

Học xác suất thống kê qua phần mềm R

Tan-Duc Nguyen

2022-05-13

Contents

Chapter 1

Stat learning



tanduc307@gmail.com

Chapter 2

How to use tools

2.1 Tạo thư mục làm việc

Khi sử dụng Rstudio thì nên tạo thư mục project để toàn bộ files sẽ nằm trong đó.

Cách thực hiện xem ở đây https://alexdl106.github.io/intro2R/howto.html#rstudio_proj-vid

Lợi thế là trong folder project này, ta có thể tạo các folder con như images, data, tables, ... để thuận tiện lưu riêng từng loại dữ liệu. [hiện tại làm sao để cấu hình cho hình ảnh nó tự động lưu theo từng folder thì chưa biết cách, thấy có package 'here' mà không biết cách dùng).

Lưu ý là khi dùng bookdown để render books thì tên của heading là ghi tiếng Anh cho lành, vì nếu ghi tiếng Việt thì nó render bị lỗi unicode.

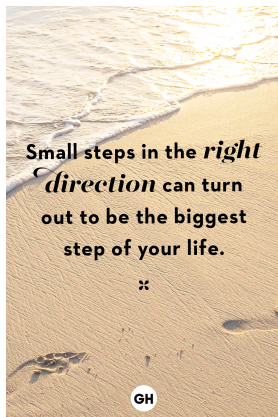
RStudio Projects from intro 2R on Vimeo.

2.2 Chèn hình ảnh

Có các cách chèn hình khi sử dụng Rstudio như sau:

- Chuyển qua chế độ Visual rồi insert hình ảnh. Sau đó hình ảnh sẽ được lưu trong folder `__book/images` khi render bằng knitr.
- Nếu gõ theo kiểu source code thì dùng đoạn mã sau ```` {r echo=TRUE, paged.print=FALSE, out.width="30%"} knitr::include_graphics("images/cover.jpg") ````

```
knitr::include_graphics("images/cover.jpg")
```



- Nếu copy và paste từ trên Internet thì dùng add-on `imageclipr` <<https://github.com/Toniiiio/imageclipr>>. Hình ảnh sẽ lưu mặc định trong working directory.
- Nếu copy và paste hình ảnh theo kiểu thủ công thì hình ảnh mặc định sẽ lưu ở `C:/Users/tandu/AppData/Local/RStudio/tmp/paste-B717C571.png` như vậy thì khi xuất bản online sẽ không thấy. Do đó phải đưa hình ảnh vào trong thư mục project (theo kiểu thủ công) hoặc dùng add-on `imageclipr` theo kiểu trực tiếp trong Rstudio (nhưng ở dạng Rmarkdown, còn Rmarkdown visual thì bị lỗi).

2.3 Chèn video

Sử dụng link này <https://video-to-markdown.marcomontalbano.com/> (theo kiểu click vào hình rồi direct qua source để play video)

Chèn từ Youtube, sử dụng code embed rồi paste vào.

```
<iframe width="560" height="315" src="https://www.youtube.com/embed/RLxV3T2b524"
title="YouTube video player" frameborder="0" allow="accelerometer;
autoplay; clipboard-write; encrypted-media; gyroscope; picture-in-picture"
allowfullscreen></iframe>
```

Tuy nhiên video sẽ không tự động resize theo khung cửa sổ

Nên dùng code này để video tự động resize. Chiều rộng và cao 100% có nghĩa là theo khung cửa sổ chứ không phải theo kích thước gốc của video.

```
<div style="position:relative;padding-bottom:56.25%;"> <iframe
style="width:100%;height:100%;position:absolute;left:0px;top:0px;"
```



```
frameborder="0" width="100%" height="100%" allowfullscreen
src="https://www.youtube.com/embed/RLxV3T2b524"> </iframe> </div>
```

Chèn từ Github, lưu ý vị trí đường dẫn của file (kể cả trong sub folder phải chính xác) và repo phải tạo Gitpage. Video sẽ play khi mở bằng trình duyệt.

```
<div style="position:relative;padding-bottom:56.25%;"> <iframe
style="width:100%;height:100%;position:absolute;left:0px;top:0px;"
frameborder="0" width="100%" height="100%" allowfullscreen
src="https://tanduc307.github.io/xstk/images/thanks.mp4"> </iframe>
</div>
```

Có thể dùng package `vembedr` <https://ijlyttle.github.io/vembedr/> để insert video cho nhanh.

