

# Development of E-learning materials using R & R markdown

Ziv Shkedy (Hasselt University)

International symposium on current trends in modeling and software development in data science and Statistics

Cape Town, 20/02/24-23/02/24



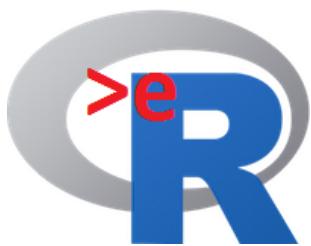
ER-BioStat

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

@erbiostat

# Overview

- Starting point: Rudradev's session,
  - 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...



Interuniversity Institute for Biostatistics  
and statistical Bioinformatics

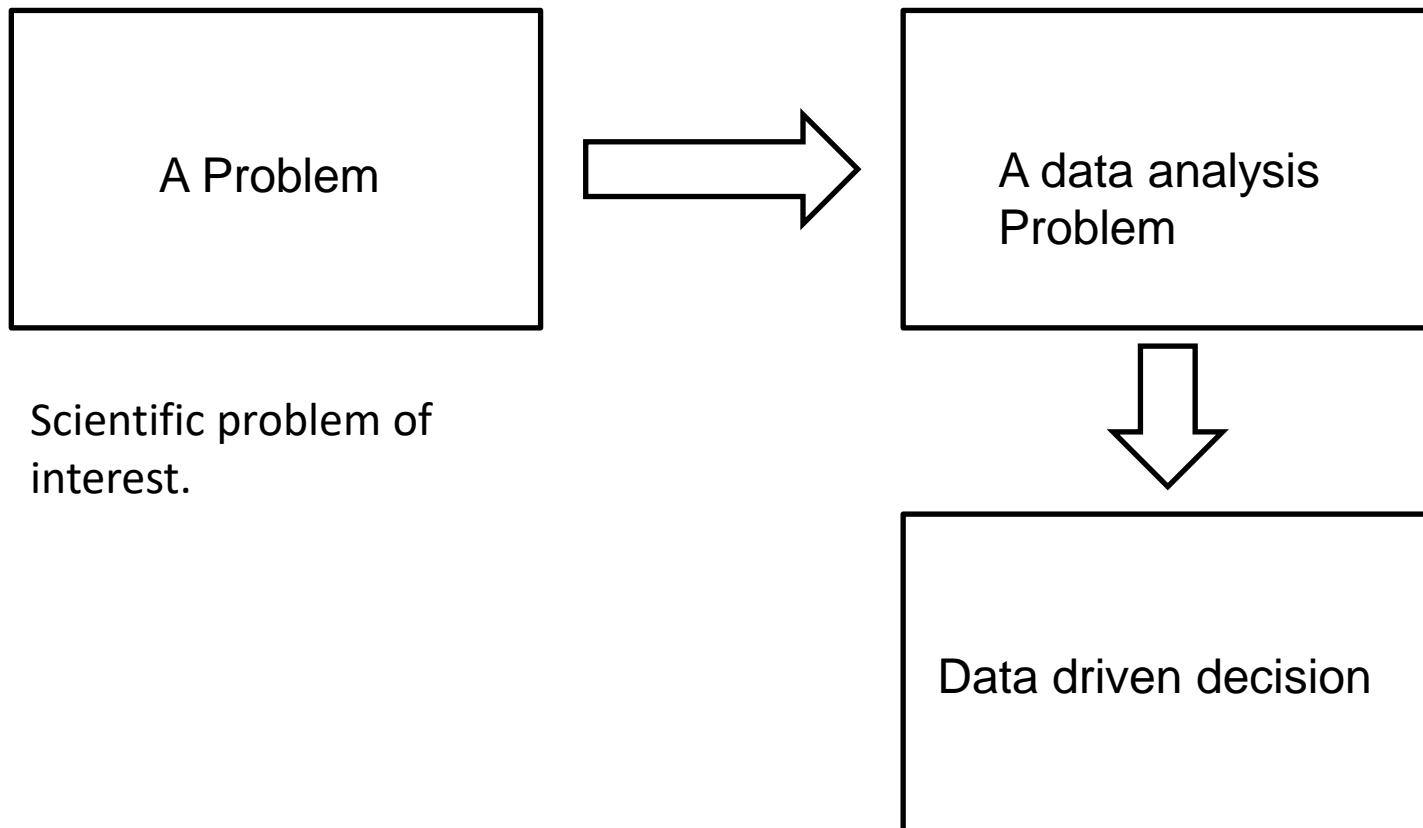


## Short summary of Rudradev's session

Linear regression using R markdown

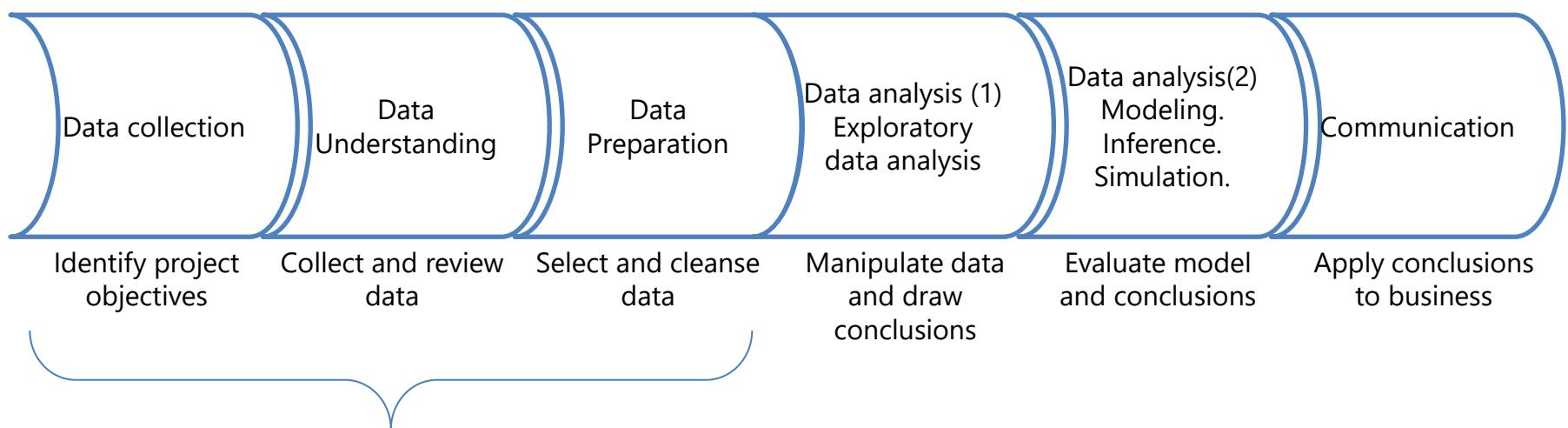
# Steps in data analysis

- Data analysis approach in the course:



# Steps in data analysis

- Steps related to data analysis:



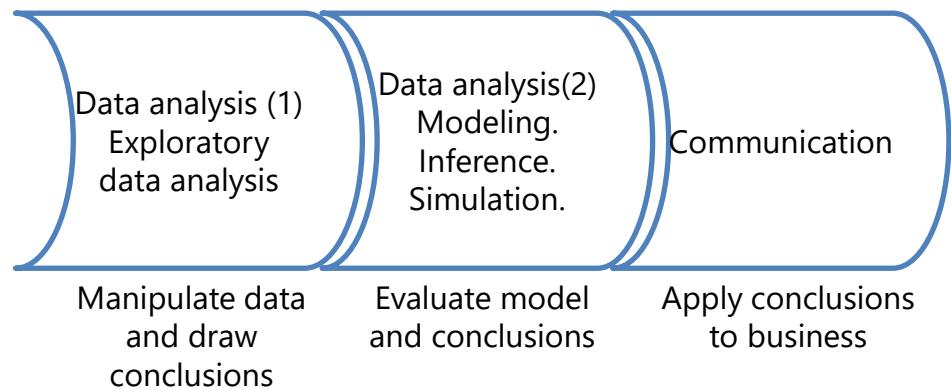
Not a part of our course

# 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 ?

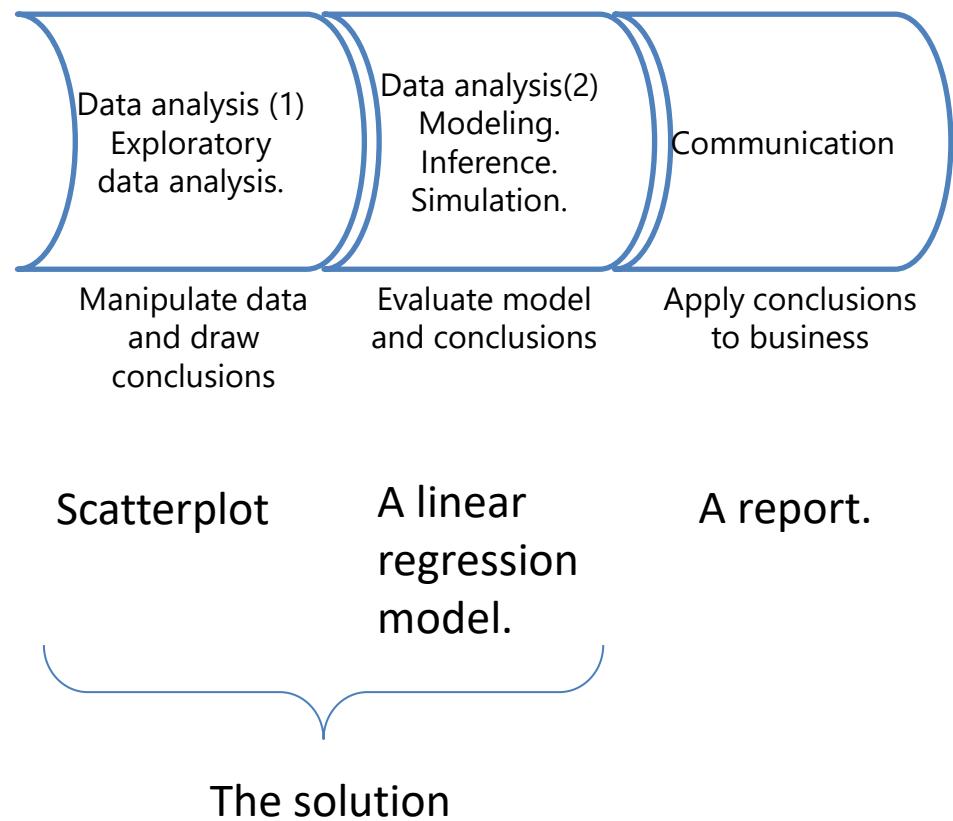


# Steps in data analysis

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

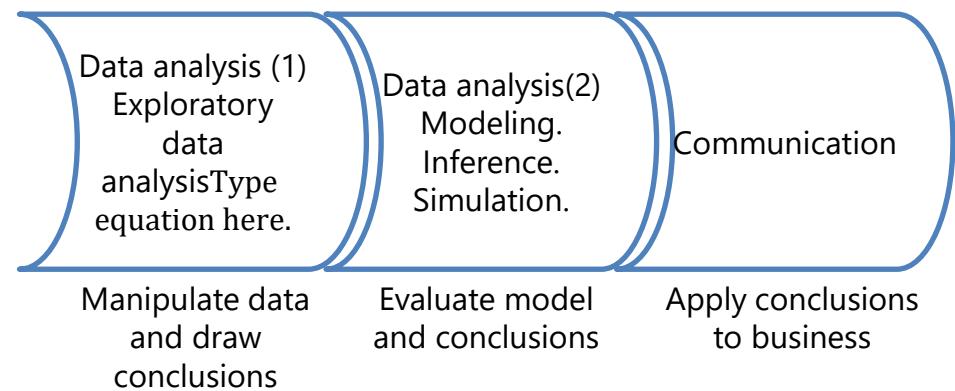
$$y_i = \beta_0 + \beta_1 x_i + \varepsilon_i$$

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.

