Econometrics with R at SciencesPo

Florian Oswald and Jean-Marc Robin

2018-04-03

Contents

Chapter 1

Econometrics at ScPo - WORK IN PROGRESS

This is our book about Introductory Econometrics for 2nd year UGs at ScPo. We follow a practical approach with minimal reliance on maths and maximal reliance on practical learning.

On this first page we can explain a bit more about the course structure:

- Who is who
- how to communicate
- how many sessions etc

Figure 1.1:

Chapter 2

R

2.1   A very quick R intro

```
stop_dist_model = lm(dist ~ speed, data = cars)
plot(dist ~ speed, data = cars,
     xlab = "Speed (in Miles Per Hour)",
     ylab = "Stopping Distance (in Feet)",
     main = "Stopping Distance vs Speed",
     pch  = 20,
     cex  = 2,
     col  = "grey")
abline(stop_dist_model, lwd = 3, col = "red")
```

```
hist(residuals(stop_dist_model),freq=FALSE)
lines(density(residuals(stop_dist_model)),col='red',lwd=3)
```

2.2   Can do Math

$$R^2 = \frac{\text{SSReg}}{\text{SST}} = \frac{\sum_{i=1}^{n}(\hat{y}_i - \bar{y})^2}{\sum_{i=1}^{n}(y_i - \bar{y})^2}$$

$$= \frac{\text{SST} - \text{SSE}}{\text{SST}} = 1 - \frac{\text{SSE}}{\text{SST}}$$

