

Insurance Pricing With GLMs

Alan Chalk

2025-01-22

Table of contents

Preface	3
1 Introduction	4
2 Chapter 3: Data preparation	5
3 Summary	6
References	7

Preface

This is a Quarto book.

To learn more about Quarto books visit <https://quarto.org/docs/books>.

1 + 1

[1] 2

1 Introduction

This is a book created from markdown and executable code.

See Knuth (1984) for additional discussion of literate programming.

```
1 + 1
```

```
[1] 2
```

2 Chapter 3: Data preparation

Introduction

This chapter will discuss various aspects of data preparation that should be addressed before modelling starts. It is very tempting once a dataset is available to jump right in to the modelling stage - because that is interesting and data preparation is boring. A typical outcome of this hastiness is getting to the end of a few days of modelling and then realizing that you included a feature which should not be included. Or you treated a feature as numeric which should been treated as a character / factor. An example of the latter is the number of doors of a vehicle (in auto insurance). Treating this as a numeric feature will often lead to the assumption that the risk of accident decreases (or increases) linearly as the number of doors increases where this might not be true.

This chapter starts of with simple things - like what is a good naming convention for our features both from the point of view of understanding what they are, and from a computer programming perspective. Character and numeric features will be discussed. And we will also look at feature screening, missing values and general sense-checking.

3 Summary

In summary, this book has no content whatsoever.

1 + 1

[1] 2

References

Knuth, Donald E. 1984. “Literate Programming.” *Comput. J.* 27 (2): 97–111. <https://doi.org/10.1093/comjnl/27.2.97>.