2.4 Văn phạm tiếng Anh

Cách sử dụng dấu ba chấm ellipses. <https://t.me/c/1605387342/140>

Cách chèn âm thanh vào theo cách upload lên github rồi play từ URL. Chưa tìm ra cách play từ source trên hard disk.

Sử dụng package `embedr` <https://github.com/mccarthy-m-g/embedr>

```
library(embedr)
embed_audio(src = "https://tanduc307.github.io/light/An%20ellipsis%20(plural_%20ellipses)%20is%20"
```

2.5 Trích dẫn

Xem hướng dẫn ở đây là làm được https://inbo.github.io/tutorials/tutorials/r_citations_markdown/

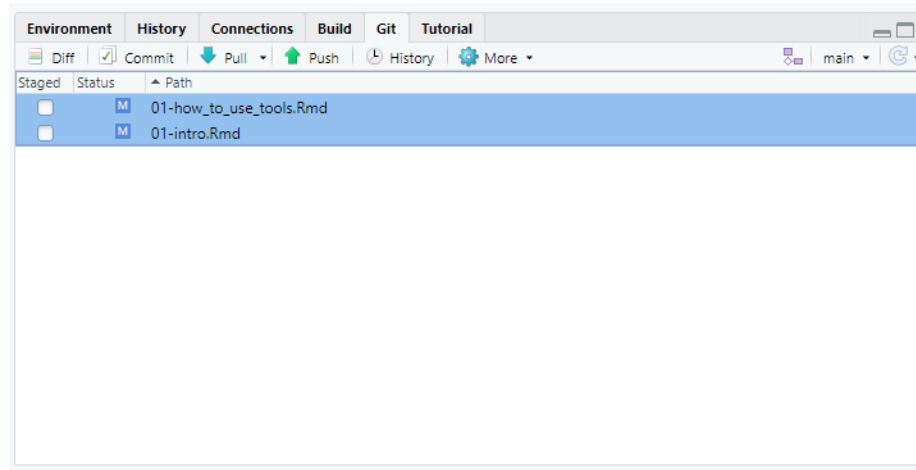
2.6 Cú pháp Markdown

Xem ở đây <https://www.markdownguide.org/basic-syntax/>

2.7 Setup Github cho RStudio

Mục đích là để thuận tiện lưu trữ toàn bộ dữ liệu của dự án R mà khỏi cần copy, save as thủ công. https://intro2r.com/github_r.html

Khi có file nào đó thay đổi thì ở Tab Git này sẽ update ngay. Ta chỉ cần chọn **Ctrl+A** rồi **Space** để select all sau đó **commit** rồi mới **push** lên server Git để lưu trữ.



2.8 Tham khảo

<https://rstats.wtf/>

Chapter 3

ANOVA RCBD

Dataset (?)

gs4_deauth() giúp cho không cần xác thực API từ googlesheet.

File raw

```
library(googlesheets4)
gs4_deauth()
data_rcbd <- read_sheet('1dFmK0hpYABrPR_e5MF27W2LSbaAGdt5dFVV3zc1H47I')
```

```
## v Reading from "raw_data".
```

```
## v Range 'Sheet1'.
```

```
print(data_rcbd, n = Inf)
```

```
## # A tibble: 48 x 3
##   block treatment yield
##   <chr>   <chr>     <dbl>
## 1 block1 TAL102     9.66
## 2 block1 TAL379     9.36
## 3 block1 TAL206     8.41
## 4 block1 TAL435     8.61
## 5 block1 TAL411     9.2
## 6 block1 ALLEN527   8.11
## 7 block1 TAL211     8.83
## 8 block1 TAL487     6.27
## 9 block1 CB1795     6.79
## 10 block1 TAL650    6.95
```

```
## 11 block1 TAL649      6.55
## 12 block1 TAL860      6
## 13 block1 TAL183      6.11
## 14 block1 TAL378      5.39
## 15 block1 CONTROL1    1.53
## 16 block1 CONTROL2    6.41
## 17 block2 TAL102     10.6
## 18 block2 TAL379      9
## 19 block2 TAL206      9.44
## 20 block2 TAL435      9.23
## 21 block2 TAL411      8.19
## 22 block2 ALLEN527    8.82
## 23 block2 TAL211      6.32
## 24 block2 TAL487      8.67
## 25 block2 CB1795      8.17
## 26 block2 TAL650      5.83
## 27 block2 TAL649      4.82
## 28 block2 TAL860      4.83
## 29 block2 TAL183      3.46
## 30 block2 TAL378      4.46
## 31 block2 CONTROL1    1.3
## 32 block2 CONTROL2    7.83
## 33 block3 TAL102     10.8
## 34 block3 TAL379     10.5
## 35 block3 TAL206     10.2
## 36 block3 TAL435      8.22
## 37 block3 TAL411      8.46
## 38 block3 ALLEN527    8.62
## 39 block3 TAL211      9.14
## 40 block3 TAL487      8.35
## 41 block3 CB1795      5.7
## 42 block3 TAL650      6.83
## 43 block3 TAL649      8.1
## 44 block3 TAL860      6.54
## 45 block3 TAL183      5.51
## 46 block3 TAL378      5.07
## 47 block3 CONTROL1    1.8
## 48 block3 CONTROL2    5.83
```