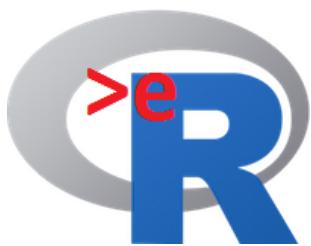
A report.

`ggplot2()`

`lm()`

R markdown to produce a HTML file.

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



KU LEUVEN

Interuniversity Institute for Biostatistics  
and statistical Bioinformatics



# 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"  
[6] "wt"    "qsec"  "vs"   "am"    "gear"  
[11] "carb"
```

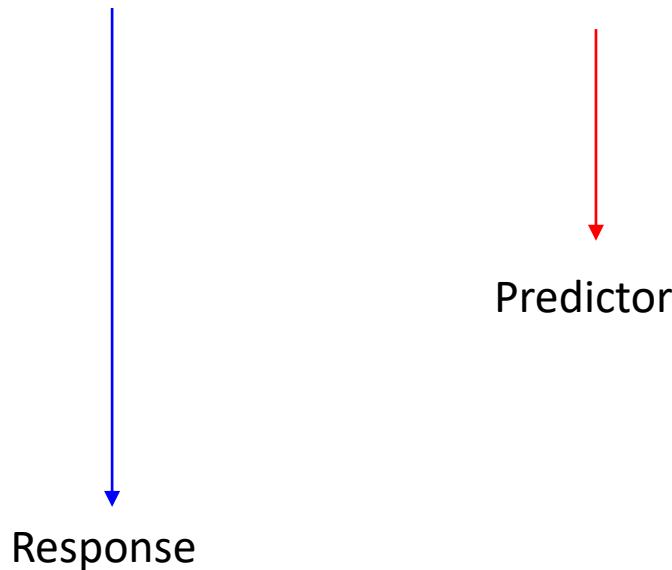
A blue arrow points from the first line of the R code (`dim(mtcars)`) to the text "The R object for the data: 32 observations and 11 variables.". Another blue arrow points from the second line of the R code (`names(mtcars)`) to the text "Variables names."

# 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

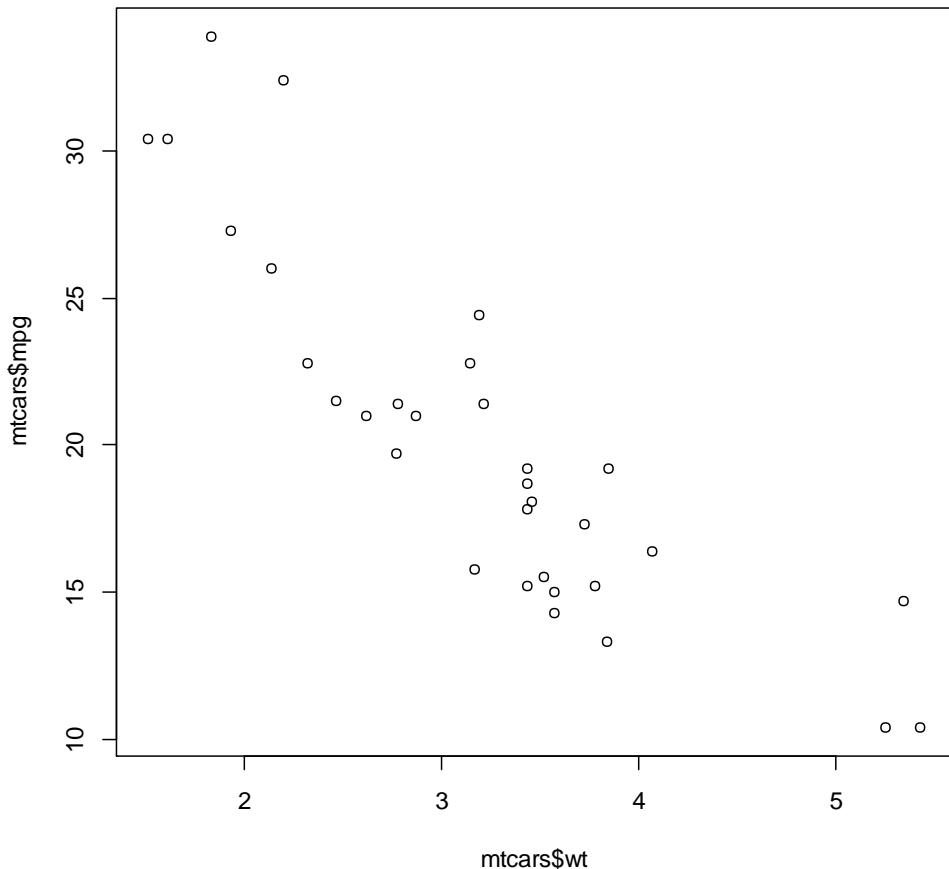
```
>
```



# The mtcars data in R

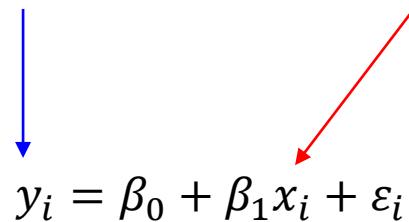
```
> plot(mtcars$wt, mtcars$mpg)  
> cor(mtcars$wt, mtcars$mpg)  
[1] -0.8676594
```

- 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



Interuniversity Institute for Biostatistics  
and statistical Bioinformatics



# 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 library(tidyverse)  
39 library(tidyverse)  
40 library(tidyverse)  
41 library(tidyverse)  
42 library(tidyverse)  
43 library(tidyverse)  
44 library(tidyverse)  
45 library(tidyverse)  
46 library(tidyverse)  
47 library(tidyverse)  
48 library(tidyverse)  
49 library(tidyverse)  
50 library(tidyverse)  
51 library(tidyverse)  
52 library(tidyverse)  
53 library(tidyverse)  
54 library(tidyverse)  
55 library(tidyverse)  
56 library(tidyverse)  
57 library(tidyverse)  
58 library(tidyverse)  
59 library(tidyverse)  
60 library(tidyverse)  
61 library(tidyverse)  
62 library(tidyverse)  
63 library(tidyverse)  
64 library(tidyverse)  
65 library(tidyverse)  
66 library(tidyverse)  
67 library(tidyverse)  
68 library(tidyverse)  
69 library(tidyverse)  
70 library(tidyverse)  
71 library(tidyverse)  
72 library(tidyverse)  
73 library(tidyverse)  
74 library(tidyverse)  
75 library(tidyverse)  
76 library(tidyverse)  
77 library(tidyverse)  
78 library(tidyverse)  
79 library(tidyverse)  
80 library(tidyverse)  
81 library(tidyverse)  
82 library(tidyverse)  
83 library(tidyverse)  
84 library(tidyverse)  
85 library(tidyverse)  
86 library(tidyverse)  
87 library(tidyverse)  
88 library(tidyverse)  
89 library(tidyverse)  
90 library(tidyverse)  
91 library(tidyverse)  
92 library(tidyverse)  
93 library(tidyverse)  
94 library(tidyverse)  
95 library(tidyverse)  
96 library(tidyverse)  
97 library(tidyverse)  
98 library(tidyverse)  
99 library(tidyverse)  
100 library(tidyverse)  
101 library(tidyverse)  
102 library(tidyverse)  
103 library(tidyverse)  
104 library(tidyverse)  
105 library(tidyverse)  
106 library(tidyverse)  
107 library(tidyverse)  
108 library(tidyverse)  
109 library(tidyverse)  
110 library(tidyverse)  
111 library(tidyverse)  
112 library(tidyverse)  
113 library(tidyverse)  
114 library(tidyverse)  
115 library(tidyverse)  
116 library(tidyverse)  
117 library(tidyverse)  
118 library(tidyverse)  
119 library(tidyverse)  
120 library(tidyverse)  
121 library(tidyverse)  
122 library(tidyverse)  
123 library(tidyverse)  
124 library(tidyverse)  
125 library(tidyverse)  
126 library(tidyverse)  
127 library(tidyverse)  
128 library(tidyverse)  
129 library(tidyverse)  
130 library(tidyverse)  
131 library(tidyverse)  
132 library(tidyverse)  
133 library(tidyverse)  
134 library(tidyverse)  
135 library(tidyverse)  
136 library(tidyverse)  
137 library(tidyverse)  
138 library(tidyverse)  
139 library(tidyverse)  
140 library(tidyverse)  
141 library(tidyverse)  
142 library(tidyverse)  
143 library(tidyverse)  
144 library(tidyverse)  
145 library(tidyverse)  
146 library(tidyverse)  
147 library(tidyverse)  
148 library(tidyverse)  
149 library(tidyverse)  
150 library(tidyverse)  
151 library(tidyverse)  
152 library(tidyverse)  
153 library(tidyverse)  
154 library(tidyverse)  
155 library(tidyverse)  
156 library(tidyverse)  
157 library(tidyverse)  
158 library(tidyverse)  
159 library(tidyverse)  
160 library(tidyverse)  
161 library(tidyverse)  
162 library(tidyverse)  
163 library(tidyverse)  
164 library(tidyverse)  
165 library(tidyverse)  
166 library(tidyverse)  
167 library(tidyverse)  
168 library(tidyverse)  
169 library(tidyverse)  
170 library(tidyverse)  
171 library(tidyverse)  
172 library(tidyverse)  
173 library(tidyverse)  
174 library(tidyverse)  
175 library(tidyverse)  
176 library(tidyverse)  
177 library(tidyverse)  
178 library(tidyverse)  
179 library(tidyverse)  
180 library(tidyverse)  
181 library(tidyverse)  
182 library(tidyverse)  
183 library(tidyverse)  
184 library(tidyverse)  
185 library(tidyverse)  
186 library(tidyverse)  
187 library(tidyverse)  
188 library(tidyverse)  
189 library(tidyverse)  
190 library(tidyverse)  
191 library(tidyverse)  
192 library(tidyverse)  
193 library(tidyverse)  
194 library(tidyverse)  
195 library(tidyverse)  
196 library(tidyverse)  
197 library(tidyverse)  
198 library(tidyverse)  
199 library(tidyverse)  
200 library(tidyverse)  
201 library(tidyverse)  
202 library(tidyverse)  
203 library(tidyverse)  
204 library(tidyverse)  
205 library(tidyverse)  
206 library(tidyverse)  
207 library(tidyverse)  
208 library(tidyverse)  
209 library(tidyverse)  
210 library(tidyverse)  
211 library(tidyverse)  
212 library(tidyverse)  
213 library(tidyverse)  
214 library(tidyverse)
```

Annotations with curly braces point to specific sections of the code:

- A brace on the left side of the code block points to the document setup section (lines 1-10).
- A brace on the right side of the code block points to the many packages listed (lines 11-214).

