

## 분석프로그래밍 I.03

### Vector, Matrix, Array, List, Data frame

국민대학교 경영학부 빅데이터경영통계학과

2018. 3. 19(월)

지난 시간에 R의 여러 데이터 타입을 살펴보았다(표 ??).

다음의 벡터 `vInteger`, `vNumeric`, `vCharacter`, `vLogical`, `vFactor`, `vDate`, `vPOSIXct`에 대해 `class`, `mode`, `attributes` 함수를 적용한 결과를 비교해보자 (표. ??).

```
1 vInteger <- c(0L, 2L, 5L, 4L)
2 vNumeric <- c(0L, 2L, 5L, 4L)+ 0.001
3 vCharacter <- c("abc","def","ghijklmn","op")
4 vLogical <- c(T, T, F, T, F)
5
6 vFactor <- factor(c("high","low","high","high"))
7 vDate <- c(as.Date(c("1990-1-21","2018-10-2")), Sys.Date())
```

Tabelle 1: R의 데이터 타입

데이터 타입	예
숫자(numeric)	0, 1, -1, 0.45, 3L, 1e-7, 0xFF, pi, exp(1)
문자(character)	"Letter", '1', "Hello", says he', "Hello?", 'Cheer up!\r\nRight Now!'
범주(factor)	factor(c('high','low','high'))
논리(logical)	TRUE, T, FALSE, F
날짜(Date)	as.Date("2018-03-12"), as.Date("2018/03/12")
날짜/시간(POSIXct)	as.POSIXct('2018/03/12 10:33')

Tabelle 2: class, mode, attributes 결과

데이터 타입	class	mode	attributes
숫자(numeric)			
문자(character)			
범주(factor)			
논리(logical)			
날짜(Date)			
날짜/시간(POSIXct)			

```

8 vPOSIXct <- c(as.POSIXct(c('2018-03-12 10:33','2017-03-11 10:33')), Sys
  .time())
9
10 class(vInteger)
11 mode(vInteger)
12 attributes(vInteger)

```

숫자(numeric), 문자(character), 논리(logical)의 데이터 타입은 기본적인 데이터 타입이라고 할 수 있고, 범주(factor), 날짜(Date, POSIXct)는 기본적인 데이터 타입을 활용하여 다른 의미를 부여한 것으로 생각할 수 있다.

## 1 매트릭스, 배열, 리스트, 데이터 프레임

CRUD는 컴퓨터가 수행하는 기본적인 데이터 처리 기능인 Create(생성), Read(읽기), Update(갱신), Delete(삭제)을 일컫는다. BREAD(Browse, Read, Edit, Add, Delete)를 얘기하기도 한다(표 ??).

Tabelle 3: 데이터 구조에 따른 생성, 참조, 수정, 추가, 삭제 함수

클래스	matrix	array	data.frame	list
생성(Create)	matrix(vec, nc, nr)	array(vec,dim=c(d1,d2,...))	data.frame(col1=, col2=, df)	list(l1=l2=,...)
생성시 변형	Date, POSIXct	는 numeric으로, factor는 character로	stringsAsFactors=T	
참조(Read)	mat[,]	arr[,,...]	df[,]	lst[c(2,4)]
참조시 변형	mat[1,,drop=T]	arr[, ,drop=T]	df[,2,drop=T]	lst[[2]]
수정(Edit)	mat[,]=	arr[,,...]=	df[,]=	lst[c(2,4)] = list(, lst[[2]] = modifyList(lst, lst2)
수정시 변형	표현력이 큰 데이터 타입으로 변형된다 (numeric<character)		데이터 타입이 일치해야 한다	
추가(Add)	rbind(mat, c(,)) cbind(mat, c(,))	abind::abind(a1, a2, along= ) rbind(df, c(,)) cbind(df, c(,))		append(lst, list(,)) append(list(,), lst) lst[ ]=, lst[[ ]]=
추가시 변형	표현력이 큰 데이터 타입으로 변형된다 (numeric<character)		rbind는 데이터 타입이 일치해야 한다	
삭제(Delete)	mat[-2,]	arr[-3,-2,,]	df [-2,-3]	lst[[2]] = NULL lst[2] = NULL
삭제시 변형			‘참조시 변형’ 참조	
이름 정보	rownames, colnames, dimnames	dimnames	names==colnames, rownames	names
구조 정보	nrow, ncol, dim	dim	nrow, ncol=length	length