

## مقدمه‌ای بر نرم‌افزار R و کاربرد آن در علم آمار (بخش جبری)

### Introduction to R Software and its Application in Statistics Science

#### چکیده

R یک زبان برنامه‌نویسی و محیط نرم‌افزاری برای تجزیه و تحلیل‌های آماری و نمایش گرافیکی است. R توسط ایهاکا و رابرت (۱۹۹۵)، در دانشگاه اوکلند نیوزیلند طراحی شد و این زبان برنامه‌نویسی به نام R، بر اساس اولین حرف نام دو نویسنده‌اش (Robert Gentleman و Ross Ihaka) نام‌گذاری شده است. امروزه با گسترش روزافزون علم آمار و کاربردهای آن در سایر علوم لزوم آشنایی با نرم‌افزارهای آماری که برای تجزیه و تحلیل داده‌ها و همچنین بسط و توسعه 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 فراوانی در بین پژوهشگران و محافل علمی دنیا برخوردار بوده است. نرم‌افزار R، بر اساس زبان آماری S نوشته شده است. زبان آماری S توسط چابرز و همکارانش (۱۹۶۰)، در لابراتوار Bell به منظور برنامه‌نویسی آماری برای تحلیل داده‌ها و مدل بندی پیشرفته ایجاد شده است. بسیاری از متخصصان علوم آماری جهت معرفی روش‌های ابداعی خود برای تحلیل داده‌ها، نتایج مطالعات خود را به راحتی و بدون هیچ هزینه‌ای به صورت پکیج‌هایی در اختیار سایرین قرار می‌دهند و همین امر R را به ابزار گسترش و پیشرفت سریع‌تر علم آمار تبدیل کرده است.

**واژه‌های کلیدی:** R، برنامه‌نویسی، داده‌های آماری

۵- امکانات گرافیکی برای تجزیه و تحلیل داده‌ها و یا نمایش به صورت مستقیم در کامپیوتر و یا چاپ در مقالات را فراهم می‌سازد.  
۶- این نرم‌افزار و پکیج‌هایش به صورت متن باز در اختیار عموم قرار گرفته است.  
نرم‌افزار R از لینکی زیر به صورت رایگان قابل دریافت است.  
<https://cran.r-project.org>  
پس از نصب نرم‌افزار، آن را اجرا کرده و صفحه‌ای به صورت زیر مشاهده خواهید کرد.



نکاتی پیش از آموزش که می‌بایستی حتماً به آنان توجه شود:

- ۱- دستورات در صفحه‌ی Console بعد از عملگر اعلان یعنی ">" وارد می‌شود.
- ۲- رعایت کوچک و بزرگ بودن حروف انگلیسی در R الزامی است.
- ۳- برای پاک کردن دستورات در Console، از کلید Ctrl+L استفاده کنید.
- ۴- برای اجرا کردن (Run) کردن هر دستور از کلید Enter استفاده کنید.
- ۵- برای معرفی کردن یک متغیر یا داده می‌توان از عبارات (=) یا (>) استفاده نمود.
- ۶- برای پاک کردن حافظه‌ی R از دستور rm(list=ls()) استفاده می‌کنیم. مثلاً برای حذف متغیر x از دستور rm(x) استفاده می‌کنیم.
- ۷- برای ذخیره‌ی داده‌ها از کلیدهای ترکیبی Ctrl+S استفاده می‌کنیم و سپس پنجره‌ای باز می‌شود و در آن مسیر ذخیره‌سازی را تعیین می‌کنیم. از کاربردهای R در بخش آمار می‌توان به محاسبه‌ی پارامترهای جمعیتی (میانگین، مد، میانه، واریانس، انحراف معیار و ...)، رگرسیون‌های خطی و