The Environment pane shows "Environment is empty". The Console pane shows the R startup message and the command prompt >|.

# 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, er\_prog4\_SA\_2024.Rmd.
- Source Editor:** Displays the Rmd code. Annotations with red arrows point to specific lines:
  - An arrow points to line 51: `# Baseline analysis` with the text "Dimension of the data, variables names and first 6 lines of the data".
  - An arrow points to line 63: `head(mtcars)` with the same text.
- Plot & correlation:** A large text box containing the code for creating a scatterplot and calculating correlation.
- The regression model:** A large text box containing the code for fitting a linear regression model.
- Environment:** Shows "Environment is empty".
- Console:** Displays the R startup message and the command `>`.
- Bottom Bar:** Includes a search bar, file icons, and system status.

# 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.

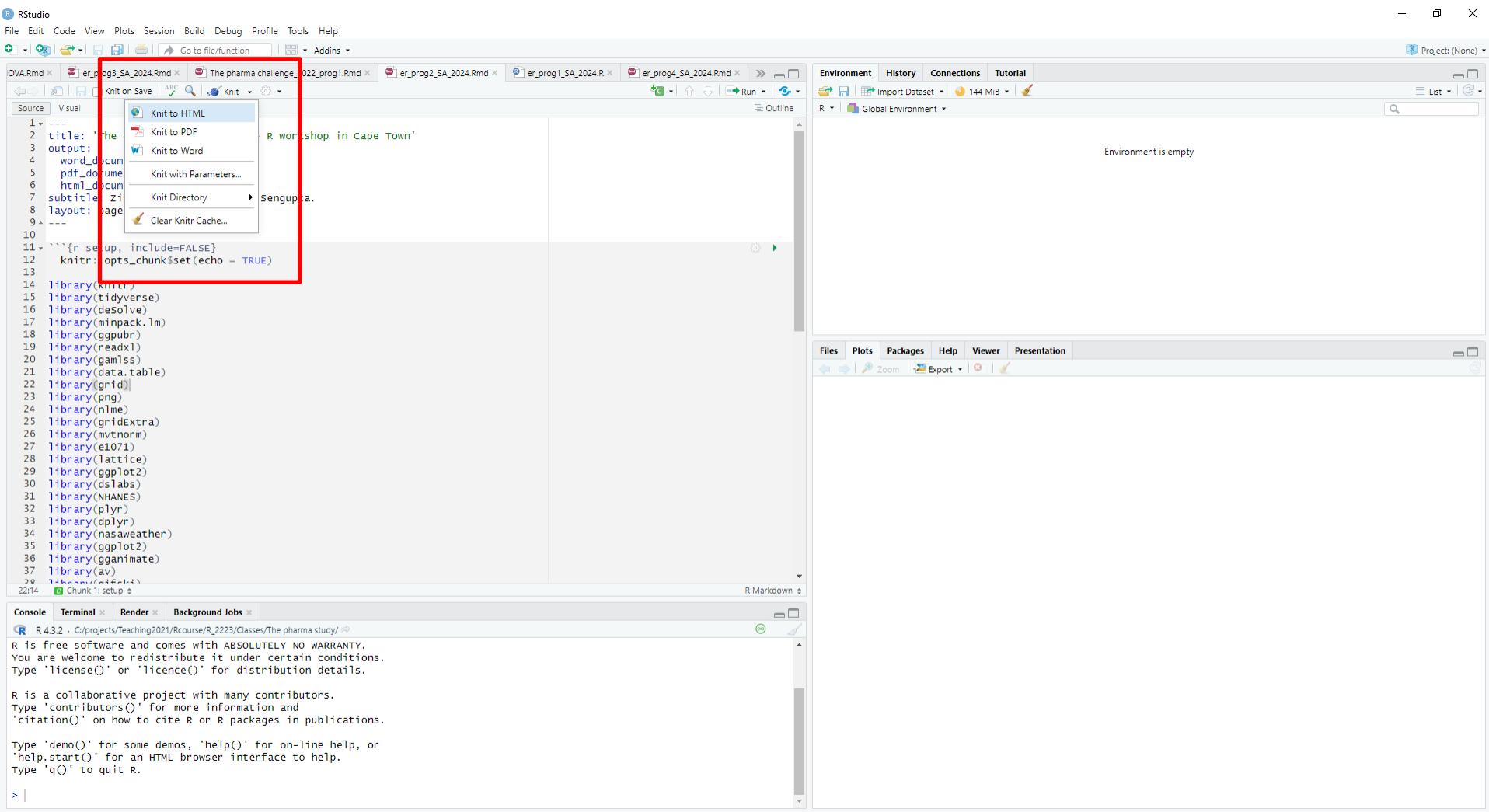
The code editor also highlights several lines of code in blue, indicating they are part of a code chunk:

- 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.

At the bottom, the R console shows 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

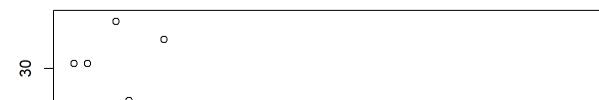
```
dim(mtcars)  
  
## [1] 32 11  
  
names(mtcars)  
  
## [1] "mpg"   "cyl"   "disp"  "hp"    "drat"  "wt"    "qsec" "vs"    "am"    "gear"  
## [11] "carb"
```

### 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

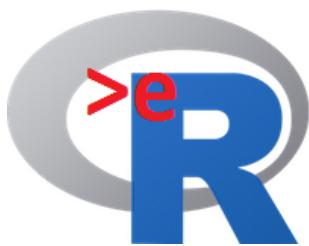
```
dim(mtcars)
## [1] 32 11
names(mtcars)
## [1] "mpg" "cyl" "disp" "hp" "drat" "wt" "qsec" "vs" "am" "gear"
## [11] "carb"
```

er\_prog2\_SA\_2024.docx: 1,277 characters (an approximate value).

Type here to search

Focus

15:32 ENG 8/02/2024



KU LEUVEN

Interuniversity Institute for Biostatistics  
and statistical Bioinformatics



# The mtcars data

Part 3

Advance HTML output

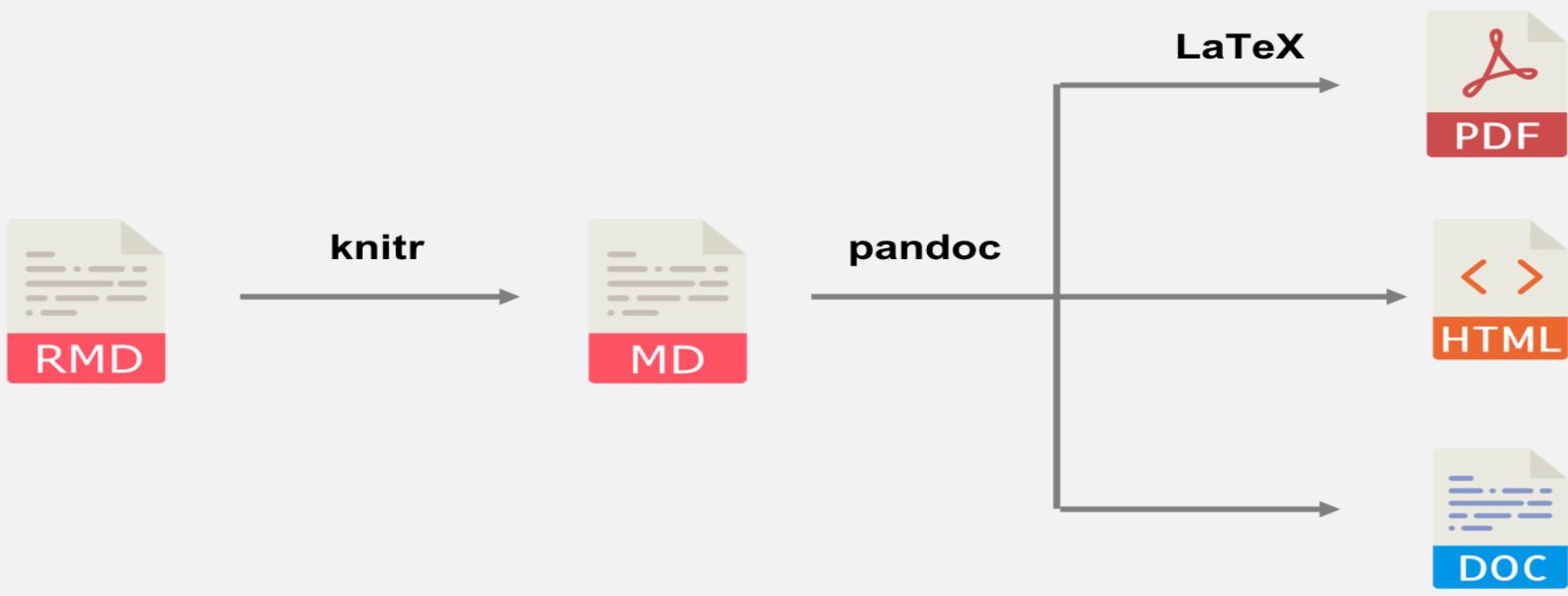
R Program: er\_prog3\_SA\_2024.Rmd

# 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

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	9.85	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

25 30 35

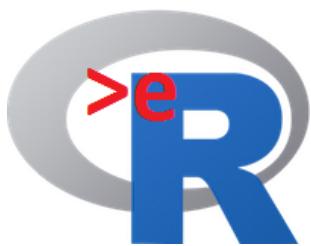
• An interactive HTML output.

• Presents the same analysis as before.

