

An introduction to Statistical Modelling using GLMs

Lindesay Scott-Hayward

27 May 2020

Contents

Preface	5
Prerequisites	5
Learning outcomes	5
Recommended reading	5
Data files	6
Acknowledgements	6
1 Introduction	7
1.1 Figures and Tables	7
1.2 Equations:	9
2 Literature	13
3 Methods	15
3.1 Shiny app	15
4 Applications	17
4.1 Example one	17
4.2 Example two	17
5 Final Words	19
A Answers	21

Preface

An introductory -workshop to the field of statistical modelling in R. The focus will be on how to fit statistical models in R, rather than on the rigorous underlying mathematics. The target audience is anyone who wants to learn how to fit linear models in R. The progression will be linear models, generalised linear models and linear mixed effects models.

Prerequisites

- Programming basics in R

Learning outcomes

- Understand the key concepts and terminology used in statistical modelling
- Use R to fit linear, generalised linear and mixed effect models in R
- Recognise practical issues with fitting these models
- Checking model fit
- Perform model comparisons

Recommended reading

I highly recommend the following books: * Statistics: An Introduction using R * Linear models with R * Data Analysis Using Regression and Multilevel/Hierarchical Models * An Introduction to Statistical Learning * Mixed Effects Models and Extensions in Ecology with R * Extending the Linear Model with R

Data files

All data files can be downloaded as a ZIP file from [here](#).

Acknowledgements

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

The **bookdown** package can be installed from CRAN or Github:

```
install.packages("bookdown")  
# or the development version  
# 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 XeLaTeX. You are recommended to install TinyTeX (which includes XeLaTeX): <https://yihui.name/tinytex/>.

Chapter 1

Introduction

You can label chapter and section titles using `{#label}` after them, e.g., we can reference Chapter 1. If you do not manually label them, there will be automatic labels anyway, e.g., Chapter 3.

1.1 Figures and Tables

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 1.1. Similarly, you can reference tables generated from `knitr::kable()`, e.g., see Table 1.1.

You can also do a text reference for a figure which is useful if you want special characters;

