

Ch03 데이터 입출력

1. 데이터 불러오기

1.1 키보드 입력

scan()함수, edit()함수

(1) scan()함수 이용

실습 (키보드로 숫자 입력)

scan() 함수를 실행하면 console창에 입력 데이터의 순서를 나타내는 프롬프트가 표시

입력된 데이터는 입력된 순서대로 벡터변수에 저장됨

```
num <- scan()
```

```
num
```

```
sum(num)
```

실습 (키보드로 문자 입력)

키보드로 문자 입력을 위해서 scan()함수에서 'what=character()'를 사용

```
name <- scan(what = character())
```

```
name
```

(2) edit() 함수를 이용한 입력

edit()함수는 표 형식의 데이터편집기를 제공

실습 (편집기를 이용한 data.frame 만들기)

```
df = data.frame()
```

```
df = edit(df)
```

```
df
```

* Mac에서는 미동작

1.2 로컬 파일 가져오기

함수를 이용하여 파일의 데이터를 가져오기

(1) read.table()함수 이용

테이블 형태로 작성

컬럼이 공백, 탭, 콜론(:), 세미콜론(;), 콤마(,) 등의 구분자가 포함된 자료 파일을 불러올 수 있는 함수

구분자가 공백 또는 tab이면 sep속성 생략가능

컬럼명이 있는 경우 header속성을 'header=TRUE'로 지정

형식: read.table(file = "경로명/파일명", sep="구분자", header = "T or F")

실습 (컬럼명이 없는 파일 불러오기)

```
getwd()
```

```
setwd("C:/Rwork/")
```

```
student <- read.table(file = "student.txt")
```

```
student
```

```
names(student) <- c("번호", "이름", "키", "몸무게")
```

```
student
```

컬럼명이 없는 경우 v1, v2, ... 형태로 기본 컬럼명이 지정됨

names() 함수를 사용하면 사용자가 원하는 컬럼명을 지정 가능

실습 (컬럼명이 있는 파일 불러오기)

```
student <- read.table(file = "student.txt", header=T)
```

```
student
```

컬럼명이 있는 파일을 불러오기 위해서 'header=TRUE' 또는 'header=T' 속성 이용

실습 (탐색기를 이용한 파일 선택)

```
student1 <- read.table(file.choose(), header = TRUE)
```

file.choose() 함수 이용 시 파일선택 대화상자에서 파일 선택 가능

실습 (구분자가 있는 경우)

```
student2 <- read.table(file = "student2.txt", sep = ";", header = TRUE)
```

```
student2 <- read.table(file = "student2.txt", sep = "\t", header = TRUE)
```

student2

구분자가 있는 경우 sep="구분자" 형식으로 지정

구분자가 tab인 경우 sep="\t" 형식으로 sep속성 값을 지정하여 파일을 불러온다.

실습 (결측치를 처리하여 파일 불러오기)

```
student3 <- read.table(file = "student3.txt", header = TRUE, na.strings = "-")
```

student3

특정 문자열을 NA로 처리하여 파일을 불러올 수 있다.

'-'문자를 결측치로 처리하기 위해서 na.string="-"속성을 사용하면 '-'문자를 NA로 변경하여 파일의 데이터를 불러온다.

(2) read.csv()함수 이용

CSV(Comma Separated Value)파일 형식은 콤마(,)를 기준으로 각 컬럼을 구분하여 저장한 데이터 형식

형식: read.csv(file="경로명/파일명" [, sep=","] [, header=TRUE])

실습 (CSV 파일 형식 불러오기)

```
student4 <- read.csv(file = "student4.txt", sep = ",", na.strings = "-")
```

```
student4
```

CSV는 콤마(,)가 구분자이기 때문에 sep속성은 생략 가능

1.3 인터넷에서 파일 가져오기

사이트에서 csv파일 형식의 데이터를 R 스크립트로 가져오기

<https://vincentarelbundock.github.io> 에서 제공하는 csv파일을 가져와 시각화하기까지의 과정 실
습

실습 (인터넷에서 파일을 가져와 시각화)

titanic 자료 가져오기

```
titanic <-
```

```
  read.csv("https://vincentarelbundock.github.io/Rdatasets/csv/COUNT/titanic.csv")
```

```
titanic
```

```
dim(titanic)
```

```
str(titanic)
```

