ACC 471 - Final Report

Subtitle

Jared Musil & Jake McNair

2017-11-29

Contents

1 Prerequisites **5**
1.1 Foreword 5
1.2 Markdown Test 5

2 Introduction **7**

3 Problem Description **11**

4 Data **13**

5 Methods Used **17**

6 Results **19**
6.1 Example one 19
6.2 Example two 19

7 Reccomentations **21**

8 Future Analysis **23**

9 Conculsion **25**

Chapter 1

Prerequisites

1.1 Foreword

This report was written using the R package `Bookdown`. This was done as it allows for reproducable research of our data, methods, and results. Where appropriate, the code has been included inline with the results. All other methods are contained within the Appendix.

It is also available as a website reading on mobile devices, and and epub for reading offline.

1.2 Markdown Test

is a *sample* book written in **Markdown**. You can use anything that Pandoc's Markdown supports, e.g., a math equation $a^2 + b^2 = c^2$.

For now, you have to install the development versions of **bookdown** from Github:

```
devtools::install_github("rstudio/bookdown")
```

Remember each Rmd file contains one and only one chapter, and a chapter is defined by the first-level heading `#`.

To compile this example to PDF, you need to install XeLaTeX.

Chapter 2

Introduction

This

Throughout this report, the columns of our dataset will be refered to as factors, and the rows of our dataset will be refered to as reccords. This is to keep

---

You can label chapter and section titles using `{#label}` after them, e.g., we can reference Chapter 2. If you do not manually label them, there will be automatic labels anyway, e.g., Chapter **??**.

Figures and tables with captions will be placed in `figure` and `table` environments, respectively.

```
par(mar = c(4, 4, .1, .1))
plot(pressure, type = 'b', pch = 19)
```

Reference a figure by its code chunk label with the `fig:` prefix, e.g., see Figure 2.1. Similarly, you can reference tables generated from `knitr::kable()`, e.g., see Table 2.1.

```
knitr::kable(
  head(iris, 20), caption = 'Here is a nice table!',
  booktabs = TRUE
)
```

You can write citations, too. For example, we are using the **bookdown** package (Xie, 2017) in this sample book, which was built on top of R Markdown and **knitr** (Xie, 2015).

Table 2.1: Here is a nice table!

Sepal.Length	Sepal.Width	Petal.Length	Petal.Width	Species
5.1	3.5	1.4	0.2	setosa
4.9	3.0	1.4	0.2	setosa
4.7	3.2	1.3	0.2	setosa
4.6	3.1	1.5	0.2	setosa
5.0	3.6	1.4	0.2	setosa
5.4	3.9	1.7	0.4	setosa
4.6	3.4	1.4	0.3	setosa
5.0	3.4	1.5	0.2	setosa
4.4	2.9	1.4	0.2	setosa
4.9	3.1	1.5	0.1	setosa
5.4	3.7	1.5	0.2	setosa
4.8	3.4	1.6	0.2	setosa
4.8	3.0	1.4	0.1	setosa
4.3	3.0	1.1	0.1	setosa
5.8	4.0	1.2	0.2	setosa
5.7	4.4	1.5	0.4	setosa
5.4	3.9	1.3	0.4	setosa
5.1	3.5	1.4	0.3	setosa
5.7	3.8	1.7	0.3	setosa
5.1	3.8	1.5	0.3	setosa

