

نرم افزار منبع باز R، ابزاری برای تحلیل داده ها و برنامه نویسی

هادی پورباقر^۱

چکیده

در مقاله حاضر به معرفی نرم افزار منبع باز R پرداخته می شود. نرم افزار R در دانشگاه اوکلند، کشور نیوزیلند، تولید و به صورت رایگان روی وب قرار گرفت. استقبال زیاد از این نرم افزار آن را به یکی از پیشرو ترین نرم افزارهای حوزه آمار تبدیل کرده است. رشد روزافزون این نرم افزار مرهون یک عزم جهانی از داوطلبانی است که حاضر به اشتراک گذاشتن یافته های خود شده اند. نظر به اینکه نرم افزار R خود یک زبان برنامه نویسی نیز می باشد، محققان قادرند بر حسب نیاز خاص خود در آن برنامه های مورد نیاز را نوشته یا برنامه های موجود در آن را تغییر دهند. علاوه بر قابلیت های محاسباتی، این نرم افزار دارای قابلیت های گرافیکی نیز می باشد که در طیف وسیعی از علوم می تواند مورد استفاده قرار گیرد. مقاله حاضر به صورت اجمالی به این قابلیت ها اشاره خواهد داشت.

واژگان کلیدی: نرم افزار R، آمار، گرافیک، برنامه نویسی، منبع باز.

۱. دانشیار گروه شیلات، دانشکده منابع طبیعی، دانشگاه تهران، کرج، ایران، تلفن: ۰۴۴۳۲۲۲۳ (+۹۸۲۶) دورنگار: ۳۲۲۲۷۷۶۵ (+۹۸۲۶)، نشانی الکترونیکی: poorbagher@ut.ac.ir

2. Open source

مقدمه

آمار آشنا هستیم. از جمله این نرم افزارها می توان به SAS، SPSS و minitab و STATA اشاره کرد. R یک ابزار آماری بسیار قدرتمند است اما کار با آن نظیر نرم افزارهای تجاری مذکور نیست که با چند کلیک و حرکت دادن موسواره^۶ کار را به اتمام رساند. در R فرامین به صورت کلماتی هستند که با نوشته شدن در خط فرمان اجرا می شوند. نظیر این حالت در سیستم عامل DOS وجود دارد که Windows جایگزین آن شد. بنابراین یادگیری R در آغاز با دشواری بیشتری نسبت به آنچه در نرم افزارهای تجاری با آن روبرو شده اید همراه است [۵].

نوشتن فرمان ها می تواند تا حدی خسته کننده باشد. اما همینکه در آن مهارت پیدا ایجاد شود، سهولت کار با این نرم افزار آشکار گردیده و کارایی بالای R برای تحلیل داده ها و ساخت برنامه مورد توجه فرد استفاده کننده قرار خواهد گرفت.

واسط گرافیکی کاربر

واسط گرافیکی کاربر که به اختصار GUI^۸ نامیده می شود بسیار ساده است. این محیط دارای چند منوی ساده است و در مقایسه با بسیاری از نرم افزارهای تجاری بسیار خلوت به نظر می آید. این خود می تواند تا حدی مانع علاقمندی کاربران جدید شود. اجرای فرامین در همین محیط ساده انجام می شود. این محیط خود یک ماشین حساب نیز هست. بدون اینکه اطلاعاتی درباره رمز نویسی داشته باشیم می توان چهار عمل اصلی را در این محیط انجام داد. جواب آنالیزها نیز در همین محیط نمایش داده میشود. چند نمونه از رمزهای پرکاربرد در ادامه نمایش داده شده است. باید توجه داشت که R به حروف کوچک و بزرگ حساس است.

برای انجام آنالیز واریانس باید رمز زیر را ماشین نویسی نمود.

$fit = aov(y \sim x); summary(fit)$

برای آنالیز واریانس دوطرفه از این رمز استفاده می شود.

$fit = aov(y \sim x1 * x2); summary(fit)$

برای رسم یک نمودار ساده رمز زیر کافی است.

$plot(x, y)$

برای رگرسیون خطی ساده رمز زیر را بکار می برند.

$fit = lm(y \sim x); summary(fit)$

در R می توان برنامه هایی را در پرونده جداگانه نوشت و ذخیره نمود.