Table of content

Title

Analysis output



KU LEUVEN

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

The screenshot shows the RStudio interface with several windows open. The top-left window displays R Markdown code for an HTML document, specifically setting up the YAML header and the `knitr` package. A red arrow points to the line `bookdown::html_document2:`, and a large red brace groups the entire YAML section from `output:` to the end of the first section. The right side of the interface features a large text box with the heading "Set up the HTML document: document\_2". Below this, the RStudio environment pane shows that the Global Environment is empty. The bottom-left window is the R Console, showing the standard R welcome message and information about the R version and license.

```
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     return(NULL) ## otherwise knitr will print delay_code_labels every time  
38   } else {  
39     if (length(delay_code_labels) > 0) {  
40       cat(delay_code_labels, sep = "\n")  
41     }  
42   }  
43 })  
44  
45 The <t>lm</t> 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.

Environment History Connections Tutorial

Import Dataset 157 MB

Global Environment

Environment is empty

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 Global Environment

Environment is empty

• toc=true: add table of content.

• toc\_float: float TOC to the left

• toc\_depth: depth of header in toc

depth=2 implies that in the TOC:

Section

Subsection

Subsubsection

Section

Subsection

- 
- 
- 

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

32

# 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 <t>lm</t> 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
```

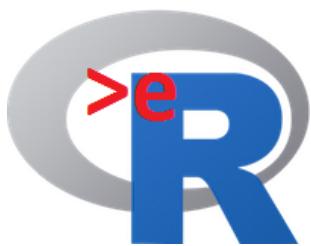
```
##          mpg cyl disp  hp drat    wt  qsec vs am gear carb
## Mazda RX4   21.0   6 160 110 3.90 2.620 16.46  0  1    4    4
## Mazda RX4 Wag 21.0   6 160 110 3.90 2.875 17.02  0  1    4    4
## Datsun 710  22.8   4 108  93 3.85 2.320 18.61  1  1    4    1
## Hornet 4 Drive 21.4   6 258 110 3.08 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



KU LEUVEN

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 7/2/2024

# Section, subsection, subsubsection

Depth=2

Only sections  
and subsections

Since depth=2, the subsubsection will appear in the TOC

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

- Miles/(US) gallon vs. the car's Weight **subsection**
- Scaterplot **subsubsection**

# 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, Start button, Icons for File Explorer, Google Chrome, Microsoft Word, R, and Powerpoint.

System tray: Network icon, Battery icon, ENG, 12:40, 7/02/2024, Mouse icon.

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.

The screenshot shows a web browser window displaying an R script execution results page. The URL is [C:/projects/Teaching2021/Rcourse/R\\_2223/Classes/The pharma study/er\\_prog3\\_SA\\_2024.html](C:/projects/Teaching2021/Rcourse/R_2223/Classes/The pharma study/er_prog3_SA_2024.html). The page title is "Simple linear regression using R". On the left, a sidebar menu lists: 1. The data, 2. Simple linear regression using R, 3. Data and estimated model, 4. Model diagnostic. The main content area shows the R command `## [1] 32 11` followed by the `mtcars` dataset. A red arrow points to the "Show" button next to the dataset. Below the dataset is the caption "Miles/(US) gallon vs. the car's Weight" and a scatterplot. The scatterplot has the x-axis labeled "wt" (Weight) and the y-axis labeled "mpg" (Miles/(US) gallon). The data points show a negative correlation, where weight increases and miles per gallon decrease.

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

Scatterplot

A scatterplot showing the relationship between weight (wt) on the x-axis and miles per gallon (mpg) on the y-axis. The x-axis ranges from approximately 2.6 to 5.0, and the y-axis ranges from approximately 10 to 45. The data points show a clear negative trend, indicating that as weight increases, miles per gallon decrease.

# 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 | X

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 ▾

## Simple linear regression using R

Ziv Shkedy et al

Show

1. The data

The mtcars dataset

dim(mtcars)

## [1] 32 11

head(mtcars)

## mpg cyl disp drat wt qsec vs am gear carb  
## Mazda RX4 21.0 6 160 3.90 2.620 16.46 0 1 4 4  
## Mazda RX4 Wag 21.0 6 160 3.90 2.875 17.02 0 1 4 4  
## Datsun 710 22.8 4 108 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 8 225 105 2.76 3.460 20.22 1 0 3 1

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

Scaterplot

Show

This screenshot shows a web-based R environment. At the top, there's a navigation bar with tabs for 'C:/projects/Teaching2021/Rcourse/R\_2223/Classes/The pharma study/er\_prog3\_SA\_2024.html' and 'er\_prog3\_SA\_2024.html'. Below the navigation is a sidebar with a table of contents: '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', and '4. Model diagnostic'. The main content area has a header 'Simple linear regression using R' and author information 'Ziv Shkedy et al'. A red arrow points from the sidebar's '1. The data' section to a red box around the R code and its output. Another red arrow points from the sidebar's '2. Simple linear regression using R' section to a scatterplot below. The R code shown is 'dim(mtcars)' and 'head(mtcars)'. The output of 'dim(mtcars)' is '## [1] 32 11'. The output of 'head(mtcars)' is a table of car data. Below the code box is the title 'Miles/(US) gallon vs. the car's Weight' and the subtitle 'Scaterplot'. The scatterplot shows a negative correlation between weight (wt) on the x-axis and miles per gallon (mpg) on the y-axis.

# 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
165
166
167
```

A red box highlights the section from line 146 to 153, which contains the code for creating a scatterplot of miles per gallon versus weight.

To the right of the code, there is a bulleted list:

- Subsection
- Subsubsections
- Plot + correlation

The RStudio environment pane shows an empty global environment. The bottom status bar indicates the R version is 4.3.2, the date is 7/02/2024, and the time is 12:45.

# 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

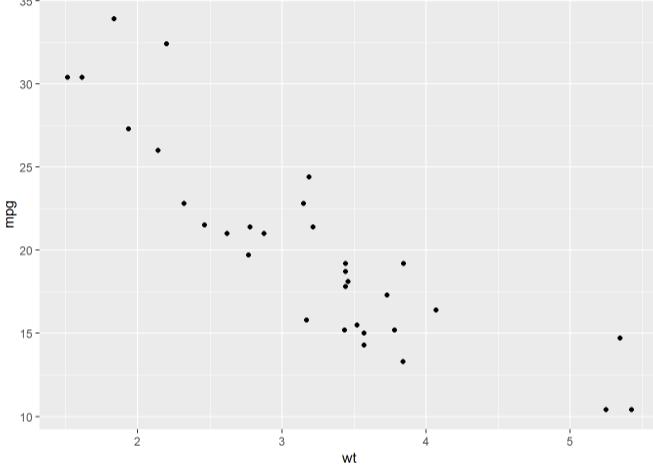


Figure 1: mpg vs. weight

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

44

# 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\_1WANOVA.Rmd \* er\_prog3\_SA\_2024.Rmd \* The pharma challenge\_2022\_prog1.Rmd \* er\_prog2\_SA\_2024.Rmd \* er\_prog1\_SA\_2024.R \* Run Outline

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 $mpg_{i}=\beta_0+\beta_1 \times weight_{i}+\varepsilon_i$.  
170  
171 ~`{r, echo=TRUE, message=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. respectivly, 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,  $\hat{y}=37.28 - 5.34 \times wt$   
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  
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  
1700 The <tt>lm()</tt> R function
```

# The linear regression model

Environment is empty

```
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 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

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

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 parameter 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 -

Type here to search

12:49 7/02/2024

# Code in the Rmd file

The screenshot shows the RStudio interface with an Rmd file open. The code editor on the left contains R code, some of which is highlighted with a red box. A red arrow points from this highlighted code to the text "Free text not a part of the R code." To the right, a brace groups the highlighted code and the text "Text text text", with the label "Program structure." below it. The R console at the bottom shows standard R startup messages.

```
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  $\$mpg_{i} = \beta_0 + \beta_1 \times weight_i + \epsilon_i$ .  
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,  $\hat{y} = 37.28 - 5.34 \times wt_i$   
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  
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
```

Environment is empty

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 → ↑

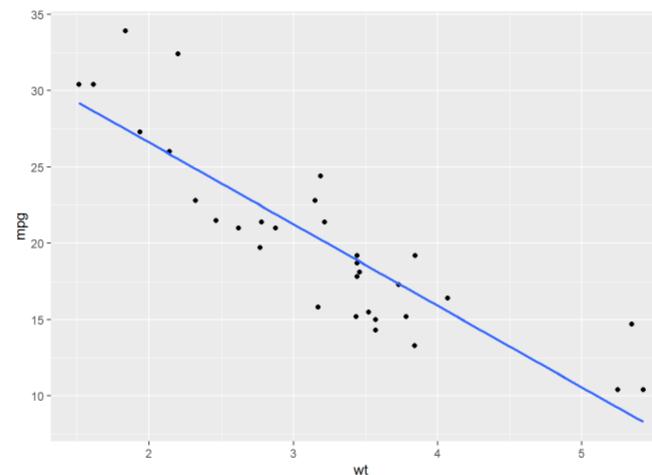


Figure 2: Data and fitted model

### 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.

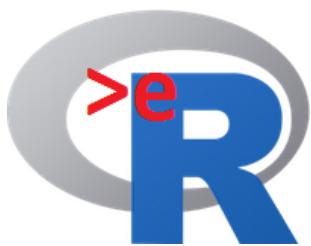
```
## 1 2 3 4 5 6 7  
## -2.2826106 -0.9197704 -2.0858521 1.2873499 -0.2001440 -0.6932545 -3.9053627  
## 8 9 10 11 12 13 14
```

↑

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.



KU LEUVEN

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

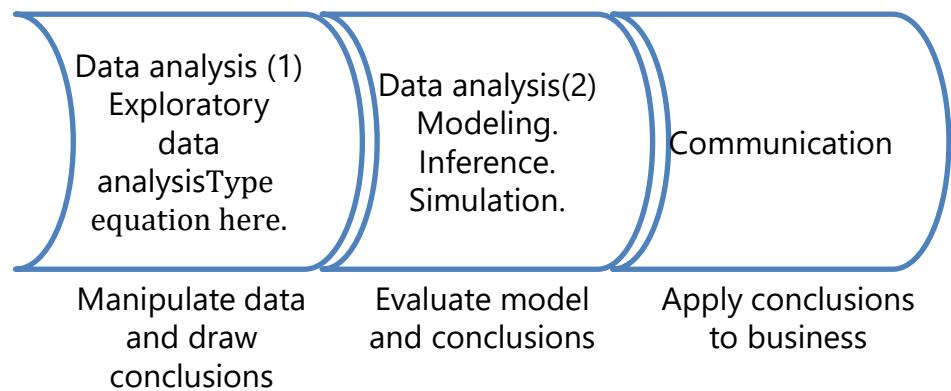
Modeling the association between the fuel consumption and the car's weight.

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

Methodology: simple linear regression.



We “translate” the methodology to software usage



Boxplot by treatment group.

A simple linear regression.

A report.

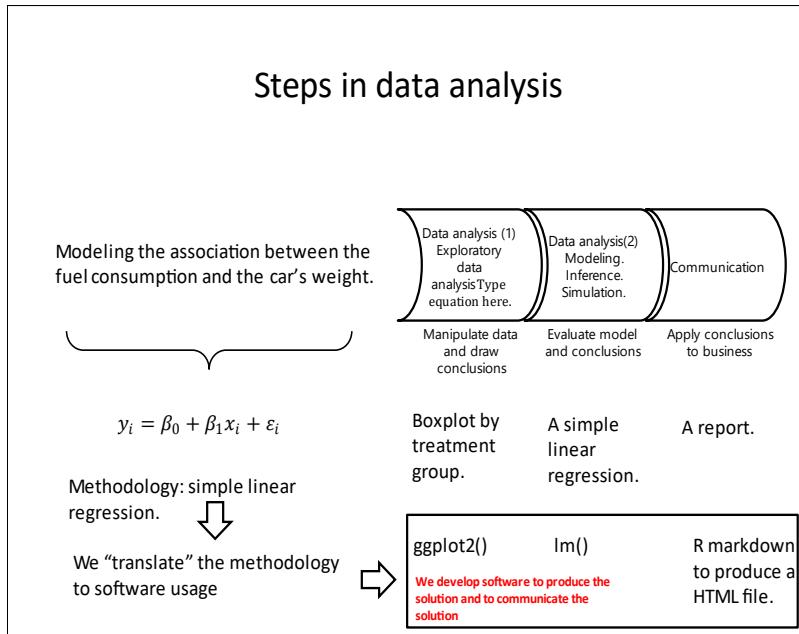
ggplot2()

lm()

R markdown to produce a HTML file.

