

Development of E-learning materials using R & R markdown

Ziv Shkedy and Bernard Osangir
Hasselt University

Updated: 05/25

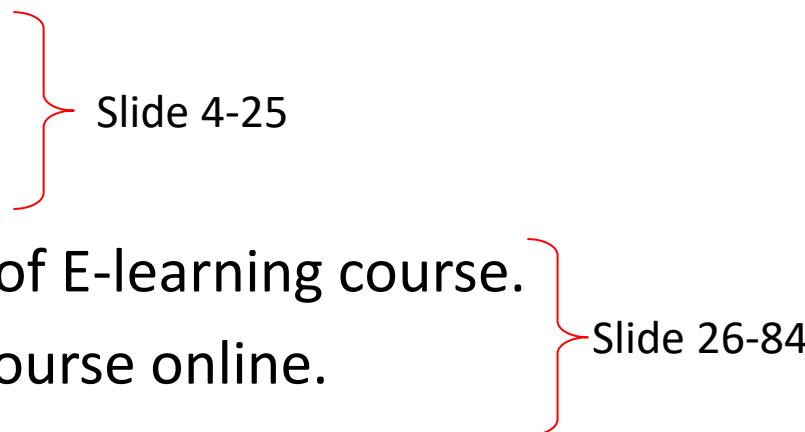


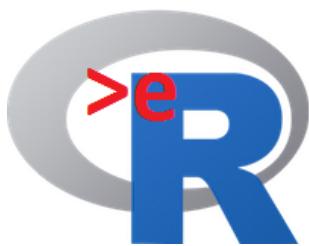
ER-BioStat

GitHub <https://github.com/eR-Biostat>

[@erbiostat](#)

Overview

- Starting point: first presentation:
 - Linear regression using R markdown.
 - Output development.
 - From R markdown to development of E-learning course.
 - From an output on my laptop to a course online.
 - Our approach: slides 85-97.
- 
- Slide 4-25
- Slide 26-84



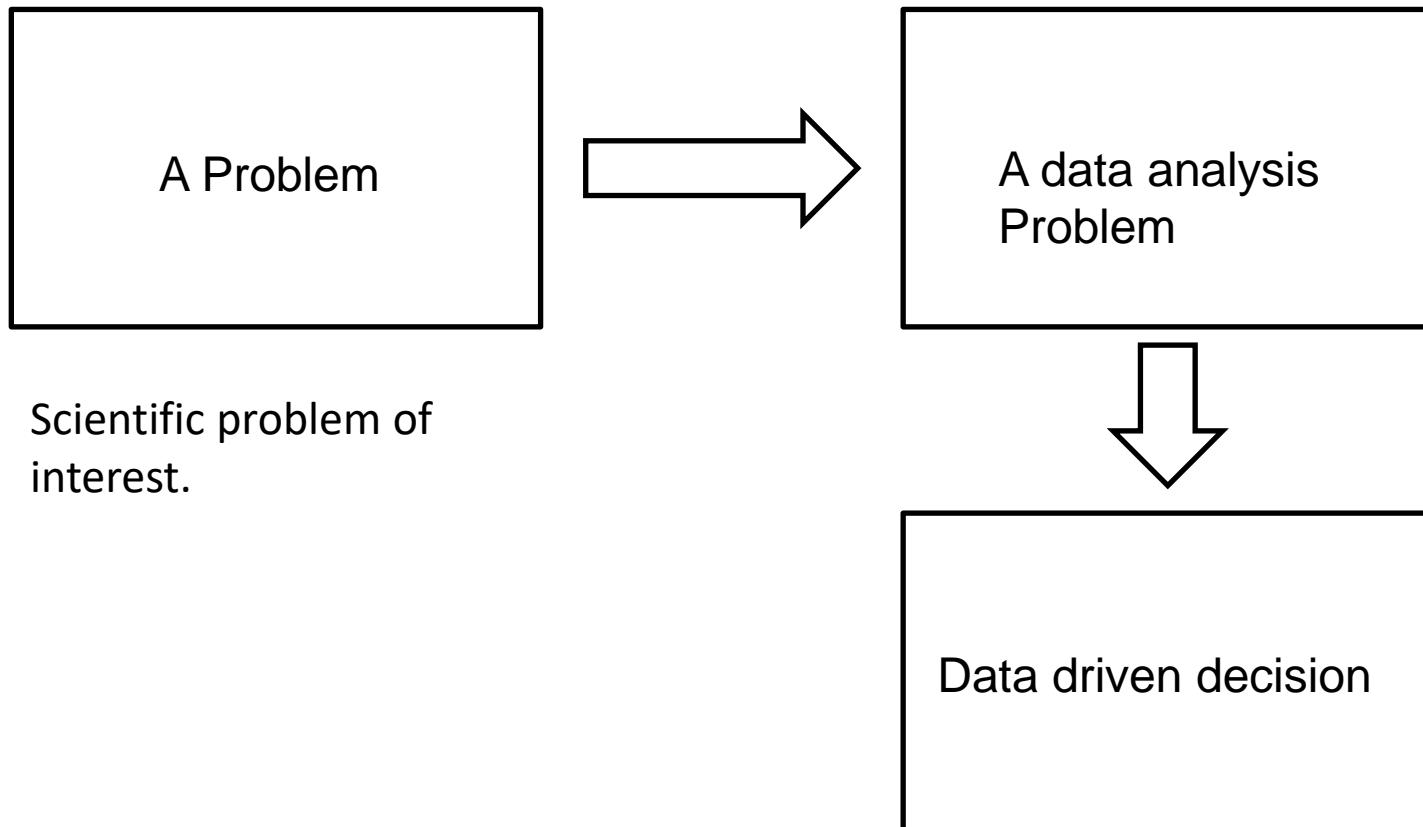
Short summary of the first session

Linear regression using R markdown

Slides 4-25: short summary of first lecture !!!

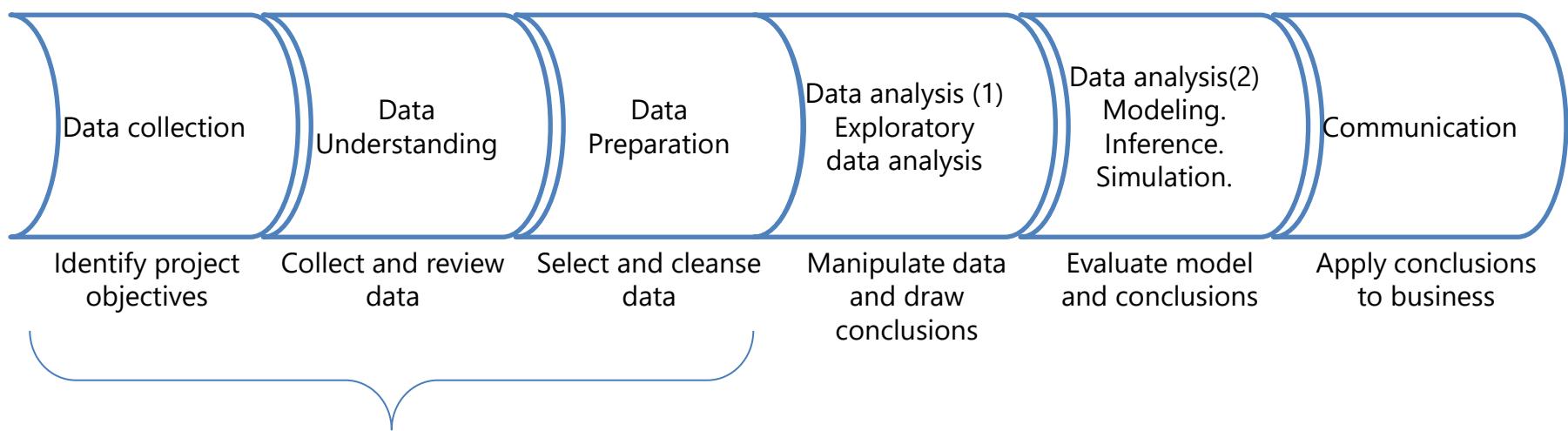
Steps in data analysis

- Data analysis approach in the course:



Steps in data analysis

- Steps related to data analysis:



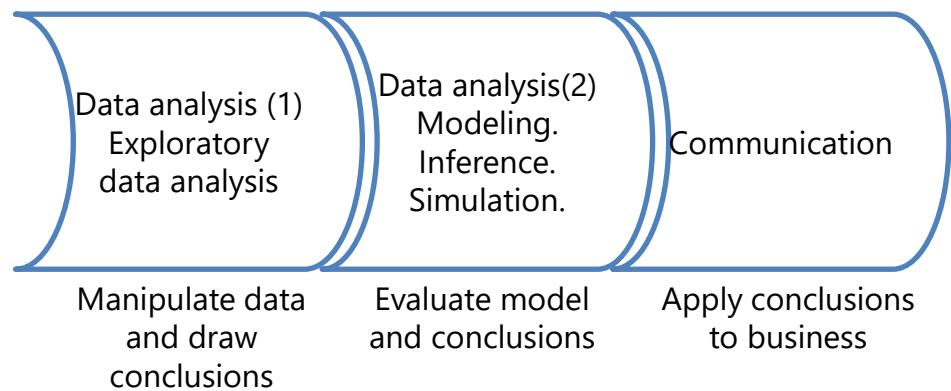
Not a part of our course

Steps in data analysis

Modeling the association between the fuel consumption and the car's weight.



Scientific problem of interest:
how to model the association ?

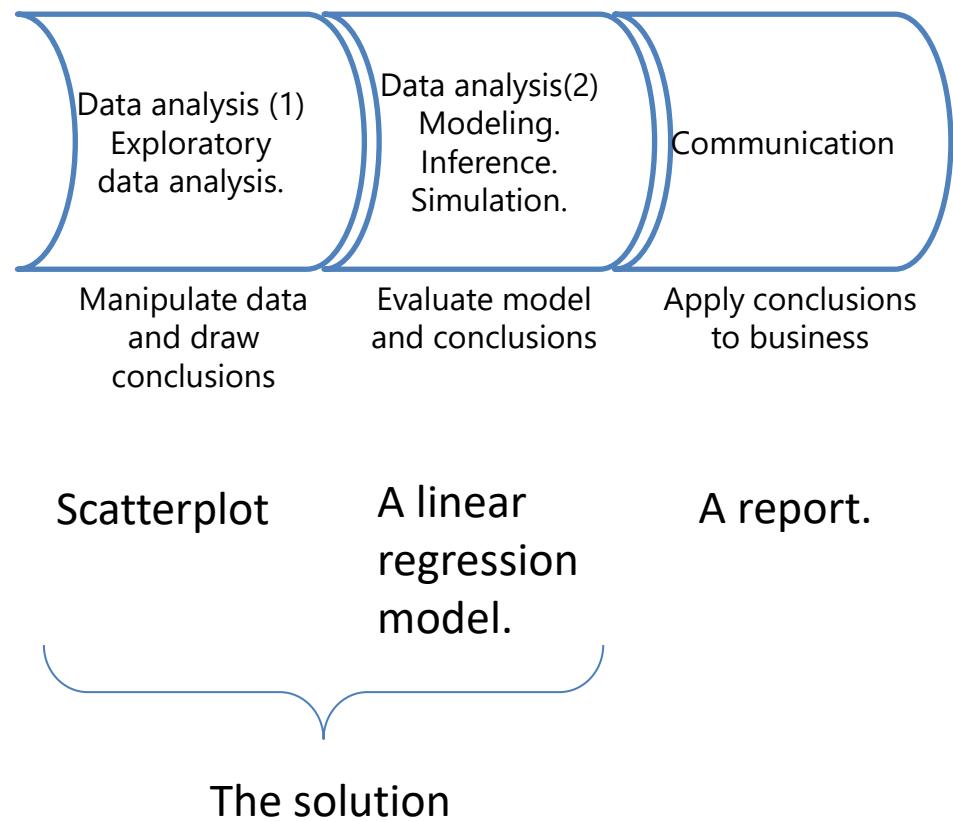


Steps in data analysis

Modeling the association between the fuel consumption and the car's weight.

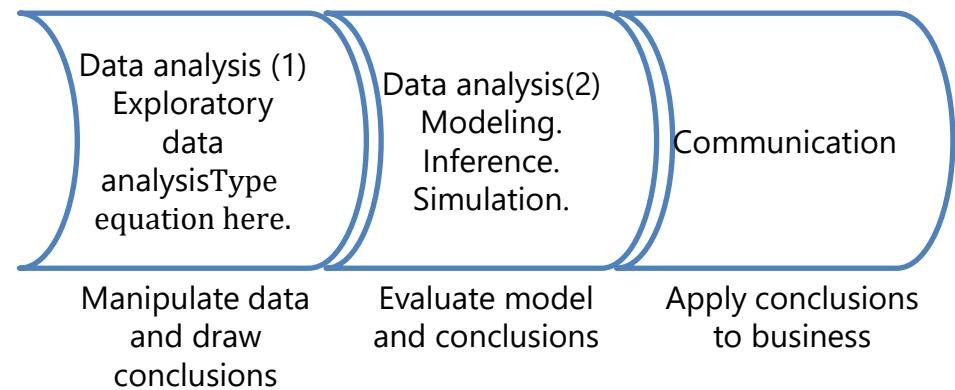
$$y_i = \beta_0 + \beta_1 x_i + \varepsilon_i$$

Methodology: simple linear regression.



Steps in data analysis

Modeling the association between the fuel consumption and the car's weight.



$$y_i = \beta_0 + \beta_1 x_i + \varepsilon_i$$

Methodology: simple linear regression.



We “translate” the methodology to software usage



Scatterplot.

Linear regression model.

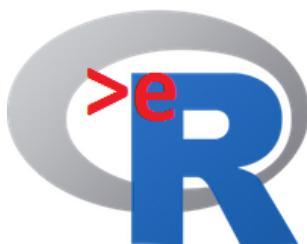
A report.

`ggplot2()`

`lm()`

R markdown to produce a HTML file.

We develop software to produce the solution and to communicate the solution



The mtcars data

Part 1

Analysis using basic R programming

R Program: er_prog1_SA_2024.R

The mtcars data in R

```
> dim(mtcars) [1] 32 11  
> names(mtcars)  
[1] "mpg"   "cyl"   "disp"  "hp"    "drat"  "wt"    "qsec"  
[8] "vs"    "am"    "gear"  
[11] "carb"
```

A diagram illustrating the mtcars dataset. It shows three main components: 1) A blue arrow pointing from the first line of R code (`dim(mtcars)`) to the text "The R object for the data: 32 observations and 11 variables.". 2) A blue arrow pointing from the second line of R code (`names(mtcars)`) to the list of variable names. 3) A blue arrow pointing from the word "carb" in the R code to the text "Variables names:".

The R object for the data: 32 observations and 11 variables.

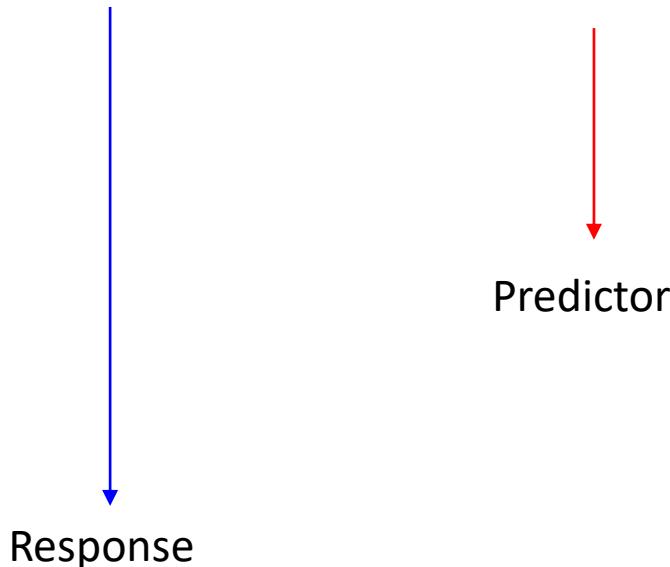
Variables names:
mpg: mile per gallon – the response.
wt: car's weight – the predictor.

The mtcars data in R

```
> head(mtcars)
```

		mpg	cyl	disp	hp	drat	wt	qsec	vs	am	gear	carb
Mazda RX4		21.0	6	160	110	3.90	2.620	16.46	0	1	4	4
Mazda RX4 Wag		21.0	6	160	110	3.90	2.875	17.02	0	1	4	4
Datsun 710		22.8	4	108	93	3.85	2.320	18.61	1	1	4	1
Hornet 4 Drive		21.4	6	258	110	3.08	3.215	19.44	1	0	3	1
Hornet Sportabout		18.7	8	360	175	3.15	3.440	17.02	0	0	3	2
Valiant		18.1	6	225	105	2.76	3.460	20.22	1	0	3	1

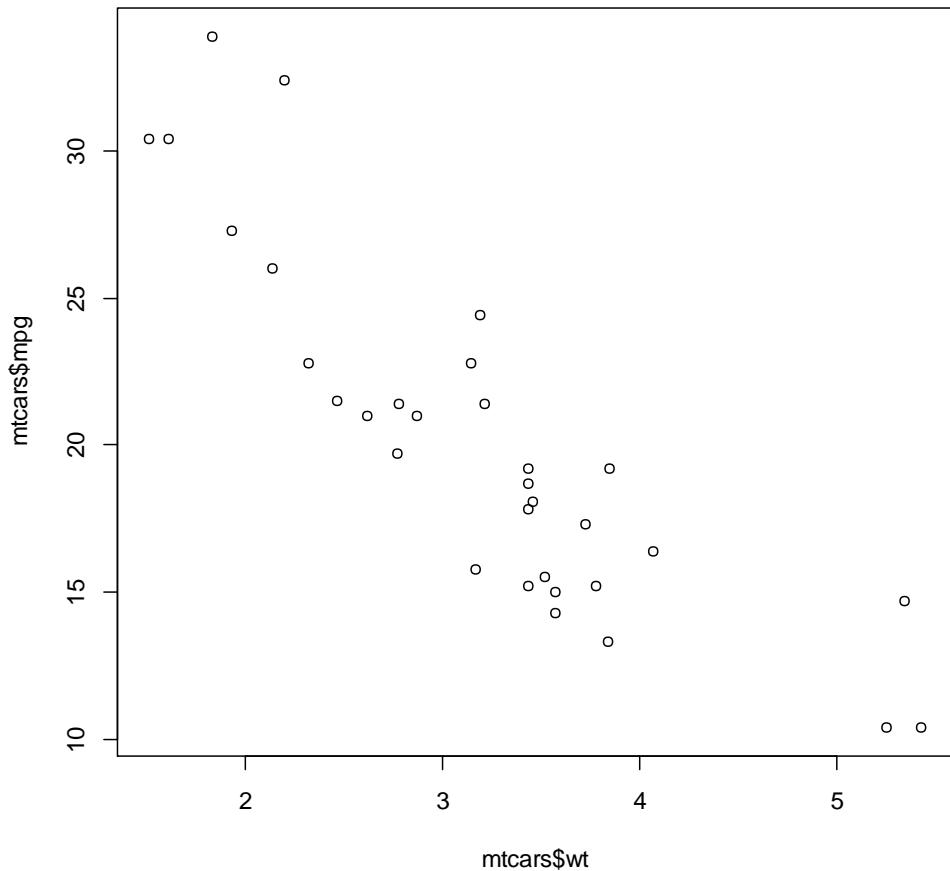
```
>
```



The mtcars data in R

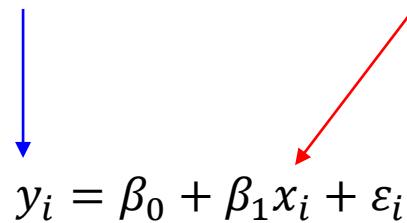
```
> plot(mtcars$wt, mtcars$mpg)  
> cor(mtcars$wt, mtcars$mpg)  
[1] -0.8676594
```

- R functions to produce the plot and the correlation.
- Basic functions in R



Simple linear regression in R

lm(response~predictor)

$$y_i = \beta_0 + \beta_1 x_i + \varepsilon_i$$


Simple linear regression in R: the mtcars data

```
>fit.lm=lm(mtcars$mpg~mtcars$wt)
```

The R object that contains the results of the fitted model.

```
> summary(fit.lm)
```

```
Call:  
lm(formula = mtcars$mpg ~ mtcars$wt)  
  
Residuals:  
    Min      1Q  Median      3Q     Max  
-4.5432 -2.3647 -0.1252  1.4096  6.8727
```

```
Coefficients:  
            Estimate Std. Error t value Pr(>|t|)  
(Intercept) 37.2851     1.8776 19.858 < 2e-16 ***  
mtcars$wt   -5.3445     0.5591 -9.559 1.29e-10 ***  
---  
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
Residual standard error: 3.046 on 30 degrees of freedom  
Multiple R-squared:  0.7528,    Adjusted R-squared:  0.7446  
F-statistic: 91.38 on 1 and 30 DF,  p-value: 1.294e-10
```

$$mpg_i = \beta_0 + \beta_1 wt_i + \varepsilon_i$$

- Output:
 - Parameter estimates etc.