نرم افزار R در طی سال های ۱۹۹۰ توسط دو پژوهشگر آمار در دانشگاه اوکلند^۱ (نیوزیلند) یعنی راس ایهاکا^۲ و رابرت جنتلمن^۳ طراحی گردید [۱].

نرم افزار R نام خود را تا حدی از حرف اول اسم این دو نفر گرفته است و تا حدی نیز نام آن اشاره به زبانی در کامپیوتر دارد که شبیه به R بوده، قبل از آن بسیار مورد استفاده قرار گرفته است، یعنی زبان S-plus. بسیاری از رمز هایی که در زبان S نوشته شده اند را می توان تقریباً بدون تغییر در R اجرا نمود. با گسترش اینترنت این زبان مورد توجه بسیاری از دانشمندان قرار گرفت، از اینرو جنتلمن و ایهاکا تصمیم گرفتند تا رمز های منبع^۴ خود را در اختیار دیگران قرار دهند [۲].

دانشمندان دیگر نیز سریعاً دست به کار شده بسته های نرم افزاری را ایجاد نمودند که خود حاوی روال^۵ و دستورهایی نرم افزاری بود که هر یک برای موضوعی خاص در پژوهش کاربرد داشتند. استفاده از R رشد افزایشی را در بسیاری از رشته های علم نظیر کشاورزی، علوم زیستی، ژنتیک، بیوشیمی، علوم زمین، زیست محیطی، ایمنی شناسی و میکروبی شناسی و ریاضیات نشان می دهد [۲]. در حال حاضر R توسط گروه مرکزی از برنامه نویسان هدایت میشود. نرم افزار R را می توان از پایگاه پروژه R [۳] دانلود و روی سیستم عامل ویندوز، مک و لینوکس نصب نمود.

R نرم افزاری برای تحلیل داده ها است که همزمان دارای قابلیت های گرافیکی برای ترسیم نمودارها بوده و مهمتر اینکه خود یک زبان برنامه نویسی نیز می باشد [۴].

در R نه تنها از هزاران برنامه و نرم افزاری که محققان از سراسر دنیا نگاشته و آن را در اختیار سایرین قرار داده اند می توان استفاده کرد بلکه امکان برنامه نویسی برای تحقیقی که دارای مشخصه های بسیار خاص مورد نظر ماست نیز فراهم شده است. نرم افزار تدوین شده توسط ما هم می تواند در اختیار سایر محققان قرار گیرد. R رایگان است و به سرعت در حال به روز شدن می باشد. اگر R را از پایگاه آن دانلود نموده باشید نباید تعجب کنید اگر پس از اندک زمانی به سراغ آن رفته و با نسخه بالاتر آن روبرو شوید. این سرعت بالای توسعه و به روز شدن آن مدیون یک عزم جهانی از جانب محققان استفاده کننده از آن است که زحمات خود را نیز به رایگان در اختیار سایر محققان قرار می دهند.

با بسیاری از نرم افزارهای تجاری^۶ تحلیل داده ها به ویژه در حوزه

1. Auckland University
2. Ross Ihaka
3. Robert Gentleman
4. Source codes

5. Routines
6. Commercial
7. Mouse
8. Graphical user interface (GUI)

نیز به R پیوند نمود [۷]. کاربران پیشرفته می توانند از ++C، C، جاوا^۳ یا پایتون^۴ برای تغییر اشیای R استفاده کنند.

قابلیت های گرافیکی

یکی از نقطه قوت های R قابلیت رسم نمودارها می باشد [۸]. توابعی برای رسم نمودارها در این نرم افزار تولید شده که هر تابع دارای پارامترهای متعددی است که امکان تغییر گراف به نحوی که مد نظر است را اجازه دهد. علاوه بر نمودارهای دو بعدی میتوان نمودارهای پیچیده و سه بعدی را نیز ترسیم نمود، نظیر خطوط کنتور، منحنی های تراکم. نمونه ای از این نمودارها در زیر ارائه شده اند (شکل ۱).

