20
Đôi lời cuối sách với bạn đọc (và tài liệu tham khảo)

Qua 15 chương sách và 3 phụ lục bạn đọc đã đi cùng tác giả một hành trình khá dài trong phân tích thống kê và biểu đồ. Trước khi "chia tay" bạn đọc, tác giả cũng muốn có đôi lời tạm biệt.

Qua kinh nghiệm giảng dạy và nghiên cứu cá nhân cho thấy phần lớn sinh viên khi tiếp cận với khoa học thống kê lần đầu chẳng mấy gì hào hứng, nếu không nói là khó khăn, vì sách giáo khoa soạn cho môn học này rất xa rời thực tế, với những ví dụ không có trong đời thường. Những khái niệm trừu tượng, những công thức rắc rối, những phép tính phức tạp và rườm rà làm cho người học cảm thấy khó khăn và từ đó cảm thấy thiếu hứng thú theo đuổi môn học. Thật vậy, có khi đọc sách giáo khoa, các bài báo nghiên cứu khoa học, chúng ta bắt gặp những phương pháp hay và những mô hình thích hợp cho nghiên cứu của chính mình, nhưng không biết làm sao tính toán các mô hình đó. Trong cuốn sách này, tác giả muốn cung cấp cho bạn đọc một phương tiện phân tích thực tế để lấp đi cái khoảng trống phương pháp và kiến thức mà có lẽ bạn đọc còn thiếu.

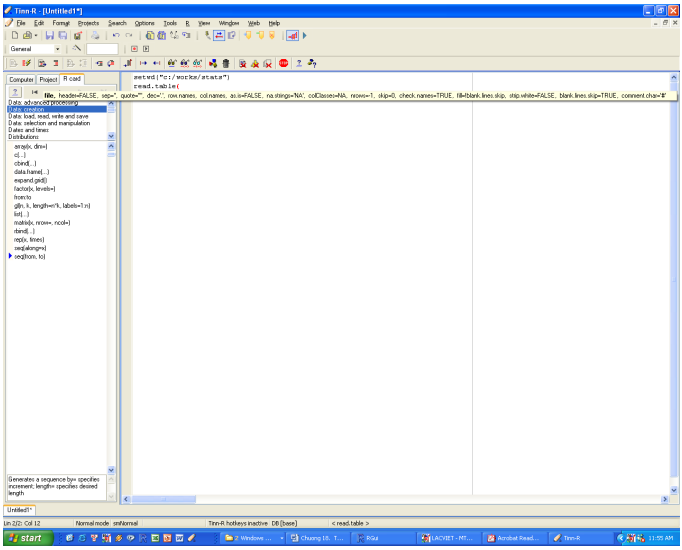
Học phải đi đôi với hành. Cách học về phương pháp hay nhất, theo tôi, là … bắt chước. R cung cấp cho bạn đọc cách học mô phỏng đó rất là tiện lợi. Trong khi đọc những chương sách này cùng với những ví dụ, bạn đọc có thể gõ những lệnh vào máy tính và xem kết quả có nhất quán với những gì mình đọc hay không. Sau khi đã biết được cách sử dụng một hàm hay một lệnh nào đó, bạn đọc có thể thêm vào (hay bớt ra) những thông số của hàm để xem kết quả. Chỉ có học như thế thì bạn đọc mới nắm vững được các khái niệm và cách sử dụng R.

Chúng ta học từ sai sót. Qua cuốn sách này, tác giả muốn bạn đọc đi một quãng đường khá … gập ghềnh, tức là bạn đọc phải tương tác với máy tính bằng những lệnh của R. Trong quá trình tương tác đó, có thể một số lệnh sẽ không chạy, vì gõ sai tên biến số hay sai chính tả, vì không để ý đến kí tự viết hoa và viết thường, vì số liệu không đầy đủ hay sai sót, v.v… Tất cả những lần sai sót đó sẽ giúp cho bạn đọc rút ra kinh nghiệm và trở nên thành thạo hơn. Đó là cách học mà người Anh hay gọi là "trial and error", học từ sai lầm và thử nghiệm.