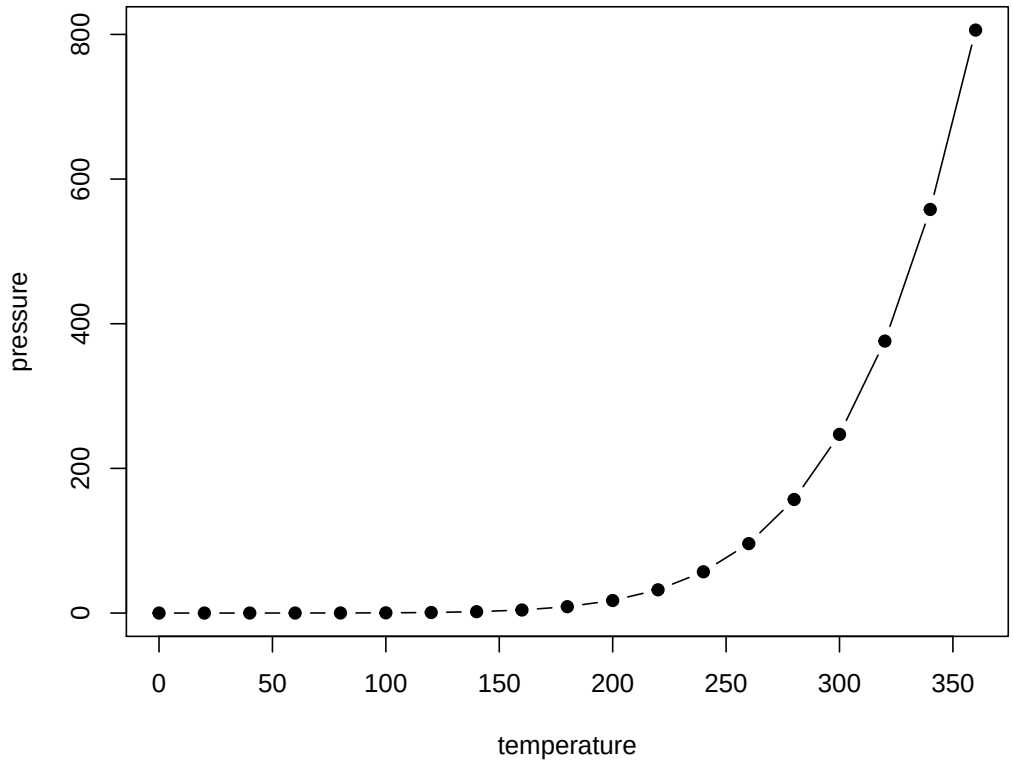

Figure 2.1: Here is a nice figure!

Chapter 3

Problem Description

The ability to utilize analytics to predict automobeile lossess is a area of active research and application throughout the insurance and fin-tech industries. All of the "big four" US domiciled auto insurrers being State Farm, Geico, Allstate, and Progressive are actively engaging in research to operationalize analytical models to increase operational efficency. [citation needed…]. This dataset is representitive of claims data common to all of these auto insurance providers, and the industry at large.

From a consumer standpoint, this has the potential to reduce average claim times, reduce premium costs, and improve claims decisions (total loss, not total loss).

Chapter 4

Data

Before doing any analysis, the feactors withing data set were first checked for missing or invalid data. Of the original 205 reccords, 41 were removed because they contained missing data for the `normalized-lossess` factor, which was coded as a `?`. This resulted in a dataset of 164 reccords of clean data. No other factors needed cleaning up, as the data was properly coded for each reccord.

```
library(readxl)
data_dict <- readxl::read_xlsx("automobile-losses-data-dictionary.xlsx")
knitr::kable(head(data_dict, 20), caption = 'Data Dictionary - Initial', booktabs = TRUE
```

Of these factors, 10 of the initial 26 were removed, resulting in the 16 factors that will be used in analysis. These factors are noted in green in `Keep` column of the above table.

The objective factor in the dataset is determined to be “.

Next, the data was partitioned into three groups named *training*, *test*, and *validation*. This was

```
data <- readxl::read_xlsx("automobile-losses.xlsx")
knitr::kable(head(data, 20), caption = 'Dataset', booktabs = TRUE)
```

```
data <- readxl::read_xlsx("automobile-losses.xlsx")
summary(data)
```

```
##        1                  2                   3
##  Min.   :-2.0000   Length:205         Length:205
##  1st Qu.: 0.0000   Class :character   Class :character
##  Median : 1.0000   Mode  :character   Mode  :character
##  Mean   : 0.8341
##  3rd Qu.: 2.0000
##  Max.   : 3.0000
##       4                   5                   6
##  Length:205         Length:205         Length:205
```

Table 4.1: Data Dictionary - Initial

#	Description	Values
1	symboling	-3, -2, -1, 0, 1, 2, 3
2	normalized-losses	continuous from [65 to 256]
3	make	alfa-romero, audi, bmw, chevrolet, dodge, honda, isuzu, jaguar, mazda, mercede
4	fuel-type	diesel, gas
5	aspiration	std, turbo
6	num-of-doors	four, two
7	body-style	hardtop, wagon, sedan, hatchback, convertible
8	drive-wheels	4wd, fwd, rwd.
9	engine-location	front, rear
10	wheel-base	continuous from [86.6 to 120.9]
11	length	continuous from [141.1 to 208.1]
12	width	continuous from [60.3 to 72.3]
13	height	continuous from [47.8 to 59.8]
14	curb-weight:	continuous from [1488 to 4066]
15	engine-type	dohc, dohcv, l, ohc, ohcf, ohcv, rotor
16	num-of-cylinders	eight, five, four, six, three, twelve, two
17	engine-size	continuous from [61 to 326]
18	fuel-system	1bbl, 2bbl, 4bbl, idi, mfi, mpfi, spdi, spfi
19	bore	continuous from [2.54 to 3.94]
20	stroke	continuous from [2.07 to 4.17]

