



آشنایی با نرم افزار R  
درس آمار حیاتی و احتمالات  
مدرس بخش نرم افزار : پویا تقی پور لنگرودی

# در این ارائه خواهیم دید :

Data frame

List

NA , NaN

Working directory

مرتب کردن بردار

Factor

Matrix



Data  
frame

R

List

NA, NaN

Working  
directory

مرتب  
کردن  
بردار

Factor

Matrix

data frame ها ماتریس‌هایی هستند که می‌توان داده‌هایی از نوع و تایپ مختلف در آنها ذخیره کرد. یکی از روش‌های تعریف آن استفاده از تابع `data.frame()` است

```
# R program to create dataframe

# creating a data frame
friend.data <- data.frame(
  friend_id = c(1:5),
  friend_name = c("Sachin", "Sourav",
                  "Dravid", "Sehwag",
                  "Dhoni"),
  stringsAsFactors = FALSE
)
# print the data frame
print(friend.data)
```

برای بازیابی داده ها میتوان از index سطر و ستون و یا نام ستونها استفاده کرد. Data frame همانند ماتریس بردارهایی با اندازه برابر را ذخیره میکنند.

Data  
frame

List

R

NA , NaN

Working  
directory

مرتب  
کردن  
بردار

Factor

Matrix

یکی از مهم ترین object های موجود در R لیست است.

از لیست ها میتوان برای ذخیره سازی مجموعه ای از object ها با اندازه و تایپ مختلف استفاده کرد.

```
# R program to create a List

# The first attributes is a numeric vector
# containing the employee IDs which is created
# using the command here
empId = c(1, 2, 3, 4)

# The second attribute is the employee name
# which is created using this line of code here
# which is the character vector
empName = c("Debi", "Sandeep", "Subham", "Shiba")

# The third attribute is the number of employees
# which is a single numeric variable.
numberOfEmp = 4

# We can combine all these three different
# data types into a list
# containing the details of employees
# which can be done using a list command
empList = list(empId, empName, numberOfEmp)

print(empList)
```

Data  
frame

List

R

NA , NaN

Working  
directory

مرتب  
کردن  
بردار

Factor

Matrix

همچنین در زمان تعریف لیست برای هر بخش نامی در نظر میگیریم.  
در صورت عدم تخصیص نام هر بخش با عدد index مربوطه نمایش داده می شود.  
با توجه به اینکه اندازه بردارها متفاوت است نمیتوان آنها را بصورت data frame داشت.  
تاکنون برای استخراج داده ها و بازیابی آنها از سینگل اسکوئر براکت استفاده کردیم. اما استفاده از  
سینگل اسکوئر براکت در لیست بخش هایی از آن را نمایش خواهد داد.  
برای استخراج داده هر بخش از دابل اسکوئر براکت و یا نماد دلار استفاده می شود.

## داده گم‌شده (NA)، مقدار بی نهایت و مقادیر غیر عددی (NaN)

در این نرم افزار همانند سایر نرم افزارهای آماری، مدیریت مشاهدات گم شده میسر است.  
جهت نمایش داده گم شده در R از نماد NA استفاده می‌شود.

جهت پیدا کردن داده گم شده از تابع `is.na` استفاده می‌کنیم. خروجی این تابع از نوع logical است و  
برای مقدار گم شده برابر true خواهد بود.

جهت مشاهده index داده گم‌شده میتوان از تابع `wich(is.na())` استفاده کرد.

مقدار غیر عددی در R با نماد NaN نشان داده می‌شود.

چند مورد مختلف از جمله استفاده از توابع ریاضی با مقادیر نادرست باعث ایجاد حالت بالا خواهند شد.

جهت بررسی این مورد نیز از `is.nan` استفاده خواهد شد.

Data  
frame

List

NA,  
NaN

R

Working  
directory

مرتب  
کردن  
بردار

Factor

Matrix

Data  
frame

List

NA , NaN

Working  
directory

R

مرتب  
کردن  
بردار

Factor

Matrix

در اغلب اوقات حجم داده‌ها و دستورهای ما بسیار بیشتر بوده و به محیطی نیاز داریم که دستور و دیتای خود را در آن تعریف کنیم و سپس با اتصال به نرم افزار آنها را تحلیل کنیم. سایتهایی وجود دارند که میتوان از دیتاست‌های آنها استفاده کرد.

[www.data.gov](http://www.data.gov)

[www.Data.ny.gov](http://www.Data.ny.gov)

[www.archive.ics.uci.edu](http://www.archive.ics.uci.edu)

ساده‌ترین راه import کردن داده‌ها در R استفاده از فایل text است. برای export کردن داده‌ها در اکسل به text ستون‌های موجود در فایل داده‌ها را که به آن field می‌گویند، با فرمت‌های مختلف می‌توان از هم جدا نمود. برای مثال با فرمت text (tab delimited) داده‌ها توسط تب از یکدیگر جدا می‌شوند و یا با فرمت comma delimited توسط کاما یا همان ویرگول متمایز می‌شوند

Data  
frame

List

NA , NaN

Working  
directory

R

مرتب  
کردن  
بردار

Factor

Matrix

تابع `read.table` جهت خواندن داده ها با فرمت `text` به R است. تابع `read.table` همیشه داده ها را به صورت یک دیتافریم برمی گرداند. بدین ترتیب این تابع جهت خواندن داده ها با تایپ مختلف مناسب است. توابع `read.csv` و `read.delim` نیز مشابه همین تابع عمل میکنند.