Một công trình phân tích số liệu cần nhiều lệnh và hàm R. Tuy nhiên, vì tính tương tác mà bạn đọc theo dõi, các lệnh này sẽ biến mất khi ngưng R. Vấn

đề đặt ra là có cách nào lưu trữ các lệnh này trong một hồ sơ để sau này sử dụng lại. Phần mềm cực kì có ích cho mục đích này là `Tinn-R` (cũng có thể tải xuống và cài đặt vào máy hoàn toàn miễn phí). Website để tải `Tinn-R` và tài liệu sử dụng là: http://www.sciviews.org/Tinn-R.

`Tinn-R` thực chất là một editor cho R (và nhiều phần mềm khác). `Tinn-R` cho phép chúng ta lưu trữ tất cả các lệnh cho một công trình phân tích trong một hồ sơ. Với `Tinn-R`, chúng ta có sẵn một chỉ dẫn trực tuyến về cách sử dụng các lệnh hay hàm trong R. Trong khi lệnh gõ sai "văn phạm" R, `Tinn-R` sẽ báo ngay và đề nghị cách sửa! Giao diện `Tinn-R` có thể giống như sau:

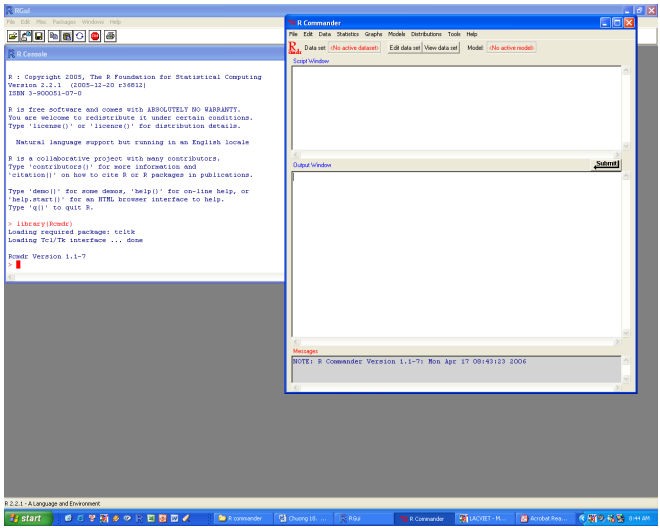


Chẳng hạn như trong giao diện trên, khi chúng ta gõ `read.table(` thì một chỉ dẫn ngay phía dưới hiện ra, với tất cả thông số của hàm `read.table`. Với `Tinn-R` chúng ta ít khi phạm phải những sai sót nhỏ trong khi chạy R. Sau khi đã xong một số lệnh, chúng ta có thể dùng chuột để tô đậm (highlight) những lệnh cần chạy và gửi sang R. Chú ý chúng ta không cần phải rời Tinn-R trong khi R chạy.

Đến đây, có lẽ bạn đọc sẽ hỏi: có cách nào sử dụng R dễ dàng hơn mà không cần phải gõ các lệnh? Câu trả lời là … có. Tại sao tôi không giới thiệu trước, ngay từ chương đầu? Tại vì tôi muốn bạn đọc đi con đường khó trước khi đi con đường dễ, nên đến bây giờ mới nói đến một phần mềm phụ khác có khả năng giúp cho bạn đọc sử dụng R một cách nhanh chóng hơn, dễ dàng hơn, và tiện lợi hơn bằng chuột thay vì bằng bàn phím.

Phần mềm để "tự động hóa" R có tên là `Rcmdr` (viết tắt từ `R commander`). Trong thực tế, `Rcmdr` là một package, mà bạn đọc có thể tải từ website chính thức của R (http://cran.au.r-project.org/src/contrib/Descriptions/Rcmdr.html) hay website của tác giả của `Rcmdr` sau đây: http://socserv.socsci.mcmaster.ca/jfox/Misc/Rcmdr. Chú ý, khi Rcmdr vận hành tốt khi có những package sau đây trong máy: `relimp`, `multcomp`, `lmtest`, `effects`, `car`, và `abind`. Nếu chưa có những package này, bạn đọc nên tải chúng về máy. Tài liệu chỉ dẫn `Rmdr` cũng có thể tải từ website http://cran.R-project.org/doc/packages/Rcmdr.pdf.

