

Development of E-learning materials using R & R markdown

Ziv Shkedy, Bernard Osangir, Leyla Kodalci, Julia Duda
Hasselt University

Updated: 08/25

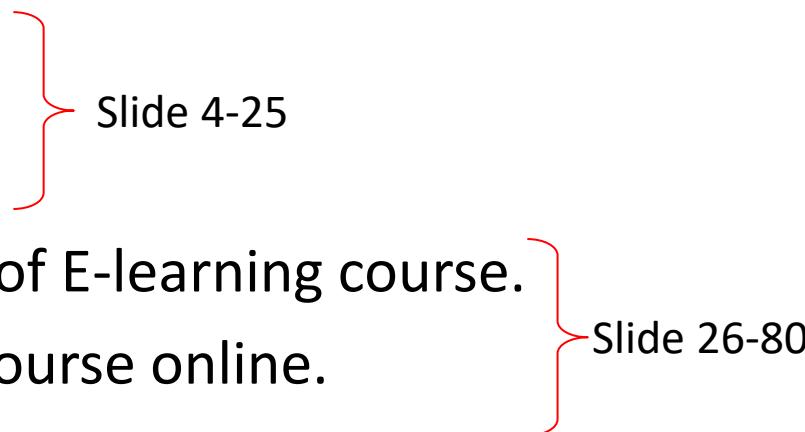


ER-BioStat

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

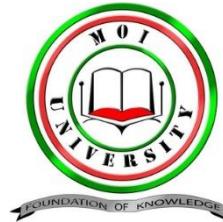
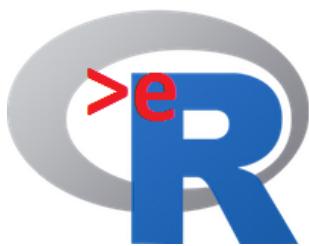
[@erbiostat](#)

Overview

- Starting point: first presentation:
 - Linear regression using R markdown.
 - Output development.
 - From R markdown to development of E-learning course.
 - From an output on my laptop to a course online.
 - Case study: 52-80
 - Our approach: slides 81-127.
- 
- Slide 4-25
- Slide 26-80

Rmd programs

- Basic analysis in R (10-16):
 - [er_prog1_SA_2024.R](#).
- Basic HTML file (17-25):
 - [er_prog2_SA_2024.Rmd](#).
- Advance HTML file (26-51):
 - [er_prog3_SA_2024.Rmd](#).
- Case study (52-80):
 - [er_prog4_VT_2024_V3a.Rmd](#).



Interuniversity Institute for Biostatistics
and statistical Bioinformatics

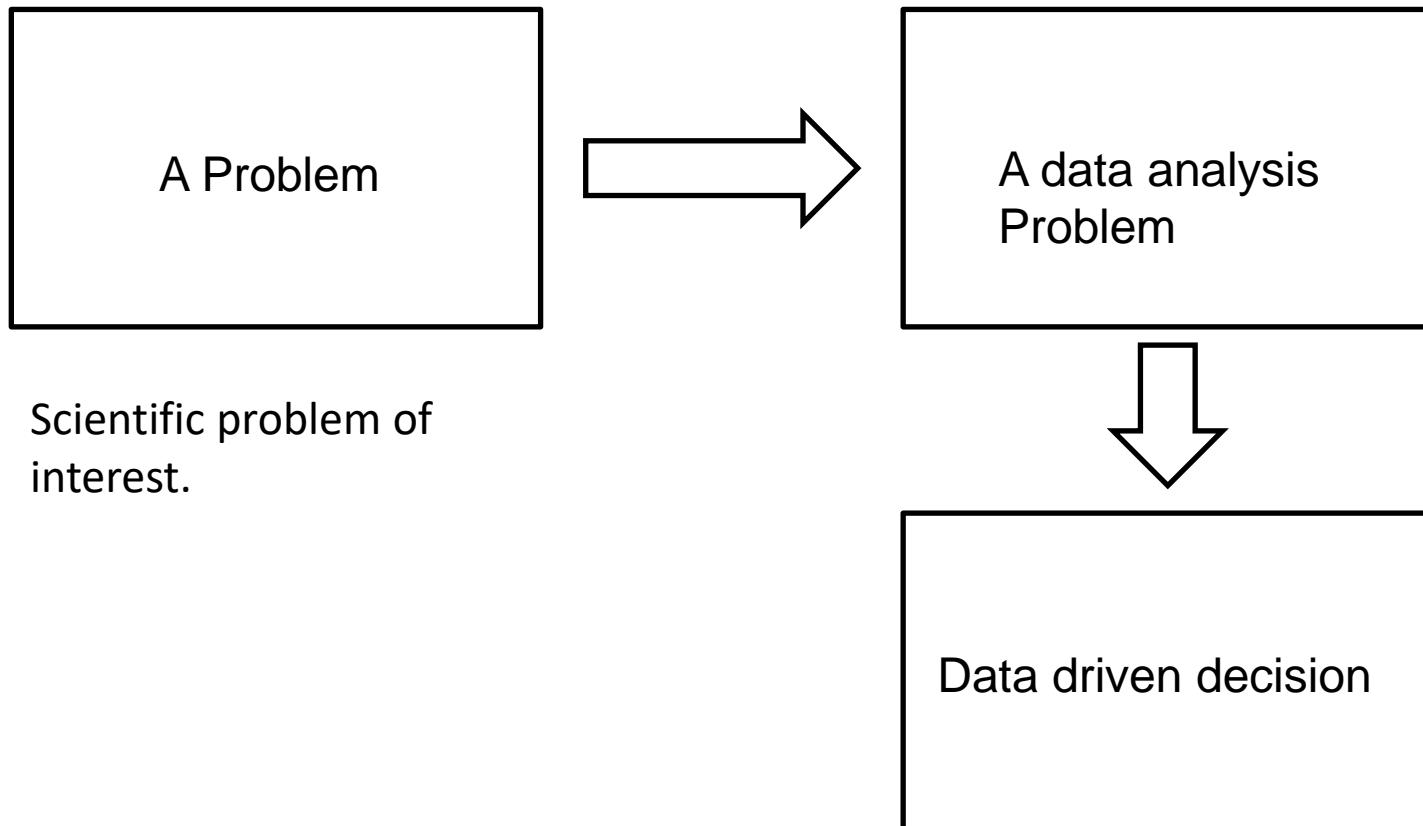
Short summary of the first session

Linear regression using R markdown

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

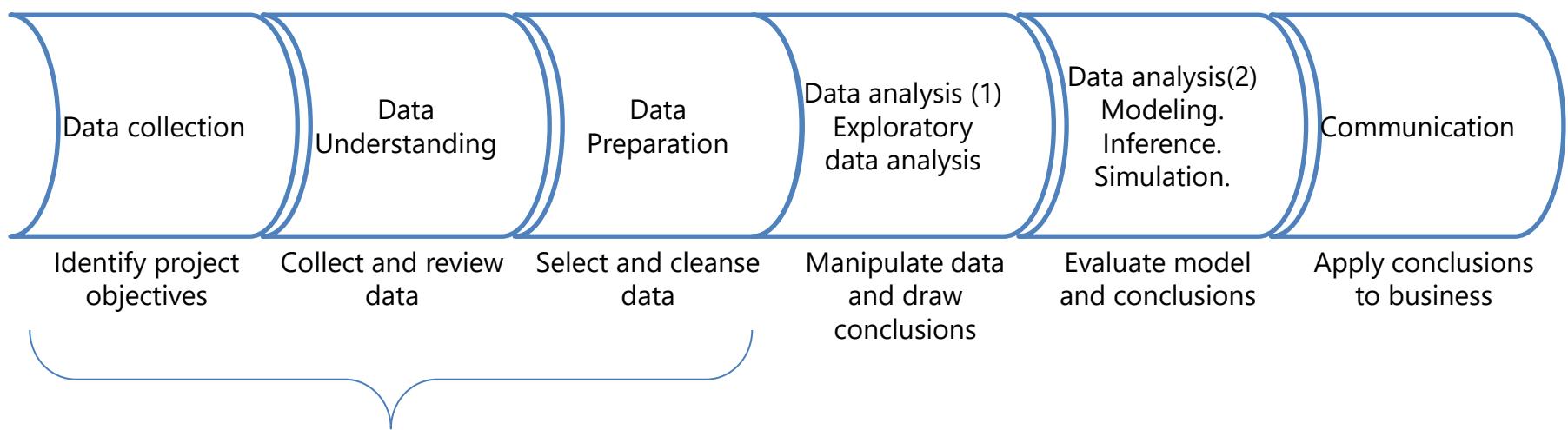
Steps in data analysis

- Data analysis approach in the course:



Steps in data analysis

- Steps related to data analysis:



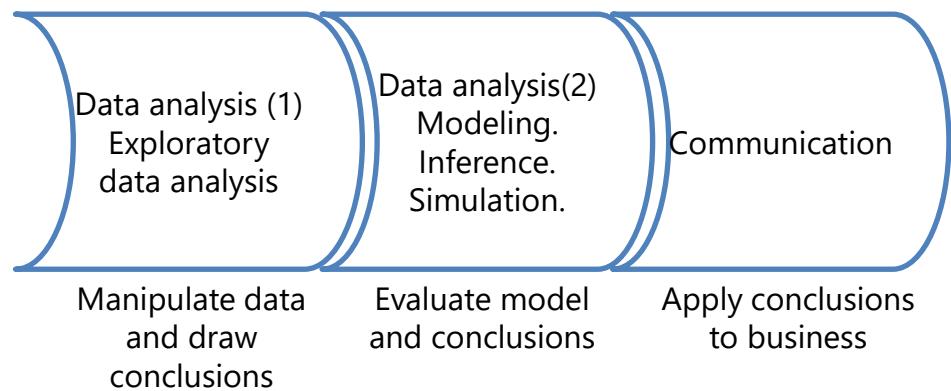
Not a part of our course

Steps in data analysis

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



Scientific problem of interest:
how to model the association ?

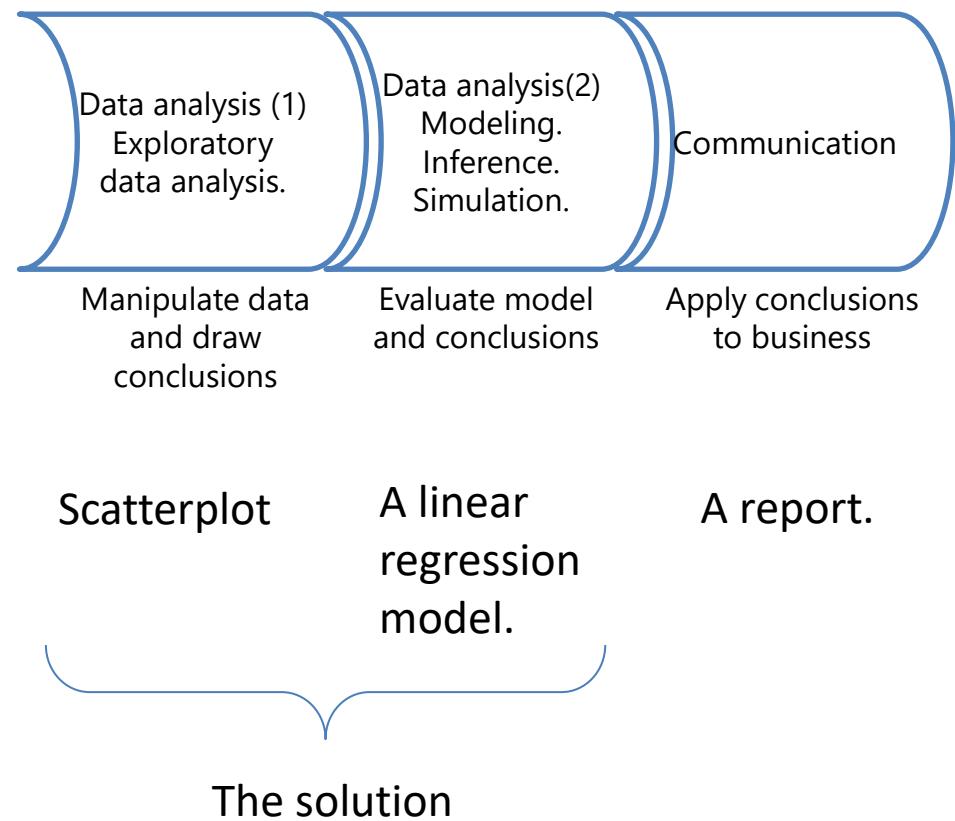


Steps in data analysis

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

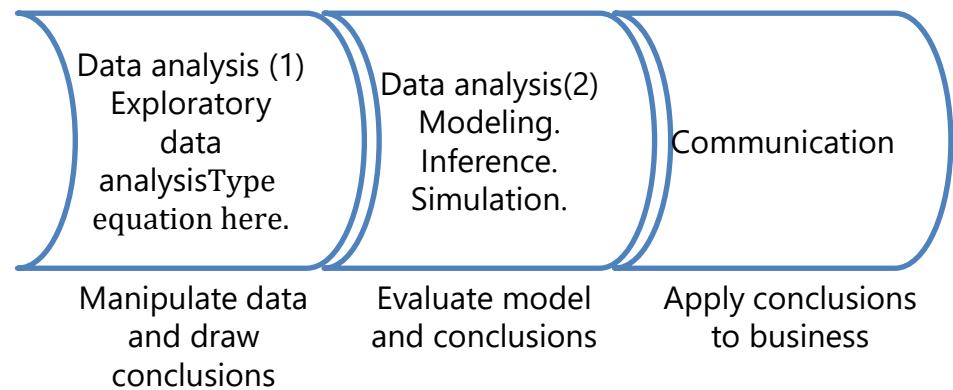
$$y_i = \beta_0 + \beta_1 x_i + \varepsilon_i$$

Methodology: simple linear regression.



Steps in data analysis

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



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

Methodology: simple linear regression.



We “translate” the methodology to software usage



Scatterplot.

Linear regression model.

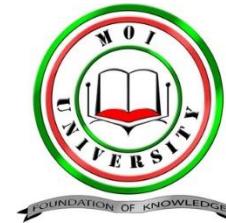
A report.

`ggplot2()`

`lm()`

R markdown to produce a HTML file.