R code for the analysis

```
dim(mtcars)
names(mtcars)
head(mtcars)
plot(mtcars$wt,mtcars$mpg)
cor(mtcars$wt,mtcars$mpg)
fit.lm=lm(mtcars$mpg~mtcars$wt)
summary(fit.lm)
```

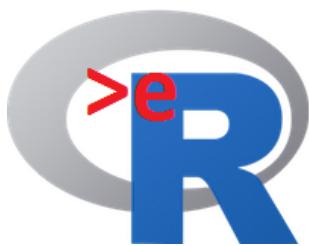


- Produce the plot.
- Calculate the correlation.
- Fit the model:

$$y_i = \beta_0 + \beta_1 x_i + \varepsilon_i$$

- Print the estimated model.

R Program: er_prog1_SA_2024.R



The mtcars data

Part 2

Analysis using basic Rmd programming

R Program: er_prog2_SA_2024.Rmd

The output

- We run the same analysis as before.
- Use R markdown.
- Produce the possible output formats:
 - HTML.
 - PDF.
 - Word doc.

The Rmd program

The screenshot shows the RStudio interface with an Rmd file open. The code in the Source tab is as follows:

```
1 ---  
2 title: 'The <tt>mtcars</tt> data - R workshop in Cape Town'  
3 output:  
4   word_document: default  
5   html_document: default  
6   pdf_document: default  
7 subtitle: Ziv Shkedy and Rudradev Sengupta.  
8 layout: page  
9 ---  
10   
11   
12   
13   
14 library(knitr)  
15 library(tidyverse)  
16 library(desolve)  
17 library(minpack.lm)  
18 library(ggpubr)  
19 library(readxl)  
20 library(gamls)  
21 library(data.table)  
22 library(grid)  
23 library(png)  
24 library(rnime)  
25 library(gridExtra)  
26 library(mvtnorm)  
27 library(e1071)  
28 library(lattice)  
29 library(ggplot2)  
30 library(cslabs)  
31 library(NHANES)  
32 library(plyr)  
33 library(dplyr)  
34 library(nasaweather)  
35 library(ggplot2)  
36 library(gganimate)  
37 library(av)  
38   
22:14 | [green] Chunk1:setup $
```

Annotations with red curly braces highlight specific sections of the code:

- A brace on the left side groups the first few lines (title, output, subtitle) under the heading "Document setup."
- A brace on the right side groups the majority of the code (library imports) under the heading "Many R packages, not all needed."

The Environment tab shows "Environment is empty". The bottom status bar indicates R 4.3.2, ENG, 15:30, 8/02/2024.

The Rmd program

The screenshot shows an RStudio interface with the following components:

- File Menu:** File, Edit, Code, View, Plots, Session, Build, Debug, Profile, Tools, Help.
- Toolbar:** Includes icons for New, Open, Save, Knit, Run, and Stop.
- Project Bar:** Shows multiple open files: OVA.Rmd, er_prog3_SA_2024.Rmd, The pharma challenge_2022_prog1.Rmd, er_prog2_SA_2024.Rmd, er_prog1_SA_2024.R, and er_prog4_SA_2024.Rmd.
- Source Editor:** Displays R code in a syntax-highlighted editor. Red arrows point from the text "Dimension of the data, variables names and first 6 lines of the data" to the lines `dim(mtcars)` and `head(mtcars)`.
- Plot & correlation:** A placeholder text area for plots and correlations.
- The regression model.:** A placeholder text area for the regression model.
- Environment Tab:** Shows "Environment is empty".
- Text Overlay:** Large red text "R code" with three dashed curly braces above it, positioned over the R code editor area.
- Console:** Displays the R startup message and a command prompt (>).
- Taskbar:** Shows the system tray with icons for battery, signal, and date/time (15:31, 8/02/2024, ENG).

The Rmd program

The screenshot shows the RStudio interface with an Rmd file open. The code editor on the left contains R code with specific sections highlighted by red arrows and labels:

- A horizontal arrow points from the word "Section" to the line `# Baseline analysis` at line 51.
- A horizontal arrow points from the word "Subsection" to the line `## First 6 lines` at line 60.
- A vertical red arrow points downwards from the "Section" and "Subsection" labels towards the "Analysis starts here" text.

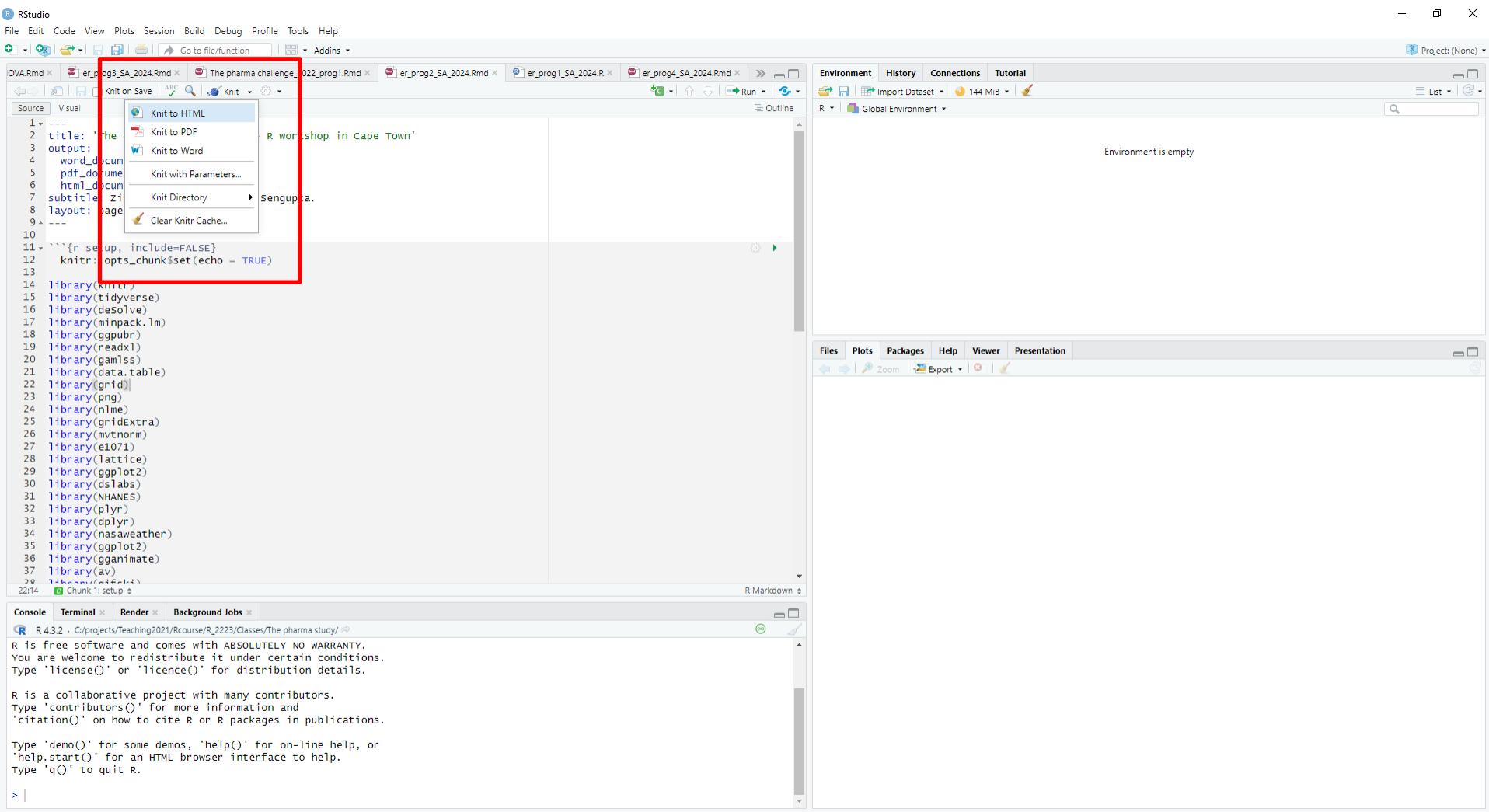
The code editor also highlights several lines of code in blue, indicating they are part of a code chunk:

- Line 51: `# Baseline analysis`
- Line 53: `## The <tt>mtcars</tt> data in R`
- Line 55: ````{r}```
- Line 56: `dim(mtcars)`
- Line 57: `names(mtcars)`
- Line 59: `````{r}```
- Line 60: `## First 6 lines`
- Line 62: ````{r}```
- Line 63: `head(mtcars)`
- Line 65: `````{r}```
- Line 66: `## Scatterplot`
- Line 68: `````{r}```
- Line 69: `plot(mtcars\$wt, mtcars\$mpg)`
- Line 70: `cor(mtcars\$wt, mtcars\$mpg)`
- Line 72: `````{r}```
- Line 74: `## Linear regression in R using the <tt>lm()</tt> function`
- Line 75: ````{r}```
- Line 76: `fit.lm=lm(mtcars\$mpg~mtcars\$wt)`
- Line 78: `summary(fit.lm)`
- Line 79: `````{r}```
- Line 80: `
- Line 81: `
- Line 82: `

The RStudio environment pane on the right shows the Global Environment is empty.

The R console at the bottom displays the standard R startup message and help information.

Choose the output



The HTML output

C:/projects/Teaching2021/Rcourse/R_2223/Classes/The pharma study/er_prog2_SA_2024.html

er_prog2_SA_2024.html | Open in Browser | Find

– □ X
Publish

The mtcars data - R workshop in Cape Town

Ziv Shkedy and Rudradev Sengupta.

Section → Baseline analysis

Subsection

The mtcars data in R

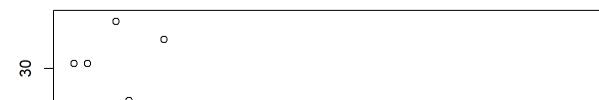
```
dim(mtcars)  
  
## [1] 32 11  
  
names(mtcars)  
  
## [1] "mpg"   "cyl"   "disp"  "hp"    "drat"  "wt"    "qsec" "vs"    "am"    "gear"  
## [11] "carb"
```

First 6 lines

```
head(mtcars)  
  
##          mpg cyl disp hp drat wt qsec vs am gear carb  
## Mazda RX4 21.0   6 160 110 3.90 2.620 16.46  0  1  4   4  
## Mazda RX4 Wag 21.0   6 160 110 3.90 2.875 17.02  0  1  4   4  
## Datsun 710 22.8   4 108 93 3.85 2.320 18.61  1  1  4   1  
## Hornet 4 Drive 21.4   6 258 110 3.08 3.215 19.44  1  0  3   1  
## Hornet Sportabout 18.7   8 360 175 3.15 3.440 17.02  0  0  3   2  
## Valiant    18.1   6 225 105 2.76 3.460 20.22  1  0  3   1
```

Scatterplot

```
plot(mtcars$wt,mtcars$mpg)
```



Type here to search



15:31
ENG
8/02/2024

The PDF output

er_prog2_SA_2024.pdf - Adobe Acrobat Reader (32-bit)

File Edit View Sign Window Help

Home Tools er_prog2_SA_2024... x

Sign In

Search tools

Export PDF

Edit PDF

Create PDF

Comment

Combine Files

Organize Pages

Request E-signatures

Fill & Sign

More Tools

The mtcars data - R workshop in Cape Town

Ziv Shkedy and Rudradev Sengupta.

Baseline analysis

The mtcars data in R

```
dim(mtcars)
## [1] 32 11
names(mtcars)
##  [1] "mpg"   "cyl"   "disp"  "hp"    "drat"  "wt"    "qsec" "vs"    "am"    "gear"
## [11] "carb"
```

First 6 lines

```
head(mtcars)
##          mpg cyl disp hp drat wt qsec vs am gear carb
## Mazda RX4     21.0   6 160 110 3.90 2.620 16.46  0  1    4    4
## Mazda RX4 Wag 21.0   6 160 110 3.90 2.875 17.02  0  1    4    4
## Datsun 710    22.8   4 108  93 3.85 2.320 18.61  1  1    4    1
## Hornet 4 Drive 21.4   6 258 110 3.08 3.215 19.44  1  0    3    1
## Hornet Sportabout 18.7   8 360 175 3.15 3.440 17.02  0  0    3    2
## Valiant       18.1   6 225 105 2.76 3.460 20.22  1  0    3    1
```

Scatterplot

```
plot(mtcars$wt,mtcars$mpg)
```

Type here to search

15:32 8/02/2024

The Word doc output

er_prog2_SA_2024.docx - Compatibility Mode - Saved to this PC

File Home Insert Design Layout References Mailings Review View Help

Cut Copy Format Painter

Font Paragraph Styles

Find Replace Select

The mtcars data -- R workshop in Cape Town

Ziv Shkedy and Rudradev Sengupta

Baseline analysis

The mtcars data in R

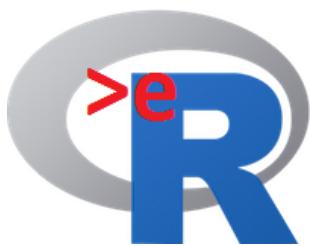
```
dim(mtcars)
## [1] 32 11
names(mtcars)
## [1] "mpg" "cyl" "disp" "hp" "drat" "wt" "qsec" "vs" "am" "gear"
## [11] "carb"
```

er_prog2_SA_2024.docx: 1,277 characters (an approximate value).

Type here to search

Focus

15:32 ENG 8/02/2024



Interuniversity Institute for Biostatistics
and statistical Bioinformatics

The mtcars data

Part 3

Advance HTML output

R Program: er_prog3_SA_2024.Rmd

What do we cover in this part ?

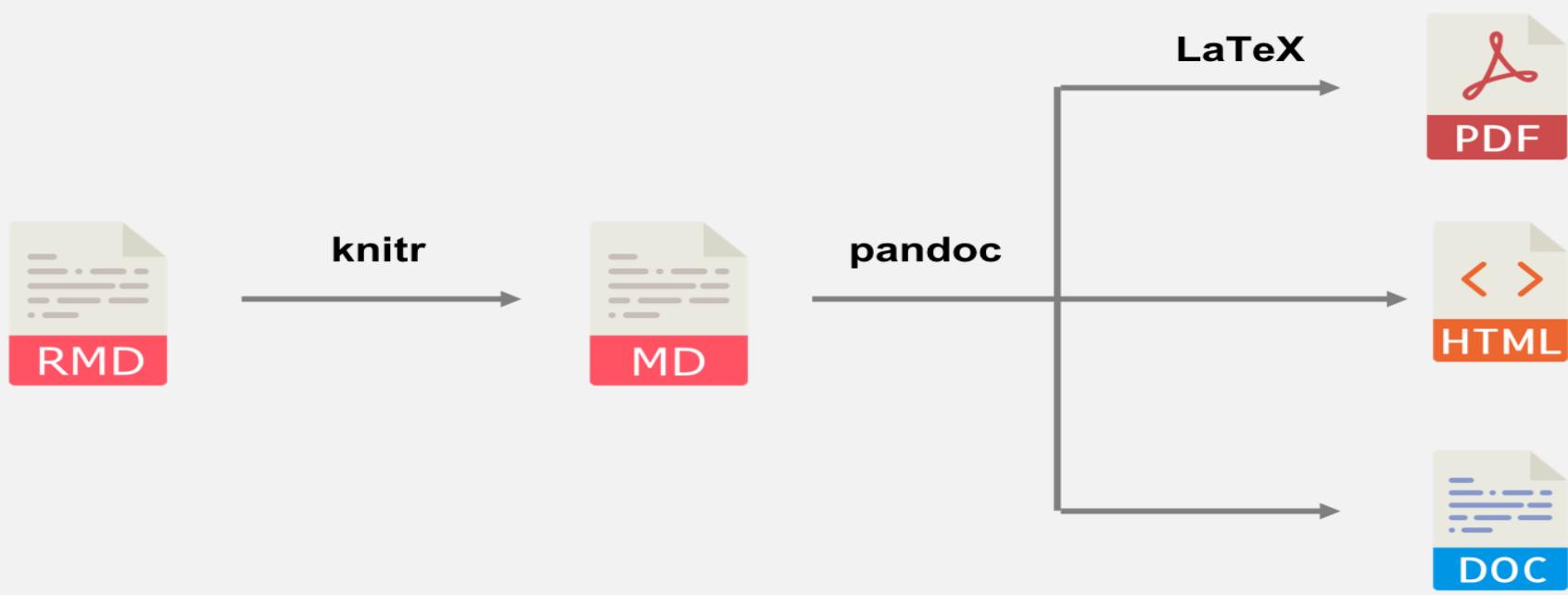
- We use the same example: the `mtcars` data and simple linear regression.
- How to produce a HTML file for the analysis we conducted before ?
- How to use the HTML file as a content for a website ?

Reproducible Research

- Aim: create an output in a different (highest) quality.
- Can be used to communicate the analysis' results with other people in the organization.
- Not all potential readers are interested on “how to do the analysis”.
- We DO NOT aim to develop a report for the analysis but to provide a document from which the results can be seen and discuss by different people in the organization.

The Rmd file

- Analyses → high quality report.
- Rmarkdown – Different dynamic and statistic formats (**html**, pdf, word, books, dashboard, e.t.c).



The HTML output

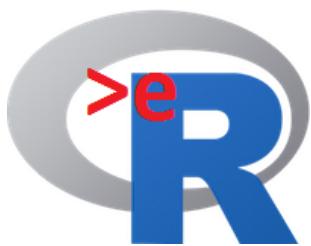
The screenshot shows an R Markdown document with the following structure:

- Header:** C:/projects/Teaching2021/Rcourse/R_2223/Classes/The pharma study/er_prog3_SA_2024.html
- Toolbar:** Open in Browser | Find | Publish
- Table of content (Sidebar):**
 - 1. The data
 - The mtcars dataset
 - Miles/(US) gallon vs. the car's Weight
 - 2. Simple linear regression using R
 - 3. Data and estimated model
 - 4. Model diagnostic
- Title:** Simple linear regression using R
- Analysis output:**
 - 1. The data
 - The mtcars dataset
 - ## [1] 32 11
 - Scatterplot
 - Miles/(US) gallon vs. the car's Weight
- Scatterplot:** A scatterplot showing the relationship between Miles/(US) gallon (Y-axis, ranging from 25 to 35) and the car's Weight (X-axis). The plot shows a negative correlation.

Annotations on the right side of the slide highlight specific parts of the document:

- A red bracket labeled "Title" points to the main title "Simple linear regression using R".
- A red bracket labeled "Analysis output" points to the section containing the scatterplot and the data table.
- A red arrow points upwards from the "Table of content" sidebar towards the main content area.

- An interactive HTML output.
- Presents the same analysis as before.



Interuniversity Institute for Biostatistics
and statistical Bioinformatics

Part 3.1: How to set up the HTML file ?

The Rmd file

- We use Rmd file to
 - Conduct the analysis.
 - Set up the document.
- We use html file to
 - Present & communicate the result.

Set up the document

RStudio

File Edit Code View Plots Session Build Debug Profile Tools Help

Go to file/function Addins

unit_05_inference_num_1WANOVA.Rmd er_prog3_SA_2024.Rmd The pharma challenge_2022_prog1.Rmd er_prog2_SA_2024.Rmd er_prog1_SA_2024.R

Source Visual Outline

```
1 ---  
2 output:  
3   bookdown::html_document2:  
4     toc: TRUE  
5     toc_float: TRUE  
6     toc_depth: 2  
7     number_sections: no  
8     css: ./lib/stylesArial.css  
9     code_folding: hide  
10  
11 params:  
12   department: ">ER-Biostat"  
13   topic: <font size = "10" > *Simple linear regression using R **</font>  
14   author: "Ziv Shkedy et al"  
15   date: "15-12-2023"  
16   endCode: FALSE  
17   RmdLocation: ""  
18 ---  
19  
20  
21 <p>  
22     
23 </p>  
24  
25  
26  
27 ```{r delaycodeprinting, message=FALSE, warning=FALSE, echo = FALSE}  
28 # You can delete this chunk if you do not want delaycodeprinting and adjust the YAML header accordingly  
29 library(knitr)  
30 # The **delaycodeprinting** chunk below allows all R code to be printed at the end of the report (endCode = TRUE)  
31 # or prints the RmdLocation from the YAML header as a code reference (endCode != TRUE)  
32 # see code chunk named 'codprint'  
33 delay_code_labels <- NULL  
34 knit_hooks$set(delay = function(before, options, envir) {  
35   if (before) {  
36     delay_code_labels <- append(delay_code_labels, options$label)  
37     return(NULL) ## otherwise knitr will print delay_code_labels every time  
38   })  
39 }  
1701 The <ttr:m>/ttr> R Function
```

Console Terminal Render Background Jobs

R 4.3.2 - C:/projects/Teaching2021/Rcourse/R_2223/Classes/The pharma study/

R is free software and comes with ABSOLUTELY NO WARRANTY.
You are welcome to redistribute it under certain conditions.
Type 'license()' or 'licence()' for distribution details.

