

Computer lab notes for *Statistics for social sciences*

II: multivariate techniques

Bachelor's Degree in International Studies

Eduardo García Portugués

2016-09-01

Contents

Some useful information	5
Software required	5
Logistics	5
1 Simple linear regression	7
1.1 Motivation; model formulation; examples; applications; assumptions	7
1.2 Estimation of model parameters; least squares; inference for model parameters; forecasting. .	7
1.3 Assessing model fit; ANOVA; model validation; model diagnostics; handling nonlinear relationships.	7
2 Multiple linear regression	9
2.1 Motivation; model formulation; examples; applications; assumptions.	9
2.2 Estimation of model parameters; least squares; inference for model parameters; forecasting. .	9
2.3 Assessing model fit; ANOVA; model validation.	9
2.4 Multicollinearity; model diagnostics.	9
3 Binomial logistic regression	11
3.1 Motivation; model formulation; examples; applications; assumptions.	11
3.2 Parameter estimation; assessing model fit; significance testing; interpreting coefficients.	11
4 Factor analysis and principal component analysis	13
4.1 Motivation; formulations; examples; applications; assumptions.	13
4.2 Principal components analysis; choosing the number of factors; analysis and interpretation. .	13
4.3 Exploratory factor analysis; design, analysis and interpretation; rotation of factors.	13
5 Cluster analysis	15
5.1 Motivation; examples; applications; hierarchical agglomerative clustering; dendrogram.	15
5.2 Choosing the number of clusters; assessing fit; interpretation of clusters.	15

Some useful information

We will cover

Software required

- *R, a free software environment for statistical computing and graphics.*
- *Deducer, a free easy to use alternative to proprietary data analysis software that provides an intuitive graphical user interface (GUI) for R, encouraging non-technical users to learn and perform analyses without programming getting in their way.*

Logistics

- Lessons in Computer Lab
- Doubts regarding theory -> Jose
- Doubts regarding implementation -> Eduardo

Regarding

Chapter 1

Simple linear regression

- 1.1 Motivation; model formulation; examples; applications; assumptions
- 1.2 Estimation of model parameters; least squares; inference for model parameters; forecasting.
- 1.3 Assessing model fit; ANOVA; model validation; model diagnostics; handling nonlinear relationships.

You can label chapter and section titles using `{#label}` after them, e.g., we can reference Chapter `??`. 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 1.1. Similarly, you can reference tables generated from `knitr::kable()`, e.g., see Table 1.1.

```
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, 2016) in this sample book, which was built on top of R Markdown and **knitr** (Xie, 2015).

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

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.

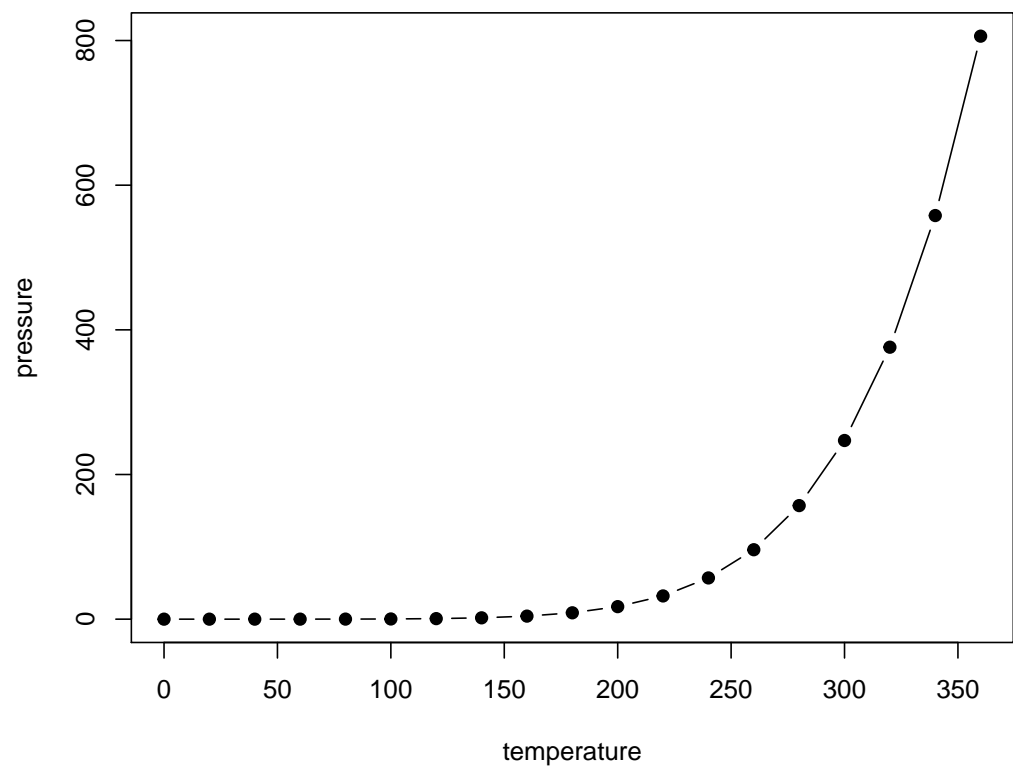


Figure 1.1: Here is a nice figure!

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

Chapter 2

Multiple linear regression

- 2.1 Motivation; model formulation; examples; applications; assumptions.
- 2.2 Estimation of model parameters; least squares; inference for model parameters; forecasting.
- 2.3 Assessing model fit; ANOVA; model validation.
- 2.4 Multicollinearity; model diagnostics.

You can label chapter and section titles using `{#label}` after them

Chapter 3

Binomial logistic regression

- 3.1 Motivation; model formulation; examples; applications; assumptions.
- 3.2 Parameter estimation; assessing model fit; significance testing; interpreting coefficients.

You can label chapter and section titles using `{#label}` after them

Chapter 4

Factor analysis and principal component analysis

- 4.1 Motivation; formulations; examples; applications; assumptions.
- 4.2 Principal components analysis; choosing the number of factors; analysis and interpretation.
- 4.3 Exploratory factor analysis; design, analysis and interpretation; rotation of factors.

You can label chapter and section titles using `{#label}` after them

Chapter 5

Cluster analysis

- 5.1 Motivation; examples; applications; hierarchical agglomerative clustering; dendrogram.
- 5.2 Choosing the number of clusters; assessing fit; interpretation of clusters.

You can label chapter and section titles using `{#label}` after them

Bibliography

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

Xie, Y. (2016). *bookdown: Authoring Books with R Markdown*. R package version 0.1.6.