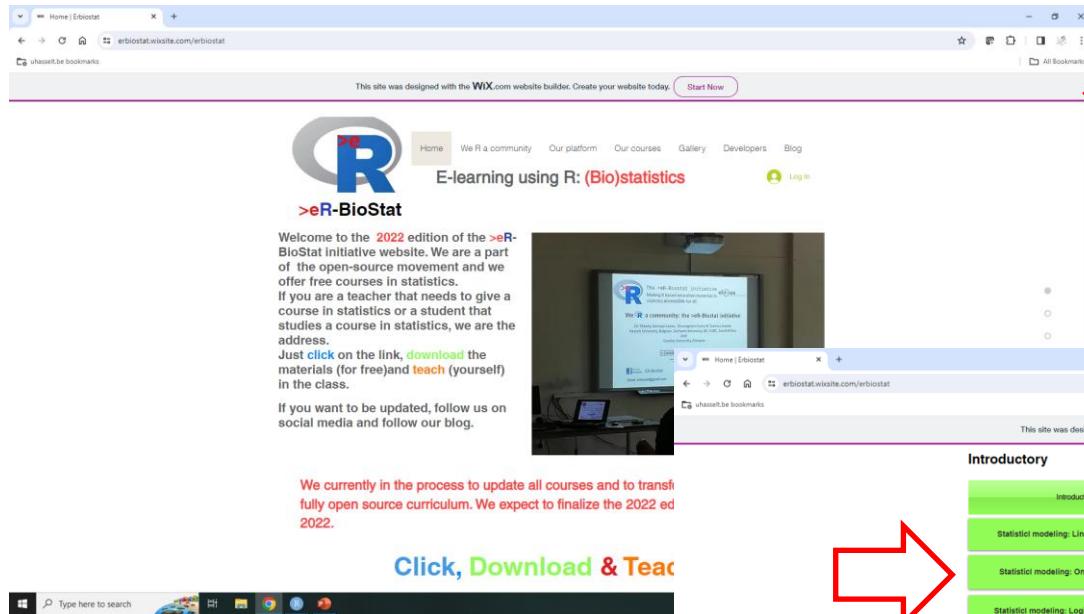
We develop software to produce the solution and to communicate the solution

# 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



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

>eR-Biostat website.

List of courses.

The screenshot shows a list of courses categorized into three sections: Introductory, Advanced, and Basic. Each section contains several course titles. A red arrow points from the left side of the page towards the course categories, and another red arrow points down towards the 'Basic' 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            |

| 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](#)

The screenshot shows a Wix website for "Introduction to Statistical modeling using R". The header includes the Wix logo and a "Start Now" button. Below the header is a large R logo. The main content area has a heading "Introduction to Statistical modeling using R" and a sub-section ">>eR-BioStat". A text box lists course topics: Simple linear regression using R, One-way ANOVA using R, and Logistic Regression using R. Another text box states that courses can be given together or separately. A third text box notes that courses were developed at an undergraduate level. A figure titled "Figure 3.1: Chick weight by diet group" is displayed, showing a boxplot of chick weight (Y-axis, 100-400) versus diet group (X-axis). The boxplot includes data points for sunflower, soybean, meatmeal, linseed, horsebean, and casein diets.

## About

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

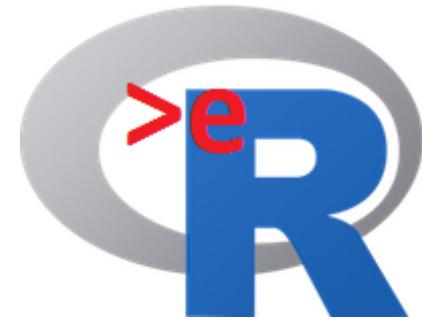
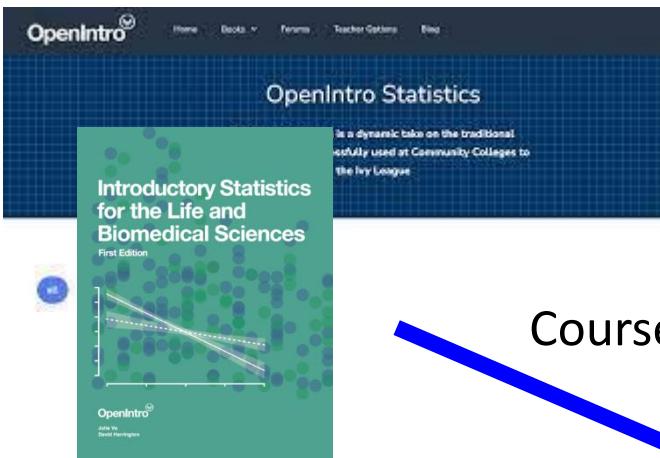
## Topics

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

# Introduction to statistical modeling using R



Course materials from two sources

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

**>eR-BioStat**

Introduction to Statistical modeling using R

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, or each one separately as a part of a specific course in statistical modelling.

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

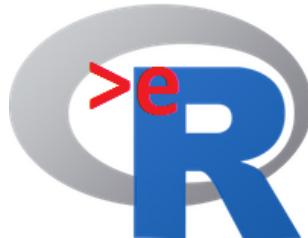
**About**  
General information about the course and course materials and the study materials available.

**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 course

# Course I

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.

## Course materials:

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

A red arrow points from the 'Datasets' link in the Simple Linear Regression section to the 'Datasets' links in the One-Way ANOVA and Simple Logistic Regression sections.

| Simple Linear Regression                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                          | One-Way ANOVA                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                         | Simple Logistic Regression                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                        |
|-------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|-----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|-----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| <p>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:</p> <ul style="list-style-type: none"><li>• Introduction and model formulation.</li><li>• Fitting a simple linear regression model using the lm() function in R.</li><li>• Model diagnostic.</li><li>• Model diagnostic in R.</li></ul> <p>External datasets are available in the data repository.</p> <p><a href="#">Slides (PDF): simple linear regression</a></p> <p><a href="#">Slides (PP): simple linear regression</a></p> <p><a href="#">R programm</a></p> <p><a href="#">Datasets</a></p> | <p>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:</p> <ul style="list-style-type: none"><li>• The one-way ANOVA model.</li><li>• Sources of Variability.</li><li>• One-way ANOVA using R: the aov() function.</li><li>• Model formulation and hypotheses testing.</li><li>• Analysis of the pharmaceutical experiment.</li><li>• Model diagnostic in R: normal probability plot.</li><li>• Multiple testing.</li></ul> <p>External datasets are available in the data repository.</p> <p><a href="#">Slides (PDF): One-Way ANOVA</a></p> <p><a href="#">Slides (PP): One-Way ANOVA</a></p> <p><a href="#">R programm</a></p> <p><a href="#">Datasets</a></p> | <p>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:</p> <ul style="list-style-type: none"><li>• Introduction and example tour.</li><li>• Fitting a simple linear logistic regression model using the glm() function in R.</li><li>• Model formulation.</li><li>• Interpretation of the model parameters.</li></ul> <p>External datasets are available in the data repository.</p> <p><a href="#">Slides (PDF): Logistic regression</a></p> <p><a href="#">Slides (PP): Logistic regression</a></p> <p><a href="#">R programm</a></p> <p><a href="#">Datasets</a></p> |

# 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](#)

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



## 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.

[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](#)

13:00  
7/02/2024 ENG

58

# Course II

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

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

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.

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

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.

Simple linear regression

Slides (PDF): simple linear regression  
Slides (PP): simple linear regression  
Slides (Rmd): simple linear regression

Slides (PDF): One-Way ANOVA  
Slides (PP): One-Way ANOVA  
Slides (Rmd): One-Way ANOVA

Slides (PDF): Logistic regression  
Slides (PP): Logistic regression  
Slides (Rmd): Logistic regression

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

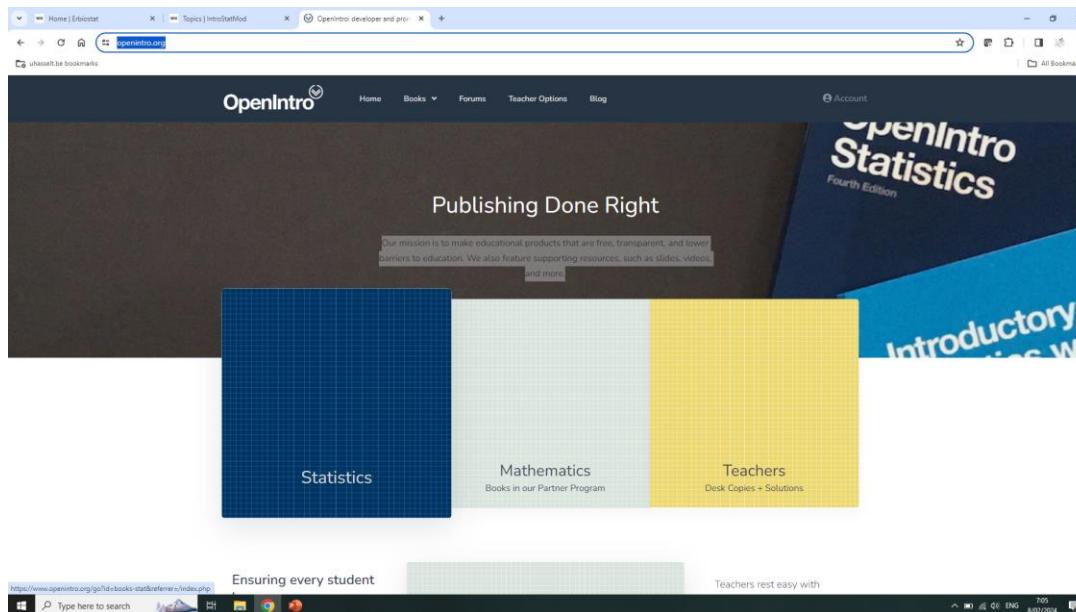
# openintro.org

- The OpenIntro project was founded in 2009 to improve the quality and availability of education by producing exceptional books and teaching tools that are 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](http://openintro.org).

