

Some R Problems Derived Me Nuts!

iHELP Working Group
Chel Hee Lee & Eugene Jung

April 18, 2013

R을 사용하게 된 이래로 여러가지 경험들을 바탕으로 도움이 되고자 하는 내용을 정리하였습니다. 이 문서의 작성자 역시 다른 사용자들의 똑똑한 테크닉들을 항상 배우고 있습니다.

본 페이지는 다양한 종류의 패키지들을 활용하는 방법보다는 R 기본 시스템만을 활용하여 할 수 있는 내용들만이 정리되었습니다. 이 문서는 매일 갱신이 되며, 문서의 최초 작성일은 2013년 4월 13일입니다.

만약, 이 페이지를 읽고 있는 사용자가 이 문서가 도움이 되었다고 생각된다면 문서에 더 많은 내용이 기재될 수 있도록 추가, 수정 및 제안에 대한 내용을 ihelp-urquestion@lists.r-forge.r-project.org의 주소로 이메일을 보내주시길 부탁드립니다. 메일링에 등록된 내용은 이 문서의 관리자가 문서를 갱신할 때마다 반영하도록 노력할 것입니다.

1 미분류 질문들

1. R은 처음부터 기존의 통계팩키지와는 다른 모습에 약간 두렵기까지 합니다. 기존의 분석은 일반적으로 [프로그램 실행 -> 데이터 불러오기 -> 분석(메뉴클릭:SPSS 또는 명령어입력:SAS) -> 실행]의 절차를 밟아 왔기에 모든 결과를 한 번에 보여주는 식입니다. 그러나 R은 그렇지 않아 이러한 점부터 생소하고 이상합니다. 데이터를 불러오기 하면 바로 데이터시트를 볼 수 있는 것도 아닙니다 (접수날짜: 2013-APR-17).

(답변) 사용방식의 다른 점에 대해서 아주 근본적인 다른 점을 알려줄 것. Introduction to R 문서에 써 있음 (단순히 링크시켜주는게 좋을 듯 함).

2. 프롬프트가 무엇인가요?

>

(답변)

3. 분석을 하고 나면 결과를 그래프나 그림으로 나타내게 되는데 R에서는 그림을 나타내는 창이 하나만 나타나서 동시에 두 개를 보지 못하는 경우가 허다한데, 이의 해결방법은 없나요? (접수: 2013-APR-13, 분류: 그래픽스 관련)

(답변) R에서는 그래픽 디바이스가 그래픽 생성시 마다 초기화되어 다시 보여줌으로서 그래픽 창이 하나만 계속 보여지는 것입니다. 새로운 그래프를 또다른 장치를 통해 보여주고자 한다면 `X11()`이라는 명령어를 이용하면 됩니다. 이 명령어는 유닉스환경에 설치된 R의 경우에 해당합니다. 윈도우즈의 경우에는 없습니다.

4. 초기에 가장 보는 에러는 “xxx 함수가 없습니다” 또는 “xxx 함수를 찾을 수 없습니다”입니다. (접수: 2013-APR-15, 분류: 패키지 관련)

(답변) 대부분의 경우는 사용하고자 하는 함수가 R 기본 배포판에 포함되어 있지 않은 사용자에게 의해서 제공된 특정한 패키지에서 존재하기 때문입니다. 이런 경우에는 먼저 사용하고자 하는 함수가 어떤 패키지에 존재하는지 알아야 합니다. 그리고, 해당 패키지를 설치했을 때에는 설치된 패키지를 사용할 수 있도록 로딩하는 과정을 거쳐야 합니다.

```
> library(pkg_name)
```

이와 반대로 현재 연결된 라이브러리를 떼어낼 수도 있습니다.

```
> detach(package:pkg_name)
```

5. 패키지를 설치 (분류: 사용자 환경)

(답변) 설치되는 패키지의 설치위치와 의존성에 대해서 반드시 알아야 합니다.

```
> install.packages("패키지명", dependencies=TRUE, )
```

6. 설치된 패키지의 목록을 확인하는 방법을 알고 싶습니다.

(답변)

2 데이터 관리에 관련하여

1. 아래와 같이 주어진 데이터에 변수 ID는 결측값 없이 모든 값이 완전하게 잘 들어가 있는데, Week 변수에는 각 ID의 첫번째 레코드에만 해당하는 부분에 값이 들어가 있고 나머지부분에는 NA값이 들어가 있습니다.

```
mydata <- data.frame(ID=c(rep(1,4), rep(2,4), rep(3,2)), Week=c(15, NA, NA, NA, 18, NA, NA, NA, 20,
```

```
> mydata
```

	ID	Week
1	1	15
2	1	NA
3	1	NA
4	1	NA
5	2	18
6	2	NA
7	2	NA
8	2	NA
9	3	20
10	3	NA

이와 같은 데이터를 아래와 같이 자동으로 채워주려면 어떻게 해야 할까요?