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



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"  "wt"    "qsec"  
[8] "vs"    "am"    "gear"  
[11] "carb"
```

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

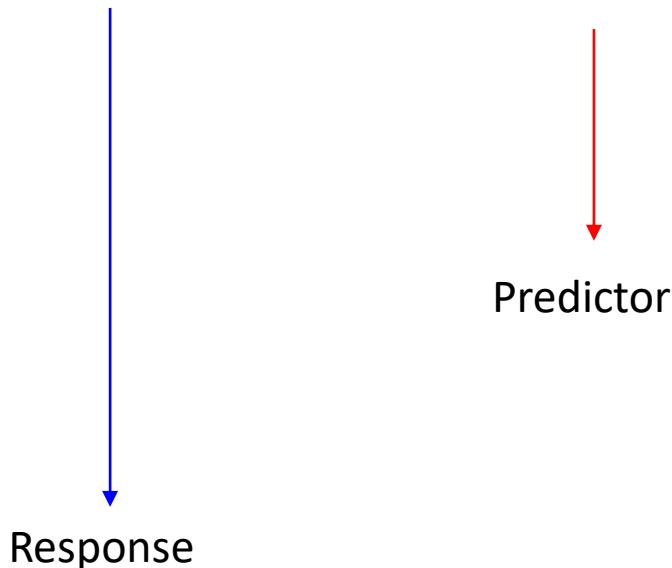
The R object for the data: 32 observations and 11 variables.
Variables names:
mpg: mile per gallon – the response.
wt: car's weight – the predictor.

The mtcars data in R

```
> head(mtcars)
```

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

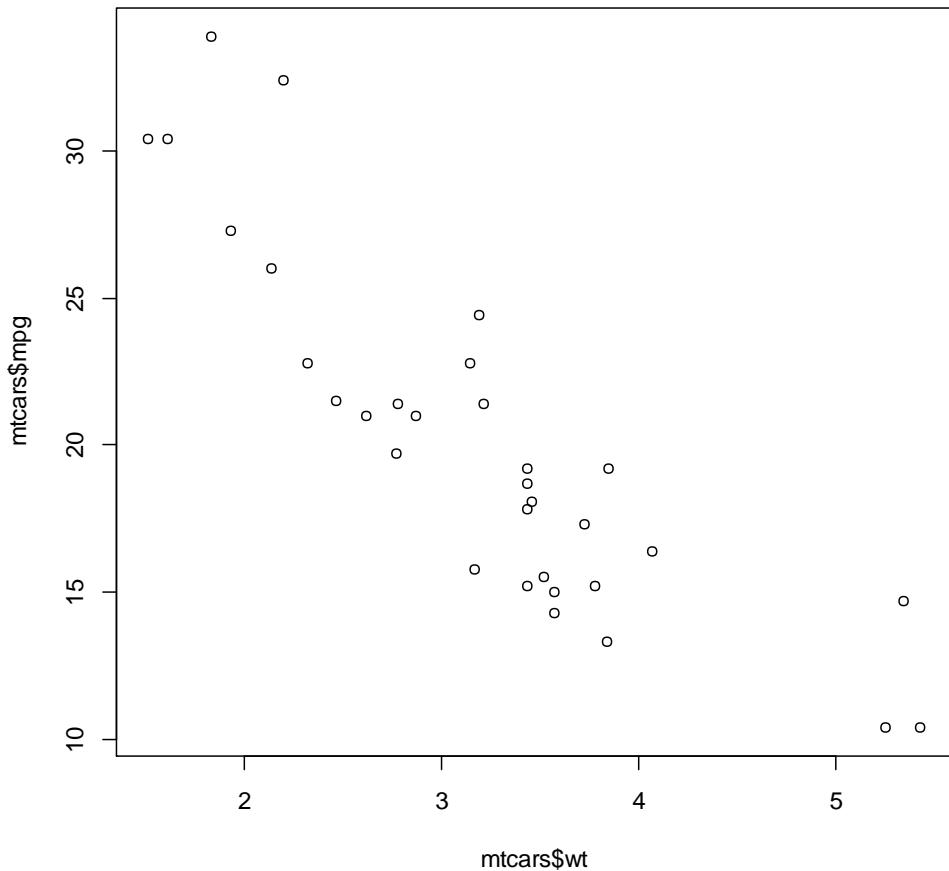
```
>
```



The mtcars data in R

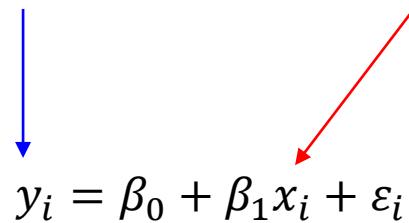
```
> plot(mtcars$wt, mtcars$mpg)  
> cor(mtcars$wt, mtcars$mpg)  
[1] -0.8676594
```

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



Simple linear regression in R

lm(response~predictor)

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


Simple linear regression in R: the mtcars data

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

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

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

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

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

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

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

- Output:
 - Parameter estimates etc.

R code for the analysis

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

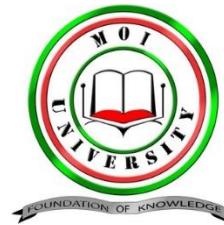
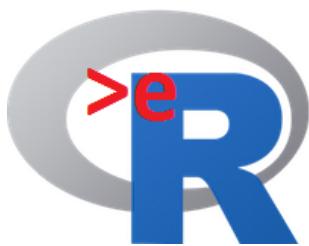


- Produce the plot.
- Calculate the correlation.
- Fit the model:

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

- Print the estimated model.

R Program: er_prog1_SA_2024.R



The mtcars data

Part 2

Analysis using basic Rmd programming

Rmd Program: er_prog2_SA_2024.Rmd

The output

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

The Rmd program

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

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

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

- A brace on the left side groups lines 1 through 9, labeled "Document setup."
- A brace on the left side groups lines 14 through 38, labeled "Many R packages, not all needed."

The Environment tab shows "Environment is empty". The bottom status bar indicates "R 4.3.2 · C:/projects/Teaching2021/Rcourse/R_2223/Classes/The pharma study/ · 15:30 8/02/2024 ENG".

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

Source Visual Outline

```
46 <!-- use_bookdown: TRUE -->
47 ---
48
49
50
51 # Baseline analysis
52
53 ## The mtcars data in R
54
55 ``{r}
56 dim(mtcars)
57 names(mtcars)
58 ...
59
60 ## First 6 lines
61
62 ``{r}
63 head(mtcars)
64 ...
65
66 ## Scatterplot
67
68
69 ``{r}
70 plot(mtcars$wt,mtcars$mpg)
71 cor(mtcars$wt,mtcars$mpg)
72 ...
73
74 ## Linear regression in R using the lm() function
75
76 ``{r}
77 fit.lm=lm(mtcars$mpg~mtcars$wt)
78 summary(fit.lm)
79 ...
80
81
82
```

22:14 Chunk 1: setup ↵ R Markdown

Console Terminal Background Jobs

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

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

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

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

>

Dimension of the data,
variables names and first
6 lines of the data

Plot & correlation

The regression model.

Dimension of the data,
variables names and first
6 lines of the data

Plot & correlation

The regression model.

X
` `` ` {r}
R code

The Rmd program

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

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

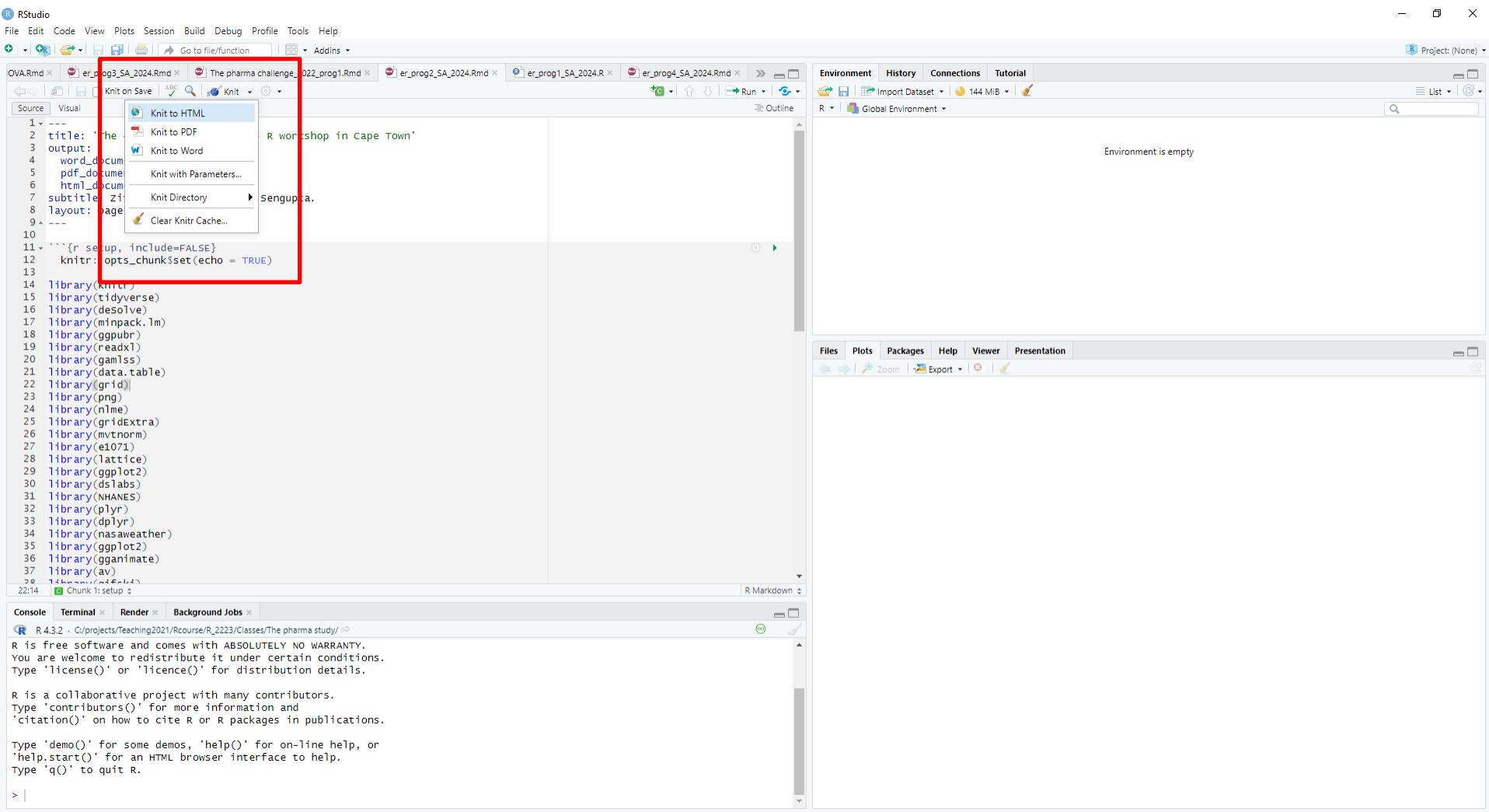
The code editor also highlights several lines of code in blue, indicating they are part of a code chunk:

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

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

The bottom console pane shows the R startup message and basic 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

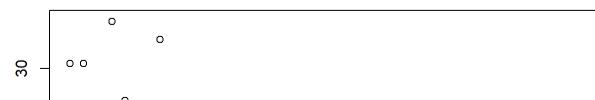
```
dim(mtcars)  
  
## [1] 32 11  
  
names(mtcars)  
  
## [1] "mpg"   "cyl"   "disp"  "hp"    "drat"  "wt"    "qsec" "vs"    "am"    "gear"  
## [11] "carb"
```

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 ENG 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

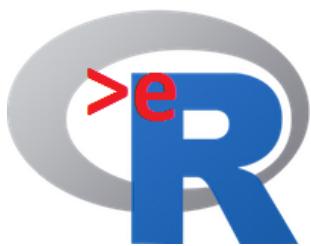
Ziv Shkedy and Rudradev Sengupta

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).



Interuniversity Institute for Biostatistics
and statistical Bioinformatics

The mtcars data

Part 3

Advance HTML output

Rmd Program: er_prog3_SA_2024.Rmd

What do we cover in this part ?

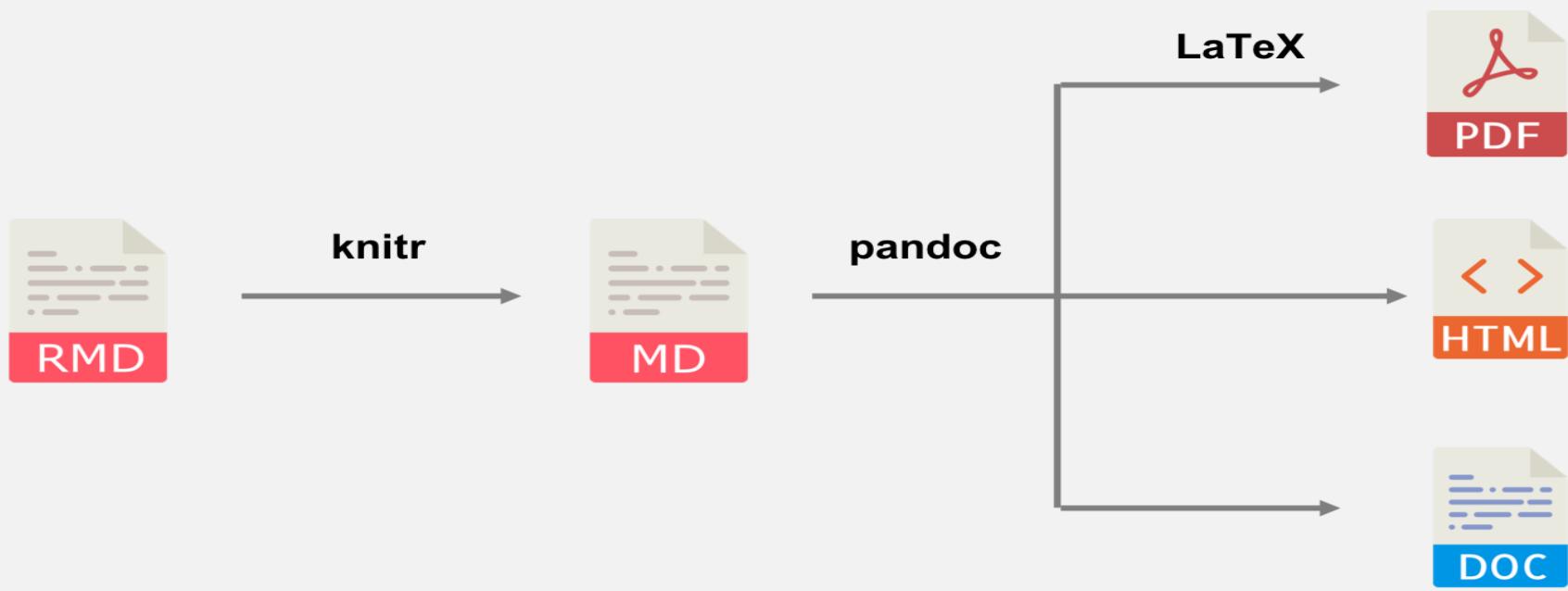
- We use the same example: the `mtcars` data and simple linear regression.
- How to produce a HTML file for the analysis we conducted before ?
- How to use the HTML file as a content for a website ?

Reproducible Research

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

The Rmd file

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



The HTML output

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

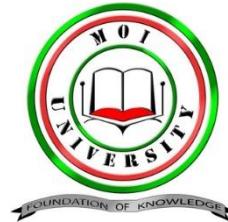
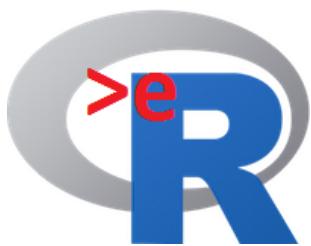
Scaterplot

Table of content

Title

Analysis output

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



Interuniversity Institute for Biostatistics
and statistical Bioinformatics

Part 3.1:

How to set up the HTML file ?

The Rmd file

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

Set up the document

RStudio

File Edit Code View Plots Session Build Debug Profile Tools Help

Go to file/function Addins

unit_05_inference_num_1ANOVA.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

Knit on Save ABC Knit

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 ````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 knitr_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     }  
40   })  
170:1 The <ttr>lm0</ttr> R Function ▾
```

