.2.1 Initiating Insurance

1.2.2 Renewing Insurance

1.2.3 Claims and Product Management

1.2.4 Loss Reserving

1.3. Case Study: Wisconsin Property Fund

1.3.1 Fund Claims Variables: Frequency and Severity

1.3.2 Fund Rating Variables

1.3.3 Fund Operations

**Proposed Changes:**

1. Loss Data and Insurance Activities

1.1 Data Driven Insurance Activities

1.1.1 Nature and Relevance of Insurance

1.1.2 Why Data Driven?

Section 1.1.2 What is Analytics? – Make this more about using data to make decision. Remove “Analytics” terminology. Need to better define what we mean by “Loss Data.” Provide some foreshadowing for the new Chapter 2.

Add some language to tie into the rest of the text. Think about summarizing the respective application objectives. Those objectives are frequency modeling (Ch2), severity modeling (Ch3), aggregate loss computation (Ch5), ratemaking (Ch7), risk classification (Ch8), experience rating (Ch9, 12), insurance portfolio management (Ch10), loss reserving (Ch11), and dependence modeling (Ch14).

### Detailed Comments

1. (1.1.1) The connection to sports sounds odd, saying insurance is not as fun though more practical.
2. (1.1.3) Add non-life in “Most commercial and personal contracts are for a year”
3. (1.1.3) Provide some details or definitions of the terms in figure 1.1
4. One route is to offer a little less insurance here and introduce only the terms necessary to understand the basic details of the examples in the text.

## New Chapter 2 – Introduction to Data Analytics

This chapter steps outside of pure loss data because many types of data may influence the outcome of loss data and we want a structure that allows one to model these different data types. This introduction to data analytic concepts emphasizes aspects that are likely to be of use within insurance activities.

The following is a rough sketch of the new Chapter.

**2.1 Key Data Analytic Concepts**

Explains key concepts for modeling in the big data era, each with a short or a minimal explanation. Maybe adapt Table 1 from Frees and Gao, 2019…

Key concepts are:

* Data Driven – refer to new Section 1.1.2. (Structural models here?)
* Big Data – refer to old Section 13.1.2
* individualization,
* predictive (vs. generative. See Donoho’s 2017 seminal paper as well as Breiman’s 2001 one), “All models are wrong, but some are useful” – refer to old Section 13.2.5
* interpretability,
* non-parametric (vs. parametric) – refer to old Section 13.2.4
* EDA (vs. CDA) – refer to old Section 13.2.2
* robustness,
* computational statistics (simulation, sampling, subsampling, resampling), etc.

Some items to be included already appear in the current Ch. 13, but others are not.

**2.2 Single Variable**

2.2.1. Loss Data – Continuous (long tail), Discrete, Count, Mixed,…. Brief introduction to different data types. Refer to more extensive discussion in 2.4…

2.2.2. EDA

Introduces very basic techniques of numerical summary and graphical summary.

The introduced summary techniques may include all the items in the current 4.1.1.

Integrate with Chapter 1 Case.

2.2.3. Model Construction. Distribution of potential outcomes - draws from a population. Types of models – parametric versus nonparametric. What are the models used for? Describe estimation and prediction

2.2.4. Model Selection. Model complexity (e.g., number of parameters). One should split the data into training and testing portions. Discuss reducible versus irreducible errors. Also provide a brief introduction to cross validation and information criteria.

**2.3 Multiple variables**

When most analysts think of “data science,” or “machine learning,” they focus on multiple variables.

2.3.1 Types of Tasks

Explain types of tasks including supervised vs. unsupervised then regression vs. classification. Refer back to estimation versus prediction.

2.3.2 EDA (as a stage in the modelling process)

Perhaps also mention in passing the concepts of PCA and clustering as in the current 13.3.1 (although we will not do so in this book).

2.3.3 Model Construction

A brief introduction to types of modeling methods focusing on basic MLE-based methods and extensions such as LM, GLM.

It partly covers the current contents of 13.3.2, but perspectives may be quite different.

Perhaps mention in passing the concepts of regularized GLM, tree-based methods, neural networks, etc. (although we will not do so in this book). May include some explanation of regularization if not in detail.

2.3.4 Model selection

A brief introduction to cross validation and information criteria.

**2.4 Typical General Process**

Starts from similar contents to the old 13.2.1. We can do a bit better – not just give old processes. For example, including data-splitting step…

Emphasize “common methods” objectives and tie to the rest of the book. These objectives include the basic concepts of distributions (Ch2, 3), EDA/summary/visualization (a bit in Ch4), statistical inference including estimation (a bit in Ch4), model selection (a bit in Ch4, 13), simulation (Ch6), bootstrap (Ch6), Bayesian methods (a bit in Ch4, 9), and data preprocessing.

**2.5 Data**

A big section on data that draws from the old Sections 13.1 and 14.1. Much of this will not needed for the text but helps the reader appreciate the many potential applications of data analytics. Use this section to underscore the reach of potential applications of data analytics.

Include discussion of Data Preparation such as a shorter version of the current 13.1. Integrate with Chapter 1 Case.

*2.5 Example of modeling process*

Exploit the work done in Section 1.3 on the LGPIF. Not sure if a separate section is needed. Maybe just integrate throughout the chapter….