ANOVA

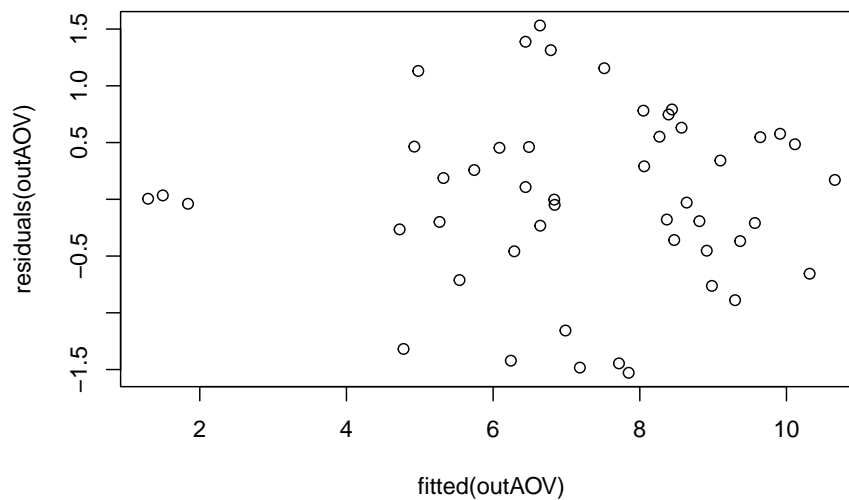
```
library(agricolae)
outAOV <- aov(yield ~ block + treatment, data = data_rcbd)
outAOV
```

```
## Call:
```

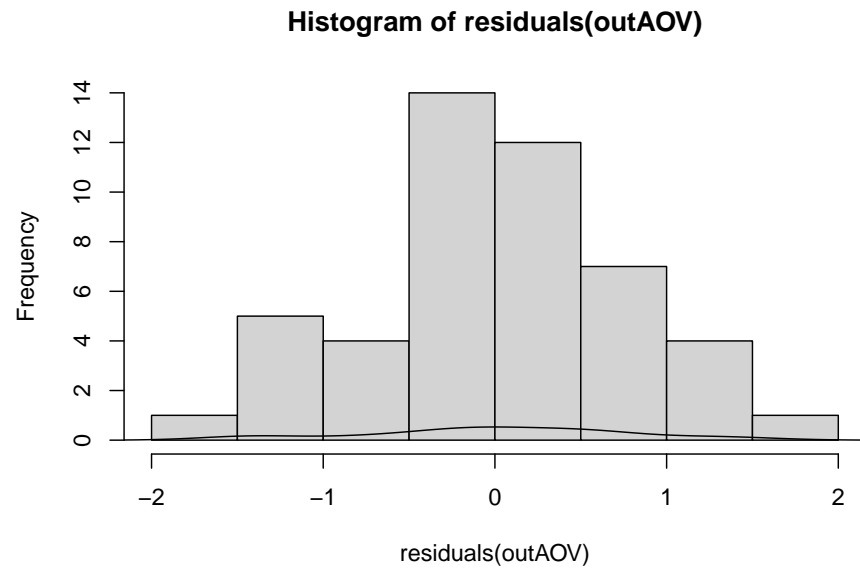
```
## aov(formula = yield ~ block + treatment, data = data_rcbd)
##
## Terms:
##           block treatment Residuals
## Sum of Squares  2.42538 217.47868 27.94669
## Deg. of Freedom    2      15      30
##
## Residual standard error: 0.9651716
## Estimated effects may be unbalanced
```

