R 데이터 파일

I. R 데이터 파일 만들기

II. R에서 Excel 파일 다루기

III. R 데이터 관리

Ⅰ. R 데이터 파일 만들기

1. 데이터 파일 만드는 방법

- Data Editor(데이터 편집기)를 이용하는 방법
- 프로그래밍 명령어를 이용하는 방법
  - 예를 들어 Excel 파일을 만들 수 있음

2. Data Editor를 이용하는 방법

- Source 창에서 다음과 같이 입력하면 Data Editor 창을 불러옴

```
mydata<-data.frame(이름=character(), 경제원론=numeric(), 미시경제=numeric(),거시경제=numeric())
mydata<-edit(mydata)
```

  - data.frame : Excel과 같이 숫자, 문자 등 데이터를 하나의 테이블에 담을 수 있는 R 자료구조 함수
  - edit : Data Editor를 불러오는 함수
- 데이터 편집기 창에 엑셀에서 데이터를 입력하는 방법과 유사하게 데이터를 입력

	이름	경제원론	미시경제	거시경제
1	김기훈	98	82	95
2	박수동	100	90	80
3	원선희	50	45	75
4	위계영	50	100	100
5	최동팔	80	95	95
6	최종열	90	60	60
7	최동수	30	30	30
8	김기팔	80	25	25
9	이상수	65	70	70
10	강창수	95	90	90
11				
12				

3. 명령어를 이용하는 방법

- 프로그래밍을 할 경우 Data Editor 창을 이용하는 방식은 불편하므로 명령어를 활용
- 변수명은 한글이나 영문 모두 사용이 가능한데 프로그래밍 편의상 변수명을 영어로 입력
- 예 : 이름(name), 경제원론(prin), 미시경제(micro), 거시경제(macro)를 보여주는 데이터는 프로그램 작성 시 다음과 같이 명령어를 입력

```
name<-c("김기훈","박수동","원선희","위계영","최동팔","최종열","최종수","김기팔","이상수","강창수")
prin<-c(98,100,50,50,80,90,30,80,65,95)
micro<-c(82,92,45,100,95,60,30,25,70,90)
macro<-c(95,80,75,100,95,60,30,25,70,90)
df<-data.frame(name,prin,micro,macro)
df
```

```
      name prin micro macro
1   김기훈   98    82    95
2   박수동  100    92    80
3   원선희   50    45    75
4   위계영   50   100   100
5   최동팔   80    95    95
6   최종열   90    60    60
7   최종수   30    30    30
8   김기팔   80    25    25
9   이상수   65    70    70
10  강창수   95    90    90
```

II. R에서 Excel 파일 다루기

1. R에서 Excel 파일 불러오기

- Excel 파일을 R에서 불러오는 방법은 다양한데 웹에 있는 xlsx 파일을 불러올 수 있는 방법을 설명
- 이를 위해서는 이 작업에 특화된 R package인 openxlsx 패키지를 install해야 함
- 패키지를 install하는 방법은 RStudio의 Console 창에서 install.packages("openxlsx")를 입력하고 Enter를 치면 패키지가 install이 됨
- install된 패키지는 프로그램 내에서 library로 불러 와야 사용할 수 있음
- 강기춘 홈페이지에 있는 실습용 Excel 데이터 파일을 R 프로그래밍에서 불러올 수 있음

b3-ch1-2.R

```
library(openxlsx)
sample1<-read.xlsx("http://kanggc.iptime.org/book/data/score.xlsx")
sample1
```

2. R package 설치

- R package는 R 함수, 코드 및 샘플 데이터의 모음
- 다양한 R package가 있는데 강기춘 홈페이지(http://kanggc.iptime.org)에서 본 강의에 필요한 package들을 한번에 설치할 수 있는 install-packages.R 파일을 다운받아 설치