Table 4.2: Dataset

1	2	3	4	5	6	7	8	9	10	11	12	13	14
3	?	alfa-romero	gas	std	two	convertible	rwd	front	88.6	168.8	64.1	48.8	2548
3	?	alfa-romero	gas	std	two	convertible	rwd	front	88.6	168.8	64.1	48.8	2548
1	?	alfa-romero	gas	std	two	hatchback	rwd	front	94.5	171.2	65.5	52.4	2823
2	164	audi	gas	std	four	sedan	fwd	front	99.8	176.6	66.2	54.3	2337
2	164	audi	gas	std	four	sedan	4wd	front	99.4	176.6	66.4	54.3	2824
2	?	audi	gas	std	two	sedan	fwd	front	99.8	177.3	66.3	53.1	2507
1	158	audi	gas	std	four	sedan	fwd	front	105.8	192.7	71.4	55.7	2844
1	?	audi	gas	std	four	wagon	fwd	front	105.8	192.7	71.4	55.7	2954
1	158	audi	gas	turbo	four	sedan	fwd	front	105.8	192.7	71.4	55.9	3086
0	?	audi	gas	turbo	two	hatchback	4wd	front	99.5	178.2	67.9	52.0	3053
2	192	bmw	gas	std	two	sedan	rwd	front	101.2	176.8	64.8	54.3	2395
0	192	bmw	gas	std	four	sedan	rwd	front	101.2	176.8	64.8	54.3	2395
0	188	bmw	gas	std	two	sedan	rwd	front	101.2	176.8	64.8	54.3	2710
0	188	bmw	gas	std	four	sedan	rwd	front	101.2	176.8	64.8	54.3	2765
1	?	bmw	gas	std	four	sedan	rwd	front	103.5	189.0	66.9	55.7	3055
0	?	bmw	gas	std	four	sedan	rwd	front	103.5	189.0	66.9	55.7	3230
0	?	bmw	gas	std	two	sedan	rwd	front	103.5	193.8	67.9	53.7	3380
0	?	bmw	gas	std	four	sedan	rwd	front	110.0	197.0	70.9	56.3	3505
2	121	chevrolet	gas	std	two	hatchback	fwd	front	88.4	141.1	60.3	53.2	1488
1	98	chevrolet	gas	std	two	hatchback	fwd	front	94.5	155.9	63.6	52.0	1874

```
##  Class :character   Class :character   Class :character
##  Mode  :character   Mode  :character   Mode  :character
##
##
##
##       7                  8                  9                  10
##  Length:205         Length:205         Length:205         Min.   : 86.60
##  Class :character   Class :character   Class :character   1st Qu.: 94.50
##  Mode  :character   Mode  :character   Mode  :character   Median : 97.00
##                                                           Mean   : 98.76
##                                                           3rd Qu.:102.40
##                                                           Max.   :120.90
##        11              12              13              14
##  Min.   :141.1   Min.   :60.30   Min.   :47.80   Min.   :1488
##  1st Qu.:166.3   1st Qu.:64.10   1st Qu.:52.00   1st Qu.:2145
##  Median :173.2   Median :65.50   Median :54.10   Median :2414
##  Mean   :174.0   Mean   :65.91   Mean   :53.72   Mean   :2556
##  3rd Qu.:183.1   3rd Qu.:66.90   3rd Qu.:55.50   3rd Qu.:2935
##  Max.   :208.1   Max.   :72.30   Max.   :59.80   Max.   :4066
##       15                 16                  17             18
##  Length:205         Length:205         Min.   : 61.0   Length:205
##  Class :character   Class :character   1st Qu.: 97.0   Class :character
##  Mode  :character   Mode  :character   Median :120.0   Mode  :character
##                                        Mean   :126.9
##                                        3rd Qu.:141.0
##                                        Max.   :326.0
##       19                 20                  21             22
##  Length:205         Length:205         Min.   : 7.00   Length:205
##  Class :character   Class :character   1st Qu.: 8.60   Class :character
##  Mode  :character   Mode  :character   Median : 9.00   Mode  :character
##                                        Mean   :10.14
##                                        3rd Qu.: 9.40
##                                        Max.   :23.00
##       23                  24              25             26
##  Length:205         Min.   :13.00   Min.   :16.00   Length:205
##  Class :character   1st Qu.:19.00   1st Qu.:25.00   Class :character
##  Mode  :character   Median :24.00   Median :30.00   Mode  :character
##                     Mean   :25.22   Mean   :30.75
##                     3rd Qu.:30.00   3rd Qu.:34.00
##                     Max.   :49.00   Max.   :54.00
```

Chapter 5

Methods Used

We utilized `X` methods in our analysis, while settlling on regression trees for our final reccomendations.

Chapter 6

Results

...

6.1 Example one

6.2 Example two

Chapter 7

Reccomentations

...

Chapter 8

Future Analysis

...

Chapter 9

Conculsion

...

Bibliography

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

Xie, Y. (2017). *bookdown: Authoring Books and Technical Documents with R Markdown*. R package version 0.4.