نرم افزار

نرم‌افزار R و کاربرد آن در علم آمار
(بخش بردار و ماترس قسمت اول)

علی‌اکبر حسن‌خانی[1*]
[1]دانشجوی کارشناسی ارشد ژنتیک و اصلاح نژاد دام گروه علوم دامی، پردیس کشاورزی و منابع طبیعی دانشگاه تهران
***نویسنده مسئول:** A.hasankhani74@ut.ac.ir

چکیده
R یک زبان برنامه‌نویسی ریاضی شی‌گرا می‌باشد که بسیار شبیه S-plus (بسته‌ی نرم‌افزاری مشهور آماری) بوده و برای محاسبات آماری طراحی شده است. زبان آماری S توسط چارز و همکارانش در سال ۱۹۶۰ در لابراتور Bell به منظور برنامه نویسی آماری برای تحلیل داده‌ها و مدل‌بندی پیشرفته ایجاد شد. امروزه با گسترش روز افزون علم آمار و کاربردهای آن در سایر علوم، لزوم آشنایی با نرم افزارهای آماری که برای تجزیه و تحلیل داده‌ها و همچنین بسط و توسعه‌ی روش‌های نوین آماری به کار می‌رود، بیش از پیش قابل درک است. در این میان و در سال‌های اخیر نرم‌افزار R از پیشرفت و محبوبیت قابل ملاحظه‌ای در بین پژوهشگران و محافل علمی دنیا برخوردار بوده است. کتاب‌ها و مقالات متعدد، سمینارهای آموزشی در دانشگاه‌های معتبر جهان و سایت‌های اینترنتی گوناگون تنها برای گسترش و آموزش این نرم‌افزار به وجود آمده است. در این مطالعه به معرفی انواع ساختار داده‌ها در نرم‌افزار R، بردارها، ماتریس‌ها و کلیاتی در خصوص عملیات‌های ماتریسی و چگونگی محاسبات ریاضی بین آن‌ها با استفاده از نرم‌افزار برنامه‌نویسی R، پرداخته شده است.

کلمات کلیدی: R، برنامه‌نویسی، ماتریس

مقدمه
پروژه R از سال ۱۹۹۵ در گروه آمار دانشگاه Auckland توسط آقایان Robert Gentleman و Ross Ihaka (به همین علت نام R برای این پروژه نام گذاری شد شروع شد و در مدت زمانی بسیار کوتاه مخاطبان زیادی در سراسر دنیا یافت. در حال حاضر این زبان را یک تیم بین‌المللی نگهداری می‌کنند و داوطلبانه توسعه می‌یابد. نشانی صفحه وب پروژه R به صورت زیر است:

http://www.r-project.org

چرا زبان برنامه‌نویسی R؟
قبل از اینکه به جزئیات بیشتری در این باره پرداخته شود، این سوال پیش می‌آید که چرا نیاز به زبان برنامه‌نویسی در محیط آماری وجود دارد، ضمن این که نرم‌افزارهای مختلف آماری (spss, sas, Minitab Statistica) نیز در دسترس است. برای پاسخ به این سوال می‌توان موارد زیر را عنوان نمود:

- برای اینکه اساس یک روش، مدل و الگوریتم آماری را به درستی درک کنید و فقط با یک جعبه سیاه کار نکنید.
- بر روی خروجی نرم‌افزار R می‌توانید کنترل بیشتری داشته باشید.
- بخواهید برای روش‌های جدید آماری و یا الگوریتم خودتان برنامه و یا نرم‌افزاری (package,extension) را بسط دهید و فقط به روش‌های کلاسیک آماری بسنده نکنید.

مزایای این زبان به شرح زیر است:

- زبان R ، رایگان است و تمامی کدهای آن باز می‌باشد. همچنین برروی سیستم عامل‌های ویندوز، یونیکس، لینوکس و مکینتاش اجرا می‌شود.
- این زبان دارای راهنمای داخلی خوبی است.
- زبان R دارای قابلیت‌های گرافیکی قابل ملاحظه‌ای است.
- آشنایی با این زبان به منزله آشنایی با زبان تجاری آماری S-plus است.
- این زبان بسیار قوی و یادگیری آن آسان است و دارای توابع

پیش ساخته آماری و package های فراوانی است.

• برجسته‌ترین ویژگی این زبان برنامه‌نویسی، این است که می‌توان با استفاده از آن، package های مختلفی را ساخت و کدهای جدیدی را وارد عرصه علم آمار کرد.