```
par(mar = c(4, 4, .1, .1))
plot(pressure, type = 'b', pch = 19, col=2)
```

Here i am referencing 1.2.

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

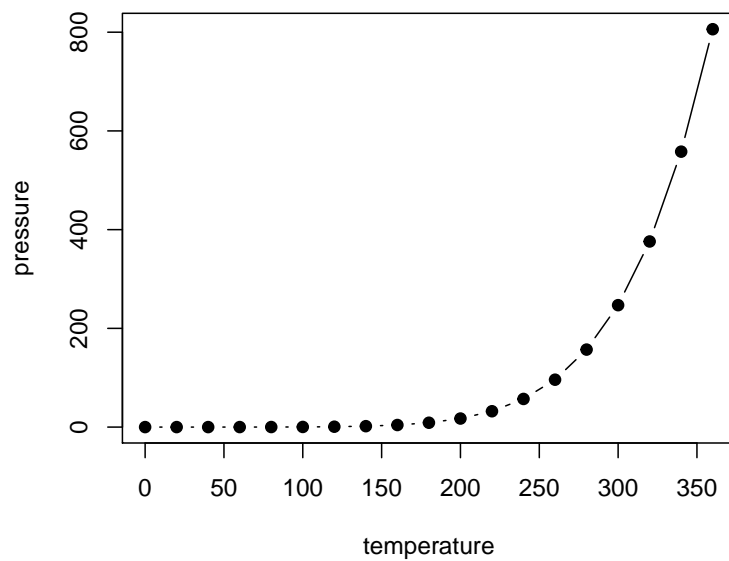


Figure 1.1: Here is a nice figure!

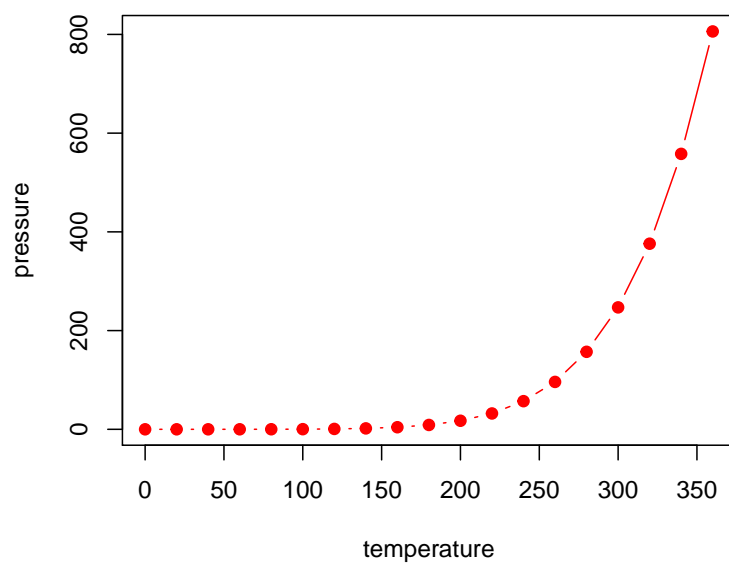
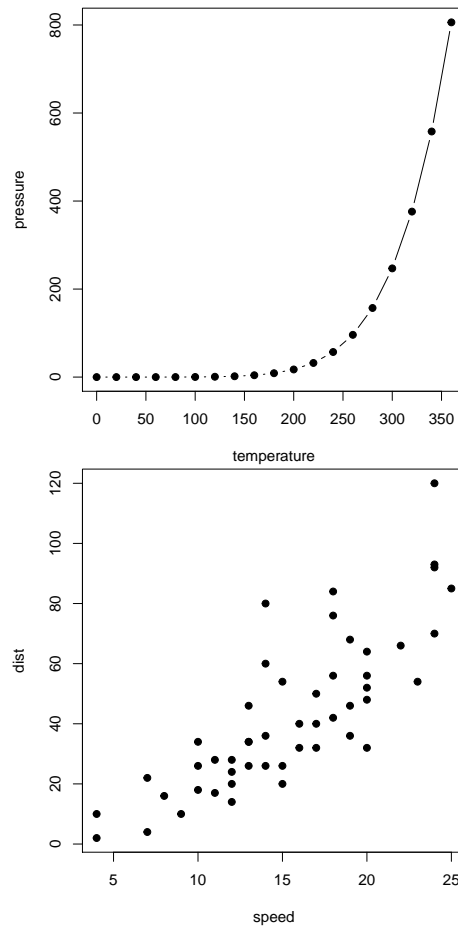


Figure 1.2: Define a text reference **here** with some math symbols like θ .



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

Reference this table using Table 1.1

You can write citations, too. For example, we are using the **bookdown** package (Xie, 2020) in this sample book, which was built on top of R Markdown and **knitr** (Xie, 2015).

1.2 Equations:

$$f(k) = \binom{n}{k} p^k (1-p)^{n-k} \quad (1.1)$$

Table 1.1: 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

You may refer to it using (1.1), e.g., see Equation (2.1).

Please make sure equations without labels are not numbered by either using the `equation*` environment or adding `\noindent` to your equations. The same rules apply to other math environments, such as `eqnarray`, `gather`, `align`, and so on (e.g., you can use the `align*` environment).

$$\begin{aligned} g(X_n) &= g(\theta) + g'(\tilde{\theta})(X_n - \theta) \\ \sqrt{n}[g(X_n) - g(\theta)] &= g'(\tilde{\theta}) \sqrt{n}[X_n - \theta] \end{aligned}$$

use `split` to have multiple lines with one equation reference

$$\begin{aligned} \text{Var}(\hat{\beta}) &= \text{Var}((X'X)^{-1}X'y) \\ &= (X'X)^{-1}X'\text{Var}(y)((X'X)^{-1}X')' \\ &= (X'X)^{-1}X'\text{Var}(y)X(X'X)^{-1} \\ &= (X'X)^{-1}X'\sigma^2IX(X'X)^{-1} \\ &= (X'X)^{-1}\sigma^2 \end{aligned}$$

there are also ‘example’ and ‘exercise’ environments (labelled with `exm` and `exr` respectively)

Example 1.1 (how to do an example). for this example, we show nothing

Example 1.2 (how to do a second example). for this example, we show something

these environments can be changed using the `css` file CSS class being the environment name, e.g., `<div class="lemma"></div>`

Chapter 2

Literature

Here is a review of existing methods.

Chapter 3

Methods

We describe our methods in this chapter.

We showed some examples in the introduction (Chapter 1)

You can also reference specific sections by giving them a label `{#sectionname}`

And you can add words to the index such as GLM

3.1 Shiny app

3.1.1 type 1 - calling from shinyapps.io

you can use a shiny app embedded in the notes Figure 3.1.

3.1.2 type 3 - creating inline

Figure 3.1: An example of a shiny app. you can see a live version at (put url here)

Chapter 4

Applications