3. R에서 Excel 파일 만들기

- R에서 작업한 데이터를 Excel 데이터 파일로 저장하면 Excel에서 그 파일을 불러올 수 있음
- 예 : R에서 학생 10명의 경제원론, 미시경제, 거시경제 성적을 입력하여 데이터를 만든 후
  합계(sum), 평균(mean), 합격/불합격(PF) 판단, 등급(grade) 부여하고, Excel 데이터 파일로 저장
- 이렇게 만들어진 Excel 데이터 파일은 Excel을 실행하여 불러와서 사용

b3-ch1-3-rev.R

```
name<-c("김기훈","박수동","원선희","위계영","최동팔","최종열","최종수","김기팔","이상수","강창수")
prin<-c(98,100,50,50,80,90,30,80,65,95)
micro<-c(82,92,45,100,95,60,30,25,70,90)
macro<-c(95,80,75,100,95,60,30,25,70,90)
df<-data.frame(name,prin,micro,macro)
df$sum<-df$prin+df$micro+df$macro
df$mean<-df$sum/3
df$PF<-ifelse(df$mean >= 60, "pass","fail")
df$grade<-ifelse(df$mean >= 90, "A", ifelse(df$mean >= 80, "B", ifelse(df$mean >= 70, "C", ifelse(df$mean >
= 60, "D", "F"))))
df
library(openxlsx)
write.xlsx(df, "C:/Users/user/Desktop/b3-ch1-3.xlsx", colNames=T, asTable=F)
```

	A	B	C	D	E	F	G	H
1	name	prin	micro	macro	sum	mean	PF	grade
2	김기훈	98	82	95	275	91.66667	pass	A
3	박수동	100	92	80	272	90.66667	pass	A
4	원선희	50	45	75	170	56.66667	fail	F
5	위계영	50	100	100	250	83.33333	pass	B
6	최동팔	80	95	95	270	90	pass	A
7	최종열	90	60	60	210	70	pass	C
8	최종수	30	30	30	90	30	fail	F
9	김기팔	80	25	25	130	43.33333	fail	F
10	이상수	65	70	70	205	68.33333	pass	D
11	강창수	95	90	90	275	91.66667	pass	A
12								

1. 변수 변환 및 변수명 변경

- 변수 변환이란 기존의 변수를 이용하여 새로운 변수를 만드는 것임
  - 예 : gdp 및 consumption에 자연로그를 취하여 lgdp 및 lconsumption 변수를 생성

    lgdp<-log(gdp)

    lconsumption<-log(consumption).
- 변수명 변경은 하나의 변수명을 변경하거나 전체 변수명을 변경
  - 예를 들어 sample1 데이터에 year, gdp, consumption 등 3개 변수가 있을 경우
  - names(sample1)[3]<-"cons" : 3번째 변수인 consumption을 cons로 변경
  - names(sample1)<-c("T","Y","C") : 세 변수명을 모두 T, Y, C로 변경

b3-ch1-4-rev.R

```
library(openxlsx)
sample1<-read.xlsx("http://kanggc.iptime.org/book/data/sample1-n.xlsx")
year<-sample1$year
gdp<-sample1$gdp
consumption<-sample1$consumption
lgdp<-log(gdp)
lconsumption<-log(consumption)
lgdp; lconsumption
names(sample1)
sample1
names(sample1)[3]<-"cons"
sample1
names(sample1)<-c("T","Y","C")
sample1
```

2. 부분자료 추출

- sample1-n.xlsx 파일에서 2000년대(2000년-2009년) 또는 2010년대(2010년-2016년) 데이터 추출
  - 예를 들어 sample1_dat 데이터에 year, gdp, consumption 등 3개 변수가 있을 경우
  - data1<-sample1_dat[1:10,] : sample1_dat의 10번째 행(2000년-2009년)까지 데이터 추출
  - data2<-sample1_dat[11:17,]: sample1_dat의 11번째부터 17번째 행(2010년-2017년)까지 데이터 추출

b3-ch1-5-rev.R

```
library(openxlsx)
sample1<-read.xlsx("http://kanggc.iptime.org/book/data/sample1-n.xlsx")
year<-sample1$year
gdp<-sample1$gdp
consumption<-sample1$consumption
data1<-sample1_dat[1:10,]
data1
data2<-sample1_dat[11:17,]
Data2
```