ساختار داده‌ها در R

در زبان R می‌توان داده‌ها را به صورت‌های زیر نگهداری نمود:

• بردار (vector)
• ماتریس (matrix)
• آرایه (array)
• داده‌های چارچوب‌دار (data frame)
• داده‌های سری زمانی (time series)
• فهرست (list)

بردارها

ساده‌ترین ساختار داده در زبان R، بردارها هستند. بردار موجودیتی است که شامل چند داده با نوع یکسان هستند که یا تماماً عدد و یا تماماً منطقی می‌باشند. هر بردار دارای ۳ ویژگی طول، حالت و نام است. طول بردار بیان‌کننده تعداد عناصر تشکیل دهنده آن است وحالت بردار بیان‌کننده نوع عناصر (عددی، کاراکتری، مختلط و منطقی) آن است. می‌توان با تابع ( ) c بردار را ساخت. به مثال زیر توجه نمایید.

```
>x < - c(۱۰,۵,۳,٦)
>x
[۱] ۱۰ ۵ ۳ ٦
```

و یا برداری را در بردار دیگری معرفی کنید.

```
>y < - c(x , ۰,۵۵ , x , x)
>y
[۱] ۱۰,۰۰ ۵,۰۰ ۳,۰۰ ٦,۰۰ ۰,۵۵ ۱۰,۰۰ ۵,۰۰ ۳,۰۰ ٦,۰۰
۱۰,۰۰ ۵,۰۰ ۳,۰۰ ٦,۰۰
```

با استفاده از تابع ( ) length می‌توانید طول بردار (تعداد عناصر یک بردار) را مشخص کنید.

```
>length(x)
[۱] ٤
```

برای مشخص‌کردن حالت بردار (عدد، کاراکتر مختلط و منطقی) از تابع ( )made استفاده می‌شود.

```
>made (x)
[۱] "numeric"
```

عناصر یک بردار را می‌توان به صورت زیر نام‌گذاری کرد.

```
>a < - c (x = ۱ , y = ۵,۸ , z = -۷۷ )
>a
  x    y    z
۱,۰   ۵,۸  -۷۷,۰
```

یک روش معادل نیز به صورت زیر است.

```
>a < - c(۱, ۵,۸ , -۷۷ )
>names (a) < - c ("x" , "y" , "z" )
```

نام عناصر بردار را می‌توان با تابع ( )names بازیابی کرد.

```
>names (a)
[۱] "x"  "y"  "z"
```

ماتریس‌ها

ماترس‌ها یا آرایه‌های دو طرفه از ساختارهای مهم در R هستند. در واقع ماتریس، بسط بردار است. در تحلیل داده ها، متغیرها به صورت ستون ماتریس و موقعیت‌ها یا واحدهای آزمایشی به صورت سطرهای یک ماتریس معرفی می‌شوند. در R می‌توان هر نوع داده‌ای را در قالب ماتریس ذخیره کرد. برای ساختن ماتریس کافی است برای عناصر تابع ( ) c سطر و ستون معرفی کرد. به مثال زیر توجه کنید.

```
>A < - matrix(c(۱,۲,۳,٤), nr=۲, nc=۲)
```

با استفاده از دستور بالا، ما عناصر مربوطه را در محیط نرم‌افزار در دو ستون و دو ردیف در متغیر A معرفی کردیم. هم اکنون اگر متغیر A را فراخوانی کنید، خروجی به صورت زیر مشاهده می‌شود.

```
>A
     [,۱]  [,۲]
[۱,]    ۱     ۳
[۲,]    ۲     ۴
```

اکنون اگر بخواهید یکی از عناصر ماتریس A را استخراج کنید، از دستور A[i,j] استفاده می‌شود.

```
>A[۱,۲]
[۱] ۳
```

با استفاده از دستور بالا، می‌توان عنصر موجود در سطر اول و ستون دوم ماترس A را استخراج کرد.

قابل ذکر است که عناصر ماتریس در نرم‌افزار R به صورت ستونی(پیش‌فرض) ذخیره می‌شود. به مثال زیر توجه کنید.

```
>xx < - matrix(۱:٦ , ncol=۳)
> xx
     [,۱]  [,۲]  [,۳]
```

```
>A
     [,۱]  [,۲]  [,۳]  [,٤]
[۱,]   ۱     ٤     ۷     ۱۰
[۲,]   ۲     ٥     ۸     ۱۱
[۳,]   ۳     ٦     ۹     ۱۲
```

برخی از توابع مفید در رابطه با ماتریس‌ها عبارتند از rownames( ) و colnames( ) که برای نامگذاری سطرها و ستون های ماتریس استفاده می‌شود، به مثال زیر توجه نمایید.

```
>x < - matrix (۱:۹ , nrow=۳)
```

تعداد ردیف، ۳ در نظر گرفته شود.

```
>rownames (x) < - LETTERS [۱:۳]
>x
  [,۱] [,۲] [,۳]
A   ۱    ٤    ۷
B   ۲    ٥    ۸
C   ۳    ٦    ۹
```

بردار کاراکتری LETTERS یک متغیر تعریف شده در R است که شامل حروف بزرگ A تا Z می‌باشد، بردار حروف کوچک نیز letter نام دارد.

برای ترکیب بردارها یا ماتریس‌ها به منظور ایجاد ماتریس جدید از دو تابع( )rbind و ( )cbind استفاده می‌شود. تابع ( )cbind ماتریس‌ها (یا بردارها) را به صورت ستونی و تابع ( )rbind آن‌ها را به صورت سطری ترکیب کند. به مثال زیر توجه نمایید.