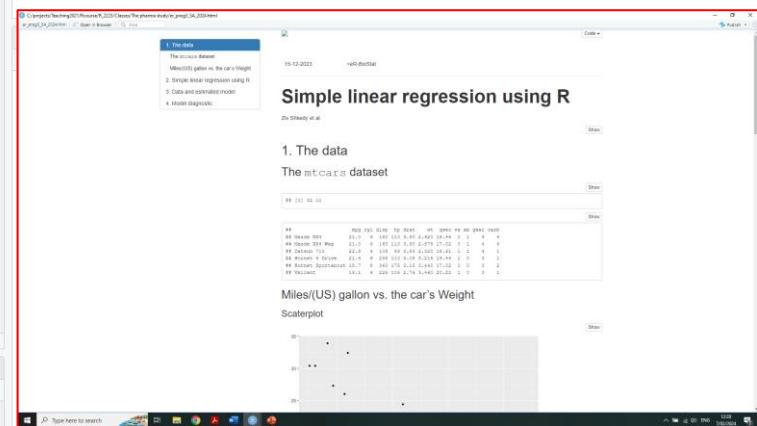
R is a collaborative project with many contributors.
Type 'contributors()' for more information and
'citation()' on how to cite R or R packages in publications.

Type 'demo()' for some demos, 'help()' for on-line help, or
'help.start()' for an HTML browser interface to help.
Type 'q()' to quit R.

> |

Set up the H document_2

Set up the HTML document:
[document 2](#)



Set up the document

RStudio

File Edit Code View Plots Session Build Debug Profile Tools Help

unit_05_inference_num_TWANOVA.Rmd er_prog3_SA_2024.Rmd The pharma challenge_2022_prog1.Rmd er_prog2_SA_2024.Rmd er_prog1_SA_2024.R

Source Visual

```
1 ---  
2 output:  
3   bookdown::html_document2:  
4     toc: TRUE  
5     toc_float: TRUE  
6     toc_depth: 2  
7     number_sections: no  
8     css: ./lib/stylesArial.css  
9     code_folding: hide  
10  
11 params:  
12   department: ">er-BioStat"  
13   topic: <font size = "10" > *simple linear regression using R **</font>  
14   author: "Ziv Shkedy et al"  
15   date: "15-12-2023"  
16   endCode: FALSE  
17   RmdLocation: ""  
18 ---  
19  
20  
21 <p>  
22     
23 </p>  
24  
25  
26  
27   
28 # You can delete this chunk if you do not want delaycodeprinting and adjust the YAML header accordingly  
29 library(knitr)  
30 # The **delaycodeprinting** chunk below allows all R code to be printed at the end of the report  
31 # or prints the RMDlocation from the YAML header as a code reference (endCode != TRUE)  
32 # see code chunk named 'codeprint'  
33 delay_code_labels <- NULL  
34 knit_hooks$set(delay = function(before, options, envir) {  
35   if (before) {  
36     delay_code_labels <- append(delay_code_labels, options$label)  
37     return(NULL) ## otherwise knitr will print delay_code_labels every time  
38   }  
39 })  
40 The <tt>lm</tt> R Function
```

Environment History Connections Tutorial

Import Dataset 157 MB R Global Environment

Environment is empty

depth=2 implies that in the TOC:

Section

Subsection

Subsubsection

Section

Subsection

- toc=true: add table of content.
- toc_float: float TOC to the left
- toc_depth: depth of header in toc

Console Terminal Render Background Jobs

R 4.3.2 . C:/projects/Teaching2021/Rcourse/R_2223/Classes/The pharma study/

R is free software and comes with ABSOLUTELY NO WARRANTY.
You are welcome to redistribute it under certain conditions.
Type 'license()' or 'licence()' for distribution details.

R is a collaborative project with many contributors.
Type 'contributors()' for more information and
'citation()' on how to cite R or R packages in publications.

Type 'demo()' for some demos, 'help()' for on-line help, or
'help.start()' for an HTML browser interface to help.
Type 'q()' to quit R.

Type here to search

12:30 7/02/2024 ENG

The HTML file

C:/projects/Teaching2021/Rcourse/R_2223/Classes/The pharma study/er_prog3_SA_2024.html
er_prog3_SA_2024.html Open in Browser Find Publish

1. The data

The mtcars dataset
Miles/(US) gallon vs. the car's Weight

2. Simple linear regression using R

3. Data and estimated model

4. Model diagnostic

15-12-2023 >eR-BioStat

Simple linear regression using R

Ziv Shkedy et al

1. The data

The mtcars dataset

```
## [1] 32 11
```

```
##          mpg cyl disp  hp drat    wt  qsec vs am gear carb
## Mazda RX4   21.0   6 160 110 3.90 2.620 16.46  0  1    4    4
## Mazda RX4 Wag 21.0   6 160 110 3.90 2.875 17.02  0  1    4    4
## Datsun 710  22.8   4 108  93 3.85 2.320 18.61  1  1    4    1
## Hornet 4 Drive 21.4   6 258 110 3.08 3.215 19.44  1  0    3    1
## Hornet Sportabout 18.7   8 360 175 3.15 3.440 17.02  0  0    3    2
## Valiant    18.1   6 225 105 2.76 3.460 20.22  1  0    3    1
```

Miles/(US) gallon vs. the car's Weight

Scaterplot

35
30
25

12:28 7/02/2024

Titles, authors and dates

RStudio

File Edit Code View Plots Session Build Debug Profile Tools Help

unit_05_inference_num_TWANOVA.Rmd er_prog3_SA_2024.Rmd The pharma challenge_2022_prog1.Rmd er_prog2_SA_2024.Rmd er_prog1_SA_2024.R

Knit on Save Knit Run Addins

Source Visual Outline

```
1 ---  
2 output:  
3   bookdown::html_document2:  
4     toc: TRUE  
5     toc_float: TRUE  
6     toc_depth: 2  
7     number_sections: no  
8     css: ./lib/stylesArial.css  
9     code_folding: hide  
10  
11 params:  
12   department: ">ER-BioStat"  
13   topic: <font size = "10" > *simple linear regression using R **</font>  
14   author: "Ziv Shkedy et al"  
15   date: "15-12-2023"  
16   endCode: FALSE  
17   RmdLocation: ""  
18 ---  
19  
20  
21 <p>  
22     
23 </p>  
24  
25  
26  
27   
28 # You can delete this chunk if you do not want delaycodeprinting and adjust the YAML header accordingly  
29 library(knitr)  
30 # The **delaycodeprinting** chunk below allows all R code to be printed at the end of the report (endCode = TRUE)  
31 # or prints the RMDlocation from the YAML header as a code reference (endCode != TRUE)  
32 # see code chunk named 'codeprint'  
33 delay_code_labels <- NULL  
34 knit_hooks$set(delay = function(before, options, envir) {  
35   if (before) {  
36     delay_code_labels <- append(delay_code_labels, options$label)  
37     return(NULL)  ## otherwise knitr will print delay_code_labels every time  
38 }  
170:1 The <tt>lm</tt> R function
```

Title

Environment History Connections Tutorial

Import Dataset 157 MB

Global Environment

Environment is empty

Files Plots Packages Help Viewer Presentation

Console Terminal Render Background Jobs

R 4.3.2 . C:/projects/Teaching2021/Rcourse/R_2223/Classes/The pharma study/

R is free software and comes with ABSOLUTELY NO WARRANTY.
You are welcome to redistribute it under certain conditions.
Type 'license()' or 'licence()' for distribution details.

R is a collaborative project with many contributors.
Type 'contributors()' for more information and
'citation()' on how to cite R or R packages in publications.

Type 'demo()' for some demos, 'help()' for on-line help, or
'help.start()' for an HTML browser interface to help.
Type 'q()' to quit R.

Type here to search

12:30 7/02/2024 ENG

Titles, authors and dates

C:/projects/Teaching2021/Rcourse/R_2223/Classes/The pharma study/er_prog3_SA_2024.html
er_prog3_SA_2024.html Open in Browser Find Publish

1. The data

The mtcars dataset
Miles/(US) gallon vs. the car's Weight

2. Simple linear regression using R

3. Data and estimated model

4. Model diagnostic

15-12-2023 >eR-BioStat

Simple linear regression using R

Ziv Shkedy et al

1. The data

The mtcars dataset

```
## [1] 32 11
```

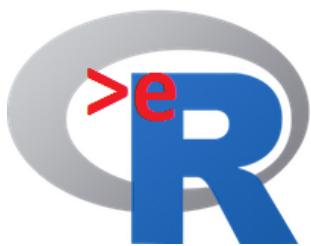
```
##          mpg cyl disp  hp drat    wt  qsec vs am gear carb
## Mazda RX4     21.0   6 160 110 3.90 2.620 16.46  0  1    4    4
## Mazda RX4 Wag 21.0   6 160 110 3.90 2.875 17.02  0  1    4    4
## Datsun 710    22.8   4 108  93 3.85 2.320 18.61  1  1    4    1
## Hornet 4 Drive 21.4   6 258 110 3.08 3.215 19.44  1  0    3    1
## Hornet Sportabout 18.7   8 360 175 3.15 3.440 17.02  0  0    3    2
## Valiant       18.1   6 225 105 2.76 3.460 20.22  1  0    3    1
```

Miles/(US) gallon vs. the car's Weight

Scaterplot

12:28 7/02/2024

Title



Interuniversity Institute for Biostatistics
and statistical Bioinformatics

Part 3.2: The HTML file and the Rmd program in details.

Section, subsection, subsubsection

RStudio

File Edit Code View Plots Session Build Debug Profile Tools Help

Unit_05_inference_num_1ANOVA.Rmd er_prog3_SA_2024.Rmd The pharma challenge_2022_prog1.Rmd er_prog2_SA_2024.Rmd er_prog1_SA_2024.Rmd Knit Run Addins

Source Visual

```
131 library(toreach)
132 library("DAAG")
133 library(DT)
134 ...
135 ...
136 ...
137 # 1. The data
138 ## The <tt>mtcars</tt> dataset
139 ...
140 ```{r, echo=TRUE, message=FALSE, warning=FALSE}
141 dim(mtcars)
142 head(mtcars)
143 ...
144 ...
145 ...
146 ...
147 ## Miles/(us) gallon vs. the car's weight
148 ...
149 ## scaterplot
150 ...
151 ```{r, echo=TRUE, message=FALSE, warning=FALSE}
152 #plot(mtcars$wt, mtcars$mpg, ylab = "Miles/(us) gallon")
153 plot(wt, mpg, data = mtcars)
154 ...
155 ...
156 ...
157 ...
158 ...
159 ...
160 ...
161 ...
162 ...
163 # 2. simple linear regression using R
164 ...
165 ## The <tt>lm()</tt> R function
166 ...
167 For the <tt>mtcars</tt> dataset, we consider the model
168 ...
169 ...
170 ...
171 The <tt>lm()</tt> R function :
```

Environment History Connections Tutorial

Import Dataset 157 MB Global Environment

Environment is empty

Only these appear in the TOC in the upper left corner

Files Plots Packages Help Viewer Presentation

R 4.3.2 . C:/projects/Teaching2021/Rcourse/R_2223/Classes/The pharma study

R is free software and comes with ABSOLUTELY NO WARRANTY.
You are welcome to redistribute it under certain conditions.
Type 'license()' or 'licence()' for distribution details.

R is a collaborative project with many contributors.
Type 'contributors()' for more information and
'citation()' on how to cite R or R packages in publications.

Type 'demo()' for some demos, 'help()' for on-line help, or
'help.start()' for an HTML browser interface to help.
Type 'q()' to quit R.

> |

Type here to search

12:38 7/02/2024 38

Section, subsection, subsubsection

Depth=2 → Only sections and subsections

Ziv Shkedy et al

1. The data section
The mtcars dataset subsection

```
## [1] 32 11
```

```
##          mpg cyl disp  hp drat    wt  qsec vs am gear carb
## Mazda RX4   21.0   6 160 110 3.90 2.620 16.46  0  1    4    4
## Mazda RX4 Wag 21.0   6 160 110 3.90 2.875 17.02  0  1    4    4
## Datsun 710  22.8   4 108  93 3.85 2.320 18.61  1  1    4    1
## Hornet 4 Drive 21.4   6 258 110 3.08 3.215 19.44  1  0    3    1
## Hornet Sportabout 18.7   8 360 175 3.15 3.440 17.02  0  0    3    2
## Valiant    18.1   6 225 105 2.76 3.460 20.22  1  0    3    1
```

Since depth=2, the subsubsection will not appear in the TOC → Miles/(US) gallon vs. the car's Weight subsection
Scatterplot subsubsection

The scatterplot shows a negative correlation between mpg and weight. The x-axis is labeled 'wt' (Weight) and ranges from approximately 2.6 to 4.4. The y-axis is labeled 'mpg' (Miles/(US) gallon) and ranges from approximately 15 to 35. Data points are scattered across the plot area, with a general downward trend from left to right.

C:/projects/Teaching2021/Rcourse/R_2223/Classes/The pharma study/er_prog3_SA_2024.html
er_prog3_SA_2024.html | Open in Browser | Find | Publish | 12:37 7/02/2024

Analysis code

- The same as before.
- In addition to the code, we can add free text in the Rmd file.

The code for the analysis

C:/projects/Teaching2021/Rcourse/R_2223/Classes/The pharma study/er_prog3_SA_2024.html
er_prog3_SA_2024.html | Open in Browser | Find

1. The data

The mtcars dataset
Miles/(US) gallon vs. the car's Weight
2. Simple linear regression using R
3. Data and estimated model
4. Model diagnostic

15-12-2023 >eR-BioStat

Code ▾
Show All Code
Hide All Code

Simple linear regression using R

Ziv Shkedy et al

1. The data

The mtcars dataset

```
## [1] 32 11
```

	mpg	cyl	disp	hp	drat	wt	qsec	vs	am	gear	carb
## Mazda RX4	21.0	6	160	110	3.90	2.620	16.46	0	1	4	4
## Mazda RX4 Wag	21.0	6	160	110	3.90	2.875	17.02	0	1	4	4
## Datsun 710	22.8	4	108	93	3.85	2.320	18.61	1	1	4	1
## Hornet 4 Drive	21.4	6	258	110	3.08	3.215	19.44	1	0	3	1
## Hornet Sportabout	18.7	8	360	175	3.15	3.440	17.02	0	0	3	2
## Valiant	18.1	6	225	105	2.76	3.460	20.22	1	0	3	1

Miles/(US) gallon vs. the car's Weight

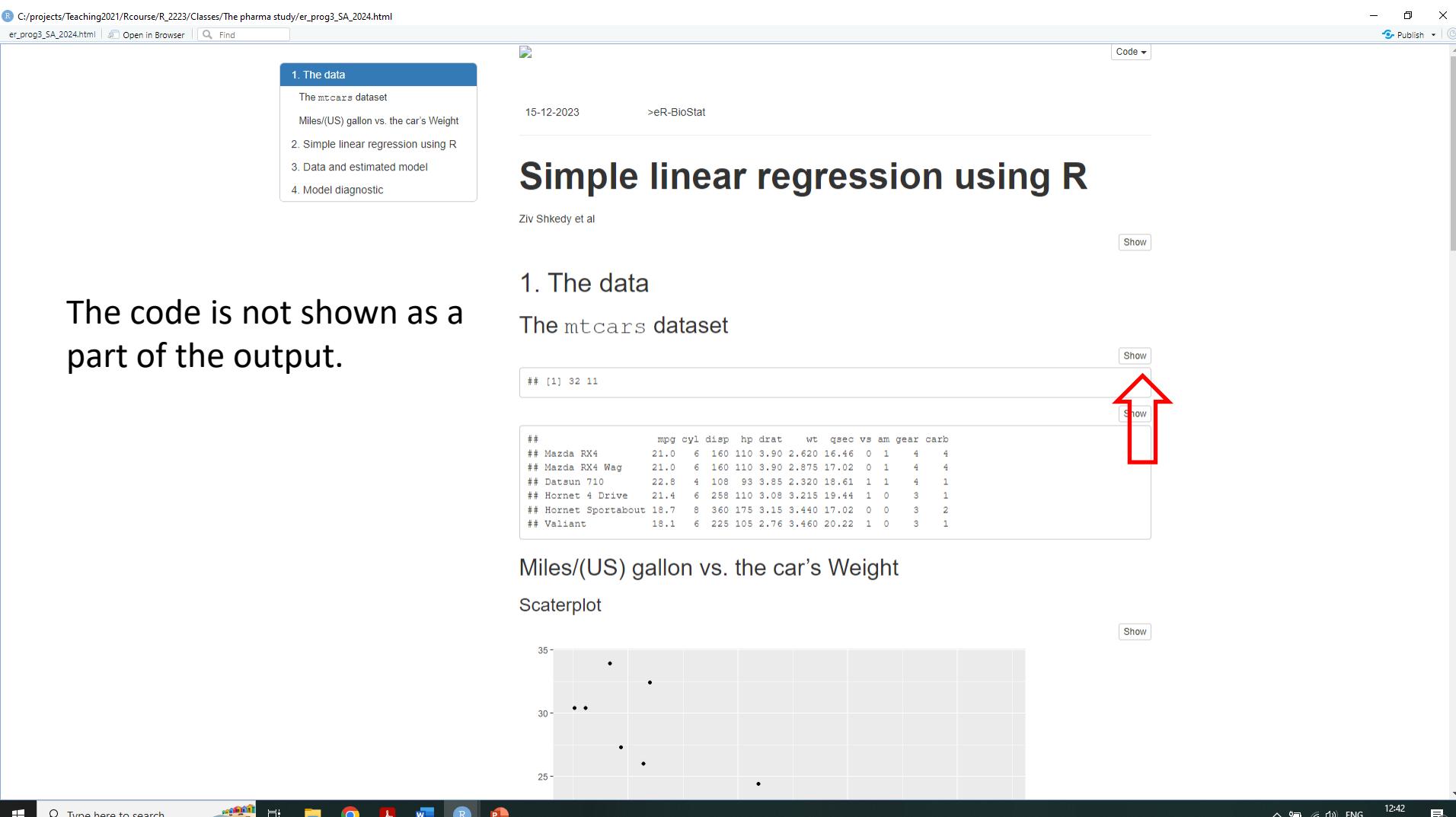
Scaterplot

Windows Taskbar: Type here to search, File, Start, Google Chrome, Microsoft Word, R, Microsoft Powerpoint, 12:40, ENG, 7/02/2024, Page 41

We can choose if we want to show the code or to hide the code.

Reading the external file

The code is not shown as a part of the output.



C:/projects/Teaching2021/Rcourse/R_2223/Classes/The pharma study/er_prog3_SA_2024.html
er_prog3_SA_2024.html | Open in Browser | Find | Code ▾

1. The data

The mtcars dataset

Miles/(US) gallon vs. the car's Weight

2. Simple linear regression using R

3. Data and estimated model

4. Model diagnostic

15-12-2023 >eR-BioStat

Simple linear regression using R

Ziv Shkedy et al

Show

1. The data

The mtcars dataset

Show

```
## [1] 32 11
```

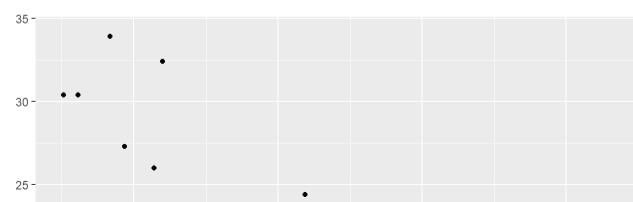
Show

```
##          mpg cyl disp  hp drat    wt  qsec vs am gear carb
## Mazda RX4   21.0   6 160 110 3.90 2.620 16.46  0  1   4   4
## Mazda RX4 Wag 21.0   6 160 110 3.90 2.875 17.02  0  1   4   4
## Datsun 710  22.8   4 108  93 3.85 2.320 18.61  1  1   4   1
## Hornet 4 Drive 21.4   6 258 110 3.08 3.215 19.44  1  0   3   1
## Hornet Sportabout 18.7   8 360 175 3.15 3.440 17.02  0  0   3   2
## Valiant    18.1   6 225 105 2.76 3.460 20.22  1  0   3   1
```

Miles/(US) gallon vs. the car's Weight

Scaterplot

Show



Windows Taskbar: Type here to search, File, Chrome, PDF, Word, R, Powerpoint, 12:42, ENG, 7/02/2024, 2 notifications

The HTML file

The code is shown as a part of the output

The screenshot shows an R HTML document titled "er_prog3_SA_2024.html". The page has a sidebar with a navigation menu:

- 1. The data
- The mtcars dataset
- Miles/(US) gallon vs. the car's Weight
- 2. Simple linear regression using R
- 3. Data and estimated model
- 4. Model diagnostic

The main content area displays the date (15-12-2023), the author (Ziv Shkedy et al), and the title "Simple linear regression using R". Below the title, there is a "Code" dropdown menu with "Show" and "Hide" options. A red arrow points from the "Show" button to a red box highlighting the code execution results.