#### مقدمه

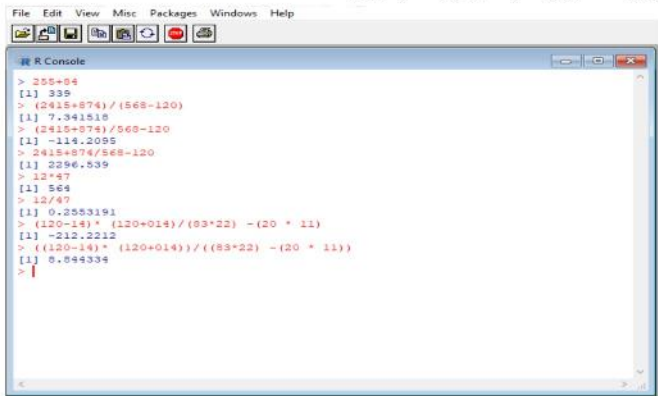
همان‌طور که قبلاً ذکر شد، R یک زبان برنامه‌نویسی و محیط نرم‌افزاری برای تجزیه و تحلیل‌های آماری و نمایش گرافیکی است. ویژگی‌های مهم R که باعث برجسته شدن آن شده عبارت‌اند از:

- ۱- یک زبان برنامه‌نویسی توسعه‌یافته، ساده و مؤثر است که شامل شرط‌ها، حلقه‌ها، توابع بازگشتی تعریف شده توسط کاربر و امکانات ورودی و خروجی است.
- ۲- دارای یک سرویس ذخیره‌سازی اطلاعات مهم است.
- ۳- مجموعه‌ای از اپراتورها را برای محاسبات آرایه‌ها، فهرست‌ها، بردارها و ماتریس‌ها به کار می‌برد.
- ۴- مجموعه‌ای بزرگ و یکپارچه از ابزارها را برای تجزیه و تحلیل داده‌ها به صورت هم‌زمان فراهم می‌کند.

دستور ۱  $> 6+9-(2*6)/(2*10)-10$  — Enter → [1] 4.4

دستور ۲  $> ((6+9)-(2*6))/((2*10)-10)$  — Enter → [1] 0.3

در تصویر زیر نمونه‌های از دستورات در محیط R انجام شده است (به اولویت‌های پرانتزهای توجه فرمایید).



### معرفی متغیرها به نرم‌افزار

برای سهولت در انجام دستورنویسی‌ها بهتر است از متغیرها استفاده کنید و داده‌ها را با استفاده از دستور ( $\rightarrow$ ) به‌عنوان یک متغیر معرفی نموده و سپس عملیات را انجام دهید. به صورت زیر:

```

> A <- (25*85*94*6358*774)/(41-24+68+74-14*852)
> A
[1] 83523505
  
```

عملیات بالا را، به‌صورت متغیر A معرفی کرده و سپس آن را فراخوانی کرده و پاسخ نهایی را مشاهده می‌کنیم. برای انجام عملیات پیچیده‌تر می‌توان متغیرهای مختلفی را معرفی و سپس آن‌ها را فراخوانی کنیم.

```

> A <- ((12658+1428)-(325*45) + (19865*31))/((8546-36958)*(3298-1243))
> A
[1] -0.01053796
> B <- ((183-12)*(86547/1546) + (12398-1654))*((9354+1368)/(164 *468))
> B
[1] 2838.189
  
```

چندگانه و لجستیک، منحنی‌های توزیع نرمال و باینومیل، آنالیز واریانس و کوواریانس، آنالیزهای سری زمانی، الگوریتم درخت تصمیم‌گیری، الگوریتم جنگل تصادفی، تجزیه و تحلیل ماندگاری، آزمون کای اسکور، وکتورها، ماتریس‌ها و ... را اشاره نمود. نکته‌ی قابل توجه این است که در این مقاله، ما به بخش جبری اشاره خواهیم کرد و در شماره‌های بعدی نشریه، دیگر بخش‌های آماری را مرحله‌به‌مرحله توضیح خواهیم داد.

### ۱- بخش جبری (عملیات ریاضی)

جدول زیر شامل عملگرهایی است که برای عملیات مختلف ریاضی در محیط R از آن‌ها استفاده می‌شود.