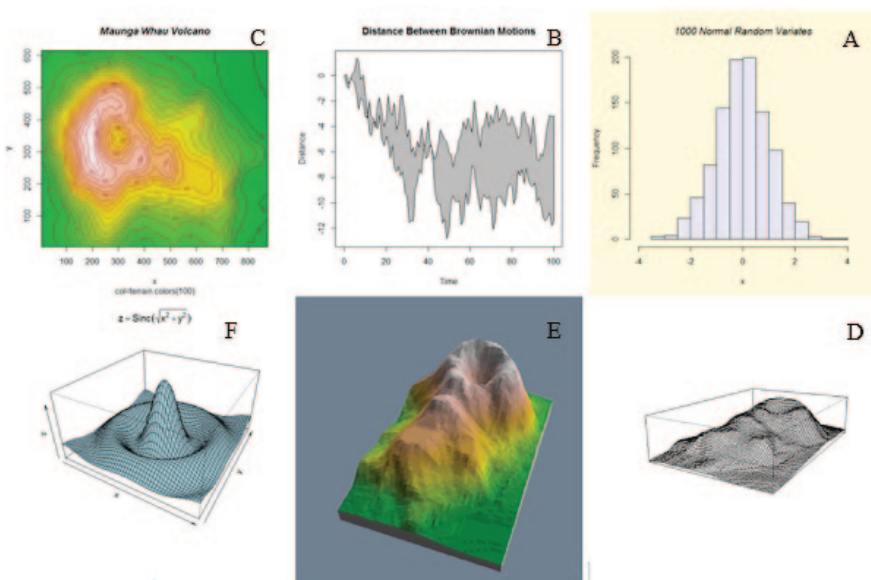
تلاش برای ایجاد واسط گرافیکی کاربر پسند^۵

نوشته فرمانها در R می تواند خسته کننده باشد، مخصوصا اینکه بسیار پیش می آید با جا انداختن یک حرف یا رعایت نکردن کوچک و بزرگ بودن حرفی از یک کلمه آن برنامه اجرا نشود. از اینرو تلاشهایی به عمل آمده است تا محیط گرافیکی کاربر پسند برای R

سپس یکباره تمامی برنامه های موجود در پرونده های جدا از هم یا قطعاتی دلخواه از آنها را اجرا نمود. نظیر بسیاری از زبان های برنامه نویسی برای اشکال زدایی^۱ برنامه ها نیز تسهیلاتی فراهم شده است. می توان در هر مرحله از اجرای تابع مقادیر محاسبه شده را که برای بخش بعدی تابع مورد استفاده قرار خواهند گرفت، دریافت کرد. توابعی نیز وجود دارند که اجرای مرحله به مرحله برنامه را نظارت می نمایند. همچنین برای استثنائاتی که در اجرای برنامه پیش می آید، پیام های خطا نمایش داد می شود.

قابلیت های آماری و ریاضی R

بسیاری از روش های آماری مورد نیاز برای تحلیل های آماری در R قرار دارد [۶]. از جمله آزمون های فرضی آمار توصیفی، آنالیز واریانس، مدل های خطی و غیر خطی. توابعی که برای این منظور بوجود آمده اند دائما در حال گسترش بوده و در قالب بسته های جدید یا بروز رسانی شده توسط داوطلبان ارائه میشوند و می توان آنها را از وب گاه^۲ CRAN (<http://cran.r-project.org>) دانلود نمود. بسیاری از توابع به زبان R نوشته شده اند اما می توان زبانهای ++C، C و فترن را



شکل ۱. برخی از انواع نمودارهایی که توسط R ترسیم شده اند. A. هیستوگرام حاصل از ۱۰۰۰ داده تصادفی؛ B. فاصله بین دو حرکت براونی؛ C. نقشه توپوگرافی؛ D. پرسپکتیو دهانه آتشفشان؛ E. پرسپکتیو دهانه آتشفشان؛ F. تغییرات یک تابع ریاضی در سه راستای x، y و z.

- | | |
|------------------------------------|------------------|
| 1. Debugging | 4. Python |
| 2. Comprehensive R archive network | 5. User-friendly |
| 3. Java | |

طور رایگان در اختیار جامعه علمی قرار گرفته است. لازمه بهره مندی از آن در درجه اول معرفی به جامعه دانشگاهی، به ویژه دانشجویان می باشد. ایجاد انجمنی علمی در درون کشور و برگزاری کارگاههای آموزشی در دانشگاه ها می تواند از گامهای موثر در این راه باشند.