$$= 1 - \frac{\sum_{i=1}^{n}(y_i - \hat{y}_i)^2}{\sum_{i=1}^{n}(y_i - \bar{y})^2} = 1 - \frac{\sum_{i=1}^{n} e_i^2}{\sum_{i=1}^{n}(y_i - \bar{y})^2}$$
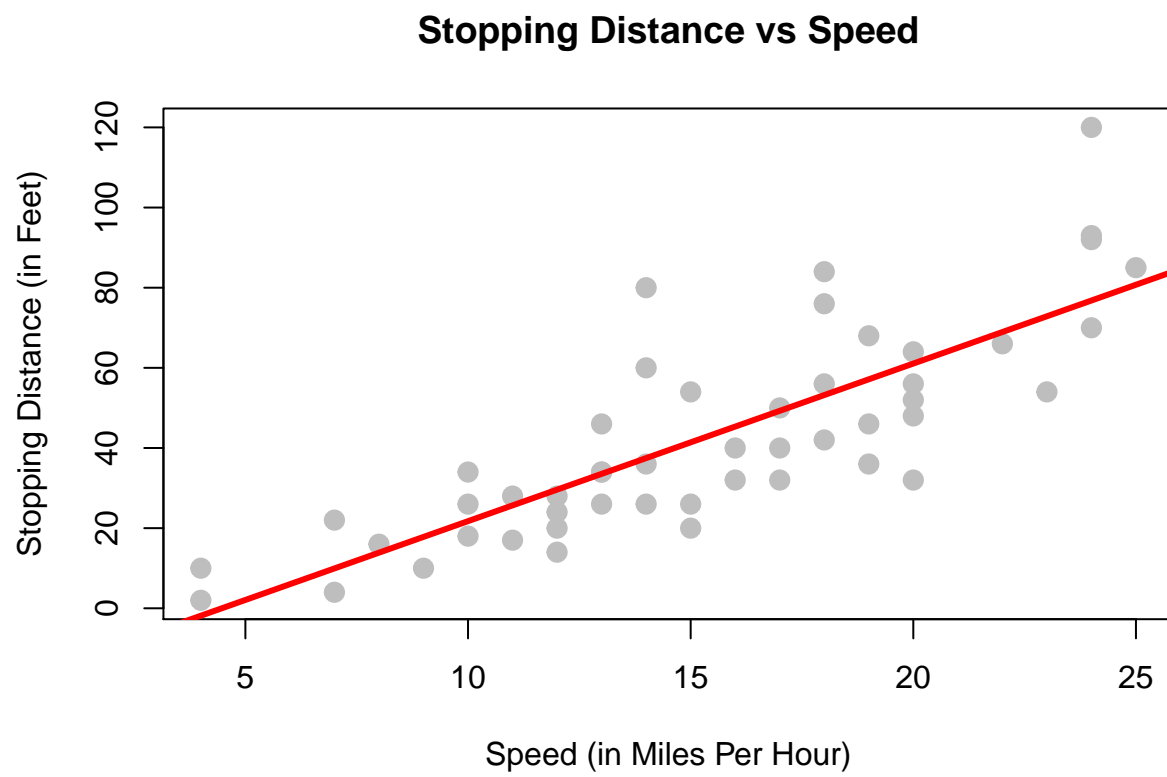
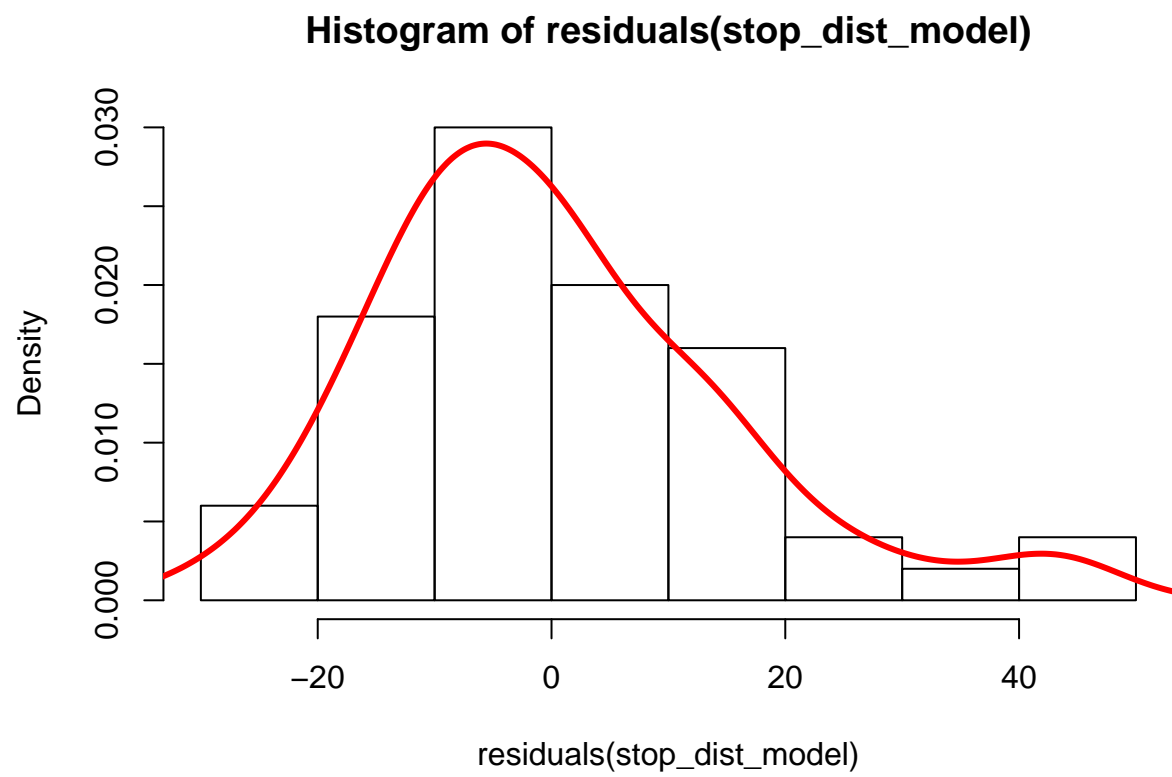But don't have to!

Figure 2.1: A simple plot in R

Figure 2.2: what you were looking for

Chapter 3

Summarizing Data

We describe our methods in this chapter.

Chapter 4

Linear Regression

1. linear regression - stop at R-squared
    1. different data: missing variable
    2. non-linear realtionship
2. scatter plot
    1. label observations
3. how do the data come to us? spreadsheet
4. approx link x and y by a line
5. OLS gives the best line for this
    1. $y_i = a + bx_i$. find a,b s.t. dist is minimal
    2. write out sum of least-squares and call it MSE: u_1 + u_2 + ... / N
6. plot fitted values - see imperfect approximation
7. R-squared: goodness of fit / measure of goodness
    1. 1 - sum of squared errors / SST
    2. how much of total variance is explained by the model?
8. regression on mean
9. How come there are residuals?
    1. measurement error?
    2. there is more to this than just x
    3. misspecification
10. There is statistical uncertainty about those estimates
11. plot a second data set with a less clear interpretation
    1. do you *really* think there is a linear relationship?
    2. SE tells us whethe rwe really think this is a positive slope
    3. poor R2 and large standard error
    4. How **confident** are you about this relationship? Is there enought data?
    5. SE is ameasure of precision depending on N