Khi đã tải `Rcmdr` xuống và cài đặt vào máy tính, bạn đọc chỉ đơn giản lệnh: `library(Rcmdr)`, và một giao diện như sau sẽ xuất hiện. Với phần "menu" (như `File`, `Edit`, `Data`, `Statistics`, `Graphs`, `Models`, `Distribution`, `Tool`, `Help`) bạn đọc có thể tự mình khám phá cách vận hành của `Rcmdr` bằng chuột.



Về nội dung lần in thứ nhất này, tác giả không có ý định bàn về những mô hình phân tích đa biến (multivariate analysis model) như phân tích yếu tố (factor analysis), phân tích tập hợp (cluster analysis), phân tích tương quan đa biến (correspondence analysis), phân tích phương sai đa biến (multivariate analysis of variance), v.v… vì đây là những phương pháp tương đối cao cấp, đòi hỏi người sử dụng phải thông thạo không những về lí thuyết thống kê, mà còn phải hiểu rất rõ những phương pháp phân tích căn bản như trình bày trong sách này. Tuy nhiên, bạn đọc có nhu cầu cho các phương pháp phân tích này cũng có

thể tìm hiểu trong trang web của R để biết thêm các package chuyên dụng cho phân tích đa biến.

Tài liệu tham khảo

Hiện nay, thư viện sách về R còn tương đối khiêm tốn so với thư viện cho các phần mềm thương mại như SAS và SPSS. Tuy nhiên, trong thời đại tiến bộ phi thường về thông tin internet và toàn cầu hóa như hiện nay, sách in và sách xuất bản trên website không còn là những khác nhau bao xa. Phần lớn chỉ dẫn về cách sử dụng R có thể tìm thấy rải rác đây đó trên các website từ các trường đại học và website cá nhân trên khắp thế giới. Phần này chỉ liệt kê một số sách mà bạn đọc nếu cần tham khảo thêm có thể tìm đọc. Trong quá trình viết cuốn sách mà bạn đọc đang cầm trên tay, tác giả cũng tham khảo một số sách và trang web sẽ liệt kê sau đây với vài lời nhận xét cá nhân.

Tài liệu tham khảo chính về R là bài báo của hai người sáng tạo ra R: Ihaka R, Gentleman R. R: A language for data analysis and graphics. *Journal of Computational and Graphical Statistics* 1996; 5:299-314.

18.1 Sách tham khảo về R

- "**Data Analysis and Graphics Using R – An Example Approach**" (Nhà xuất bản Cambridge University Press, 2003) của John Maindonald nay đã xuất in lại lần thứ 2 với thêm một tác giả mới John Braun. Đây là cuốn sách rất có ích cho những ai muốn tìm hiểu và học về R. Năm chương đầu của sách viết cho bạn đọc chưa từng biết về R, còn các chương sau thì viết cho các bạn đọc đã biết cách sử dụng R thành thạo.

- "**Introductory Statistics With R**" (Nhà xuất bản Springer, 2004) của Peter Dalgaard là một cuốn sách loại căn bản cho R nhắm vào bạn đọc chưa biết gì về R. Sách tương đối ngắn (chỉ khoảng 200 trang) nhưng giá khá đắt!

- **"Linear Models with R"** (Nhà xuất bản Chapman & Hall/CRC, 2004) của Julian Faraway. Sách hiện có thể tải từ internet xuống miễn phí tại website sau đây: http://www.stat.lsa.umich.edu/~faraway/book/pra.pdf hay http://cran.r-project.org/doc/contrib/Faraway-PRA.pdf. Tài liệu dài 213 trang.