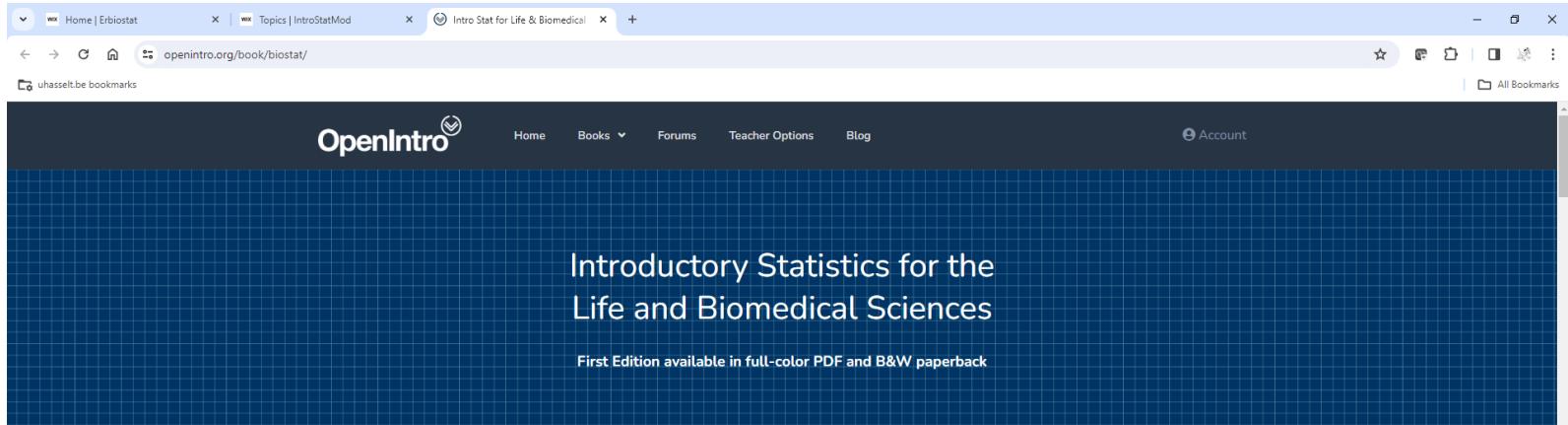
# openintro.org

<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.

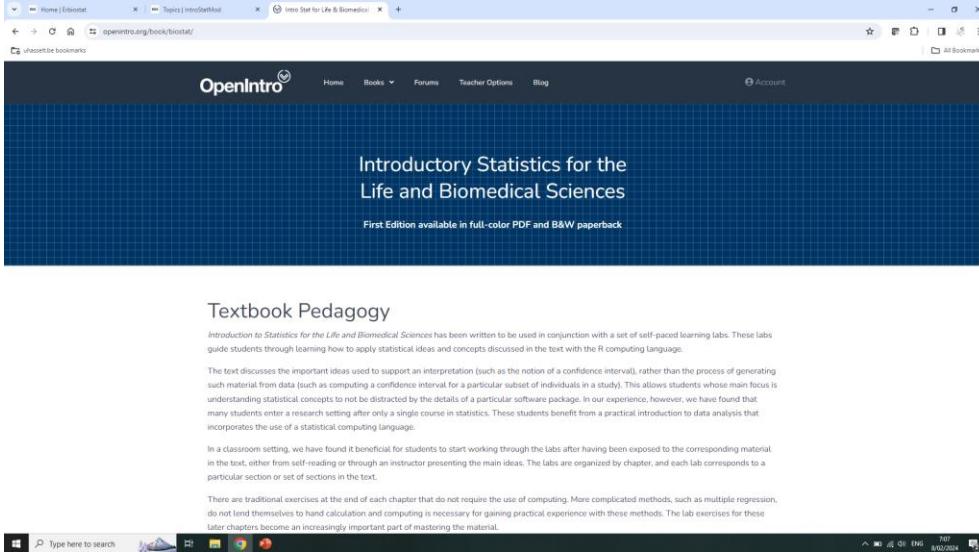
In a classroom setting, we have found it beneficial for students to start working through the labs after having been exposed to the corresponding material in the text, either from self-reading or through an instructor presenting the main ideas. The labs are organized by chapter, and each lab corresponds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



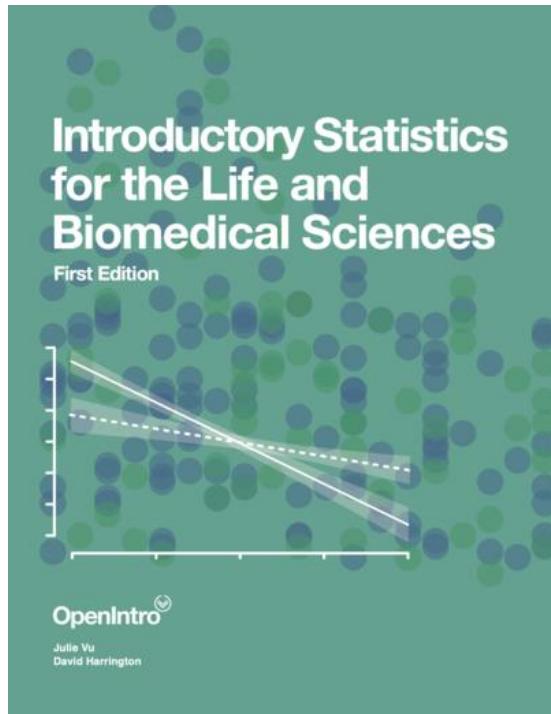
OpenIntro Biostatistics:  
developing open source materials  
for teaching and learning applied  
statistics with R.

Developed by Dave Harrington  
and Julie Vu.

Planning to work 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 Tex are available.
- Presentations and practical sessions (labs) are available online for free.

# Course II

This site was designed with the **Wix.com** website builder. Create your website today.

[Start Now](#)

respectively, available online here:

Online book

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.

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

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): simple linear regression

Slides (PP): simple linear regression

Slides (Rmd): simple linear regression

Simple linear regression

Slides (PDF): One-Way ANOVA

Slides (PP): One-Way ANOVA

Slides (Rmd): One-Way ANOVA

Slides (PDF): Logistic regression

Slides (PP): Logistic regression

Slides (Rmd): Logistic regression

# 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

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

# The Rmd file for the slides

A screenshot of a GitHub repository interface. The repository is named 'eR-Biostat / Courses'. The current view is the 'Code' tab for the file 'unit\_06\_simple\_linear\_regression.Rmd'. The code is an R Markdown document. It starts with a YAML header:

```
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.
```

The GitHub interface shows standard navigation tools like back, forward, and search at the top, and a file tree on the left. The bottom of the screen shows the Windows taskbar and system tray.

# 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 of the section is "THE MAIN IDEAS". It states: "Linear regression provides methods for examining the association between a quantitative response variable and a set of possible predictor variables." A bullet point below it says: "• Linear regression should only be used with data that exhibit linear or approximately linear relationships."

The next section is titled "Simple linear regression" and describes it as "used to estimate the linear relationship between a response variable  $y$  and a single predictor  $x$ ". Two bullet points follow: "• The response variable  $y$  can be referred to as the *dependent* variable, and the predictor variable  $x$  the *independent* variable." and "• 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 of the section is "THE MAIN IDEAS ...".

The bottom of the screen shows the Windows taskbar with icons for File Explorer, Google Chrome, and Microsoft Edge. A search bar says "Type here to search". The system tray shows battery level, signal strength, ENG, 7:10, 8/02/2024, and a battery icon.

# 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](#)



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

uhasselt.be bookmarks

15-12-2023    >eR-BioStat

Code ▾

## 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

70

# A part of the website of the course

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  $x$  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

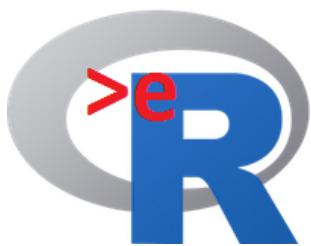
#### The mtcars dataset

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



KU LEUVEN

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 side-by-side screenshots of a website and its corresponding GitHub preview.

**Left Screenshot:** A screenshot of a Microsoft Edge browser window showing the homepage of "Introduction to Statistical modeling using R". The page features a large R logo, navigation links for Home, About, Topics, Online tutorials, and Contact. Below the navigation, there's a section titled ">>eR-BioStat" with a brief description of the course focus on statistical modeling using R, mentioning topics like Simple linear regression, One-way ANOVA, and Logistic Regression. It also notes that the course can be given together or separately as a part of a specific course in statistical modeling. The courses are developed at an undergraduate level. There's a note about the availability of open source files and online slides. At the bottom, there are links for About, Topics, and Online tutorials, along with a search bar.

**Right Screenshot:** A screenshot of a Microsoft Edge browser window showing a GitHub preview of an Rmd file titled "Tens\_2024\_Proj1\_V2.Rmd". The page displays the R code for a linear regression model, starting with  $y = \hat{y}_0 + \hat{y}_1x_1$ . It includes a section titled "1. General Introduction" and "Linear regression models". Below the code, there's a section titled "Examining scatterplots" with a heading "The mtcars dataset". It shows the first few rows of the mtcars dataset:

| # | mpg  | cyl | disp | hp  | drat | wt    | qsec  | vs | am | gear | carb |
|---|------|-----|------|-----|------|-------|-------|----|----|------|------|
| 1 | 21.0 | 6   | 160  | 110 | 3.99 | 2.875 | 18.08 | 0  | 1  | 4    | 4    |
| 2 | 21.0 | 6   | 160  | 110 | 3.99 | 2.875 | 17.02 | 0  | 1  | 4    | 4    |
| 3 | 22.8 | 4   | 108  | 93  | 3.85 | 2.32  | 18.6  | 1  | 1  | 4    | 4    |
| 4 | 21.4 | 6   | 256  | 110 | 3.08 | 3.215 | 15.44 | 1  | 0  | 3    | 1    |
| 5 | 18.7 | 8   | 360  | 175 | 3.08 | 3.44  | 16.02 | 0  | 1  | 5    | 4    |
| 6 | 18.1 | 8   | 227  | 108 | 4.93 | 3.44  | 16.52 | 1  | 0  | 3    | 1    |

# The Rmd program

The screenshot shows the RStudio interface with several panes:

- Code Editor (Source tab):** Displays an R Markdown file containing YAML header settings and R code. A red curly brace on the left side of the code editor highlights the first few lines of the YAML header, which define the output type as an HTML document with a table of contents.
- Environment Pane:** Shows the "Global Environment" tab with the message "Environment is empty".
- Console:** Displays the R startup message, the R version (R 4.3.2), and the current working directory (C:/projects/Teaching2021/Rcourse/R\_2223/Classes/The pharma study).

A large black text overlay "Document setting" is positioned in the center-left area of the screen, pointing towards the YAML header in the code editor.