R Markdown

Console Terminal × Render × Background Jobs ×

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

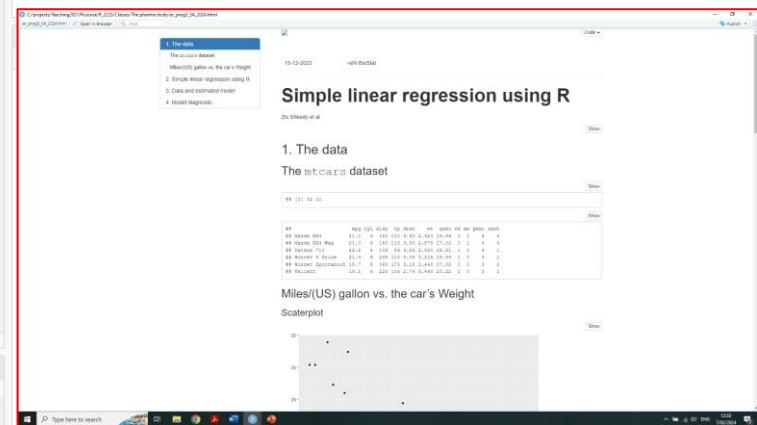
R is free software and comes with ABSOLUTELY NO WARRANTY.
You are welcome to redistribute it under certain conditions.
Type 'license()' or 'licence()' for distribution details.

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.

> |

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

Knit on Save Knit Run Addins

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: *simple linear regression using R **
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
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 <t>lm</t> R Function

Environment History Connections Tutorial

Import Dataset 157 MB Global Environment

Outline

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

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

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
```

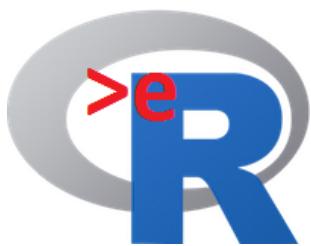
```
##          mpg cyl disp  hp drat    wt  qsec vs am gear carb
## Mazda RX4     21.0   6 160 110 3.90 2.620 16.46  0  1    4    4
## Mazda RX4 Wag 21.0   6 160 110 3.90 2.875 17.02  0  1    4    4
## Datsun 710    22.8   4 108  93 3.85 2.320 18.61  1  1    4    1
## Hornet 4 Drive 21.4   6 258 110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

35
30
25

12:28 7/02/2024



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 ## scatterplot
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 ENG 2

Section, subsection, subsubsection

Depth=2 → Only sections and subsections

Ziv Shkedy et al

1. The data section
The mtcars dataset subsection

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

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

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

Type here to search ENG 12:37 7/02/2024

Analysis code

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

The code for the analysis

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

1. The data

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

15-12-2023 >eR-BioStat

Code ▾
Show All Code
Hide All Code

Simple linear regression using R

Ziv Shkedy et al

1. The data

The mtcars dataset

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

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

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

Scaterplot

Windows taskbar: Type here to search, Start button, File Explorer, Google Chrome, File, Word, R, Powerpoint, 12:40, ENG, 7/02/2024, Page 42

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

Reading the external file

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



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

1. The data

The mtcars dataset

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

2. Simple linear regression using R

3. Data and estimated model

4. Model diagnostic

15-12-2023 >eR-BioStat

Simple linear regression using R

Ziv Shkedy et al

Show

1. The data

The mtcars dataset

Show

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

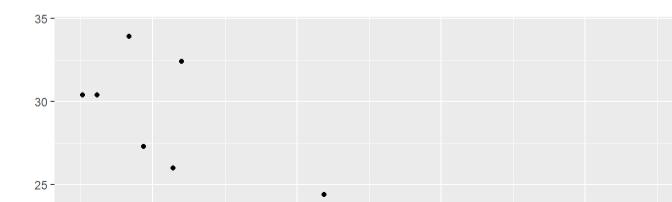
Show

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

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

Scaterplot

Show



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

The HTML file

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 down from the 'Show' button next to the 'dim(mtcars)' code result to a red box around the entire code execution results section. Another red arrow points down from the 'Show' button next to the 'head(mtcars)' code result to the same red box. The code execution results show the dimensions of the mtcars dataset (32 observations, 11 variables) and the first six rows of the dataset. Below this, there's a scatterplot titled 'Miles/(US) gallon vs. the car's Weight' with a 'Scaterplot' label. A red arrow points down from the 'Show' button next to the scatterplot to the plot itself.

Code in the Rmd file

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

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

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

The Environment pane shows an empty global environment.

The Plot pane shows a scatterplot of weight (0.000 lbs) on the x-axis versus mpg on the y-axis.

The Subsection, Subsubsections, and Plot + correlation sections are listed on the right side of the slide.

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

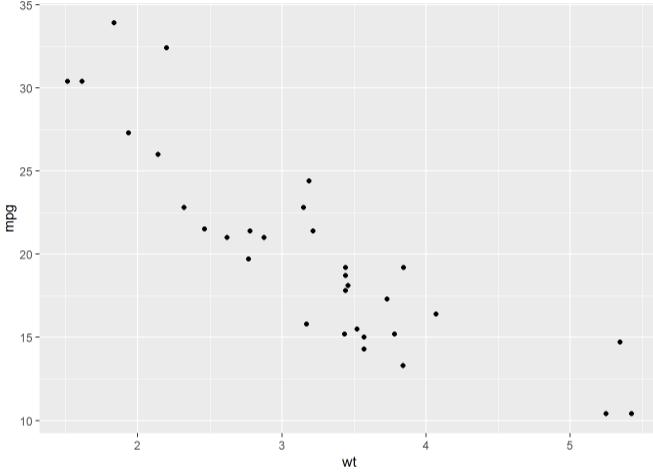


Figure 1: mpg vs. weight

Correlation

Show

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

- Subsection
- Subsubsection
- Plot + correlation

2. Simple linear regression using R

The `lm()` R function

For the `mtcars` dataset, we consider the model

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

Type here to search  12:46 ENG 7/02/2024 46

Code in the Rmd file

RStudio

File Edit Code View Plots Session Build Debug Profile Tools Help

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

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

Environment History Connections Tutorial

Import Dataset 154 MB

Global Environment

The linear regression model

Environment is empty

Files Plots Packages Help Viewer Presentation

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

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

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

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

> |

Type here to search

The output in the HTML file

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

Correlation

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

2. Simple linear regression using R

The lm() R function

For the mtcars dataset, we consider the model

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

The model:
free text.

The output

Code for the
model is
shown.

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

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

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

3. Data and estimated model

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

35 -

Windows Taskbar: Type here to search, File, Start, Task View, Chrome, Word, R, Powerpoint, etc.

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

Code in the Rmd file

RStudio

File Edit Code View Plots Session Build Debug Profile Tools Help

unit_05_inference_num_IWANOVAn.Rmd er_prog3_SA_2024.Rmd The pharma challenge_2022_prog1.Rmd er_prog2_SA_2024.Rmd er_prog1_SA_2024.Rmd

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  
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  
$\hat{\beta}_1=-5.34$  
178  
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 History Connections Tutorial

Global Environment

Environment is empty

Free text not a part of the R code.

Text text text

``{r}
R code
...}

Program structure.

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

Console Terminal Render Background Jobs

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

Type here to search

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 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, $\hat{mpg}_i = 37.28 - 5.34 \times wt_i$

Free text → ↑

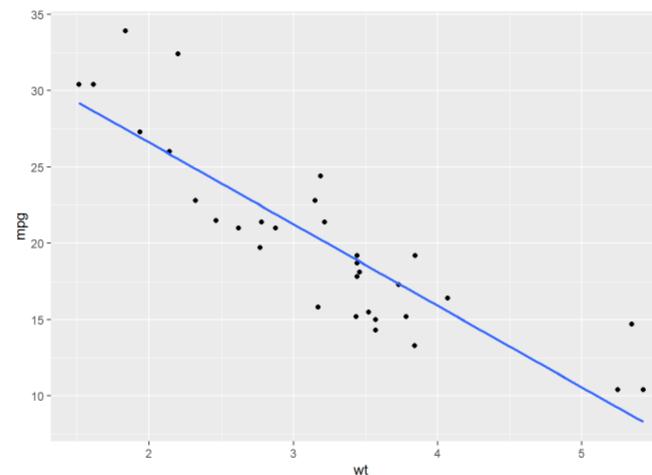


Figure 2: Data and fitted model

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.

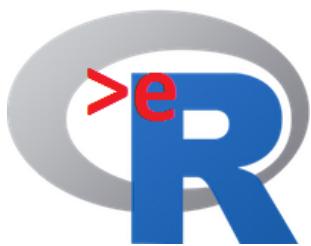
```
## 1 2 3 4 5 6 7  
## -2.2826106 -0.9197704 -2.0858521 1.2873499 -0.2001440 -0.6932545 -3.9053627  
## 8 9 10 11 12 13 14
```

Windows Type here to search | Start | File | Chrome | Downloads | WPS Office | R | P | Show | 12:50 7/02/2024

Short discussion

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

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



Interuniversity Institute for Biostatistics
and statistical Bioinformatics

Part 4

Case study: The NHANES dataset

The number of sleep hours per night

Rmd Program: er_prog4_VT_2024_V3a.Rmd.

The HTML book

The screenshot shows a Microsoft Edge browser window displaying an R Markdown document. The title of the page is "Foundations for inference using R". On the left, there is a sidebar with a table of contents:

1. Variability in estimates
A point estimate for the population parameter
Example: the wind speed in the airquality dataset
The variability of the sample mean
The sampling distribution for the mean
the Central Limit Theorem
2. Standard error of the mean
3. Confidence intervals
4. Hypothesis testing
5 Hypothesis testing and confidence intervals
6 Decision error (Type I and Type II error)

The main content area includes a timestamp (03-05-2024), a header (eR-BioStat), and a section titled "Foundations for inference using R". Below the title, it says "Ziv Shkedy and Thi Huyen Nguyen based on Chapter 4 in the book of Julie Vu and Dave Harrington *Introductory Statistics for the Life and Biomedical Sciences* (<https://www.openintro.org/book/biostat/>)". A "Show" button is present.

1. Variability in estimates

A point estimate for the population parameter

A natural way to estimate features of the population, such as the population mean weight, is to use the corresponding summary statistic calculated from the sample. For example, the sample mean \bar{x} is a point parameter estimate for the population (unknown) mean μ and the sample variance s^2 is a point parameter estimate for the population variance σ^2 .

ample: the wind speed in the airquality dataset

a and point estimates

airquality dataset gives information about 153 daily air quality measurements in New York, May to September 1973.

```
1] 153 6
```

Ozone	Solar.R	Wind	Temp	Month	Day
41	190	7.4	67	5	1
36	118	8.0	72	5	2
12	149	12.6	74	5	3
18	313	11.5	62	5	4
## 5	NA	NA	14.3	56	5
## 6	28	NA	14.9	66	5

The variable Wind is the average wind speed in miles per hour at 0700 and 1000 hours at LaGuardia Airport. The mean wind speed is $\bar{\mu} = \bar{x} = 9.95$ and the sample standard deviation is $\hat{\sigma} = s = 3.52$. This sample mean is a point estimate of the population mean. If

The HTML book

The screenshot shows a web browser window displaying an R Markdown document. The title of the document is "Foundations for inference using R". The left sidebar contains a table of contents with the following sections:

1. Variability in estimates
2. Standard error of the mean
3. Confidence intervals
4. Hypothesis testing
5. Hypothesis testing and confidence intervals

The main content area starts with the heading "1. Variability in estimates". Below it, there is a section about the population parameter, mentioning the sample mean \bar{x} as a point estimate for the population mean. It also notes that the sample standard deviation s is a point estimate for the population variance σ^2 . A link to the "airquality" dataset is provided.

Further down, there is a section about wind speed in the airquality dataset, stating that it contains about 153 daily air quality measurements in New York, May to September 1973. A code block shows the R command to load the dataset:

```
library(ggplot2)
library(dplyr)
library(tidyverse)
```

The status bar at the bottom right of the browser window shows the date as 25/02/2025 and the time as 16:04.

- Covers:
 - Theory:
 - One population & continuous response.
 - Confidence intervals.
 - Hypotheses testing.
 - Examples.
- R code for all examples in the book.

The NHANES data set

- The NHANES dataset consists of data from the US National Health and Nutrition Examination Study.
- Information about 76 variables is available for 10000 individuals included in the study.
- The 10000 individuals are considered as the **population**.

The NHANES data set: analysis of the number of sleep hours per night

- The variable of interest is the number of sleeping hours per night (the variable SleepHrsNight).
- Continuous variable.
- Information about the number of sleeping hours per night is available for 7755 individuals (i.e., the population).