Check assumptions

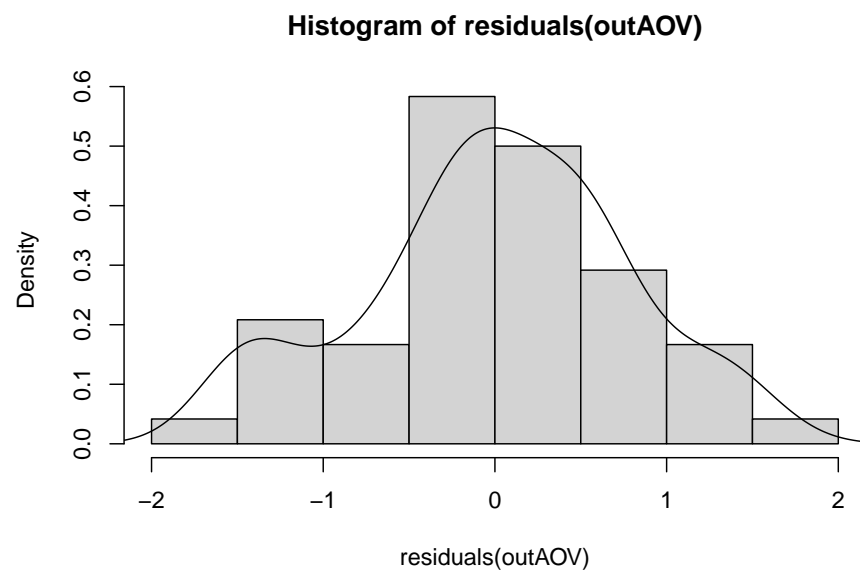
```
plot(fitted(outAOV), residuals(outAOV))
```



```
hist(residuals(outAOV))
lines(density(residuals(outAOV)))
```



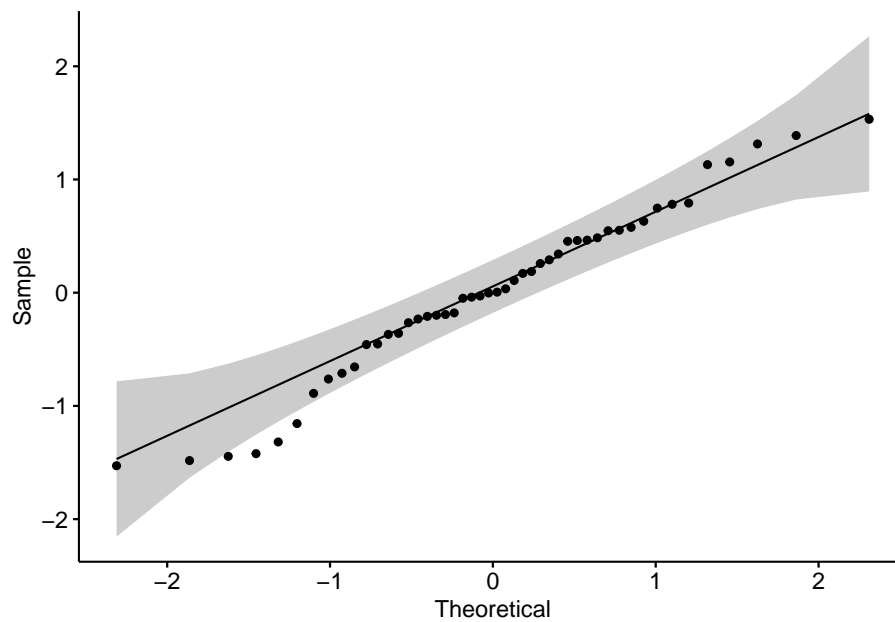
```
hist(residuals(outAOV), prob = TRUE)  
lines(density(residuals(outAOV)))
```



```
library(ggpubr)
```

```
## Loading required package: ggplot2
```

```
ggqqplot(residuals(outAOV))
```