# 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 ▾

## 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.

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    |
| ## 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 
$$y = b_0 + b_1 x \quad \text{newine}$$

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 
$$y = b_0 + b_1 x_1 + b_2 x_2 + \dots + b_p x_p \quad \text{.$$

```

146 ## Examining scatterplots

147

148 ### The `mtcars` dataset

149

150 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.

151

152 ````{r, echo=TRUE, message=FALSE, warning=FALSE}`

153 `dim(mtcars)`

154 `head(mtcars)`

155 ``````

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 `mpg`) and the car's weight (in 1000 lbs), the R object `wt`.

161

162 ````{r, echo=TRUE, message=FALSE, warning=FALSE, fig.cap="Investigating the relationship between fuel consumption and car weight."}`

9:15 Title :

Console Terminal x Render x Background Jobs x

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.

> |

Project: (None) x

Environment History Connections Tutorial

Import Dataset 156 MB

R Global Environment

• Section

Environment is empty

• Subsection

• Free text...

Files Plots Packages Help Viewer Presentation

Zoom Export

7:25 8/02/2024 ENG

# 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 | 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

- 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.

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 RX4W  21.0   6 160 110 3.90 2.620 16.46  0  1    4    4
## Datsun 710  22.8   4 108  93 3.85 2.320 18.00  1  1    4    1
## Hornet 4-Door 21.4   6 250 110 3.08 3.730 17.82  1  0    3    1
## Hornet Sportabout 18.7   8 360 175 3.15 3.440 17.30  0  1    5    2
## Valiant    18.1   8 225 105 2.76 3.460 18.90  0  1    5    2
## Fiat 128     32.4   4  78  62 4.93 1.615 21.40  1  0    4    1
## Fiat 130     15.2   8 196 122 3.92 3.440 18.00  0  1    5    2
## Fiat 150     15.2   8 200 120 3.92 3.440 18.00  0  1    5    2
## Dodge Challenger 14.3   8 304 160 3.08 4.070 17.40  0  0    3    3
## AMC Javelin  14.3   8 304 160 3.08 4.070 17.40  0  0    3    3
## Camaro Z28   13.9   8 350 160 3.08 4.070 17.40  0  0    3    3
## Pontiac Firebird 12.8   8 350 160 3.08 4.070 17.40  0  0    3    3
## Fiat X1-9     27.3   2  95  52 4.93 1.615 21.40  1  0    4    1
## Porsche 914-2 26.0   4 120  90 4.43 2.320 18.00  1  0    4    1
## Toyota Corolla 18.0   4  95  52 4.93 1.615 21.40  1  0    4    1
## Lincoln Continental 10.4   8 360 180 3.08 4.930 17.40  0  0    5    4
## Chrysler New Yorker 14.7   8 360 180 3.08 4.930 17.40  0  0    5    4
## Dodge Dart    17.8   8 225 105 3.08 4.070 17.40  0  0    3    3
## AMC Rebel DT 14.3   8 304 160 3.08 4.070 17.40  0  0    3    3
## Fiat 127     31.8   4  70  60 4.93 1.415 21.40  1  0    4    1
## Mercedes-Benz 190 15.2   8 256 120 3.43 3.440 18.00  0  1    5    2
## Volvo 142D   18.1   4 121  93 3.92 2.875 19.20  1  0    4    2
## Volvo 142E   18.1   4 121  93 3.92 2.875 19.20  1  0    4    2
## Toyota Corona 17.8   4 121  93 3.92 2.875 19.20  1  0    4    2
## Honda Civic   18.7   4  95  52 4.93 1.615 21.40  1  0    4    1
## Toyota Corona 17.8   4 121  93 3.92 2.875 19.20  1  0    4    2
## Toyota Corona 17.8   4 121  93 3.92 2.875 19.20  1  0    4    2
## Toyota Corona 17.8   4 121  93 3.92 2.875 19.20  1  0    4    2
```

# The Rmd program

RStudio

File Edit Code View Plots Session Build Debug Profile Tools Help

Go to file/function Addins

OVA.Rmd x er\_prog3\_SA\_2024.Rmd x The pharma challenge\_2022\_prog1.Rmd x er\_prog2\_SA\_2024.Rmd x er\_prog1\_SA\_2024.R x er\_prog4\_SA\_2024.Rmd x Run

Source Visual Outline

1.1 The MOTOR TREND CAR RƯU 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 <code>mtcars</code>. The dataset contains information about 11 variables and 32 cars. Use <code>help(mtcars)</code> to get more information about the data.

152  
153 ~`{r, echo=TRUE, message=FALSE, warning=FALSE}  
154 dim(mtcars)  
155 head(mtcars)  
156 ~`  
  
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 <code>wt</code>) and the car's weight (in 1000 lbs), the R object <code>wt</code>.  
161  
162 ~`{r, echo=TRUE, message=FALSE, warning=FALSE, fig.cap="mog vs. weight"}  
163 #plot(mtcars\$wt,mtcars\$mpg, ylab = "mpg", xlab = "weight (1000 lbs)")  
164 qplot(wt,mpg,data = mtcars)  
165 ~`  
166  
167 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.  
168  
169 ~`{r, echo=TRUE, message=FALSE, warning=FALSE}  
170 cor(mtcars\$wt,mtcars\$mpg)  
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 relationship shown in a scatterplot.  
180  
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 Miles/(US) gallon vs. the car's Weight

- Analysis in R:
  - Scatterplot + correlation.

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

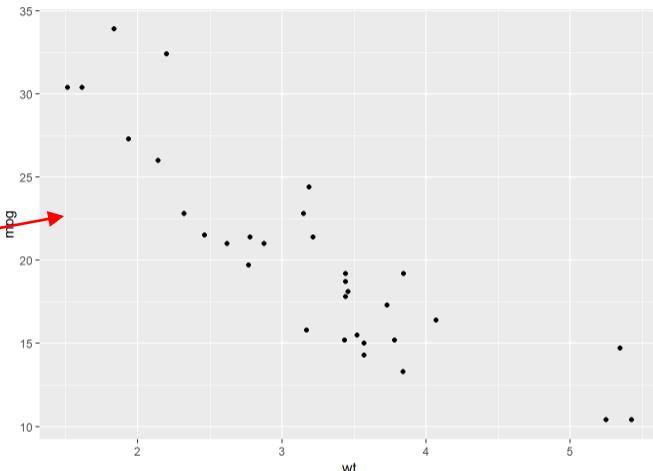
C:/projects/Teaching2021/Rcourse/R\_2223/Classes/The pharma study/er\_prog4\_SA\_2024.html  
er\_prog4\_SA\_2024.html | Open in Browser | Find | Publish | X

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`.



A scatterplot showing the relationship between car weight (wt) on the x-axis and fuel consumption (mpg) on the y-axis. The x-axis ranges from approximately 1.6 to 5.4, and the y-axis ranges from 10 to 35. The data points show a clear negative linear trend, indicating that as weight increases, fuel efficiency decreases.

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
```

- Analysis in R:
  - Scatterplot.
  - Correlation.

A red double-headed vertical arrow pointing upwards, indicating the transition from the analysis steps to the resulting scatterplot.

Text without the code.

## 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

Type here to search       7:26 ENG 8/02/2024

79

# 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)
```

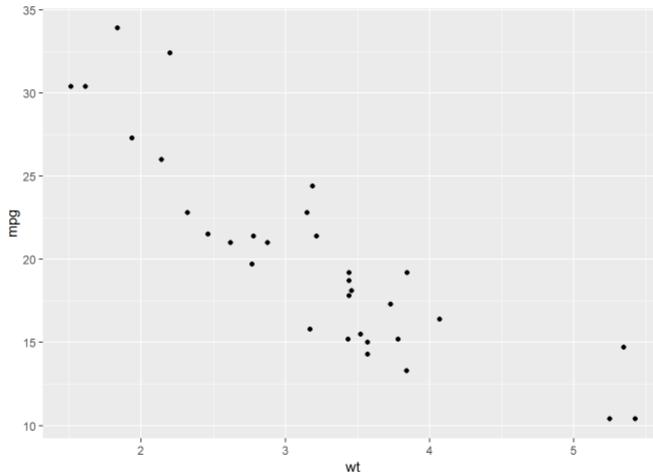


Figure 1: mpg vs. weight

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  $s_x$  and  $s_y$ .  
220  
221  $r$ : correlation between  $s_x$  and  $s_y$ .  
222  
223  
224 # 3. simple linear regression using R  
225  
226 ## The  $\text{lm}()$  R function  
227  
228 For the  $\text{mtcars}$  dataset, we consider the model  
229  
230  $\text{mpg}_i = \beta_0 + \beta_1 \times \text{weight}_i + \varepsilon_i$ .  
231  
232 In the above model, the variable  $\text{mpg}$  is the response and  $\text{weight}$  is the predictor. In R, we can fit  
the simple linear regression model using the R function  $\text{lm}$ . The function has the general call of  
 $\text{lm}(y \sim x)$ . The output for the  $\text{mtcars}$  data is shown below.  
233  
234 ````{r, echo=TRUE, message=FALSE, warning=FALSE}  
235  $\text{fit.lm} \leftarrow \text{lm}(\text{mpg} \sim \text{wt}, \text{data} = \text{mtcars})$   
236  $\text{summary(fit.lm)}$   
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,  $\hat{y} = 37.28 - 5.34 \times \text{wt}$   
244  
245 ````{r, echo=TRUE, message=FALSE, warning=FALSE, fig.cap="Data and fitted model"}  
246  $\text{qplot}(\text{wt}, \text{mpg}, \text{data} = \text{mtcars}) +$   
247  $\text{geom_smooth(method = "lm", se = F)}$   
248 ````  
249  
250 ## Parameter estimates  
251  
252 Parameter estimates for  $s_y$  and  $s_x$  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

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

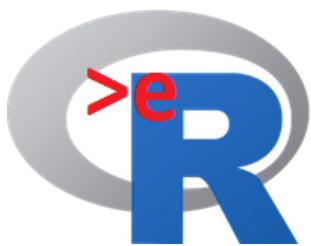
Figure 2 shows the data (mpg vs. weight) and fitted regression line,  $mpg_i = 37.28 - 5.34 \times wt_i$



Free text about regression.

- The regression model:
  - Code.
  - Output.

# Short summary

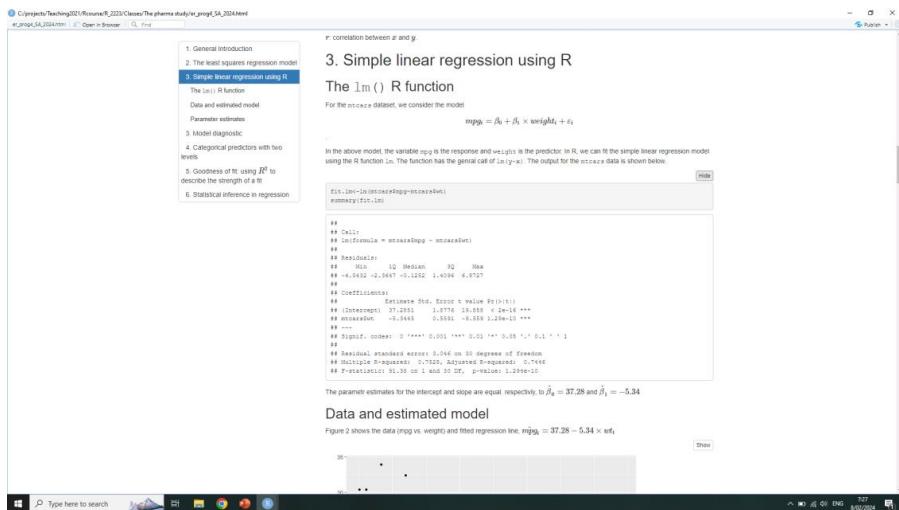


# Development of E-learning materials using R markdown

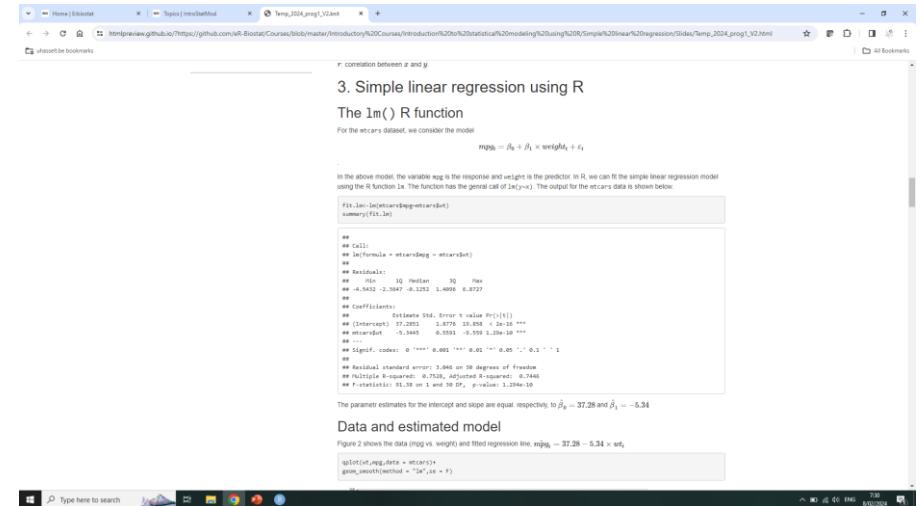
## Part 5b: the >eR-BioStat approach

How can we create a course online??