عدد ۹ تا ۱۲ با سر ستون C نام گذاری شود.

```
>cbind(A=۱:٤ , B=٥:۸ , C=۹:۱۲)
     A   B   C
[۱,] ۱   ٥   ۹
[۲,] ۲   ٦   ۱۰
[۳,] ۳   ۷   ۱۱
[٤,] ٤   ۸   ۱۲
>rbind (A=۱ :٤ , B= ٥:۸ , C= ۹:۱۲)
  [,۱] [,۲] [,۳] [,٤]
A   ۱    ۲    ۳    ٤
B   ٥    ٦    ۷    ۸
C   ۹   ۱۰   ۱۱   ۱۲
```

```
[,۱]  ۱  ۳  ٥
[,۲]  ۲  ٤  ٦
```

اگر بخواهید نحوهٔ ذخیره شدن عناصر به صورت سطری باشد، از دستور زیر استفاده کنید.

```
>xx <-  matrix(۱:٦, ncol=۳, byrow=T)
>xx
     [,۱] [,۲] [,۳]
[۱,]   ۱    ۲    ۳
[۲,]   ٤    ٥    ٦
```

اکنون با استفاده از دستور B [i,j] ستون‌های مد نظرتان را از ماتریس استخراج کنید.

```
>B < - matrix(۱:٦, nr=۲)
>B
     [,۱] [,۲] [,۳]
[۱,]   ۱    ۳    ٥
[۲,]   ۲    ٤    ٦
>B[,۲:۳]
     [,۱] [,۲]
[۱,]   ۳    ٥
[۲,]   ٤    ٦
```

و یا با استفاده از دستور زیر به سطر معینی از ماتریس دسترسی پیدا کنید.

```
>B [,۲]
[۱]  ۲  ٤  ٦
```

در صورتی‌که بخواهید یک ماتریس را به صورت یک بردار تعریف کنید، از تابع as.vector استفاده کنید.

```
>xx < - matrix(۱:٦, ncol=۳)
```

تعداد ستون، ۳ در نظر گرفته شود.

```
>xx
     [,۱] [,۲] [,۳]
[۱,]   ۱    ۳    ٥
[۲,]   ۲    ٤    ٦
>x < - as.vector(xx)
>x
[۱] ۱ ۲ ۳ ٤ ٥ ٦
```

برای تبدیل یک بردار به یک ماتریس نیز از تابع dim( ) استفاده می‌شود. به این مثال توجه کنید.

```
>A < - ۱:۱۲
>dim(A) < - c(۳,٤)
```

نتیجه‌گیری

با استفاده از محیط نرم‌افزار R می‌توان بسیاری از ماترس‌ها و بردارهای حجیم که محاسبه دستی آن‌ها برای انسان غیرممکن است را در کمترین زمان ممکن و با دقت بسیار بالایی حل کرد و خروجی دقیقی را از عملیات‌های ماتریسی و یا برداری بدست آورد. لازم به ذکر است که در بخش دوم این مقاله (چاپ بعدی نشریه) به عملیات‌های ریاضی بین ماتریس‌ها و بردارها (جمع، ضرب و سایر عملیات‌ها) پرداخته خواهد شد.

منابع

Crawley, M.J., 2012. *The R book*. John Wiley & Sons.

Jones, O., Maillardet, R. and Robinson, A., 2014. *Introduction to Scientific Programming and Simulation Using R*. Chapman and Hall/CRC.

Kuhnert, P., Venables, B. and Zocchi, S.S., 2005. *An introduction to R: Software for Statistical Modelling & Computing*. USP/ESALQ/LCE.

Ng, H.T., 2006. Statistics: An Introduction Using R

Seefeld, K. and Linder, E., 2007. Statistics using R with biological examples. *Durham: University of New Hampshire*.

Verzani, J., 2018. *Using R for Introductory Statistics*. Chapman and Hall/CRC.

R Software and Its Application in Statistical Science (Vector and Matrix Part I)

Aliakbar Hasankhani[1*]

[1] M.Sc. Animal Breeding and Genetics, Department of Animal Science, College of Agriculture and Natural Resources at University of Tehran

***Corresponding Author E-mail:** a.hasankhani74@ut.ac.ir

Abstract

R is an object-oriented mathematical programming language much like S-plus (the famous statistical software package) and designed for statistical computing. The S statistical language was developed by Charles and his colleagues at Bell Laboratory in 1960 for statistical programming for data analysis and advanced modeling. Nowadays, with the increasing spread of the science of statistics and its applications in other sciences, the necessity of acquaintance with the statistical software used for data analysis as well as the development of new statistical methods is becoming more and more understood.in recent years, R software has gained considerable popularity among researchers and scholars worldwide. Numerous books and articles, training seminars at prestigious universities around the world, and various websites have been created solely to develop and train this software this study introduces the types of data structures in R, vectors, matrices, and generalizations about matrix operations and how mathematical calculations are performed between them using R programming software.

Keyword(s): R, Programming, Matrix