The NHANES data set: analysis of the number of sleep hours per night

er_prog4_VT_2024-V3.knit

C:/Ziv_Temp_2023/Workshop_Vietnam_2025/ShortCourse/er_prog4_VT_2024-V3.html

uhasselt.be bookmarks

Example 2: The NHANES data set analysis of the number of sleep hours per night

The population

In this section, the variable of interest is the number of sleeping hours per night (the variable SleepHrsNight). Information about the number of sleeping hours per night is available for 7755 individuals (i.e., the population). The population mean and variance are $\mu = 6.927$ and $\sigma^2 = 1.813$, respectively.

Population parameters: μ, σ

[1] 7755

[1] 6.927531

[1] 1.81368

• Analysis:

- Point estimates.
- Confidence intervals.
- Hypothesis testing in one population.
- Continuous response (number of sleep hours).

Frequency

1607
25/02/2025

The NHANES data set analysis of the number of sleep hours per night

- 1. Variability in estimates
- 2. Standard error of the mean
- 3. Confidence intervals
- 4. Hypothesis testing

The Formal Approach to Hypothesis Testing

Example 1: wind speed in New York 1973

Example 2: The NHANES data set analysis of the number of sleep hours per night

5 Hypothesis testing and confidence intervals

6 Decision error (Type I and Type II error)

The population

In this section, the variable of interest is the number of sleeping hours per night (the variable `SleepHrsNight`). Information about the number of sleeping hours per night is available for 7755 individuals (i.e., the population). The population mean and variance are $\mu = 6.927$ and $\sigma^2 = 1.813$, respectively.

Hide

```
library(NHANES)
data(NHANES)
#dim(NHANES)
sleep<-na.omit(NHANES$SleepHrsNight)
length(sleep)
```

```
## [1] 7755
```

Hide

```
mean(sleep)
```

Population mean

```
## [1] 6.927531
```

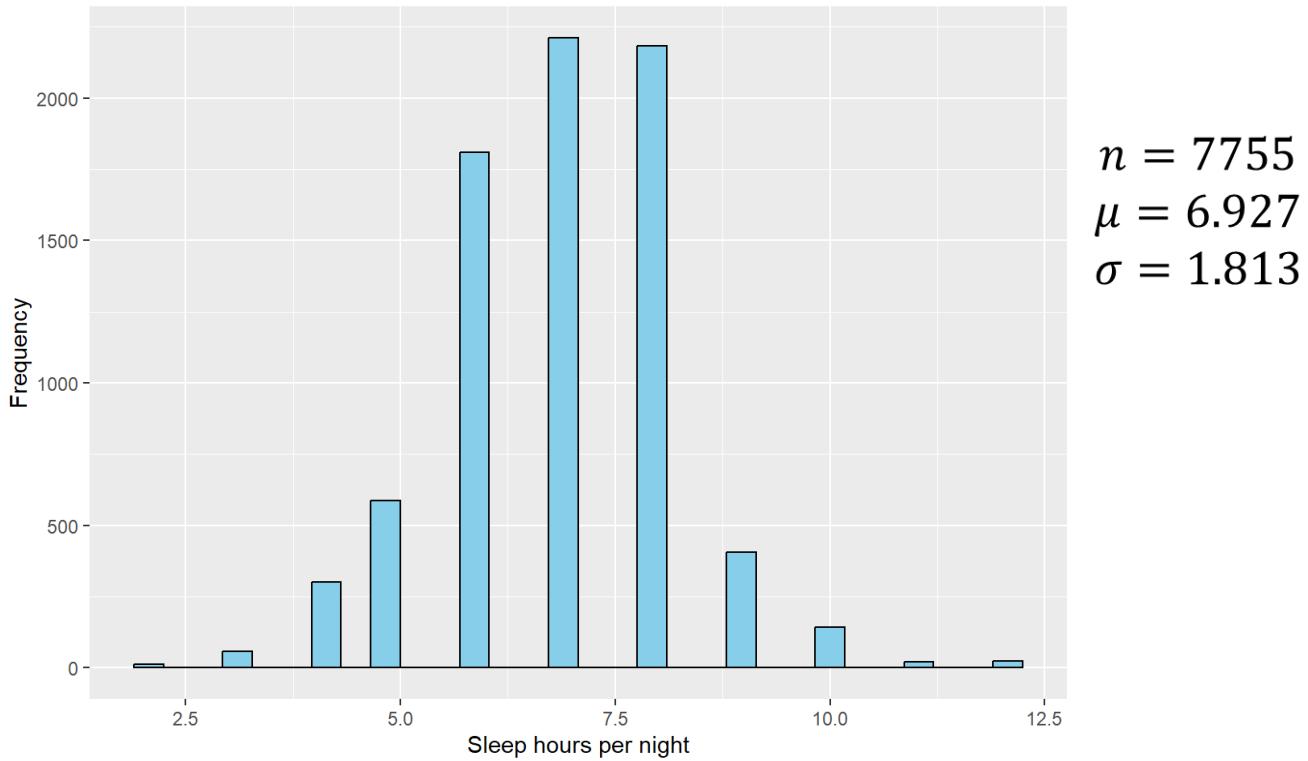
Hide

```
var(sleep)
```

Population variance

```
## [1] 1.81368
```

The number of sleep hours per night in the population



Visualization

1. Variability in estimates

2. Standard error of the mean

3. Confidence intervals

4. Hypothesis testing

The Formal Approach to Hypothesis Testing

Example 1: wind speed in New York 1973

Example 2: The NHANES data set analysis of the number of sleep hours per night

5 Hypothesis testing and confidence intervals

6 Decision error (Type I and Type II error)

Hide

```
ggplot(NHANES, aes(x = SleepHrsNight)) +  
  geom_histogram(fill = "skyblue", color = "black") +  
  ylab("Frequency") +  
  xlab("Sleep hours per night")
```

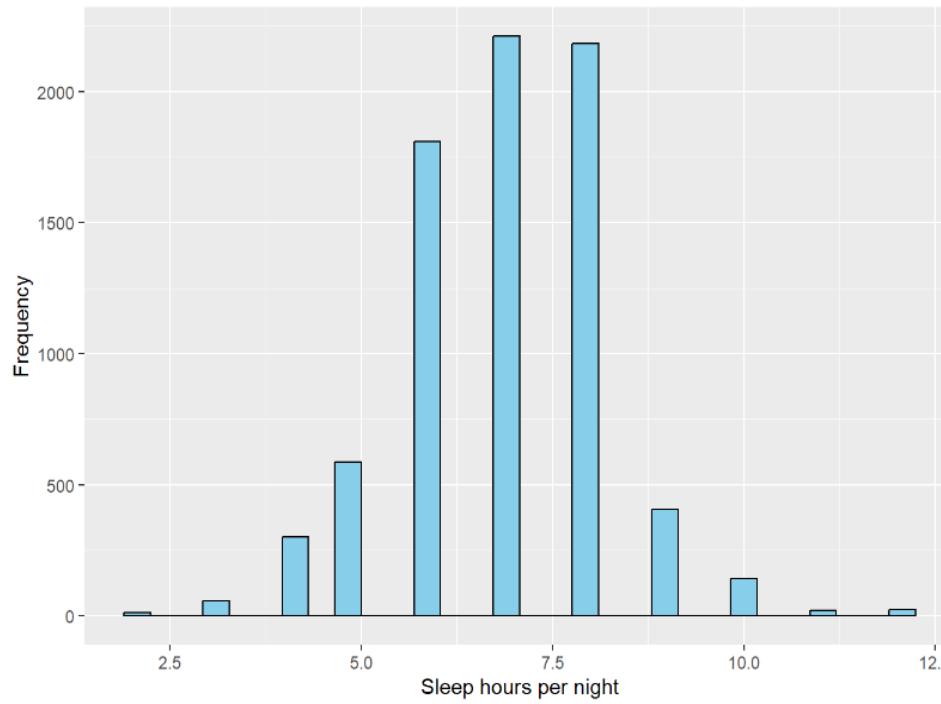


Figure 19: Histogram of sleep hours per night.

Case study: The NHANES dataset

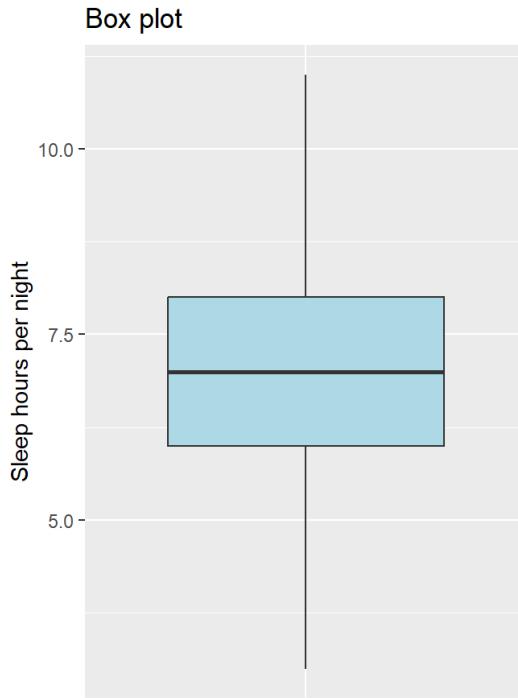
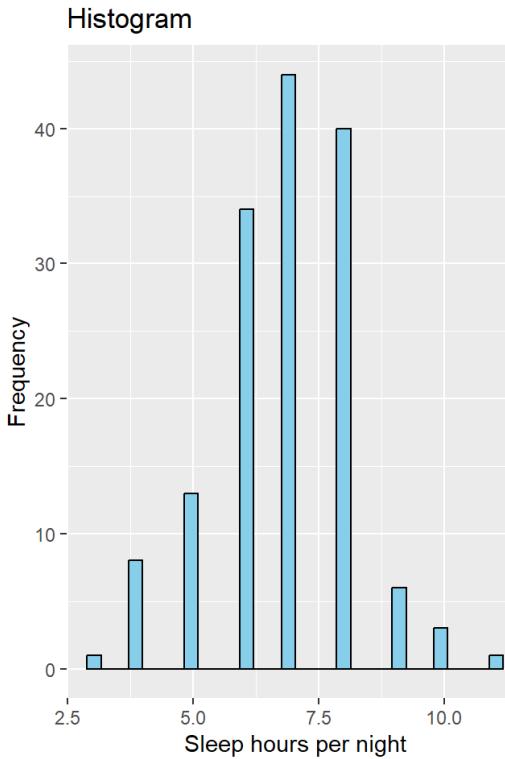
The number of sleep hours per night

Point estimates

A random sample from the population

- Population size: 7755.
- We draw a random sample from the population.
- Sample size: 150.

A random sample from the population



- A random sample from the population:

$$n = 150$$

$$\bar{x} = 6.846$$

$$s^2 = 1.862$$

A random sample from the population

1. Variability in estimates

2. Standard error of the mean

3. Confidence intervals

4. Hypothesis testing

The Formal Approach to Hypothesis Testing

Example 1: wind speed in New York 1973

Example 2: The NHANES data set analysis of the number of sleep hours per night

5 Hypothesis testing and confidence intervals

6 Decision error (Type I and Type II error)

A random sample of size 150 from the population

We draw a sample of 150 individuals from the population ($n = 150$). The point estimates for the sample are $\bar{x} = 6.8466$ and $\sigma^2 = 1.8622$.

Hide

```
set.seed(456789)
x.sleep<-sample(na.omit(NHANES$SleepHrsNight),size=150,replace=FALSE)
length(x.sleep)
```



```
## [1] 150
```

Show

$$\bar{X} = \frac{1}{n} \sum_{i=1}^n X_i$$

```
var(x.sleep)
```

Hide

$$S^2 = \frac{1}{n-1} \sum_{i=1}^n (X_i - \bar{X})^2$$

Sample mean and variance.

A random sample from the population

1. Variability in estimates

2. Standard error of the mean

3. Confidence intervals

4. Hypothesis testing

5 Hypothesis testing and confidence intervals

Two sided alternatives and confidence intervals

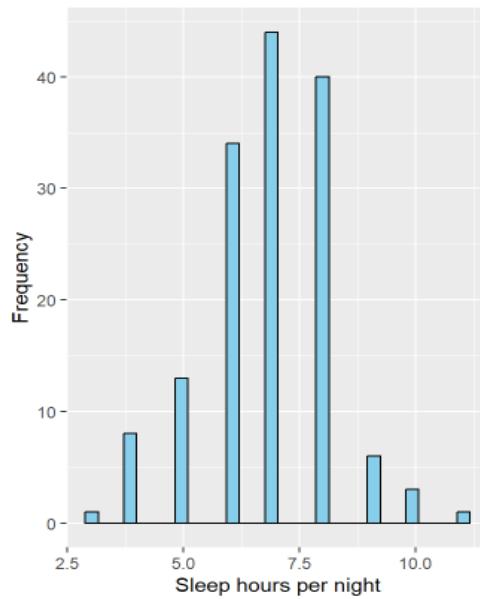
Example: The NHANES dataset - analysis of the total cholesterol level

6 Decision error (Type I and Type II error)

```
box_sleep = ggplot(data.frame(SleepHrsNight = x.sleep), aes(x = "", y = SleepHrsNight)) +  
  geom_boxplot(fill = "lightblue") +  
  xlab("") +  
  ylab("Sleep hours per night") +  
  ggtitle("Box plot")
```

```
grid.arrange(hist_sleep, box_sleep, ncol = 2)
```

Histogram



Box plot

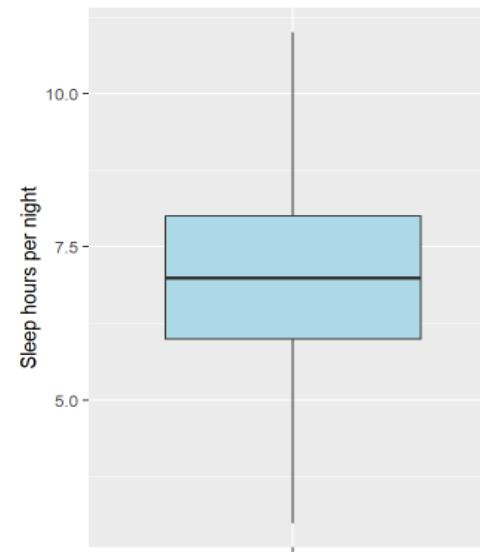


Figure 21: Histogram and box plot of sleep hours per night in the sample.