	ID	Week
1	1	15
2	1	15
3	1	15
4	1	15
5	2	18
6	2	18
7	2	18
8	2	18
9	3	20
10	3	20

이를 수행하는데에는 여러 가지 종류의 함수들이 다양한 패키지 안에 존재합니다. 그러나, 이를 수행하는 기본 알고리즘은 동일하며, R 기본시스템만으로 작성이 가능합니다. 아래의 함수를 복사하여 사용하시면 됩니다.

```
fill <- function(x, first, last){
  n <- last-first+1
  for(i in c(1:length(first))) x[first[i]:last[i]] <- rep(x[first[i]], n[i])
  return(x)
}
```

- 위에서 주어진 데이터에서 ID 변수에서 보이는 것처럼 같은 관측치가 여러번 반복 측정되어 ID가 반복적으로 입력이 되었을 때, SAS에서처럼 각 아이디별로 첫번째와 마지막 레코드를 알수 있는 .FIRST 와 .LAST 같은 기능이 R에서는 어떻게 해야 하나요?

```
mydata$first <- !duplicated(mydata$ID)
mydata$last <- !duplicated(mydata$ID, fromLast=TRUE)
```

```
> mydata
  ID Week first last
1  1   15  TRUE FALSE
2  1   NA FALSE FALSE
3  1   NA FALSE FALSE
4  1   NA FALSE  TRUE
5  2   18  TRUE FALSE
6  2   NA FALSE FALSE
7  2   NA FALSE FALSE
8  2   NA FALSE  TRUE
9  3   20  TRUE FALSE
10 3   NA FALSE  TRUE
```

- wide format 데이터를 long format 으로 바꿀 수 있나요?
- 여러개의 엑셀시트로 구성되어 있는 엑셀파일을 불러와 하나의 데이터셋으로 합치기
- 가끔 리스트형으로 받아진 데이터가 중첩된 구조를 가지고 있어서, 한 번에 이를 불러오기를 해야할 때는 어떻게 해야할지.
- do.call() 함수를 사용하는 법에 대해서..
- which.max()와 which.min()을 사용하는 방법
- list()과 data.frame()과의 관계
- apply(), lapply(), sapply(), mapply()의 사용방법
- 현재 패키지를

3 수치해석 및 시뮬레이션에 관련하여

- R에도 C와 같은 switch문이 존재하며, 그렇다면 어떻게 사용할 수 있나요?
- warning(경고)와 error(에러)를 이용하는 법
- try() 함수를 이용하여 에러를 컨트롤 해보기
- tryCatch() 함수를 이용해서 에러를 컨트롤하기

```

result <- tryCatch(
{
    수행하고자 하는 표현식
},
warning = function(w) {
    위에서 수행한 표현식이 경고를 발생시킬때 어떻게 처리하고자 하는지에 대한 표현식
},
error = function(e) {
    위에서 수행한 표현식이 에러를 발생시킬때 어떻게 처리하고자 하는지에 대한 표현식
}, finally {
    위에서 수행한 표현식에 대한 최종적 처리를 위한 표현식
}