```
library(ggplot2)
```

Some *significant* applications are demonstrated in this chapter.

4.1 Example one

4.2 Example two

Aside: using the model to predict outside of the range of the observed data is called **extrapolating**. Oftentimes, statistical models are developed *in order* to be used to extrapolate (e.g. climate modelling). This is however, dangerous, as we can only assess the assumptions of the model over the range of the observed data. When extrapolating we have to assume that these relationships hold beyond the range of the data, which may or may not be reasonable (hence why weather forecast over short-time periods are OK, but climate forecasts are much more uncertain). Hence, we should always view the model as an approximation of the data generating process. In this particular case, the interpretation of the parameters is not sensible when $x = 0$ (weight = 0 kg), but makes sense in the range that we are interested in exploring. If the case where $x = 0$ is important, then we would have to change the model to ensure that the predictions made sense at those values of x .

Task 1

Produce a scatterplot of `mpg` against `hp`. What does the relationship look like?

[Show Solution on P21](#)**Task 2**

This is another task with multi solution

[Show Solution on P23](#)

Chapter 5

Final Words

We have finished a nice book.

If you want to split the book by sections, so each html page is a section you can use `split_by='section'` in the YAML header.

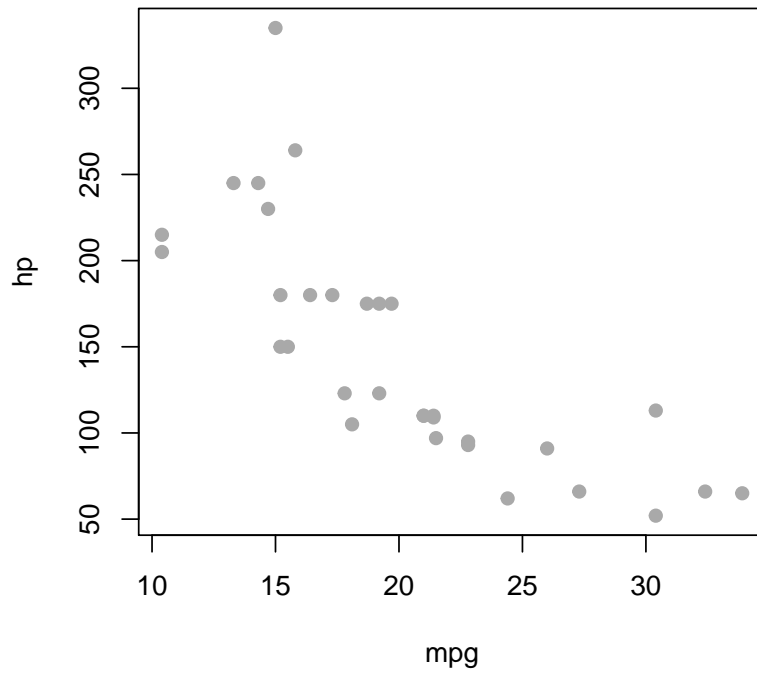
```
output:
  bookdown::gitbook:
    lib_dir: "book_assets"
  config:
    sharing: null
  split_by: section
```


Appendix A

Answers

Solution 1

```
plot(hp ~ mpg, data = mtcars,  
     pch=19, col='darkgrey')
```



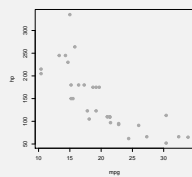
This is my solution

[Return to task on P18](#)

Solution 2

Base R

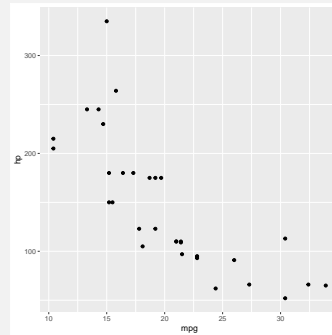
```
plot(hp ~ mpg, data = mtcars,
     pch=19, col='darkgrey')
```



The plot suggests that a linear relationship might exist between the two variables. So we can proceed by fitting a linear model in R.

tidyverse

```
ggplot(mtcars) +
  geom_point(aes(x = mpg, y = hp))
```



The plot suggests that a linear relationship might exist between the two variables. So we can proceed by fitting a linear model in R.

[Return to task on P18](#)

Bibliography

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

Xie, Y. (2020). *bookdown: Authoring Books and Technical Documents with R Markdown*. R package version 0.19.

Index

equations, 11

GLM, 15