4.1 Try to find the Slope!

```
knitr::include_url("https://gallery.shinyapps.io/simple_regression/",height = "600px")
```

Chapter 5

Standard Errors

1. Standard Errors
    1. True data
    2. play with N and see how errors behave

Chapter 6

Multiple Regression

- Same as simple regression but more variables
- income ~ age is not linear

Chapter 7

Categorial Variables: Dummies and Interactions

1. if you only have educ categories, the estimate is like the conditional mean by educ
2. if you have too many categories you will have empty bins

Chapter 8

Quantile Regression

1. before you were modelling the mean. the average link
2. now what happens to **outliers**? how robust is the mean to that
3. what about the entire distribution of this?

Chapter 9

Panel Data

- scanner data on breakfast cereals, (Q_{it}, D_{it})
- why does D vary with Q
- pos relation ship
- don't observe the group identity!
- unobserved het alpha is correlated with Q
- within group estimator
- what if you don't have panel data?

Chapter 10

Instrumental Variables

1. wage = educ
2. educ = ability
3. ols estimator captures your ability
4. you predict x by z and you only keep the variation that is in the error

Chapter 11

Logit and Probit

Chapter 12

Principal Component Analysis

Chapter 13

Notes

this creates a library for the used R packages.

13.1 Book usage

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 4.

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

```r
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 13.1. Similarly, you can reference tables generated from `knitr::kable()`, e.g., see Table 13.1.

```r
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, 2018) in this sample book, which was built on top of R Markdown and **knitr** (Xie, 2015).
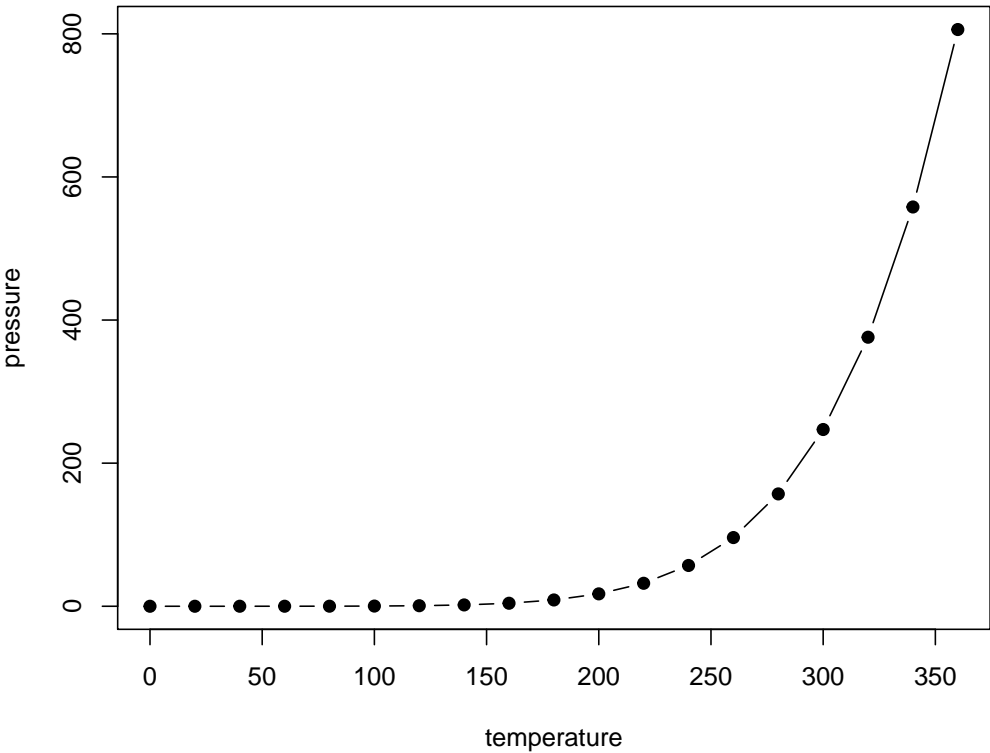
Figure 13.1: Here is a nice figure!

Table 13.1: 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

Bibliography

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

Xie, Y. (2018). *bookdown: Authoring Books and Technical Documents with R Markdown.* R package version 0.7.5.