실습 (자료의 차원 정보와 자료구조 보기 및 범주의 빈도수 확인)

타이타닉 자료는 1,316개의 관측치와 5개 변수의 컬럼으로 구성

```
table(titanic$age)
```

```
table(titanic$sex)
```

```
table(titanic$survived)
```

단계별 실습 (데이터프레임에서 문자형 컬럼)

R 3.x버전에서 데이터프레임의 칼럼으로 사용되는 자료형이 문자형(chr)인 경우 요인형(Factor)으로 자동 변환됨.

R 4.0버전에서는 문자형을 그대로 유지

```
head(titanic)
```

```
tail(titanic)
```

head()함수: 관측치의 앞부분 6개 보여줌

tail()함수: 관측치의 뒷부분 6개 보여줌

실습 (단계4: 교차분할표 작성)

```
tab <- table(titanic$survived, titanic$sex)
```

```
tab
```

table()함수에서 첫번째 인수는 교차분할표에서 행으로 나타나고, 두번째 인수는 열로 나타난다.

실습 (단계5: 범주의 시각화 – 막대 차트 그리기)

```
barplot(tab, col = rainbow(2), main = "성별에 따른 생존 여부")
```

barplot()함수: 막대그래프 그리기

2. 데이터 저장하기

콘솔에 출력하거나 처리가 완료된 결과물을 특정 파일에 저장하는 방법

2.1 화면(console) 출력

처리결과가 저장된 변수를 화면(console)에 출력하는 R 함수: cat()함수, print()함수

실습 (cat()함수 이용 변수 출력)

```
x <- 10
```

```
y <- 20
```

```
z <- x * y
```

```
cat("x * y의 결과는 ", z, "입니다.\n")
```

```
cat("x * y = ", z)
```

cat()함수는 출력할 문자열과 변수를 함께 결합하여 콘솔에 출력해 준다.

실습 (print()함수 이용 변수 출력)

```
print(z)
```

print()함수는 문자열을 함께 사용할 수 없다. 변수의 값 또는 수식의 결과만을 출력할 수 있다.

2.2 파일 저장

결과를 특정 파일에 저장하는 방법

(1) sink() 함수 이용

sink()함수를 실행하면 이후에 작업한 모든 내용이 지정된 파일에 저장된다.

sink()함수의 기능 종료를 위해서 인수없이 sink()함수를 한 번 더 실행

실습 (sink()함수를 이용한 파일 저장)

```
setwd("C:/Rwork/Part-I")
```

```
library(RSADBE)
```

```
data("Severity_Counts")
```

```
sink("severity.txt")
```

```
severity <- Severity_Counts
```

```
severity
```

```
sink()
```

(2) write.table()함수 이용

처리된 결과(변수)를 테이블 형식으로 파일에 저장하는 함수

'row.names' 속성: 행 번호 제거

'quote' 속성: 따옴표를 제거

실습 (write.table()함수를 이용한 파일 저장)

titanic

```
write.table(titanic, "titanic.txt", row.names = FALSE)
```

'row.names = F'는 행의 이름을 제거

실습 (Write.table()함수로 저장한 파일 불러오기)

```
titanic_df <- read.table(file = "titanic.txt", sep = "", header = T)
```

titanic_df

Write.table()함수로 저장된 데이터는 read.table()함수를 이용하여 텍스트 파일을 데이터프레임 형식으로 불러올 수 있다.

(3) write.csv()함수 이용

데이터프레임 형식의 데이터를 CSV형식으로 파일에 저장

실습 (write.csv()함수를 이용한 파일 저장)

```
setwd("C:/Rwork/")
```

```
st.df <- studentx
```

```
write.csv(st.df, "stdf.csv", row.names = F, quote = F)
```

Write.csv() 함수를 이용하여 CSV형식으로 파일에 저장

#엑셀에서 텍스트 파일 대상 csv파일 만들기 참고