`dim(mtcars)`

`## [1] 32 11`

`head(mtcars)`

	mpg	cyl	disp	hp	drat	wt	qsec	vs	am	gear	carb
## Mazda RX4	21.0	6	160	110	3.90	2.620	16.46	0	1	4	4
## Mazda RX4 Wag	21.0	6	160	110	3.90	2.875	17.02	0	1	4	4
## Datsun 710	22.8	4	108	93	3.85	2.320	18.61	1	1	4	1
## Hornet 4 Drive	21.4	6	258	110	3.08	3.215	19.44	1	0	3	1
## Hornet Sportabout	18.7	8	360	175	3.15	3.440	17.02	0	0	3	2
## Valiant	18.1	6	225	105	2.76	3.460	20.22	1	0	3	1

Miles/(US) gallon vs. the car's Weight

Scaterplot

Windows taskbar at the bottom:

- Type here to search
- File Explorer icon
- Google Chrome icon
- Recycle Bin icon
- Word icon
- R icon (highlighted)
- PowerPoint icon
- System tray icons: battery, signal, volume, language (ENG), date (7/02/2024), and notifications.

Code in the Rmd file

The screenshot shows the RStudio interface with an Rmd file open. The code in the Rmd file is:

```
125 library(ggplot2)
128 library(gganimate)
129 library(av)
130 library(gifski)
131 library(foreach)
132 library("DAAG")
133 library(DT)
134
135 ``
136
137 # 1. The data
138
139
140 ## The <tt>mtcars</tt> dataset
141
142 ```{r, echo=TRUE, message=FALSE, warning=FALSE}
143 dim(mtcars)
144 head(mtcars)
145 ``
146
147 ## Miles/(us) gallon vs. the car's weight
148
149 ### Scatterplot
150
151 ```{r, echo=TRUE, message=FALSE, warning=FALSE, fig.cap="mpg vs. weight"}
152 #plot(mtcars$wt, mtcars$mpg, ylab = "mpg", xlab = "weight (0.000 lbs)")
153 qplot(wt, mpg, data = mtcars)
154 ``
155
156 ### Correlation
157
158 ```{r, echo=TRUE, message=FALSE, warning=FALSE}
159 cor(mtcars$wt, mtcars$mpg)
160 ``
161
162
163 # 2. simple linear regression using R
164
170:1 The <tt>lm()</tt> R function
```

A red box highlights the following sections of the code:

- ## The <tt>mtcars</tt> dataset
- ### Scatterplot
- ### Correlation
- # 2. simple linear regression using R

The RStudio environment pane shows:

- Environment: Environment is empty
- Files, Plots, Packages, Help, Viewer, Presentation tabs

The R console output is:

```
R 4.3.2 : C:\projects\Teaching2021\Course\R_2223\Classes/The pharma study/
R is free software and comes with ABSOLUTELY NO WARRANTY.
You are welcome to redistribute it under certain conditions.
Type 'license()' or 'licence()' for distribution details.

R is a collaborative project with many contributors.
Type 'contributors()' for more information and
'citation()' on how to cite R or R packages in publications.

Type 'demo()' for some demos, 'help()' for on-line help, or
'help.start()' for an HTML browser interface to help.
Type 'q()' to quit R.
```

The taskbar at the bottom shows various application icons.

The output in the HTML file

C:/projects/Teaching2021/Rcourse/R_2223/Classes/The pharma study/er_prog3_SA_2024.html
er_prog3_SA_2024.html | Open in Browser | Find | Publish

1. The data
The mtcars dataset
Miles/(US) gallon vs. the car's Weight
2. Simple linear regression using R
3. Data and estimated model
4. Model diagnostic



```
## Hornet Sportabout 18.7   8   360 175 3.15 3.440 17.02 0   0   3   2
## Valiant        18.1   6   225 105 2.76 3.460 20.22 1   0   3   1
```

Miles/(US) gallon vs. the car's Weight

Scaterplot

Show

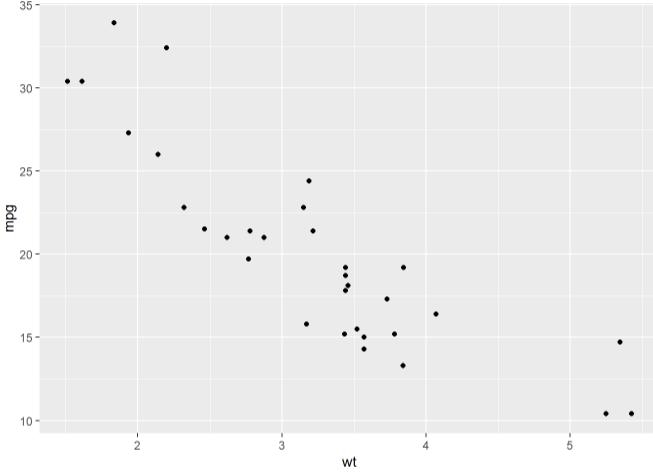


Figure 1: mpg vs. weight

Correlation

Show

```
## [1] -0.8676594
```

2. Simple linear regression using R

The `lm()` R function

For the `mtcars` dataset, we consider the model

$$mpg_i = \beta_0 + \beta_1 \times weight_i + \varepsilon_i$$

- Subsection
- Subsubsection
- Plot + correlation

45

Code in the Rmd file

RStudio

File Edit Code View Plots Session Build Debug Profile Tools Help

unit_05_inference_num_IWANOVAR.Rmd er_prog3_SA_2024.Rmd The pharma challenge_2022_prog1.Rmd er_prog2_SA_2024.Rmd er_prog1_SA_2024.Rmd

Go to file/function Addins

Source Visual

```
161  
162  
163 # 2. simple linear regression using R  
164  
165 ## The <tt>lm()</tt> R function  
166  
167 For the <tt>mtcars</tt> dataset, we consider the model  
168  
169  $\text{mpg}_{\text{i}} = \beta_0 + \beta_1 \times \text{weight}_{\text{i}} + \varepsilon_{\text{i}}$ .  
170  
171 ~~~{r, echo=TRUE, message=FALSE, warning=FALSE}  
172 fit.lm<-lm(mtcars$mpg~mtcars$wt)  
173 summary(fit.lm)  
174 ~~~  
175  
176  
177 The parametr estimates for the intercept and slope are equal, respectively, to  $\hat{\beta}_0 = 37.28$  and  
178  $\hat{\beta}_1 = -5.34$   
179 # 3. Data and estimated model  
180  
181 Figure 2 shows the data (mpg vs. weight) and fitted regression line,  $\hat{y} = 37.28 - 5.34 \times \text{wt}$ .  
182  
183 ~~~{r, echo=TRUE, message=FALSE, warning=FALSE, fig.cap="Data and fitted model"}  
184 ggplot(wt,mpg,data = mtcars)+  
185 geom_smooth(method = "lm",se = F)  
186 ~~~  
187  
188  
189 # 4. Model diagnostic  
190  
191  
192  
193 ## The <tt>mtcars</tt> dataset  
194  
195 For the <tt>mtcars</tt> data, the residuals from the model can be obtained by calling to the object <tt>resid</tt>,  
Figure 5 shows the diagnostic plots for the regression model.  
196  
170:1 The <tt>lm()</tt> R function
```

R Markdown

Console Terminal Render Background Jobs

R 4.3.2 - C:/projects/Teaching2021/Rcourse/R_2223/Clases/The pharma study/

R is free software and comes with ABSOLUTELY NO WARRANTY.
You are welcome to redistribute it under certain conditions.
Type 'license()' or 'licence()' for distribution details.

R is a collaborative project with many contributors.
Type 'contributors()' for more information and
'citation()' on how to cite R or R packages in publications.

Type 'demo()' for some demos, 'help()' for on-line help, or
'help.start()' for an HTML browser interface to help.
Type 'q()' to quit R.

> |

Project: (None)

Environment History Connections Tutorial

Import Dataset 154 MB

Global Environment

Environment is empty

Files Plots Packages Help Viewer Presentation

Zoom Export

Windows Taskbar

Type here to search

12:48 ENG 7/02/2024

The linear regression model

The output in the HTML file

C:/projects/Teaching2021/Rcourse/R_2223/Classes/The pharma study/er_prog3_SA_2024.html
er_prog3_SA_2024.html | Open in Browser | Find | Publish | ↗

Correlation

1. The data
2. Simple linear regression using R
The lm() R function
3. Data and estimated model
4. Model diagnostic

2. Simple linear regression using R

The lm() R function

For the mtcars dataset, we consider the model

$mpg_i = \beta_0 + \beta_1 \times weight_i + \varepsilon_i$

The model:
free text.

The output

Code for the
model is
shown.

fit.lm<-lm(mtcars\$mpg~mtcars\$wt)
summary(fit.lm)

```
##  
## Call:  
## lm(formula = mtcars$mpg ~ mtcars$wt)  
##  
## Residuals:  
##    Min     1Q   Median     3Q    Max  
## -4.5432 -2.3647 -0.1252  1.4096  6.8727  
##  
## Coefficients:  
##             Estimate Std. Error t value Pr(>|t|)  
## (Intercept) 37.2851   1.8776 19.858 < 2e-16 ***  
## mtcars$wt   -5.3445   0.5591 -9.559 1.29e-10 ***  
## ---  
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1  
##  
## Residual standard error: 3.046 on 30 degrees of freedom  
## Multiple R-squared:  0.7528, Adjusted R-squared:  0.7446  
## F-statistic: 91.38 on 1 and 30 DF,  p-value: 1.294e-10
```

The parametr estimates for the intercept and slope are equal, respectively, to $\hat{\beta}_0 = 37.28$ and $\hat{\beta}_1 = -5.34$

3. Data and estimated model

Figure 2 shows the data (mpg vs. weight) and fitted regression line, $\hat{mpg}_i = 37.28 - 5.34 \times wt_i$

35 -

Windows Taskbar: Type here to search, File, Home, Chrome, Word, R, Powerpoint, etc.

System tray: ENG, 12:49, 7/02/2024, battery icon, signal strength, etc.

Code in the Rmd file

RStudio
File Edit Code View Plots Session Build Debug Profile Tools Help
Go to file/function Addins
unit_05_inference_num_IWANOVAn.Rmd er_prog2_SA_2024.Rmd The pharma challenge_2022_prog1.Rmd er_prog2_SA_2024.R R Knit Run Outline Project: (None)
Source Visual
161
162
163 # 2. simple linear regression using R
164
165 ## The `lm()` R function
166
167 For the `mtcars` dataset, we consider the model
168
169 $\text{mpg}_{\text{i}} = \beta_0 + \beta_1 \times \text{weight}_{\text{i}} + \epsilon_{\text{i}}$.
170
171
172 ~~~{r, echo=TRUE, message=FALSE, warning=FALSE}
173 fit.lm<-lm(mtcars\$mpg~mtcars\$wt)
174 summary(fit.lm)
175 ~~~
176
177 The parametr estimates for the intercept and slope are equal, respectively, to $\hat{\beta}_0 = 37.28$ and
178 $\hat{\beta}_1 = -5.34$
179 # 3. Data and estimated model
180
181 Figure 2 shows the data (mpg vs. weight) and fitted regression line, $\hat{\text{mpg}}_{\text{i}} = 37.28 - 5.34 \times \text{wt}_{\text{i}}$
182
183 ~~~{r, echo=TRUE, message=FALSE, warning=FALSE, fig.cap="Data and fitted model"}
184 qplot(wt,mpg,data = mtcars)+
185 geom_smooth(method = "lm",se = F)
186 ~~~
187
188 # 4. Model diagnostic
189
190
191
192 ## The `mtcars` dataset
193
194 For the `mtcars` data, the residuals from the model can be obtained by calling to the object `resid`,
195 Figure 5 shows the diagnostic plots for the regression model.
196
170:1 The `lm()` R function
R 4.3.2 - C:/projects/Teaching2021/Rcourse/R_2223/Clases/The pharma study/
R is free software and comes with ABSOLUTELY NO WARRANTY.
You are welcome to redistribute it under certain conditions.
Type 'license()' or 'licence()' for distribution details.
R is a collaborative project with many contributors.
Type 'contributors()' for more information and
'citation()' on how to cite R or R packages in publications.
Type 'demo()' for some demos, 'help()' for on-line help, or
'help.start()' for an HTML browser interface to help.
Type 'q()' to quit R.
|
Console Terminal Render Background Jobs
R 43.2 - C:/projects/Teaching2021/Rcourse/R_2223/Clases/The pharma study/
Environment History Connections Tutorial
Import Dataset 154 MB Global Environment
Environment is empty
Files I
Free text not a part of the R code.
Text text text
~~{r}
R code
~~
Program structure.

The output in the HTML file

C:/projects/Teaching2021/Rcourse/R_2223/Classes/The pharma study/er_prog3_SA_2024.html | Open in Browser | Find | Publish |

1. The data
2. Simple linear regression using R
3. Data and estimated model
4. Model diagnostic

```
##  
## Residual standard error: 3.046 on 30 degrees of freedom  
## Multiple R-squared:  0.7528, Adjusted R-squared:  0.7446  
## F-statistic: 91.38 on 1 and 30 DF, p-value: 1.294e-10
```

The parametr estimates for the intercept and slope are equal, respectively, to $\hat{\beta}_0 = 37.28$ and $\hat{\beta}_1 = -5.34$

3. Data and estimated model

Figure 2 shows the data (mpg vs. weight) and fitted regression line, $\hat{mpg}_i = 37.28 - 5.34 \times wt_i$

Free text → ↑

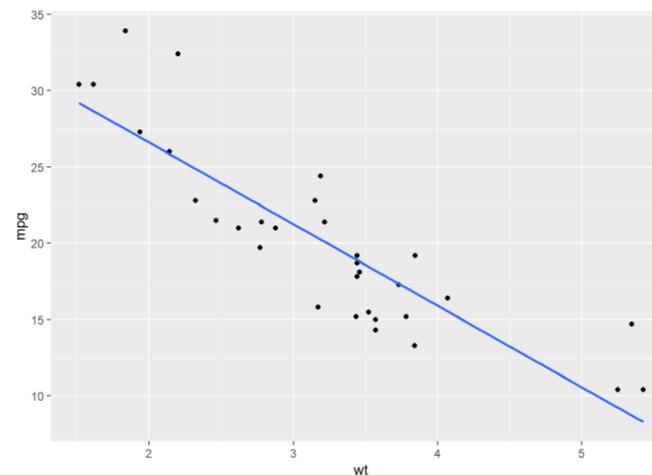


Figure 2: Data and fitted model

4. Model diagnostic

The mtcars dataset

For the mtcars data, the residuals from the model can be obtained by calling to the object `resid`. Figure 5 shows the diagnostic plots for the regression model.

```
## 1 2 3 4 5 6 7  
## -2.2826106 -0.9197704 -2.0858521 1.2873499 -0.2001440 -0.6932545 -3.9053627  
## 8 9 10 11 12 13 14
```

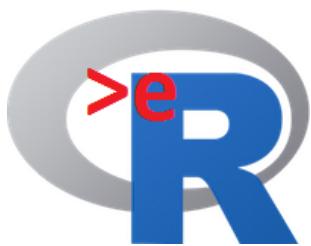
↑

Type here to search   ENG 12:50 7/02/2024

Short discussion

- R Studio + R markdown:
- Easy to use.
- Text + code.
- Output:
 - Standard: HTML, PDF, DOC.
 - Advanced: HTML.

- So far: simple analysis that produce an output.
 - How do we create a course ?



Interuniversity Institute for Biostatistics
and statistical Bioinformatics

Development of E-learning materials using R markdown

Part 4

The course online (1)

Steps in data analysis

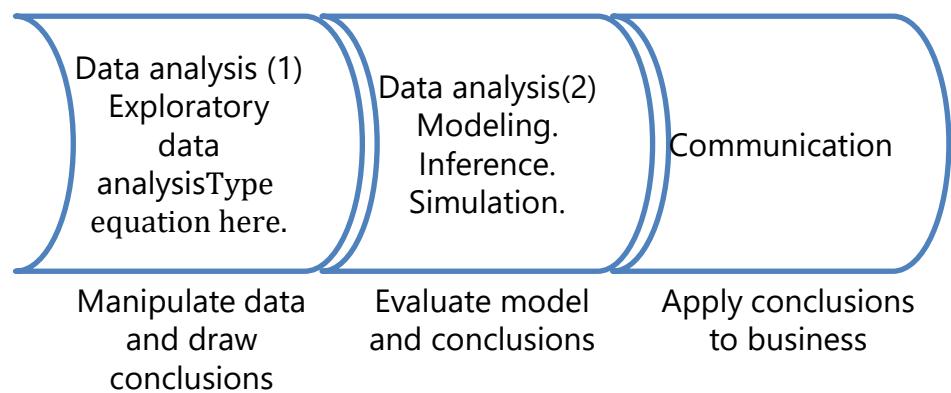
Modeling the association between the fuel consumption and the car's weight.

$$y_i = \beta_0 + \beta_1 x_i + \varepsilon_i$$

Methodology: simple linear regression.



We “translate” the methodology to software usage



Boxplot by treatment group.

A simple linear regression.

A report.

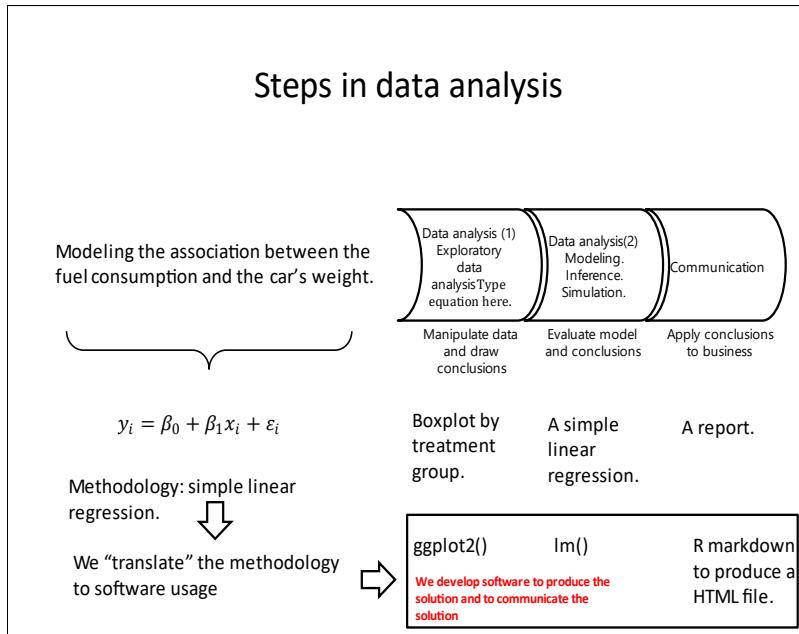
ggplot2()

lm()

R markdown to produce a HTML file.

We develop software to produce the solution and to communicate the solution

Developing a course about linear regression



- Suppose that we do not need to produce a report for an analysis but...
- Our aim: development of education materials for a course about linear regression.

Introduction to statistical modeling using R

This screenshot shows the homepage of the >eR-Biostat website. At the top, there's a navigation bar with links for Home, We R a community, Our platform, Our courses, Gallery, Developers, and Blog. Below the navigation is a large logo featuring a stylized 'R' and the text '>eR-BioStat'. The main content area has a heading 'E-learning using R: (Bio)statistics'. It includes a welcome message for the 2022 edition, a photo of a classroom, and a call-to-action button 'Click, Download & Teach'. A red arrow points from the right side of the slide towards this section.

<https://erbiostat.wixsite.com/erbiostat>

>eR-Biostat website.

List of courses.

This screenshot shows a list of courses categorized into three sections: Introductory, Advanced, and Basic. Each category contains several course titles represented by colored boxes. A red arrow points from the left side of the slide towards the 'Basic' section. Another red arrow points from the right side of the slide towards the 'Advanced' section.

Introductory	Advanced
Introduction to R	Applied Generalized Linear Models (GLM) using R
Statistical modeling: Linear regression using R	Modeling Binary Data using R
Statistical modeling: One-way ANOVA using R	Longitudinal data analysis (LDA) using R
Statistical modeling: Logistic regression using R	Linear models using R
Vizualizing data using R: an introduction	Survival Analysis using R
Basic concepts of statistical inference using R	An introduction to bootstrap using R
	Sample size calculation using R
	Exploratory multivariate data analysis using R
	Survival Analysis using R (A)

Basic

Basic concept in statistical inference using R (1)
Basic concept in statistical inference using R (2)
Linear Regression using R

Online books

Introduction to statistical modeling using R

This site was designed with the **WIX**.com website builder. Create your website today. [Start Now](#)

This group of courses is focused on statistical modelling and covers the following topics:

- Simple linear regression using R.
- One-way ANOVA using R.
- Logistic Regression using R.