عملگر	شرح	عملگر	شرح	عملگر	شرح
=>	کوچک‌تر یا مساوی	>	کوچک‌تر		
<=	بزرگ‌تر یا مساوی	<	بزرگ‌تر		
Min()	می‌نیم	=	مساوی		
Max()	ماکسیمم	!=	نامساوی		
pmin()	می‌نیم موازی	Obs()	قدر مطلق		
pmax()	ماکسیمم موازی	sum	مجموع		
Round()	گرد کردن اعداد	prod	حاصل ضرب		
Sin() cos() tan()	توابع مثلثاتی	Floor()	جز صحیح یک عدد		
		+	جمع		
		-	تفریق		
		*	ضرب		
		/	تقسیم		
		** یا ^	توان		
		Sqrt()	جذر		
		% / %	خارج قسمت تقسیم		
		% %	باقیمانده تقسیم		

در بخش جبری، R همانند یک ماشین حساب عمل می‌کند. برای مثال عبارات زیر را در صفحه‌ی دستورات R یا همان (console) وارد نموده و سپس کلید Enter را زده تا دستور اجرا شود و پس از آن، پاسخ دستور وارد شده در زیر آن ظاهر گردد. مثال:

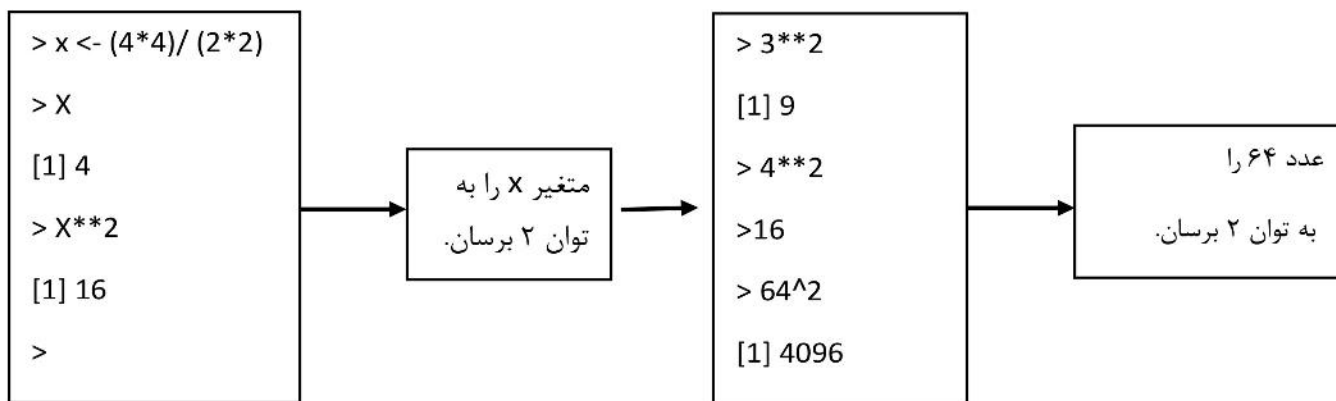
```

> 6+8
[1] 14
  
```

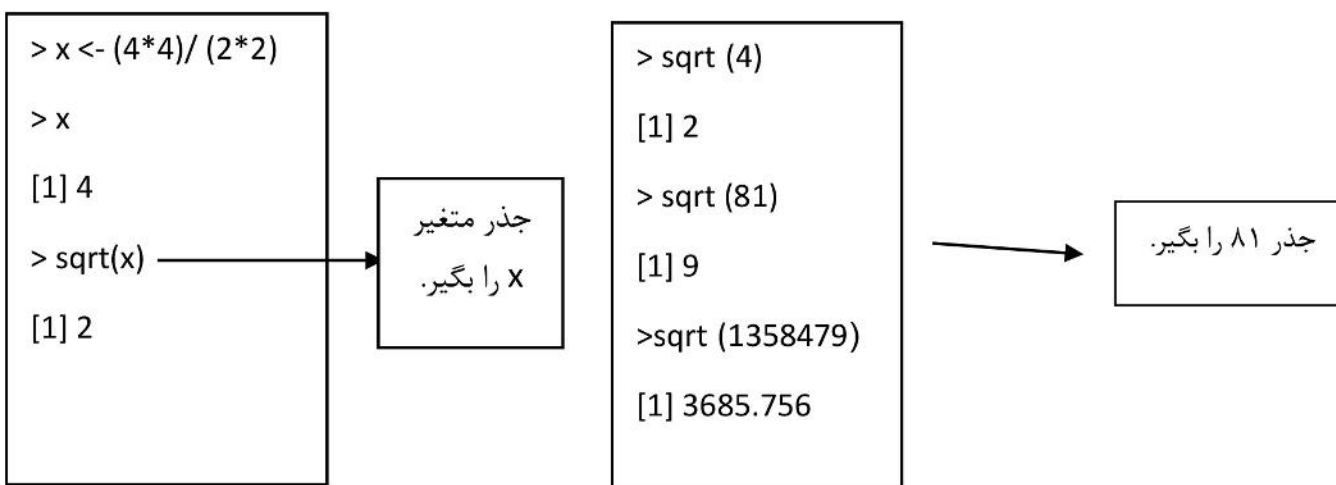
عدد ۱ نشان‌دهنده‌ی تعداد سطر

توجه داشته باشید که به‌هنگام محاسبه‌ی چند عملیات باهم، باید به‌هنگام دستورنویسی ترتیب پرانتزها رعایت شود و در غیر این صورت پاسخ نهایی اشتباه خواهد شد. نمونه‌ای از این تفاوت را در زیر خواهیم دید:

به توان رساندن اعداد با استفاده از عملگر "\*" یا "8" است.



جذر گرفتن از اعداد و متغیرها با استفاده از دستور sqrt() صورت می‌گیرد.



پس از اینکه عملیات جبری ما به صورت متغیرهای مختلفی معرفی شدند می‌توان بین متغیرهای معرفی شده نیز عملیات جبری انجام داد.

همانگونه که مشاهده کردید ابتدا عملیات پیچیده‌ی جبری را به صورت متغیرهایی معرفی کرده و پس از فراخوانی متغیرها، بین خود متغیرها رابطه‌ی جبری برقرار کرده و آن‌ها را به صورت متغیرهای دیگری معرفی و فراخوانی کرده‌ایم. حال پس انجام عملیات مختلف جبری در محیط برنامه‌ی R، چنانچه نیاز به ذخیره‌سازی اطلاعات بود می‌توان با استفاده از کلیدهای ترکیبی Ctrl+S اطلاعات موردنظر را ذخیره‌سازی کرد.