Case study: The NHANES dataset

The number of sleep hours per night

Confidence interval for the
population mean

Confidence interval for the population mean (Case 2)

If $X \sim F$

Then $\bar{X} \sim N(\mu, \frac{S^2}{n})$
:

and $T_{\bar{X}} = \frac{\bar{X} - \mu}{\sqrt{\frac{S^2}{n}}} \sim N(0,1)$

X has an unknown distribution,
but we have a **large sample**
($n > 30$)

$$E(X) = \mu$$

$$Var(X) = \sigma^2$$

The same as case 1 but we replace σ^2 by S^2 .

C.I. for case 2

Step 1: example, choose $1-\alpha = 0.95$

Step 2: case 2, so :

$$\frac{\bar{X} - \mu}{\sqrt{\frac{\sigma^2}{n}}} \sim N(0,1)$$

or

$$\frac{\bar{X} - \mu}{\sqrt{\frac{s^2}{n}}} \sim N(0,1)$$

Step 3: critical points: -1.96 and 1.96

(the same as in Case 1, since we are still using the
standard normal distribution function)

Step 4: Calculate the point estimator (s) (and possibly s^2)

C.I. for case 2

Step 5: In the same manner as in Case 1:

The $(1-\alpha)$ CI for μ is :

$$\left[\bar{x} - z\sqrt{\frac{\sigma^2}{n}}, \bar{x} + z\sqrt{\frac{\sigma^2}{n}} \right] \quad \text{or} \quad \left[\bar{x} - z\sqrt{\frac{s^2}{n}}, \bar{x} + z\sqrt{\frac{s^2}{n}} \right]$$

Example for case 2

- Suppose $X = \text{number of sleep hours per night.}$
- X has an unknown distribution with unknown variance.
- But large sample ($n = 150 \gg 30$).

The 95% CI for μ : the mean number of sleep hours per night in the population.

Step 1: choose confidence level $1-\alpha = 0.95$

Step 2: case 2, so :
$$\frac{\bar{X} - \mu}{\sqrt{\frac{S^2}{n}}} \sim N(0,1)$$

Step 3: critical points: -1.96 and 1.96

(the same as in Case 1, since we are still using the standard normal distribution function).

Example for case 2

Step 4 : Calculate the point estimators:

$$\bar{x} = 6.8466 \text{ and } s^2 = 1.8622$$

Step 5 : In the same manner as in Case 1:

The $(1-\alpha)$ CI for μ is :

$$\left[\bar{x} - z \sqrt{\frac{s^2}{n}}, \bar{x} + z \sqrt{\frac{s^2}{n}} \right]$$

$$\left[6.8466 - 1.96 \sqrt{\frac{1.8622}{150}}, 6.8466 + 1.96 \sqrt{\frac{1.8622}{150}} \right]$$

$$[6.6283, 7.0650]$$

Example for case 2

- A 95% CI for the population mean μ of the number of sleep hours per night [6.6283, 7.0650]
- **Interpretations:**
 - Based on our sample, we are 95% confident that the true mean of number of sleep hours per night lie between 6.6283 and 7.0650.

A 95% C.I. for the mean sleep hours per night

1. Variability in estimates

2. Standard error of the mean

3. Confidence intervals

4. Hypothesis testing

5 Hypothesis testing and confidence intervals

Two sided alternatives and confidence intervals

Example: The NHANES dataset - analysis of the total cholesterol level

6 Decision error (Type I and Type II error)

A 95% C.I for the mean sleep hours per night

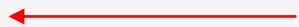
The sample standard deviation and the standard error of the sample mean are equal to 1.3646 and 0.1114, respectively.

Hide

```
n<-length(x.sleep)
SD.x<-sqrt(var(x.sleep))
SD.x
```

```
## [1] 1.364638                      Standard deviation
```

```
SE<-SD.x/sqrt(n)
SE
```



$$SE = \sqrt{\frac{s^2}{n}}$$

```
## [1] 0.1114222                      Standard error of the sample mean
```

For the sample, the error margin for a 95% confidence interval is $m = 1.96 \times SE = 1.96 \times 0.1114$ and the confidence interval is given by

$$\bar{x} \pm m = 6.8466 \pm 0.2183 = (6.628279, 7.065054).$$

Hide

```
LL<-mean(x.sleep)-1.96*SE
UL<-mean(x.sleep)+1.96*SE
c(LL,UL)
```

```
## [1] 6.628279 7.065054
```

A 95% C.I. for the mean sleep hours per night

- A 95% Confidence interval for the population mean using the R function `z.test()`.

1. Variability in estimates

2. Standard error of the mean

3. Confidence intervals

4. Hypothesis testing

5 Hypothesis testing and confidence intervals

Two sided alternatives and confidence intervals

Example: The NHANES dataset - analysis of the total cholesterol level

6 Decision error (Type I and Type II

Hide

```
z.test(x.sleep, sd=SD.x)
```

```
##  
## One Sample z-test  
##  
## data: x.sleep  
## z = 61.448, n = 150.00000, Std. Dev. = 1.36464, Std. Dev. of the sample  
## mean = 0.11142, p-value < 2.2e-16  
## alternative hypothesis: true mean is not equal to 0  
## 95 percent confidence interval:  
## 6.628283 7.065050  
## sample estimates:  
## mean of x.sleep  
## 6.846667
```

Case study: The NHANES dataset

The number of sleep hours per night

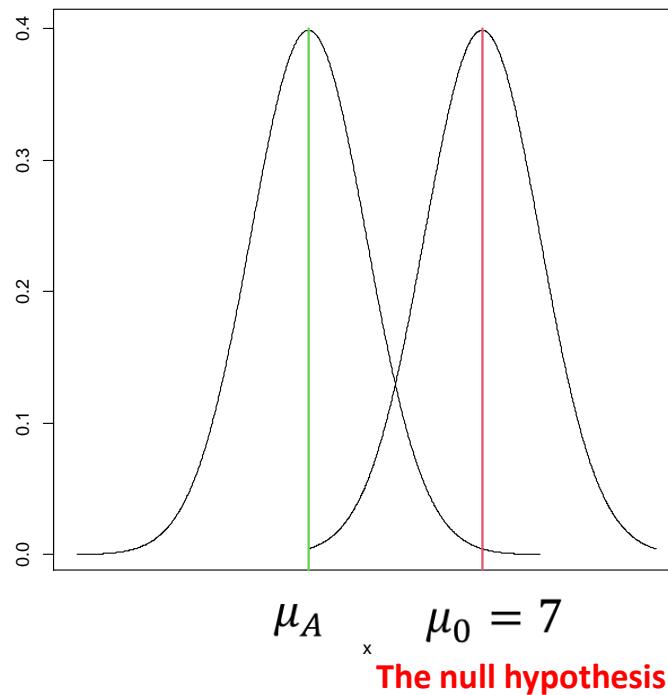
Hypotheses testing

Test of hypothesis: a one sided test

$$H_0: \mu = 7$$

$$H_A: \mu < 7$$

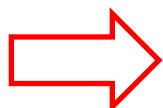
- We test the null hypothesis versus a one sided alternative.
- In our case, under the alternative the mean is smaller than 7 (but not specified).



Test statistic

$$t = \frac{\bar{x} - \mu_0}{\sqrt{\frac{1.3646^2}{150}}} = -1.3761$$

The population variance σ^2 is unknown
but... $n=150$.



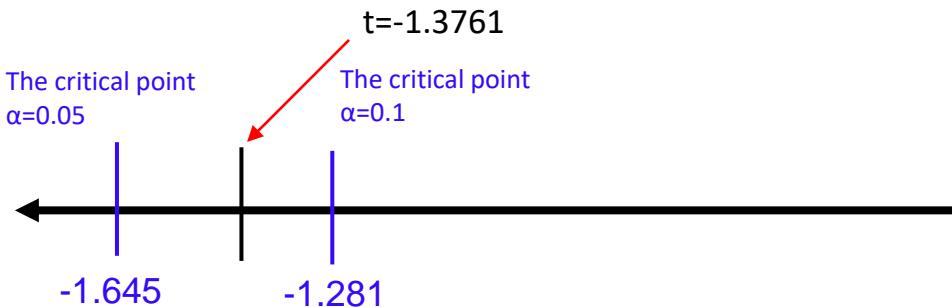
$$\frac{\bar{X} - \mu}{\sqrt{\frac{S^2}{n}}} \sim N(0,1)$$

The critical points and the test statistic

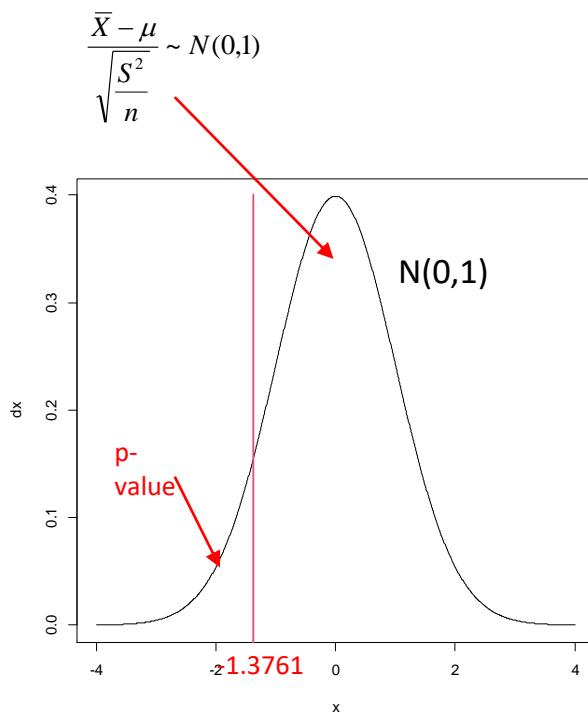
For one sided test and $\alpha=0.05$, $Z=-1.645$.

For one sided test and $\alpha=0.1$, $Z=-1.281$.

For $\alpha=0.1$ We reject $H_0 : -1.3761 < -1.281$.



p-value



$$H_0: \mu = 7$$
$$H_A: \mu < 7$$

$$P(Z < -1.3761) = 0.08439$$

- For $\alpha=0.05$, we DO NOT reject the null hypothesis.
- For $\alpha=0.1$, we reject the null hypothesis.

Hypothesis testing

1. Variability in estimates

2. Standard error of the mean

3. Confidence intervals

4. Hypothesis testing

5 Hypothesis testing and confidence intervals

Two sided alternatives and confidence intervals

Example: The NHANES dataset - analysis of the total cholesterol level

6 Decision error (Type I and Type II error)

Hypothesis testing

We wish to test the null hypothesis $\mu = 7$ against a one sided alternative $H_1 : \mu < 7$. This can be done using the argument `alternative = 'less'` in the function `z.test`. Note that we assume that in the population, $\sigma = 1.3646$. As can be seen in the panel below, for the sample, the mean number of sleeping hours is equal to $\bar{x} = 6.8466$ and the test statistic is equal to -1.3761 . The $p=0.08439 > 0.05$. We cannot reject the null hypothesis and conclude that $\mu = 7$.

[Hide](#)

```
mean(x.sleep)
```

[Hide](#)

```
## [1] 6.846667
```

[Hide](#)

```
sqrt(var(x.sleep))
```

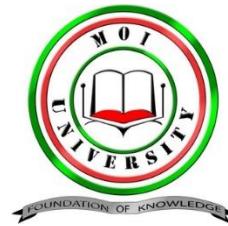
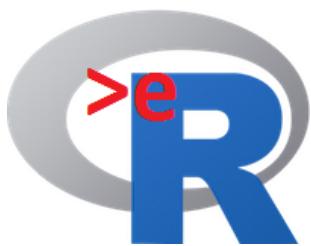
[Hide](#)

```
## [1] 1.364638
```

[Hide](#)

```
z.test(x.sleep,mu=7, 1.364638, alternative = 'less')
```

```
##  
## One Sample z-test  
##  
## data: x.sleep  
## z = -1.3761, n = 150.00000, Std. Dev. = 1.36464, Std. Dev. of the  
## sample mean = 0.11147, p-value = 0.08439  
## alternative hypothesis: true mean is less than 7  
## 95 percent confidence interval:  
##      -Inf 7.02994  
## sample estimates:  
## mean of x.sleep  
##                 6.846667
```



Interuniversity Institute for Biostatistics
and statistical Bioinformatics

Development of E-learning materials using R markdown

Part 5

The course online (1)

Steps in data analysis

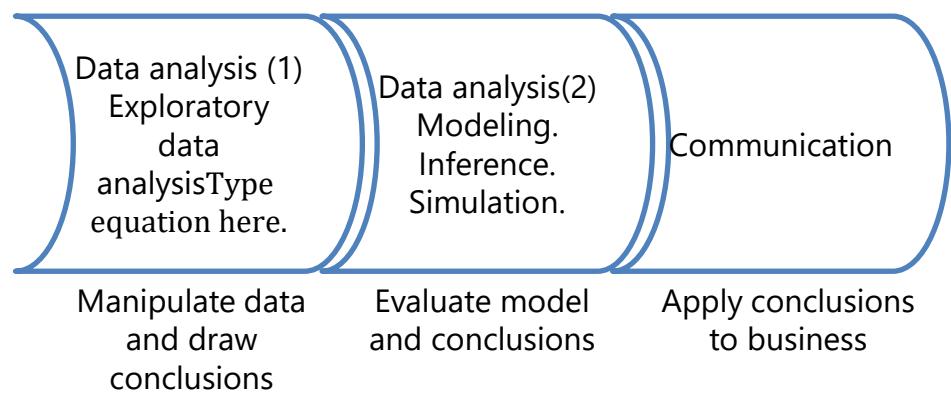
Modeling the association between the fuel consumption and the car's weight.

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

Methodology: simple linear regression.



We “translate” the methodology to software usage



Boxplot by treatment group.

A simple linear regression.

A report.

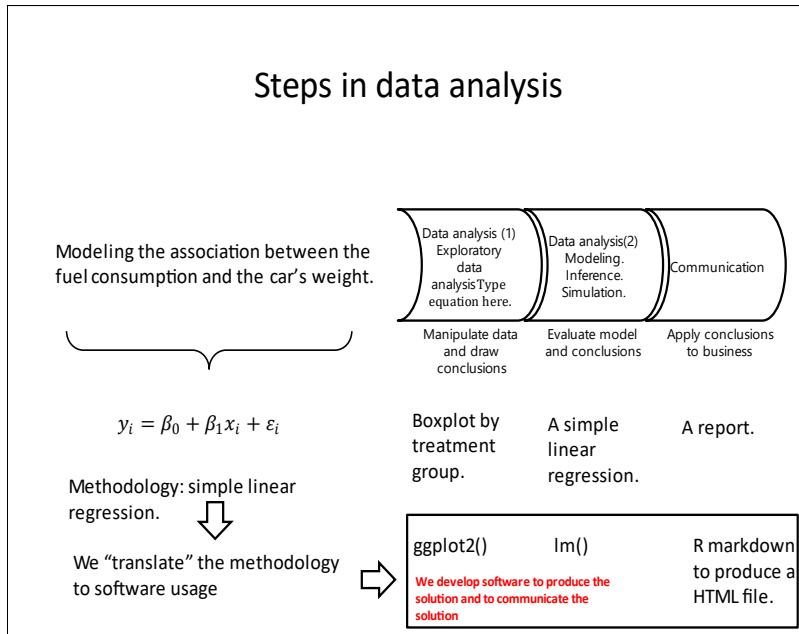
ggplot2()

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

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

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

>eR-Biostat website.