در R می توان داده ها را مستقیماً از اکسل `import` کرد. لازم به ذکر است برای استفاده از تابع `read.xls` در سیستم عامل ویندوز، لازم است که مفسر `Perl` نصب شود.

تابع `read.xls` جهت خواندن فایل اکسل و تبدیل آن به `data frame` است.

تابع `gzfile` یک کانکشن با فایل زیپ برقرار میکند و این امکان را به تابع `read.csv` می دهد که فعالیت خود را با دسترسی به محتوای فایل زیپ انجام دهد.

`read.csv("clipboard")` چنانچه قصد داشتید دیتای کپی شده را ذخیره سازی کنید میتوانید از دستور استفاده کنید.



Data  
frame

List

NA , NaN

Working  
directory

R

مرتب  
کردن  
بردار

Factor

Matrix

8

يك روش استفاده از تابع dput است.

در اين تابع آرگومان اول نام data set و آرگومان دوم نام فايل خروجی است.

با استفاده از اين تابع می توان يك شیی را ذخیره نمود. فايل ذخیره شده را می توان با استفاده از تابع dget به R وارد و در يك متغیر ذخیره نمود

فايل حاصل از dput را میتوان توسط dget لود کرد و سپس توسط dget.data ذخیره میکنیم.

## قانون recycling، توابع پرکاربرد و مرتب کردن بردارها

زمانی که یک محاسبه بین دو بردار با اندازه متفاوت انجام گیرد بردار کوچکتر recycle می شود به این معنا که طول خود را به اندازه بردار بزرگتر افزایش می دهد. بنابراین اندازه بردار خروجی برابر بردار بزرگتر خواهد بود.

تابع min کوچکترین مقدار، تابع max بزرگترین و تابع sum جمع مقادیر موجود در یک بردار را نمایش می دهد.

تابع mean میانگین، median میانه، تابع range برداری حاوی مقادیر ماکس و مین، تابع var واریانس، تابع sd در واقع standard deviation و تابع quantile که بصورت پیش فرض چهارک ها را نمایش می دهد.

تابع cumsum بردار جمع تجمعی، تابع cumprod ضرب تجمعی، تابع cummin برداری نزولی حاوی مقادیر مینیموم تجمعی است و در نهایت cummax برداری صعودی حاوی مقادیر ماکزیموم تجمعی را برمی گرداند.

Data  
frame

List

NA , NaN

Working  
directory

مرتب  
کردن  
بردار

Factor

Matrix

Data  
frame

List

NA , NaN

Working  
directory

مرتب  
کردن  
بردار

Factor

R

Matrix

در برخی مواقع لازم است که سطوح (levels) يك فاکتور بر اساس ویژگی از متغیر مورد اندازه گیری نظیر میانگین مرتب شود. در این صورت می توان از تابع **reorder** استفاده نمود. این تابع به سه آرگومان نیاز دارد. **اولی فاکتور، دوم برداری که میخواهیم سطوح فاکتور را براساس آن مرتب کنیم و آرگومان آخر تابع مورد نظر ما خواهد بود.**

در مدل های خطی، بردار ضرایب حاوی solution برای هر سطح به صورت انحراف از سطح مرجع (reference level) است. در صورت عدم تعیین شرطی برای سطوح يك فاکتور، سطوح فاکتور بر اساس حروف الفبا مرتب و اولین سطح، سطح مرجع است. جهت تغییر سطح مرجع از تابع **relevel** استفاده می شود.

توابع `rowSums` و `colSums` به ترتیب برای محاسبه مجموع مقادیر سطر و ستون‌های ماتریس بکار می‌رود و همچنین توابع `rowMeans` و `colMeans` نیز برای محاسبه میانگین سطر و ستون ماتریس استفاده می‌شود.

توجه داشته باشید که شرط اصلی برای جمع و تفریق ماتریس‌ها آن است که از یک مرتبه باشند. جهت تعریف یک ماتریس قطری از تابع `diag()` استفاده می‌شود. برای محاسبه ترانهاده (transpose) یک ماتریس از تابع `t()` استفاده می‌کنیم. به مثال زیر توجه کنید. برای ضرب اسکالر میتوان به طریق زیر عمل کرد.

```
mat3  
2 * mat3
```

برای ضرب دو ماتریس در هم باید تعداد ستون‌های ماتریس اول برابر تعداد سطرهای ماتریس دوم باشد. برای ضرب Hadamard زمانی میتوانیم اقدام کنیم که مرتبه هر دو ماتریس بایکدیگر برابر باشد. برای ضرب Kronecker که در آن هر عنصر ماتریس اول در همه عناصر ماتریس دوم ضرب میشود میتوانیم از تابع Kronecker استفاده کنیم. برای ضرب خارجی نیز میتوان از تابع `outer` استفاده کرد.



خسته نباشید!