### نتیجه‌گیری

با استفاده از محیط نرم‌افزاری R می‌توان بسیاری از محاسبات پیچیده‌ی آماری را در کمترین زمان ممکن محاسبه و با دقت بالایی آن‌ها را ارزیابی کرد.

```

> A <- ((12658+1428)-(325*45)+(19865*31))/((8546-36958)*(3298-1243))
> A
[1] -0.01053796
> B <- ((183-12)*(86547/1546)+(12398-1654))*((9354+1368)/(164 *468))
> B
[1] 2838.189
> C <- -(13698*4687)
> C
[1] -64202526
> D <- A+B
> D
[1] 2838.178
> E <- C-B
> E
[1] -64205364
> F <- ((E/D)*A)
> F
[1] 238.39
> G <- -F-C+(B*A)
> G
[1] 64202258

```

عملیات جبری را به صورت یک متغیر معرفی می کنیم.

سپس متغیر را فراخوانی می کنیم.

پس از فراخوانی متغیر موردنظر، پاسخ عملیات مذکور را مشاهده می کنیم.

می توان بین متغیرها نیز رابطه‌ی جبری برقرار کرد و آن‌ها را به صورت یک متغیر دیگری معرفی و فراخوانی کرد.

language for data analysis and graphics." Journal of computational and graphical statistics, 5(3), 299-314.

5. Martin, T. (2009). "The Undergraduate Guide to R." A beginner's introduction to.

6. Paradis, E. (2005). "R for Beginners." Institut des Sciences de l'Evolution, Univerisite Montpellier II. France.

7. Verzani, J. (2014). "Using R for introductory statistics." Chapman and Hall/CRC Publication.

منبع

1. Chambers, J. (2008). "Software for data analysis: programming with R." Springer Science & Business Media Publication.

2. Crawley, M.J. (2005). "An introduction using R." Á Wiley Publication.

3. Gentleman, R. (2008). "R programming for bioinformatics." Chapman and Hall/CRC Publication.

4. Ihaka, R. and Gentleman, R. (1996). "R: a