# 2강. R데이터처리(2)

◈ 담당교수 : 김성수 교수

## ■ 주요용어

용어	해설
변수이름 바꾸기	변수이름을 바꾸고자 할 때 R에서는 fix() 명령이나,
	names(), colnames() 를 이용한다. 이외에도 R
	패키지 reshape가 이용된다.
변수 값 라벨	s숫자로 입력된 변수 값을 라벨로 바꾸고자 할 때,
	명목형(nominal data)인 경우는 factor() 함수를, 순
	서형(ordered data)인 경우는 ordered() 함수를 이
	용하면 된다.
변수 값 변환	변수값을 변환하는 방법은 직접 자료를 변환하는 방법
	과 car(Companion to Applied Regression) 패키지
	의 recode( ) 함수를 이용하는 방법이 있다.
dplyr 패키지	dplyr 패키지는 데이터 처리를 위한 R 패키지이다. 데
	이터분석을 위한 사전처리 작업의 대부분은 Filtering
	(데이터 추출), Selecting cloumns (변수 선택),
	Adding new variables (변수 추가), Sorting (정렬),
	Aggregating (집계) 등으로 이루어져 있다. 이러한 작
	업을 쉽게 작업할 수 있는 기능을 제공하는 패키지가
	dplyr 이다.

### ■ 연습문제

- 1. 데이터개체 nwd 의 변수 x2가 0.11 보다 작은 경우를 99로 바꾸고, 이를 결측치로 처리하고자 한다. R 명령 (a)는?
- > nwd[nwd\$x2 < 0.11, "x2"] = 99

정답 및 해설: nwd[nwd == 99] = NA

2. 데이터개체 nwd 의 변수가 6개 있다. 이들 변수의 이름을 지정하고자 한다. 명령 (b)는 ?

```
> (b)(nwd) = c("a1", "a2", "a3", "a4", "a5", "newy")
```

정답 및 해설 : colnames

3. 데이터개체 insurance 의 변수 job에서 1 = 근로자, 2 = 사무직, 3 = 전문가로 변수 값 라벨을 주고자 한다. R 명령 (c)는 ?

> insurance\$job = (c)(insurance\$job, levels=c(1:3),

labels=c("근로자","사무직","전문가"))

정답 및 해설: factor

- 4. 변수 age의 값을 40이하 = 1, 40~60 = 2, 60~Hi = 3 으로 변환하고자 한다. R 명령 (d)는?
- > library(car)
- > drug\$agr2 = drug\$age
- > drug\$agr2 = (d)(drug\$age, "lo:40=1; 40:60=2; 60:hi=3")

정답 및 해설: recode

- 5. dsplyr 패키지의 chaining 기능을 이용하여 두 변수 UniqueCarrier, DepDelay 를 택하고, DepDelay > 60 인 케이스를 선택하고자 한다. 적합한 명령 (e)는 ?

정답 및 해설: %>%

#### ■ 참고자료

- 양우성 (2014), 데이터처리의 새로운 강좌, dplyr 패키지 https://wsyang.com/2014/02/introduction-to-dplyr/
- Hadley Wickham's dplyr tutorial at useR! 2014

http://datascience.la/hadley-wickhams-dplyr-tutorial-at-user-2014-part-1/

- Oscar T.-R. (2013), Introduction to R Studio
- Pruim, R. (2011), Computational Statistics Using R and R Studio
- W.N.Venables, D.M.Smith and the R Core Team(2014), An Introduction to R. Notes on R: A Programming Environment and
- Data Analysis and Graphics.
- R: https://www.r-project.org/
- R Studio: https://www.rstudio.com/
- dplyr package

https://cran.rstudio.com/web/packages/dplyr/vignettes/introduction.html

- dplyr 매뉴얼:
- dplyr 튜토리얼: <a href="https://rpubs.com/justmarkham/dplyr-tutorial">https://rpubs.com/justmarkham/dplyr-tutorial</a>

#### ■ 코스웨어

- R의 기본 코스웨어 학습방법 안내

다음은 R에 대한 기본적인 학습을 할 수 있는 방법을 안내 해 드립니다. 아직 R에 대하여 익숙하지 못한 학습자는 "R의 입문 코스웨어"를 통하여 R에 대하여 익숙해지기 바랍니다.

1) [학교 홈페이지에 접속] - [우측하단 "학과" 선택]



2) ["정보통계학과" 선택]



3) [상단메뉴 학습정보 내의 "학습자료실" 선택]



4) [좌측메뉴의 "코스웨어" 선택]



5) [좌측메뉴의 "R의 입문 코스웨어" 선택]



6) "R의 입문 코스웨어"는 총10장으로 구성됨



## 7) 각 장의 좌측메뉴의 학습목차를 참고하여 학습



"R의 입문 코스웨어" 과정을 통하여 R의 기본개념 및 구조를 이해하고 실제 작업에 필요한 자료처리 및 분석기법을 응용할 수 있는 능력을 학습하기 바랍니다.