The courses can be given together, as three parts of a course about statistical modelling or separately as a part of a specific course in statistical modelling.

The courses were developed at an [undergraduate](#) level courses (for both statistician and non statisticians).

This is an open source course and all source files used to produce the slides are available online (in PP, Tex or Rmd formats).

Figure 3.1: Chick weight by diet group.

```
boxplot(chicks$weight ~ chicks$feed)
```

About

General information about the course and course materials and the study methods used in the course.

Topics

In this page, the course is presented in a typical slides format. The course

Online tutorials

In this page, supporting online tutorials are given in different formats. The online

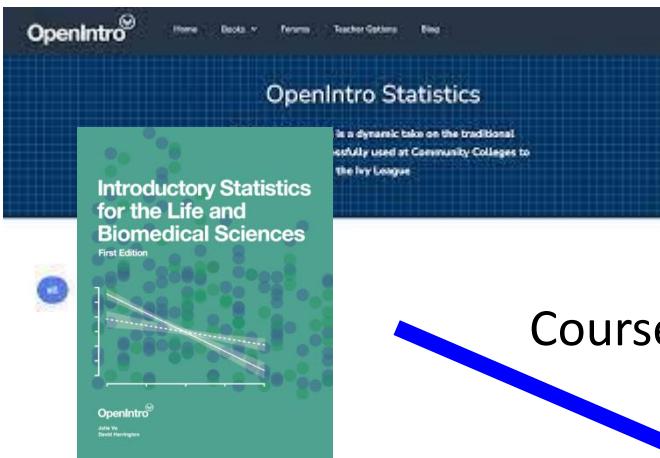


Type here to search



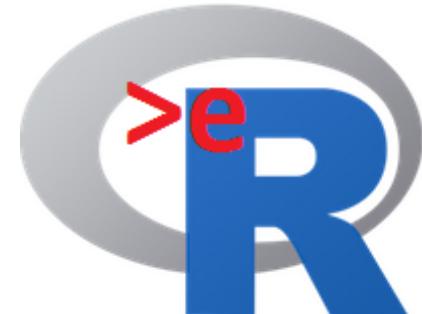
13:00
ENG
7/02/2024
2

Introduction to statistical modeling using R



Course II

Course materials from two sources



Course I

A screenshot of the >eR-BioStat website. The header includes the R logo, the site name 'Introduction to Statistical modeling using R', and navigation links for Home, About, Topics (which is circled in red), Online tutorials, and Contact. The main content area discusses the course focus on statistical modeling and lists topics like Simple linear regression using R, One-way ANOVA using R, and Logistic Regression using R. It also mentions the availability of source files and online tutorials. A sidebar shows a thumbnail of a presentation slide with a scatter plot. The footer contains links for About, Topics, and Online tutorials, along with standard footer text.

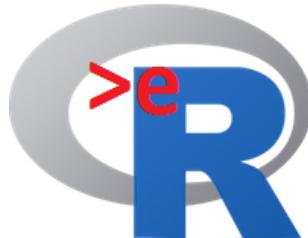
Course I

wx Home | Erbiostat wx Topics | IntroStatMod +

erbiostat.wixsite.com/introstatmod/topics

uhasselt.be bookmarks All Bookmarks

This site was designed with the WIX.com website builder. Create your website today. [Start Now](#)



External datasets for illustration are included in the data repositories.

Simple Linear Regression

This course covers the topic of simple linear regression using the R function lm(). Topics (all presented at a basic level) covered in the course include:

- Introduction and model formulation.
- Fitting a simple linear regression model using the lm() function in R.
- Model diagnostic.
- Model diagnostic in R.

External datasets are available in the data repository.

Slides (PDF): simple linear regression
Slides (PP): simple linear regression
R programm
Datasets

One-Way ANOVA

This course covers the topic of one way ANOVA models using the R function aov(). Topics (all presented at a basic level) covered in the course include:

- The one-way ANOVA model.
- Sources of Variability.
- One-way ANOVA using R: the aov() function.
- Model formulation and hypotheses testing.
- Analysis of the pharmaceutical experiment.
- Model diagnostic in R: normal probability plot.
- Multiple testing.

External datasets are available in the data repository.

Slides (PDF): One-Way ANOVA
Slides (PP): One-Way ANOVA
R programm
Datasets

Simple Logistic Regression

This course covers the topic of simple logistic regression using the R function glm(). Topics (all presented at a basic level) covered in the course include:

- Introduction and example tour.
- Fitting a simple linear logistic regression model using the glm() function in R.
- Model formulation.
- Interpretation of the model parameters.

External datasets are available in the data repository.

Slides (PDF): Logistic regression
Slides (PP): Logistic regression
R programm
Datasets

Course materials:

- Slides (PDF)
- Slides(PowerPoint)
- R program for the examples in the course.
- Datasets.

→



Type here to search



13:00
7/2/2024

Examples of the slide

A screenshot of a web browser displaying a PDF document titled "eR-Biostat_Introduction to Statistical Modeling using R_Regressionin_2022_V1.pdf". The browser has three tabs open: "Home | Erbiostat", "Topics | IntroStatMod", and the current tab showing the PDF.

The PDF content includes:

- The eR-Biostat initiative logo, which is a stylized "R" with a red "e" and a blue ">" symbol.
- The text: "The >eR-Biostat initiative" and "Making R based education materials in statistics accessible for all".
- The title of the document: "Basic concepts in statistical modeling using R: simple linear regression".
- The text: "Developed by Legesse Kassa Debusho (UNISA, South Africa) and Ziv Shkedy (Hasselt University)".
- The URL: <https://erbiostat.wixsite.com/erbiostat>
- The text: "LAST UPDATED: 2022"
- Social media links: Facebook, GitHub, and Twitter.
- Email address: erbiostat@gmail.com
- The PDF file size is 1.21 MB and it was generated with GitHub Copilot.

The browser interface shows a sidebar with "uhasselt.be bookmarks" and a search bar at the bottom. The taskbar at the very bottom shows icons for File Explorer, Google Chrome, and Microsoft Edge.

Course materials

This site was designed with the **WIX.com** website builder. Create your website today. [Start Now](#)

Simple Linear Regression

This course covers the topic of simple linear regression using the R function lm(). Topics (all presented at a basic level) covered in the course include:

- Introduction and model formulation.
- Fitting a simple linear regression model using the lm() function in R.
- Model diagnostic.
- Model diagnostic in R.

External datasets are available in the data repository.

One-Way ANOVA

This course covers the topic of one way ANOVA models using the R function aov(). Topics (all presented at a basic level) covered in the course include:

- The one-way ANOVA model.
- Sources of Variability.
- One-way ANOVA using R: the aov() function.
- Model formulation and hypotheses testing.
- Analysis of the pharmaceutical experiment.
- Model diagnostic in R: normal probability plot.
- Multiple testing.

External datasets are available in the data repository.

Simple Logistic Regression

This course covers the topic of simple logistic regression using the R function glm(). Topics (all presented at a basic level) covered in the course include:

- Introduction and example tour.
- Fitting a simple linear logistic regression model using the glm() function in R.
- Model formulation.
- Interpretation of the model parameters.

External datasets are available in the data repository.

Course I

Slides (PDF): simple linear regression
Slides (PP): simple linear regression
R programm
Datasets

Slides (PDF): One-Way ANOVA
Slides (PP): One-Way ANOVA
R programm
Datasets

Slides (PDF): Logistic regression
Slides (PP): Logistic regression
R programm
Datasets

- Basic course about simple linear regression, One-Way ANOVA and logistic regression.
- Developed as a part of the >eR-BioStat initiative.

Type here to search 13:00 ENG 7/02/2024

59

Course materials

This site was designed with the **WIX.com** website builder. Create your website today. [Start Now](#)

The website layout includes:

- Left Sidebar:** Shows the OpenIntro Statistics book cover for "Introductory Statistics for the Life and Biomedical Sciences, First Edition".
- Header:** Includes a "Start Now" button for Wix.com.
- Content Sections:**
 - Simple linear regression:** Based on unit 6, covers topics like scatterplots, least squares regression, and statistical inference in regression. Includes links for Slides (PDF), Slides (PP), and Slides (Rmd).
 - One-Way ANOVA:** Based on unit 5, covers topics like ideas behind ANOVA, assumptions, and pairwise comparisons. Includes links for Slides (PDF) and Slides (Rmd).
 - Logistic regression:** Based on unit 9, covers odds and probabilities, logistic regression models, and inference for simple logistic regression. Includes links for Slides (PDF), Slides (PP), and Slides (Rmd).
- Bottom Navigation:** Includes a search bar and various system icons.

Red annotations:

- A red box highlights the first two sections: Simple linear regression and One-Way ANOVA.
- A red bracket on the right groups the Simple linear regression and One-Way ANOVA sections under the heading "Course II".

List of course features:

- Basic course about simple linear regression, One-Way ANOVA and logistic regression.
- Developed using materials available online from the OpenIntro consortium.

© 2020 by eR-BioStat. Proudly created with Wix.com

Course II

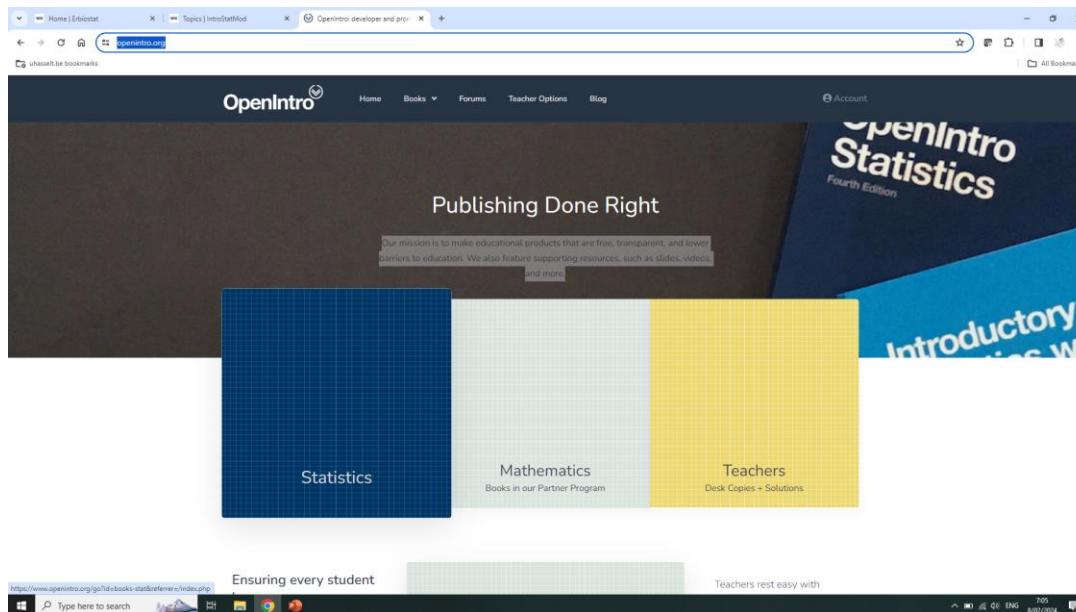
openintro.org

- The OpenIntro project was founded in 2009 to improve the quality and availability of education by producing:
 - Exceptional books.
 - Teaching tools.
- Free to use and easy to modify.
- “Our inaugural effort is *OpenIntro Statistics*. Probability is optional, inference is key, and we feature real data whenever possible”.
- Files for the entire book are freely available at openintro.org.

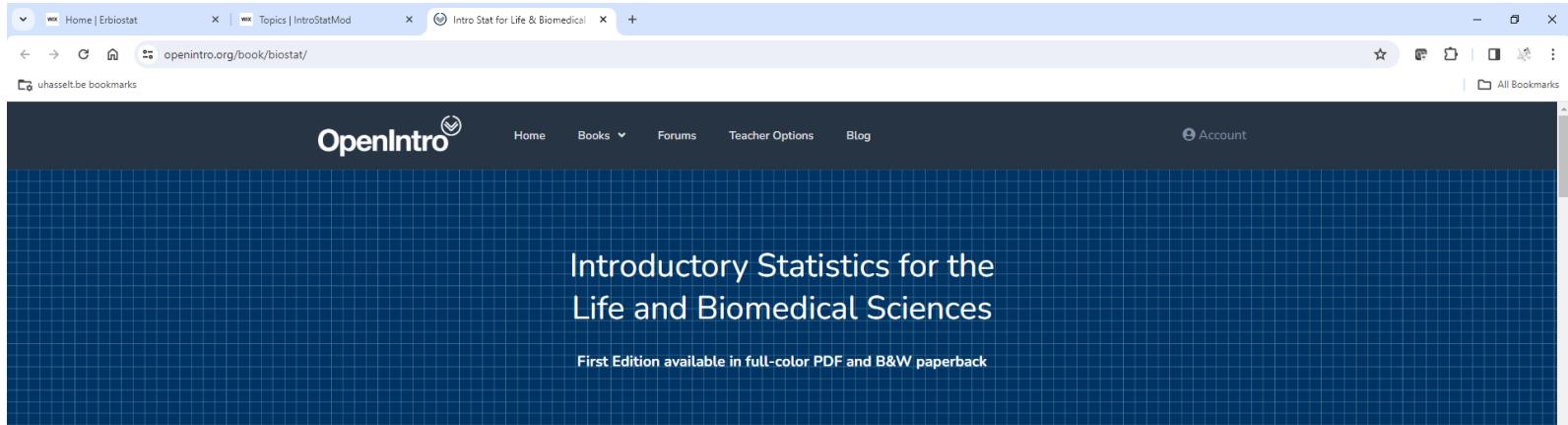
openintro.org

<https://www.openintro.org/>

“Our mission is to make educational products that are **free**, **transparent**, and lower barriers to education. We also feature supporting resources, such as slides, videos, and more.”



Introductory Statistics for the Life and Biomedical Sciences



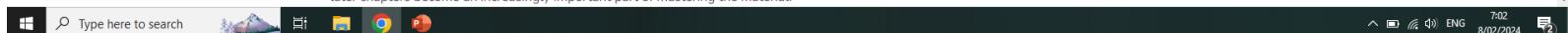
Textbook Pedagogy

Introduction to Statistics for the Life and Biomedical Sciences has been written to be used in conjunction with a set of self-paced learning labs. These labs guide students through learning how to apply statistical ideas and concepts discussed in the text with the R computing language.

The text discusses the important ideas used to support an interpretation (such as the notion of a confidence interval), rather than the process of generating such material from data (such as computing a confidence interval for a particular subset of individuals in a study). This allows students whose main focus is understanding statistical concepts to not be distracted by the details of a particular software package. In our experience, however, we have found that many students enter a research setting after only a single course in statistics. These students benefit from a practical introduction to data analysis that incorporates the use of a statistical computing language.

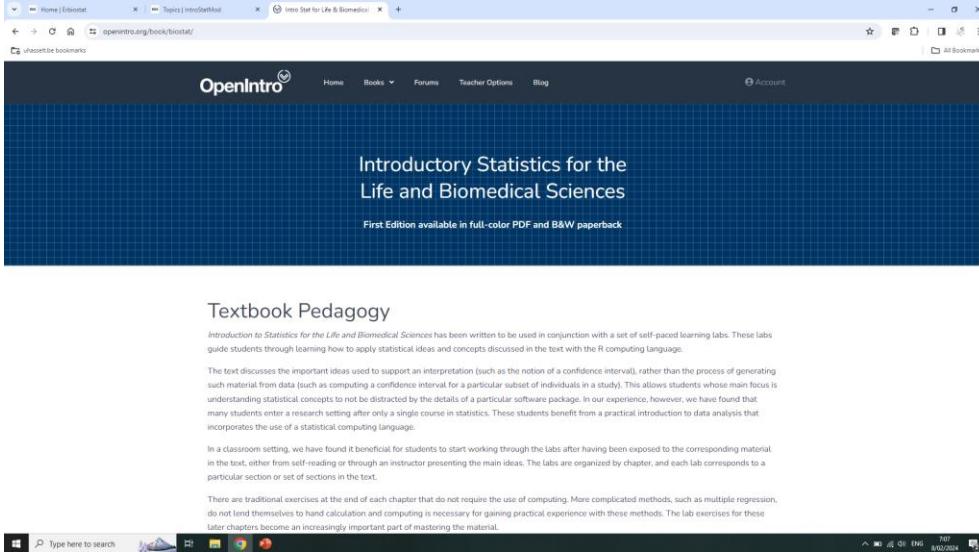
In a classroom setting, we have found it beneficial for students to start working through the labs after having been exposed to the corresponding material in the text, either from self-reading or through an instructor presenting the main ideas. The labs are organized by chapter, and each lab corresponds to a particular section or set of sections in the text.

There are traditional exercises at the end of each chapter that do not require the use of computing. More complicated methods, such as multiple regression, do not lend themselves to hand calculation and computing is necessary for gaining practical experience with these methods. The lab exercises for these later chapters become an increasingly important part of mastering the material.



<https://www.openintro.org/book/biostat/>

Introductory Statistics for the Life and Biomedical Sciences



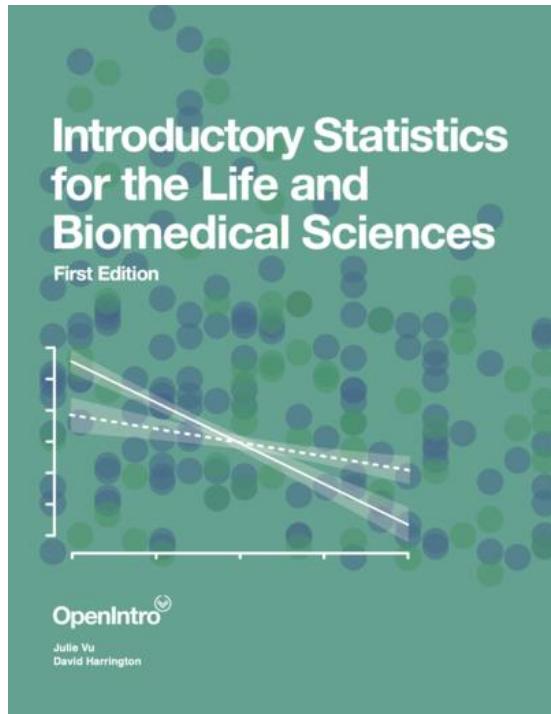
OpenIntro Biostatistics:
developing open source materials
for teaching and learning applied
statistics with R.

Developed by [Dave Harrington](#)
and [Julie Vu](#).

Developed for students in health or life sciences:

- Motivated more by application than theory.
- No previous statistics courses.
- Familiarity with basic algebra.
- No or limited experience with computing.

Introductory Statistics for the Life and Biomedical Sciences



- Written by Dave Harrington and Julie Vu.
- Available on line (for free) on PDF format.
- Focus on theory and practice.
- Undergraduate/graduate levels.
- Available (for free) online:
 - Source files for the book in.
 - Presentations and practical sessions (labs).

Reference for Course I & II

This site was designed with the **WIX**.com website builder. Create your website today. [Start Now](#)

Reference for course I & II

Course I

Course II

Course III

Online book

Online book

Online book

Simple linear regression

One-Way ANOVA

Logistic regression

ANOVA

Start Now

uhasselt.be bookmarks

Home | Erbiostat

Topics | IntroStatMod

erbiostat.wixsite.com/introstatmod/topics

uhasselt.be bookmarks

All Bookmarks

© 2020 by >eR-BioStat. Proudly created with Wix.com

Course II: Slides

Home | Eriostat Topics | IntroStatMod Courses/Introductory Courses/

github.com/eR-Biostat/Courses/blob/master/Introductory%20Courses/Introduction%20to%20statistical%20modeling%20using%20R/Online%20materials/unit_06_simple_linear_regression.pdf

uhasselt.be bookmarks All Bookmarks

Files

master + Search

Go to file

Basic courses
Coordination
Data Analysis
ICP Workshop
Inference
Introductory Courses
Introduction to R
Introduction to statistical mode...
Logistic regression
One way ANOVA
Online materials
README.md
unit_05_inference_num_1WA...
unit_05_inference_num_1WA...
unit_05_inference_num_1WA...
unit_06_simple_linear_regres...
unit_06_simple_linear_regres...
unit_06_simple_linear_regres...
unit_09_logistic_regression.R...
unit_09_logistic_regression.pdf
unit_09_logistic_regression.p...
Simple linear regression
README.md
Visualizing data using R- an intr...

664 KB Code 55% faster with GitHub Copilot

Slides in PDF.

PP files are available online.