منابع و مآخذ

- [1]. Ihaka, R. and R. Gentleman. (1996). "R: a Language for Data Analysis and Graphics", Journal of Computational and Graphical Statistics, 5: 299-314.
- [2]. Tippmann, S. (2015). "Programming Tools: Adventures with R", Nature, 517: 109-110.
- [3]. Foundation, T. R. (2015). "The R Project for Statistical Computing", Available from: <http://www.r-project.org>.
- [4]. Crawley, M. J. (2013). "The R Book", John Wiley & Sons. 1051 p.
- [5]. De Micheaux, P. L., R. Drouilhet and B. Liqueur (2013). "The R Software: Fundamentals of Programming and Statistical Analysis", Springer Science & Business. 628 p.
- [6]. Cohen, Y. and J. Y. Cohen (2008). "Statistics and Data with R: An Applied Approach through Examples", John Wiley & Sons. 599 p.
- [7]. Chambers, J. (2008). "Software for Data Analysis: Programming with R", Springer Science & Business Media. 498 p.
- [8]. Maindonald, J. and J. Braun (2006). "Data Analysis and Graphics Using R: An Example-based Approach", Cambridge University Press. 525 p.
- [9]. Fox, J. (2005). "Getting Started with the R Commander: a Basic-Statistics Graphical User Interface to R", Journal of Statistical Software, 14: 1-42.
- [10]. Gandrud, C. (2013). "Reproducible Research with R and R Studio", CRC Press. 294 p.

تهیه شود. باید توجه داشت هیچیک از محیط های کاربر که تاکنون برای R تهیه شده است کارایی چشمگیری ندارند. علت این موضوع شاید تعداد بسیار زیاد بسته های آماری در R و همچنین ازدیاد و به روز رسانی سریع این بسته های نرم افزاری باشد. از جمله این محیط ها می توان به Rstudio و Rcommander اشاره کرد که خود یک بسته در R می باشد [9,10]. در بسته هم میتوان فرمان ها را نوشت اما منوهایی وجود دارند که تا حدی امکان انتخاب یک آنالیز در آن فراهم می اشد.

حمایت های برخط^۱ برای R

شاید یکی از مزایای R وجود تعداد بسیار زیاد انجمن ها^۲ و وبلاگ های اینترنتی است که در آن افراد به مطرح کردن مشکلات خود پرداخته و افراد دیگر به آن جواب داده یا به بحث پیرامون آن پرداخته اند. از این گذشته خود R نیز با نوشتن علامت سوال در برابر یک (مثلا یک تابع) کلمه به جستجوی اطلاعات پیرامون آن می پردازد. نشریه علمی نیز با نام The R Journal وجود دارد که آن هم به صورت منبع باز مقالات علمی درباره نوآوری ها و تحقیقات صورت گرفته در مورد R را به چاپ می رساند.

کتاب های آموزشی

تا کنون ده ها جلد کتاب پیرامون R نگاشته شده است. بیشتر کتاب ها در زمینه آمار می باشند اما در زمینه های دیگر نظیر رسم نمودارها، مرفومتربیک و زمینه های بسیار دیگری نیز کتب مختلفی تحریر شده است. امروزه با توجه به استقبال گسترده از R بسیاری از ناشران معطوف به چاپ کتاب در این زمینه گشته اند.

نتیجه گیری

هر چند بسیاری از محققان ما به نرم افزارهای تجاری و منبع بسته عادت نموده اند، لازم است تا کاربردهای نرم افزار R به عنوان ابزاری قدرتمند برای تحلیل داده ها مورد توجه جامعه علمی ما قرار گیرد. ساختار این نرم افزار به گونه ای است که کاربر می تواند تغییر لازم متناسب با نیازهای خود را در نرم افزار اعمال نماید. بسیاری از کاربران ممکن است با دیدن رابط گرافیکی آن مایوس شده و عدم آموزش صحیح اولیه و دریافت پیام های خطای مکرر از نرم افزار آنها را از علاقمند شدن به آن باز دارد. R یک محصول مفید می باشد که به

1.On-line
2.Forum