빅데이터 R 분석 김경민

●해결문제

□ BMI 계산

- BMI는 몸무게와 키를 이용하여 체지방율을 측정하는 지수이다. 자신의 몸무게와 키를 입력 받아 각각 변수 weigh와 heigh에 저장하고 BMI지수를 계산해 본다. 단, 키는 cm로 입력 받아서 처리한다.
- BMI = 체중(kg) / (키(m) x키(m))

• 기본 입출력

- scan()
 - 키보드로부터 숫자 입력
- scan(what ="") , scan(what=character())
 - 키보드로부터 문자 입력
- □ readline("메시지")
 - 한 줄단위로 입력되며 문자열 입력
 - · 문자열 숫자변환 : as.numeric()
- print(변수 or 수식)
 - 변수나 수식 출력
- □ cat(문자열, 변수)
 - 문자열과 변수를 함께 출력

• 자료형

□스칼라(Scalar)

• 단일 차원의 값으로 길이가 1인 벡터와 같은 동일

□벡터(Vector)

- 한 가지의 스칼라 데이터타입의 데이터들을 저장
- 벡터에 저장된 요소들을 인덱스를 사용하여 접근
 - ◦인덱스는 1부터 시작
- 연속되거나 반복되는 벡터
 - oc(): 여러 개의 값을 하나로 결합
 - · seq(시작값, 끝값, 증가치) : 증가치는 생략 가능하며 시퀀스를 생성 , 시작값:끝값만 써도 동일
 - ∘ rep(반복할 값이 저장된 벡터, 전체 반복횟수, each=개별 반복횟수) : 주어진 값을 반복하는 벡터를 생성

• 자료형

□ 리스트(list)

- · 성격이 다른 자료형(문자열, 숫자형, 논리형 등)과 자료구조(벡터, 행렬, 리스트, 데이터프레임 등)를 객체로 생성
- 키(key)와 값(value) 쌍으로 저장

□ 데이터프레임(DataFrame)

- 테이블 구조와 유사하며 리스트와 벡터의 혼합형으로 열은 리스트 열내의 데이터는 벡터 자료구조
- 데이터프레임 생성함수 : data.frame(), read.table(), read.csv()

• 자료형 확인

□ 변수에 저장된 자료형 확인

• is.numeric(), is.integer(), is.logical(), is.double(), is.character(), is.complex(), is.data.frame(), is.factor(), is.na(), is.nan()

□ 자료형 변환

- as.numeric(), as.integer(), ax.logical(), as.double(), as.character(), as.data.frame(), as.factor(), as.list(), as.vector(), ax.array(), as.Data()
- 리스트를 벡터로 변환 : unlist()

┗ 자료형과 자료구조 확인

- mode(): 자료형 확인
- class() : 자료구조 확인

● 문자열 패키지: stringr

┗ 설치 및 사용

- install.packages("stringr")
- library(stringr)

stringr 기본함수

- str_lengh() : 문자열 길이 반환
- str_c() : 문자열 연결
- str_sub() : 부분 문자열 추출
- str_split(): 문자열 분리
- str_replace() : 문자열 교체

edit() 함수를 이용한 입력

□ 데이터프레임 입력시 표 형식의 데이터 편집기 제공

파일	편집 도움말						
	varl	var2	var3	var4	var5	var6	var7
1	170	60					
2	160	70					
3	180	80					
4							
5							
6							
7							
8							
9							
10							
11							
12							
13							
14							
15							
16							
17							
18							
19							

●해결문제

□국민건강보험공단 자료를 이용하여 BMI와 비만도를 구하시오.

- 비만도
 - ◦저체중 20 미만
 - •정상 20 24
 - ∘ 과체중 25 29
 - ◦비만 30 이상

```
dfbmi <- read.csv("./data/01_국민건강보험공단500.csv",
header = T,
stringsAsFactors = F,
fileEncoding = 'euc-kr')
```

· 성별 비만도 교차 빈도표를 구하고 csv파일로 저장

• 해결문제

□ Ifelse()함수

- 벡터단위로 연산이 필요할 때 사용
- ifelse (조건, 조건을 만족했을시 값, 조건을 불만족 했을 시 값)

□ CSV 파일 읽어오기

- read.csv(파일명, [,sep=","][,header=T][stringsAsFactors=FALSE])
 - stringsAsFactors=FALSE : 문자열로 읽어오기

┗ CSV 파일 저장하기

- write.csv(데이터프레임, 저장파일명[, quote = FALSE])
 - quote = FALSE : " 생략

table()

• 기준이 될 데이터에 대해 빈도표(도수 분포표)나 교차 빈도표

●해결문제

- □ 국민건강보험 관리공단의 건강검진 자료를 이용하여 대사증후군을 판별하시오.
 - · 높은 혈압(130/85mmHg 이상)
 - · 높은 혈당(공복 혈당 100mg/dL 이상)
 - · 높은 중성지방(트리글리세라이드 150mg/dL 이상)
 - · 낮은 HDL 콜레스테롤 수치(남성은 40mg/dL 미만, 여성은 50mg/dL 미만)
 - · 복부 비만(남성 90cm 이상, 여성 85cm 이상)
 - 판별
 - 0 : 정상, 1~2 : 주의군, 3~5 : 위험군

• 해결문제