Unit 6: Simple Linear Regression

Statistics 102 Teaching Team

March 30, 2020

1 / 40

Introduction

Examining scatterplots

Least squares regression

Type here to search 13:06 ENG 7/02/2024

Course II: The Rmd file for the slides

The screenshot shows a GitHub repository for 'eR-Biostat / Courses'. The repository contains an Rmd file named 'unit_06_simple_linear_regression.Rmd'. The code block displays the following content:

```
1  ---
2  title: "Unit 6: Simple Linear Regression"
3  author: "Statistics 102 Teaching Team"
4  date: "March 30, 2020"
5  output:
6    beamer_presentation:
7      includes:
8        in_header: ../slides_header.tex
9        fig_width: 3.25
10       fig_height: 3
11       fig_caption: false
12       toc: true
13       keep_tex: true
14       classoption: "aspectratio=169"
15       slide_level: 3
16   ...
17
18   # Introduction
19
20   ### The main ideas
21
22   \small
23
24   Linear regression provides methods for examining the association between a quantitative response variable and a set of possible predictor variables.
25
26   - Linear regression should only be used with data that exhibit linear or approximately linear relationships.
27
28   **Simple linear regression** is used to estimate the linear relationship between a response variable  $y$  and a single predictor  $x$ .
29
30   - The response variable  $y$  can be referred to as the *dependent* variable, and the predictor variable  $x$  the *independent* variable.
```

Course II: A PDF output

The screenshot shows a Microsoft Edge browser window with three tabs open:

- Home | Erbiostat
- Topics | IntroStatMod
- Courses/Introductory Courses/

The main content area displays a PDF titled "Courses / Introductory Courses / Introduction to statistical modeling using R / Online materials / unit_06_simple_linear_regression.pdf". The PDF has a file size of 664 KB and was generated 55% faster with GitHub Copilot.

The page number is 3 / 40. The title "THE MAIN IDEAS" is visible. The text states: "Linear regression provides methods for examining the association between a quantitative response variable and a set of possible predictor variables." Below this is a bulleted list:

- Linear regression should only be used with data that exhibit linear or approximately linear relationships.

The text then defines "Simple linear regression" as "used to estimate the linear relationship between a response variable y and a single predictor x ". Below this is another bulleted list:

- The response variable y can be referred to as the *dependent* variable, and the predictor variable x the *independent* variable.
- The statistical model for simple linear regression is based on the straight line relationship

$$y = b_0 + b_1x$$

The page number is 4 / 40. The title "THE MAIN IDEAS ..." is visible. The text states: "Multiple linear regression is used to estimate the linear relationship between a".

The left sidebar shows a file tree for the "master" branch of the repository, listing various course units and files such as "Basic courses", "Coordination", "Data Analysis", "ICP Workshop", "Inference", "Introductory Courses", "Introduction to R", "Introduction to statistical mode...", "Logistic regression", "One way ANOVA", "Online materials", "README.md", "unit_05_inference_num_1WA...", "unit_05_inference_num_1WA...", "unit_05_inference_num_1WA...", "unit_06_simple_linear_regres...", "unit_06_simple_linear_regres...", "unit_06_simple_linear_regres...", "unit_09_logistic_regression.R...", "unit_09_logistic_regression.pdf", "unit_09_logistic_regression.p...", "Simple linear regression", "README.md", and "Visualizing data using R- an intr...".

Course I + II: online text about linear regression using R

This site was designed with the **WIX.com** website builder. Create your website today. [Start Now](#)

respectively, available online here:

Online book

Online book

This part of the course is based on [unit 6](#) in Vu & Harrington course and it covers the following topics:

- Examining scatterplots.
- Least squares regression.
- Interpreting a linear model.
- Statistical inference in regression.

Slides (PDF): simple linear regression

Slides (PP): simple linear regression

Slides (Rmd): simple linear regression

This part of the course is based on [unit 5](#) in Vu & Harrington course and it covers the following topics:

- Ideas behind One-Way ANOVA..
- Assumptions for ANOVA.
- Normal probability plots (Q-Q plots).
- Pairwise comparisons.
- ANOVA model in R using the `aov()` function

Slides (PDF): One-Way ANOVA

Slides (PP): One-Way ANOVA

Slides (Rmd): One-Way ANOVA

This part of the course is based on [unit 9](#) in Vu & Harrington course and it covers the following topics:

- Odds and probabilities.
- Introduction to logistic regression.
- Simple logistic regression.
- Logistic versus linear regression.
- Inference for simple logistic regression.

Slides (PDF): Logistic regression

Slides (PP): Logistic regression

Slides (Rmd): Logistic regression

Written by Dave H. and Jullie V.

Developed by ZS based on text of Dave H. and Julie V.

Simple linear regression

© 2020 by >eR-BioStat. Proudly created with Wix.com

A part of the website of the course

wx Home | Eriostat wx Topics | IntroStatMod Temp_2024_prog1_V2.knit

https://htmlpreview.github.io/?https://github.com/eR-Biostat/Courses/blob/master/Introductory%20Courses/Introduction%20to%20statistical%20modeling%20using%20R/Simple%20linear%20regression/Slides/Temp_2024_prog1_V2.html

uhasselt.be bookmarks All Bookmarks

Code ▾

15-12-2023 >eR-BioStat

Simple linear regression using R

Ziv Shkedy et al

```
## load/install libraries
.libPaths(c("./Rpackages", .libPaths()))
library(knitr)
library(tidyverse)
library(deSolve)
library(minpack.lm)
library(ggpubr)
library(readxl)
library(gamlss)
library(data.table)
library(grid)
library(png)
library(lme)
library(gridExtra)
library(mvtnorm)
library(e1071)
library(lattice)
library(ggplot2)
library(dslabs)
library(NHANES)
library(plyr)
library(dplyr)
library(nasawebster)
library(ggplot2)
library(gganimate)
library(av)
library(gifski)
library(foreach)
library("DAAG")
library(DT)
```

1. General Introduction

Linear regression models

Type here to search     

7:11 ENG 8/02/2024

71

A part of the website of the course

- Text + example about simple linear regression.
- Rmd file to produce the HTML.

The screenshot shows a web browser window with the URL https://htmlpreview.github.io/?https://github.com/eR-Biostat/Courses/blob/master/Introductory%20Courses/Introduction%20to%20statistical%20modeling%20using%20R/Simple%20linear%20regression/Slides/Temp_2024_prog1_V2.html. The page content includes:

- library(DT)**
- ## 1. General Introduction

Linear regression models

Linear regression provides methods for examining the association between a quantitative response variable and a set of possible predictor variables. Linear regression should only be used with data that exhibit linear or approximately linear relationships. **Simple linear regression** is used to estimate the linear relationship between a response variable y and a single predictor x . The response variable y can be referred to as the *dependent* variable, and the predictor variable x as the *independent* variable. The statistical model for simple linear regression is based on the straight line relationship

$$y = b_0 + b_1 x$$

Multiple linear regression is used to estimate the linear relationship between a response variable y and several predictors x_1, x_2, \dots, x_p . The statistical model for multiple linear regression is based on

$$y = b_0 + b_1 x_1 + b_2 x_2 + \dots + b_p x_p$$
- ### Examining scatterplots
- #### The mtcars dataset

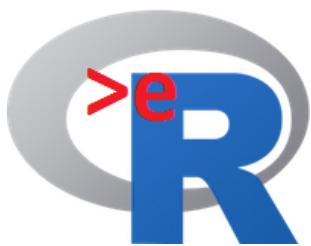
The *Motor Trend Car Road Tests* data was extracted from the 1974 Motor Trend US magazine, and comprises fuel consumption and 10 aspects of automobile design and performance for 32 automobiles (1973–74 models). It is available in R as `mtcars`. The dataset contains information about 11 variables and 32 cars. Use `help(mtcars)` to get more information about the data.

```
dim(mtcars)
## [1] 32 11
head(mtcars)
```

	mpg	cyl	disp	hp	drat	wt	qsec	vs	am	gear	carb
## Mazda RX4	21.0	6	160	110	3.90	2.620	16.46	0	1	4	4
## Mazda RX4 Wag	21.0	6	160	110	3.90	2.875	17.02	0	1	4	4
## Datsun 710	22.8	4	108	93	3.85	2.320	18.61	1	1	4	1
## Hornet 4 Drive	21.4	6	258	110	3.08	3.215	19.44	1	0	3	1
## Hornet Sportabout	18.7	8	360	175	3.15	3.440	17.02	0	0	3	2
## Valiant	18.1	6	225	105	2.76	3.460	20.22	1	0	3	1

Miles/(US) gallon vs. the car's Weight

Windows taskbar icons: Search, File Explorer, Google Chrome, Microsoft Edge, Task View, Volume, Network, Battery, ENG, 7/11, 8/02/2024.



Interuniversity Institute for Biostatistics
and statistical Bioinformatics

Development of E-learning materials using R markdown

Part 5a: developing the content

R Program: er_prog4_SA_2024.Rmd

Content on the website

- The Rmd file for the online book: the program and output.

The image shows two windows side-by-side on a Windows desktop. The left window is a web browser displaying the 'Introduction to Statistical modeling using R' website, which is built with WIX.com. It features an 'R' logo, navigation links for Home, About, Topics, Online tutorials, and Contact, and a section titled '>eR-BioStat'. The right window is a code editor showing an R Markdown (Rmd) file named 'Temp_2024/prog_v2.Rmd'. The code includes a title '1. General Introduction' and a section on 'Linear regression models' with a mathematical formula for simple linear regression: $y = b_0 + b_1x$. Below this, there is a section on 'Examining scatterplots' with a code snippet for the 'mtcars' dataset, including the command `head(mtcars)` and its output, which lists various car models with their fuel consumption and other performance metrics.

Can be produced using the Rmd program:
er_prog4_SA_2024.Rmd

The Rmd program

Document setting



```
1 ---  
2 output:  
3   bookdown::html_document2:  
4     toc: TRUE  
5     toc_float: TRUE  
6     toc_depth: 2  
7     number_sections: no  
8     css: ./lib/stylesArial.css  
9     code_folding: hide  
10  
11 params:  
12   department: ">eR-Biostat"  
13   topic: <font size = "10" > **simple linear regression using R **</font>  
14   author: "Ziv Shkedy et al"  
15   date: "15-12-2023"  
16   endCode: FALSE  
17   RmdLocation: ""  
18 ---  
19  
20  
21 <p>  
22     
23 </p>  
24  
25  
26  
27 ```{r delaycodeprinting, message=FALSE, warning=FALSE, echo = FALSE}  
28 # you can delete this chunk if you do not want delaycodeprinting and adjust the YAML header accordingly  
29 library(knitr)  
30 # The **delaycodeprinting** chunk below allows all R code to be printed at the end of the report (endCode = TRUE)  
31 # or prints the RMDlocation from the YAML header as a code reference (endCode != TRUE)  
32 # see code chunk named 'codeprint'  
33 delay_code_labels <- NULL  
34 knit_hooks$set(delay = function(before, options, envir) {  
35   if (before) {  
36     delay_code_labels <- append(delay_code_labels, options$label)  
37   }  
38   return(NULL) ## otherwise knitr will print delay_code_labels every time  
39 })  
40
```

Environment is empty

Files Plots Packages Help Viewer Presentation

Console Terminal Render Background Jobs

R 4.3.2 · C:\Projects\Teaching2021\Rcourse\R_2223\Classes/The pharma study/

R is free software and comes with ABSOLUTELY NO WARRANTY.
You are welcome to redistribute it under certain conditions.
Type 'license()' or 'licence()' for distribution details.

R is a collaborative project with many contributors.
Type 'contributors()' for more information and
'citation()' on how to cite R or R packages in publications.

Type 'demo()' for some demos, 'help()' for on-line help, or
'help.start()' for an HTML browser interface to help.
Type 'q()' to quit R.

The HTML output

C:/projects/Teaching2021/Rcourse/R_2223/Classes/The pharma study/er_prog4_SA_2024.html
er_prog4_SA_2024.html | Open in Browser | Find | Code ▾ | Publish | X

1. General Introduction

- 2. The least squares regression model
- 3. Simple linear regression using R
- 3. Model diagnostic
- 4. Categorical predictors with two levels
- 5. Goodness of fit: using R^2 to describe the strength of a fit
- 6. Statistical inference in regression

15-12-2023 >eR-BioStat

Simple linear regression using R

Ziv Shkedy et al Show

1. General Introduction

Linear regression models

Linear regression provides methods for examining the association between a quantitative response variable and a set of possible predictor variables. Linear regression should only be used with data that exhibit linear or approximately linear relationships. **Simple linear regression** is used to estimate the linear relationship between a response variable y and a single predictor x . The response variable y can be referred to as the *dependent* variable, and the predictor variable x as the *independent* variable. The statistical model for simple linear regression is based on the straight line relationship

$$y = b_0 + b_1 x$$

Multiple linear regression is used to estimate the linear relationship between a response variable y and several predictors x_1, x_2, \dots, x_p . The statistical model for multiple linear regression is based on

$$y = b_0 + b_1 x_1 + b_2 x_2 + \dots + b_p x_p$$

Examining scatterplots

The mtcars dataset

The *Motor Trend Car Road Tests* data was extracted from the 1974 Motor Trend US magazine, and comprises fuel consumption and 10 aspects of automobile design and performance for 32 automobiles (1973–74 models). It is available in R as `mtcars`. The dataset contains information about 11 variables and 32 cars. Use `help(mtcars)` to get more information about the data.

```
## [1] 32 11
```

	mpg	cyl	disp	hp	drat	wt	qsec	vs	am	gear	carb
## Mazda RX4	21.0	6	160	110	3.90	2.620	16.46	0	1	4	4
## Datsun 710	21.0	6	160	110	3.90	2.875	17.02	0	1	4	4

The Rmd program

RStudio
File Edit Code View Plots Session Build Debug Profile Tools Help
unit_05_inference_num_1ANOVA.Rmd x er_prog3_SA_2024.Rmd x er_prog4_SA_2024.Rmd x The pharma challenge_2022_prog1.Rmd x er_prog2_SA_2024.Rmd x
Go to file/function Addins x
Source Visual
136
137 # 1. General Introduction
138
139 ## Linear regression models
140
141 Linear regression provides methods for examining the association between a quantitative response variable and a set of possible predictor variables.
142 Linear regression should only be used with data that exhibit linear or approximately linear relationships. **Simple linear regression** is used to estimate the linear relationship between a response variable y and a single predictor x . The response variable y can be referred to as the "dependent" variable, and the predictor variable x the "independent" variable. The statistical model for simple linear regression is based on the straight line relationship
143
$$y = b_0 + b_1x \quad \text{newline}$$

144 **Multiple linear regression** is used to estimate the linear relationship between a response variable y and several predictors x_1, x_2, \dots, x_p . The statistical model for multiple linear regression is based on
145
$$y = b_0 + b_1x_1 + b_2x_2 + \dots + b_px_p \quad \text{newline}$$

146
147 ## Examining scatterplots
148
149 ### The `mtcars` dataset
150
151 The "Motor Trend Car Road Tests" data was extracted from the 1974 Motor Trend US magazine, and comprises fuel consumption and 10 aspects of automobile design and performance for 32 automobiles (1973-74 models). It is available in R as `mtcars`. The dataset contains information about 11 variables and 32 cars. Use `help(mtcars)` to get more information about the data.
152
153 ```{r, echo=TRUE, message=FALSE, warning=FALSE}
154 dim(mtcars)
155 head(mtcars)
156```
157
158 ### Miles/(us) gallon vs. the car's weight
159
160 our aim is to investigate the relationship between the fuel consumption (in Miles/(us) gallon, the R object
161 `mpg`) and the car's weight (in 1000 lbs), the R object `wt`.
162
163 ```{r, echo=TRUE, message=FALSE, warning=FALSE, fig.cap="Investigating the relationship between fuel consumption and car weight."}
164
165 Title c
Environment History Connections Tutorial
Import Dataset 156 MB
Global Environment
• Section
• Subsection
• Free text...
Environment is empty
Files Plots Packages Help Viewer Presentation
Zoom Export
Console Terminal Render Background Jobs
R 4.3.2 - C:/projects/Teaching2021/Rcourse/R_2223/Classes/The pharma study/
R is free software and comes with ABSOLUTELY NO WARRANTY.
You are welcome to redistribute it under certain conditions.
Type 'license()' or 'licence()' for distribution details.
R is a collaborative project with many contributors.
Type 'contributors()' for more information and
'citation()' on how to cite R or R packages in publications.
Type 'demo()' for some demos, 'help()' for on-line help, or
'help.start()' for an HTML browser interface to help.
Type 'q()' to quit R.
Type here to search

The HTML output

The screenshot shows a web browser displaying an R Markdown document. The title of the document is "Simple linear regression using R". The page includes a sidebar with a navigation menu, a header with the date and author information, and a main content area with sections on simple linear regression models and the mtcars dataset.

• Section → 1. General Introduction

• Subsection → Linear regression models

• Free text... →

Linear regression provides methods for examining the association between a quantitative response variable and a set of possible predictor variables. Linear regression should only be used with data that exhibit linear or approximately linear relationships. **Simple linear regression** is used to estimate the linear relationship between a response variable y and a single predictor x . The response variable y can be referred to as the *dependent* variable, and the predictor variable x the *independent* variable. The statistical model for simple linear regression is based on the straight line relationship

$$y = b_0 + b_1 x$$

Multiple linear regression is used to estimate the linear relationship between a response variable y and several predictors x_1, x_2, \dots, x_p . The statistical model for multiple linear regression is based on

$$y = b_0 + b_1 x_1 + b_2 x_2 + \dots + b_p x_p$$

Examining scatterplots

The mtcars dataset

The *Motor Trend Car Road Tests* data was extracted from the 1974 Motor Trend US magazine, and comprises fuel consumption and 10 aspects of automobile design and performance for 32 automobiles (1973–74 models). It is available in R as `mtcars`. The dataset contains information about 11 variables and 32 cars. Use `help(mtcars)` to get more information about the data.

```
## [1] 32 11
```