List of courses.

The screenshot shows a list of courses categorized into three sections: Introductory, Advanced, and Basic. Each category has a corresponding color-coded background: green for Introductory, orange for Advanced, and blue for Basic. The courses listed are:

Introductory	Advanced	Basic
Introduction to R	Applied Generalized Linear Models (GLM) using R	Basic concept in statistical inference using R (1)
Statistical modeling: Linear regression using R	Modeling Binary Data using R	Basic concept in statistical inference using R (2)
Statistical modeling: One-way ANOVA using R	Longitudinal data analysis (LDA) using R	Linear Regression using R
Statistical modeling: Logistic regression using R	Univariate models using R	
Vizualizing data using R: an introduction	Survival Analysis using R	
Basic concepts of statistical inference using R	An introduction to bootstrap using R	
	Sample size calculation using R	
	Exploratory multivariate data analysis using R	
	Survival Analysis using R (A)	

A red arrow points from the left side of the slide towards the 'Introductory' section, and another red arrow points from the right side towards the 'Advanced' section.

Introduction to statistical modeling using R

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

The figure is a boxplot titled "Figure 3.1: Chick weight by diet group". The y-axis lists six diet groups: sunflower, soybean, meatmeal, linseed, horsebean, and casein. The x-axis represents weight, ranging from 100 to 400. The horsebean group has a significantly lower median weight than the other groups.

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

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

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

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

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

About

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

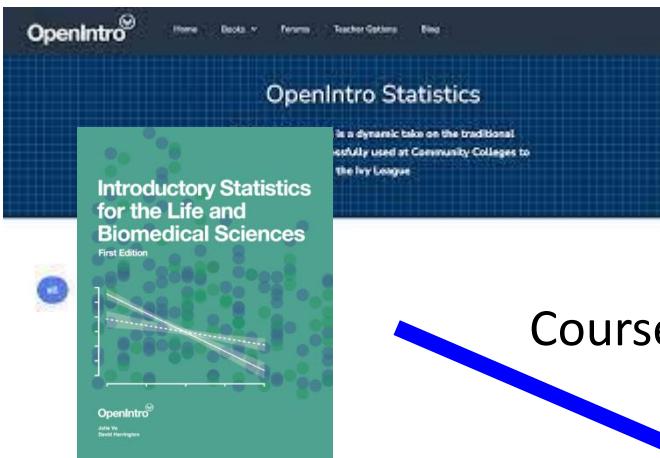
Topics

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

Online tutorials

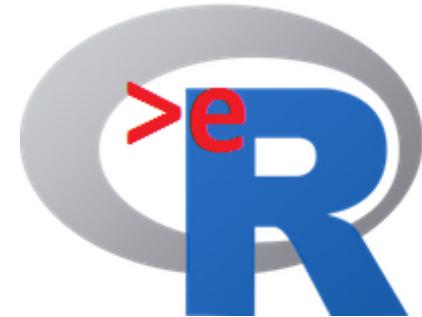
In this page, supporting online tutorials are given in different formats. The online

Introduction to statistical modeling using R



Course II

Course materials from two sources



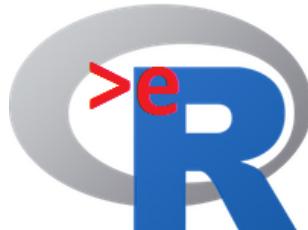
Course I

A screenshot of the >eR-BioStat website. The header includes the R logo and the text 'Introduction to Statistical modeling using R'. A red circle highlights the 'Topics' menu item. Below the header, there's a list of topics: Simple linear regression using R, One-way ANOVA using R, and Logistic Regression using R. A note states that the courses can be given together as three parts of a course or separately as individual modules or as a part of a specific course in statistical modeling. Another note indicates the courses were developed at an undergraduate level. A statement says the course is open source and source files are available online. The 'Topics' section shows a thumbnail of a scatter plot with regression lines. The footer contains links for 'About', 'Topics', and 'Online tutorials', along with copyright information and a search bar.

Course I

The screenshot shows a web browser window with the URL erbiostat.wixsite.com/introstatmod/topics. The page title is "Topics | IntroStatMod". A banner at the top right says "External datasets for illustration are included in the data repositories." Below the banner, there are three main sections: "Simple Linear Regression", "One-Way ANOVA", and "Simple Logistic Regression". Each section has a brief description and a list of topics. At the bottom of each section is a green button labeled "Slides (PDF)" followed by other links for "Slides (PP)", "R programm", and "Datasets". A red arrow points from the text "examples in the course." to the "Slides (PDF)" button in the Simple Linear Regression section.

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



Course materials:

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

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">• Introduction and model formulation.• Fitting a simple linear regression model using the lm() function in R.• Model diagnostic.• Model diagnostic in R. <p>External datasets are available in the data repository.</p> <p>Slides (PDF): simple linear regression</p> <p>Slides (PP): simple linear regression</p> <p>R programm</p> <p>Datasets</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">• 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. <p>External datasets are available in the data repository.</p> <p>Slides (PDF): One-Way ANOVA</p> <p>Slides (PP): One-Way ANOVA</p> <p>R programm</p> <p>Datasets</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">• Introduction and example tour.• Fitting a simple linear logistic regression model using the glm() function in R.• Model formulation.• Interpretation of the model parameters. <p>External datasets are available in the data repository.</p> <p>Slides (PDF): Logistic regression</p> <p>Slides (PP): Logistic regression</p> <p>R programm</p> <p>Datasets</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 document is hosted on GitHub and is 1.21 MB in size. The GitHub Copilot logo is visible at the top of the page.

The slide content includes:

- The >eR initiative**
- Making R based education materials in statistics accessible for all**
- Basic concepts in statistical modeling using R:
simple linear regression**
- Developed by**
Legesse Kassa Debusho (UNISA, South Africa) and Ziv Shkedy (Hasselt University)
- https://erbiostat.wixsite.com/erbiostat**
- LAST UPDATED: 2022**
- Social media links: Facebook, GitHub, Twitter
- Email: erbiostat@gmail.com
- Page number: 2

The browser interface shows multiple tabs open, including "Home | Erbiostat", "Topics | IntroStatMod", and "Courses/introductory Courses/". The left sidebar of the browser displays a file tree for the GitHub repository, showing various course materials like "Basic courses", "Coordination", "Data Analysis", etc., and specific files like "README.md", "Temp_2024_prog1_V2.html", and "eR-Biostat_Introduction to ...".

Course materials

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.

- 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.

Course I

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

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

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

Course materials

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

uhasselt.be bookmarks All Bookmarks

OpenIntro Home Books Forum Teacher Details Blog

OpenIntro Statistics

OpenIntro Statistics is a dynamic take on the traditional introductory statistics textbook. It is used at Community Colleges to IV League

Introductory Statistics for the Life and Biomedical Sciences First Edition

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

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

Simple linear regression

Course II

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

Type here to search

13:02 7/02/2024 ENG 90

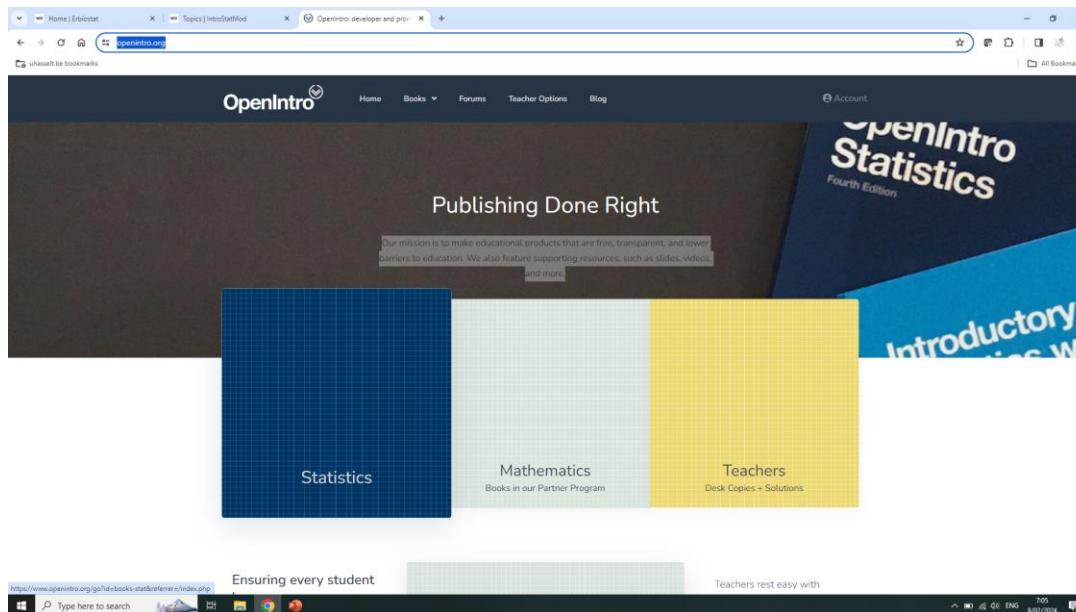
openintro.org

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

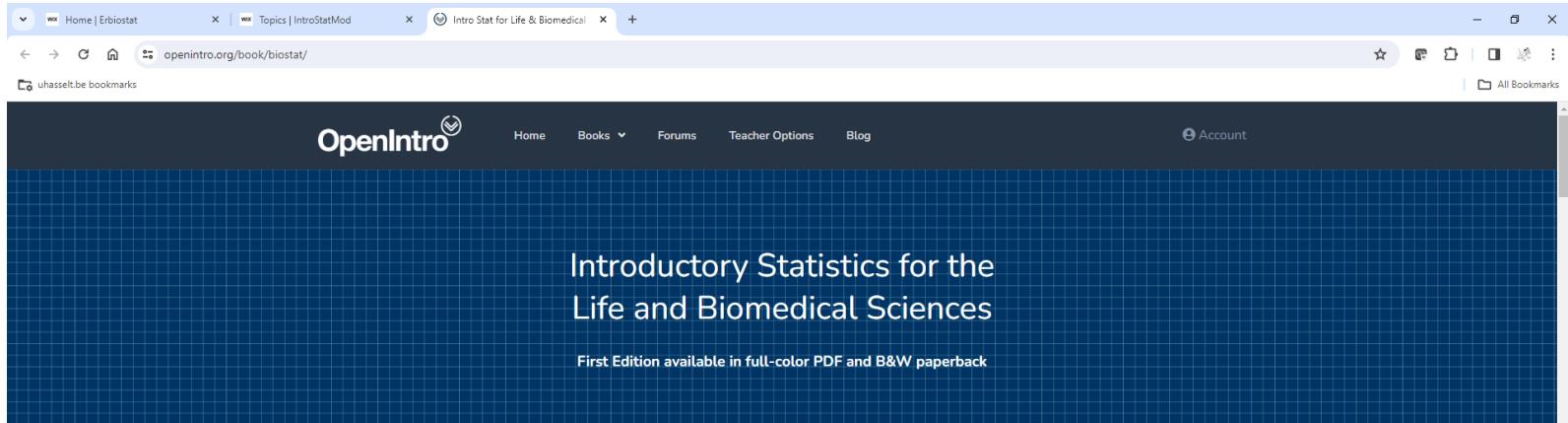
openintro.org

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

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



Introductory Statistics for the Life and Biomedical Sciences



Textbook Pedagogy

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

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

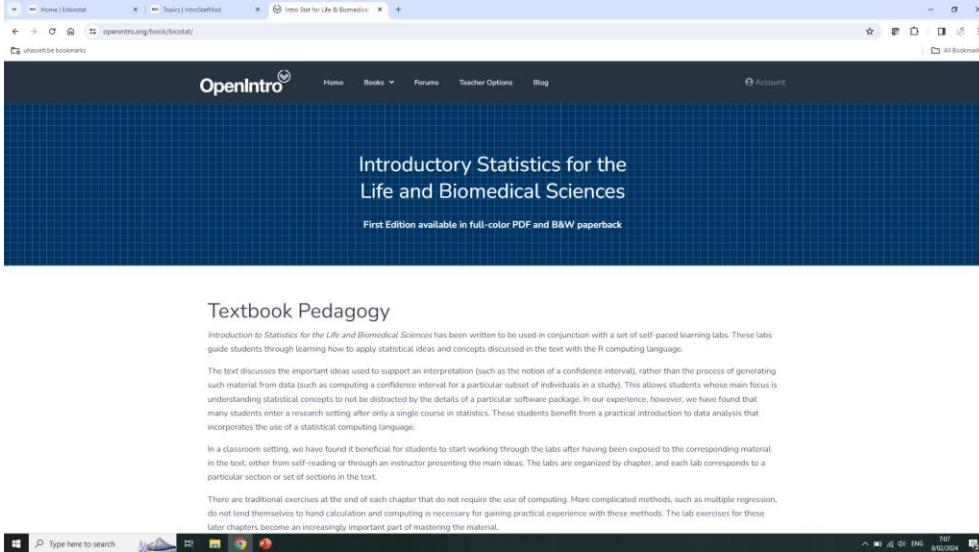
In a classroom setting, we have found it beneficial for students to start working through the labs after having been exposed to the corresponding material in the text, either from self-reading or through an instructor presenting the main ideas. The labs are organized by chapter, and each lab corresponds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

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



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

Introductory Statistics for the Life and Biomedical Sciences



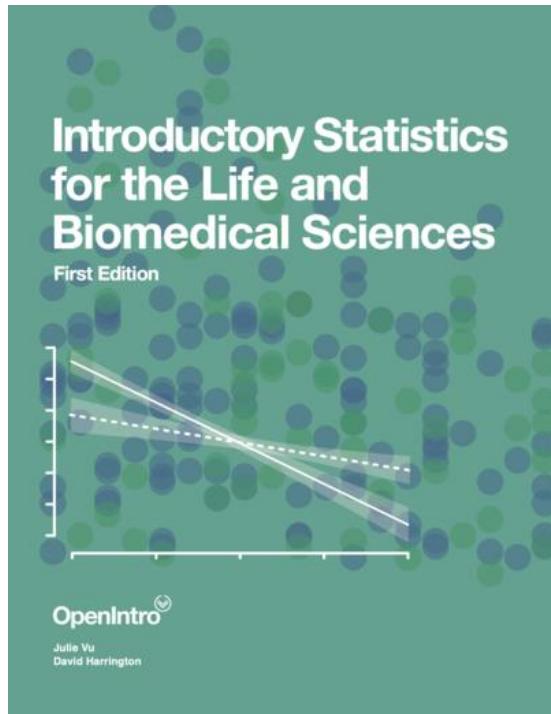
OpenIntro Biostatistics:
developing open source materials
for teaching and learning applied
statistics with R.