# From a laptop to a website..

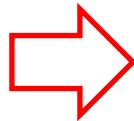


The screenshot shows a Windows desktop with a browser window open to a local file. The page title is "3. Simple linear regression using R". It contains an R script for fitting a linear model to the "mtcars" dataset. The script includes code for loading the data, fitting the model using the `lm` function, and summarizing the results. A note at the bottom states: "The parameter estimates for the intercept and slope are equal, respectively, to  $\hat{\beta}_0 = 37.28$  and  $\hat{\beta}_1 = -5.34$ ". Below the code, there is a figure showing a scatter plot of "mpg" vs "wt" with a regression line.



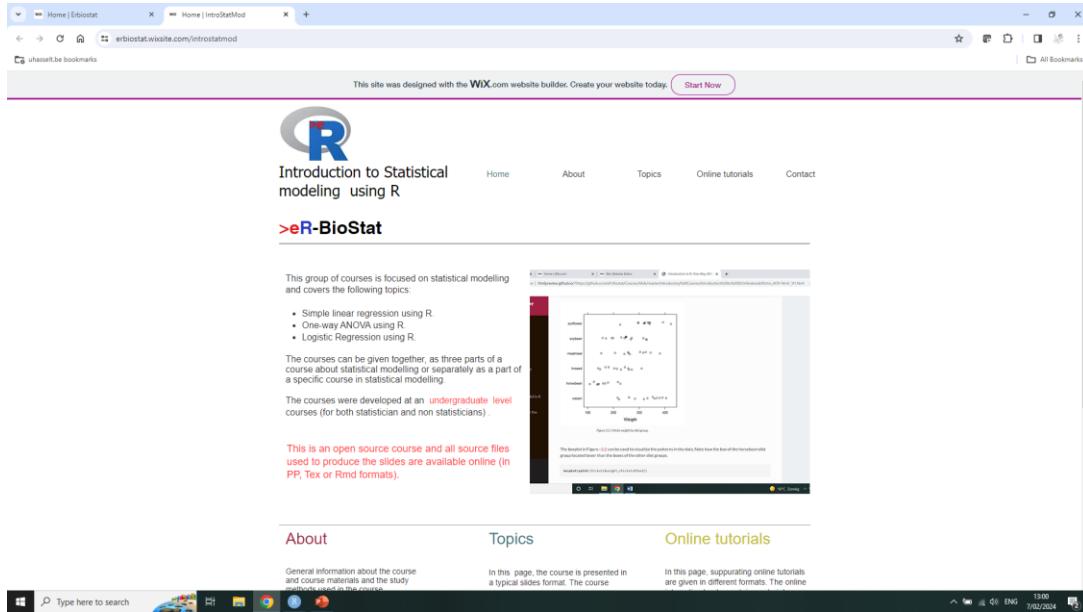
The screenshot shows a Windows desktop with a browser window open to a GitHub page. The URL is "https://github.com/ab-Bisrat/Courses/blob/master/Introductory%20to%20Statistical%20Modeling%20using%20R%20near%20Regression/slides/Temp\_2024\_prog1\_V2.html". The page content is identical to the one on the laptop, displaying the R script for simple linear regression using the "mtcars" dataset. The GitHub interface includes a sidebar with navigation links like "Home", "Issues", "Topics", and "Pull requests".

The HTML file on the laptop....



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

# Introduction to statistical modeling using R



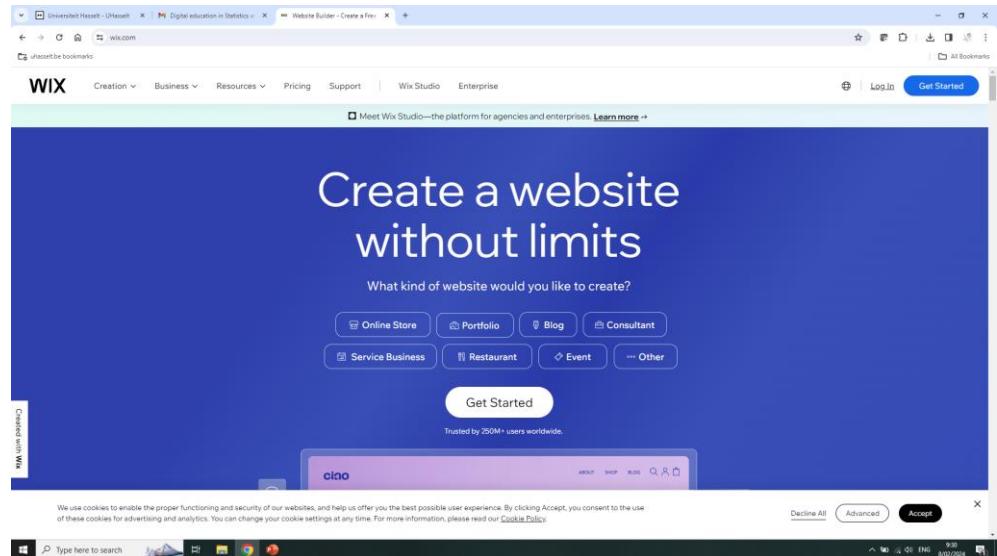
- Website.
- Storage of materials:
  - Slides.
  - Programs.
  - Datasets.
  - Etc.
- Software for the analysis ?
- Storage space & cost ?

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

- Our approach: bring costs to zero...

# How can I develop a website for my course ?

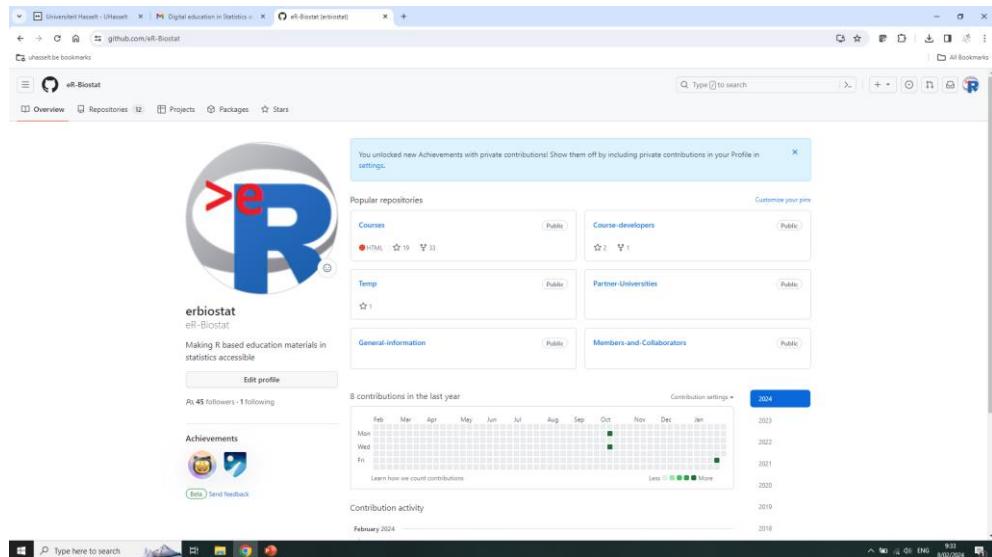
- Websites for the courses were developed using WIX.
- Free.
- Easy to use and learn.



<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.
- 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

.. Add files via upload 2 years ago

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

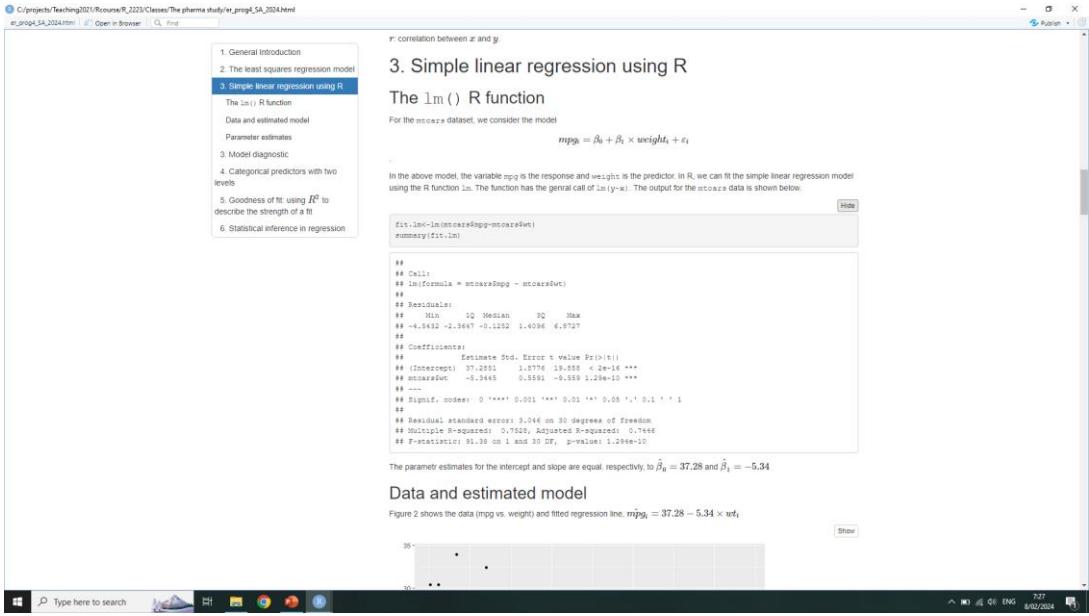
9:36 8/02/2024 ENG

<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. The R code shown is:

```
r: correlation between wt and mpg  
  
3. Simple linear regression using R  
  
The lm() R function  
  
For the mtcars dataset, we consider the model  
  
mpg = β₀ + β₁ × weight₁ + ε₁  
  
In the above model, the variable mpg is the response and weight₁ 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 -0.1282 1.4098 6.8727  
##  
## Coefficients:  
## Estimate Std. Error t value Pr(>|t|)  
## (Intercept) 37.8786 1.8778 19.888 < 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: 81.38 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:  $mpg = 37.28 - 5.34 \times wt$

- 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.

# Discussion

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