	mpg	cyl	disp	hp	drat	wt	qsec	vs	am	gear	carb
## Mazda RX4	21.0	6	160	110	3.90	2.620	16.46	0	1	4	4
## Datsun 710	22.8	4	108	93	3.85	2.875	17.02	0	1	4	1
## Hornet 4 Drive	18.7	6	167.6	123	3.08	3.435	17.82	0	1	4	4
## Hornet Sportabout	18.3	8	225.8	140	3.15	3.440	17.30	0	0	5	4
## Valiant	14.3	8	236.2	140	3.08	3.850	17.40	0	0	5	4
## Fiat 128	22.8	4	94.1	62	3.00	2.200	19.46	0	1	4	1
## Lincoln Continental	10.4	8	255.2	120	3.00	4.400	20.00	0	0	5	4
## Chrysler New Yorker	15.2	8	307.0	150	3.00	4.400	20.00	0	0	5	4
## Dodge Dart	17.8	8	167.6	123	3.00	3.440	17.00	0	0	4	4
## AMC Rebel	14.3	8	167.6	123	3.00	3.440	17.00	0	0	4	4
## Pontiac Firebird	16.4	8	236.2	140	3.00	3.440	17.00	0	0	4	4
## Fiat X1/9	19.2	4	90.4	62	3.00	2.200	19.46	0	1	4	1
## Toyota Corolla	17.8	4	95.1	62	3.00	2.200	19.46	0	1	4	1
## AMC Gremlin	17.3	4	93.7	62	3.00	2.200	19.46	0	1	4	1
## Volvo 142	14.3	4	120.8	70	3.00	2.200	19.46	0	1	4	1
## AMC Hornet	16.4	4	120.8	70	3.00	2.200	19.46	0	1	4	1
## Ford Mustang	15.2	8	236.2	140	3.00	3.440	17.00	0	0	4	4
## Fiat 131	17.8	4	90.4	62	3.00	2.200	19.46	0	1	4	1
## Toyota Celica	15.2	4	95.1	62	3.00	2.200	19.46	0	1	4	1
## AMC Pacer	14.3	4	93.7	62	3.00	2.200	19.46	0	1	4	1
## Chevrolet Vega	17.8	4	108.0	70	3.00	2.200	19.46	0	1	4	1
## AMC Gremlin X	14.3	4	108.0	70	3.00	2.200	19.46	0	1	4	1
## Ford Pinto	16.4	4	108.0	70	3.00	2.200	19.46	0	1	4	1
## Toyota Corolla II	17.8	4	108.0	70	3.00	2.200	19.46	0	1	4	1
## AMC Hornet SC	15.2	4	108.0	70	3.00	2.200	19.46	0	1	4	1
## Plymouth Cricket	17.8	4	90.4	62	3.00	2.200	19.46	0	1	4	1
## Fiat 127	17.8	4	90.4	62	3.00	2.200	19.46	0	1	4	1
## Toyota Starlet	15.2	4	95.1	62	3.00	2.200	19.46	0	1	4	1
## AMC Gremlin X	14.3	4	93.7	62	3.00	2.200	19.46	0	1	4	1
## Ford Pinto	16.4	4	93.7	62	3.00	2.200	19.46	0	1	4	1
## Toyota Corolla III	17.8	4	93.7	62	3.00	2.200	19.46	0	1	4	1
## AMC Hornet GL	15.2	4	93.7	62	3.00	2.200	19.46	0	1	4	1
## Fiat 128 GTC	17.8	4	93.7	62	3.00	2.200	19.46	0	1	4	1
## Toyota Starlet II	15.2	4	93.7	62	3.00	2.200	19.46	0	1	4	1
## AMC Hornet GS	14.3	4	93.7	62	3.00	2.200	19.46	0	1	4	1
## Ford Pinto II	16.4	4	93.7	62	3.00	2.200	19.46	0	1	4	1
## Toyota Corolla IV	17.8	4	93.7	62	3.00	2.200	19.46	0	1	4	1
## AMC Hornet GSX	15.2	4	93.7	62	3.00	2.200	19.46	0	1	4	1
## Fiat 128 GTCX	17.8	4	93.7	62	3.00	2.200	19.46	0	1	4	1
## Toyota Starlet III	15.2	4	93.7	62	3.00	2.200	19.46	0	1	4	1
## AMC Hornet SE	14.3	4	93.7	62	3.00	2.200	19.46	0	1	4	1
## Ford Pinto III	16.4	4	93.7	62	3.00	2.200	19.46	0	1	4	1
## Toyota Corolla V	17.8	4	93.7	62	3.00	2.200	19.46	0	1	4	1
## AMC Hornet S	15.2	4	93.7	62	3.00	2.200	19.46	0	1	4	1
## Fiat 128 GTE	17.8	4	93.7	62	3.00	2.200	19.46	0	1	4	1
## Toyota Starlet IV	15.2	4	93.7	62	3.00	2.200	19.46	0	1	4	1
## AMC Hornet GLS	14.3	4	93.7	62	3.00	2.200	19.46	0	1	4	1
## Ford Pinto IV	16.4	4	93.7	62	3.00	2.200	19.46	0	1	4	1
## Toyota Corolla VI	17.8	4	93.7	62	3.00	2.200	19.46	0	1	4	1
## AMC Hornet GSX	15.2	4	93.7	62	3.00	2.200	19.46	0	1	4	1
## Fiat 128 GTE	17.8	4	93.7	62	3.00	2.200	19.46	0	1	4	1
## Toyota Starlet V	15.2	4	93.7	62	3.00	2.200	19.46	0	1	4	1
## AMC Hornet GSX	14.3	4	93.7	62	3.00	2.200	19.46	0	1	4	1
## Ford Pinto V	16.4	4	93.7	62	3.00	2.200	19.46	0	1	4	1
## Toyota Corolla VII	17.8	4	93.7	62	3.00	2.200	19.46	0	1	4	1
## AMC Hornet GSX	15.2	4	93.7	62	3.00	2.200	19.46	0	1	4	1
## Fiat 128 GTE	17.8	4	93.7	62	3.00	2.200	19.46	0	1	4	1
## Toyota Starlet VI	15.2	4	93.7	62	3.00	2.200	19.46	0	1	4	1
## AMC Hornet GSX	14.3	4	93.7	62	3.00	2.200	19.46	0	1	4	1
## Ford Pinto VI	16.4	4	93.7	62	3.00	2.200	19.46	0	1	4	1
## Toyota Corolla VIII	17.8	4	93.7	62	3.00	2.200	19.46	0	1	4	1
## AMC Hornet GSX	15.2	4	93.7	62	3.00	2.200	19.46	0	1	4	1
## Fiat 128 GTE	17.8	4	93.7	62	3.00	2.200	19.46	0	1	4	1
## Toyota Starlet VII	15.2	4	93.7	62	3.00	2.200	19.46	0	1	4	1
## AMC Hornet GSX	14.3	4	93.7	62	3.00	2.200	19.46	0	1	4	1
## Ford Pinto VII	16.4	4	93.7	62	3.00	2.200	19.46	0	1	4	1
## Toyota Corolla IX	17.8	4	93.7	62	3.00	2.200	19.46	0	1	4	1
## AMC Hornet GSX	15.2	4	93.7	62	3.00	2.200	19.46	0	1	4	1
## Fiat 128 GTE	17.8	4	93.7	62	3.00	2.200	19.46	0	1	4	1
## Toyota Starlet VIII	15.2	4	93.7	62	3.00	2.200	19.46	0	1	4	1
## AMC Hornet GSX	14.3	4	93.7	62	3.00	2.200	19.46	0	1	4	1
## Ford Pinto VIII	16.4	4	93.7	62	3.00	2.200	19.46	0	1	4	1
## Toyota Corolla X	17.8	4	93.7	62	3.00	2.200	19.46	0	1	4	1
## AMC Hornet GSX	15.2	4	93.7	62	3.00	2.200	19.46	0	1	4	1
## Fiat 128 GTE	17.8	4	93.7	62	3.00	2.200	19.46	0	1	4	1
## Toyota Starlet IX	15.2	4	93.7	62	3.00	2.200	19.46	0	1	4	1
## AMC Hornet GSX	14.3	4	93.7	62	3.00	2.200	19.46	0	1	4	1
## Ford Pinto IX	16.4	4	93.7	62	3.00	2.200	19.46	0	1	4	1
## Toyota Corolla XI	17.8	4	93.7	62	3.00	2.200	19.46	0	1	4	1
## AMC Hornet GSX	15.2	4	93.7	62	3.00	2.200	19.46	0	1	4	1
## Fiat 128 GTE	17.8	4	93.7	62	3.00	2.200	19.46	0	1	4	1
## Toyota Starlet XII	15.2	4	93.7	62	3.00	2.200	19.46	0	1	4	1
## AMC Hornet GSX	14.3	4	93.7	62	3.00	2.200	19.46	0	1	4	1
## Ford Pinto XII	16.4	4	93.7	62	3.00	2.200	19.46	0	1	4	1
## Toyota Corolla XIII	17.8	4	93.7	62	3.00	2.200	19.46	0	1	4	1
## AMC Hornet GSX	15.2	4	93.7	62	3.00	2.200	19.46	0	1	4	1
## Fiat 128 GTE	17.8	4	93.7	62	3.00	2.200	19.46	0	1	4	1
## Toyota Starlet XIV	15.2	4	93.7	62	3.00	2.200	19.46	0	1	4	1
## AMC Hornet GSX	14.3	4	93.7	62	3.00	2.200	19.46	0	1	4	1
## Ford Pinto XIV	16.4	4	93.7	62	3.00	2.200	19.46	0	1	4	1
## Toyota Corolla XV	17.8	4	93.7	62	3.00	2.200	19.46	0	1	4	1
## AMC Hornet GSX	15.2	4	93.7	62	3.00	2.200	19.46	0	1	4	1
## Fiat 128 GTE	17.8	4	93.7	62	3.00	2.200	19.46	0	1	4	1
## Toyota Starlet XVI	15.2	4	93.7	62	3.00	2.200	19.46	0	1	4	1
## AMC Hornet GSX	14.3	4	93.7	62	3.00	2.200	19.46	0	1	4	1
## Ford Pinto XVI	16.4	4	93.7	62	3.00	2.200	19.46	0	1	4	1
## Toyota Corolla XVII	17.8	4	93.7	62	3.00	2.200	19.46	0	1	4	1
## AMC Hornet GSX	15.2	4	93.7	62	3.00	2.200	19.46	0	1	4	1
## Fiat 128 GTE	17.8	4	93.7	62	3.00	2.200	19.46	0	1	4	1
## Toyota Starlet XVIII	15.2	4	93.7	62	3.00	2.200	19.46	0	1	4	1
## AMC Hornet GSX	14.3	4	93.7	62	3.00	2.200	19.46	0	1	4	1
## Ford Pinto XVIII	16.4	4	93.7	62	3.00	2.200	19.46	0	1	4	1
## Toyota Corolla XIX	17.8	4	93.7	62	3.00	2.200	19.46	0	1	4	1
## AMC Hornet GSX	15.2	4	93.7	62	3.00	2.200	19.46	0	1	4	1
## Fiat 128 GTE	17.8	4	93.7	62	3.00	2.200	19.46	0	1	4	1
## Toyota Starlet XX	15.2	4	93.7	62	3.00	2.200	19.46	0	1	4	1
## AMC Hornet GSX	14.3	4	93.7	62	3.00	2.200	19.46	0	1	4	1
## Ford Pinto XX	16.4	4	93.7	62	3.00	2.200	19.46	0	1	4	1
## Toyota Corolla XXI	17.8	4	93.7	62	3.00	2.200	19.46	0	1	4	1
## AMC Hornet GSX	15.2	4	93.7	62	3.00	2.200	19.46	0	1	4	1
## Fiat 128 GTE	17.8	4	93.7	62	3.00	2.200	19.46	0	1	4	1
## Toyota Starlet XXII	15.2	4	93.7	62	3.00	2.200	19.46	0	1	4	1
## AMC Hornet GSX	14.3	4	93.7	62	3.00	2.200	19.46	0	1	4	1
## Ford Pinto XXII	16.4	4	93.7	62	3.00	2.200	19.46	0	1	4	1
## Toyota Corolla XXIII	17.8	4	93.7	62	3.00	2.200	19.46	0	1	4	1
## AMC Hornet GSX	15.2	4	93.7	62	3.00	2.200	19.46	0	1	4	1
## Fiat 128 GTE	17.8	4	93.7	62	3.00	2.200	19.46	0	1	4	1</td

The Rmd program

RStudio
File Edit Code View Plots Session Build Debug Profile Tools Help
File Edit Code View Plots Session Build Debug Profile Tools Help
Project: (None)

Source Visual Environment History Connections Tutorial
Import Dataset 146 MB
Global Environment
Environment is empty

152
153 `## Miles/(us) gallon vs. the car's weight`
154 `## Our aim is to investigate the relationship between the fuel consumption (in Miles/(us) gallon, the R object`
155 `<tt>mpg</tt>) and the car's weight (in (1000 lbs), the R object <tt>wt</tt>).`
156 `##`
157
158 `### Miles/(us) gallon vs. the car's weight`
159 `##`
160 `## Our aim is to investigate the relationship between the fuel consumption (in Miles/(us) gallon, the R object`
161 `<tt>mpg</tt>) and the car's weight (in (1000 lbs), the R object <tt>wt</tt>).`
162 `##`
163 `##`
164 `##`
165 `##`
166 `##`
167 `## The relationship between the car's weight and mpg, shown in Figure 1, appears linear. A line might provide a useful`
168 `## summary of this association. Pearson correlation is equal to -0.867, indicates, on a negative association.`
169 `##`
170 `##`
171 `##`
172 `##`
173
174 `# 2. The least squares regression model`
175
176
177 `## Model assumptions`
178
179 `## There are 4 assumptions that should be satisfied for a line to be considered a reasonable approximation for a`
180 `## relationship shown in a scatterplot.`
181 `## 1. Linearity: the data show a linear trend.`
182 `## 2. Constant variability: the variability of the response variable about the line remains roughly constant as the`
159:1 Miles/(US) gallon vs. the car's Weight

Console Terminal Background Jobs
R 4.3.2 · C:\projects\Teaching2021\Rcourse\R_2223\Classes/The pharma study
R is free software and comes with ABSOLUTELY NO WARRANTY.
You are welcome to redistribute it under certain conditions.
Type 'license()' or 'licence()' for distribution details.
R is a collaborative project with many contributors.
Type 'contributors()' for more information and
'citation()' on how to cite R or R packages in publications.
Type 'demo()' for some demos, 'help()' for on-line help, or
'help.start()' for an HTML browser interface to help.
Type 'q()' to quit R.

Type here to search

- Analysis in R:
 - Scatterplot + correlation.

The HTML output

C:/projects/Teaching2021/Rcourse/R_2223/Classes/The pharma study/er_prog4_SA_2024.html

er_prog4_SA_2024.html | Open in Browser | Find

- □ X
Publish

1. General Introduction

Linear regression models

Examining scatterplots

2. The least squares regression model

3. Simple linear regression using R

3. Model diagnostic

4. Categorical predictors with two levels

5. Goodness of fit: using R^2 to describe the strength of a fit

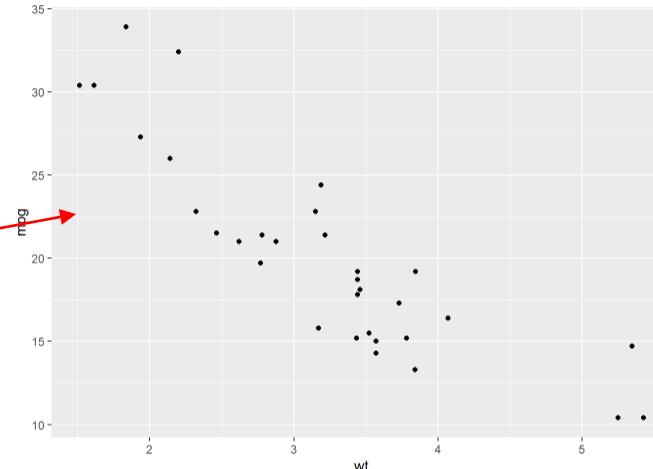
6. Statistical inference in regression

```
## Datsun 710      22.8   4 108 93 3.85 2.320 18.61 1 1 4 1
## Hornet 4 Drive 21.4   6 258 110 3.08 3.215 19.44 1 0 3 1
## Hornet Sportabout 18.7   8 360 175 3.15 3.440 17.02 0 0 3 2
## Valiant        18.1   6 225 105 2.76 3.460 20.22 1 0 3 1
```

Miles/(US) gallon vs. the car's Weight

Our aim is to investigate the relationship between the fuel consumption (in Miles/(US) gallon, the R object `mpg`) and the car's weight (in 1000 lbs), the R object `wt`.

Show



- Analysis in R:
 - Scatterplot.
 - Correlation.

- The example:
 - `mtcars` data.
 - Not in Dave & Julie files.

Show

Text without the code.

Figure 1: mpg vs. weight

The relationship between the car's weight and mpg, shown in Figure 1, appears linear. A line might provide a useful summary of this association. Pearson correlation is equal to -0.867, indicates, on a negative association.

```
## [1] -0.8676594
```

2. The least squares regression model

Model assumptions

There are 4 assumptions that should be satisfied for a line to be considered a reasonable approximation for a relationship shown in a



7:26 ENG 8/02/2024

The HTML output

R C:/projects/Teaching2021/Rcourse/R_2223/Classes/The pharma study/er_prog4_SA_2024.html
er_prog4_SA_2024.html | Open in Browser | Find | Publish | @

1. General Introduction
Linear regression models
Examining scatterplots
2. The least squares regression model
3. Simple linear regression using R
3. Model diagnostic
4. Categorical predictors with two levels
5. Goodness of fit: using R^2 to describe the strength of a fit
6. Statistical inference in regression

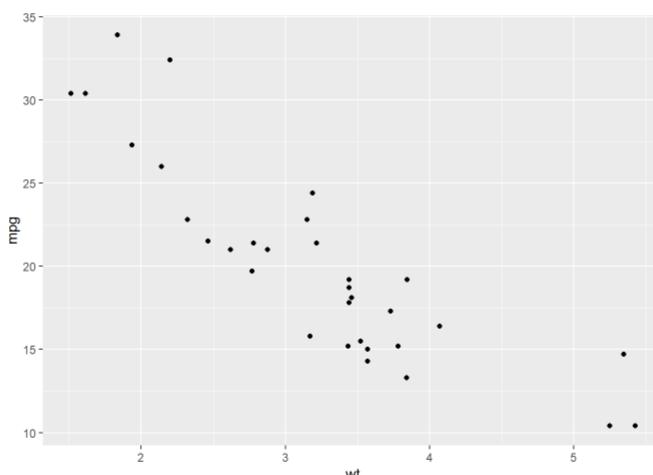
```
## Datsun 710      22.8   4 108 93 3.85 2.320 18.61  1  1   4   1
## Hornet 4 Drive 21.4   6 258 110 3.08 3.215 19.44  1  0   3   1
## Hornet Sportabout 18.7  8 360 175 3.15 3.440 17.02  0  0   3   2
## Valiant       18.1   6 225 105 2.76 3.460 20.22  1  0   3   1
```

Miles/(US) gallon vs. the car's Weight

Our aim is to investigate the relationship between the fuel consumption (in Miles/(US) gallon, the R object `mpg`) and the car's weight (in 1000 lbs, the R object `wt`).

[Hide]

```
#plot(mtcars$wt,mtcars$mpg, ylab = "mpg", xlab = "weight (1000 lbs)")
qplot(wt,mpg,data = mtcars)
```



[Show]

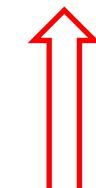
Figure 1: mpg vs. weight

The relationship between the car's weight and mpg, shown in Figure 1, appears linear. A line might provide a useful summary of this association. Pearson correlation is equal to -0.867, indicates, on a negative association.

[Show]

```
## [1] -0.8676594
```

Text with the code.



2. The least squares regression model

The Rmd program

RStudio
File Edit Code View Plots Session Build Debug Profile Tools Help
unit_05_inference_num_1ANOVA.Rmd | er_prog3_SA_2024.Rmd | er_prog4_SA_2024.Rmd | The pharma challenge_2022_prog1.Rmd | er_prog2_SA_2024.Rmd
Source Visual Knit Addins
219 $\$s_x$, $\$s_y$: sample standard deviations of $\$x$ and $\$y$.
220
221 $\$r$: correlation between $\$x$ and $\$y$.
222
223
224 # 3. simple linear regression using R
225
226 ## The $\text{lm}()$ R function
227
228 For the mtcars dataset, we consider the model
229
230 $\$mpg_i = \beta_0 + \beta_1 \times weight_i + \epsilon_i$.
231
232 In the above model, the variable mpg is the response and weight is the predictor. In R, we can fit
the simple linear regression model using the R function lm . The function has the general call of
 $\text{lm}(y \sim x)$. The output for the mtcars data is shown below.
233
234 ````{r, echo=TRUE, message=FALSE, warning=FALSE}
235 fit.lm<-lm(mtcars\$mpg~mtcars\$wt)
236 summary(fit.lm)
237 ````
238
239 The parameter estimates for the intercept and slope are equal, respectively, to $\hat{\beta}_0 = 37.28$ and
 $\hat{\beta}_1 = -5.34$
240
241 ## Data and estimated model
242
243 Figure 2 shows the data (mpg vs. weight) and fitted regression line, $\hat{y} = 37.28 - 5.34 \times wt$
244
245 ````{r, echo=TRUE, message=FALSE, warning=FALSE, fig.cap="Data and fitted model"}
246 qplot(wt, mpg, data = mtcars)+
247 geom_smooth(method = "lm", se = F)
248 ````
249
250 ## Parameter estimates
251
252 Parameter estimates for $\$y$ and $\$x$ are given by
253
159:1 Miles/(US) gallon vs. the car's Weight
R Markdown
Console Terminal Render Background Jobs
R 4.3.2 · C:\Projects\Teaching2021\Rcourse\R_2223\Classes\The pharma study
R is free software and comes with ABSOLUTELY NO WARRANTY.
You are welcome to redistribute it under certain conditions.
Type 'license()' or 'licence()' for distribution details.
R is a collaborative project with many contributors.
Type 'contributors()' for more information and
'citation()' on how to cite R or R packages in publications.
Type 'demo()' for some demos, 'help()' for on-line help, or
'help.start()' for an HTML browser interface to help.
Type 'q()' to quit R.
Type here to search
Project: (None)
Environment History Connections Tutorial
Import Dataset 156 MB Global Environment
Environment is empty
Free text about regression.
The regression model.

The HTML output

R C:/projects/Teaching2021/Rcourse/R_2223/Classes/The pharma study/er_prog4_SA_2024.html
er_prog4_SA_2024.html | Open in Browser | Find | Publish |

3. Simple linear regression using R

The lm() R function

For the `mtcars` dataset, we consider the model

$$mpg_i = \beta_0 + \beta_1 \times weight_i + \epsilon_i$$

In the above model, the variable `mpg` is the response and `weight` is the predictor. In R, we can fit the simple linear regression model using the R function `lm`. The function has the general call of `lm(y~x)`. The output for the `mtcars` data is shown below.

```
fit.lm<-lm(mtcars$mpg~mtcars$wt)
summary(fit.lm)
```

```
## 
## Call:
## lm(formula = mtcars$mpg ~ mtcars$wt)
## 
## Residuals:
##    Min     1Q   Median     3Q    Max 
## -4.5432 -2.3647 -0.1252  1.4096  6.8727 
## 
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)    
## (Intercept) 37.2851   1.8776 19.858 < 2e-16 ***
## mtcars$wt   -5.3445   0.5591 -9.559 1.29e-10 ***
## ---      
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
## 
## Residual standard error: 3.046 on 30 degrees of freedom
## Multiple R-squared:  0.7528, Adjusted R-squared:  0.7446 
## F-statistic: 91.38 on 1 and 30 DF,  p-value: 1.294e-10
```

The parametr estimates for the intercept and slope are equal, respectively, to $\hat{\beta}_0 = 37.28$ and $\hat{\beta}_1 = -5.34$

Data and estimated model

Figure 2 shows the data (mpg vs. weight) and fitted regression line, $mpg_i = 37.28 - 5.34 \times wt_i$

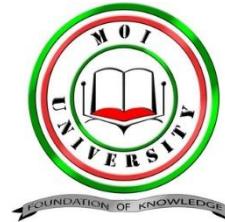
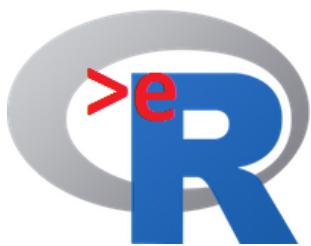


Free text about regression.