ANOVA table

```
anova(outAOV)
```

```
## Analysis of Variance Table
##
## Response: yield
##          Df Sum Sq Mean Sq F value    Pr(>F)
## block      2   2.425   1.2127   1.3018    0.287
## treatment 15 217.479  14.4986  15.5638 3.284e-10 ***
## Residuals 30   27.947   0.9316
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

t-test LSD

```

outFactorial <-LSD.test (outAOV, c("treatment"), main = "yield ~ block + treatment",com
##
## Study: yield ~ block + treatment
##
## LSD t Test for yield
##
## Mean Square Error: 0.9315563
##
## treatment, means and individual ( 95 %) CI
##
##          yield          std r          LCL          UCL Min    Max
## ALLEN527 8.516667 0.3661056 3 7.3786265 9.654707 8.11 8.82
## CB1795 6.886667 1.2378341 3 5.7486265 8.024707 5.70 8.17
## CONTROL1 1.543333 0.2502665 3 0.4052932 2.681374 1.30 1.80
## CONTROL2 6.690000 1.0289801 3 5.5519598 7.828040 5.83 7.83
## TAL102 10.363333 0.6198656 3 9.2252932 11.501374 9.66 10.83
## TAL183 5.026667 1.3895443 3 3.8886265 6.164707 3.46 6.11
## TAL206 9.346667 0.8936629 3 8.2086265 10.484707 8.41 10.19
## TAL211 8.096667 1.5464260 3 6.9586265 9.234707 6.32 9.14
## TAL378 4.973333 0.4724757 3 3.8352932 6.111374 4.46 5.39
## TAL379 9.616667 0.7774531 3 8.4786265 10.754707 9.00 10.49
## TAL411 8.616667 0.5229085 3 7.4786265 9.754707 8.19 9.20
## TAL435 8.686667 0.5093460 3 7.5486265 9.824707 8.22 9.23
## TAL487 7.763333 1.3031245 3 6.6252932 8.901374 6.27 8.67
## TAL649 6.490000 1.6408230 3 5.3519598 7.628040 4.82 8.10
## TAL650 6.536667 0.6149255 3 5.3986265 7.674707 5.83 6.95
## TAL860 5.790000 0.8741281 3 4.6519598 6.928040 4.83 6.54
##
## Alpha: 0.05 ; DF Error: 30
## Critical Value of t: 2.042272
##
## least Significant Difference: 1.609432
##
## Treatments with the same letter are not significantly different.
##
##          yield groups
## TAL102 10.363333 a
## TAL379 9.616667 ab
## TAL206 9.346667 abc
## TAL435 8.686667 bc
## TAL411 8.616667 bc
## ALLEN527 8.516667 bc
## TAL211 8.096667 bcd
## TAL487 7.763333 cd

```


##	CB1795	6.886667	de
##	CONTROL2	6.690000	de
##	TAL650	6.536667	def
##	TAL649	6.490000	def
##	TAL860	5.790000	ef
##	TAL183	5.026667	f
##	TAL378	4.973333	f
##	CONTROL1	1.543333	g

Chapter 4

Guide

4.1 Inline formatting

work or *work*

yes or **yes**

H₂O

Ca²⁺

`code`

SMALL CAPS

RStudio

Footnote¹

If you do not want a certain heading to be numbered, you can add {-} after the heading, e.g. **# Preface {-}**

Blockquotes are written after >, e.g.,

“Love the life you live,
Live the life you love.” — Bob Marley

Code blocks

¹footnote

4.2 Listings

Unordered list items start with `*`, `-`, or `+`, and you can nest one list within another list by indenting the sub-list by four spaces, e.g.,

- one item
- one item
- one item
 - one item
 - one item


Ordered list items start with numbers (the rule for nested lists is the same as above), e.g.,

1. the first item
2. the second item
3. the third item

4.3 Media

Mình cưới thôi em

4.4 Images

- Chèn theo syntax trong markdown 
- Chèn từ link sử dụng thẻ HTML ``

Xem thêm <https://www.markdownguide.org/hacks/#image-size>

4.5 Math expressions

Ma trận dãy số

$$\begin{matrix} x_{11} & x_{12} & x_{13} \\ x_{21} & x_{22} & x_{23} \end{matrix}$$