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

Developed for students in health or life sciences:

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

Introductory Statistics for the Life and Biomedical Sciences



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

Reference for Course I & II

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

Reference for course I & II

Course I

Course II

Course III

Online book

Online book

Online book

Simple linear regression

One-Way ANOVA

Logistic regression

ANOVA

Start Now

WIX.com

Home | Erbiostat

Topics | IntroStatMod

erbiostat.wixsite.com/introstatmod/topics

uhasselt.be bookmarks

All Bookmarks

Course II: Slides

Home | Eriostat Topics | IntroStatMod Courses/Introductory Courses/

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

uhasselt.be bookmarks All Bookmarks

Files

master + 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

Course II: The Rmd file for the slides

The screenshot shows a GitHub repository interface for the 'eR-Biostat' organization. The repository is named 'Courses / Introductory Courses / Introduction to statistical modeling using R / Online materials / unit_06_simple_linear_regression.Rmd'. The code editor displays the Rmd file content, which includes YAML front matter and R code for generating a presentation.

```
title: "Unit 6: Simple Linear Regression"
author: "Statistics 102 Teaching Team"
date: "March 30, 2020"
output:
  beamer_presentation:
    includes:
      in_header: ./slides_header.tex
    fig_width: 3.25
    fig_height: 3
    fig_caption: false
    toc: true
    keep_tex: true
    classoption: "aspectratio=169"
    slide_level: 3
  ...
# Introduction
## The main ideas
\small
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.
```

Course II: A PDF output

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

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

The main content area displays a PDF titled "Courses / Introductory Courses / Introduction to statistical modeling using R / Online materials / unit_06_simple_linear_regression.pdf". The PDF page number is 3 / 40.

THE MAIN IDEAS

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

$$y = b_0 + b_1x$$

THE MAIN IDEAS . . .

Multiple linear regression is used to estimate the linear relationship between a

The left sidebar shows a file tree for the "master" branch of the repository:

- 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_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...

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

wix Home | Erbiostat wix Topics | IntroStatMod

erbiostat.wixsite.com/introstatmod/topics

uhasselt.be bookmarks

This site was designed with the WIX.com website builder. Create your website today. Start Now

Online book

Online book

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

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

Slides (PDF): simple linear regression

Slides (PP): simple linear regression

Slides (Rmd): simple linear regression

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

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

Slides (PDF): One-Way ANOVA

Slides (PP): One-Way ANOVA

Slides (Rmd): One-Way ANOVA

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

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

Slides (PDF): Logistic regression

Slides (PP): Logistic regression

Slides (Rmd): Logistic regression

Written by Dave H. and Jullie V.

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

Simple linear regression

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

A part of the website of the course

wx Home | Eriostat wx Topics | IntroStatMod Temp_2024_prog1_V2.knit

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 101

A part of the website of the course

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

library(DT)

1. General Introduction

Linear regression models

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

$$y = b_0 + b_1 x$$

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

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

Examining scatterplots

The mtcars dataset

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

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

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

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

Type here to search



Interuniversity Institute for Biostatistics
and statistical Bioinformatics

Development of E-learning materials using R markdown

Part 6: 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.

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

Introduction to Statistical modeling using R

Home About Topics Online tutorials Contact

>eR-BioStat

This group of courses is focused on statistical modeling 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 part can be given separately as a part of a specific course in statistical modeling.

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

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

About **Topics** **Online tutorials**

General information about the course and course materials and the study materials in the course

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

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

Home | Edit control | Home | IntroStatMod | +
erBiostat.wixsite.com/introstatmod
This site was designed with the WIX.com website builder. Create your website today. Start Now

Temp_2024/prog4_v2.Rmd | Temp_2024/prog4_v2.html
Unknown bookmarks

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 a statistical method for examining the 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:

$$y = \hat{y}_0 + \hat{y}_1 x$$

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

$$y = \hat{y}_0 + \hat{y}_1 x_1 + \hat{y}_2 x_2 + \dots + \hat{y}_p x_p$$

Examining scatterplots

The **mtcars** dataset

The Motor Trend Car Road Test 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 `head(mtcars)` to get more information about `mtcars`.

```
## [1] 32 11  
head(mtcars)  
#> #>   mpg cyl disp  hp drwt wt gear carb  
#> mazda RX4 21.0 6 160 120 3.90 2.620 4 4 4  
#> mazda RX4 Wag 21.0 6 160 120 3.90 2.875 4 4 4  
#> datsun 710 21.4 6 108 93 3.85 2.320 4 1 1  
#> hornet 4 Drive 21.4 8 250 120 3.08 3.215 3 1 1  
#> hornet Sportabout 18.7 8 250 120 3.08 3.440 3 1 2  
#> valiant 18.1 8 225 120 3.08 3.440 3 1 1
```

Can be produced using the Rmd program:
`er_prog4_SA_2024.Rmd`

The Rmd program

Document setting

```
1 ---  
2 output:  
3   bookdown::html_document2:  
4     toc: TRUE  
5     toc_float: TRUE  
6     toc_depth: 2  
7     number_sections: no  
8     css: ./lib/stylesArial.css  
9     code_folding: hide  
10  
11 params:  
12   department: ">eR-Biostat"  
13   topic: <font size = "10" > **simple linear regression using R **</font>  
14   author: "Ziv Shkedy et al"  
15   date: "15-12-2023"  
16   endCode: FALSE  
17   RmdLocation: ""  
18 ---  
19  
20 <p>  
21     
22 </p>  
23  
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     print(delay_code_labels)  
40   }  
41 }  
42 }```
```

Environment is empty

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.

The HTML output

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

1. General Introduction

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

15-12-2023 >eR-BioStat

Simple linear regression using R

Ziv Shkedy et al Show

1. General Introduction

Linear regression models

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

$$y = b_0 + b_1 x$$

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

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

Examining scatterplots

The mtcars dataset

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

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

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

The Rmd program

RStudio

File Edit Code View Plots Session Build Debug Profile Tools Help

unit_05_inference_num_1ANOVA.Rmd x er_prog3_SA_2024.Rmd x er_prog4_SA_2024.Rmd x The pharma challenge_2022_prog1.Rmd x er_prog2_SA_2024.Rmd x

Go to file/function Addins x

Source Visual

```
136
137 # 1. General Introduction
138
139 ## Linear regression models
140
141 Linear regression provides methods for examining the association between a quantitative response variable and a set of possible predictor variables.
142 Linear regression should only be used with data that exhibit linear or approximately linear relationships. **Simple linear regression** is used to estimate the linear relationship between a response variable $y$ and a single predictor $x$. The response variable $y$ can be referred to as the "dependent" variable, and the predictor variable $x$ the "independent" variable. The statistical model for simple linear regression is based on the straight line relationship
143 
$$y = b_0 + b_1 x \quad \text{newine}$$

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 `<tt>mtcars</tt>`. The dataset contains information about 11 variables and 32 cars. Use `<tt>help(mtcars)</tt>` 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 `<tt>mpg</tt>`) and the car's weight (in 1000 lbs), the R object `<tt>wt</tt>`.

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.

7:25 8/02/2024 ENG

Environment History Connections Tutorial

Import Dataset 156 MB Global Environment

• Section

Environment is empty

• Subsection

• Free text...

Files Plots Packages Help Viewer Presentation

Zoom Export

Type here to search

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.

```
## [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 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   4 121 109 4.08 3.210 19.80  1  0     3     1
## Hornet Sportabout 18.7   6 143 108 4.93 3.440 17.00  0  1     3     2
## Valiant    18.1   6 143 108 4.93 3.440 17.00  0  1     3     2
## Fiat 128    29.0   4  78  66 4.08 2.200 18.60  1  0     4     0
## Fiat 130    15.8   8 120 108 4.93 3.330 18.00  0  1     5     1
## Fiat 150L   15.0   8 120 108 4.93 3.330 18.00  0  1     5     1
## Fiat 150    15.2   8 120 108 4.93 3.330 18.00  0  1     5     1
## Dodge Dart 17.8   8 120 108 4.93 3.330 18.00  0  1     5     1
## Hornet 2000G 19.2   8 143 108 4.93 3.330 18.00  0  1     5     1
## Hornet 2000T 17.8   8 143 108 4.93 3.330 18.00  0  1     5     1
## Toyota Corolla 17.4   4  76  62 4.08 2.320 18.60  1  0     4     0
## Toyota Corona 17.3   4  78  66 4.08 2.320 18.60  1  0     4     0
## Dodge Challenger 15.2   8 143 108 4.93 3.330 18.00  0  1     5     1
## AMC Javelin 14.3   8 143 108 4.93 3.330 18.00  0  1     5     1
## Camaro Z28 13.9   8 143 108 4.93 3.330 18.00  0  1     5     1
## Pontiac Firebird 14.3   8 143 108 4.93 3.330 18.00  0  1     5     1
## Fiat X1-9 27.3   4  66  91 4.08 1.935 18.90  1  0     4     0
## Porsche 914-2 26.0   4  95 110 4.93 2.835 18.00  1  0     4     0
## Lotus Europa 21.4   4  95 110 4.93 2.835 18.00  1  0     4     0
## Ford Pantera L 15.8   8 143 108 4.93 3.330 18.00  0  1     5     1
## Ferrari Dino 19.7   6 143 108 4.93 3.330 18.00  0  1     5     1
## Maserati Bora 15.0   8 143 108 4.93 3.330 18.00  0  1     5     1
## Volvo 142D 21.4   4  95 110 4.93 2.835 18.00  1  0     4     0
```

The Rmd program

RStudio
File Edit Code View Plots Session Build Debug Profile Tools Help
File Edit Code View Plots Session Build Debug Profile Tools Help
OVARmd 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.R
Go to file/function Addins
Source Visual
1 The motor trend car 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 <tt>mtcars</tt>. The dataset contains information about 11 variables and 32 cars. Use <tt>help(mtcars)</tt> to get more information about the data.
152
153 `r, echo=TRUE, message=FALSE, warning=FALSE}`
154 `dim(mtcars)`
155 `head(mtcars)`
156 `...`
157
158 `## Miles/(us) gallon vs. the car's weight`
159
160 Our aim is to investigate the relationship between the fuel consumption (in Miles/(us) gallon, the R object <tt>mpg</tt>) and the car's weight (in (1000 lbs), the R object <tt>wt</tt>).
161
162 `r, echo=TRUE, message=FALSE, warning=FALSE, fig.cap="Mpg vs. weight"`
163 `#plot(mtcars$wt, mtcars$mpg, ylab = "mpg", xlab = "weight (1000 lbs)")`
164 `plot(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
159:1 Miles/(US) gallon vs. the car's Weight
Console Terminal Background Jobs
R 4.3.2 · C:\projects\Teaching2021\Rcourse\R_2223\Classes\The pharma study
R is free software and comes with ABSOLUTELY NO WARRANTY.
You are welcome to redistribute it under certain conditions.
Type 'license()' or 'licence()' for distribution details.
R is a collaborative project with many contributors.
Type 'contributors()' for more information and
'citation()' on how to cite R or R packages in publications.
Type 'demo()' for some demos, 'help()' for on-line help, or
'help.start()' for an HTML browser interface to help.
Type 'q()' to quit R.
Project: (None)
Environment History Connections Tutorial
Import Dataset 146 MB Global Environment
Environment is empty
• Analysis in R:
• Scatterplot + correlation.

The HTML output

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

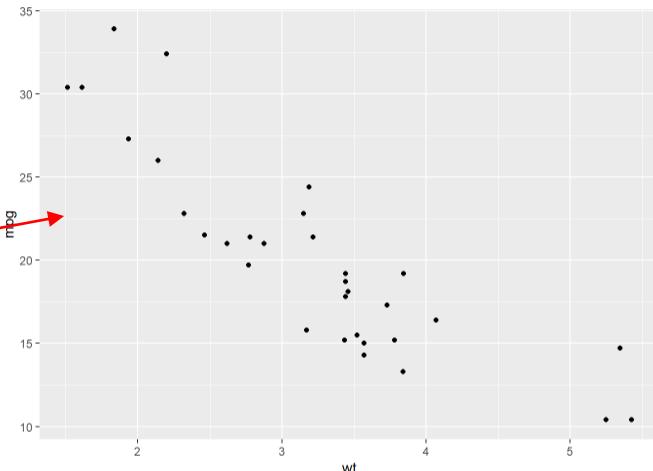
1. General Introduction
Linear regression models
Examining scatterplots
2. The least squares regression model
3. Simple linear regression using R
3. Model diagnostic
4. Categorical predictors with two levels
5. Goodness of fit: using R^2 to describe the strength of a fit
6. Statistical inference in regression

```
## 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`.

Show



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 correlation, with fuel efficiency decreasing as weight increases.

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

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

Windows Taskbar: Type here to search, File, Start, Google Chrome, Microsoft Edge, R, File, Print, 7:26, ENG, 8/02/2024

Text without the code.

Red arrows point from the text "Text without the code." to the "Show" button above the scatterplot and to the correlation coefficient in the bottom left box.