- The regression model:
 - Code.
 - Output.

Short summary

- Use Rmd to create the document.
- Upload the document online as part of the course.
- Text: written by Dave & Jullie (and available online).
- Examples written by Ziv (and available online).



Interuniversity Institute for Biostatistics
and statistical Bioinformatics

Development of E-learning materials using R markdown

The >eR-BioStat approach

How can we create a course online??

The >eR-BioStat approach

- Content development:
 - R/R markdown.
- Storage:
 - Github.
- Website development:
 - WIX.



All have publicly available
and free versions.

From a laptop to a website..

The screenshot shows a Windows desktop with a browser window open to a local file. The page title is "3. Simple linear regression using R". It contains R code for fitting a linear model to the "mtcars" dataset:

```
## correlation between x and y
## correlation between x and y

## The lm() R function

For the mtcars dataset, we consider the model


$$mpg = \beta_0 + \beta_1 \times weight + \epsilon_1$$


In the above model, the variable mpg is the response and weight is the predictor. In R, we can fit the simple linear regression model using the R function lm. The function has the general call of lm(y~x). The output for the mtcars data is shown below.
```

```
fit.lm<-lm(mpg~weight, data=mtcars)
summary(fit.lm)

## Call:
## lm(formula = mpg ~ weight, data = mtcars)
##
## Residuals:
##   Min   1Q   Median   3Q   Max
## -9.983 -2.3647 -0.1252 14.934 4.1871
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept) 37.881 1.8776 19.955 < 2e-16 ***
## weight      -0.4493  0.0991 -4.5284 0.000122 ***
## ---
## Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.1 '.' 1
##
## Residual standard error: 3.046 on 30 degrees of freedom
## Multiple R-squared: 0.7520, Adjusted R-squared: 0.7464
## F-statistic: 21.09 on 1 and 30 DF, p-value: 1.299e-12
```

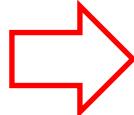
The parameter estimates for the intercept and slope are equal, respectively, to $\hat{\beta}_0 = 37.88$ and $\hat{\beta}_1 = -0.4493$.

Data and estimated model

Figure 2 shows the data (mpg vs. weight) and fitted regression line. $mpg = 37.88 - 0.4493 \times weight$

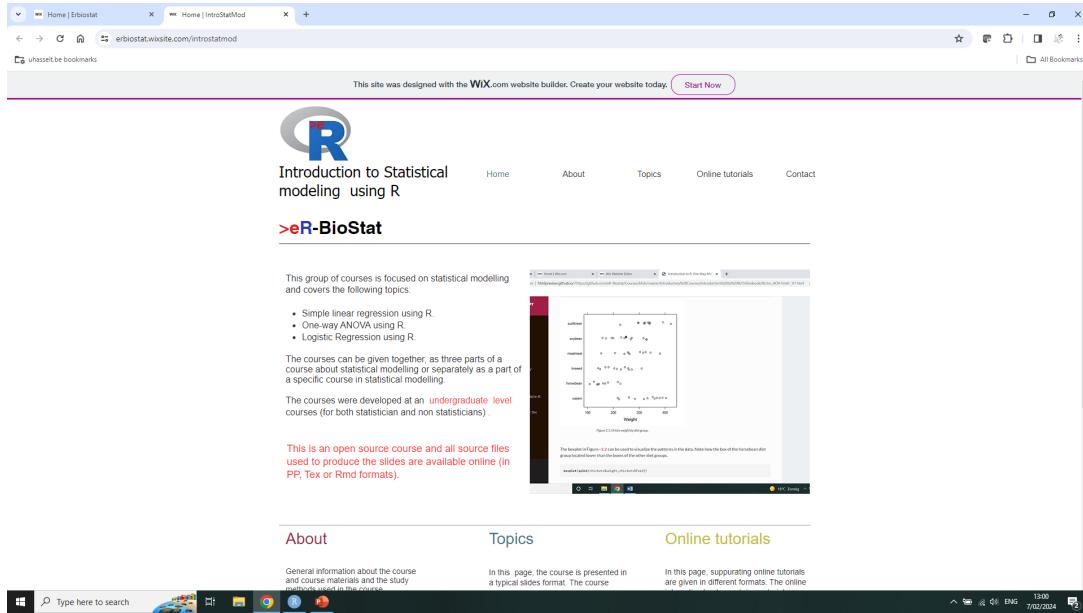
The screenshot shows a Windows desktop with a browser window open to a GitHub page. The URL is "https://github.com/ab-Bisrat/Courses/blob/master/Introductory%20to%20Statistical%20Modeling%20using%20R/Simple%20Linear%20Regression/slides/Temp_2024_prog1_V2.html". The page content is identical to the one on the laptop, showing the R code and its output for simple linear regression on the mtcars dataset.

The HTML file on the laptop....
Can be produced using the Rmd
program: er_prog4_SA_2024.Rmd.



The HTML file as a part of the course's website.

Introduction to statistical modeling using R



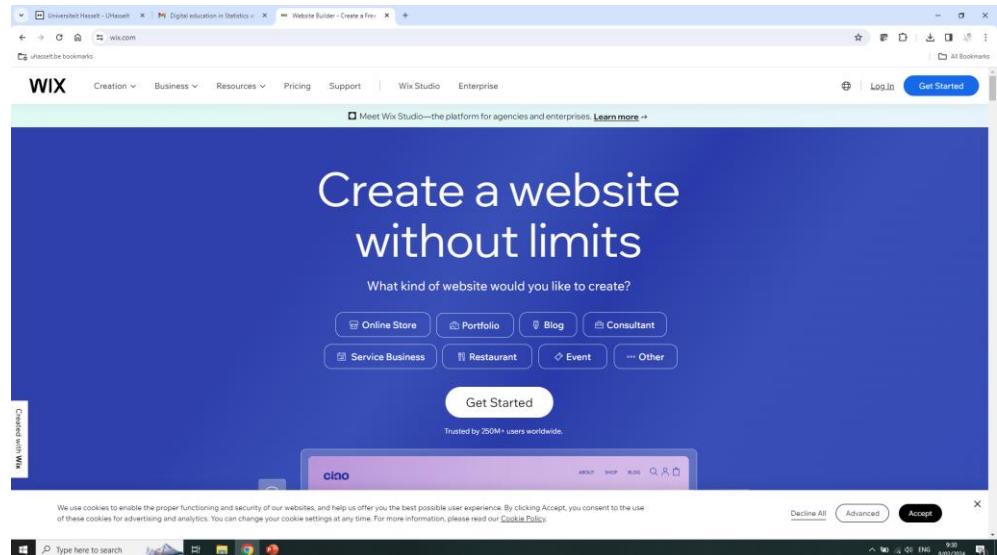
- Website.
- Storage of materials:
 - Slides.
 - Programs.
 - Datasets.
 - Etc.
- Software for the analysis ?
- Storage space & cost ?

<https://erbiostat.wixsite.com/introstatmod>

- Our approach: bring costs to zero...

How can I develop a website for my course ?

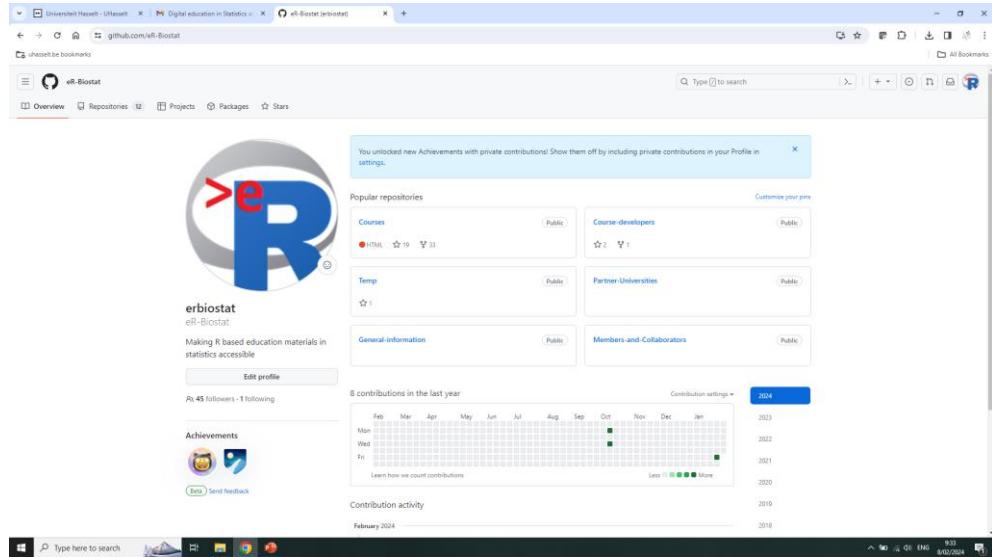
- Websites for the courses were developed using WIX.
 - Free.
 - Easy to use and learn.



<https://www.wix.com/>

Where can I store the course materials (slides, programs, notes....?)

- All course materials are stored on Github:
 - Slides.
 - Programs.
 - Datasets.
 - R markdown programs.
 - **HTML files....**
- Free and unlimited.



<https://github.com/eR-Biostat>

Universiteit Hasselt - UHasselt X | Digital education in Statistics ar X | Courses/Introductory Courses/ +

github.com/eR-Biostat/Courses/tree/master/Introductory%20Courses/Introduction%20to%20statistical%20modeling%20using%20R

uhasselt.be bookmarks

Files

master

Go to file

Basic courses
Coordination
Data Analysis
ICP Workshop
Inference
Introductory Courses
Introduction to R
Introduction to statistical mode...
Logistic regression
One way ANOVA
Online materials
Simple linear regression
README.md
Visualizing data using R- an intr...
README.md
Modeling Infectious diseases
Statistical modeling (1)
Statistical modeling (2)
.gitignore
Courses.Rproj
README.md
Systematic Review and Meta An...
Systematic Review and Meta.docx
_config.yml

Name	Last commit message	Last commit date
..		
Logistic regression	Add files via upload	2 years ago
One way ANOVA	Add files via upload	2 years ago
Online materials	Add files via upload	2 years ago
Simple linear regression	Add files via upload	last week
README.md	Update README.md	7 years ago

README.md

The >eR-Biostat initiative

Introduction to statistical modeling in R

This course is an introductory course about statistical modeling in R. The course can be given as a two-days workshop or as a course of 5 classes (3 hours per class). Topics (all presented at a basic level) covered in the course include:

- Simple linear regression (<https://github.com/eR-Biostat/Courses/tree/master/Introductory%20Courses/Introduction%20to%20statistical%20modeling%20using%20R/Simple%20linear%20regression>).
- One-way ANOVA (<https://github.com/eR-Biostat/Courses/tree/master/Introductory%20Courses/Introduction%20to%20statistical%20modeling%20using%20R/One%20way%20ANOVA>).
- Two-way ANOVA.
- Logistic regression (<https://github.com/eR-Biostat/Courses/tree/master/Introductory%20Courses/Introduction%20to%20statistical%20modeling%20using%20R/Logistic%20regression>).

Homework assignments and an example of an exam are NOT available as a part of the course but practical sessions are included as a part of each topic.

Type here to search

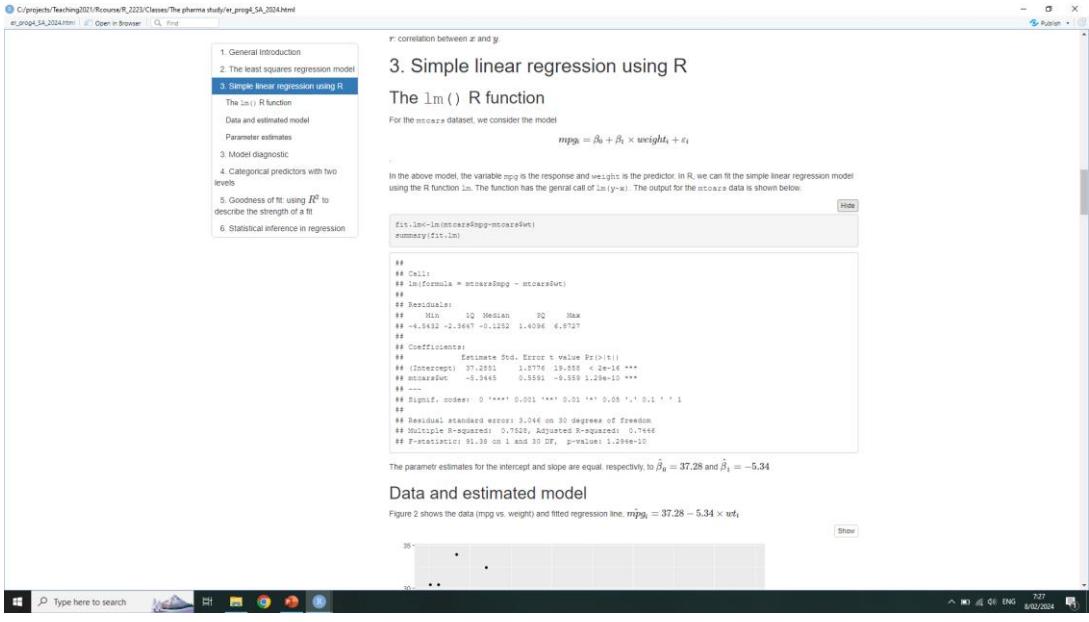
9:36 8/02/2024 ENG

<https://github.com/eR-Biostat/Courses/tree/master/Introductory%20Courses>

Software for the analysis

- We use R but.....

Summary



The screenshot shows a Windows desktop with a browser window open. The address bar indicates the URL is `C:/projects/Teaching2021/Rcourse/R_2223/Classes/The pharma study/r_progs_SA_2024.html`. The page content is an R Markdown document titled "Simple linear regression using R". It includes a sidebar with navigation links and a main area with R code and its output.

R: correlation between x_i and y_i

3. Simple linear regression using R

The `lm()` R function

For the `mtcars` dataset, we consider the model

$$\text{mpg}_i = \beta_0 + \beta_1 \times \text{weight}_i + \varepsilon_i$$

In the above model, the variable `mpg` is the response and `weight` is the predictor. In R, we can fit the simple linear regression model using the R function `lm`. The function has the general call of `lm(y~x)`. The output for the `mtcars` data is shown below:

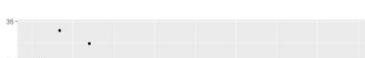
```
fit.lm<-lm(mtcars$mpg ~ mtcars$wt)
summary(fit.lm)

##
## Call:
## lm(formula = mtcars$mpg ~ mtcars$wt)
## ...
## Residuals:
##   Min   1Q Median   3Q Max
## -4.5432 -2.3647 -0.1252  1.4098  6.8727
## ...
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept) 37.8788  1.8778 19.888 4.2e-16 ***
## mtcars$wt -5.3448  0.5833 -8.933 1.29e-15 ***
## ...
## Signif. codes: 0 '****' 0.001 '**' 0.05 '*' 0.1 ' ' 1
## Residual standard error: 3.046 on 32 degrees of freedom
## Multiple R-squared: 0.7928, Adjusted R-squared: 0.7446
## F-statistic: 91.38 on 1 and 30 DF, p-value: 1.29e-10
```

The parameter estimates for the intercept and slope are equal, respectively, to $\hat{\beta}_0 = 37.28$ and $\hat{\beta}_1 = -5.34$.

Data and estimated model

Figure 2 shows the data (`mpg` vs. `weight`) and fitted regression line: $\text{mpg}_i = 37.28 - 5.34 \times \text{wt}_i$



- Content produced using R markdown.
- Store online on GitHub.
- Presented online in a website developed using WIX.
- Data analysis using R.
- Costs=0 !!!

What is available to the users ?

- Who are the users ? Teachers & students & others.
- What is available ? Everything.
- An open source approach:
 - Slides.
 - R programs for examples.
 - R programs for the slides.
 - PowerPoints files.
 - HTML files.

Users

21/02/23-20/02/24

Traffic Overview | Wix.com

manage.wix.com/dashboard/e53545a0-4f7d-4f89-92b7-6ebf2f6764cc/analytics/overviews/traffic?referralInfo=sidebar

uhasselt.be bookmarks All Bookmarks

WIX IntroStatMod Explore Help Hire a Professional Upgrade Search for tools, apps, help & more... Notifications (1) Subscriptions (3)

Let's set up your business > 1/4 completed

Setup Home Site & App Subscriptions Contacts Communications (2) Automations Marketing & SEO Analytics & Reports Traffic Overview Real-time Sales Overview Marketing Overview Behavior Overview Reports Insights Benchmarks Site Speed Uptime & Security Alerts Email Updates Billing & Payments > Quick Access

Type here to search

Traffic Overview

Last 365 days (Feb 21, 2023 - Today) compared to previous period (Feb 21, 2022 - Feb 20, 2023)

Site sessions 109 ↓ 80% Unique visitors 69 ↓ 83%

Sessions over time

Sessions by traffic source

Traffic Source	Sessions	Change (%)	Actions
Direct	94	-57%	Get traffic
Facebook	15	-95%	Get traffic
Unknown	0	-100%	Get traffic
Google.com	Get traffic		
Wix email marketing	Get traffic		

See Full Report

New vs returning visitors

Unique visitors 69

Sessions by device

Site sessions 109

Avg. sessions by day

Sun Mon Tue Wed Thu Fri Sat

See Full Report

19:55 ENG 20/02/2024

Users

21/02/23-20/02/24

Traffic Overview | Wix.com

manage.wix.com/dashboard/e53545a0-4f7d-4f89-92b7-6ebf2f6764cc/analytics/overviews/traffic?referralInfo=sidebar

uhasselt.be bookmarks All Bookmarks

WIX IntroStatMod Explore Help Hire a Professional Upgrade Search for tools, apps, help & more... Notifications (1) Subscriptions (1)

Let's set up your business > 1/4 completed

Setup Home Site & App Subscriptions Contacts Communications (2) Automations Marketing & SEO Analytics & Reports Traffic Overview Real-time Sales Overview Marketing Overview Behavior Overview Reports Insights Benchmarks Site Speed Uptime & Security Alerts Email Updates Billing & Payments Quick Access Start Type here to search

Traffic Overview Last 365 days (Feb 21, 2023 - Today) compared to previous period (Feb 21, 2022 - Feb 20, 2023)

Unique visitors 69 97% * 67 Returning 3% * 2 See Full Report

Site sessions 109 94% * 102 Mobile 6% * 7 See Full Report

See Full Report See Full Report

Traffic insights See All Insights

Sessions by country Countries South Africa > 22 Kenya > 21 Ethiopia > 15 Belgium > 14 Serbia > 12 Ghana > 8 Kazakhstan > 4

1 2 3 >

19:57 20/02/2024 ENG

The screenshot shows the Wix Analytics dashboard for a site named 'uhasselt.be'. The main header includes the title 'Traffic Overview | Wix.com' and the URL 'manage.wix.com/dashboard/e53545a0-4f7d-4f89-92b7-6ebf2f6764cc/analytics/overviews/traffic?referralInfo=sidebar'. Below the header is a navigation bar with links for 'Explore', 'Help', 'Hire a Professional', 'Upgrade', and a search bar. A sidebar on the left lists various analytics categories like 'Traffic Overview', 'Real-time', 'Sales Overview', etc., with 'Traffic Overview' currently selected. The main content area displays two circular dashboards: one for 'Unique visitors' (69 total, 97% new, 3% returning) and one for 'Site sessions' (109 total, 94% mobile, 6% desktop). Below these are sections for 'Sessions by country' (a world map showing session distribution) and 'Traffic insights' (a list of countries with the highest session counts: South Africa, Kenya, Ethiopia, Belgium, Serbia, Ghana, Kazakhstan). The bottom of the screen shows the Windows taskbar with icons for Start, Search, Task View, File Explorer, Google Chrome, and Powerpoint, along with system status icons for battery, signal, and volume.

Discussion

- R Studio + R markdown:
- Easy to use.
- Text + code.
- Output:
 - Standard: HTML, PDF, DOC.
 - Advanced: HTML.