**What the book assumes:**

Nội dung cuốn sách này không định dạy người đọc các kiến thức về lập trình hay sử dụng công cụ phân tích. Cuốn sách yêu cầu người đọc cần có các kiến thức cơ bản về sử dụng công cụ R & Rstudio. Trong đa số các trường hợp, người sử đọc chỉ cần sử dụng các đoạn mã R cơ bản là đủ mà không cần đòi hỏi các kỹ năng lập trình.

Nếu người đọc chưa có các kiến thức về R có thể đọc các tài liệu về R cơ bản như ….. Bạn cũng có thể tham khảo các tài liệu khác trên trang chính thức của R tại:

<https://www.r-project.org/doc/bib/R-books.html>

Code R sử dụng trong tài liệu này sẽ được đóng trong khung ví dụ:

|  |
| --- |
| print("hello world!")  [1] "hello world!" |

**Software download and installation:**

*R & Rstudio:*

*RStan & Bayesvl packages:*

* Cài đặt devtools

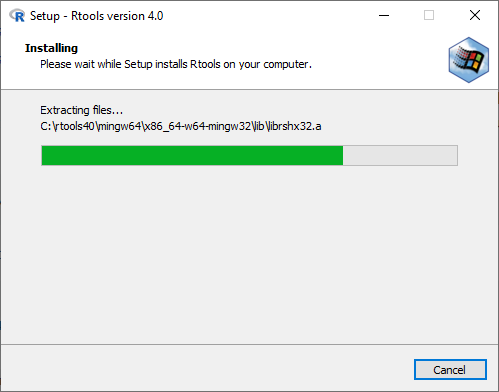
|  |
| --- |
| install.packages("devtools") |

* Cài đặt rstan

Trên windows, download bộ cài RTools tại

<https://cran.r-project.org/bin/windows/Rtools/>

Chạy bộ cài



|  |
| --- |
| install.packages("rstan", repos = "https://cloud.r-project.org/", dependencies = TRUE) |

Kiểm tra rstan đã chạy tốt

|  |
| --- |
|  |

* Cài đặt package bayesvl từ github

|  |
| --- |
| devtools::install\_github("sshpa/bayesvl") |

Kiểm tra bayesvl chạy tốt

|  |
| --- |
| library(bayesvl)  dag <- bayesvl()  dag <- bvl\_addNode(dag, "y", "bern","beta(1, 1)")  data <- data.frame(y= c(1,0,1,1,0,1,0,0,0,0,1,0,0,1,1,0,1,1,1,1))  fit <- bvl\_modelFit(dag, data, chains = 2, cores = 4)  summary(fit) |

|  |
| --- |
| Model Info:  nodes: 1  arcs: 0  scores: NA  formula: y ~ bernoulli(theta\_y)  Estimates:  Inference for Stan model: 906a162c0c1845dd25fe9de50735ce0a.  2 chains, each with iter=5000; warmup=1000; thin=1;  post-warmup draws per chain=4000, total post-warmup draws=8000.  mean se\_mean sd 2.5% 25% 50% 75% 97.5% n\_eff Rhat  theta\_y 0.55 0 0.11 0.34 0.47 0.55 0.62 0.75 2743 1  Samples were drawn using NUTS(diag\_e) at Mon May 4 01:34:14 2020.  For each parameter, n\_eff is a crude measure of effective sample size,  and Rhat is the potential scale reduction factor on split chains (at  convergence, Rhat=1).  elapsed time: 108.462465047836 secs |

**Một số ví dụ đơn giản trên R:**

Lệnh trên R được nhập trực tiếp trên cửa sổ console của R hoặc Rstudio tại vị trí dấu nhắc “>”

|  |
| --- |
| > 1 + 3 + 5  [1] 9 |

Trong ví dụ trên lệnh yêu cầu thực hiện 1 + 3 + 5, ở dòng tiếp theo kết quả trả về “[1] 9” cho ta thấy component đầu tiên “[1]” của kết quả có giá trị là 9

Tạo 1 object trên R:

|  |
| --- |
| x <- c(3, 5, 7, 8) |

Lệnh này sẽ tạo 1 object kiểu vector

Một chuỗi các lệnh liên tiếp có cấu trúc sẽ cho ta 1 program hay còn gọi là 1 script (đoạn mã)

|  |
| --- |
| a <- c(3, 5, 7, 8)  b <- c(1, 2, 9, 6)  if (sum(a) > sum(b)) {  print("sum of vector a is greater than sum of vector b.")  } else {  print("sum of vector a is equal or less than sum of vector b.")  } |

Vẽ đồ thị trên R

|  |
| --- |
| x=seq(from=-2, to=2, by=0.1) # Specify vector of x values.  y = sin(x) # Calculate corresponding y values by sine function.  plot( x , y , col="skyblue" , type="l" ) # Plot the x, y points as a blue line. |



Fig 1.1

Làm việc với dữ liệu:

Đọc dữ liệu từ tệp .csv

|  |
| --- |
| df <- read.csv("STEM5000.csv", header = T)  head(df) |

Summary descriptive statistics toàn bộ các cột của dataset

|  |
| --- |
| summary(df) |