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)
```

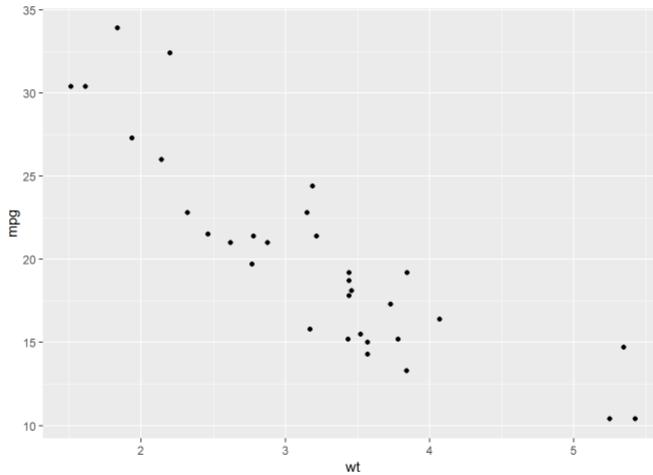


Figure 1: mpg vs. weight

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

The HTML output

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

3. Simple linear regression using R

The lm() R function

For the `mtcars` dataset, we consider the model

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

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

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

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

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

Data and estimated model

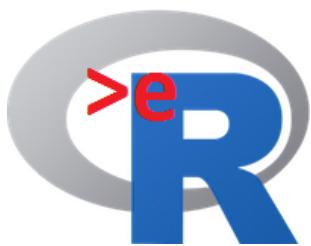
Figure 2 shows the data (mpg vs. weight) and fitted regression line, $mpg_i = 37.28 - 5.34 \times wt_i$

Free text about regression.

- The regression model:
 - Code.
 - Output.

Short summary

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



Interuniversity Institute for Biostatistics
and statistical Bioinformatics

Development of E-learning materials using R markdown

The >eR-BioStat approach

How can we create a course online??

The >eR-BioStat approach

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



All have publicly available
and free versions.

From a laptop to a website..

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

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

## The lm() R function

For the mtcars dataset, we consider the model


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


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

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

## Call:
## lm(formula = mpg ~ weight, data = mtcars)
##
## Residuals:
##   Min   1Q   Median   3Q   Max
## -10.432 -2.587 -0.502 14.976 4.187
##
## Coefficients:
##   Estimate Std. Error t value Pr(>|t|)
## (Intercept) 37.885 1.8776 19.955 < 2e-16 ***
## weight      -0.4445 0.6994 -0.639 0.26842
## ---
## 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: 0.7520, Adjusted R-squared: 0.7464
## F-statistic: 31.93 on 1 and 30 DF, p-value: 1.299e-12
```

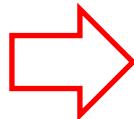
The parameter estimates for the intercept and slope are equal, respectively, to $\hat{\beta}_0 = 37.88$ and $\hat{\beta}_1 = -0.44$.

Data and estimated model

Figure 2 shows the data (mpg vs. weight) and fitted regression line. $mpg = 37.88 - 0.44 \times weight$

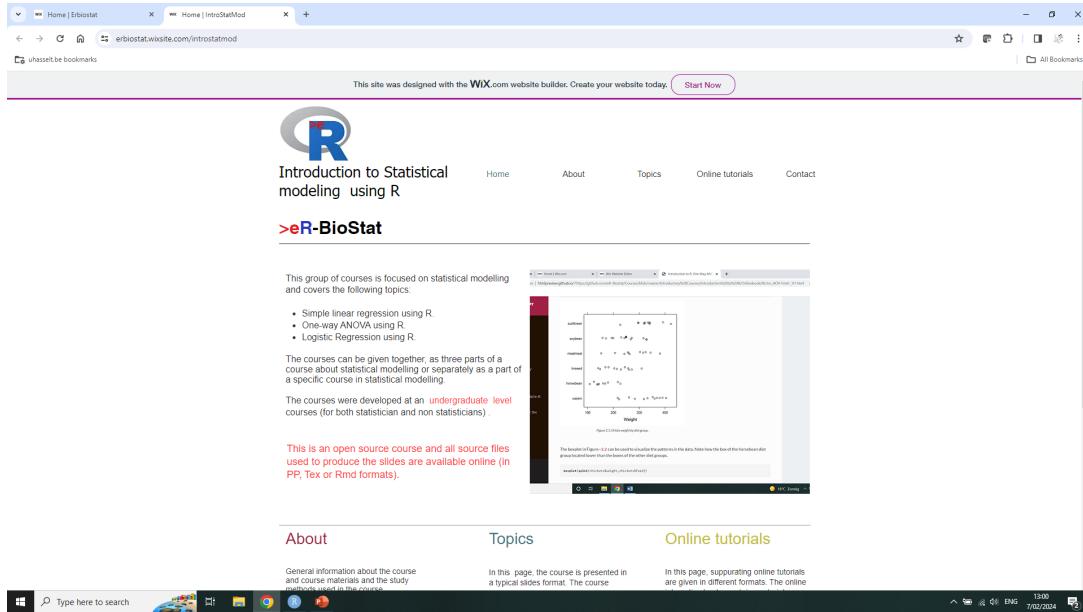
The screenshot shows a Windows desktop with a browser window open to a GitHub page. The URL is "https://github.com/ab-Bisrat/Courses/blob/master/Introductory%20to%20Statistical%20Modeling%20using%20R/Simple%20Linear%20Regression/slides/Temp_2024_prog1_V2.html". The page content is identical to the one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

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



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

Introduction to statistical modeling using R



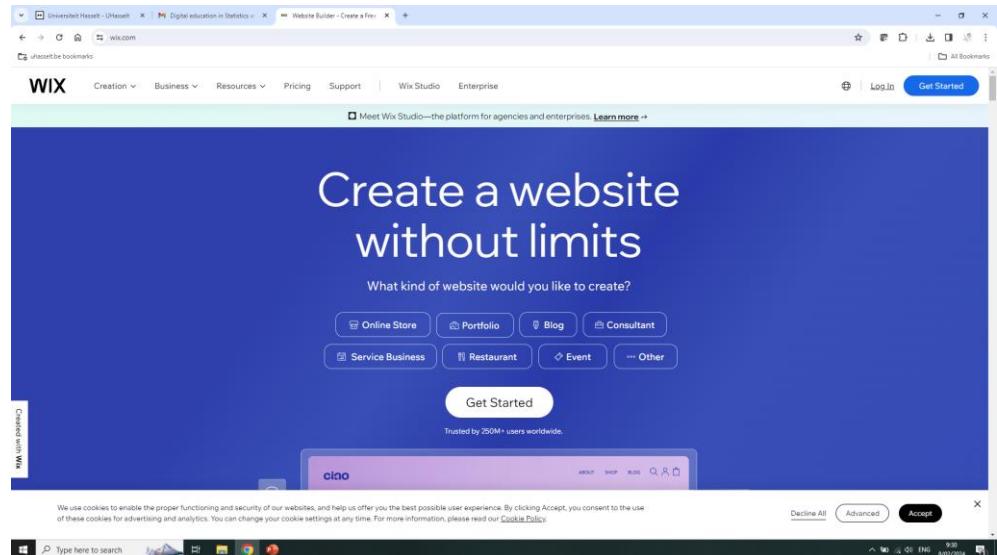
- Website.
- Storage of materials:
 - Slides.
 - Programs.
 - Datasets.
 - Etc.
- Software for the analysis ?
- Storage space & cost ?

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

- Our approach: bring costs to zero...

How can I develop a website for my course ?

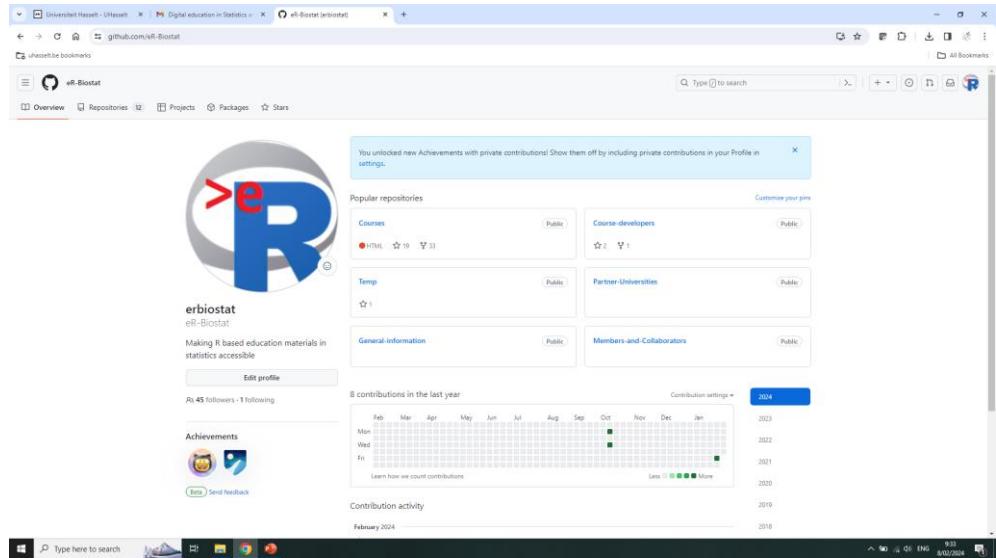
- Websites for the courses were developed using WIX.
 - Free.
 - Easy to use and learn.



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

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

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



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

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

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

uhasselt.be bookmarks

Files

master

Go to file

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

Name Last commit message Last commit date

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

README.md

The >eR-Biostat initiative

Introduction to statistical modeling in R

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

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

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

Type here to search

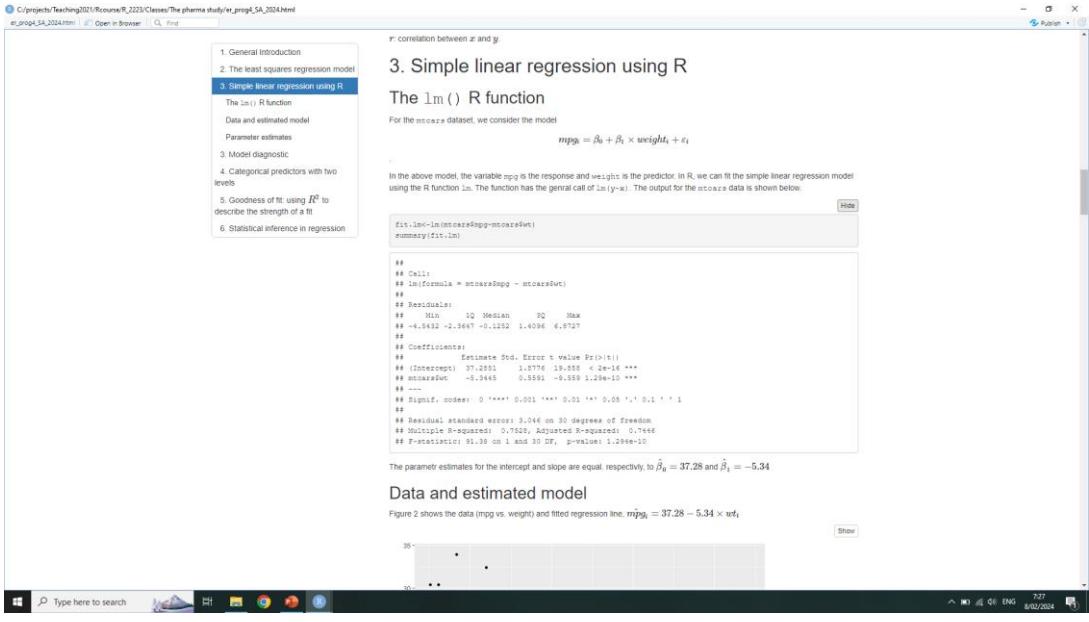
9:36 8/02/2024 ENG

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

Software for the analysis

- We use R but.....

Summary



The screenshot shows a Windows desktop with a browser window open. The address bar indicates the URL is `C:/projects/Teaching2021/Rcourse/R_2223/Classes/The pharma study/r_progr_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 x and y  
  
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 -0.1252  1.4098  6.8727  
##  
## Coefficients:  
##             Estimate Std. Error t value Pr(>|t|)  
## (Intercept) 37.8785  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: 91.38 on 1 and 30 DF, p-value: 1.29e-10
```

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

Data and estimated model

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

- 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.
 - HTML files.

Users

21/02/23-20/02/24

Traffic Overview | Wix.com

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

uhasselt.be bookmarks All Bookmarks

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

Let's set up your business 1/4 completed

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

Traffic Overview

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

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

Sessions over time Sessions Jul 23, 2023 0

New vs returning visitors Unique visitors 69 New 97% * 67 Returning 3% * 2

Sessions by device Site sessions 109 Desktop 94% * 102 Mobile 6% * 7

Sessions by traffic source Direct 94 Facebook 15 Unknown 0 Google.com Get traffic Wix email marketing Get traffic See Full Report

Avg. sessions by day Sun Mon Tue Wed Thu Fri Sat

Type here to search

19:55 ENG 20/02/2024

The screenshot shows the Wix Traffic Overview dashboard. A red oval highlights the top section displaying 'Site sessions' (109) and 'Unique visitors' (69). Below this is a line chart titled 'Sessions over time' showing fluctuations in daily sessions from February 21 to January 28. To the right are two circular summary cards: 'New vs returning visitors' (69 unique visitors, 97% new, 3% returning) and 'Sessions by device' (109 total, 94% desktop, 6% mobile). Further down are sections for 'Sessions by traffic source' (Direct, Facebook, Unknown, Google.com, Wix email marketing) and 'Avg. sessions by day' (a bar chart showing sessions per day from Sunday to Saturday).

Users

21/02/23-20/02/24

Traffic Overview | Wix.com

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

uhasselt.be bookmarks All Bookmarks

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

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

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

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

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

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

See Full Report See Full Report

Traffic insights See All Insights

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

1 2 3 >

19:57 20/02/2024 ENG

The screenshot shows the Wix Analytics dashboard for a site named 'uhasselt.be'. The main header includes the title 'Traffic Overview | Wix.com' and the URL 'manage.wix.com/dashboard/e53545a0-4f7d-4f89-92b7-6ebf2f6764cc/analytics/overviews/traffic?referralInfo=sidebar'. Below the header is a navigation bar with links for 'Explore', 'Help', 'Hire a Professional', 'Upgrade', and a search bar. A sidebar on the left lists various analytics categories like 'Traffic Overview', 'Real-time', 'Sales Overview', etc., with 'Traffic Overview' currently selected. The main content area displays two circular dashboards: one for 'Unique visitors' (69, 97% returning) and one for 'Site sessions' (109, 94% mobile). Below these are sections for 'Sessions by country' (a world map showing session distribution) and 'Traffic insights' (a list of countries with session counts: South Africa 22, Kenya 21, Ethiopia 15, Belgium 14, Serbia 12, Ghana 8, Kazakhstan 4). The bottom of the screen shows the Windows taskbar with icons for Start, Search, Task View, File Explorer, Google Chrome, and Powerpoint, along with system status icons for battery, signal, and volume.

Discussion

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