```

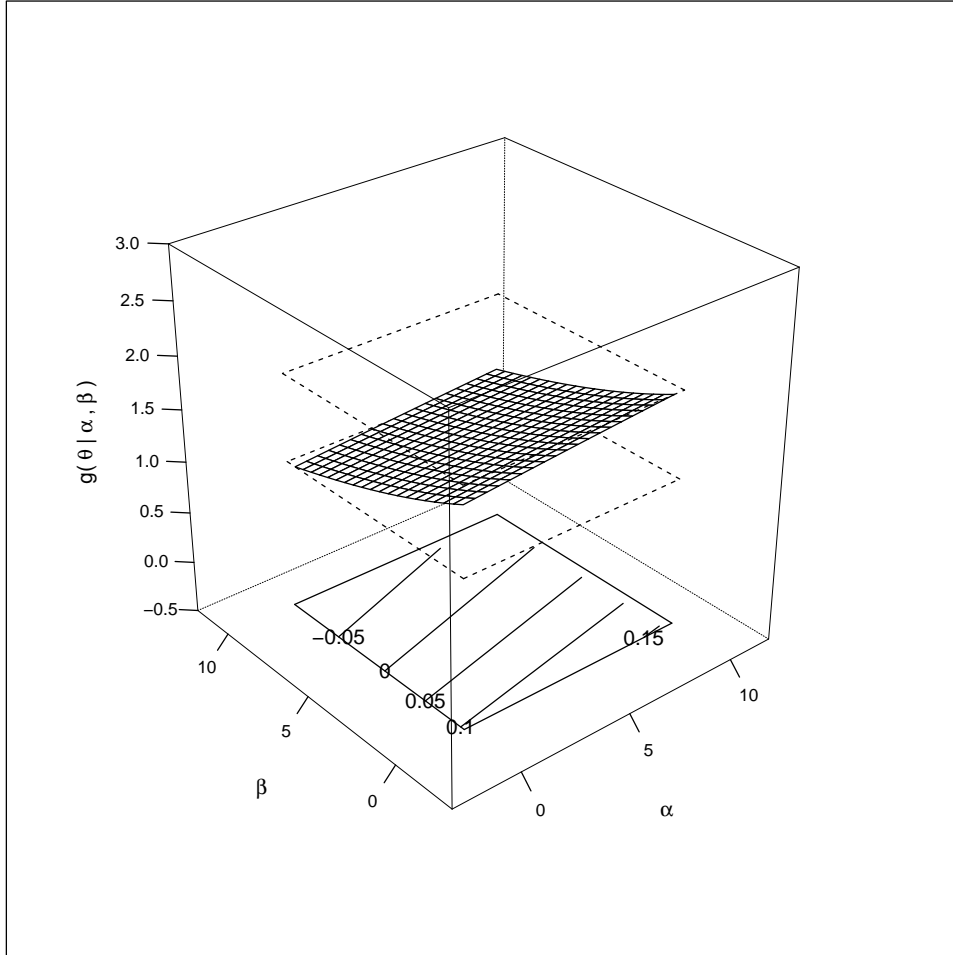
예제는 내일 시간날때 작성

5. `combn()` 함수를 이용하여 모든 조합을 찾기
6. Metropolis-Hastings 알고리즘을 구현하는 프레임 워크
7. Newton-Raphson 알고리즘을 구현하는 방법
8. Laplace Approximation 알고리즘을 구현하는 방법
9. Bootstrap 방법

4 비주얼라이제이션

1. `coordinating system`을 활용하기
2. Lattice 패키지를 이용하여 아래와 같은 그림을 생성해보기 (가장 단순한 예제임 - 팁 보다는 튜토리얼 형식으로?)

3D Plot Customization



Lattice Example 1

3. L^AT_EX의 문서에 포함될 .eps 그래픽을 R에서 뽑았을 때는 아무런 문제가 없어 보였는데, 정작 pdf로 문서를 뽑아 보니까 이 그래픽이 들어간 페이지가 90도로 돌아가 있거나 혹은 그래픽이 90도로 회전되어 있을 경우에는 아래와 같이 하면 됩니다.

```
postscript(file='filename.eps', onefile=FALSE, horizontal=FALSE)
```

이 문제에 대한 출처는 postscript 도움말입니다.

이문제를 다른 방법으로도 해결할 수 있습니다. (대충 서너개 더 있음).

5 데이터 입출력 및 파일관리 유틸리티

1. `read.table` 계열의 함수를 이용하여 데이터를 불러올 때 첫번째 인자는 파일의 위치와 파일명이 입력된 문자열이어야 합니다. 그런데, 간혹 문법에서 틀린 점도 없고, 불러오고자 하는 데이터 파일도 올바른 파일경로에 위치하고 있음에도 불구하고, 데이터를 찾을 수 없다고 하는 경우가 있습니다. 이것은 내부적으로 파일경로에 띄어쓰기, 특수문자, 혹은 특수한 인코딩 등 다양한 이유로 인하여 파일경로가 올바르게 처리되지 않았기 때문입니다. 아래와 같은 방법으로 `read.table()` 함수 사용시 `file.choose()` 함수를 함께 사용하면 이러한 문제를 해결이 가능합니다.

```
mydata <- read.table(file.choose(), header=TRUE, sep=",")
```

6 기타내용들

1. 불러오고자 하는 데이터의 인코딩이 UTF-8가 아닐때 이를 확인하고, 데이터를 올바르게 불러오기 위한 내용은 <http://lists.r-forge.r-project.org/pipermail/ihelp-urquestion/2013-April/000017.html> 를 읽어보시길 바랍니다.
2. R을 한국어가 아닌 영문으로 사용하고 싶습니다 (버전에 관계없이 일반적으로 통용되는 방법 - 윈도우즈 사용자에게 맞추어 작성됨). 이를 설정하는 방법에 대해서는 <http://lists.r-forge.r-project.org/pipermail/i> 을 읽어보시길 바랍니다.