- **"R Graphics (Computer Science and Data Analysis)"** (Nhà xuất bản Chapman & Hall/CRC, 2005) của Paul Murrell. Đây là cuốn sách chuyên về phân tích biểu đồ bằng R. Sách có rất nhiều mã để bạn đọc có thể tự thiết kế các biểu đồ phức tạp và … màu mè.

- "**Modern Applied Statistics with S-Plus**" (Nhà xuất bản Springer, 4th Edition, 2003) của W. N. Venables và B. D. Ripley được viết cho ngôn ngữ S-Plus nhưng tất cả các lệnh và mã trong sách này đều có thể áp dụng cho R mà không cần thay đổi. (S-Plus là tiền thân của R, nhưng S-Plus là một phần mềm thương mại, còn R thì hoàn toàn miễn phí!) Đây là cuốn sách có thể nói là cuốn sách tham khảo cho tất cả ai muốn phát triển thêm về R. Hai tác giả cũng là những chuyên gia có thẩm quyền về ngôn ngữ R. Sách dành cho bạn đọc với trình độ cao về máy tính và thống kê học.

18.2 Các website quan trọng hay có ích về R

- Rất nhiều tài liệu tham khảo có thể tải từ website chính thức của R sau đây: http://cran.R-project.org/other-docs.html

  Trong đó có một số tài liệu quan trọng như "**An Introduction to R**" của W. N. Venables và B. D. Ripley.
  Địa chỉ internet: http://cran.r-project.org/doc/manuals/R-intro.pdf.

- Vài tài liệu hướng dẫn cách sử dụng R có thể tải (miễn phí) và tham khảo như sau:

  "**R for Beginners**" (57 trang) của Emmanuel Paradis. Tài liệu được soạn cho bạn đọc mới làm quen với R.
  Địa chỉ internet: http://cran.r-project.org/doc/contrib/Paradis-rdebuts_en.pdf.

  "**Using R for Data Analysis and Graphics**: Introduction, Code and Commentary" (35 trang) của John Maindonald là một tóm lược các lệnh và hàm căn bản của R cho phân tích số liệu và biểu đồ. Chủ đề của tài liệu này rất gần với cuốn sách mà bạn đang đọc.
  Địa chỉ internet: http://cran.r-project.org/doc/contrib/usingR.pdf

  "**Statistical Analysis with R – a quick start"** (46 trang) của Oleg Nenadic và Walter Zucchini. Tài liệu hướng dẫn cách ứng dụng R cho phân tích thống kê và biểu đồ.
  Địa chỉ internet: http://www.statoek.wiso.uni-goettingen.de/mitarbeiter/ogi/pub/r_workshop.pdf

  "**A Brief Guide to R for Beginners in Econometrics**" (31 trang) của M. Arai. Tài liệu chủ yếu soạn cho giới phân tích thống kê kinh tế.
  Địa chỉ internet: http://people.su.se/~ma/R_intro

"**Notes on the use of R for psychology experiments and questionnaires"** (39 trang) của Jonathan Baron và Yuelin Li. Web. Tài liệu được soạn cho giới nghiên cứu tâm lí học và xã hội học. Có ví dụ về log-linear model và một số mô hình phân tích phương sai trong tâm lí học.
Địa chỉ internet: http://www.psych.upenn.edu/~baron/rpsych/rpsych.html

- StatsRus gồm một sưu tập về các mẹo để sử dụng R hữu hiệu hơn (dài khoảng 80 trang). Địa chỉ internet:
http://lark.cc.ukans.edu/pauljohn/R/statsRus.html

- Và sau cùng là một tải liệu "**Hướng dẫn sử dụng R cho phân tích số liệu và biểu đồ**" (khoảng 110 trang – thường xuyên cập nhật hóa) do chính tác giả viết bằng tiếng Việt. Website: www.R.ykhoanet.com thực chất là tóm lược một số chương chính của cuốn sách này. Trang web này còn có tất cả các dữ liệu (datasets) và các mã sử dụng trong sách để bạn đọc có thể tải xuống máy tính cá nhân để sử dụng.